Origins and Evolution of Imbalance in Synoptic-Scale Baroclinic Wave Life Cycles

ANDREW B. G. BUSH, JAMES C. MCWILLIAMS,* AND W. RICHARD PELTIER

Department of Physics, University of Toronto, Toronto, Ontario, Canada

(Manuscript received 14 January 1994, in final form 27 July 1994)

ABSTRACT

A set of balance equations is derived that is appropriate for analysis of the three-dimensional anelastic system and is based on expansions in Rossby and Froude number similar to those employed in the study of the shallow-water equations by Spall and McWilliams. Terms that constitute the usual balance equations are formally retained here in addition to non-Boussinesq terms of the same order arising from the vertical variation of the background density field. The authors apply the derived set of equations diagnostically to the analysis of three-dimensional, anelastic numerical simulations of a synoptic-scale baroclinic wave. Of particular interest in this analysis is the degree to which and the time at which the flow becomes appreciably unbalanced, as well as the form of the imbalance itself. Unbalanced motions are here defined as departures from solutions of the balance equations. Application of this analysis procedure allows us to identify two classes of unbalanced motion, respectively: 1) unbalanced motion that is slaved to the balanced motion and is therefore characterized by the same time and length scales as the balanced motion (i.e., higher-order corrections on the "slow" manifold) and 2) unbalanced motion that is on a faster timescale than the large-scale balanced motion but is nevertheless forced by these same balanced motions (e.g., forced internal gravity waves). It will be shown in the analysis that both forms of imbalance arise in the frontal zones generated during the numerical simulation, but that the gravity wave generation is probably a numerical artifact of insufficient vertical resolution as the slope of the surface front decreases below the threshold required for consistent horizontal and vertical resolution. The total unbalanced motion field is dominated by the slower advective motion, but the numerically generated gravity waves nevertheless reach a peak amplitude comparable to that of the slower unbalanced motion. Whether internal wave radiation would persist, or perhaps become more intense, with increased spatial resolution is an issue that is left unresolved in the present analysis.

1. Introduction

The observationally motivated fact that large-scale atmospheric and oceanic motions approximately satisfy the gradient wind balance equation or have a small characteristic Rossby number R ($=V/fL$, where V and L are typical velocity and length scales, respectively, and f is the Coriolis parameter) eventually led to the development of what are now known as the balance equations (Charney 1962; Lorenz 1960; Gent and McWilliams 1983). These equations are also formally valid in the limit of small Froude number F ($=V/NH$, where N and H are typical values of the buoyancy frequency and height scale, respectively) and large R (McWilliams 1985) as well as for flows including frontal structures in which neither F nor R is small (Gent et al. 1993); therefore, we may reasonably expect the balance equations to be accurate throughout a significant area of the $R-F$ plane. The balance equations result from a consistent truncation of the divergence and vorticity equations, although the balance equations themselves are not consistently derivable by taking the divergence and curl of a particular set of momentum equations. There are many other forms of intermediate models in which the physics ranges in complexity between that characteristic of the quasigeostrophic equations and that embodied within the so-called primitive equations, the most currently prevalent of which would probably be the semigeostrophic system of Hoskins and Bretherton (1972). The balance equations have, however, been shown to be the most accurate of the full range of such intermediate models for the limited number of flows for which explicit comparisons have been made (McWilliams and Gent 1980; Gent and McWilliams 1982; Norton et al. 1986; Barth et al. 1990; Allen et al. 1990; Snyder et al. 1991; Whitaker 1993).

Differences between a primitive equation solution and an intermediate model solution are unambiguously attributable to the terms neglected in the intermediate model. Snyder et al. (1991) have demonstrated that a prognostic integration of the semigeostrophic set of equations leads to large differences in the structural

* Current affiliation: National Center for Atmospheric Research,† Boulder, Colorado.

† The National Center for Atmospheric Research is sponsored by the National Science Foundation.

Corresponding author address: Dr. W. Richard Peltier, Department of Physics, University of Toronto, 60 St. George Street, Toronto, Ontario M5S 1A7, Canada.

characteristics of the predicted flows because of an accumulation over time of this error. An alternate method of testing a set of intermediate equations is to apply them diagnostically to the solutions delivered by a more fundamental set of equations. In this way one can, at any time, determine the degree to which the total flow is balanced, with the definition of what constitutes balance being provided by the intermediate model. This is the approach we will follow in the present study.

Recently, there have been attempts to confute the existence of a globally invariant slow manifold (where "slow" refers to those low-frequency motions that evolve on advective timescales) for various sets of equations (Warn and Menard 1986; Lorenz and Krishnamurthy 1987; Lorenz 1992) by demonstrating the existence of counterexamples wherein the solutions invariably develop a high-frequency gravity wave component. The concept of a fuzzy manifold (Warn and Menard 1986), in which the solution is imagined to remain in a region "close" to some hypothetical globally invariant slow manifold, may be a more appropriate idea but one that is of limited utility because of its fuzziness. If one is considering large-scale flows that generate unbalanced motion whose feedback on the large-scale flow is relatively insignificant, then balanced models will describe the flow adequately. If, on the other hand, the unbalanced flow can feed back significantly on the dynamics of the large-scale flow then balanced models will no longer be applicable. Given the success of modeling large-scale motions using nonlinear normal-mode initialization (Baer and Tribbia 1977; Leith 1980), inherent in which is the assumption of the existence of a slow manifold, it would appear that any fuzziness of the manifold does not usually significantly alter the large-scale motion. On the other hand, it is now well known on the basis of general circulation model experiments that the gravity wave drag exerted on the large-scale flow by the breaking of internal waves, especially above mountainous topography (Peltier and Clark 1979), appears to play an important role; so the unimportance of gravity waves cannot be universally true.

As well as this topographic source of unbalanced motion, however, there is also the possibility that when a synoptic-scale disturbance develops mesoscale frontal zones, significant imbalance might develop along the entire length of the frontal zone, particularly since two-dimensional frontal studies have revealed substantial frontal imbalance of advective origin (e.g., Keyser and Pecnick 1985; Reeder and Keyser 1988). The present analysis will explore the unbalanced motion in a fully three-dimensional front. It remains unclear as to what impact this frontal imbalance might have on frontal evolution. Gravity wave generation during both viscid and inviscid frontal collapse has been a focus of considerable research (Ley and Peltier 1978; Gall et al. 1987; Garner 1989; Snyder et al. 1993). Mesoscale pressure anomalies characteristic of internal gravity waves have been observationally associated with weather fronts for some time [Herron et al. (1969); Eckerman and Vincent (1993), who report unbalanced vertical velocities attributable to gravity waves of magnitude as large as 0.5 m s^{-1}]. Numerical simulations of frontal evolution, however, pose a technical problem. Models that employ a diffusive parameterization of subgrid-scale processes and whose physics allow for the existence of internal waves will inevitably predict the collapse of the surface front to a diffusion limited scale that subsequently provides a constant and perhaps intense mechanism for internal wave generation. The ad hoc nature of such diffusive parameterizations suggests that internal waves excited in the model in this circumstance may be numerical artifacts rather than the product of a physically realistic spontaneous emission process. Snyder et al. (1993) provide a careful study of these issues.

It is the purpose of this present research to develop a set of balance equations that are consistent with the three-dimensional anelastic equations and that may be employed to diagnostically determine the degree of balance in a particular class of atmospherically important flows. The numerical simulation on which we have chosen to focus is that of a midlatitude synoptic-scale baroclinic wave life cycle, recently described in detail by Bush and Peltier (1994). The scaling variables are chosen in accordance with this large-scale flow. The frontal zones that develop in the course of the evolution will be shown to become a source of unbalanced motion on both fast and slow timescales as the nonlinear wave begins to occlude.

As the motions that develop in the simulation have neither small Rossby nor small Froude numbers, we do not expect our scaling analysis to be formally valid over an especially wide range of times during the evolution of the wave. Similar analyses (e.g., Gent et al. 1993), however, indicate that an expansion in Rossby or Froude number works reasonably well, even if the expansion parameter is O(1). We do not, therefore, implement an explicit frontal scaling in those mesoscale frontal regions that eventually arise during the life cycle of the parent wave.

The plan of the paper is as follows: in section 2 we will present the scaling analysis and the derivation of the set of balance equations to be employed thereafter; in section 3 we provide a brief description of the evolution of the baroclinic wave as predicted by the full anelastic model; section 4 contains a full discussion of the diagnostic analysis procedure that we will employ; results are discussed in section 5; and our conclusions are presented in section 6.

2. Scale analysis of the anelastic equations

Spall and McWilliams (1992, hereafter SM) have recently developed a scale analysis technique to determine the degree of imbalance in the shallow-water

equations for a wide range of Rossby and Froude numbers, and their methods will be extended here for application to the anelastic equations. Their scaling predictions for the amplitude of the unbalanced motion appear to agree well with the results from numerical simulation in the regimes $R \ll 1$, $F^2 < R$ and $R > 1$, $F \ll 1$. Our goal here will be to develop a similar methodology for application to the three-dimensional anelastic primitive equations originally derived by Batchelor and more formally justified by Ogura and Philips. The anelastic equations are similar to the Boussinesq equations, but they retain additional, and often important, effects associated with the existence of a vertical density stratification in the background state. This system may be written in quasi-flux form as

$$\bar{\rho}_a \frac{D}{Dt}(\bar{\rho}_a \mathbf{u}) - (\bar{\rho}_a w)(\bar{\rho}_a \mathbf{u}) \frac{d z \bar{\rho}_a}{\bar{\rho}_c} + f \bar{\rho}_a \mathbf{k} \times \bar{\rho}_a \mathbf{u} = -\bar{\rho}_a \nabla_{3D} p + \bar{\rho}_a \rho \mathbf{g} \quad (2.1a,b,c)$$

$$\nabla_{3D} \cdot (\bar{\rho}_a \mathbf{u}) = 0 \quad (2.2)$$

$$\frac{1}{c_s^2} \frac{Dp}{Dt} - \frac{D\rho}{Dt} = 0, \quad (2.3)$$

where the last equation has implicitly assumed adiabatic motion of an ideal gas whose adiabatic sound speed is c_s. The operator ∇_{3D} is employed above and in all that follows to represent the three-dimensional (x, y, z) gradient operator. We denote by the operator ∇ the horizontal (x, y) form of the gradient operator. Here $\bar{\rho}_a(z)$ is an assumed adiabatic background density stratification, and ρ is the total density field. The second term on the left-hand side of the momentum equation is a by-product of assimilating the background density field under the material derivative. The numerical simulation we will analyze has a constant value for the Coriolis parameter $f = 10^{-4}\ \mathrm{s}^{-1}$, so that Rossby wave propagation effects associated with baroclinic wave development will not be captured in the solution.

The anelastic approximation to the continuity equation (2.2) (which, although retaining the influence of compressibility in the basic state, has filtered acoustic waves by setting the local time derivative of density to zero), combined with the fact that most large-scale motions are dominated by horizontal flow, suggests the use of a poloidal/toroidal decomposition of the solenoidal mass flux vector $\bar{\rho}_a \mathbf{u}$, in the form:

$$\bar{\rho}_a \mathbf{u} = \nabla_{3D} \times \mathbf{k}\Psi + \epsilon \nabla_{3D} \times (\nabla_{3D} \times \mathbf{k}\chi), \quad (2.4)$$

in which the toroidal and poloidal potentials (Ψ and χ, respectively) are analogous to the streamfunction and velocity potential for incompressible flow. They will be referred to here as the rotational and divergent components of the mass flux, respectively; ϵ is considered to be a small parameter whose definition will become precise after we have formed the balance equations.

We nondimensionalize equations (2.1)–(2.3) with the following parameters appropriate to synoptic-scale motion:

$$(x, y) \sim L, \quad z \sim H_s, \quad (u, v) \sim V, \quad w \sim \epsilon \frac{H_s V}{L} \quad t \sim \frac{L}{V}$$

$$\Psi \sim \rho_0 V L, \quad \chi \sim \epsilon \rho_0 H_s V L. \quad (2.5a)$$

Furthermore, we decompose the density and pressure fields into nondimensional components as follows:

$$\rho = \rho_0 \bar{\rho}_a(z) + \rho_0 \frac{N^2 H_s}{g} \bar{\rho}_s(z) + \rho_0 \frac{R}{B}[1, R] \frac{N^2 H_s}{g} \rho(x, y, z, t)$$

$$p = \rho_0 g H_a \bar{p}_a(z) + \rho_0 (N H_s)^2 \bar{p}_s(z) + \rho_0 f V L [1, R] p(x, y, z, t). \quad (2.5b)$$

The scaling for each component of these two fields has been chosen to ensure hydrostatic balance between them. Note that we use a common vertical scale H_s for the background stratification and its deviation, although in other circumstances it might be more appropriate to distinguish them. In the above expansions and throughout this paper we use the notation $[a, b]$ to denote the larger of the arguments. The Burger number B is defined as

$$B = \left(\frac{R}{F}\right)^2 = \left(\frac{N H_s}{f L}\right)^2. \quad (2.6)$$

For the synoptic-scale motions on which we intend to focus here, the following values for the above-referenced scales are selected: $L \sim 3590$ km (the wavelength of the baroclinic wave, see below), $V \sim 10\ \mathrm{m\ s}^{-1}$, $N \sim 10^{-2}\ \mathrm{s}^{-1}$, $H_s \sim 9$ km, $\rho_0 \sim 1\ \mathrm{kg\ m}^{-3}$. The adiabatic density profile has a characteristic scale height $H_a \sim 28$ km, while stratification effects have a characteristic height of $H_s \sim 9$ km. We do not know that the above scaling will remain apt when the baroclinic wave develops frontal zones. However, we anticipate that the majority of the unbalanced motion will occur in the mesoscale frontal zones (both surface and upper level), and we employ a multiple timescale and length-scale expansion to capture this behavior.

The dependent variables may then be decomposed into what we will here refer to as balanced and unbalanced (primed) components; namely,

$$\Psi = \Psi(x, y, z, t) + \frac{\mu}{\omega R} \Psi'(\nu x, \nu y, \nu z, \omega t)$$

$$\epsilon\chi = \epsilon\chi(\delta x, \delta y, \delta z, t) + \frac{\mu}{\nu} \chi'(\nu x, \nu y, \nu z, \omega t)$$

$$\rho = \rho(x, y, z, t) + \mu\Pi\rho'(\nu x, \nu y, \nu z, \omega t)$$

$$p = p(x, y, z, t) + \mu\Pi p'(\nu x, \nu y, \nu z, \omega t), \quad (2.7)$$

where the relative amplitude for the unbalanced pressure and density fluctuations is

$$\Pi = \frac{\omega RB[1, \nu^2]}{[1, R][1, B]}.$$

In employing the above expansions we have introduced a time and length scale separation between the unbalanced and the balanced components, whose relative scales are characterized by frequency ω and wavenumber ν. A relative wavenumber δ has also been introduced to distinguish between the length scales of the balanced divergent flow and the balanced rotational flow. We expect δ to be of O(1) in the results to be presented here, however, since the majority of the divergent motion in a baroclinic wave is on the same scale as the wave itself (at least before the fronts become intense). The factors multiplying the unbalanced components have been chosen so as to be consistent with the linear internal wave balances in the following equations.

a. The vorticity equation

After substitution of (2.4), (2.5), and (2.7) into (2.1), we construct the vorticity equation by forming $\mathbf{k}\cdot\nabla_{3D}\times$ (2.1) to obtain

$$\begin{aligned}\bar{\rho}_a\nabla^2\Psi_t = \bar{\rho}_a\delta^3\frac{\epsilon}{R}\nabla^2\chi_z - J(\nabla^2\Psi, \Psi) - \epsilon\delta^2\Big[&\nabla(\nabla^2\Psi)\cdot\nabla\chi_z - \nabla^2\chi\nabla^2\Psi_z + \delta\nabla^2\Psi\nabla^2\chi_z - \nabla\Psi_z\cdot\nabla(\nabla^2\chi)\\ &- \frac{d_z\bar{\rho}_a}{\bar{\rho}_a}(\delta\nabla\Psi\cdot\nabla(\nabla^2\chi) + \nabla^2\Psi\nabla^2\chi)\Big] + \epsilon^2\Big[-\delta^6 J(\chi_{zz}, \nabla^2\chi) + \delta^5\frac{d_z\bar{\rho}_a}{\bar{\rho}_a}J(\nabla^2\chi, \chi_z)\Big]\\ &- \mu\Big[-\bar{\rho}_a\frac{\nu^2}{R}\nabla^2\Psi_t' + \bar{\rho}_a\frac{\nu^2}{R}\nabla^2\chi_z'\Big] + O(\mu, \mu\epsilon, \mu^2). \quad (2.8)\end{aligned}$$

The Jacobian operator is $J(a, b) = \partial_x a\partial_y b - \partial_y a\partial_x b$. In Eq. (2.8) and others to follow, we retain three sets of terms: 1) those comprising the balance components for this system (all terms up to and including $O(\epsilon, R)$; 2) higher-order advective terms of $O(\epsilon^2, \epsilon R)$; and 3) linear gravity wave contributions of $O(\mu)$ that are forced by the higher-order advective terms. Because they are not germane to the arguments below, we do not retain any nonlinear terms involving unbalanced components.

The balanced vorticity equation for the anelastic system [which consists of those terms in (2.8) of $O(1, \epsilon, R)$] is equivalent to that of the three-dimensional hydrostatic system employed by Norton et al. (1986) with the addition of the two terms involving the effects of background density variation in the form $\partial_z\bar{\rho}_a/\bar{\rho}_a$. From the $O(\mu)$ terms, one may identify the linear gravity wave terms multiplied by a factor of $\mu\bar{\rho}_a\nu^2R$; namely,

$$-\nabla^2\Psi_t' + \nabla^2\chi_z'.$$

b. The divergence equation

The divergence equation is obtained by forming $\nabla_{3D}\cdot$(2.1a,b,c) and assuming hydrostatic balance to eliminate the vertical pressure gradient. In this manner, the continuity equation (2.2) may be used to eliminate the time derivative of the three-dimensional mass flux divergence. The result is

$$\begin{aligned}&\bar{\rho}_a\nabla^2 p + [\bar{\rho}_a\nabla^2\Psi + 2RJ(\Psi_y, \Psi_x)]\\ &- \epsilon R\Big[2\delta^3 J(\Psi_x, \chi_{xz}) - 2\delta^3 J(\Psi_y, \chi_{yz})\\ &+ 2\delta^3 J(\Psi_z, \nabla^2\chi) + \frac{d_z\bar{\rho}_a}{\bar{\rho}_a}\delta^3 J(\nabla^2\chi, \Psi)\\ &- \delta^2 d_z\bar{\rho}_a\nabla^2\chi_t\Big] + \mu\Big[\Pi\nu^2\bar{\rho}_a[1, R]\nabla^2 p'\\ &+ \frac{\nu^2\bar{\rho}_a}{\omega R}\nabla^2\Psi' - \nu\omega R d_z\bar{\rho}_a\nabla^2\chi_t'\Big]\\ &+ O(\mu, \mu\epsilon, \mu^2) = 0. \quad (2.9)\end{aligned}$$

The balanced form of this equation consists of those terms of $O(1, \epsilon, R)$ that describe a simple gradient wind balance.

The linear gravity wave terms at $O(\mu)$ may be identified as

$$\frac{\nu\Pi[1, R]}{\omega R}\nabla^2 p' + \frac{\nu}{(\omega R)^2}\nabla^2\Psi' - \frac{d_z\bar{\rho}_a}{\bar{\rho}_a}\nabla^2\chi_t',$$

after factoring $\mu\nu\omega R\bar{\rho}_a$. It is these terms that, forced by the terms of $O(\epsilon R)$, govern the evolution of forced internal gravity waves in the system (see section 2e).

c. Energy equation

The internal energy equation (2.3) becomes, after substitution of the above expansions,

$$\frac{B\delta^2}{R[1,R]}\left[\left(d_z\bar{\rho}_a-\left(\frac{M}{V}\right)^2 gH_a d_z\bar{\rho}_a\right)+\left(d_z\bar{\rho}_s-\left(\frac{M}{V}\right)^2 gH_s d_z\bar{\rho}_s\right)\right]\epsilon\nabla^2\chi-\bar{\rho}_a\rho_t$$
$$-J(\rho,\Psi)+\left(\frac{M}{V}\right)^2(NH_s)^2[\bar{\rho}_a p_t+J(p,\Psi)]$$
$$-\epsilon\left[\delta^2\nabla\rho\cdot\nabla\chi_z+\delta^2\rho_z\nabla^2\chi+\left(\frac{M}{V}\right)^2(NH_s)^2\right.$$
$$\left.\times[\bar{\rho}_a p_t+J(p,\Psi)+\delta^2\nabla p\cdot\nabla\chi_z+\delta^2 p_z\nabla^2\chi]\right]$$
$$+\mu\left[\frac{B\nu}{R[1,R]}\left[\left(d_z\bar{\rho}_a-\left(\frac{M}{V}\right)^2 gH_a d_z\bar{\rho}_a\right)+\left(d_z\bar{\rho}_s-\left(\frac{M}{V}\right)^2 gH_s d_z\bar{\rho}_s\right)\right]\nabla^2\chi'-\Pi\omega\bar{\rho}_a\rho_t'\right.$$
$$\left.+\left(\frac{M}{V}\right)^2(NH_s)^2\Pi\omega\bar{\rho}_a p_t'\right]=0, \quad (2.10)$$

in which $M = V/c_s$ is the Mach number.

This equation, although quite different from the mass equation of the shallow-water system given in SM, nevertheless has the same type of linear operator acting on $\epsilon\chi$, appropriately modified to include vertical variations of the background density. The Mach number terms, which are of nondimensional order 0.1 for $c_s = 337$ m s^{-1}, express the weakly compressible nature of the energy equation (2.3).

The linear gravity wave terms may be extracted from the $O(\mu)$ term in (2.10) after factoring $\mu\Pi\omega\bar{\rho}_a$; namely,

$$\frac{B\nu}{\Pi\omega\bar{\rho}_a R[1,R]}\left[\left(d_z\bar{\rho}_a-\left(\frac{M}{V}\right)^2 gH_a d_z\bar{\rho}_a\right)+\left(d_z\bar{\rho}_s-\left(\frac{M}{V}\right)^2(NH_s)^2 d_z\bar{\rho}_s\right)\right]\nabla^2\chi'$$
$$-\rho_t'+\left(\frac{M}{V}\right)^2 gH_s p_t'.$$

d. The omega equation

To construct the omega equation for our scaled anelastic system, we use the (appropriately nondimensionalized) hydrostatic balance relation

$$p_z+\rho=0, \quad (2.11)$$

and then form a linear combination of Eqs. (2.8)–(2.11). In particular the operation

$$\frac{[1,R]}{\bar{\rho}_a}\nabla^2(2.10)-[1,R]\nabla^2\partial_t(2.11)+\partial_{tz}\left(\frac{(2.9)}{\bar{\rho}_a}\right)+\partial_z\left(\frac{(2.8)}{\bar{\rho}_a}\right)$$

yields

$$\left[-\frac{B\delta^2}{\bar{\rho}_a}\left[\left(d_z\bar{\rho}_a-\left(\frac{M}{V}\right)^2 gH_a d_z\bar{\rho}_a\right)+\left(d_z\bar{\rho}_s-\left(\frac{M}{V}\right)^2(NH_s)^2 d_z\bar{\rho}_s\right)\right]\nabla^2+\partial_{zz}\right]\frac{\epsilon}{R}\delta^2\nabla^2\chi$$
$$=\left[\left(\frac{J(\nabla^2\Psi,\Psi)}{\bar{\rho}_a}\right)_z-2R\left(\frac{J(\Psi_y,\Psi_x)_t}{\bar{\rho}_a}\right)_z\right]-\frac{\nabla^2 J(\rho,\Psi)}{\bar{\rho}_a}+\left(\frac{M}{V}\right)^2(NH_s)^2\left[\nabla^2 p_t+\frac{\nabla^2 J(p,\Psi)}{\bar{\rho}_a}\right]$$
$$+\epsilon\delta^2\left[\frac{\nabla^2(\nabla\rho\cdot\nabla\chi_z)}{\bar{\rho}_a}+\frac{\nabla^2(\rho_z\nabla^2\chi)}{\bar{\rho}_a}+\left(\left(\frac{\nabla(\nabla^2\Psi)\cdot\nabla\chi_z}{\bar{\rho}_a}\right)_z-\left(\frac{\nabla^2\chi\nabla^2\Psi_z}{\bar{\rho}_a}\right)_z+\left(\delta\frac{\nabla^2\chi_z\nabla^2\Psi_z}{\bar{\rho}_a}\right)_z\right.\right.$$
$$\left.-\left(\frac{\nabla\Psi_z\cdot\nabla(\nabla^2\chi)}{\bar{\rho}_a}\right)_z\right)-\left(\frac{d_z\bar{\rho}_a}{\bar{\rho}_a^2}\right)_z[\delta\nabla\Psi\cdot\nabla(\nabla^2\chi)+\nabla^2\Psi\nabla^2\chi]$$
$$\left.-\left(\frac{d_z\bar{\rho}_a}{\bar{\rho}_a^2}\right)[\delta(\nabla\Psi\cdot\nabla(\nabla^2\chi))_z+(\nabla^2\Psi\nabla^2\chi)_z]+\left(\frac{M}{V}\right)^2\frac{gH_s}{\bar{\rho}_a}[\nabla^2(\nabla p\cdot\nabla\chi_z)+\nabla^2(\nabla^2\chi p_z)]\right]$$
$$+\epsilon R\left[-\delta^3\partial_z\left(\frac{J(\Psi_x,\chi_{xz})_t}{\bar{\rho}_a}\right)+\delta^3\partial_z\left(\frac{J(\Psi_y,\chi_{yz})_t}{\bar{\rho}_a}\right)-\delta^3\partial_z\left(\frac{J(\Psi_z,\nabla^2\chi)_t}{\bar{\rho}_a}\right)+\delta^3\partial_z\left(\frac{d_z\bar{\rho}_a}{\bar{\rho}_a}J(\nabla^2\chi\cdot\Psi)_t\right)\right.$$
$$\left.-\delta^2\partial_z\left(\frac{d_z\bar{\rho}_a}{\bar{\rho}_a}\nabla^2\chi_t\right)\right]+\epsilon^2\left[-\delta^6\partial_z\left(\frac{J(\chi_{zz},\nabla^2\chi)}{\bar{\rho}_a}\right)+\delta^5\partial_z\left(\frac{d_z\bar{\rho}_a}{\bar{\rho}_a^2}J(\nabla^2\chi,\chi_z)\right)\right]+O(\mu). \quad (2.12)$$

It is through this rather onerous equation that we are obliged to calculate the balanced divergent potential after neglecting the terms of $O(\mu, \epsilon^2, \epsilon R)$. Note that for brevity we have neglected a number of higher-order terms retained in (2.8)–(2.10).

e. Amplitude estimates

Estimates for the amplitudes of the balanced divergent field and the unbalanced fields may be obtained based on the leading-order balances apparent in Eqs. (2.9) and (2.12).

Recall that the dynamical variables have been nondimensionalized in deriving equation (2.12) and are therefore of O(1); the terms composing the leading-order balance in (2.12) are then determined by the nondimensional coefficients R, B, δ, and ϵ, which are not necessarily of O(1). The leading-order balance in Eq. (2.12) results in an estimate for the scaling parameter ϵ; namely,

$$\epsilon = \frac{R[1, R]}{\delta^2[1, \delta^2 B]}. \tag{2.13}$$

This is equivalent to the shallow-water multiscale estimate of SM [see their Eq. (22)], and it has the limits of $\epsilon \sim R/\delta^2$ for $R \ll 1$ and $\epsilon \sim F^2/\delta^4$ for $F \ll 1$ and $R \geqslant 1$.

We will now derive two estimates (based on slow or fast timescales) of the unbalanced amplitude μ. An amplitude estimate for the higher-order advective motion, which has inherently the same time and length scales of the lower-order balanced motion, may be obtained from the omega equation at $O(\epsilon R, \epsilon^2, \mu)$ after setting $\omega = \nu = 1$. From Eq. (2.12) the dominant balance at this order is

$$\begin{aligned}
&\left[\frac{B}{\bar{\rho}_a}\left[\left(d_z\bar{\rho}_a - \left(\frac{M}{V}\right)^2 gH_a d_z\bar{\rho}_a\right) + \left(d_z\bar{\rho}_s - \left(\frac{M}{V}\right)^2 gH_s d_z\bar{\rho}_s\right)\right]\nabla^2 - \partial_{zz}\right]\frac{\mu_s}{R}\nu^3\nabla^2\chi' \\
&\sim \epsilon R\left[-2\delta^3\partial_z\left(\frac{J(\Psi_x, \chi_{xz})_t}{\bar{\rho}_a}\right) + \delta^3\partial_z\left(\frac{J(\Psi_y, \chi_{yz})_t}{\bar{\rho}_a}\right) - \delta^3\partial_z\left(\frac{J(\Psi_z, \nabla^2\chi)_t}{\bar{\rho}_a}\right)\right. \\
&\quad \left. + \delta^3\partial_z\left(\frac{d_z\bar{\rho}_a}{\bar{\rho}_a}J(\nabla^2\chi, \Psi)_t\right) - \delta^2\partial_z\left(\frac{d_z\bar{\rho}_a}{\bar{\rho}_a}\nabla^2\chi_{tt}\right)\right] \\
&\quad + \epsilon^2\left[-\delta^6\partial_z\left(\frac{J(\chi_{zz}, \nabla^2\chi)}{\bar{\rho}_a}\right) - \delta^5\partial_z\left(\frac{d_z\bar{\rho}_a}{\bar{\rho}_a}J(\nabla^2\chi, \chi_z)\right)\right],
\end{aligned} \tag{2.14}$$

from which a scaling estimate for the single-scale amplitude μ_s may be determined, by balancing the nondimensional coefficients, to be

$$\mu_s = \frac{\epsilon R[\delta^3\epsilon, R]\delta^2[1, \delta]}{\nu^3[1, B]}. \tag{2.15}$$

Equation (2.15) provides an upper limit to the amplitude of the higher-order advective corrections to the balanced motion.

We note that it is possible to invert Eq. (2.14) to solve for $\mu_s\nabla^2\chi'$ in a manner analogous to that described in section 4 for $\epsilon\nabla^2\chi$. In this manner one could determine the contributions by the balanced motion to the unbalanced motion, although this course is not pursued in the present analysis. Instead we will determine unbalanced motions by subtracting the balanced component from the total.

To estimate the multiscale amplitude of the fast unbalanced motion, we assume that gravity waves dominate the unbalanced motion field and that they are being forced by advective terms of higher order than those retained in the balance equations. We further assume that initialization gravity waves have been damped sufficiently so as not to influence the solution, so that all gravity waves present are being forced by the higher-order advective terms. We here exploit the multitime and length-scale expansion and retain the frequency ratio ω and the wavenumber ratios δ and ν. Since all the advective terms in the vorticity and energy equations are included within the framework of balanced dynamics, one must consider the higher-order advective terms of the divergence equation (or, equivalently, the omega equation, since this contains the same information) as being those that force the gravity waves. We consider the balance between the amplitude factor multiplying the linear gravity wave terms with the advective terms of $O(\epsilon R)$ in the divergence equation (2.9) to estimate the multiscale unbalanced amplitude μ_m as

$$\mu_m = \frac{R[1, R][1, \delta]}{\omega\nu[1, \delta^2 B]}. \tag{2.16}$$

This is, of course, an upper bound insofar as the balance equations are a valid leading order description and the actual forcing of the gravity waves may be deferred to higher order. The dispersion relation for linear gravity waves under this particular scaling may be obtained by

a linearization of the nondimensional anelastic equations, assuming wavelike solutions. It may be written as

$$\omega = \left[\frac{\nu^{1/2}}{R}, \frac{\nu}{F}\right]. \tag{2.17}$$

3. The numerical simulation

The reader is referred to Bush and Peltier (1994) for a complete description of the model simulation, but we will reiterate the salient points here for completeness. The model equations are the same three-dimensional, nonhydrostatic anelastic system (e.g., Clark 1977) that has been previously employed in baroclinic wave life cycle analyses on the basis of the assumption of rigid-lid upper boundary conditions (Polavarapu and Peltier 1990). The baroclinic wave life cycle that we will consider here was initialized using a mixed baroclinic/barotropic mean state (Fig. 1), upon which we superimpose the structure of the fastest-growing mode of linear theory calculated from a Galerkin-type stability analysis of the mean state. This life cycle is more realistic than those analyzed by Polavarapu and Peltier (1990) in that both tropospheric and stratospheric components are included in the model. From the sequence of analyses presented in Bush and Peltier (1994), we have selected the intermediate baroclinicity case for discussion here because of the strong frontal development that we expect may lead to significant unbalanced motion. The model domain has x, y, z dimensions of 3590, 8000, 18 km, respectively, and the integrations were performed with a constant Coriolis parameter. The model has a channel configuration with periodic boundary conditions in the zonal direction, the zonal dimension being chosen to be equal to the wavelength of the fastest-growing linear normal mode. The spatial discretization is of moderate resolution with Δx, Δy, $\Delta z = 52.8, 63.5, 0.6$ km, and a horizontal ∇^6 diffusion operator is employed in the interior of the model domain. Time differencing is centered and second-order accurate, and an Euler backward step is performed every 20 time steps to prevent the buildup of splitting errors. We note here that this choice of grid resolution should be roughly consistent with the aspect ratio of the expected frontal zones; namely,

$$\Delta z \sim \frac{f}{N}\Delta L \tag{3.1}$$

in which Δz and ΔL are the vertical and horizontal grid scales and f/N is a geostrophic estimate for the slope of the front that will be induced by the growth of the baroclinic wave. Using typical midlatitude tropospheric values for f and N, we therefore anticipate adequate vertical resolution for any frontal motions (that inherently collapse to the horizontal grid scale) whose slope is not less than 0.01.

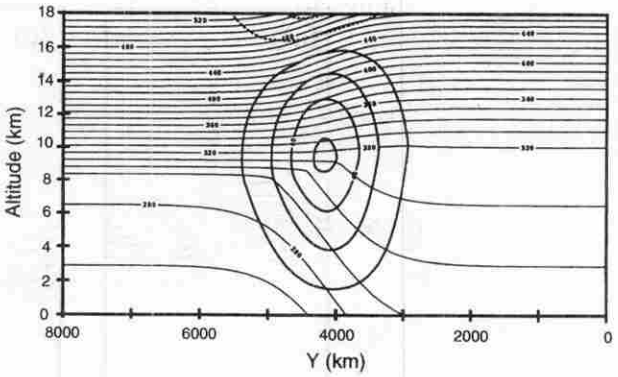

FIG. 1. The two-dimensional mean state used to grow a synoptic-scale baroclinic wave. Thick contours are zonal velocities and thin contours are potential temperature. The jet maximum is 57 m s^{-1}; a contour interval of 15 m s^{-1} is used starting at 10 m s^{-1}.

The vertical and meridional dimensions of the model have been chosen so that both upper- and lower-level frontogenesis may be examined without horizontal boundary influences. A deep tropopause fold forms during the life cycle of this wave [see Bush and Peltier (1994) for a detailed discussion], and it will be of interest in the present analysis to investigate the differences, if any, in the degree of imbalance that develops between the upper- and the lower-level fronts. Since tropopause folding events have to some degree been successfully analyzed using semigeostrophic theory (e.g., Reeder and Keyser 1988; Koshyk and Cho 1992), we expect the upper-level frontal motion to be, for the most part, balanced.

The evolution of the surface potential temperature is displayed in Fig. 2, and strong frontal zones are evident within this field by day 3.5 on the cold air side of the cusp from which the surface low is beginning to occlude. We will refer to this region of intense frontogenesis as the cold front. In Figs. 3a–d we display the energetics of the life cycle: the eddy kinetic energy, the vertical heat flux, and the vertical and horizontal Reynolds stresses [see Bush and Peltier (1994) for the definitions of these quantities]. The maximizing of the eddy kinetic energy (defined to be the departure from a zonal average) at day 6 demonstrates that this is the time at which the baroclinic wave saturates. The conversion terms, however, show that after day 4 the baroclinic conversion is no longer adding to disturbance energy. The occlusion of the low occurs at day 4, and we will see from the following analysis that the balanced rotational motion does not saturate until day 6. One may therefore conclude that the kink in the eddy kinetic energy at day 4 is associated with the actual pinch-off of the low and the termination of baroclinic growth, whereas day 6 marks the time at which the rotational flow around the low and along the cusp reaches its maximum amplitude. The inertia–gravity oscillation observed between days 0 and 2 is a conse-

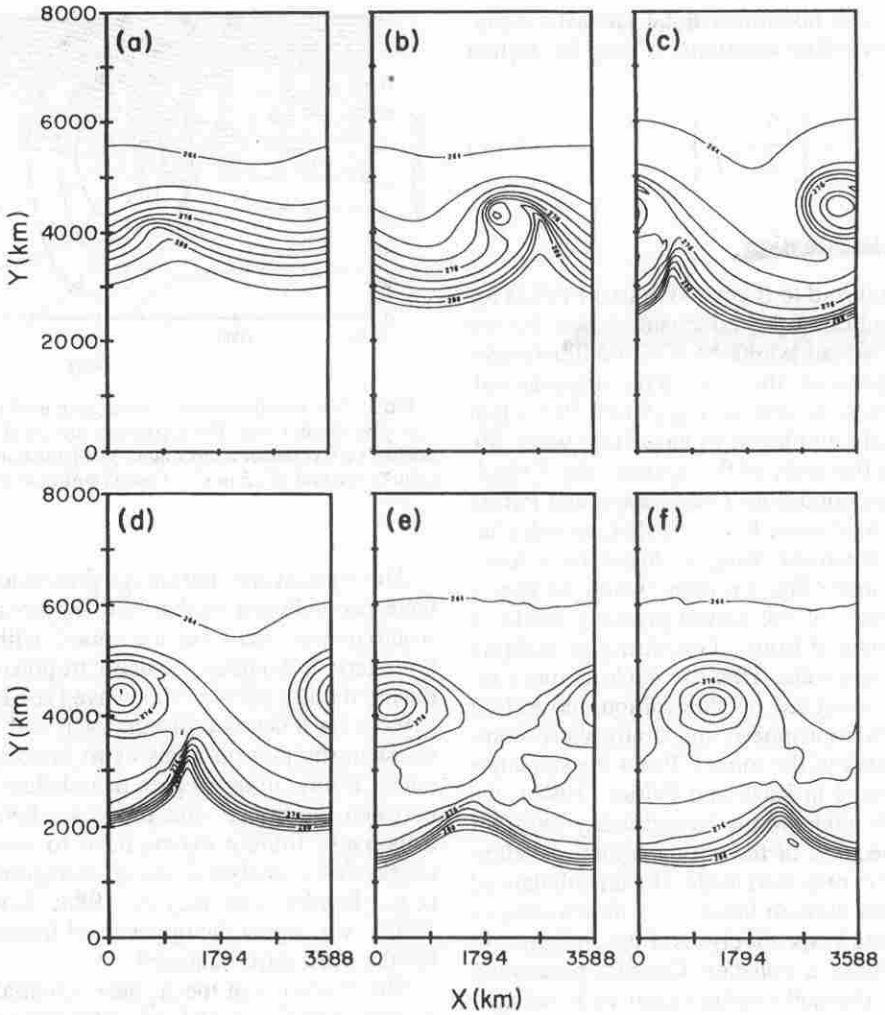

FIG. 2. Evolution of the surface potential temperature on (a) day 2.5, (b) day 3.5, (c) day 4.5, (d) day 5, (e) day 8.5, (f) day 9.5. The domain is 8000 × 3590 km. The maximum temperature in the core of the low is 282 K with a contour interval of 3 K.

quence of the fact that the simulation was initialized with an imperfectly balanced initial state. The reader is referred to Bush and Peltier (1994) for a discussion of the start-up phase of the time integrations.

4. The balance analysis

The procedure to be employed for the analysis of balance in the above described flow hinges on the approximation that the initial Ψ field is balanced, that is, that any unbalanced rotational motion present initially is included in the balanced component. Motivation for this approximation is based on the scaling argument (see SM) that small errors in Ψ (or p) result in negligible errors in the balanced divergent flow for $R \ll 1$ or $F \ll 1$. The unbalanced divergent flow, determined by the residual, then has a negligible error in this parameter regime. We will show in section 5 that these errors are indeed small in our analysis.

From Eq. (2.4), it is readily apparent that the poloidal/toroidal scalars may be computed from the model data by inversion of the (nondimensional) Poisson equations,

$$\nabla^2\Psi = -\bar{\rho}_a\zeta \tag{4.1}$$

$$\epsilon\nabla^2\chi = -\bar{\rho}_a w, \tag{4.2}$$

in which ζ is the vertical component of the vorticity vector and w is the vertical component of the velocity vector. The appendix discusses the boundary conditions required to solve Eqs. (4.1), (4.2), and all other Poisson equations that arise in the diagnostic analysis of balance.

The method of diagnostic analysis that we will employ is as follows (in this section, when we quote the number of a particular equation, we are referring to the balanced form of the equation):

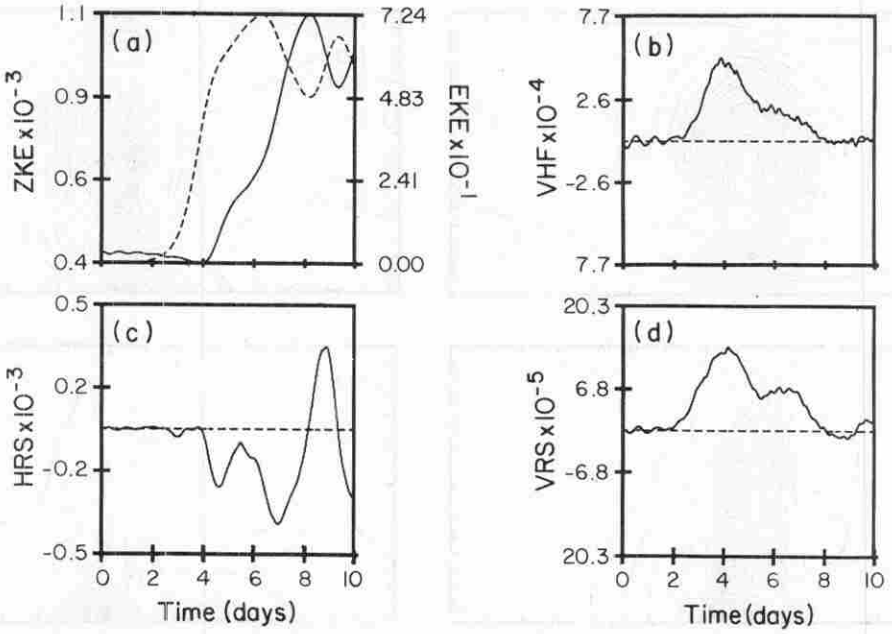

FIG. 3. Time series of the eddy kinetic energy and conversion terms during the nonlinear simulation. Terms were calculated every 1/2 hour of model time. (a) Zonal kinetic energy ($\times 10^3$) (solid) and eddy kinetic energy ($\times 10$), (b) vertical heat flux ($\times 10$), (c) horizontal Reynold's stress ($\times 10^3$), and (d) vertical Reynold's stress ($\times 10^5$). Units for (a) are joules per cubic meter and for (b), (c), and (d) are watts per cubic meter. Note the different scales on the plots.

1) solve (4.1) for Ψ from the model ζ; solve (4.2) for a first guess for χ from the model w;

2) assume the resulting Ψ is balanced;

3) solve Eq. (2.9) for the balanced pressure, then Eq. (2.11) for the balanced density field;

4) calculate the balanced χ field by iteration on Eqs. (2.8) and (2.12) until the solution for $\nabla^2\chi$ converges;

5) calculate the balanced vertical mass flux according to (4.2).

Convergence is monitored by calculating the maximum (normalized) difference between successive iterations. The solution generally converges to a difference that is less than 1% in less than 7 iterations, although later in the simulation (i.e., after day 5) convergence requires upward of 20 iterations. We will focus most of our attention in what follows on the unbalanced vertical mass flux field $\overline{\rho}_a w'$, since this is a field in which gravity waves have a strong signature. All unbalanced flux fields $\overline{\rho}_a \mathbf{u}'$ are calculated as differences between the model data $\rho_a \mathbf{u}_{\text{tot}}$ and the derived balanced fields $\overline{\rho}_a \mathbf{u}_b$; that is, $\overline{\rho}_a \mathbf{u}' = \overline{\rho}_a \mathbf{u}_{\text{tot}} - \overline{\rho}_a \mathbf{u}_b$.

To investigate the degree to which the initial fields are balanced, we ran the above-described diagnostic balance procedure on the two-dimensional mean state of Fig. 1 without the added perturbation. The mean state temperature and zonal velocity fields are constructed so as to be in thermal wind balance to machine precision (see Polavarapu and Peltier 1990), and a diagnostic analysis of the mean state revealed unbalanced meridional and vertical velocities that were 11 orders of magnitude smaller than those in the jet core. The normal-mode perturbation, however, is not balanced in any sense. The arbitrary magnitude of the perturbation is chosen such that the potential temperature deviation is 2 K, so we expect the unbalanced contributions to Ψ to be small. A balanced diagnostic analysis of the initial mean state plus the perturbation delivers the mass fluxes displayed in Fig. 4. The contributions to the horizontal mass fluxes are negligible compared to the magnitude of the flux in the jet core, and the amplitude of these unbalanced fluxes is a factor of 10 smaller than the unbalanced magnitudes that arise later in the simulation in the frontal zones. We therefore conclude that the magnitude of the initial imbalance is small compared to the unbalanced motion that evolves as the baroclinic wave matures. A similar diagnostic analysis of the inertio–gravity waves that arise upon time integration of the initial fields reveals that the amplitude of these waves is likewise an order of magnitude smaller than the unbalanced motion associated with the growing baroclinic wave.

The balance equations are solved diagnostically every 5 hours of model time for the first 6 days of a 10-day simulation. As in SM, estimates of our scaling

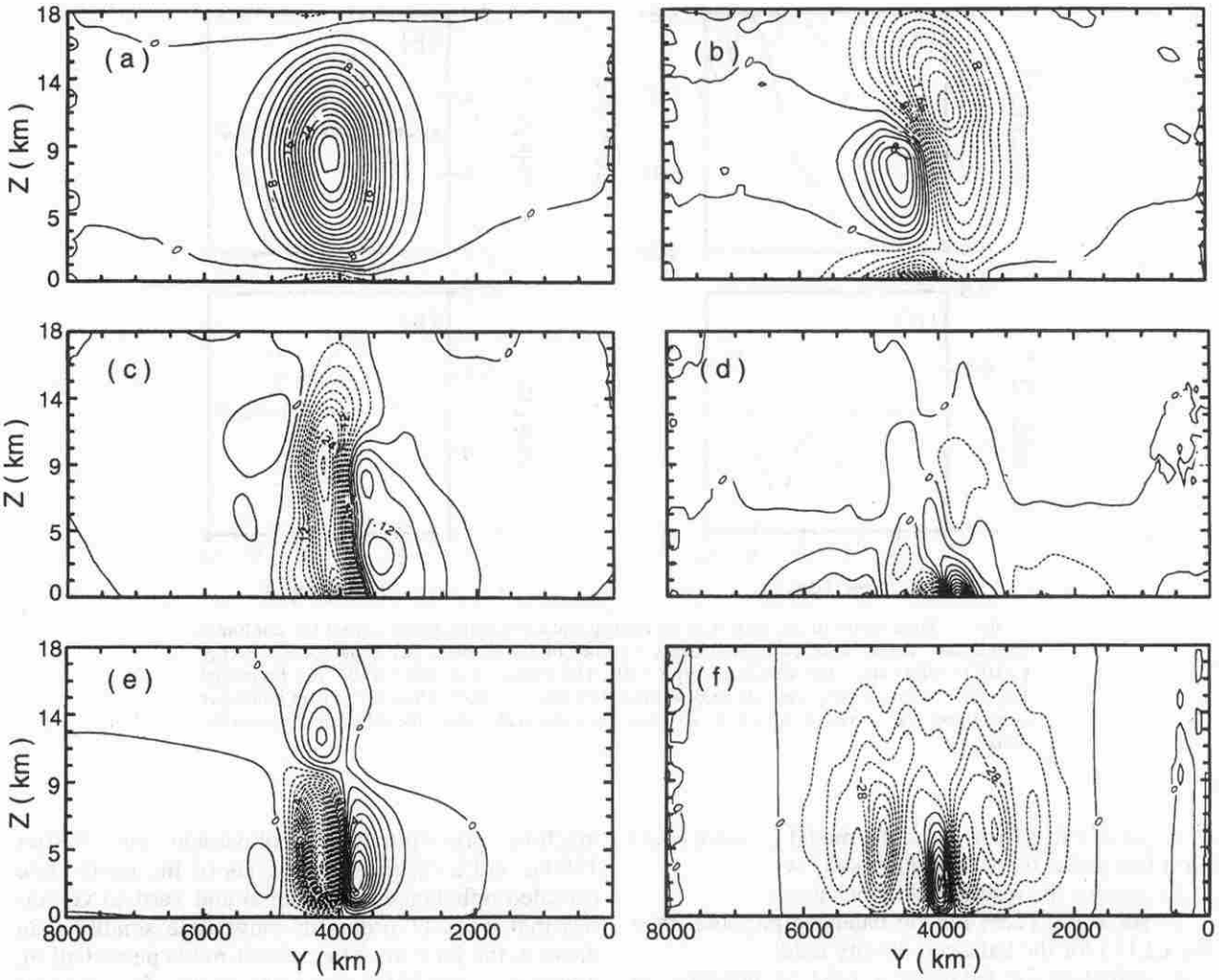

FIG. 4. Contour plots of a meridional cross section of the initial balanced and unbalanced mass flux fields $\bar{\rho}_a\mathbf{u}$. (a) $\bar{\rho}_a u$ (contour interval of 2 kg m^{-2} s^{-1}), (b) $\bar{\rho}_a u'$ (contour interval of 0.2 kg m^{-2} s^{-1}), (c) $\bar{\rho}_a v$ (contour interval of 0.03 kg m^{-2} s^{-1}), (d) $\bar{\rho}_a v'$ (contour interval of 0.1 kg m^{-2} s^{-1}), (e) $\bar{\rho}_a \omega$ (contour interval of 10^{-4} kg m^{-2} s^{-1}), and (f) $\bar{\rho}_a \omega'$ (contour interval of 7×10^{-4} kg m^{-2} s^{-1}).

parameters may be calculated according to the following expressions:

$$\epsilon = \frac{\langle \chi \rangle}{\langle \Psi \rangle} \tag{4.3}$$

$$\mu = \frac{\langle \chi' \rangle}{\langle \Psi \rangle}, \tag{4.4}$$

in which the norm $\langle \cdot \rangle$ of a field A is defined by

$$\langle A \rangle = \left[\frac{1}{(nx \cdot ny \cdot nz)} \sum_{i=1}^{nx} \sum_{j=1}^{ny} \sum_{k=1}^{nz} A^2(i,j,k) \right]^{1/2} \tag{4.5}$$

for an (nx, ny, nz) grid. We can also estimate the relative wavenumbers according to

$$\delta = \left(\frac{\langle \nabla^2 \chi \rangle \langle \Psi \rangle}{\langle \nabla^2 \Psi \rangle \langle \chi \rangle} \right)^{1/2} \tag{4.6}$$

$$\nu = \left(\frac{\langle \nabla^2 \chi' \rangle \langle \Psi \rangle}{\langle \nabla^2 \Psi \rangle \langle \chi' \rangle} \right)^{1/2}. \tag{4.7}$$

Similarly, the spatial distribution of the kinetic energy densities of the balanced rotational, the balanced divergent, and the unbalanced divergent components may be calculated according to

$$\mathrm{KE}_{br}(i,j,k) = \frac{1}{2} \bar{\rho}_a^{-1} (\Psi_x^2 + \Psi_y^2) \tag{4.8}$$

$$\mathrm{KE}_{bd}(i,j,k) = \frac{1}{2} \bar{\rho}_a^{-1} (\chi_{xz}^2 + \chi_{yz}^2 + (\nabla^2 \chi)^2) \tag{4.9}$$

$$\mathrm{KE}_{ud}(i,j,k) = \frac{1}{2} \bar{\rho}_a^{-1} (\chi_{xz}'^2 + \chi_{yz}'^2 + (\nabla^2 \chi')^2). \tag{4.10}$$

Note that the sum of (4.8)–(4.10) does not equal the total kinetic energy because of nonvanishing cross

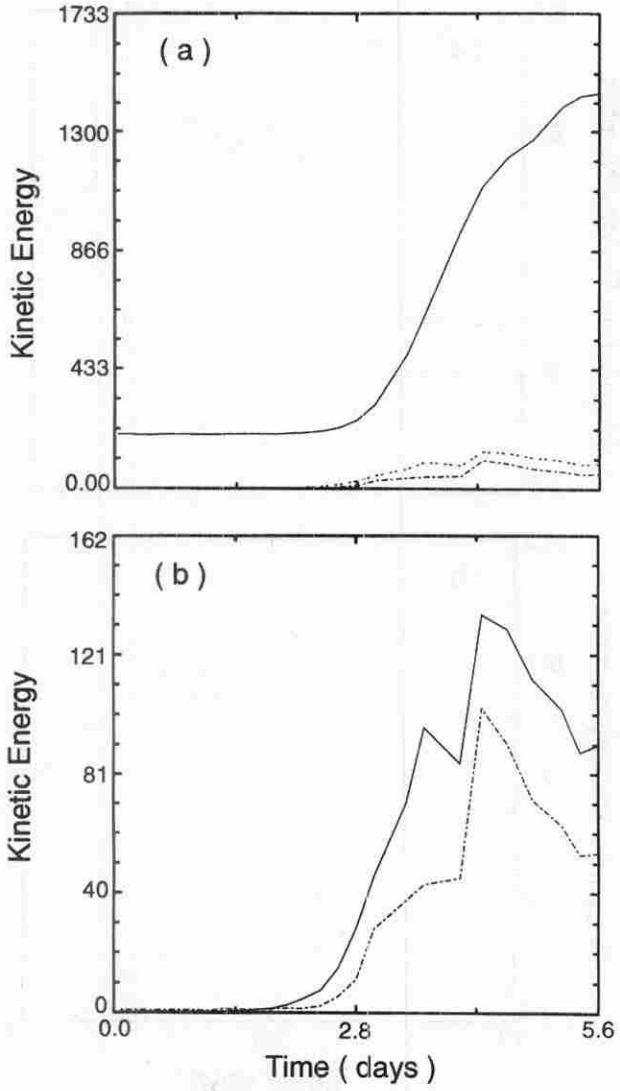

FIG. 5. (a) Time series of the balanced rotational kinetic energy (solid), the balanced divergent kinetic energy (dashed), and the unbalanced divergent kinetic energy (dash-dot). (b) An enlargement of the divergent components: balanced (solid) and unbalanced (dashed).

terms among components. From these expressions, we may estimate a characteristic frequency for each component according to

$$\omega = \frac{\langle \partial_t \mathrm{KE}(i,j,k) \rangle}{\langle \mathrm{KE}(i,j,k) \rangle}. \qquad (4.11)$$

5. Discussion

In this section we will apply the techniques of section 4 to the numerical simulation described in section 3 to determine the degree of balance characteristic of the life cycle of a synoptic-scale baroclinic wave. The kinetic energy components, calculated by application of Eqs. (4.8)–(4.10), are given in Fig. 5a, with an enlargement of the divergent components in Fig. 5b. The exponential growth of the wave is clearly evident in all components, the balanced rotational and balanced divergent components growing together and the unbalanced rotational component lagging by roughly 0.4 days during this phase of growth. This lag is consistent with the fact that the higher-order advective terms of $O(\epsilon R, \epsilon^2)$ are driving the unbalanced motion, so that when ϵ becomes large enough, μ will increase accordingly. The exponential growth of this baroclinic wave is a result of the conversion of available potential energy into rotational and divergent eddy kinetic energy. The degree to which available potential energy is converted into divergent kinetic energy depends on the zonal asymmetry of the flow. When the asymmetric flow is small compared to the symmetric flow, balanced dynamics will adequately describe the motion. As the asymmetry increases, however, balanced dynamics fails to capture the intense divergent motion, and we see an increase in the unbalanced components of the flow. We note that the balanced rotational kinetic energy does not saturate until day 6, although the balanced and unbalanced divergent kinetic energies appear to saturate well before this time at day 4, the time at which the baroclinic conversions reach their maximum. Since divergent/convergent motion is required to produce a vertical heat flux, it is no surprise that the divergent component of the flow is correlated with the baroclinic conversion terms.

In Fig. 6 we display the volume-averaged balanced (ϵ) and unbalanced (μ) divergent potentials during the simulation. The initial inertial oscillation caused by the

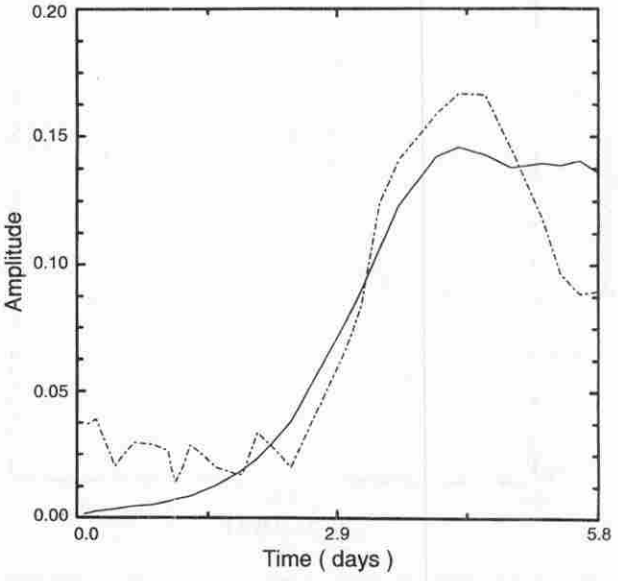

FIG. 6. Amplitude estimates of the balanced divergent (solid) and the unbalanced divergent (dashed) components of the flow, calculated by Eqs. (4.3) and (4.4).

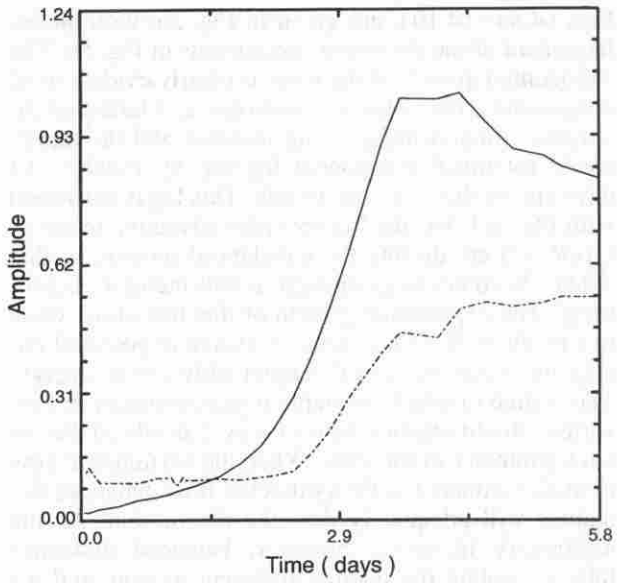

FIG. 7. Same as Fig. 6 but for $\nabla^2\chi$ and $\nabla^2\chi'$ in Eqs. (4.3) and (4.4) instead of χ and χ'. The disappearance of the inertial oscillation suggests that it is occuring on a length scale much smaller than that of the rotational motion.

imbalance in the normal-mode perturbation is clearly evident, although it appears to have little influence on the following evolution. The early exponential growth of the baroclinic wave is described by the rise of the amplitude of the balanced components. This is more clearly seen in the volume-averaged $\overline{\rho}_a w$ fields, shown

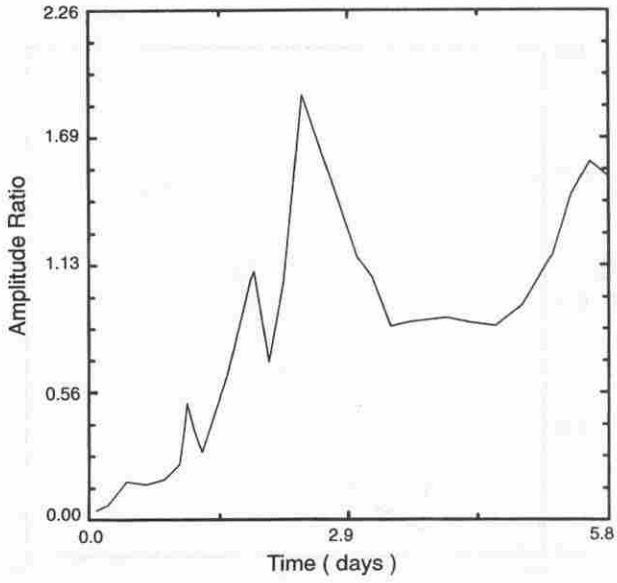

FIG. 8. The ratio $\epsilon:\mu$, where ϵ and μ are calculated according to (4.3) and (4.4). The ratio increases as the balanced divergent component rises exponentially, then saturates, and decreases to roughly 1.

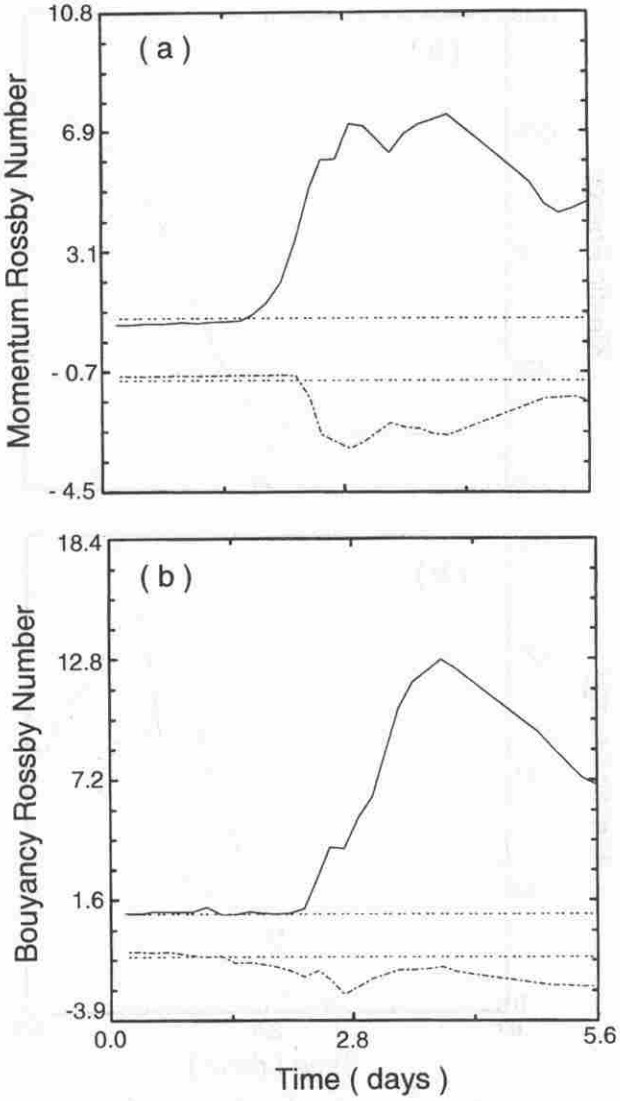

FIG. 9. (a) Time series of the maximum (solid) and minimum (dashed) momentum Rossby number defined by (5.1). The dashed horizontal lines denote 1 and -1. (b) As in (a), only for the buoyancy Rossby number defined by (5.2). Note the O(1) values of these Rossby numbers in the initial flow, and their subsequent increase as the frontal regions form. The threshold for inertial instability in (a) is -1, and values of F_θ in the surface front drop below this threshold at roughly day 3.

in Fig. 7, since the length scale separation for $\overline{\rho}_a w$ and $\overline{\rho}_a w'$ is more pronounced than for χ and χ'. The saturation of both the balanced and the unbalanced fields at approximately day 4 coincides with the saturation of the energy conversion terms of Fig. 3 and the equilibration of the baroclinic wave.

The ratio ϵ/μ is plotted in Fig. 8. Initially, there are only extremely small values of the balanced divergent potential (i.e., the structure of the normal mode is mainly unbalanced), but ϵ rapidly increases to a value greater than μ, after which time approximate equality

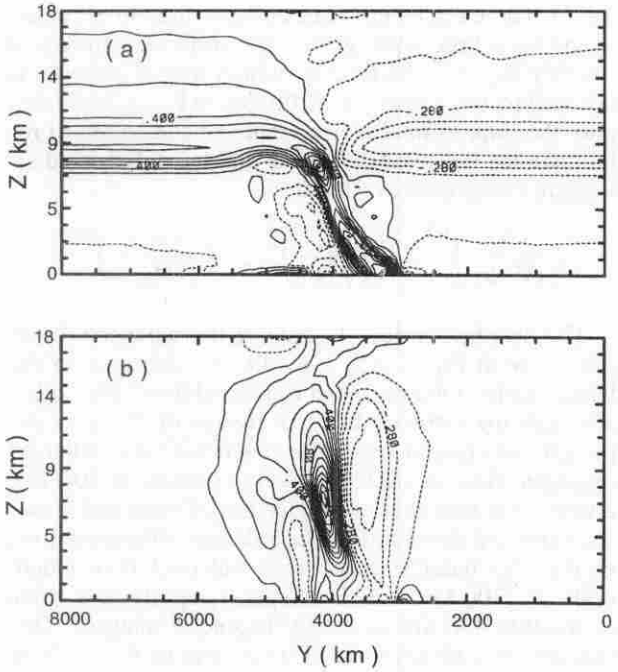

FIG. 10. The meridional structure of (a) R_θ and (b) R_ζ at day 3.4, the cross section being taken at $x = 1285$ kilometers. The contour interval is 0.17 for both plots. Note in (a) the large values of R_θ in the surface front and in (b) the maximum and minimum values of R_ζ in both the upper and lower level fronts.

between the two obtains. This result is at first glance contradictory to the predictions of the scaling analysis, which say that μ should be at the most of $O(\epsilon R, \epsilon^2)$ (although $\mu \geq \epsilon$ does not imply a loss of evolutionary control by the balanced dynamics unless there is a significant feedback from the unbalanced motion onto the rotational and balanced divergence components). To explain this discrepancy we plot, respectively, in Figs. 9a and 9b calculations of the maximum and the minimum momentum Rossby number R_ζ and buoyancy Rossby number R_θ, defined by

$$R_\zeta = \frac{\zeta}{f} \tag{5.1}$$

$$R_\theta = \frac{\partial_z \theta - \langle \partial_z \theta \rangle_{xy}}{\langle \partial_z \theta \rangle_{xy}}, \tag{5.2}$$

in which $\langle A \rangle_{xy}$ denotes a horizontal average of a field A. As can be seen in both figures, the initial Rossby number of the flow is not small. It is, in fact, of $O(1)$. As the baroclinic wave develops, the mesoscale frontal zones produce large values of vertical vorticity and large values of buoyancy.

The maximum momentum Rossby number saturates at a value around 7 at roughly day 4, when the baroclinic conversions have ceased. This is consistent with the fact that the frontal zones relax and their cross-front scale increases after the occlusion of the low (cf. Figs. 2d–f), decreasing the local frontal vertical vorticity. The minimum momentum Rossby number has a constant value of about -0.7 until the time of intense frontal formation, at which time it decreases below the threshold value for inertial instability of -1 (e.g., Holton 1979). Examination of the spatial distribution of the buoyancy Rossby number (Fig. 10a) reveals that the maximum values occur in the surface frontal zone, although large values are also found in the upper-level front. Figure 10b reveals that the momentum Rossby number has its largest and smallest values in both the upper- and the lower-level fronts as the wave is growing. The smallest values of the momentum Rossby number are found in the surface front at the tip of the cusp (cf. Fig. 2). The contribution to frontolysis by vortex stretching has been noted by Reeder and Keyser (1988) in their study of anelastic equation and semigeostrophic upper-level frontogenesis, although they found no evidence of such a mechanism at work in the surface front and some evidence that this occurred in conjunction with diffusion in upper levels. At this point in our simulation, however, diffusion is setting the frontal scale and we see no definite evidence of a dynamical inhibitor to frontal collapse (such as, for example, strong internal wave emission), given our relatively moderate resolution of the mesoscale.

Since the maximum Rossby number of the flow during the evolution of the baroclinic wave is so large, we cannot expect the previous scaling analysis to hold rigorously in the frontal regions wherein one finds these large values. However, the multiscale amplitude esti-

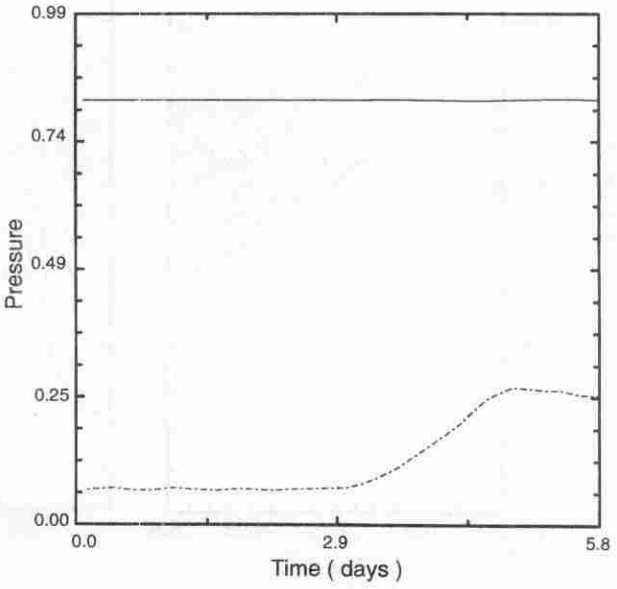

FIG. 11. Time series of the nondimensional volume-averaged balanced (solid) and unbalanced (dashed) pressure. Note the rise of the unbalanced pressure at day 3 resulting from an underestimation by the balance equations of the depth of the surface low (see Fig. 12).

mate for the unbalanced motion (2.15) implies that for Rossby and Burger numbers of O(1) the unbalanced amplitude μ will be of O(ϵ), and this is what we find in Fig. 8.

The volume-averaged balanced and unbalanced pressure (again normalized by Ψ) are shown in Fig. 11, and the constancy of the balanced pressure is, through the balanced divergence equation, a simple reflection of the fact that the volume-averaged rotational motion is relatively constant. The cyclostrophic acceleration term $2RJ(\Psi_y, \Psi_x)/\bar{\rho}_a$ in the balanced divergence equation is seen to have a magnitude of roughly 17% of the Coriolis term $\nabla^2\Psi$. A marked increase in the unbalanced pressure starting at day 3 is evident, this unbalanced pressure being associated with an underestimate by the balance equations of the rapidly deepening low pressure center (Fig. 12). Since the model employs the nonhydrostatic equations of motion and the balance analysis assumes hydrostatic balance, the unbalanced fields will include implicitly all nonhydrostatic motion. As demonstrated by Polavarapu and Peltier (1990), the surface low and the cold front are regions in which nonhydrostatic accelerations are the greatest and the errors in assuming hydrostatic balance are the largest.

In Fig. 13a, we show the length scale ratio estimates (i.e., δ^{-1} and ν^{-1}) as calculated from Eqs. (4.6)–(4.7). There is a length scale collapse in both estimates as the baroclinic wave grows; the collapse saturates at roughly day 3.5, the time by which frontal collapse is arrested by the numerical diffusion. In Fig. 13b we display the dimensional length scales of the balanced rotational, the balanced divergent, and the unbalanced divergent components of the flow:

$$\left(\frac{\langle\Psi\rangle}{\langle\nabla^2\Psi\rangle}\right)^{1/2}, \quad \left(\frac{\langle\chi\rangle}{\langle\nabla^2\chi\rangle}\right)^{1/2}, \quad \left(\frac{\langle\chi'\rangle}{\langle\nabla^2\chi'\rangle}\right)^{1/2}.$$

The decrease in length scale of the balanced divergent flow in Fig. 13a is a result of a decrease in the length scale of the balanced rotational flow (Fig. 13b), although the estimate for δ is always of O(1) as expected. The length scale associated with the balanced divergent flow alone is relatively constant at 500 km, a value that would be characteristic of balanced mesoscale up- and downdrafts. The balanced rotational flow, on the other hand, starts off with a characteristic length scale of 2300 km, which decreases rapidly to a value of roughly 870 km as the cyclogenesis matures. This decrease is characteristic of a cascade of the scale of the dominant motions from the synoptic scale to the mesoscale, that is, from the scale of the growing baroclinic wave to that of frontal flow. We note that the

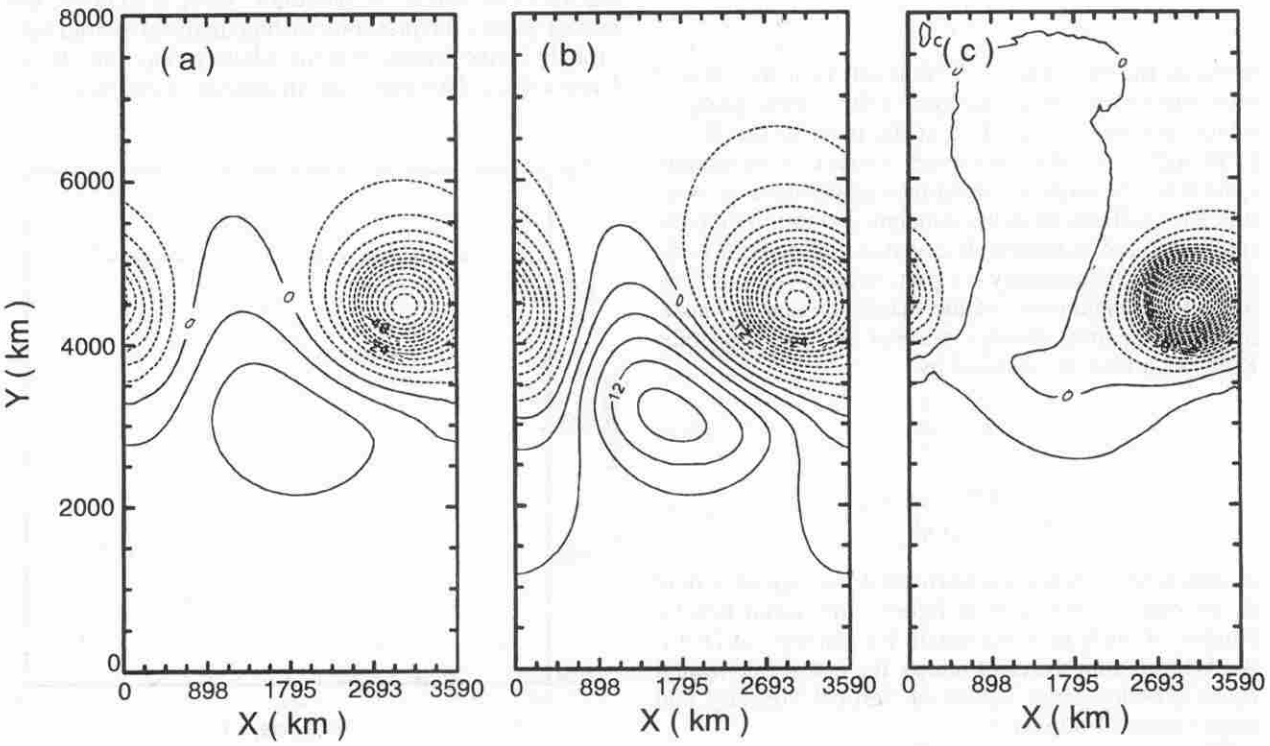

FIG. 12. Horizontal cross sections taken at $z = 1$ km of (a) total pressure, (b) balanced pressure, and (c) unbalanced pressure at day 4 in units of 10^{-1} mb and plotted as a deviation from the initial adiabatic profile (which is 2000 mb at the surface). The magnitude of the deviations in (a) and (b) are 8 mb and 4 mb, respectively. The unbalanced pressure has a maximum amplitude of 4 mb. Contour intervals are 0.6 mb, 0.3 mb, and 0.2 mb, respectively.

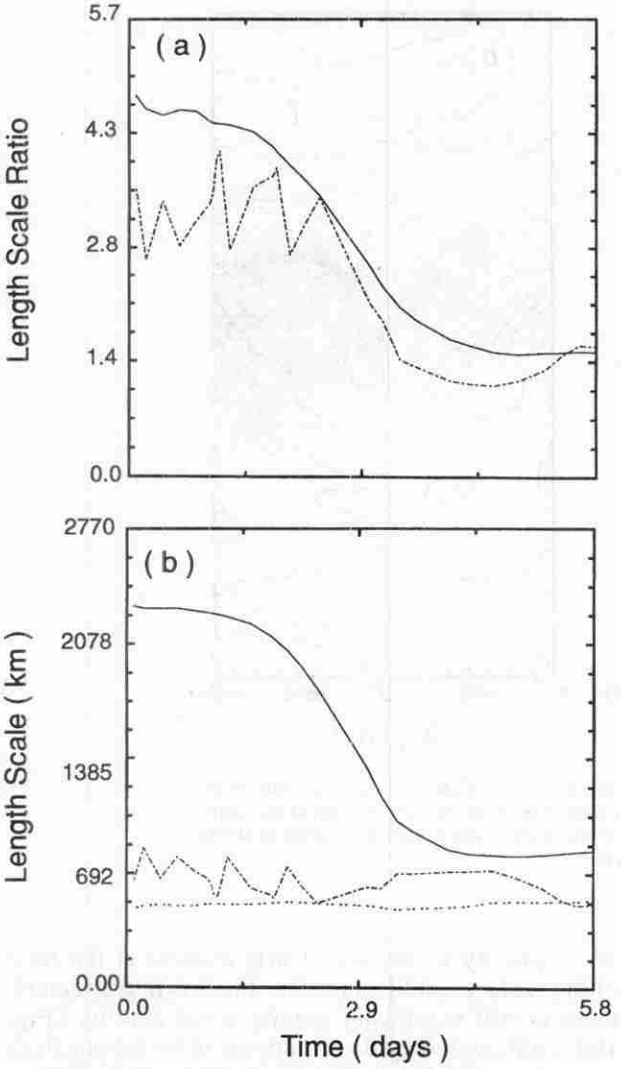

FIG. 13. (a) Time series of estimates for the length scale ratios δ^{-1} (solid) and ν^{-1} (dashed), calculated according to (4.6) and (4.7), respectively. (b) Time series of length scale estimates for the balanced rotational component (solid), the balanced divergent component (dash-dot), and the unbalanced divergent component (dashed). Note that the decrease in the length scale ratio noted in (a) is caused by a collapse in the length scale of the balanced rotational motion. Units in (b) are kilometers.

occlusion of the surface low (cf. Fig. 2) occurs at roughly day 4, shortly after the collapse in length scale of the balanced rotational flow has ceased. The length scale of the unbalanced χ' field is fairly constant at a value of 700 km, indicating predominantly mesoscale frontal imbalance. We may explore further the length scale separation between the balanced and the unbalanced motions by plotting the length scale ratio estimates for the vertical mass flux (Fig. 14a). Here we see a relatively constant value for the balanced divergent length scale ratio δ, which since we know that the estimate for Ψ decreases between days 1 and 3, must imply a decreasing scale for the balanced divergent potential during this same time. Examination of Fig. 14b reveals a length scale estimate of roughly 1000 km for the first 2.7 days, after which there is a rapid scale decrease to the mesoscale (roughly 300 km). The unbalanced length scale, on the other hand, decreases continuously from initialization to occlusion, finally saturating at a value of 210 km after day 3. This collapse of the divergent length scales is consistent with the fact that the majority of the balanced and unbalanced divergent motion occurs in the collapsing frontal zones.

A comparison of the unbalanced vertical mass flux at $z = 1$ km and at $z = 8$ km is shown in Fig. 15. The magnitude of $\overline{\rho}_a w'$ in the upper levels is a factor of 3 smaller than that found in the lower levels and, if back-

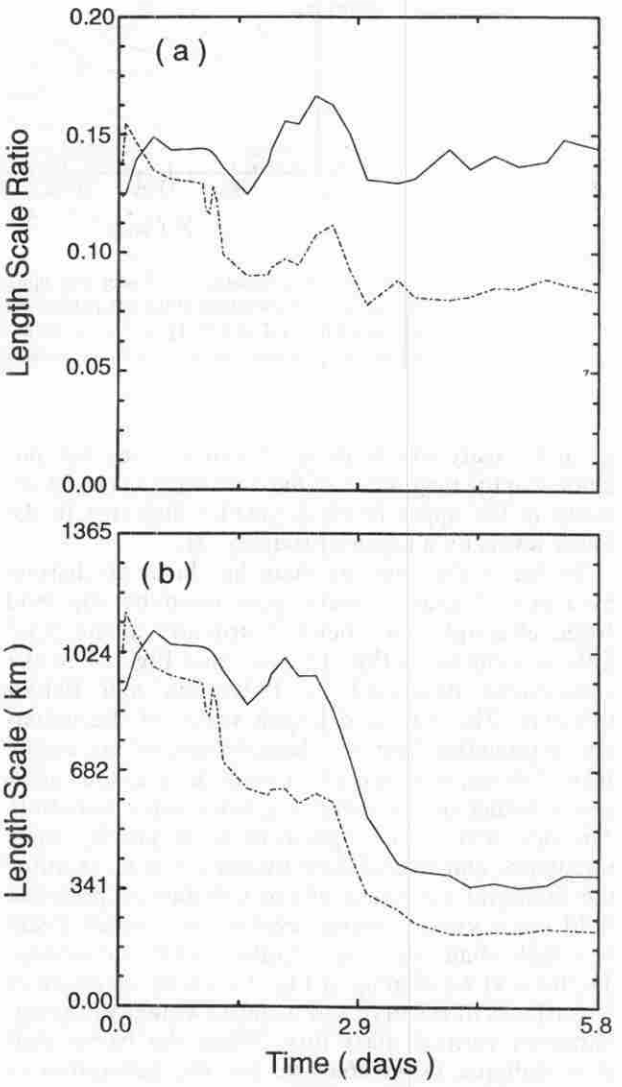

FIG. 14. Same as Fig. 13 except the calculations are performed using the vertical mass flux ($\nabla^2\chi$) rather than the divergent potential (χ). Estimates for the balanced divergent flow are the solid lines. Estimates for the unbalanced divergent flow are the dashed lines.

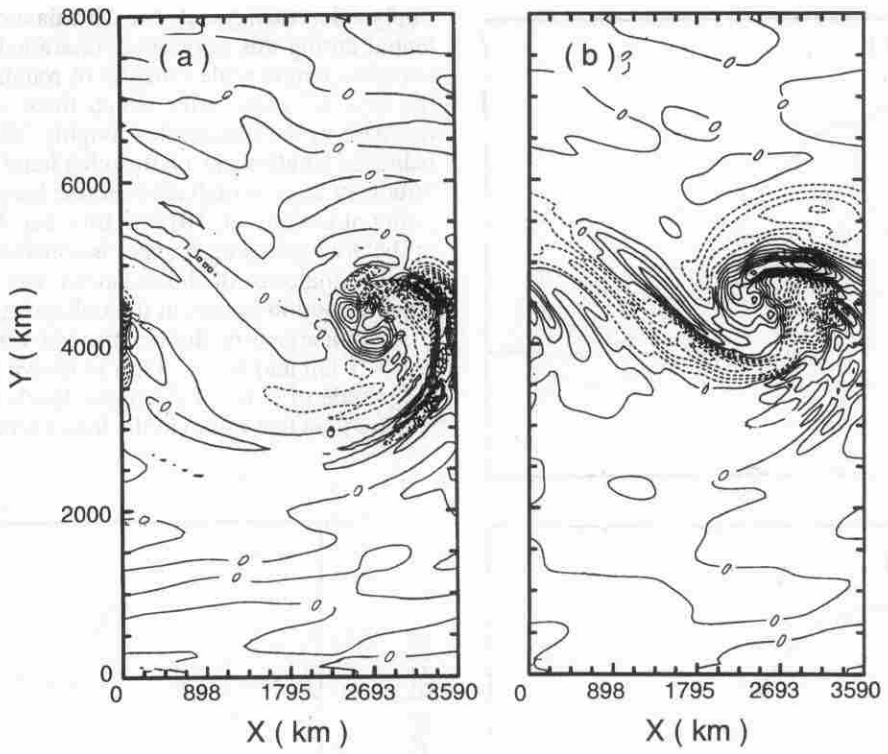

FIG. 15. A comparison of lower and upper level imbalance in the vertical mass flux field. Horizontal cross sections of the unbalanced vertical mass flux are plotted at $z = 1$ km in (a) (with a contour interval of 0.01 kg m^{-2} s^{-1}) and at $z = 8$ km in (b) (with a contour interval of 0.004 kg m^{-2} s^{-1} the cross-sections being taken at day 4).

ground density effects are taken into account, we conclude that the magnitude of the unbalanced vertical velocity in the upper levels is smaller than that in the lower levels by a factor of roughly 2/3.

So far in the analysis there has been no distinct evidence of gravity wave generation by the cold front, although some banded structure in the $\overline{\rho}_a w'$ field is evident in Fig. 15 [see also Fig. 14 in the simulations described by Polvarapu and Peltier (1990)]. The time and length scales of the unbalanced potential flow are characteristic of mesoscale frontal flows, so it is unclear from the previous analysis whether or not there is gravity wave radiation. The best way to distinguish between gravity wave imbalance and frontal flow imbalance is to examine the temporal evolution of the unbalanced potential field since gravity waves evolve on a much faster timescale than the slower higher-order corrections. To this end we display in Fig. 16 a time sequence of isosurfaces of positive and negative values of the unbalanced vertical mass flux. When the fronts start their collapse (Fig. 16a) we see the generation of unbalanced velocities along the entire frontal zone. As the low deepens and the fronts tighten further, we see the initial generation of gravity waves at the cold front (Fig. 16b). At the point of occlusion (Fig. 16c), gravity waves are clearly evident at the base of the surface cold front. After the low has detached, there is still significant gravity wave activity (Fig. 16d), although they do not appear to be propagating away from the front. The horizontal wavelength of these waves is approximately 250 km (roughly $4\Delta x$) and the vertical wavelength is slightly greater than 1 km.

The origin of these gravity waves is most probably numerical, although higher-resolution solutions will be required to establish this definitively. In Figs. 17a,b we show two cross-sectional plots of potential temperature and vertical mass flux through the front at days 4 and 6, respectively. The frontal slope in the lower 2 km is markedly more shallow, especially at the later time, than its slope at upper levels; we roughly estimate the aspect ratio in the lower 2 km to be 0.002, whereas above 2 km the ratio is slightly greater than 0.01. We are therefore, according to (3.1), underresolved in the vertical in the lower 2 km by a factor of roughly 5, and this underresolution may be producing spurious gravity waves near the surface that do not propagate away from their source. The magnitude of these gravity waves approaches 50% of that of the slower advective flow, which itself has a maximum magnitude of 0.15

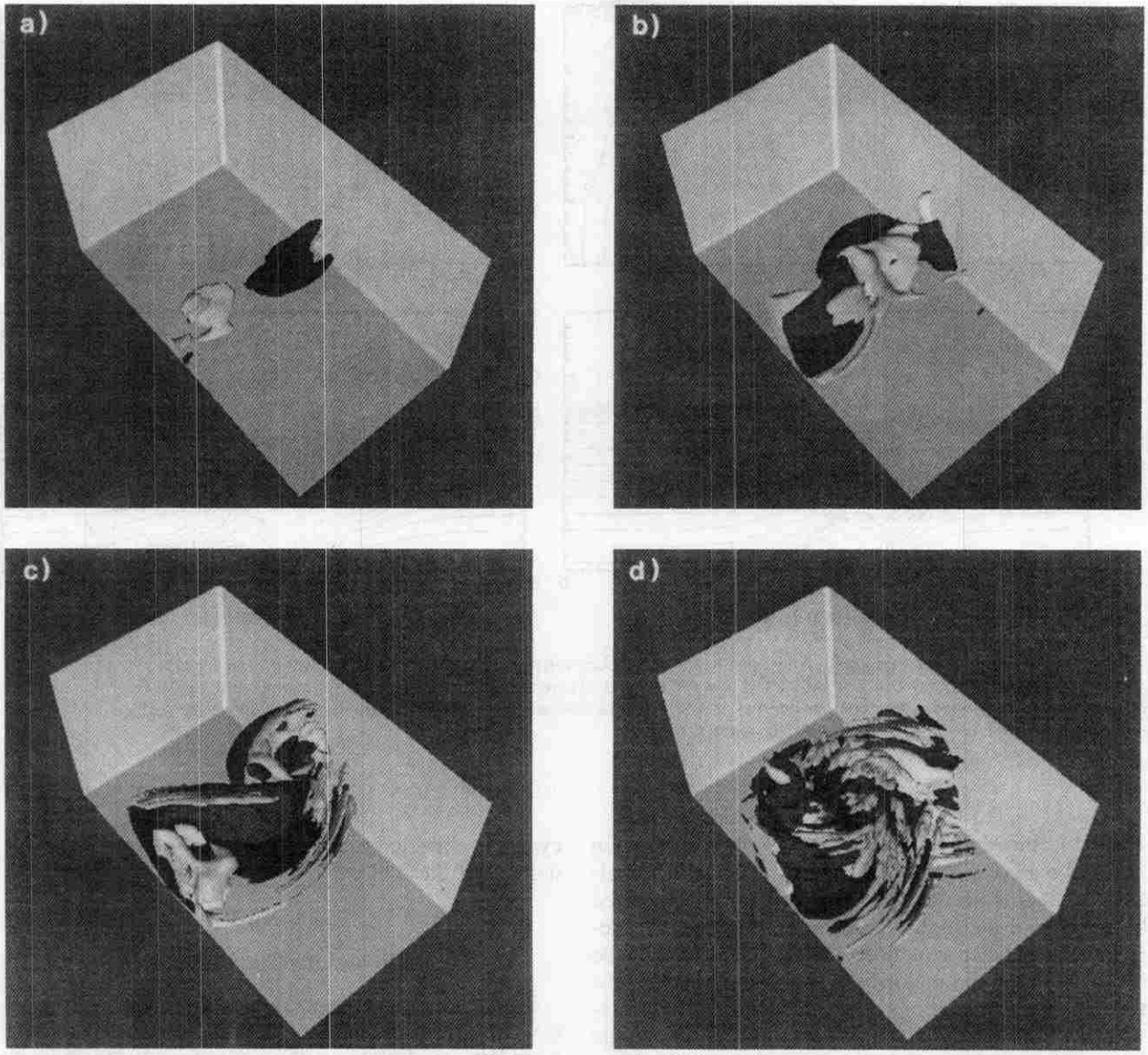

FIG. 16. Isosurface plots of positive (light) and negative (dark) unbalanced vertical mass flux at (a) day 2.3, (b) day 3.2, (c) day 4, and (d) day 6. Light (dark) surfaces enclose values above (below) 0.06 (-0.06) kg m^{-2} s^{-1}. The perspective is such that we are looking downward and northeastward. The solid light planes denote the model domain. Note in (c) and (d) the appearance of gravity waves (most likely of numerical origin) of horizontal wavelength $\sim$ 250 kilometers in the vicinity of the surface cold front.

kg m^{-2} s^{-1} in the center of the low. Typical frontal values of the unbalanced vertical flux field are 0.07 kg m^{-2} s^{-1}.

6. Conclusions

We have developed a set of balance equations for the three-dimensional anelastic primitive equations based on an expansion in Rossby and Froude numbers. The accuracy of the balance approximation for a strong midlatitude baroclinic wave life cycle has been tested by the application of these equations diagnostically. We conclude that, although the scaling assumptions formally break down during the life cycle of the baroclinic wave, the implications of this analysis are robust in the sense that they deliver what would be expected according to other analyses (Spall and McWilliams 1992; Snyder et al. 1993; Gent et al. 1993). The broad (R, F) parameter regime in which this study is performed spans both the slow synoptic-scale development of the baroclinic wave and the mesoscale frontal regime to which the flow eventually collapses. According to the criteria proposed by Spall and McWilliams (1992) to determine through scaling arguments whether the unbalanced motion is predominantly slow or fast (i.e., respectively, $R \leqslant 1$, $F^2 < R$ or $R > 1$ and

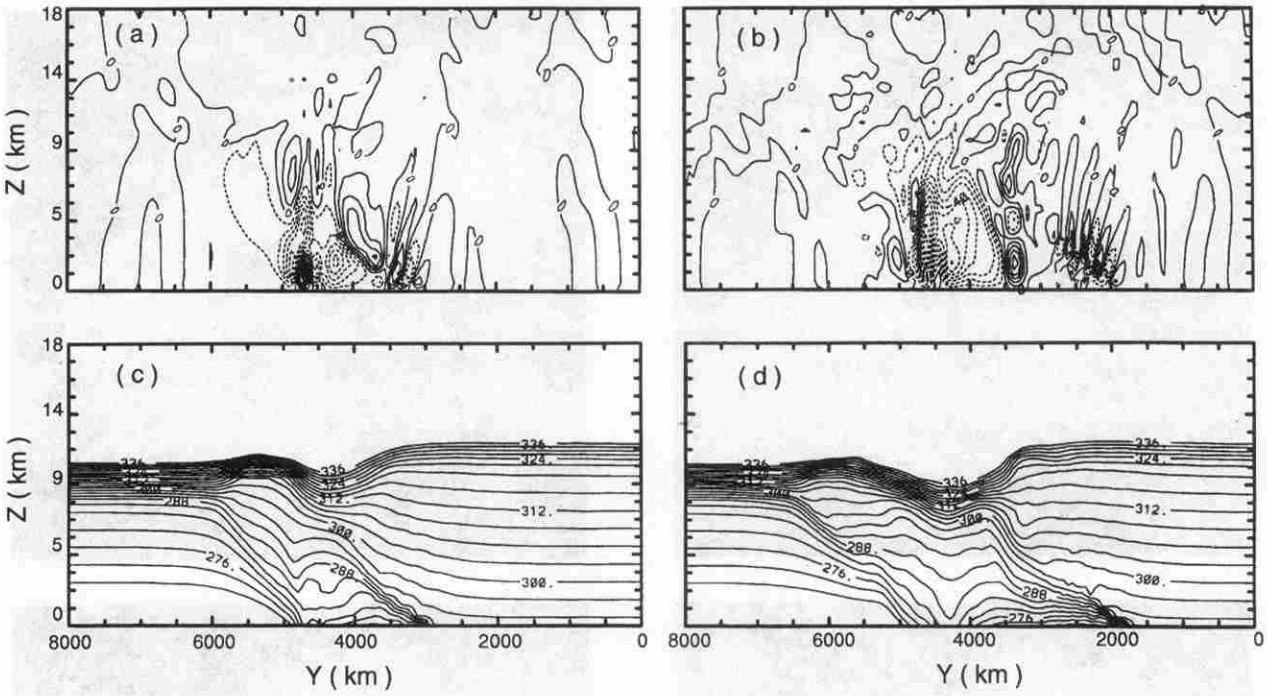

FIG. 17. Meridional cross sections (looking zonally) of unbalanced vertical mass flux (contour interval of 0.01 kg m^{-2} s^{-1} at (a) day 4 at $x = 3161$ km, and at (b) day 6 at $x = 1661$ km. Shown in (c) and (d) are the corresponding potential temperature fields. Only those values of $\Theta \leq 336$ K are plotted with a contour interval of 3 K. Note the appearance of numerical gravity waves above the surface cold front in those regions where the slope of the front falls below f/N.

$F \leq 1$), the unbalanced motion in this simulation should be a combination of both slow and fast imbalance. The slow timescale unbalanced motion is caused by neglect of the higher-order advective terms in deriving the balance equations and is found here to be responsible for the majority of the unbalanced amplitude. Spatial distributions of the unbalanced motion reveal the largest imbalance in the frontal zones, with the largest magnitude being in the core of the surface low. Evidence of fast timescale gravity waves is clear, but these are, most probably, numerically spurious, being generated when the slope of the front becomes shallow enough to violate the consistency criterion between the horizontal and the vertical grid spacing used. These gravity waves have an amplitude that reaches a maximum of roughly 50% of that of the slower unbalanced motion, so they may not be considered insignificant. Further analyses are in the planning stage that will attempt to track the nondiffusive collapse of the frontal-scale flow to smaller scales than was possible in this preliminary analysis. The issue remains as to whether we will then begin to see the spontaneous emission of internal waves from the frontal zone that is expected on both observational and theoretical grounds (e.g., Ley and Peltier 1978). The technical apparatus that we have developed here for the analysis of the origins and nature of imbalance in anelastic baroclinic wave life cycles is expected to prove useful in the further development of these ideas.

APPENDIX

Boundary Conditions

The domain is periodic in the x direction, with walls at $y = y_1, y_2$ km and at $z = z_1, z_2$ km, so we need explicit boundary conditions on the walls for all the Poisson equations in section 2.

a. Toroidal scalar

The inversion of Eq. (4.1) requires knowledge of $\partial_y \Psi$ on the meridional walls at $y = y_1, y_2$. We assume that the poloidal component of the zonal mass flux is negligible at the meridional walls, so that

$$\partial_y \Psi = \bar{\rho}_a u \quad \text{on} \quad y = y_1, y_2. \tag{A1}$$

b. Poloidal scalar

To invert Eq. (4.2) for $\epsilon\chi$, we require boundary conditions for $\epsilon\partial_y\chi$ on the meridional walls at $y = y_1, y_2$. We impose no normal flow at the meridional walls, so from Eq. (2.4) we see that

$$\epsilon\chi_y = \int_0^z \Psi_x dz' \quad \text{on} \quad y = y_1, y_2. \tag{A2}$$

c. Vorticity equation

The vorticity equation (2.8) involves a two-dimensional Laplacian operating on $\partial_t\Psi$, so explicit boundary conditions are required for $y = y_1, y_2$. We make the approximation that the time variation of the toroidal component of the zonal mass flux in the domain is small; namely,

$$\Psi_{yt} = 0 \quad \text{on} \quad y = y_1, y_2. \tag{A3}$$

This approximation has been validated for an instantaneous diagnostic analysis [Ψ_{yt} on the meridional boundaries is of nondimensional $O(10^{-2})$]. This approximation may not be desirable, however, in a time integration of these balance equations.

d. Divergence equation

For the divergence equation (2.9), we again have a two-dimensional Laplacian operating on the pressure, so $\partial_y p$ is required on $y = y_1, y_2$. We make the approximation of gradient wind balance at the walls, and the zonal velocity field is completely determined by its toroidal component, which delivers in nondimensional form

$$p_y = -\Psi_y \quad \text{on} \quad y = y_1, y_2. \tag{A4}$$

e. The omega equation

The omega equation (2.12) is a three-dimensional elliptic equation, requiring boundary conditions on $\epsilon\nabla^2\chi$ on $y = y_1, y_2$ and on $z = z_1, z_2$. Since $\epsilon\nabla^2\chi = -\overline{\rho}_a w$, then trivially,

$$\epsilon\nabla^2\chi = 0, \quad \text{on} \quad z = z_1, z_2. \tag{A5}$$

Further, we impose no meridional flow at $y = y_1, y_2$ so that from Eq. (2.4) we deduce that

$$\epsilon\nabla^2\chi_y = \int_0^z \nabla^2\Psi_x dz' \quad \text{on} \quad y = y_1, y_2. \tag{A6}$$

REFERENCES

Allen, J. S., J. A. Barth, and P. A. Newberger, 1990: 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.

Baer, F., and J. J. Tribbia, 1977: On complete filtering of gravity modes through nonlinear initialization. *Mon. Wea. Rev.*, **105,** 1536–1539.

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.

Bush, A. B. G., and W. R. Peltier, 1994: Tropopause folds and synoptic-scale baroclinic wave life cycles. *J. Atmos. Sci.*, **51,** 1581–1604.

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

Clark, T., 1977: A small-scale dynamic model using a terrain-following coordinate transformation. *J. Comput. Phys.*, **24,** 186–215.

Gall, R. L., R. T. Williams, and T. L. Clark, 1987: On the minimum scale of surface fronts. *J. Atmos. Sci.*, **44,** 2562–2574.

Garner, S. T., 1989: Fully Lagrangian numerical solutions of unbalanced frontogenesis and frontal collapse. *J. Atmos. Sci.*, **46,** 717–739.

Gent, P. R., and J. C. McWilliams, 1982: Intermediate model solutions to the Lorenz equations: Strange attractors and other phenomena. *J. Atmos. Sci.*, **39,** 3–13.

——, and ——, 1983: Regimes of validity for balanced models. *Dyn. Atmos. Oceans,* **7,** 167–183.

——, ——, and C. Snyder, 1993: Scaling analysis of curved fronts: Validity of the balance equations and semigeostrophy. *J. Atmos. Sci.*, **51,** 160–163.

Herron, T. J., I. Tolstoy, and D. W. Kraft, 1969: Atmospheric pressure background fluctuations in the mesoscale range. *J. Geophys. Res.*, **74,** 1321–1329.

Holton, J. R., 1979: *An Introduction to Dynamic Meteorology.* 2d ed. International Geophysical Series, Vol. 23, Academic Press.

Hoskins, B. J., and F. P. Bretherton, 1972: Atmospheric frontogenesis models: Mathematical formulation and solution. *J. Atmos. Sci.*, **29,** 11–37.

Keyser, D., and M. J. Pecnick, 1985: Diagnosis of ageostrophic circulations in a two-dimensional primitive equation model of frontogenesis. *J. Atmos. Sci.*, **42,** 1283–1305.

Koshyk, J. N., and H.-R. Cho, 1992: Dynamics of a mature front in a uniform potential vorticity semigeostrophic model. *J. Atmos. Sci.*, **49,** 497–510.

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

Ley, B. E., and W. R. Peltier, 1978: Wave generation and frontal collapse. *J. Atmos. Sci.*, **35,** 3–17.

Lindzen, R. S., and M. Fox-Rabinovitz, 1989: Consistent vertical and horizontal resolution. *Mon. Wea. Rev.*, **117,** 2575–2583.

Lorenz, E. N., 1960: Energy and numerical weather prediction. *Tellus,* **12,** 364–373.

——, 1992: The slow manifold—What is it? *J. Atmos. Sci.*, **49,** 2449–2451.

——, and V. Krishnamurthy, 1987: On the nonexistence of a slow manifold. *J. Atmos. Sci.*, **44,** 2940–2950.

McWilliams, J. C., 1985: On a uniformly valid model spanning the regimes of geostrophic and isotropic, stratified turbulence: Balanced turbulence. *J. Atmos. Sci.*, **42,** 1773–1774.

——, and P. R. Gent, 1980: Intermediate models of planetary circulations in the atmosphere and ocean. *J. Atmos. Sci.*, **37,** 1657–1678.

Norton, N. J., J. C. McWilliams, and P. R. Gent, 1986: A numerical model of the balance equations in a periodic domain and an example of balanced turbulence. *J. Comput. Phys.*, **67,** 439–471.

Peltier, W. R., and T. L. Clark, 1979: The evolution and stability of finite-amplitude mountain waves. Part II: Surface drag and severe downslope windstorms. *J. Atmos. Sci.*, **36,** 1498–1529.

Persson, P. O. G., and T. T. Warner, 1991: Model generation of spurious gravity waves due to inconsistency of the vertical and horizontal resolution. *Mon. Wea. Rev.*, **119,** 917–935.

Polavarapu, S. M., and W. R. Peltier, 1990: The structure and nonlinear evolution of synoptic scale cyclones: Life cycle simulations with a cloud-scale model. *J. Atmos. Sci.*, **47,** 2645–2672.

Reeder, M. J., and D. Keyser, 1988: Balanced and unbalanced upper-level frontogenesis. *J. Atmos. Sci.*, **45,** 3366–3386.

Snyder, C., W. C. Skamarock, and R. Rotunno, 1991: A comparison of primitive equation and semigeostrophic simulations of baroclinic waves. *J. Atmos. Sci.*, **48,** 2179–2194.

——, ——, and ——, 1993: Frontal dynamics near and following frontal collapse. *J. Atmos. Sci.*, **50,** 3194–3212.

Spall, M. A., and J. C. McWilliams, 1992: Rotational and gravitational influences on the degree of balance in the shallow-water equations. *Geophys. Astrophys. Fluid Dyn.*, **64,** 1–29.

Warn, T., and R. Menard, 1986: Nonlinear balance and gravity–inertial wave saturation in a simple atmospheric model. *Tellus,* **38A,** 285–294.

Whitaker, J. S., 1993: A comparison of primitive and balance equation simulations of baroclinic waves. *J. Atmos. Sci.*, **50,** 1519–1530.