JOURNAL OF COMPUTATIONAL PHYSICS 67, 439–471 (1986)

A Numerical Model of the Balance Equations in a Periodic Domain and an Example of Balanced Turbulence

N. J. NORTON, J. C. MCWILLIAMS, AND P. R. GENT

National Center for Atmospheric Research, P.O. Box 3000, Boulder, Colorado 80307

Received September 20, 1985; revised March 5, 1986

The Balance Equations represent rotating, stratified fluid motions and therefore are of potential utility in studies of geophysical phenomena. In this paper we present numerical algorithms for this model. The most novel aspect resolves the implicit structure of the time-derivative terms in the Balance Equations with iterative, predictor-corrector methods which combine time-stepping and iteration. The discrete spatial operations are low-order, finite-differences which exactly conserve certain integral properties of the model. The numerical performance of the model is evaluated, and comparisons with other geophysical models are made—the more highly approximated Quasigeostrophic Equations (QG) and the more fundamental Primitive Equations (PE). We find that the Balance Equations are an excellent Intermediate Model: their computational cost is closer to that of the less expensive QG, but they quite accurately represent the evolution of the vorticity field of the more accurate PE, even when fluid advection rates are comparable to the Coriolis frequency (i.e., the Rossby number is order one) and a substantial gravity wave component is generated in the PE. In addition, the existence of a solvability barrier at large Rossby number is demonstrated. As an illustration of a physical phenomenon for which the Balance Equations are well suited, solutions are examined for the slow viscous decay of broad-band, large-scale, baroclinic initial conditions at finite Rossby number (i.e., balanced turbulence). Several departures from the more familiar geostrophic turbulence are found: the enstrophy cascade is stronger, the baroclinic energy conversion is stronger, and there is a tendency to form isolated frontal structures in the vorticity field.

1. INTRODUCTION

Large-scale fluid motions in the oceans and atmosphere are significantly influenced by the earth's rotation. This regime is characterized by small or moderate Rossby number,

$$R \equiv V/fL. \tag{1}$$

Here V is a characteristic horizontal velocity scale, f a Coriolis frequency, and L a horizontal length. The Quasigeostrophic Equations (QG) are derived as the leading-order asymptotic expansion in R of the more fundamental fluid equations, and during the past three decades Quasigeostrophy has been found to be a

0021-9991/86 $3.00

qualitatively (and sometimes quantitatively) useful model. Not only does it yield apt analogs to observed large-scale phenomena, but it also excludes other phenomena, such as inertia-gravity waves, which are additional solutions to the most common of the more fundamental models (i.e., the Primitive Equations (PE); see the Appendix), but which also appear to be nearly extraneous for large-scale flows. The presence of these extraneous components to model solutions imposes burdens both of interpretation (greater complexity) and of computation (faster physical time scales, shorter than $1/f$, which must be resolved).

The term Intermediate Models denotes the general class of models which are more accurate than Quasigeostrophy at finite, but not large, R, yet also exclude the extraneous phenomena of the Primitive Equations. A review of this class is presented in [13], and a comparison of their solutions for a wave-triad evolution is made in [4]. In general, however, too few numerical realizations of Intermediate Models have been made, hence few are yet available to adequately assess their relative accuracy and utility.

One of the most attractive of the Intermediate Models is the Balance Equations (BE) in the form first proposed by Lorenz in [11]. In addition to superior accuracy in the wave-triad solution comparison, its virtues include (1) second-order accuracy in small R[1]; (2) conservation, in the absence of dissipation, of volume integrals of energy and all powers of the potential temperature (or potential density if the fluid has a multi-component equation of state, as in seawater); and (3) the minimum number of terms retained in the truncation of the PE consistent with these properties. Several variants of the BE have been formulated, with different degrees of accuracy in the spatial variation of the Coriolis frequency and different independent vertical coordinates, physical height (pressure in a compressible fluid), or potential temperature. More information on the history and formal properties of the BE is given in [5–7].

In this paper, we describe the design and computational performance of a numerical model of a particular variant of the BE (designated βBE in [5]) in order to be able to explore its value as an Intermediate Model of large-scale geophysical phenomena. This particular choice is made for its compatibility between the geophysically important gradient of Coriolis frequency and horizontally periodic boundary conditions (the latter are simpler to implement than solid boundary conditions).

Intermediate Models in general, and the Balance Equations in particular, have a mathematical structure which requires novel computational algorithms (i.e., a partial differential equation system with implicitly defined time derivatives). The algorithms which we report are relatively simple ones, yet their performance appears to be satisfactory in accuracy and efficiency for the regime with moderate R values.

Since the model is too complex for rigorous and complete theoretical analysis, we demonstrate the numerical performance of the BE with an initial-value problem for

[1] The Balance Equations are also formally valid for strongly stratified flow even if R is large [12].

a random realization of a broad-band spectrum of vorticity. At finite R the solutions, which we call balanced turbulence, exhibit substantial departures from geostrophic turbulence ($R=0$). Both the numerics and physics are also examined in relation to the QG and PE models. Analogous studies of physical behavior for the evolution of balanced isolated vortices, rather than balanced turbulence, are presented in companion papers [14, 15].

The order of the paper is as follows: the BE are defined in Section 2, numerical methods are given in Section 3, the initial-value problem is posed in Section 4, numerical performance is evaluated in Section 5, and physical behavior is analyzed in Section 6.

2. Balance Equations

We shall consider a non-dimensional form of the BE with horizontal velocities scaled by V, horizontal coordinates by L, vertical coordinate z by H, and time coordinate t by L/V. Other quantities are scaled compatibly [14]. There are several mathematically equivalent alternatives in the formulation of the model [5]. We have chosen one in which the primary dependent variables of the model are the streamfunction, ψ; the cumulative divergence potential, X; the deviation in potential temperature about a horizontal average, θ; and the horizontally averaged static stability profile, $S(z, t)$. The corresponding governing equations are (5)–(9) below. S is proportional to the vertical derivative of the horizontally averaged potential temperature,

$$S = B\Theta_z, \tag{2}$$

where B is the Burger number, equal to $(NH/fL)^2$, with N a characteristic Brunt–Vaisala frequency. Θ is scaled relative to θ such that the total potential temperature is $BR^{-1}\Theta + \theta$. ψ and X are related to east and north horizontal velocities by

$$u = -\psi_y - RX_{xz}, \qquad v = \psi_x - RX_{yz}; \tag{3}$$

the geopotential field ϕ is calculated from hydrostatic balance,

$$\phi_z = \theta; \tag{4}$$

and the vertical velocity w is calculated from mass continuity,

$$w = \Delta X. \tag{5}$$

Here x and y are east and north coordinates, coordinate subscripts denote partial derivatives, and Δ is the horizontal Laplacian operator. The primary equations of the model are, respectively, the vorticity, thermal balance, omega, and stability equations [5]:

$$\Delta\psi_t = -J(\psi, \Delta\psi) - \beta\psi_x + w_z + R\nabla\cdot(\Delta\psi\nabla X_z - w\nabla\psi_z) + D, \tag{6}$$

$$\Delta\theta = \Delta\psi_z + 2RJ(\psi_x, \psi_y)_z, \tag{7}$$

$$\begin{aligned} w_{zz} + S\Delta w = {} & J(\psi, \Delta\psi)_z + \beta\psi_{xz} - \Delta J(\psi, \theta) + R\nabla\cdot(w\nabla\psi_z - \Delta\psi\nabla X_z)_z \\ & - 2RJ(\psi_x, \psi_y)_{zt} + R\Delta(\nabla\cdot(\theta\nabla X_z) - (\theta w)_z) - D_z, \end{aligned} \tag{8}$$

$$S_t = -R^2\langle\theta w\rangle_{zz}, \tag{9}$$

where J is the horizontal Jacobian operator, ∇ is the horizontal gradient operator, β is the non-dimensional Coriolis gradient (rescaled by V/L^2), and $\langle\cdot\rangle$ denotes a horizontal average. D is the curl of the horizontal vector representing momentum sources or dissipation (rescaled by V^2/L^2); e.g., the Newtonian dissipation form for D would be

$$D = \nu\Delta\Delta\psi, \tag{10}$$

where ν is the non-dimensional viscosity, and a common alternative, which we use here, is

$$D = -\nu\Delta\Delta\Delta\psi, \tag{11}$$

with ν referred to as hyperviscosity. The model is assumed to be adiabatic. More general forms for D and diabatic heating could easily be incorporated, but this will not be done here. The dimensionless parameters β and R are assumed to be small since terms of order βR and R^2 are neglected in deriving this model. All of the preceding fields except S and Θ have zero horizontal average in a periodic domain.

Lateral boundary conditions are periodicity in x and y over a distance W. The vertical boundary condition is no normal flow; for plane boundaries this implies

$$w = X = 0 \qquad \text{at} \quad z = 0, 1. \tag{12}$$

In addition, θ is conserved under horizontal advection on the vertical boundaries. A simpler form of this condition is that these boundaries are isothermal. If this is true initially, it will remain so by the more general condition; thus, we shall adopt

$$\theta = 0 \qquad \text{at} \quad z = 0, 1. \tag{13}$$

An appropriate initial condition is the specification of $\psi(x, y, z, 0)$ and $S(z, 0)$. This is sufficient to allow the time integration of the model schematically as follows: first, θ is calculated from (7); next ψ_t and X (or w) are calculated from the coupled equations (5), (6), and (8); finally, S_t is calculated from (9). Using ψ_t and S_t the solution can be advanced in time, after which the cycle can be repeated.[2] The key

[2] This is Method 1 discussed in [5].

difficulty in integrating the BE, which is unusual in geophysical fluid computations, is the implicit definition of the time-derivative terms in the middle step above: no explicit expressions can be written for them without involving other quantities which themselves are unknown until the time derivatives are known. Thus, coupled, non-linear equations must be solved simultaneously, in general by iterative techniques (see Section 3.A).

Finally, we record the integral conservation laws for the model. They provide important constraints upon the choice of the algorithms, allow tests of the correctness of the model construction (augmenting known analytic solution tests), and provide useful measures of its computational accuracy. The integral quantities are the kinetic and potential energies and moments of the potential temperature field:

$$K \equiv \frac{1}{2} \int dz \langle \nabla\psi \cdot \nabla\psi \rangle, \tag{14}$$

$$P \equiv BR^{-2} \int dz(-z\Theta), \tag{15}$$

$$M^{(n)} \equiv \int dz \langle BR^{-1}\Theta + \theta)^n \rangle, \tag{16}$$

for any positive integer n. The corresponding conservation laws are

$$K_t = \int dz \langle \theta w \rangle - \int dz \langle \psi D \rangle, \tag{17}$$

$$P_t = -\int dz \langle \theta w \rangle, \tag{18}$$

$$M_t^{(n)} = 0. \tag{19}$$

3. Numerical Model

Several general considerations have guided the development of our numerical model of the BE, (5)–(9). First, and most important, the implicit time-derivative structure of the BE system requires the use of iterative methods in its numerical solution. However, for a system as complicated as the BE, we have no a priori knowledge of the convergence properties of any such methods. We therefore considered several combined time-stepping and iteration methods (which we shall refer to simply as time methods), in hopes that at least one would prove successful. Preliminary testing on idealized ODE systems provided the basis for narrowing our choice of time methods to two, both of which were implemented and tested in the BE.

Second, geophysical fluid motions frequently are spatially complex. Their numerical calculation often requires high spatial resolution, which puts a premium on computer storage for in-core calculations. In particular, on a Cray-1 computer, only low-order time methods are practical, since they require the retention of very few time levels of information.

Third, experience with geophysical computations suggests that it is advantageous to choose spatial discretizations which conserve certain integral properties of the continuous equations (in the limit of vanishing time step). For the BE we selected methods which conserve $K+P$ and $M^{(n)}$ (for $n=1$ and 2) when $D=0$ (see (17)–(19)). These are the same integrals Lorenz's scheme [11] conserves, and consequently our vertical discretization formulae are similar to his (see Section 3.B).

To illustrate the time methods, we symbolically represent the BE as

$$\mathscr{L}_1\psi_t = \mathscr{N}_1[\psi, X] \tag{20}$$

$$\mathscr{L}_2 X = \mathscr{N}_2[\psi, X, \psi_t], \tag{21}$$

where the $\mathscr{L}$'s are easily invertible, spatial, linear differential operators, and the $\mathscr{N}$'s are non-invertible, spatial, non-linear differential operators. In forming (20)–(21), w has been replaced by $\mathscr{L}_1 X$ from (5), and θ has been replaced by a non-linear functional of ψ from (7). Equation (9) becomes

$$S_t = \mathscr{N}_3[\psi, X], \tag{22}$$

which is evaluated after (20)–(21) have been solved.

The remainder of this section is split into two parts: in the first we discuss the two time methods implemented in the BE model, and in the second we discuss the spatially discrete forms of the operators $\mathscr{L}$ and $\mathscr{N}$.

A. *Time Methods*

We further simplify the notation of the model system (20)–(21) by replacing $\mathscr{L}_1^{-1}\mathscr{N}_1$ by $\mathscr{F}$ and $\mathscr{L}_2^{-1}\mathscr{N}_2$ by $\mathscr{G}$:

$$\psi_t = \mathscr{F}(\psi, X) \tag{23}$$

$$X = \mathscr{G}(\psi, X, \psi_t). \tag{24}$$

A general time method algorithm for (23)–(24) is

Given ψ^N, ψ_t^N, and X^N at $t_{n-1}, t_{n-2},\ldots$:

1. PREDICT $\psi^0(t_n)$ and $X^0(t_n)$
2. EVALUATE $\psi_t^0(t_n) = \mathscr{F}(\psi^0(t_n), X^0(t_n))$
3. ITERATE N times to obtain converged values of $\psi^N(t_n)$, $\psi_t^N(t_n)$, and $X^N(t_n)$ using Algorithm M1 or M2 (defined below). (25)

The superscripts 0 and N represent the starting and final values of the iterates. We choose to fix N rather than perform costly iteration convergence testing at each time step; the success of this approach is shown in Section 5. Model times are

$$t_n = t_0 + n\,\Delta t, \qquad n = 1, 2, \ldots, \tag{26}$$

where t_0 is the initial time and Δt is the time-step size. We defer the discussion of (25) for the start-up times until after our discussion of the specific time methods M1 and M2.

Although any of several methods could be used to predict ψ^0 and X^0 in (25.1), computer-storage considerations led us to choose second-order methods. For t_n, $n \geqslant 2$, we use

$$X^0(t_n) = 2X(t_{n-1}) - X(t_{n-2}) \tag{27}$$

and

$$\psi^0(t_n) = \psi(t_{n-2}) + 2\Delta t\,\psi_t(t_{n-1}). \tag{28}$$

We propose two methods to be used in (25.3). In the first algorithm, M1, the value of ψ remains fixed throughout an iteration cycle and then is corrected outside this cycle using the resultant ψ_t, thereby partly separating the iterative solution for the implicit time derivative from the time-stepping process. In the second algorithm, M2, ψ is updated as soon as a new iterate for ψ_t is available, thus combining iteration and time-stepping.[3]

Algorithm M1

3. REPEAT for $\lambda = 1, 2, \ldots, L$:
    1. REPEAT for $\nu = 1, 2, \ldots, N(\lambda)$:
        1. EVALUATE $X^\nu \leftarrow \mathscr{G}(\psi^{\lambda-1}, X^{\nu-1}, \psi_t^{\nu-1})$
        2. EVALUATE $\psi_t^\nu \leftarrow \mathscr{F}(\psi^{\lambda-1}, X^\nu)$
    2. CORRECT $\psi^\lambda \leftarrow \mathscr{H}(\psi_t^{N(\lambda)}, \psi_t(t_{n-1}), \psi_t(t_{n-2}), \ldots, \psi(t_{n-1}), \ldots)$
    3. REPLACE $X^0 \leftarrow X^{N(\lambda)}$; $\psi_t^0 \leftarrow \psi_t^{N(\lambda)}$. (29)

Algorithm M2

3. REPEAT for $\nu = 1, 2, \ldots, N$:
    1. EVALUATE $X^\nu \leftarrow \mathscr{G}(\psi^{\nu-1}, X^{\nu-1}, \psi_t^{\nu-1})$
    2. CORRECT $\psi^\nu \leftarrow \mathscr{H}(\psi_t^{\nu-1}, \psi_t(t_{n-1}), \psi_t(t_{n-2}), \ldots, \psi(t_{n-1}), \ldots)$
    3. EVALUATE $\psi_t^\nu \leftarrow \mathscr{F}(\psi^\nu, X^\nu)$. (30)

Within the iteration cycles in Algorithms M1 and M2, all old iterates are overwritten (stored in the same location) by new iterates as soon as they become available; we write "a overwrites b" as $b \leftarrow a$. $\mathscr{H}$ is a linear algebraic time-stepping operator. Although an appropriate method of any order in Δt might be used in the

[3] A variant of M2, where steps (30.3.1) and (30.3.3) were interchanged, was also implemented and tested. In all but a very few cases, performance of the variant was inferior to that of M2.

correction steps (29.3.2) and (30.3.2), we have chosen the following second-order method:

$$\mathscr{H}(\psi_t(t_n), \psi_t(t_{n-1}), \psi(t_{n-1})) \equiv \psi(t_{n-1}) + \frac{\Delta t}{2} [\psi_t(t_n) + \psi_t(t_{n-1})], \tag{31}$$

where $\psi_t\ (t_n)$ is the current estimate as designated in (29) or (30). The integers L, $N(\lambda)$, and N regulate the iteration cycles and their appropriate values may be solution-dependent. Sensitivity to the choice of these parameters and comparison of M1 and M2 solutions are examined in Section 5. Because we are unable to state appropriate functional properties of $\mathscr{F}$ and $\mathscr{G}$, such as Lipschitz conditions, we have no a priori assurances of convergence of the iterations in M1 or M2, but several a posteriori demonstrations.

Algorithm (25) must be modified for the start-up times $t_0, t_1, ..., t_r$, where r is one less than the order of the predictor in (25.1). For compatibility with (27)–(28), $r = 1$. At t_0, the initial value problem is defined by specifying ψ, and $X(t_0)$ and $\psi_t(t_0)$ are obtained as follows:

1. SET $X^0 = 0$, $\psi_t^0 = 0$
2. REPEAT for $\nu = 1, 2, ..., N_0$:
   1. $X^\nu \leftarrow \mathscr{G}(\psi, X^{\nu-1}, \psi_t^{\nu-1})$
   2. $\psi_t^\nu \leftarrow \mathscr{F}(\psi, X^\nu)$. (32)

Note that ψ remains fixed throughout the iteration cycle. Because (32.1) contains poor starting values compared to (27)–(28), we have sometimes found it necessary to choose $N_0 \gg N$ in order to obtain well-converged solutions.

At t_1, we replace (27)–(28) by

$$X^0(t_1) = X(t_0) \qquad \text{and} \qquad \psi^0(t_1) = \psi(t_0) + \Delta t\, \psi_t(t_0), \tag{33}$$

and otherwise apply (25).

Finally, we have found that modification of the iterations in M1 and M2 to include relaxation improves their convergence properties under some circumstances (e.g., when R is large; see Section 5). This is accomplished by adding the step

$$X^\nu \leftarrow \alpha X^{\nu-1} + (1-\alpha) X^\nu, \tag{34}$$

with a similar expression for ψ_t^ν after the ψ_t evaluation in (29) or (30).

The algorithms M1 and M2 are extensions of familiar predictor-corrector schemes to implicit time-derivative systems. In the simpler case of explicit systems, they reduce to $\mathrm{p(ec)}^L$ and $\mathrm{pe(ce)}^N$, respectively. Here p denotes PREDICT a new field, e denotes EVALUATE a new time derivative, and c denotes CORRECT the new field. Integer exponents indicate repetition of the operation sequence in parentheses. We use these reduced forms of M1 and M2 to integrate the explicit QG and PE models.

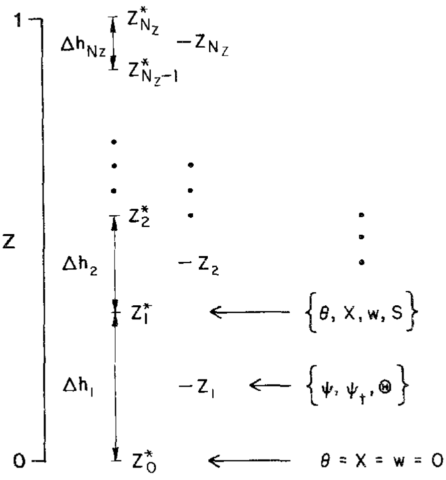

FIG. 1. Model vertical grid.

B. *Spatial Discretization*

We select discrete forms of the spatial differential operators in (5)–(11) with the following properties: they are finite-difference with quadratic errors in the horizontal grid spacing and linear errors in the vertical[4]; they can be manipulated through discrete operations so that discrete forms of the conservation laws (17)–(19), with $n = 1$ and 2, hold exactly except for time-method errors; and they have a relatively narrow reach across adjacent grid points. These properties do not uniquely determine the discrete forms, but they do strongly constrain them.

The horizontal grid is uniform with spacing Δs. The vertical grid is staggered and non-uniform (Fig. 1). We partition the vertical domain, $0 \leqslant z \leqslant 1$, into N_z intervals $\{\Delta h_i,\ i = 1,..., N_z\}$. We define two sets of grid coordinates, $\{z_i^*,\ i = 0,..., N_z\}$ and $\{z_i,\ i = 1,..., N_z\}$, as follows:

$$\left.\begin{aligned} z_0^* &= 0 \\ z_i^* &= z_{i-1}^* + \Delta h_i \\ z_i &= z_{i-1}^* + \tfrac{1}{2}\Delta h_i \end{aligned}\right\} \qquad i = 1,..., N_z. \tag{35}$$

The dependent variables θ, X, w, and S are defined on $\{z_i^*\}$, and ψ, ψ_t, and Θ are defined on $\{z_i\}$. Vertical boundary conditions (12)–(13) become

$$\theta = X = w = 0 \qquad \text{at} \quad z = z_0^*,\ z_{N_z}^*. \tag{36}$$

This grid is similar to that of Lozenz [11], except that we have partitioned the

[4] It is common practice in rotating fluid computations to use less resolution and accuracy in vertical operations than in horizontal ones. Note that the errors in our vertical operators improve to quadratic when the vertical grid is uniform.

potential temperature into average and deviation components and located them on different grids. The criterion for locating a variable on a particular grid is to achieve the desired narrow reach.

The vertical derivatives in (5)–(11) can be constructed from two primary operators, differencing and averaging:

$$V[q]=\frac{q(z_i^*)-q(z_{i-1}^*)}{\Delta h_i}, \qquad V^*[q]=\frac{q(z_{i+1})-q(z_i)}{z_{i+1}-z_i}$$
$$A[q]=\tfrac{1}{2}(q(z_i^*)+q(z_{i-1}^*)), \qquad A^*[q]=\tfrac{1}{2}(q(z_{i+1})+q(z_i)). \tag{37}$$

Operators with a superscript asterisk yield a product at level z_i^*, and ones without an asterisk yield a product at level z_i. With appropriate combinations of (37), the governing equations can be written so as to require variables only on their defining grids. Equations (5) and (10)–(11) conform to this since only fields on the same vertical grid level are involved; (5) is evaluated at $\{z_i^*\}$ and (10)–(11) at $\{z_i\}$. The remaining equations, (6)–(9), conform in a non-trivial way when written as follows:

$$\Delta\psi_t=-J(\psi,\Delta\psi)-\beta\psi_x+V[w]+R\nabla\cdot(\Delta\psi V[\nabla X]$$
$$-A[wV^*[\nabla\psi]])+D, \tag{38}$$

$$\Delta\theta=V^*[\Delta\psi]+2RV^*[J(\psi_x,\psi_y)], \tag{39}$$

$$V^*[V[w]]+S\,\Delta w=V^*[J(\psi,\Delta\psi)+\beta\psi_x]-\Delta J(A^*[\psi],\theta)$$
$$+R\nabla\cdot(V^*[A[wV^*[\nabla\psi]]-\Delta\psi\ V[\nabla X]])$$
$$-2RV^*[J(\psi_x,\psi_y)_t]+R\Delta(\nabla\cdot(\theta V^*[A[\nabla X]])$$
$$-V^*[A[\theta]\,A[w]])-V^*[D], \tag{40}$$

$$S_t=-R^2\langle V^*[V[\theta w]]\rangle. \tag{41}$$

Equation (38) is evaluated at $\{z_i\}$, and (39)–(41) at $\{z_i^*, i=1,..., N_z-1\}$.

For the horizontal operators in (38)–(41), we choose the following discrete forms. Δ is represented as a five-point cross,

$$\Delta q\rightarrow(q_{m+1,n}+q_{m-1,n}+q_{m,n+1}+q_{m,n-1}-4q_{m,n})/(\Delta s)^2, \tag{42}$$

where m and n are horizontal grid indices and $\rightarrow$ denotes the discrete definition of the left-hand-side quantity. The horizontal derivative in the terms with coefficient β is

$$q_x\rightarrow(q_{m+1,n}-q_{m-1,n})/(2\Delta s). \tag{43}$$

The form $\nabla\cdot(a\nabla b)$ is evaluated from the identity

$$\nabla\cdot(a\nabla b)=\tfrac{1}{2}(\Delta ab-b\Delta a+a\Delta b), \tag{44}$$

with Δ from (42). The Jacobian operator is evaluated by the formula of Arakawa [1] in all cases except where its arguments are ψ_x and ψ_y; in the latter cases, it is evaluated from the identity

$$J(\psi_x, \psi_y) = \tfrac{1}{2}\nabla \cdot (\Delta\psi\, \nabla\psi) + \tfrac{1}{4}\Delta(\psi\, \Delta\psi) - \tfrac{1}{8}\Delta\Delta(\psi^2), \tag{45}$$

with application of (42) and (44).

Finally, the conservation laws (17)–(19) can be derived from the above with the following discrete forms:

$$\begin{gathered}
K \to \frac{1}{2}\sum_{z_i} \Delta h_i \langle (\nabla\psi)^2 \rangle, \qquad P \to BR^{-2}\sum_{z_i} \Delta h_i(-z_i\Theta), \\
M^{(1)} \to \sum_{z_i} \Delta h_i \Theta, \qquad M^{(2)} \to B^2R^{-2}\sum_{z_i} \Delta h_i \Theta^2 + \sum_{z_i^*} (z_{i+1} - z_i)\langle \theta^2 \rangle, \\
\int dz \langle w\theta \rangle \to \sum_{z_i^*} (z_{i+1} - z_i)\langle w\theta \rangle, \qquad \int dz \langle \psi D \rangle \to \sum_{z_i} \Delta h_i \langle \psi D \rangle.
\end{gathered} \tag{46}$$

$\langle \cdot \rangle$ is an average over m and n, and

$$\begin{aligned}
(\nabla\psi)^2 \to \tfrac{1}{2}((\psi_{m+1,n} - \psi_{m,n})^2 &+ (\psi_{m-1,n} - \psi_{m,n})^2 \\
&+ (\psi_{m,n+1} - \psi_{m,n})^2 + (\psi_{m,n-1} - \psi_{m,n})^2)/(\Delta s)^2.
\end{aligned} \tag{47}$$

The quantity $\langle \psi D \rangle$ can be shown to be positive definite from (42) and (10) or (11).

4. Posing a Numerical Initial-Value Problem

We select a particular initial-value problem in order to illustrate the numerical and physical behavior of the BE. It is the turbulent evolution and slow viscous decay of a band-limited white vorticity spectrum. The initial spectrum is centered about the baroclinic deformation radius (defined below). We choose a low-order truncation in the vertical, $N_z = 2$. The gradient of the Coriolis frequency is taken to be only moderately small, viz., $\beta = 0.5$. Thus, the illustrative problem is generic, with most physical parameters of order unity (ν is an exception) and general initial conditions.

The vertical domain is partitioned unequally in imitation of the shallow depth of a typical oceanic pycnocline: we choose $\Delta h_1 = 5/6$ and $\Delta h_2 = 1/6$. Also, for the Burger number $B = 1$, $\Theta_1(t=0) = 1.0$, and $\Theta_2(t=0) = 8.2$, the eigensolutions of the vertical structure equation,

$$\frac{d}{dz}\left(S^{-1}\frac{dF}{dz}\right)+\lambda^2 F=0, \qquad 0<z<1,$$
$$\frac{dF}{dz}=0, \qquad z=0,1, \tag{48}$$

are

$$F_1^{(1)}=F_2^{(1)}=1, \qquad \lambda^{(1)}=0;$$
$$F_1^{(2)}=-1/\sqrt{5}, \qquad F_2^{(2)}=\sqrt{5}, \qquad \lambda^{(2)}=1. \tag{49}$$

Subscripts refer to z_i level and parenthetical superscripts to mode index. Mode 1 is called barotropic or external, and mode 2 is called first baroclinic or internal. The deformation radius is λ^{-1}.

The initial vorticity field is evaluated from

$$\zeta_i(x,y)\to\gamma\cdot\sum_{j=1}^{N_z}\sum_{k,l=1}^{4}F_i^{(j)}$$
$$\times\{a_{jkl}\cos 2\pi[(kx+ly)/W+\xi_{jkl}]+b_{jkl}\cos 2\pi[(kx-ly)/W+\eta_{jkl}]\}, \tag{50}$$

where the random variables (a, b) are from a normal distribution and (ξ, η) are from a uniform distribution over a unit interval. The constant γ is used to enforce the normalization condition

$$\int_0^1 dz\langle\zeta^2\rangle\to\sum_{i=1}^{N_z}\Delta h_i\langle\zeta_i^2\rangle=1. \tag{51}$$

W has the value 4π; hence the range of horizontal scales in (50) is between 0.5 and 2.0 in each coordinate. Of course, solutions vary with the particular realization for the random numbers. We shall show results from only a single realization, but we have checked that our conclusions are qualitatively valid over many realizations.

All solutions in Sections 5 and 6 have the same initial vorticity field (50) and the same $\Theta_i(0)$. Initial values for other dependent variables are calculated as described in Sections 2 and 3 and the Appendix. The parameters specified above are fixed, while others will vary among solutions (in particular, the principal physical parameter R). The model integration time, $T=10$, exceeds the times of strongest adjustment in spectrum shape, turbulent cascade rate, and energy dissipation rate (see Section 6).

For a given Δs, the time step Δt is chosen small enough for accuracy and computational stability, and the hyperviscosity ν is chosen just large enough for stability (see Section 5). For the study of numerical behavior, many different solutions are calculated and a larger $\Delta s=W/100$ is used for economy. Accompanying values are $\nu=2.5\times10^{-5}$ and $\Delta t\leqslant 1/12$. We fix R at an intermediate value of 0.25. In the study of physical behavior, however, we are most interested in a few solutions with the highest available resolution for different R values and models. Accordingly, we choose a smaller $\Delta s = W/144$, $\nu = 1.2 \times 10^{-6}$, and $\Delta t \leqslant 1/24$.

5. Numerical Properties

In this section we evaluate the time methods M1 and M2, compare the BE model with the PE and QG models, and explore a BE solvability barrier for large values of R. All method evaluations and model intercomparisons are based on the size of a global error as a function of the cost of calculating the solution.

Several measures of numerical error are conceivable. Differences between an analytic solution of the continuous equations and a numerical solution would provide an absolute measure. However, no such solutions are known for the physical regime of interest. An alternative standard might be a particularly well resolved numerical solution, but we are unable to achieve a resolution greatly better than the minimum required for a physically credible solution. A disadvantage with any absolute standard in a regime of turbulence is the difficulty in distinguishing between cumulative numerical error and the effects of rapid divergence between solutions with initially small differences.

We therefore choose a relative measure of error rate based on a discretized form of the appropriate kinetic energy tendency equation, (17) or (A7). We make a centered, second-order finite-difference approximation to the time derivative, evaluate all other terms at the middle time, then difference the two quantities to obtain the residual, $\varepsilon(t)$. We sample ε at appropriately frequent fixed times over a period which includes the development of peak vorticity amplitudes in order to form a global error measure

$$E = \left[\frac{1}{M} \sum_{m=1}^{M} \varepsilon^2(t_m) \right]^{1/2}. \tag{52}$$

For our particular realization, choices of M and t_m which give sufficient resolution are $M = 10$ and $t_m = m/2$, $m = 1, 2, \ldots, M$. Note that the choice of fixed sampling times restricts the time step to the set of all Δt's of the form $1/(2j)$, for integer j.

We define the cost C of a solution in Cray-1 CPU seconds measured from the first to last time step (for $t \in (0, t_M]$; initialization costs are not included). All timing tests were run on a fully loaded, multiprogramming machine (i.e., a single, shared CPU), and hence measured costs are somewhat variable; however, costs for all repeated cases were within 0.5% of their original cost.

Note that on all cost-error figures presented in this section, the highest-cost cases represent an arbitrarily chosen small value of Δt, and the lowest-cost cases represent the largest value of Δt such that the solution is stable (see below). The largest-Δt solutions almost always have sufficiently small E values so that the important physical behavior is represented correctly. Both the energy imbalance estimator ε and our time methods are of second-order accuracy in small Δt (large C), and all numerical parameter sets tested have been carried to this asymptotic limit.

Our time methods are subject to various computational instabilities for inappropriate choices of numerical parameters. An important restriction, analogous to the familiar C.F.L. condition, is that Δt must be small enough to resolve the physical rates of advection, diffusion, and wave propagation (Rossby waves for all models with $\beta \neq 0$ and gravity waves for the PE model). This restriction determines the largest Δt values on the cost-error curves below. Another important restriction is that non-linear computational instability be prevented. This instability results from the accumulation of excess variance on small spatial scales in the solution, no matter how small Δt [8]. Commonly this is prevented either by conserving quadratic integral invariants, as is possible in the QG [10], or by choosing ν large enough to limit the small-scale variance. The latter can be accomplished with physically small values of ν in the PE, non-conserving QG, and the BE at small and moderate R values. However, for large R the ν required to eliminate the BE model solvability barrier is so large as to grossly distort the physical character of the problem (see Section 5.C).

A numerical solution is rejected if it fails to meet any one of three acceptance criteria. In practice, these criteria are most often violated because of a computational instability. First, the global error must be small; solutions with $E > E_{\max}$ are rejected (for our realization, we choose $E_{\max} = 10^{-4}$). Second, local errors must be bounded; we test for $\varepsilon > \varepsilon_{\max}$ every $3\,\Delta t$, with $\varepsilon_{\max} \geqslant E_{\max}$. Third, we monitor the L_2-norm of X at every time step and reject a solution if

$$\|X(t)\|_2 > B_*, \qquad 0 \leqslant B_* < \infty. \tag{53}$$

The appropriate value of B_* is highly solution-dependent, but it can be chosen to be much larger than a representative value of $\|X\|_2$ for given initial conditions, yet still be a fairly sensitive indicator of instability. Other fields could be monitored in addition to or instead of X; however, this seems to be unnecessary, since once unstable growth occurs in one field, all other fields tend to follow. In most instances, solutions fail by the second and third criteria within a few time steps of each other.

A. *Time-Method Evaluation*

Here we document the numerical behavior of time methods M1 and M2, determine optimal parameters for each, and select one for later use in the model comparisons and physical studies. The exploration of numerical behavior occurs under conditions of moderate numerical difficulty: an intermediate R value (0.25) and stable Δt values.

Numerical parameters unique to M1 are L, the number of iteration-correction cycles, and $N(\lambda)$, the number of (X, ψ_t) iterations for each $\lambda = 1, 2, \ldots, L$ (see (29)).

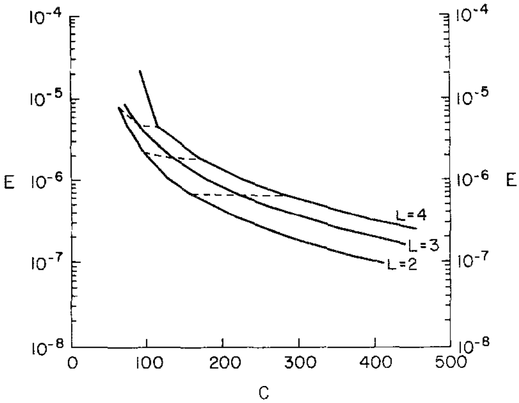

FIG. 2. Cost-error curves for time-method M1 with $N(\lambda) \equiv 1$ and L as labeled. Dashed lines connect solutions with common Δt.

For each value of $L = 1, 2, 3, 4$ and several Δt, we considered all possible combinations of $N(\lambda)$ such that

$$\sum_{\lambda=1}^{L} N(\lambda) = L + i, \qquad i = 0, 1, \ldots, I, \tag{54}$$

where

$$I = \begin{cases} 2, & L = 1 \\ 3, & L = 2 \\ 1, & L = 3, 4. \end{cases}$$

We found that for a fixed value of $L > 1$ and Δt, each additional iteration cycle (i increases by one) increases the cost by 30% of the minimum-iteration case ($i = 0$), while the resulting errors vary surprisingly little—less than 8%, often no more than 2%—from minimum-iteration errors (which are not always the largest errors). Further, for a given value of L, the size of the largest-allowable time step does not increase for $i > 0$. Hence, for $L > 1$, we determine $N(\lambda) \equiv 1$ is optimal.[5]

Optimal M1 cost-error curves for $L = 2, 3, 4$ and several Δt are shown in Fig. 2. Dashed lines connect cost-error points for fixed values of Δt and three values of L and illustrate that, except for the largest-Δt cases, an increase in L results in an

[5] The $L = 1$ cases behave peculiarly. One is the only value of L for which $i > 0$ results in accuracy improvement: $i = 1$ accuracy is much greater than that for $i = 0$, although a further increase in i results in no further error reduction. For small Δt, $i = 0$ solutions are ill-behaved in that errors begin to increase as Δt decreases, but $O((\Delta t)^2)$ error improvement is restored with the application of relaxation. No $L = 1$ cases are shown in Fig. 2, because they are all less accurate for a fixed cost and more expensive for a fixed accuracy than $L > 1$ cases (i.e., the $L = 1$ curve is above and to the right of the $L = 4$ curve).

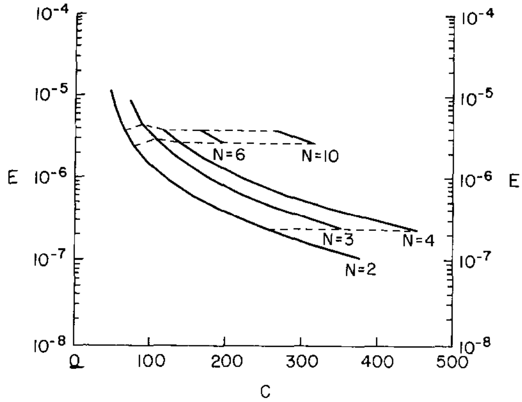

FIG. 3. Cost-error curves for time-method M2 with N as labeled. Dashed lines connect solutions with common Δt.

increased cost with little or no gain in accuracy. Thus, the error appears to be time-step-controlled, not iteration-controlled. Because costs increase with increased L and accuracy does not increase significantly, if at all, we conclude $L=2$, $N(1)=N(2)=1$ is the optimal choice of M1 numerical parameters.

In method M2, N controls the number of iteration-correction cycles. We selected $N=1, 2, 3, 4, 6, 10$ and many Δt; the resulting constant-N curves are shown in Fig. 3. For a fixed value of Δt, errors decrease substantially, by nearly a factor of 10, as N is increased from one (not shown[6]) to two, but errors are nearly unchanged as N increases from two to ten (dashed lines on Fig. 3). For Δt's near their largest-allowable values, we actually see a slight unexplained increase in error as N is increased from two to three, but this behavior does not occur for smaller values of Δt. The cost for each additional iteration-correction cycle is 65% of the one-cycle cost (with little or no error improvement for $N>2$); hence, for M2, $N=2$ is optimal for any value of Δt. This conclusion also holds under the numerically difficult situation of R values near the large-R solvability barrier (see Section 5.C).

Figure 4 is a comparison of optimal M1 and M2 cost-error curves. M2 errors are somewhat smaller than M1 errors for large Δt values, but the curves are nearly indistinguishable for small Δt's. We believe that the coincidence of the curves for small Δt may not be a universal result; relative positions of the curves in the small-Δt regime may depend on a particular realization or global error measure. However, the slight superiority of M2 over M1 in the larger-Δt regime has been consistently observed; it occurred in our earlier time-method testing on ODE systems and the companion studies of isolated vortices.

[6] Ill-behaved solutions similar to those in the M1, $L=1$, $N(1)=1$ case also occur for small-Δt values in the M2, $N=1$ case; relaxation restores $O((\Delta t)^2)$ error improvement with decreasing Δt, but does not make the solutions competitive with $N>1$ solutions. The similarity in M1 and M2 method behavior is not surprising in these cases, because the methods are quite similar for these minimum-iteration parameters.

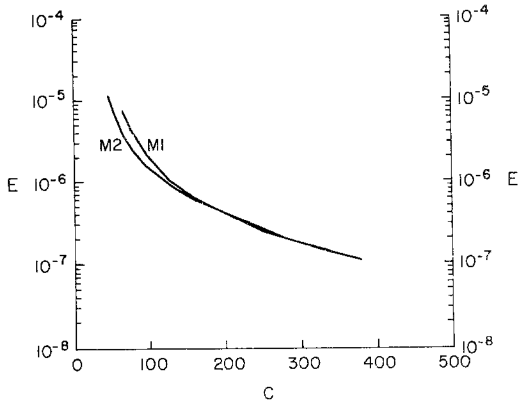

FIG. 4. Optimal M1 and M2 solutions ($L=2$, $N(\lambda)\equiv 1$ and $N=2$, respectively).

B. *Model Intercomparisons*

The QG and PE systems may be solved by the time methods $p(ec)^L$ and $pe(ce)^N$. [illegible] is the first member ($N=0$) of the latter class of methods. Note that p, e, and c have slightly different meanings in the QG and PE models (see Appendix): p, e, and c each apply to X, ψ, and θ in the PE model; p and c refer to ψ only and e implies evaluate both ψ_t and X in the QG model.

Once computations were begun, it became obvious that only small values of L and N needed to be considered: we limit our discussion to $N=0, 1, 2$ and $L=1, 2$. Both the QG and PE models perform poorly in the pe mode; QG solutions all fail by condition (53) before completion for all Δt values tested, while all leapfrog PE solutions have unacceptably large global errors (both are manifestations of splitting instability [10]). A correction step added at each time step (i.e., $p(ec)^1$) prevents failure in the QG model and improves the PE global error to an acceptable level with only a negligible ($<1\%$) additional cost. Hence, we see no need to consider the ad hoc leapfrog method with an occasional correction step.

Figure 5 clearly shows the superiority of the $p(ec)^1$ scheme for the PE model. Although $p(ec)^1$ requires the smallest Δt value of any of the successful schemes, we presume $p(ec)^1$ outperforms $pe(ce)^1$ because the accuracy gained by a final evaluation cannot offset its great cost.

The minimum cost of a PE solution is much greater than the BE minimum cost, partly because the largest successful PE time step is so much smaller than that of the BE. This greatly reduced time step is required to resolve gravity wave

[7] In some applications, use of so-called "semi-implicit" time methods is advocated for the PE [9]. These have the property of allowing larger Δt, hence possibly smaller C, by numerically damping the high-frequency solution components. Since this is an alteration of the physical problem being calculated, biasing both the numerical and physical model intercomparisons, we do not consider such methods here.

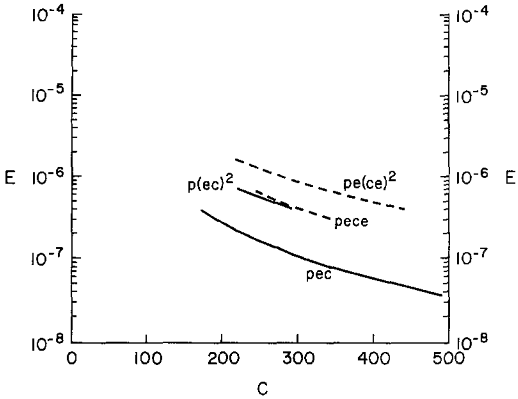

FIG. 5. PE solutions with time methods as labeled.

propagation in the PE model. The gravity wave frequencies are proportional to $R^{-1}k$ for large k, where k is the horizontal wavenumber [14]; this is an order of magnitude larger than the advective rates for $R = 0.25$ and the k values excited here (see Fig. 18). For larger R values, the ratio of costs for the BE and PE models increases but remains well below one for all cases examined.

Figure 6 shows that the choice of the optimal QG method is much less clear than for the PE: $\mathrm{p(ec)}^2$ is the best method in the low-cost regime, but the $\mathrm{p(ec)}^1$ accuracy improves more rapidly with decreasing Δt in the intermediate-cost regime than does $\mathrm{p(ec)}^2$, so that $\mathrm{p(ec)}^1$ is the best method in the high-cost regime. Indeed, accuracy for both $\mathrm{p(ec)}^1$ and $\mathrm{pe(ce)}^1$ improves faster than $O((\Delta t)^2)$ over a very large range of Δt values; only at very small Δt values ($C > 500$) do these accuracy improvement rates slow and approach the expected asymptotic limit.

The relative performance of the three models is summarized in Fig. 7. The QG scheme is $\mathrm{p(ec)}^2$, the BE is M2 (more complicated than, but derived from, a $\mathrm{pe(ce)}^2$

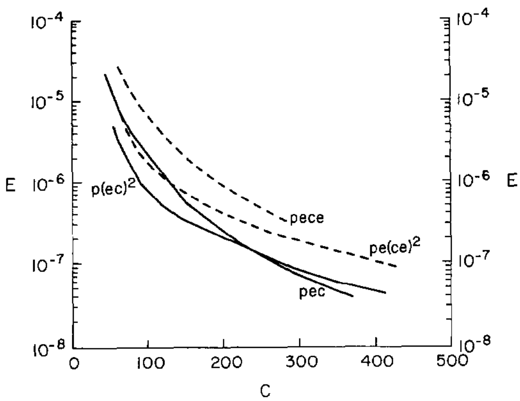

FIG. 6. QG solutions with time methods as labeled.

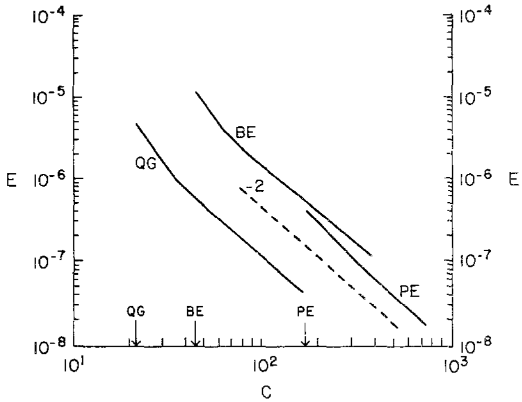

FIG. 7. Model intercomparison based upon optimal time methods (see text). A dashed line $\propto C^{-2} \propto (\Delta t)^2$ is drawn for reference. Least costs for each model are marked on the abscissa with arrows.

scheme), and the PE is p(ec)[1]. (For comparably small Δt values, the optimal scheme for both the QG and PE is p(ec)[1].) Note that all three models closely parallel the dashed $O((\Delta t)^2)$ curve in their small-Δt limits. As noted above, all E values on this figure are acceptably small by criteria of physical analysis (see Section 6); hence, in practice there is no need to improve upon the minimum-cost solutions. Arrows point to the minimum cost for each model; roughly, the ratio of PE : BE costs is 4 : 1 and BE : QG is 2 : 1.[8] These ratios show that the BE model is much closer to the QG than the PE in cost.

C. *Solvability Barrier for Large R*

Intermediate Models in general, and the BE in particular, are subject to solvability barriers [5]. In general no explicit specification of the barrier location is known for the BE, except in certain simple situations such as a steady, circular vortex ($\beta = \nu = 0$ [13]) and the Lorenz Triad Equations [3]. We find we are unable to obtain valid numerical solutions of the BE at sufficiently large values of R, both for the present initial-value problem as well as for isolated vortex solutions [14]. We presume this occurs quite generally, although it has only been demonstrated for these two solution classes. Thus far, the inaccessible R values have been found to be moderately large, large enough to permit a study of BE solutions for Rossby numbers ranging from zero to order unity. The numerical solvability barrier occurs at different values of R in different types of initial-value problems. A numerical solvability failure, of course, is not necessarily indicative of a solvability barrier for the continuous equations; it might rather only be a failure of a particular numerical

[8] This cost comparison is more appropriate than one for a fixed error level (which yields a cost ratio of about 3 : 1 for both BE : QG and PE : QG), since in practice only the PE would be used in the large C regime where this comparison is possible.

technique. Nevertheless, the numerical solvability barrier is interesting for the possibility of its representing the continuous barrier.

To locate the barrier we have made use of several alternative definitions of solution failure: (1) inability to achieve convergence in the initial iteration step for N_0 arbitrarily large in (32); (2) excessive energy imbalance, with $\varepsilon_{\max}$ chosen at a level of several percent of K_t; and (3) the occurrence of (53), with B_* chosen at a level of several times the value of $\|X(t=0)\|_2$. We designate the Rossby number at which solution failure occurs as $R_*(t)$. Definition (1) is independent of Δt, and, for definiteness, we set $\nu=0$. Definitions (2) and (3) depend upon these parameters as well as the particular threshold values; we test for them with ν sufficiently large and Δt sufficiently small so as to distinguish a solvability barrier in R from non-linear computational instability due to insufficient viscosity.

In practice we have found that all three definitions give quite similar barrier locations: (2) and (3) always occur quite close together in time, and, near where (1) occurs, (2) and (3) occur at very early t. In part this coincidence is because the solvability is insensitive to the particular values of Δt, Δs, ν, time method, and L or N (but not N_0), as long as they conform to the conditions for satisfactory solutions in typical circumstances (see Section 5.A).[9] Similarly the values for $\varepsilon_{\max}$ and B_* are not very important since solution failure takes the form of explosive computational instability. We have not explored the sensitivity to N_z, lacking sufficient computer memory to do so while keeping Δs small enough to resolve the turbulent cascade.

The only numerical parameter we have found to be important near the barrier is α in (34). Away from the barrier, relaxation is usually neither needed nor advantageous; for example, in solutions with $R=0.25$ and small Δt, the least error for a given computational cost occurs at $\alpha=0$, i.e., no relaxation. However, the barrier can be displaced in R or t by increasing α up to a limit of 1. This is illustrated in Fig. 8a for definition (1) above: $R_*(0)$ is monotonically increasing with α throughout the range plotted. As α approaches 1 from below, N_0 must be taken increasingly large to determine R_* satisfactorily; we have only plotted points which are well determined for $N_0 \leqslant 1000$.

$R_*(t)$ decreases with time, as shown in Fig. 8b. The rate of decrease of $R_*(t)$ is rapid at early times (on the order of a characteristic advection time) but later diminishes to zero at a finite value of $R_*(\infty) \approx 1.1$. Thus solutions with $R < R_*(\infty)$ can be integrated without limit in time. $R_*(t)$ increases with α, as in Fig. 8a; however, its range of variation with α is greatly reduced as t increases.

We can offer only a partial interpretation of this solvability barrier. It is based upon the facts that, for a given ψ or ζ field, the concomitant BE w field has an increasing amount of small-scale variance as R increases, and the evolving ζ field has an increasingly frontal character and larger small-scale variance at a given t as R increases (see Section 6). It is even possible that the mode of solvability failure for

[9] A specific illustration of the insensitivity of the barrier location to ν is the following: for $R=1.5$ a 1000% increase in ν delays the time of failure in Fig. 8b by less than 10%.

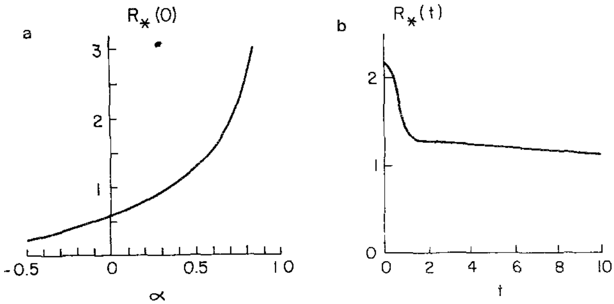

FIG. 8. Solvability barriers: (a) $R_*(t=0, \alpha)$ and (b) $R_*(t)$ for fixed values of the numerical parameters (in particular, $\alpha = 0.75$ and $\Delta t = 0.00625$).

the continuous BE is the occurrence of a spatial singularity at finite R, although this cannot be demonstrated in discrete computations. The R dependence of w is shown at $t=0$ in Fig. 9; the rate of increase of small-scale variance in w is quite rapid with R, and to an increasing degree with wavenumber.

Since excessive variance at small scales is a cause of non-linear computational instability, model failure by definitions (2) and (3) is consistent with the preceding interpretation, although we have not attempted to quantify "excessive" here. [redacted] and conversely success by (1) most probably implies success by (2) and (3) for at least some time interval after $t=0$. Thus, the close correspondence among barrier locations determined by (2) and (3) at small t and (1) at $t=0$ is at least not inconsistent with failure due to excessive small-scale content.

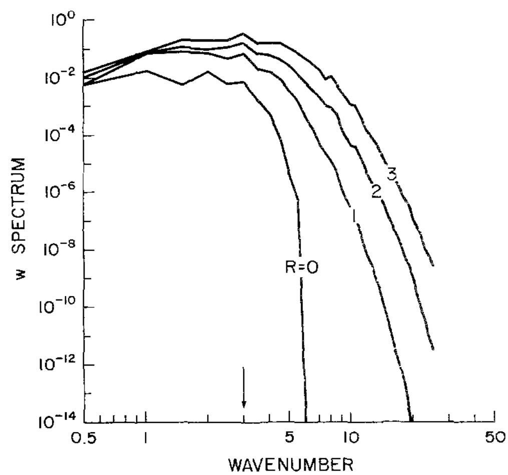

FIG. 9. Azimuthally averaged wavenumber spectra of $w(x, y, 0)$ for different R values. The arrow marks the highest wavenumber for which the spectrum of $\zeta(x, y, 0)$ from (50) is non-trivial.

The decrease of R_* with t also is consistent with this interpretation: from the relatively large-scale initial conditions (50), a turbulent cascade transfers ζ to smaller scales, which also implies an increased small-scale content in w with time for fixed R. This enhancement by cascade increases for several characteristic advective times, after which the $R_*(t)$ curve approaches a horizontal asymptote (see Fig. 8b).

We are, of course, uncertain how close the numerical solvability barrier might be to a barrier for the continuous BE. We can make the hypothesis that the failure of the BE in crossing the barrier will be associated in the PE with the enhanced excitation of gravity waves, which are excluded from the BE. Evidence is presented in Section 6 that the PE gravity wave component grows rapidly with larger R and t, even where the companion BE solution lies within its solvability domain. Near the barrier itself (e.g., $R = 1.5$, $t = 1.0$—see Fig. 8b), PE solutions exhibit a rapidly increasing gravity wave component with t, but present no particular difficulties in integrations across it. Thus, at least this aspect of PE solution behavior may exhibit a parametrically or temporally broad, rather than abrupt, transition across a BE solvability barrier.

6. Physical Properties

The problem posed in Section 4 lies in the regime of baroclinic, geostrophic turbulence when $R = 0$, which in turn has some properties in common with two-

t = 0

t = 5

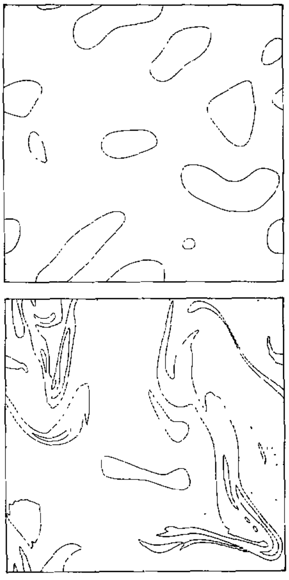

FIG. 10. $\zeta_2(x, y)$ for $R = 0$ at $t = 0$ and 5. The contour interval is 2.5. In all contour plots, the contours straddle zero, positive contours are solid, and negative contours are dashed.

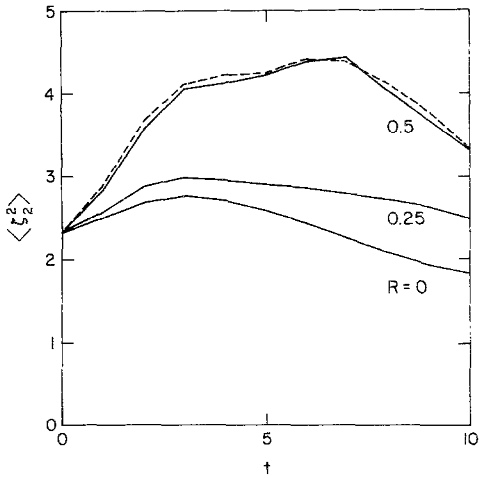

FIG. 11. Upper-level vorticity variance for several R values. Solid curves are from the BE or QG models and the dashed curve is from the PE.

dimensional turbulence [2, 17–19]. These are regimes where the turbulent cascade has a tendency for energy to be transferred to larger spatial scales (in all three dimensions) and enstrophy (i.e., $\langle\zeta^2\rangle$) to be transferred to smaller scales where it is dissipated. For initial conditions with sufficient energy in the baroclinic mode on horizontal scales not small compared to $1/\lambda^{(2)}$, as in (50), a conversion of energy from potential to kinetic (i.e., $\langle w\theta\rangle > 0$ in (17)–(18)) may be expected; this process is often referred to as baroclinic instability. Together the cascade and baroclinic instability imply an increase in the energy in the barotropic mode on large scales.

Some of this behavior can be seen in Fig. 10. The upper level has the largest vorticity amplitude because $F^{(2)}$ has its peak there. Consequently it is the site of the

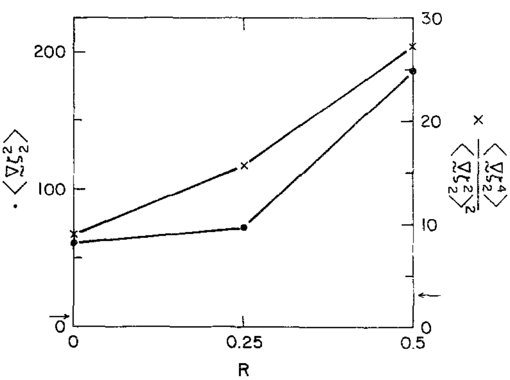

FIG. 12. Maximum values for upper-level vorticity gradient variances and kurtosis for $t \in [0, 10]$. Values at $t = 0$ (valid for all R) are indicated by the arrows on the respective ordinates.

most intense cascade. The vorticity field tends to develop elongated structures with time, while its peak amplitude changes very little. These developments have occurred to their maximum degree by the time $t = 5$ shown in Fig. 10. The elongation is the characteristic form by which enstrophy is transferred to small scales. As it occurs vorticity gradients increase in magnitude and both energy and enstrophy dissipation increase ($\langle \nabla\zeta^2 \rangle$ is proportional to the kinetic energy dissipation rate in (17) with D defined by (11)). The other processes listed in the preceding paragraph are also occurring in this solution.

We give the name balanced turbulence to the extension of geostrophic turbulence to finite R. In the present example several changes in behavior occur as R increases. One is the generation of more enstrophy, particularly at the upper level (see Fig. 11). The temporal peak has a larger magnitude and occurs later for larger R. These tendencies are found even more strongly in $\nabla\zeta$ (see Fig. 12). Even the QG solution shows an increase in variance by nearly a factor of ten over the initial value, and this is further increased by a factor of more than three in the BE solution with $R = 0.5$. Figure 12 also includes the kurtosis of the upper-level vorticity gradient and shows that spatial intermittency increases with R: it reaches a value nearly ten times the Gaussian value of three. Thus, the kinetic energy dissipation rate is larger and more intermittent in balanced turbulence than in geostrophic turbulence. The same is true for the enstrophy dissipation rate, $\langle \Delta\zeta^2 \rangle$, yet again more strongly. Collectively these measures justify the conclusion that the cascade is stronger in balanced turbulence.[10]

The spatial patterns of vorticity are shown in Fig. 13 at a time near the enstrophy maximum. Together with Fig. 10, these illustrate the dependence upon R. The increasing amplitude, gradient, and intermittency are concentrated in narrow, elongated zones, which might be called vorticity fronts in analogy with meteorological fronts.

Other fields exhibit different degrees of variation with R. In general, fields more indicative of the larger scales of motion have a lesser R dependence than does upper-level vorticity, with the reverse for fields more indicative of smaller scales. For example, streamfunction patterns are quite similar at the lower level and moderately so at the upper level, even after the longer evolution time of $t = 10$ (Fig. 14). One can also see correlations between the upper and lower streamfunction fields, which is indicative of a dominant barotropic component having evolved, at least for the larger horizontal scales. (From (50), the expected value for the initial correlation between the ψ_i is zero.) There does not appear to be a strong R dependence in the degree of barotropic dominance.

[10] An artificially enhanced cascade can occur in the QG if the spatial discretization formulae do not conserve both energy and enstrophy (R. Salmon, personal communication). It is therefore possible that the enchanced cascades seen in the BE and PE are due more to the non-conservative discrete formulae used than to the intrinsic non-conservation of enstrophy in their continuous forms. However, we used an alternate, non-conservative Jacobian operator in the QG and found that it does indeed lead to an enhanced cascade, but the degree of enhancement is much less than that in the BE and PE as R increases from zero.

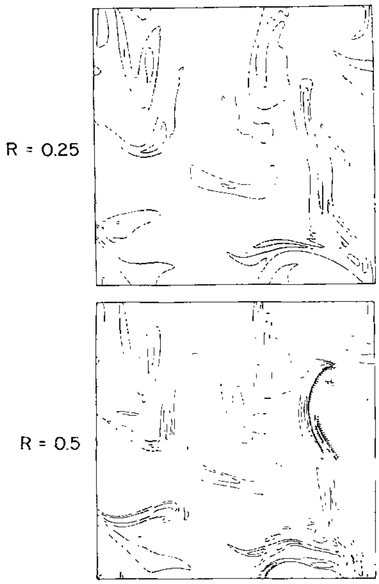

FIG. 13. $\zeta_2(x, y)$ at $t = 5$ for $R = 0.25$ and 0.5. The contour interval is 2.5.

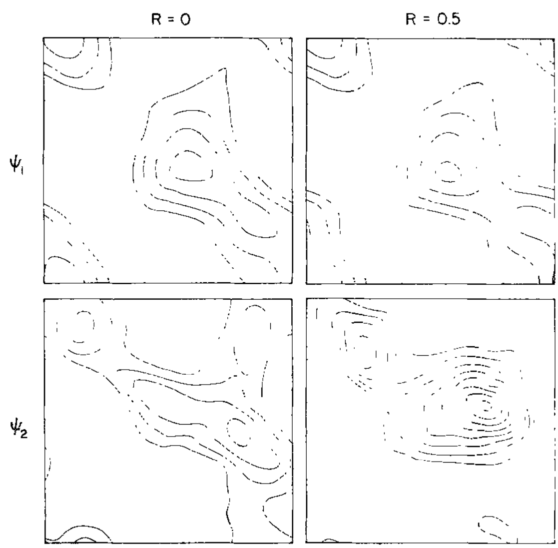

FIG. 14. Lower- and upper-level streamfunction at $t = 10$ for $R = 0$ and 0.5. The contour interval is 0.6.

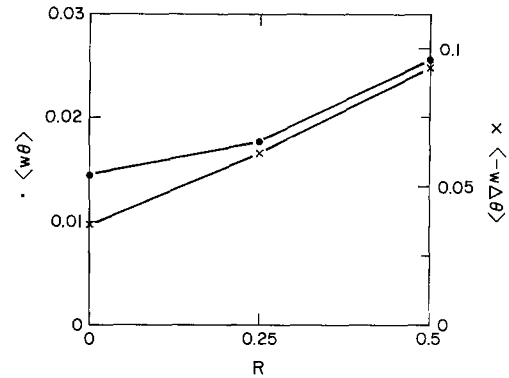

FIG. 15. Maximum values for the indicated source terms for kinetic energy and enstrophy for $t \in [0, 10]$.

The potential energy conversion rate in (17)–(18), $\langle w\theta\rangle$, ascends to a peak at moderately early times ($t = 2$–3) and thereafter descends to irregular oscillations about zero. Its peak magnitude is also an increasing function of R (see Fig. 15). The maximum rate of kinetic energy generation, over all R and t values calculated here, is about 30 times larger than the maximum energy dissipation rate discussed above. The analogous source term for enstrophy, $\langle -w\,\Delta\theta\rangle$, is also plotted in Fig. 15 and shows a slightly greater increase with R. However, in contrast to kinetic energy, there is more than a single inviscid source term for enstrophy, although the others are $O(R)$ relative to $\langle -w\,\Delta\theta\rangle$; this non-uniqueness makes its interpretation somewhat ambiguous for finite R.

An important issue, of course, is how accurately the BE are representing the effects due to finite R. This we assess by comparison with PE solutions having identical initial conditions and physical parameter values. For the BE solutions shown above, the correspondence with PE solutions is remarkably close for fields reflecting primarily the influence of the larger scales of motion. The same is true even for the upper level vorticity field at $R = 0.5$. One measure of this, $\langle\zeta^2\rangle$, is shown in Fig. 11 [illegible] integration, even when the differences with QG values are quite large. A more detailed comparison is shown in Fig. 16. At $t = 5$ differences between the BE and PE can only be detected by very careful scrutiny, whereas by $t = 10$ differences are more easily detectable, although even then they are primarily deformations and small translations of identifiable structures. (Recall that turbulence has the property of limited predictability; i.e., small differences between solutions tend to grow exponentially in time even within the same model.) Streamfunction patterns for the PE are indistinguishable from those of the BE in Fig. 14.

On the other hand, there is at least one major difference between the BE and PE: the PE have gravity wave solutions, while the BE do not. A useful indicator of gravity wave behavior is vertical velocity. Time sequences of $\langle w^2\rangle$ are shown in

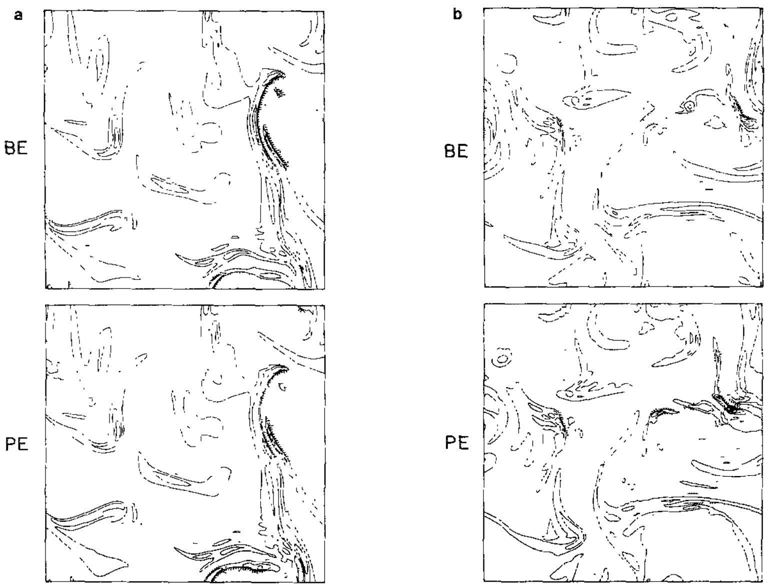

FIG. 16. $\zeta_2(x, y)$ at $R = 0.5$ from the BE and PE solutions: (a) $t = 5$ and (b) $t = 10$. The contour interval is 2.5.

Fig. 17. At the smaller R value, $\langle w^2 \rangle$ for the PE is only slightly larger than that for the BE, indicating the presence of a weak gravity wave component whose amplitude appears bounded in time. For the larger R value, however, $\langle w^2 \rangle$ for the PE diverges systematically in time until it is more than an order of magnitude larger by $t = 10$. The gravity waves appear in the solutions in spatially localized bursts which subsequently propagate as fairly well-defined wave packets. With time more sites exhibit bursts, and the domain becomes filled with propagating waves. An illustration at an early time is shown in Fig. 18. The contour interval is chosen so that the corresponding BE plot would be empty, while the PE plot shows a first strong burst in the lower right corner of the domain. In spite of the discrepancy in the BE and PE w fields, the energy conversion rate, $\langle w\theta \rangle$, differs little in the two solutions during the period of systematic kinetic energy generation. Also, the divergent kinetic energy in the PE (see (A8)), which is closely related to w, remains a very small fraction of the rotational kinetic energy (14), always less than 1% in the $R = 0.5$ solution. Even the divergent velocity variance in the PE is only about 1.5 times that of the BE, because the large-amplitude gravity waves are quite small in horizontal scale.

Thus, while there are substantial differences between the BE and PE w fields when R is large enough, due to PE gravity wave bursts, there are quite close correspondences in the energy cycle and rotational flow field evolution.

Finally, we remark that there are many alternative definitions of Rossby number. Given (51), our parameter R is equivalent to the ratio of initial r.m.s. vorticity and

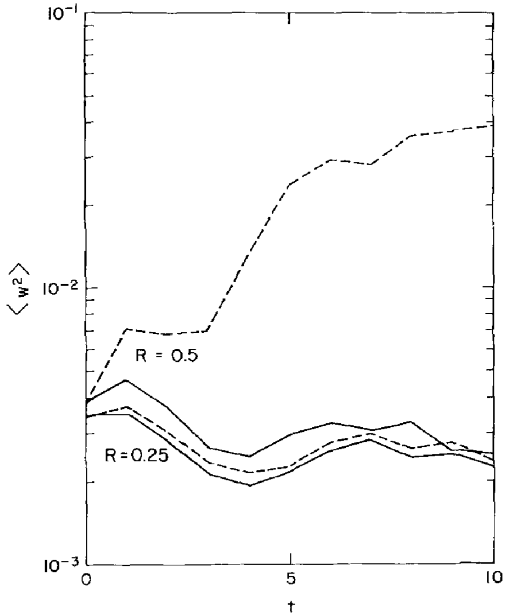

FIG. 17. Vertical velocity variance for two R values and both the BE (solid line) and PE (dashed line) model solutions.

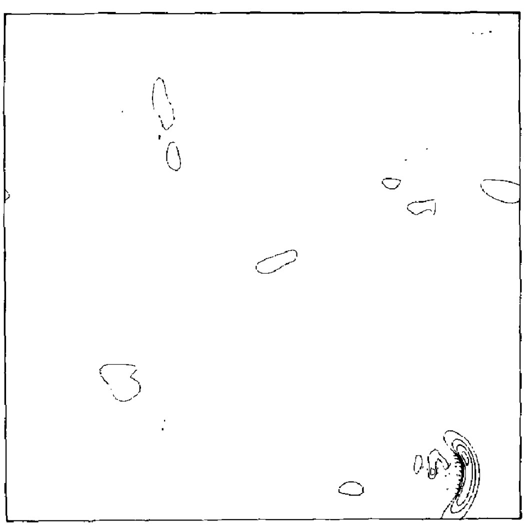

FIG. 18. $w(x, y)$ at $t = 4$ and $R = 0.5$ in the PE solution. The contour interval is 0.4.

Coriolis frequency. A measure based upon r.m.s. velocity and the length scale of its spectrum peak would be somewhat smaller. On the other hand, normalized vorticity measures using r.m.s. upper-level vorticity or the maximum vorticity would be larger, the former by at least 50% and the latter by a factor of ten or more at the time of greatest enstrophy. Thus, the R values we have considered here are not small by at least some measures. Furthermore, they are large enough to violate the sufficient condition for inertial instability derived for axisymmetric vortices by Ooyama [16], viz., $1 + R\zeta < 0$ at some locations in the domain. Whether or not this condition is relevent in more general flow geometries, our solutions excise symmetric inertial instability both by physical approximation (for the QG and BE) and by coarseness of vertical resolution (the condition is never violated in the PE at both levels in the same horizontal location).

7. Summary and Prospects

In this paper numerical methods are presented for the solution of the Balance Equations. These methods, while novel in detail, are relatively familiar in type: low-order finite-differences, predictor-corrector time integration, iterative solution of implicitly defined time derivatives, and conservation of some integral properties of the continuous model. The numerical behavior of the model is satisfactory, both in accuracy and cost, in a physical regime of significant environmental rotation where the use of the Balance Equations is appropriate.

The Balance Equations are shown to be an excellent Intermediate Model in this regime. They accurately represent the changes in solution behavior as R increases from zero (Quasigeostrophy) to values of order one, judged by the standard of the rotational velocity component, or vorticity, of the Primitive Equations. The Balance Equations do not include the gravity wave component of the Primitive Equations, which increases in strength with R but does not here become comparable in energy to the rotational component. In this regime the Balance Equations are much less expensive to integrate than the Primitive Equations, although they are more expensive than Quasigeostrophy. The success of the Balance Equations as an Intermediate Model for the phenomenon examined here (balanced turbulence) is also found for isolated vortices at intermediate R values [14, 15].

However, with further increases with R (at fixed Burger number of $B = 1$), advection rates surpass some of the gravity wave rates. gravity wave amplitudes greatly increase, computational expenses become more similar among the different models, and the accuracy of the Balance Equations can be expected to diminish, even to the point of encountering a numerical solvability barrier. (In our examples the Balance Equations remain quite accurate, in the sense above, even for R and t values quite close to the barrier. This barrier does not yet have a satisfactory physical interpretation.) Thus, in some regimes with large R the model advantage must shift to the Primitive Equations. Here we have done little to explore this frontier to the regime of utility for the Balance Equations, except to document some occurrences of the numerical solvability barrier. In particular, we have not explored large values of B (strong stratification), where the Balance Equations may be a good model even for large R [12].

Finally, the physical nature of balanced turbulence is illustrated in several solutions. As R increases from zero, the rates of enstrophy transfer to small scales, kinetic energy dissipation, and conversion of potential to kinetic energy all increase. Also, the vorticity field becomes more spatially intermittent, with increasingly isolated, horizontally anisotropic (frontal) zones of large amplitude and gradient. This physical regime is mostly unexplored, and our solutions are a first, albeit very incomplete, survey of its phenomena. Nevertheless, it is a very important regime, connecting the largest scales (dominated by rotation : geostrophic waves and turbulence) to the smallest ones (with negligible rotation : gravity waves and stratified turbulence). The present results suggest that the Balance Equations may play an important role in its exploration.

APPENDIX: QUASIGEOSTROPHY AND THE PRIMITIVE EQUATIONS

The QG model equations are a subset of the BE, defined by (4)–(13) with $R \equiv 0$. This system has the energy cycle (12)–(18) for K defined by (14) and an alternative P, viz.,

$$P \equiv \frac{1}{2} \int dz\, S^{-1} \langle \theta^2 \rangle. \tag{A1}$$

S and Θ are invariant with time, but the $M^{(n)}$, $n > 1$, are not conserved. In their place all integral powers of potential vorticity, $\zeta + (S^{-1}\theta)_z + \beta y$, are conserved when $D \equiv 0$. Since the QG have an explicit time-derivative structure, the time methods are the reduced ones described at the end of Section 3.A. The primary dependent variable is ψ. We evaluate its derivative ψ_t from (6) with w evaluated from (8) and $\Theta = \psi_z$. All necessary spatial discretization formulae are given in Section 3.B. These imply that the discrete conservation laws for energy, $M^{(1)}$, and the first and second powers of potential vorticity are satisfied exactly apart from time-method errors.

The PE model is not a subset of the BE; rather the opposite is true. The governing equations include (2)–(5); a vorticity equation which adds to (6) the

The specific formulae include the following:

$$
\begin{aligned}
J(w, X_{zz}) &= \bar{A}[J(w, V^*[V[X]])] \\
w_{zzt} &= V^*[V[w_t]] \\
J(\psi, w_z)_z &= V^*[J(\psi, V[w])] \\
J(\nabla\psi; \nabla X_z)_z &= V^*[J(\nabla\psi; V[\nabla X])] \\
J(\psi_z, w)_z &= V^*[\bar{A}[J(V^*[\psi], w)]] \\
\tfrac{1}{2}\Delta(\nabla X_z^2)_z &= \Delta(V^*[V[\nabla X]] \cdot V^*[A[\nabla X]]) \\
\nabla \cdot (w\nabla X_{zz})_z &= \nabla \cdot (V^*[A[wV^*[V[\nabla X]]]]) \\
\nabla \cdot (\theta\nabla X_z) &= \nabla \cdot (\theta V^*[A[\nabla X]]) \\
(\theta w)_z &= V^*[A[\theta]\, A[w]].
\end{aligned} \tag{A9}
$$

The operator $\bar{A}$ is an alternate average to that defined in (37). Its definition is

$$\bar{A}[q] = \frac{1}{2}\left(\frac{z_{i+1} - z_i}{\Delta h_i} q(z_i^*) + \frac{z_i - z_{i-1}}{\Delta h_i} q(z_{i-1}^*)\right), \tag{A10}$$

and it yields a product at level z_i. Horizontal operators here are the Arakawa Jacobian and (42), (44), (A5), and an extension of (47), viz.,

$$
\begin{aligned}
\nabla a \cdot \nabla b \to \tfrac{1}{2}((a_{m+1,n} - a_{m,n})(b_{m+1,n} - b_{m,n}) + (a_{m-1,n} - a_{m,n})(b_{m-1,n} - b_{m,n}) \\
+ (a_{m,n+1} - a_{m,n})(b_{m,n+1} - b_{m,n}) + (a_{m,n-1} - a_{m,n})(b_{m,n-1} - b_{m,n}))/(\Delta s)^2.
\end{aligned} \tag{A11}
$$

The energy equation (A7) holds with (46) and

$$
\begin{aligned}
\hat{K} &\to \frac{1}{2} R^2 \sum_{z_i} \Delta h_i \langle (\nabla V[X])^2 \rangle, \\
\int dz \langle X\hat{D} \rangle &\to \sum_{z_i^*} (z_{i+1} - z_i) \langle X\hat{D} \rangle.
\end{aligned} \tag{A12}
$$

References

1. A. Arakawa, *J. Comput. Phys.* **1**, 119 (1966).
2. J. Charney, *J. Atmos. Sci.* **28**, 1087 (1971).
3. J. H. Curry and D. Winsand, *J. Atmos. Sci.*, in press.
4. P. R. Gent and J. C. McWilliams, *J. Atmos. Sci.* **39**, 3 (1982).
5. P. R. Gent and J. C. McWilliams, *Dyn. Atmos. and Oceans* **7**, 67 (1983).
6. P. R. Gent and J. C. McWilliams, *Dyn. Atmos. and Oceans* **7**, 167 (1983).

7. P. R. Gent and J. C. McWilliams, *Tellus* **36A**, 166 (1984).
8. H. Kreiss and J. Oliger, *GARP Pub. Series No. 10* (World Meteorological Organization, 1973), pp. 1–107.
9. M. Kwizak and A. J. Robert, *Mon. Weather Rev.* **99**, 32 (1971).
10. D. K. Lilly, *Mon. Weather Rev.* **93**, 11 (1965).
11. E. N. Lorenz, *Tellus* **12**, 364 (1960).
12. J. C. McWilliams, *J. Atmos. Sci.* **42**, 1773 (1985).
13. J. C. McWilliams and P. R. Gent, *J. Atmos. Sci.* **37**, 1657 (1980).
14. J. C. McWilliams, P. R. Gent, and N. J. Norton, *J. Phys. Oceanogr.* **16**, 838 (1986).
15. J. C. McWilliams and P. R. Gent, *Geophys. Astrophys. Fluid. Dyn.* **35**, 235 (1986).
16. K. Ooyama, *J. Atmos. Sci.* **23**, 43 (1966).
17. P. B. Rhines, *Ann. Rev. Fluid Mech.* **11**, 401 (1979).
18. R. Salmon, *Geophys. Astrophys. Fluid Dyn.* **10**, 25 (1978).
19. R. Salmon, *Topics in Ocean Physics LXXX* (North-Holland, New York, 1982), pp. 30–78.