A Comparison of Primitive-Equation and Semigeostrophic Simulations of Baroclinic Waves

CHRIS SNYDER, WILLIAM C. SKAMAROCK AND RICHARD ROTUNNO

*National Center for Atmospheric Research, Boulder, Colorado**

(Manuscript received 25 January 1991, in final form 18 April 1991)

ABSTRACT

In the course of adapting a nonhydrostatic cloud model [or primitive-equation model (PE)] for simulations of large-scale baroclinic waves, we have encountered systematic discrepancies between the PE solutions and those of the semigeostrophic (SG) equations. Direct comparisons using identical, uniform potential vorticity jets show that 1) the linear modes of the PE have distinctively different structure than the SG modes; 2) at finite amplitude, the PE pressure field develops lows that are deeper, and highs that are weaker, than in the SG solution; and 3) the nonlinear PE wave produces a characteristic "cyclonic wrapping" of the temperature contours on both horizontal boundaries and has an associated "bent-back" frontal structure at the surface, while in the SG solutions (for this particular basic state jet) there is an equal tendency to pull temperature contours anticyclonically around highs and cyclonically around lows. An analysis of the vorticity and potential vorticity equations for small Rossby number reveals that the SG model errs in its treatment of terms involving the ageostrophic vorticity. Simulations based on an equation set that includes the leading-order dynamical contributions of the ageostrophic vorticity agree more closely with the PE simulations.

1. Introduction

Because of the central role of baroclinic waves in producing everyday weather in midlatitudes, meteorologists have long sought to model the development of such disturbances. Like other slow motions of the atmosphere or ocean, baroclinic disturbances may be simulated with approximate, balanced models, which filter high-frequency gravity and inertial oscillations, or by using some form of the primitive equations (PE). To date, the most sophisticated approximate-model simulations of baroclinic waves have been those using the semigeostrophic (SG) equations (Hoskins 1976; Hoskins and West 1979; Heckley and Hoskins 1980; Davies et al. 1991). At the same time, PE models (Mudrick 1974; Takayabu 1986; Polavarapu and Peltier 1991) have provided increasingly detailed, if conceptually more complicated, simulations of baroclinic disturbances. While developing our own elastic, nonhydrostatic PE model, we have had occasion to compare our PE solutions with previous SG results for baroclinic development; these comparisons reveal significant errors in the SG solutions. This paper presents a comparison of PE and SG simulations in a specific case, along with a small–Rossby-number analysis of the errors in the SG equations. Although the comparison focuses on baroclinic waves, the analysis is general enough to apply to a variety of geophysical flows.

Semigeostrophy was first proposed by Hoskins (1975) and depends, like quasi-geostrophy (QG), on approximate geostrophic balance. However, SG includes advection by the ageostrophic wind and tilting and stretching of relative vorticity and can produce a frontal singularity in a finite time. Indeed, SG now underlies much of the current understanding of frontogenesis (e.g., Hoskins and Bretherton 1972) due to its notable success in the description of two-dimensional flows. For example, PE and SG simulations of two-dimensional frontogenesis agree closely in both the deformation-induced (Koclas et al. 1986) and shear-induced (Hoskins and Bretherton 1972; Volkert and Bishop 1990) cases. Most analyses of the errors in the SG approximation have consequently focused on the local breakdown of the approximation under extreme conditions, such as near frontal collapse (Davies and Müller 1988; Holt and Shutts 1990) or in regions of small stability to moist stability to moist slantwise ascent (Fantini 1990). Allen et al. (1990) have, however, presented solutions for SG, and a variety of other approximate models, in the case of shallow-water flow over O (1) topography. They find that SG is more accurate than QG in such cases, but still makes significant quantitative errors when compared with the full shallow-water equations. In fully three-dimensional flows, comparisons against PE solutions have been limited to

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

Corresponding author address: Dr. Chris Snyder, NCAR/MMM, P.O. Box 3000, Boulder, CO 80307-3000.

the frequencies and structures of linear modes for the frontal stability problems (Duffy 1976; Moore and Peltier 1988) and for planetary Rossby waves on a sphere (Shutts 1989; Magnusdottir and Schubert 1991). Although SG has been used to simulate three-dimensional flows (in particular, the nonlinear development of baroclinic waves), it has not been directly compared with the PE in such cases.

Primitive-equation simulations lack the conceptual simplicity of balanced models but, of course, retain a more accurate version of atmospheric dynamics. We are currently modifying the nonhydrostatic Klemp–Wilhelmson (1978) cloud model for high-resolution, adaptive-grid simulations of large-scale baroclinic waves. The numerical aspects of this work are due to Skamarock and J. B. Klemp (personal communication, 1991). Our long-range goal is to resolve explicitly certain processes that are normally parameterized in numerical simulations of large-scale flows. In particular, we hope to use the nonhydrostatic nature and adaptive-gridding capability of the model to simulate explicitly regions of moist convection embedded within an idealized baroclinic disturbance. In this paper, we present some preliminary results from the model that do not require the adaptive-gridding capability.

During the course of developing this model, we compared PE and SG simulations of baroclinic waves, using the $\mu = 1$ uniform potential vorticity jet of Hoskins and West (1979, HW). We find systematic differences in the baroclinic-wave development between the PE and SG models. Relative to those in the SG simulations, the most unstable PE mode has a NW–SE phase tilt and the finite amplitude PE solutions exhibit lower minimum pressures and weaker maximum pressures. More importantly, the large-amplitude wave in the PE simulations produces a distinctive tongue of warm air at the surface that penetrates northward and westward into the cold air and a similar cold tongue moving southward and eastward at the lid.

The errors produced by SG in these three-dimensional flows were anticipated in the scaling analysis of McWilliams and Gent (1980). They show that, in three-dimensional flows, SG has the same asymptotic accuracy as QG and correctly includes only the leading corrections to the basic geostrophic balance, while in frontal flows with sufficiently small alongfront variability, SG retains an additional order of accuracy in an expansion in powers of the ratio of cross- to alongfront velocity. We complement McWilliams and Gent's analysis by showing that, at leading order in Rossby number, the errors in SG arise from the neglect of the dynamical contributions of the ageostrophic vorticity. Moreover, we confirm this asymptotic result by demonstrating that solutions to an extended model, which correctly includes the leading-order contributions of the ageostrophic vorticity in the potential vorticity equation, accurately portray the early evolution of a three-dimensional baroclinic wave.

2. Model equations

We consider the adiabatic, Boussinesq, f-plane development of perturbations to a zonal jet. The flow will have initially uniform potential vorticity, will satisfy periodic horizontal boundary conditions at $x = 0$, x_L and $y = 0$, y_L, and will be confined in the vertical between rigid surfaces at $z = 0$, H. The Cartesian coordinates (x, y, z) measure eastward and northward distance and physical height, respectively.

For our purposes, the PE are the inviscid, adiabatic, elastic, Boussinesq equations on an f plane. Replacing pressure by $\phi = c_p\theta_0(p/p_0)^{(\gamma-1)/\gamma}$, the PE become

$$\frac{d\mathbf{v}}{dt} + f\mathbf{k}\times\mathbf{v} + \nabla\phi = \frac{g\theta}{\theta_0}\mathbf{k}, \tag{1a}$$

$$\frac{d\phi}{dt} = -c_s^2\nabla\cdot\mathbf{v}, \tag{1b}$$

$$\frac{d\theta}{dt} = 0, \tag{1c}$$

where θ_0 is a reference temperature, p_0 is a reference surface pressure, c_s^2 is the constant speed of sound, and all other notation is conventional. Although vertical accelerations appear in (1a) and the compressible pressure term is retained in (1b) (for numerical reasons), the flow is, in fact, both hydrostatic and anelastic with high accuracy in the simulations presented here.

Semigeostrophy is obtained from (1) by making the anelastic approximation in (1b), the hydrostatic approximation in the vertical component of (1a) and the geostrophic momentum (GM) approximation in the horizontal components of (1a). Under the GM approximation, the momentum is evaluated, as in QG, from the geostrophic velocity $\mathbf{v}_g$, where $\mathbf{v}_g = (u_g, v_g, 0) = f^{-1}(-\phi_y, \phi_x, 0)$, but unlike QG, ageostrophic advections are retained.

As discussed in Hoskins (1975), with the GM approximation the appropriate vorticity equation is

$$\frac{d\zeta_{SG}}{dt} = \zeta_{SG}\cdot\nabla\mathbf{v} - \mathbf{k}\times\frac{g}{\theta_0}\nabla\theta, \tag{2}$$

where

$$\zeta_{SG} = (-v_{gz}, u_{gz}, f + v_{gx} - u_{gy}) + f^{-1}\mathbf{J}_{xyz}(u_g, v_g).$$

We have used the additional notation

$$J_{cd}(a, b) = \frac{\partial a}{\partial c}\frac{\partial b}{\partial d} - \frac{\partial a}{\partial d}\frac{\partial b}{\partial c},$$

$$\mathbf{J}_{cde}(a, b) = [(J_{de}(a, b), J_{ec}(a, b), J_{cd}(a, b)].$$

Similarly, the potential vorticity equation is

$$\frac{dq_{SG}}{dt} = 0, \tag{3}$$

where

$$q_{SG} = \frac{g}{f\theta_0}\zeta_{SG}\cdot\nabla\theta.$$

A transformation to the geostrophic coordinates,

$$X = x + f^{-1}v_g, \quad Y = y - f^{-1}u_g, \quad Z = z, \quad T = t,$$

simplifies the GM equations considerably; the principal advantage being that the horizontal advections become purely geostrophic. The transformed equations are discussed in detail in Hoskins (1975) and McWilliams and Gent (1980). Although the semigeostrophic equations are strictly the geostrophic momentum equations in geostrophic coordinates, we will henceforth follow common practice and use "SG" to refer to both the transformed equations and to the GM equations in physical coordinates.

In geostrophic coordinates, the definition of q_{SG} may be rewritten as

$$\frac{f^2}{q_{SG}}\Phi_{ZZ} + \Phi_{XX} + \Phi_{YY} - \frac{1}{f^2}J_{XY}(\Phi_X, \Phi_Y) = f^2. \quad (4)$$

The quantity $\Phi = \phi + \frac{1}{2}(u_g^2 + v_g^2)$ is the potential function in the new coordinates; i.e., $(\Phi_X, \Phi_Y, \Phi_Z) = (\phi_x, \phi_y, \phi_z)$. For the flows considered here, with initially uniform potential vorticity, q_{SG} is by (3) permanently fixed and uniform. SG then reduces to diagnosing Φ from (4) given $\theta = (\theta_0/g)\Phi_Z$ on the horizontal boundaries. The time evolution of these boundary conditions is determined requiring $w = 0$ at the surface and at the lid, which implies

$$\Phi_{ZT} = \frac{1}{f}\left(\Phi_Y\frac{\partial}{\partial X} - \Phi_X\frac{\partial}{\partial Y}\right)\Phi_Z, \text{ on } Z = 0, H. \quad (5)$$

In order to reproduce the SG solution of Hoskins and West (1979), we follow them and neglect the nonlinear term in (4). This term is generally small (Hoskins 1976), but its neglect does produce noticeable changes in the development, as discussed in appendix A.

3. Numerical methods

We integrate the nonhydrostatic equations (1) using an integration scheme based on that of Klemp and Wilhelmson (1978). Briefly, the equations are differenced on a spatially staggered grid with the velocity components u_i located one-half Δx_i from the (ϕ, θ) points. All spatial derivatives are calculated with second-order centered differencing except for the horizontal advection terms, which employ fourth-order centered differencing. Although the presence of sound waves requires a very small time step, computational efficiency is maintained by the use of a time-splitting procedure. The pressure gradient terms, the divergence term, the buoyancy term in (1b), and the vertical advection of the mean potential temperature $\langle\theta\rangle(z)$ in (1c) are integrated with small time steps that are stable for the acoustic modes using a semi-implicit scheme in the vertical. All other terms are integrated over a large time step that is limited only by the advection velocity.

The PE simulations presented here are dry and adiabatic; all microphysical and subgrid-scale processes have been turned off. The only explicit dissipations are a weak time filter, as described by Robert (1966), with a filter coefficient of 0.2, and a weak horizontal ∇^4-diffusion which is turned on at day 5 of the integrations. Without the horizontal diffusion, the PE solutions become poorly resolved and noisy when the upper-level front collapses shortly after day 5.

While neither acoustic nor gravity modes are dynamically important for the phenomena considered here, using the compressible nonhydrostatic equations (1) has certain advantages. All the equations are prognostic and the set may be integrated as efficiently as the hydrostatic PE. In addition, adaptive-grid techniques, which will be necessary in our future work to resolve the small scales associated with fronts and moist convection, are greatly facilitated by the use of the fully prognostic set (1).

Numerical integration of the SG model is straightforward. After dropping the nonlinear term, (4) is discretized with second-order differences and then solved by fast Fourier transforms in each horizontal direction and inversion of the resulting tridiagonal equation in each vertical column. The potential-temperature advections on the horizontal boundaries are second-order in time and space and are computed by the positive-definite scheme of Smolarkiewicz (1984). The transformation back to physical coordinates is necessary only for plotting and is accomplished by simply transforming the computational grid to physical coordinates and interpolating to a uniform grid.

Unstable modes for both models are calculated iteratively by integrating forward for a day and then normalizing to a small reference value. The process is repeated until the resulting disturbance converges to the most unstable mode at a given zonal wavelength.

Both models produce accurate solutions to their respective equation sets. The PE solutions may be checked, at least qualitatively, against solutions for a similar basic-state jet in Polavarapu and Peltier (1990, their case HWF). For the SG model, our solutions show excellent quantitative agreement with HW. Thus, differences in the numerical techniques between the models can be ruled out as culprits in the comparisons.

4. PE and SG simulations

In this section, we present a direct comparison of PE and SG simulations of baroclinic waves. The SG simulation is identical to the "$\mu = 1$" case of HW, aside from the different numerical techniques used to obtain the solution. The initial conditions consist of identical zonal jets plus a small-amplitude perturbation,

which has the structure, for each model separately, of the linearly most unstable mode at the chosen zonal wavelength. In all cases, we choose the amplitude of the initial perturbation so that $\max(v_g) = 1.7$ m s^{-1}. Because the motions in the PE model are effectively hydrostatic and anelastic, and because both models use the same physical coordinates, differences in the simulations arise principally from the GM approximation and, to a lesser extent, from the approximate form of (4).

The basic state is a uniform potential vorticity zonal jet in geostrophic and hydrostatic balance. The zonal wind is given by Eq. (17) in HW, with their parameter $\mu = 1$, and is shown in Fig. 1, along with the corresponding potential temperature field. We choose all dimensional parameters as in HW, including: $q_{SG} = 1.28 \times 10^{-4}$ s^{-2}, $x_L = 4090$ km, $y_L = 5623$ km, and $H = 9$ km.

For these computations, the PE model uses 40 points in x, 60 in y, and 20 in z, which gives a grid spacing of roughly 100 km horizontally and 0.45 km vertically. The resolution in the SG model is the same, both in the geostrophic-coordinate computational grid and in the physical-coordinate grid used for interpolation. In any case, we have performed similar comparisons using other resolutions and have confirmed that the results are not sensitive to the resolution.

a. *The linear modes*

For a zonal wavelength of 4090 km, the most unstable PE mode grows with a 1.43 ± 0.04 d doubling time. The SG mode grows somewhat more quickly; its doubling time is 1.34 ± 0.02 d, which compares well with the 1.3 days quoted by HW. Moore and Peltier (1988) and Bannon (1989) have also shown that SG overestimates growth rates.

The structure of the modes is compared in Fig. 2. The data from both models is plotted in physical coordinates as deviations from the basic state, and the full (basic state plus perturbation) geostrophic velocities

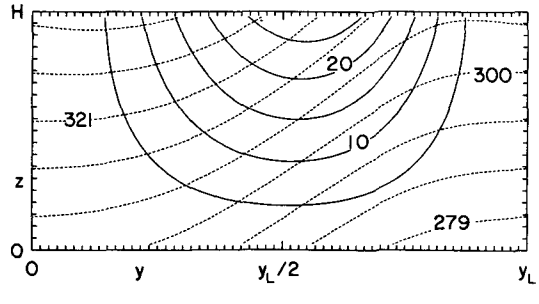

FIG. 1. The zonal velocity (solid lines; contour interval 3 m s^{-1}) and potential temperature (dashed lines; contour interval 7 K) for the HW basic-state jet with $\mu = 1$. The top and bottom of the plot correspond to the lowest ($z = 0.225$ km) and highest ($z = 8.775$ km) interior grid levels, respectively.

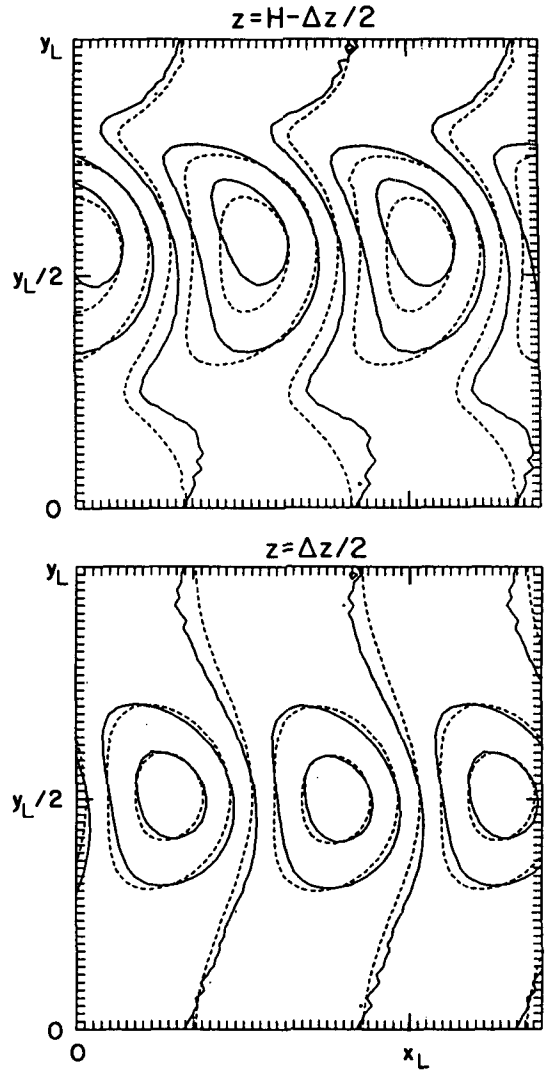

FIG. 2. The perturbation ϕ for the most unstable modes in the PE (solid lines) and SG (dashed) models. Contours of both fields are shown in physical coordinates for x, y planes at the levels $z = 0.225$ km (bottom) and $z = 8.775$ km (top). The contour interval is 2/5 of the (arbitrary) maximum amplitude of each mode and the modes' relative phase was chosen so that the zero contours cross at $y = y_L/2$, $z = 0.225$ km.

have been used in the transformation of the SG mode to physical coordinates. Both modes are essentially linear, having very small (but otherwise arbitrary) amplitude.

The amplitude of the PE mode is largest near the center of the domain at low levels and is shifted north of the jet axis near the lid, while the maximum amplitude of the SG mode lies on the jet axis at all levels. In addition, for the PE mode the NW–SE phase tilt and total phase change north of the jet axis dominates that to the south, in contrast to the SG mode, which has the same net phase change both north and south of its maximum amplitude.

This general structure appears to be characteristic of PE instabilities on baroclinic jets in a planar geometry. Gall (1977) used a jet with north–south symmetry about a vertical plane in physical coordinates and still found a PE mode similar to that in Fig. 2.

b. *Nonlinear evolution*

As the initial perturbations grow, the differences between the PE and SG simulations become more pronounced and continue to evolve. The perturbation pressure and full potential temperature in both models after three days are shown in Fig. 3, for levels near the surface and the lid. Although the disturbance is still only weakly nonlinear, the pressure field in the PE wave is distinctly asymmetric with the gradients and magnitude of ϕ' largest in regions of low pressure, as is commonly observed on synoptic maps. (Here, primes denote deviations from the basic state.) In contrast, the maximum and minimum perturbation pressure, as well as the maximum and minimum pressure gradient, have equal magnitudes in the SG wave, a property that follows directly from the inherent symmetries of the SG solution in geostrophic coordinates [see (23) in HW].

A more subtle, but perhaps more important distinction between the two simulations can be seen in the shape of the surface pressure contours. The PE wave already shows a tendency for northwesterly flow in the cold air west of the surface low and almost pure southerly flow in the warm air east of the low, while the strongest surface flow in the SG wave has an east–west component of the same magnitude to either side of the surface low. This behavior stems from the bias toward NW–SE phase tilts in ϕ' for the PE solution. Because u_g is negatively correlated with v_g, cold air is advected southward and eastward (warm air northward and westward) at the surface, and the accompanying surface low moves westward, in a frame of reference moving with the wave, toward the following high. The resulting geostrophic flow then becomes northwesterly to the west of the low and southerly to the east.

By day 6.3 (Fig. 4), the magnitude of the pressure

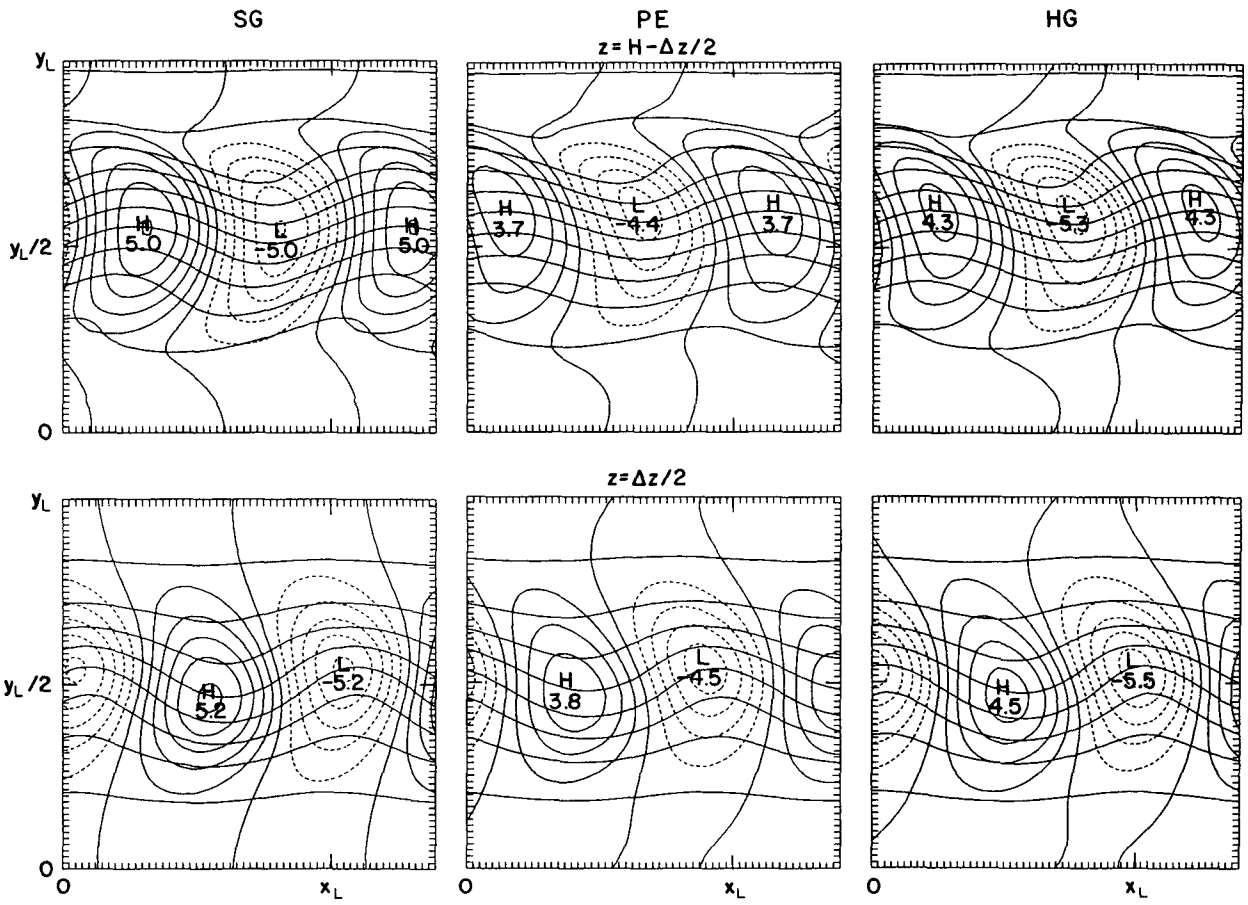

FIG. 3. SG (left), PE (center), and HG (right; see § 5) solutions at day 3. Contours of both the full θ (solid, "horizontal" lines; contour interval 5 K) and the perturbation ϕ (contour interval 100 m^2 s^{-2}) are shown for x, y planes at z = 0.225 km (bottom) and z = 8.775 km (top); maxima and minima in ϕ are labeled with units of 100 m^2 s^{-2}. To obtain ϕ in meters as in HW, divide by g = 9.8 m s^{-1}. Perturbation here refers to a deviation from the basic state.

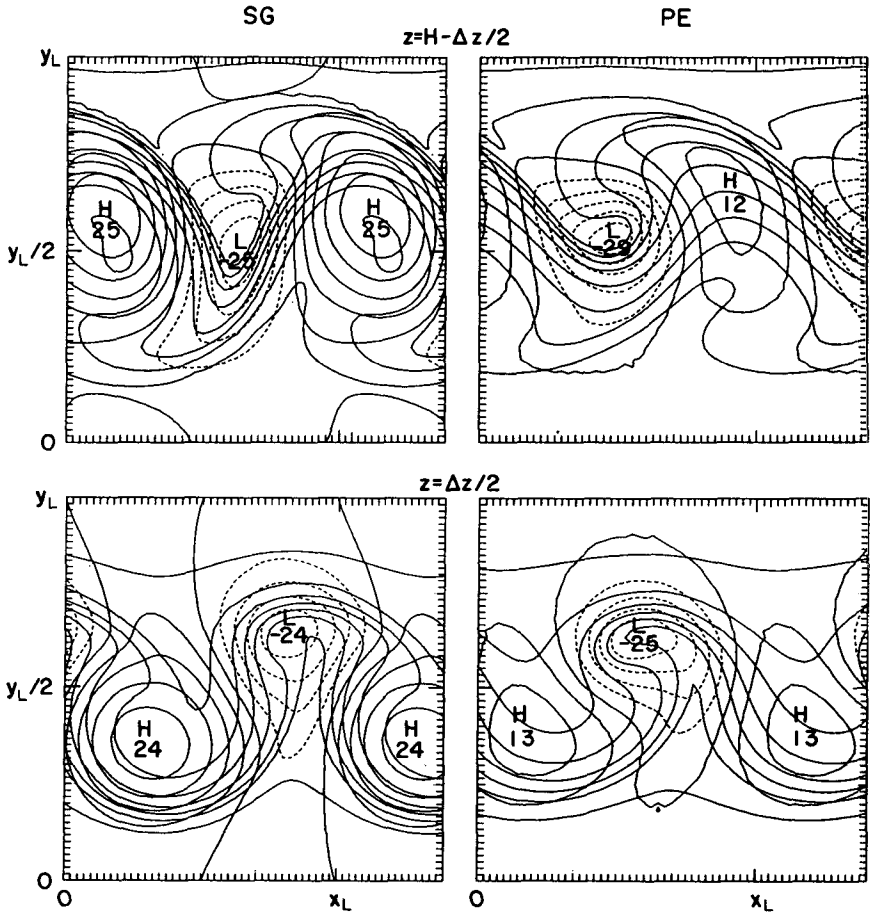

FIG. 4. SG (left) and PE (right) solutions at day 6.3. Plotting conventions as in Fig. 3, but with a contour interval of 500 $m^2 s^{-2}$ for ϕ.

perturbation in the trough at the lid has, in the PE wave, increased to more than twice that in the ridge, and a similar disparity has developed at the surface. Overall, the regions of low pressure have come to dominate the flow pattern, overwhelming the broad, weak highs.

At the same time, the PE wave has produced a pronounced cyclonic wrapping of the potential temperature contours at both the surface and the lid. At the surface, a warm "seclusion" pushes northward into the cold air while bending back to the west, and at the lid, a cold tongue extends southeastward into the warm air. The strongest frontal region lies downstream from the ridge at upper levels. In addition, prominent fronts are embedded within the surface temperature field, with the cold front stretching southwestward from just south of the low and a bent-back warm front situated to the northeast of the surface low, much as in the conceptual frontal model proposed by Shapiro and Keyser (1990). The vertical vorticity at the surface (not shown) extends in a continuous comma shape along the warm front and into the cold front.

Accompanying this evolution of the boundary temperature field is an alignment of the upper trough with the surface low, as the growth of the wave slows. In the ridge, however, the surface and upper-level patterns are increasingly separated by the pairing of the surface low (upper trough) with the high to its west (east). By day 6.3, the primary feature of the wave is a vertically aligned, cyclonic vortex.

The comparison to the SG solution at day 6.3 is also striking and illustrates how SG distorts even the qualitative features of the flow. As can be seen from Fig. 4, the SG wave maintains maximum and minimum pressures of equal magnitude separated by a zonal distance of one-half wavelength and bends the θ contours into the same qualitative shape, irrespective of the sign of the temperature anomaly, except for a relative reduction in the area of warm anomalies. The SG solution does not capture the character of the potential temperature field, does not exhibit a pairing of the cyclones and anticyclones near either the surface or the lid, and does not reproduce the strength of the troughs or the weakness of the highs. Moreover, the SG sim-

ulation produces a surface warm front only as a secondary phenomenon late in the development, and as the SG wave equilibrates, the westward phase shift with height vanishes in both the troughs and ridges. Indeed, although this and other SG simulations of baroclinic waves (Hoskins 1976; Heckley and Hoskins 1981) clearly demonstrate that SG includes physical processes, which are absent in QG models, the foregoing comparison with PE does not reveal significantly more quantitative accuracy than the comparison of QG with PE in Mudrick (1974).

We emphasize that the outcome of this comparison is not peculiar to the HW case. We have performed a variety of other simulations, using asymmetric basic-state jets as well as basic states with nonuniform potential vorticity and a realistic tropopause. In all cases the SG solutions exhibit similar deviations from the PE solutions, in the sense that these deviations are always in the same "direction" relative to the PE solutions.

5. Analysis

At this point, it is not completely apparent what has gone wrong in the SG solutions, nor is it obvious why these deficiencies have not appeared in previous two-dimensional studies. For example, the well-known shortcomings of QG are usually thought to originate in the lack of ageostrophic advection and stretching of relative vorticity, yet SG includes both of these at least approximately (Hoskins 1975). We now identify, using an expansion of the vorticity equation for small Rossby number, the crucial term neglected by SG.

a. Asymptotic analysis

Before scaling the equations, we split the potential temperature into a time-independent, horizontally uniform component, $\langle\theta\rangle$, and a deviation from this state, $\theta - \langle\theta\rangle$. Both the PE and the SG equations are then nondimensionalized using the following scales:

$$\left.\begin{aligned} &x, y, X, Y \sim L, \qquad z, Z \sim H, \qquad t \sim 1/\mathrm{Ro}f, \\ &u, v, u_g, v_g \sim U, \qquad w \sim HU/\mathrm{RiRo}L, \quad \phi \sim fUL, \\ &\langle\theta\rangle \sim \Theta, \quad \theta - \langle\theta\rangle \sim \Theta/\mathrm{RiRo}, \end{aligned}\right\} \tag{6}$$

where the Rossby number and Richardson number are, respectively,

$$\mathrm{Ro} = U/fL, \quad \mathrm{Ri} = g\Theta H/\theta_0 U^2.$$

Consistent with the typical midlatitude synoptic scaling, we also choose the horizontal length scale to be the radius of deformation,

$$L = (g\Theta H/f^2\theta_0)^{1/2};$$

this fixes the relative magnitudes of the Rossby and Richardson numbers at $\mathrm{Ro} = \mathrm{Ri}^{-1/2}$.

In what follows, our procedure will be to separate the velocity, vorticity, and advections into geostrophic and ageostrophic parts, denoted by subscripts g and a, respectively, and then to expand both the PE and the SG equations in small Rossby number. As in McWilliams and Gent (1980), we do not make the formal step of writing the dependent variables as power series in Ro. It should also be noted that the dimensional parameters in section 4a give a Rossby number of roughly 0.3, although inferences based on the small Ro expansions will prove quite useful.

With the choice of scales given in (6), the horizontal ageostrophic velocities, u_a and v_a, scale as RoU. The inertial terms in the PE horizontal momentum equations then become in nondimensional form (McWilliams and Gent 1980),

$$\mathrm{Ro}d_g\mathbf{v}_g + \mathrm{Ro}^2[\mathbf{v}_a\cdot\nabla\mathbf{v}_g + d_g(u_a, v_a)] + \mathrm{O}(\mathrm{Ro}^3), \tag{7}$$

where $d_g = \partial/\partial t + \mathbf{v}_g\cdot\nabla$ and $\mathbf{v}_a = (u_a, v_a, w)$. Both SG and QG retain the O(Ro) geostrophic advection of geostrophic momentum. However, at O(Ro^2), QG neglects all terms, while SG keeps the first term in parentheses but omits the second, and thus has the same formal accuracy as QG [i.e., through O(Ro)].[1] Diagnosing the O(Ro^2) terms from the PE solution confirms this formal analysis; as shown in Fig. 5, $d_g\mathbf{v}_a$ and $\mathbf{v}_a\cdot\nabla\mathbf{v}_g$ have comparable magnitude at day 3. Moreover, the relative magnitude of these terms is much the same throughout the integration from day 0 to day 6.3 (not shown).

Readers familiar with Hoskins (1975) may recall at this point that SG is justified there by a Lagrangian analysis of the momentum equations. However, while the resulting equations retain certain terms neglected by QG, the terms neglected by SG have the same order of magnitude as those neglected by QG, as verified in Fig. 5. In appendix B, we show that QG may be derived with the same Lagrangian analysis as long as only the leading-order, geostrophic velocity is used in the transformation of the equations to Eulerian form.

The dynamical effects of these small errors in the SG momentum equations are most easily isolated in the vorticity and potential vorticity equations. Separating the geostrophic contribution, the nondimensional PE and SG vorticities are by definition

$$\zeta = \mathbf{k} + \mathrm{Ro}\zeta_g + \mathrm{Ro}^2\zeta_a;$$

$$\zeta_{\mathrm{SG}} = \mathbf{k} + \mathrm{Ro}\zeta\boldsymbol{v}_g + \mathrm{Ro}^2\mathbf{J}_{xyz}(\phi_x, \phi_y).$$

The Jacobian terms in ζ_{SG} are formally the same size as the ageostrophic vorticity; since the PE vorticity equation has the same form as (2), asymptotic expan-

[1] McWilliams and Gent (1980) consider QG to be the leading-order theory and corrections to QG to be O (Ro). In our nomenclature, geostrophy is the leading-order balance, QG is an O (Ro) theory, and corrections to QG occur at O (Ro^2).

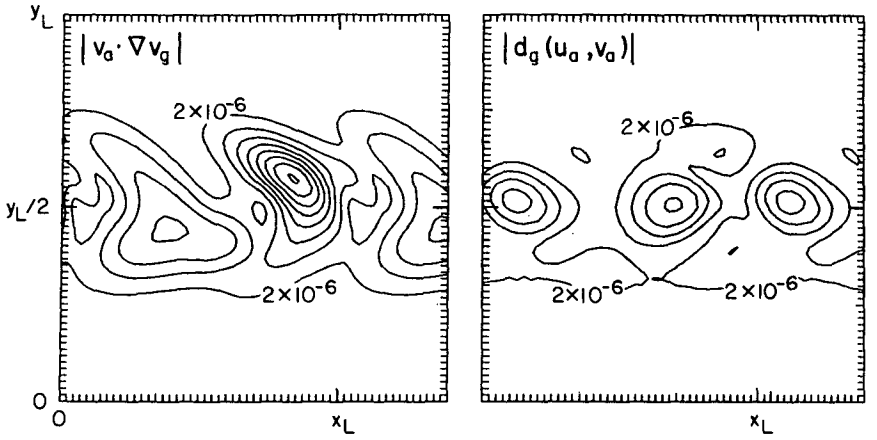

FIG. 5. Contours of the magnitude of the $O(Ro^2)$ terms in (7) for the x, y plane at $z = 0.225$ km in the PE solution. Shown are $|v_a \cdot \nabla v_g|$ (left) and $|d_g(u_a, v_a)|$ (right) at day 3; the contour interval for both fields is 2×10^{-6} m s^{-2}.

sions of each will differ only in their treatment of terms involving ζ_a [or $J_{xyz}(\phi_x, \phi_y)$].

For further comparison, the leading-order contribution to the ageostrophic vorticity must be expressed in terms of ϕ. After nondimensionalization, the momentum equations (1a) yield

$$\zeta_a = -(d_g u_{gz} + \mathbf{v}_{gz} \cdot \nabla u_g,\ d_g v_{gz} + \mathbf{v}_{gz} \cdot \nabla v_g, \\ -\mathbf{v}_{gx} \cdot \nabla u_g - \mathbf{v}_{gy} \cdot \nabla v_g) + O(Ro).$$

Using the thermodynamic equation (1c) to evaluate $d_g \mathbf{v}_{gz}$ to leading order then produces the unexpected result that

$$\zeta_a = -2\mathbf{J}_{xyz}(\phi_x, \phi_y) + O(Ro). \tag{8}$$

Surprisingly, the nonlinear term in ζ_{SG} has the same form as the ageostrophic, $O(Ro^2)$ contribution to the PE vorticity, but has the wrong coefficient, which should be -2 instead of $+1$.

The nondimensional PE vorticity satisfies

$$Ro d_g \zeta_g + Ro^2(\mathbf{v}_a \cdot \nabla \zeta_g + d_g \zeta_a) \\ = \mathbf{v}_{gz} - \mathbf{k} \times \nabla\theta + Ro(\mathbf{v}_{az} + \zeta_g \cdot \nabla \mathbf{v}_g) \\ + Ro^2(\zeta_g \cdot \nabla \mathbf{v}_a + \zeta_a \cdot \nabla \mathbf{v}_g) + O(Ro^3). \tag{9}$$

We present the three-dimensional equation for generality, but remind the reader that the vertical component of vorticity is the controlling factor in the dynamics when the scaling (6) is appropriate, since vertical gradients of θ are then $O(Ro^{-1})$ larger than horizontal gradients and the vertical component of vorticity enters the potential vorticity equation at lower order than the horizontal components [e.g., see (10)]. The balance through $O(Ro)$ in (9) reproduces the QG vorticity equation and is independent of the ageostrophic vorticity. In the vertical component, the first corrections to QG are the ageostrophic advection, tilting and stretching of geostrophic vorticity, *and* the geostrophic advection of ageostrophic vorticity.

A similar expansion through $O(Ro^2)$ of the nondimensional form of (2) looks exactly like (9), except ζ_a is replaced by $\mathbf{J}_{xyz}(\phi_x, \phi_y)$. By (8), SG thus has the correct form for all $O(Ro^2)$ contributions to the vorticity equation but retains the wrong coefficient on terms involving $\mathbf{J}_{xyz}(\phi_x, \phi_y)$, which in the PE are derived from the ageostrophic vorticity. [McWilliams and Gent (1980) discuss a similar result for the potential vorticity equation.] Hoskins (1975) suggests that, consistent with the neglect of $J_{xy}(\phi_x, \phi_y)$ in (4), the Jacobian terms in ζ_{SG} are small and should be omitted; but this analysis shows that these terms enter at the first correction to QG, with dynamical effects potentially as large, for example, as the stretching of relative vorticity. It should be emphasized that our analysis strictly pertains to the GM equations; the effects of the further approximation of omitting the nonlinear terms in (4) are discussed in appendix A.

It is also interesting to consider the nature of the errors in the SG potential vorticity equation. Given (8), it is easy to show

$$q_{SG} = q - \frac{3}{2} Ro^2 \zeta_a \langle\theta\rangle_z + O(Ro^3), \tag{10}$$

where $q = \zeta \cdot \nabla\theta$ is the Ertel potential vorticity for the PE, and has been nondimensionalized by $f\Theta/H$. Since the PE conserve q, and q_{SG} is conserved in SG, (10) implies that in the SG model there is an $O(Ro^2)$ source, proportional to ζ_a, for q. Again, we see that errors in SG are closely linked to the ageostrophic vertical vorticity and, through (8), to the presence of curvature in the flow. For example, if the flow is nearly straight (say, if the coordinate axes can be rotated so that $\partial/\partial x \sim Ro\ \partial/\partial y$), then $J_{xy}(\phi_x, \phi_y)$ is $O(Ro)$; the ageostrophic contribution to the vertical vorticity is $O(Ro^3)$;

and the SG potential vorticity equation is accurate through $O(Ro^2)$.

b. Inferences

Even if the initial conditions in both the PE and SG models were in some sense identical, the simulations would immediately begin to diverge because of the differences between the models in the treatment of the ageostrophic vorticity. These differences are clearly the original cause of the deviations of the SG solution from the PE solution in section 4, although after several days the solutions have diverged sufficiently so that the subsequent evolution of the deviations is not determined solely by differences in the model equations. Differences between the PE and SG solutions, or at least their initial tendencies, can therefore be explained by considering the effects of the Jacobian terms in the vertical vorticity and potential vorticity equations.

Before beginning, it is convenient to illustrate the relation between ϕ, $J_{xy}(\phi_x, \phi_y)$, and the ageostrophic vorticity with three simple examples. First consider an axisymmetric vortex with $\phi = (\alpha/2)(x^2 + y^2)$, where $\alpha > 0$ implies low central pressure and cyclonic flow, while $\alpha < 0$ implies high pressure and anticyclonic flow. For this flow,

$$J_{xy}(\phi_x, \phi_y) = \alpha^2, \quad \zeta_a = -2\alpha^2 + O(Ro).$$

Thus, the Jacobian is positive and the ageostrophic vorticity is negative regardless of the sense of the flow (cyclonic or anticyclonic), consistent with the familiar rule that flow in gradient balance around a low (high) is subgeostrophic (supergeostrophic). Next, consider a horizontally periodic pressure field of the form $\phi = P(x, y)$, where

$$P(x, y) = \sin kx(1 - \cos ly), \qquad (11)$$

and $k = 2\pi/x_L$, $l = 2\pi/y_L$. Retaining only the leading-order term, it follows that

$$\zeta_a = -2k^2l^2[(1 - \cos ly)\sin^2 kx - \sin^2 ly];$$

ϕ and ζ_a are shown in Fig. 6a. Note that near extrema (of either sign) in the pressure field, where the flow approximates the axisymmetric vortex of the previous example, $J_{xy}(\phi_x, \phi_y)$ is positive and ζ_a is negative, but that near the saddle points in ϕ, where the flow is highly confluent $J_{xy}(\phi_x, \phi_y)$ is negative and ζ_a is cyclonic. Further discussion may be found in Xu (1990, section 3c). Finally, let ϕ be the superposition of an x-independent component, corresponding to a barotropic zonal jet and a small perturbation with the same periodic form as above:

$$\phi = -y + l^{-1}\sin ly + \epsilon P(x, y), \quad \epsilon \ll 1. \qquad (12)$$

Neglecting terms of order ϵ^2 and order Ro, we have

$$\zeta_a = -2\epsilon k^2 l \sin ly P.$$

As shown in Fig. 6b, the ageostrophic vorticity in this case is antisymmetric across the jet, having the same sign as the "perturbation" ϕ on the cyclonic-shear (north) side of the jet, and the opposite sign on the anticyclonic-shear side.

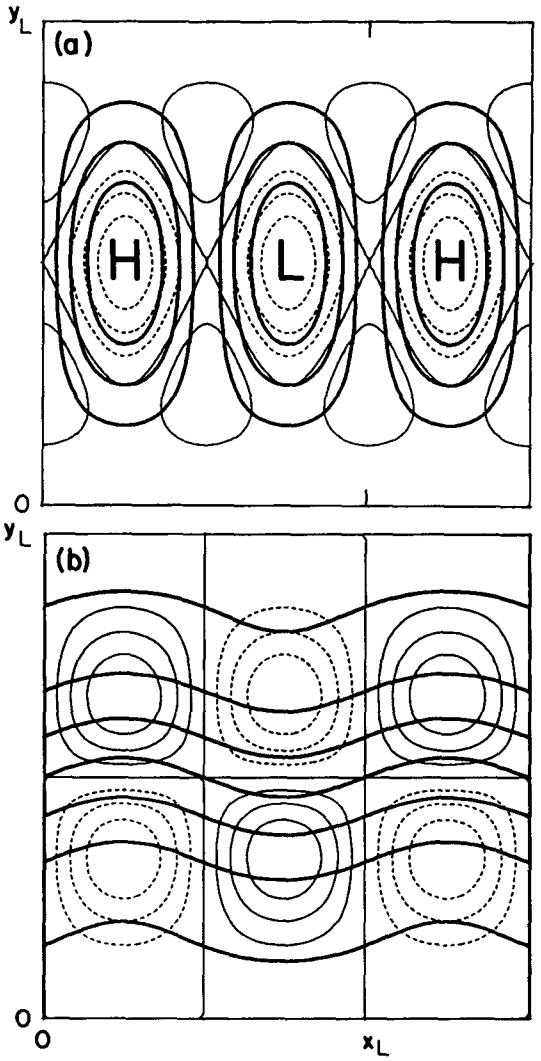

FIG. 6. The relation of ϕ (heavy solid lines) and ζ_a (thin lines, negative contours dashed) in two simple cases. (a) ϕ, given by (11), is horizontally periodic ϕ. The contour interval for both fields is one-fourth of the respective maximum amplitudes and the zero contour for ϕ is suppressed. (b) ϕ, given by (12), is the superposition of an x-independent and a small-amplitude, horizontally periodic component. Contours as in (a), except the contour interval for ϕ, are slightly less than one-sixth of its total change across the domain.

Using these examples and the differences between PE and SG in the vorticity equation at $O(Ro^2)$, one can understand the tendency, relative to the SG solutions, for lows to be deeper and highs to be weaker in the PE solutions. Our arguments are based on the assumption that, since $\zeta_g = \nabla^2\phi$, extrema in ϕ generally follow the location and magnitude of extrema in ζ_g.

We first note that there is a strong asymmetry be-

tween high and low pressure within the PE solutions. However, the stretching of relative vorticity provides an obvious mechanism for producing a field of ζ_g in which cyclonic vorticity is relatively strong and anticyclonic vorticity relatively weak, and an associated pressure field in which lows are deep and highs are weak. Of course, this suggests that even in the SG simulations lows should be deeper than highs, since SG includes stretching of ζ_g; reference to the SG solutions in appendix A, which use the nonlinear form of (4), shows that this is indeed the case.

The relative tendency for deeper lows in the PE solutions, compared to the SG solutions, may now be qualitatively explained as follows. Suppose that both the PE and SG models were initialized with identical mass fields, so that all geostrophic quantities are initially the same in both models. The subsequent evolution of the geostrophic vorticity is governed by

$$\frac{d\zeta_g}{dt} = \begin{cases} (\mathbf{k} + \mathrm{Ro}\zeta_g)\cdot\nabla w - \mathrm{Ro}\dfrac{d\zeta_a}{dt} + \mathrm{O}(\mathrm{Ro}^2), & \text{in PE;} \\ (\mathbf{k} + \mathrm{Ro}\zeta_g)\cdot\nabla w + \dfrac{1}{2}\mathrm{Ro}\dfrac{d\zeta_a}{dt} + \mathrm{O}(\mathrm{Ro}^2), & \text{in SG,} \end{cases} \tag{13}$$

where we have canceled a factor of Ro relative to (9). Because the ageostrophic velocities are correct to leading order in SG, the material derivative, d/dt, in SG agrees with that of the PE through O(Ro) initially; therefore, only the terms involving ζ_a contribute at leading order to the difference in the initial change of ζ_g between the two models. If we now follow a fluid parcel, say one that is approaching a nearby minimum in ϕ and maximum in ζ_g, the stretching and tilting terms will produce changes of ζ_g that are essentially identical in both models. However, since ζ_a is most negative near extrema in ϕ and ζ_g, $d\zeta_a/dt$ will be negative as the parcel approaches the low and these terms will thus in the PE tend to increase the already positive geostrophic vorticity, while having the opposite tendency in the SG model. For a parcel approaching a maximum in ϕ, $d\zeta_a/dt$ will also be negative, but now will tend to decrease (increase) the magnitude of the geostrophic vorticity in the PE (SG) solution, since ζ_g is locally negative. We conclude that inverting ζ_g for ϕ soon after initialization will reveal an instantaneous tendency for relatively deeper lows and weaker highs in the PE simulations than in the SG.

One can also imagine initializing both models with identical fields of their respective vorticities (ζ in the PE, and ζ_{SG} in SG). In this case, the initial ϕ field will exhibit relatively deeper lows and weaker highs in the PE model than in SG because of the different balances inherent in the two models. More specifically, the divergence equation can be written as

$$\nabla^2\phi = \begin{cases} \zeta + 2\,\mathrm{Ro}J_{xy}(\phi_x, \phi_y) + \mathrm{O}(\mathrm{Ro}^2), & \text{in PE;} \\ \zeta_{SG} - \mathrm{Ro}J_{xy}(\phi_x, \phi_y), & \text{in SG.} \end{cases}$$

Again, the fact that $J_{xy}(\phi_x, \phi_y)$ is generally positive (or equivalent, that ζ_a is generally negative), near extrema in the vorticity and ϕ will cause the PE solutions to have a bias toward lower pressure compared to those of SG.

Consideration of the potential vorticity equation leads to an alternative explanation of the bias, relative to the SG solution, toward low pressure in the PE solution. As previously discussed, (10) shows that there is an $\mathrm{O}(\mathrm{Ro}^2)$ source of q_{SG} in the PE model. Thus, our PE solutions for the $\mathrm{HW}\mu = 1$ case maintain uniform q (up to truncation errors) but should possess significant internal anomalies of q_{SG}, which by (10) should have the same structure and opposite sign to ζ_a. If the ageostrophic vorticity is anticyclonic as expected near the centers of both lows and highs, there will then be positive q_{SG} anomalies over extrema in ϕ; these anomalies will tend to reduce strength of highs and increase strength of lows, compared to the SG simulation where there are no interior anomalies in q_{SG}.

Figure 7 shows $q_{SG} - q$ near the surface, diagnosed from the PE solution at day 6. No assumption of small Ro has been made in the calculation and $q - q_{SG}$ includes all the higher-order terms beyond $\mathrm{O}(\mathrm{Ro}^2)$ that are implicit in (10). Nevertheless, the dominant features are qualitatively as predicted by the structure of ζ_a: an intense positive anomaly above surface low and

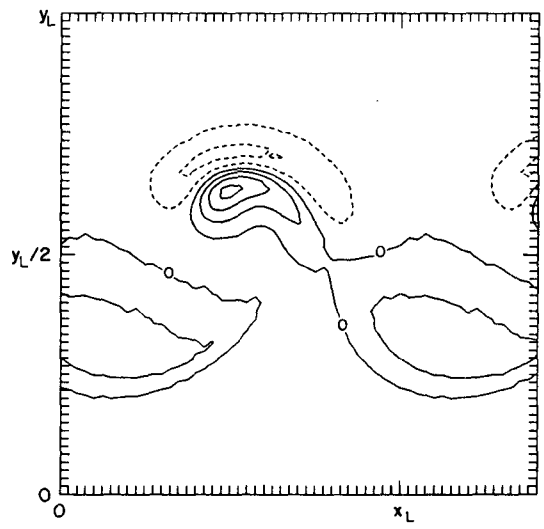

FIG. 7. The deviation of the SG potential vorticity from the Ertel potential vorticity ($q_{SG} - q$) diagnosed from the PE solution at day 6. The field is shown at $z = 0.225$ km with a contour interval of 0.03 PVU and labels in 10^4 PVU. The Ertel potential vorticity is 0.345 PVU.

a broader positive anomaly above the surface high. Note also that along both the warm and cold fronts, where the flow is highly confluent, ζ_a is cyclonic and anomalies of q_{SG} are negative.

The nonuniformity of q_{SG} in the PE solutions will affect traditional two-dimensional analyses of the fronts. Indeed, since the anomalies of q_{SG} originate in the three-dimensional large-scale flow associated with the wave, variation of q_{SG} across the front will be characteristic of fronts embedded in realistic baroclinic waves. These ideas deserve further attention.

Other aspects of the differences between the PE and SG simulations, especially when the comparison is made at large amplitude, have a less direct connection with the instantaneous divergence of the PE and SG dynamics. A prime example of this is the characteristic, "cyclonically wrapped" structure of the boundary-temperature field in the mature PE wave. This structure is partially dependent on the shape of the linear PE wave; both the NW–SE phase tilts and the displacement of the maximum wave amplitude into the cyclonic-shear side of the jet produce westward advection of warm anomalies and an eastward advection of cold anomalies relative to the wave (at the lid, the sense of the temperature anomalies is reversed), and favor the wrapping of temperature contours preferentially around the surface low and the upper trough. Starting from a small-amplitude perturbation, SG cannot duplicate these effects because it fails to reproduce the structure of the linear wave. However, the characteristic cyclonic wrapping in the PE simulations is not produced only by the modal structure. Because of the intense circulation associated with the regions of low pressure, which develops largely independent of the modal shape, the mature PE disturbance inherently favors the bent-back temperature and frontal structure seen in Fig. 4, as the temperature contours are advected around the northern side of the surface cyclone and the southern side of the upper trough. Indeed, we have examined the development of unstable waves on other jets, for which the linear PE wave has no horizontal tilts, and the PE simulations still exhibit cyclonic wrapping of the boundary temperature and bent-back warm-frontal structure in the final stages of their growth. In contrast, location of the fronts within the SG solutions in these cases is much more sensitive to the details of the basic-state jet, particularly the sense of the barotropic shear across the baroclinic zone. Davies et al. (1991) present a detailed discussion of this sensitivity to barotropic shear in SG simulations of baroclinic waves.

6. A hypogeostrophic model

If the foregoing asymptotic analysis has some validity, solutions generated by a theory, which includes all $O(Ro^2)$ corrections to QG, should, when compared with those of the SG model, agree more closely with the PE solutions. We will call such a theory hypogeostrophy (HG). McWilliams and Gent (1980) defined HG to be an asymptotic expansion of the PE, retaining only terms up to $O(Ro^2)$, but we use HG to mean simply that the model is accurate through $O(Ro^2)$.

a. *The HG model*

Instead of the SG vorticity equation (2), we propose

$$Ro\frac{d\zeta_{HG}}{dt} = \zeta_{HG}\cdot\nabla\mathbf{v} - \mathbf{k}\times\nabla\theta, \tag{14}$$

where

$$\zeta_{HG} \equiv \mathbf{k} + Ro\zeta_g - 2\,Ro^2\mathbf{J}_{xyz}(\phi_x, \phi_y)$$
$$= \zeta_{SG} - 3\,Ro^2\mathbf{J}_{xyz}(\phi_x, \phi_y).$$

Since ζ_{HG} includes the correct leading-order contribution from ζ_a [see (8)], asymptotic expansions for small Ro of the HG and PE vorticities agree through $O(Ro^2)$, and hence (14) has $O(Ro^2)$ accuracy as well. The form of (14) and the definition of ζ_{HG} were chosen to provide a potential vorticity analog conserved following parcels, and to ensure that the equations remained uncomplicated in geostrophic coordinates.

Using (14) and the full thermodynamic equation (1c), we may form a potential vorticity equation similar to (3):

$$\frac{dq_{HG}}{dt} = 0; \quad q_{HG} = \zeta_{HG}\cdot\nabla\theta. \tag{15}$$

As with the vorticity equation, asymptotic expansions of (15) and the conservation equation for Ertel potential vorticity agree through $O(Ro^2)$. At $O(Ro^3)$, (15) retains the correct contribution from the horizontal ageostrophic vorticity, but does not include the higher-order corrections to the vertical vorticity necessary for $O(Ro^3)$ accuracy.

The HG equations assume a more convenient form in geostrophic coordinates. The material derivative and the HG potential vorticity become

$$d/dt = \partial/\partial T + (u_g + O(Ro^2))\partial/\partial X + (v_g + O(Ro^2))\partial/\partial Y + Row\partial/\partial Z,$$

and

$$q_{HG} = \frac{Ro(\Phi_{ZZ} - 3\,Ro^2\mathbf{J}_{XYZ}(\Phi_X, \Phi_Y)\cdot\nabla\Phi_Z)}{1 - Ro\,(\Phi_{XX} + \Phi_{YY}) + Ro^2J_{XY}(\Phi_X, \Phi_Y)}. \tag{16}$$

Notice that without GM the horizontal advections in geostrophic coordinates are not purely geostrophic. We will make the approximation of neglecting the non-geostrophic terms in the horizontal advections; since these terms are $O(Ro^2)$ compared to $\mathbf{v}_g$, such an approximation does not degrade the asymptotic accuracy of HG. With this approximation in the material derivative, the boundary conditions for HG are given by

(5), and thus HG in geostrophic coordinates is mathematically equivalent to SG, except for the different relation between Φ and the potential vorticity.

If q_{HG} is uniform, we set $\Phi = \frac{1}{2}\mathrm{Ro}^{-1}q_{HG}Z^2 + \tilde{\Phi}$ and, after canceling a factor of Ro, (16) may be written as

$$\frac{1}{q_{HG}}\tilde{\Phi}_{ZZ} + \tilde{\Phi}_{XX} + \tilde{\Phi}_{YY} - 4\,\mathrm{Ro}J_{XY}(\tilde{\Phi}_X, \tilde{\Phi}_Y) - \frac{3\,\mathrm{Ro}^2}{q_{HG}}\frac{\partial(\tilde{\Phi}_X, \tilde{\Phi}_Y, \tilde{\Phi}_Z)}{\partial(X, Y, Z)} = 0. \quad (17)$$

Comparing the above to the nondimensional form of (4) (after a similar decomposition of Φ) shows that the SG model again has the correct form, but the wrong coefficient for the O(Ro) term. We will, for convenience, neglect the $O(\mathrm{Ro}^2)$ term in (17), since it corresponds to an $O(\mathrm{Ro}^3)$ contribution to q_{HG} and we are primarily concerned with maintaining the $O(\mathrm{Ro}^2)$ asymptotic accuracy of HG. However, integrations using the full (17) do have certain advantages (principally, slower growth of perturbations), as might be expected given that the $O(\mathrm{Ro}^2)$ terms in (17) are present in an asymptotic expansion of the Ertel potential vorticity.

The terms through O(Ro) in (17) could also be derived as an asymptotic expansion of the Ertel potential vorticity in geostrophic coordinates, using (8) transformed to geostrophic coordinates to express the ageostrophic vertical vorticity in terms of Φ. The derivation given has the advantage of providing, in physical coordinates, vorticity and potential vorticity equations identical in form to those of the PE.

b. *Hypogeostrophic simulations*

The HG model may now be applied to the HW $\mu = 1$ case, allowing us to evaluate the validity of the foregoing asymptotic analysis. The numerical code for the HG model derives directly from that of the SG model, except that (17) is solved iteratively, by evaluating the nonlinear Jacobian terms from Φ at the previous iteration.

The characteristics of the linear HG mode agree closely with those of the PE mode. For a wavelength of 4090 km, the doubling time of the HG mode, at 1.39 ± 0.04 d, falls between the PE and SG growth rates. More importantly, the HG mode accurately approximates the structure of the PE mode. As can be seen from Fig. 8, which compares the perturbation ϕ near the lid from HG and the PE, the linear HG wave reproduces both the NW–SE tilts and the displacement of the maximum amplitude to the north of the jet axis. At the surface, the structures of the HG and PE modes are equally close.

The linear HG mode differs from that of the SG model only for basic-state flows with horizontal shear. Linearizing (17) about a zonal basic state gives

$$\frac{1}{q_{HG}}\Phi_{ZZ} + (1 - 4\,\mathrm{Ro}\bar{\Phi}_{YY})\Phi_{XX} + \Phi_{YY} = 0, \quad (18)$$

where the basic-state quantities are denoted with bars, and terms $O(\mathrm{Ro}^2)$ have been neglected. The linearized boundary conditions [from (5)] are the same as in the SG model, but compared to the linear form of (4), (18) includes a variable coefficient multiplying Φ_{XX}. This coefficient appears because, when $\bar{\Phi}_{YY}$ is nonzero, the linearization of $J_{XY}(\Phi_X, \Phi_Y)$ is nonzero and there is a contribution to the ageostrophic vorticity at first order in perturbation amplitude (see Fig. 6b). When the Jacobian terms are kept in the SG model (see appendix A), (4) has a linear form identical to (18), but of course has the incorrect coefficient on the O(Ro) term. Both Eliassen (1983) and Joly and Thorpe (1990) retain this variable coefficient in their linearized equations for the SG potential vorticity.

The differences between the HG and SG modes can be qualitatively explained using (13), which has the same form for both the HG and PE models. Our arguments are essentially the same as in section 5, where the bias toward low pressure in PE solutions relative to SG solutions was diagnosed, except that the relationship between ϕ and ζ_a has changed.

At levels where the basic-state flow is nonzero, the ageostrophic vorticity will have the opposite sign to the perturbation pressure (positive in the troughs) on the cyclonic-shear side of the jet, but the relation will be reversed on the anticyclonic-shear side; the resulting pattern is much as in Fig. 6b. If we again follow a parcel moving toward a trough, $d\zeta_a/dt$ will be negative to the north of the jet axis and positive to the south. Reference to (13) then shows that the perturbation ζ_g (and presumably the magnitude of the perturbation ϕ) in the

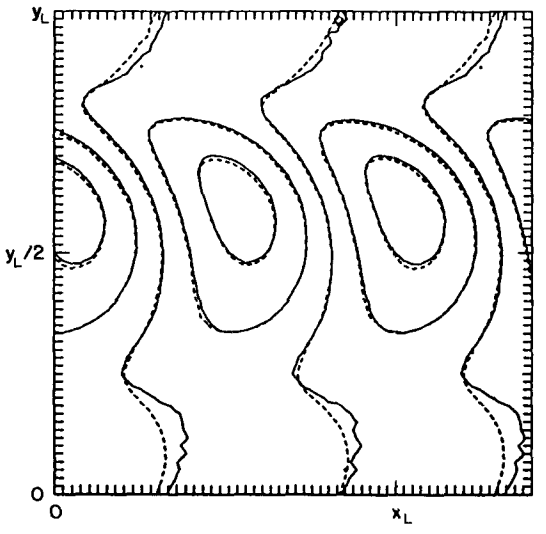

FIG. 8. As in Fig. 2, but for the PE (solid lines) and HG (dashed) modes at $z = 8.775$ km.

HG mode will be shifted toward the cyclonic-shear side of the jet, relative to the linear SG wave. This leads directly to the bias in the PE or in HG toward maximum amplitude north of the jet axis and NW–SE phase tilts, as shown in Fig. 2. The full modal structure for the PE and HG modes, in which NW–SE phase tilts are dominant even at the surface (where the linear ageostrophic vorticity is zero), further depends on two effects that we have so far ignored: the vertical coupling inherent in the inversion of q_{HG} and the increased tilting of the perturbation θ at the lid when its maximum is displaced into the cyclonic shear of the jet.

Although the disturbance in the HG model has grown too quickly, the comparison of HG with PE at day 3 is otherwise excellent. As shown in Fig. 3, the HG solution again captures the characteristic structure produced by the PE simulation, including the relative strength of the low pressure at both the surface and the lid, the continuing bias toward NW–SE phase tilts, and the development of warm, southerly flow and cold, northwesterly flow at the surface. In light of our previous assertion that the nonlinear evolution in the PE simulation to day 3 is determined in large part by the NW–SE phase tilts of the linear mode, which inherently produce westward (eastward) advection of warm (cold) air at the boundaries, the continuing similarity of structure in the HG and PE solutions is not surprising.

Unfortunately, we have not integrated the HG model much beyond day 3.5. Because of its nonlinearity, (17) changes from elliptic to hyperbolic type in the region of the nascent upper-level front after roughly 3.5 d. The iterative procedure for solving (17) fails to converge at this point, although the solutions continue to look reasonable, even a single time step before the iterative solver fails. Although the change of type of (17) may correspond to some real physical phenomenon, such a change can occur only when and where the nonlinear terms in (17) are comparable to the linear terms. Thus, (17) will cease to be elliptic in precisely those locations where the small-Ro expansion upon which HG is based converges most slowly and where the higher-order terms that HG neglects are largest. In these regions, the HG dynamics need not correspond to the PE dynamics and will instead depend mainly on the approximations used to derive HG.

6. Conclusion

Side-by-side PE and SG simulations of a three-dimensional baroclinic wave are significantly different throughout the integration. Errors are present in the linearized SG modes and these errors grow and evolve as the wave develops beyond the linear regime. In contrast to the SG solutions, the PE wave has distinctive NW–SE phase tilts, especially at upper levels, throughout much of its development, and at finite amplitude, develops deep lows and weak highs. At large amplitude, the PE wave has a strong tendency to "seclude" tongues of warm air at the surface and cold air at the lid, eventually resulting in a vertically aligned, cyclonic vortex. These characteristics of the PE solutions appear in many other PE channel integrations (Mudrick 1974; Takayabu 1986; Keyser et al. 1989; Polavarapu and Peltier 1991) and the embedded frontal structure is similar to the conceptual model proposed by Shapiro and Keyser (1990).

The distinctive large-amplitude structure of the PE wave is influenced by the strong tendency in the smaller amplitude wave for horizontal streamlines to tilt NW–SE and to be displaced into the cyclonic-share side of the jet. As discussed in section 5, the resulting transport of warm (cold) air northward and westward (southward and eastward) on the horizontal boundaries contributes to the characteristic cyclonic wrapping of the temperature contours in the PE solutions. Here we are in partial agreement with HW, who conclude that the horizontal phase tilts in the small-amplitude wave are crucial in determining the ultimate structure of the SG solutions. However, the intense circulation around regions of low pressure that develops in the PE also strongly influences the mature structure of the disturbance and is probably the dominant influence at large amplitude.

A small-Ro asymptotic analysis of the vorticity and potential vorticity equations, following McWilliams and Gent (1980), demonstrates that SG includes all the leading-order corrections to QG, except those involving the ageostrophic vertical vorticity. This asymptotic analysis is not restricted to baroclinic waves and we expect that there will be similar differences between PE and SG solutions for other three-dimensional flows with similar Rossby numbers. We confirm the analysis by deriving and solving equations for an HG model, which correctly includes the leading-order dynamical contributions of the ageostrophic vorticity. Although our solution techniques for HG do not converge for large amplitude disturbances, the HG solutions reproduce the characteristic features of the first 3.5 days of the PE development.

We stress that differences between the PE and SG simulations are not caused by a "break down" of the SG approximation, such as Hoskins and Bretherton (1972) predict near frontal collapse. Rather, since SG has limited asymptotic accuracy (formally, the same as QG), the SG and PE solutions steadily diverge due to the accumulation of small errors by SG. A fair estimate of the size of these errors and thus, of the applicability of the SG model, is the magnitude of the ratio of vertical ageostrophic vorticity to f.

When compared with PE solutions, SG simulations have been markedly more successful for two-dimensional frontal flows than for the three-dimensional baroclinic wave case presented here. However, it must be remembered that frontal flow, which is characterized by small ratios of cross- to alongfront velocity and length scales, and a Rossby number of order unity,

represents a different dynamical regime than the flow within a developing baroclinic wave, which has a small or moderate Rossby number and nearly isotropic horizontal scales except near fronts. In fact, Hoskins and Bretherton (1974) derive the SG equations for frontal flow by neglecting the curvature of the front and then systematically retaining only terms of leading or first order in ϵ, where ϵ is the ratio of cross- to alongfront velocities (see also Pedlosky 1979). McWilliams and Gent (1980) formalize and extend the Hoskins–Bretherton analysis. They show that the SG dynamics are accurate only through $O(\epsilon)$ (in their terminology, through leading order) when the ratio of l, the cross-front length scale to r—the radius of curvature of the front—is $O(\epsilon)$ but, when the front is nearly straight, i.e., $l/r \geq O(\epsilon^2)$, SG gains an additional order of accuracy in an expansion for small ϵ. Thus, SG is a higher order approximation for two-dimensional fronts than for three-dimensional baroclinic waves, making errors of $O(\epsilon^3)$ in a nearly straight front but errors of $O(\mathrm{Ro}^2)$ in the three-dimensional wave.

Finally, we note that other weak effects are probably important in determining the structure and evolution of dry baroclinic disturbances. Specifically, spherical geometry should be influential, especially for long waves. Evidence for the importance of sphericity exists, for instance, in the linear stability analysis of Simmons and Hoskins (1976) for the PE on a sphere; their modes agree closely with QG solutions and have, if anything, a tendency to tilt from SW to NE. In addition, large amplitude baroclinic waves on the sphere appear to differ from those in channel integrations. For example, Thorncroft and Hoskins (1990) and that on the sphere cold air (or high potential vorticity air) aloft is often advected westward relative to the upper-level trough and eventually cut off, a phenomenon that we have not observed in any of our simulations in a channel.

APPENDIX A

Semigeostrophic Solutions with Exact q_{SG}

The SG solutions shown in Figs. 2, 3, and 4 result from the GM equations transformed to geostrophic coordinates, with the additional approximation that the Jacobian term is dropped in (4). On the other hand, our asymptotic analysis of SG was based on the exact form of the SG vorticity and potential vorticity equations, (2) and (3). To distinguish between these models, we modify our terminology as in McWilliams and Gent (1980), reserving SG for the model using the exact form of (4) and denoting by H (for Hoskins) the model that omits the nonlinear terms in (4). In this appendix, we discuss SG solutions for the HW $\mu = 1$ case and provide an asymptotic analysis of the effects of approximating q_{SG} in the H model.

Solutions for the SG model may be obtained with the same numerical techniques used for the HG model. In addition, we use the initialization procedures and, resolutions are identical to those used for the H model and discussed in section 4.

The exact SG simulation is similar to the H simulation, but falls between H and HG in its agreement with the PE simulation. This might be expected, since the nondimensional form of (4) agrees more closely with the asymptotically correct HG equation (17) when the nonlinear terms are retained.

The most unstable meridional mode for the SG model is compared with the H mode in Fig. 9. The SG mode exhibits a slight tendency toward NW–SE phase tilts, but those tilts are much less pronounced than in the PE or HG modes (Fig. 8).

At larger amplitudes, the SG solution continues to deviate slightly away from the H solution and toward the PE solution. Figure 10 shows the SG solution at day 6.3. While it is still close to the H solution shown in Fig. 4, the SG solution has not maintained maximum and minimum perturbation pressures of equal magnitude, developing instead somewhat deeper lows and weaker highs like the PE wave. The SG solution has also begun a slight, but noticeable wrapping of the temperature contours cyclonically around the surface low, a behavior that is again reminiscent of the PE.

The differences between the H and SG developments may be understood by applying the small-Ro analysis of section 5 to the truncated form of (4). For the HW case, the dynamics of both the H and SG models consist of requiring that a potential vorticity analog remains constant and advecting temperature on the horizontal boundaries; thus, the models differ only in the definition of potential vorticity. In its exact form, (4) is simply the definition of q_{SG} written in geostrophic coordinates. However, when the nonlinear terms are omitted, (4) implies that the potential vorticity analog

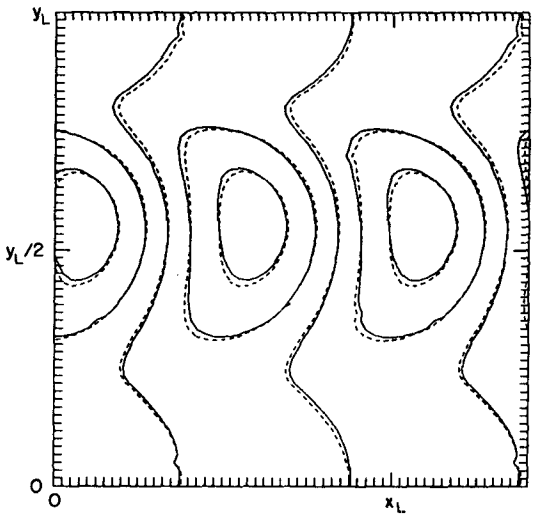

FIG. 9. As in Fig. 2, but for the SG [solid lines; uses the exact (4)] and H [dashed; uses the truncated (4)] modes at $z = 8.775$ km.

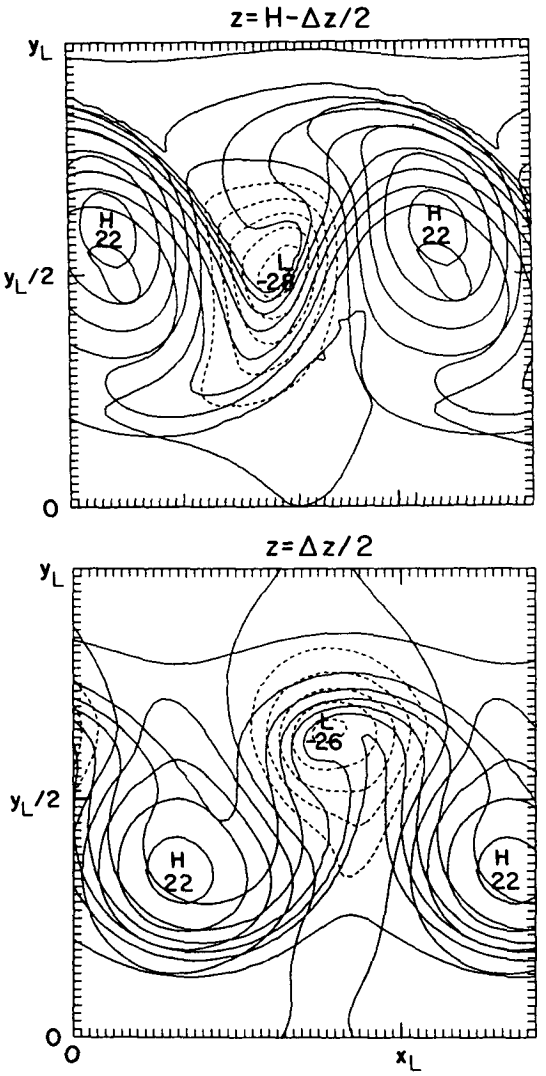

FIG. 10. As in Fig. 4, but for the SG model using the exact (4).

conserved by the H model is (with the nondimensionalizations of section 5)

$$q_{\mathrm{H}} = \frac{\mathrm{Ro}\Phi_{ZZ}}{1 - \mathrm{Ro}(\Phi_{XX} + \Phi_{YY})} = \frac{q_{\mathrm{SG}}}{1 - \mathrm{Ro}^2 J_{xy}(\phi_x, \phi_y)}, \quad \text{(A1)}$$

where the second equality was obtained by writing q_{H} in terms of q_{SG} and then, after some manipulation, returning to physical coordinates using an identity from Hoskins (1975, p. 239).

It is now easy to check the asymptotic accuracy of q_{H}. Much as in (10), where q_{SG} was written in terms of the Ertel potential vorticity, (A1) becomes for small Ro

$$q_{\mathrm{H}} = q_{\mathrm{SG}}[1 + \mathrm{Ro}^2 J_{xy}(\phi_x, \phi_y) + \mathrm{O}(\mathrm{Ro}^3)] = q - 2\,\mathrm{Ro}^2 \zeta_a \langle\theta\rangle_z + \mathrm{O}(\mathrm{Ro}^3). \quad \text{(A2)}$$

Comparison with (10) shows that the errors in H have the same leading-order form as those in SG but with a larger coefficient. Clearly, approximating (4) decreases the accuracy of the H model and these additional errors have the same nature as those already implicit in the GM approximation. This is consistent with the fact that the H solutions differ from the SG solutions in the same "direction" as the SG solutions differ from the PE.

For an alternative but essentially equivalent derivation of the asymptotic accuracy of the H model, see the discussion accompanying (7.1) and (7.2) in McWilliams and Gent (1980).

APPENDIX B

The Lagrangian Nature of GM

Through a procedure of successive approximations to the Lagrangian momentum equations, Hoskins (1975, section 3) argues that the GM approximation appears as the next correction to geostrophy. How is this to be reconciled with the Eulerian analysis (7) which shows that QG is the next correction to geostrophy?

Consider the nondimensional Lagrangian momentum equation at the same level of generality as in Hoskins [1975; see the equations preceding his (9)]:

$$\partial_t \mathbf{x} = \mathbf{v}_g + \mathrm{Ro}_L \mathbf{k} \times \partial_t^2 \mathbf{x}, \quad \mathrm{Ro}_L \equiv \frac{1}{fT}, \quad \text{(B1)}$$

where the dependent variable is particle position $\mathbf{x}(a, t)$ and T is a Lagrangian time scale. Let

$$\mathbf{x} = \mathbf{x}_0 + \mathrm{Ro}_L \mathbf{x}_1 + \cdots. \quad \text{(B2)}$$

Substituting (B2) into (B1), and collecting terms of equal powers in Ro_L gives the following ordered set of problems:

$$\mathrm{Ro}_L{}^0: \quad \partial_t \mathbf{x}_0 = \mathbf{v}_g \quad \text{(B3)}$$

$$\mathrm{Ro}_L{}^1: \quad \partial_t \mathbf{x}_1 = \mathbf{k} \times \partial_t^2 \mathbf{x}_0. \quad \text{(B4)}$$

The claim is that (B4) with (B3) gives the Lagrangian GM momentum equation, which in turn implies the Eulerian GM momentum equation given by (10) of Hoskins (1975).

If we adhere rigorously to the ordering in Ro_L of the expansion given by (B2), the latter claim is not true. The connection between Lagrangian velocity $\partial_t \mathbf{x}$ and the Eulerian velocity $\mathbf{v}$ is

$$\partial_t \mathbf{x} = \mathbf{v}[\mathbf{x}(\mathbf{a}, t), t], \quad \text{(B5)}$$

(e.g., Monin and Yaglom 1971, p. 529), and so the relation between the Lagrangian and Eulerian acceleration is

$$\partial_t{}^2\mathbf{x} = \partial_t\mathbf{v} + \partial_t\mathbf{x}\cdot\nabla\mathbf{v}. \qquad \text{(B6)}$$

The only systematic way to convert the approximate Lagrangian set (B3)–(B4) to an equivalent Eulerian set is to substitute (B2) into (B5) and (B6), and then to expand the Eulerian velocity field as

$$\mathbf{v} = \mathbf{v}_0 + \mathrm{Ro}_L\mathbf{v}_1 + \cdots. \qquad \text{(B7)}$$

Collecting terms of equal powers in Ro_L, one obtains the velocities,

$$\partial_t\mathbf{x}_0 = \mathbf{v}_0 \quad \partial_t\mathbf{x}_1 = \mathbf{v}_1, \qquad \text{(B5′)}$$

and the acceleration,

$$\partial_t{}^2\mathbf{x}_0 = \partial_t\mathbf{v}_0 + \partial_t\mathbf{x}_0\cdot\nabla\mathbf{v}_0. \qquad \text{(B6′)}$$

Thus with (B5′) and (B6′), the approximate Lagrangian equations (B3)–(B4) agree with the approximate Eulerian equation (7) (with $\epsilon_L = \mathrm{Ro}$). Thus, a Lagrangian analysis too yields QG, not GM, as the next correction to geostrophy.

REFERENCES

Allen, J. S., J. A. Barth and P. A. Newberger, 1990: On intermediate models for barotropic continental shelf and slope flow fields. Part III: Comparisons of numerical model solutions in periodic channels. *J. Phys. Oceanogr.*, **20,** 1949–1973.

Bannon, B. R., 1989: Linear baroclinic instability with the geostrophic momentum approximation. *J. Atmos. Sci.*, **46,** 402–409.

Davies, H. C., and J. C. Müller, 1988: Detailed description of deformation-induced semigeostrophic frontogenesis. *Quart. J. Roy. Meteor. Soc.*, **114,** 1201–1220.

——, C. Schär and H. Wernli, 1991: The palette of fronts and cyclones within a baroclinic wave-development. *J. Atmos. Sci.*, **48,** 1666–1689.

Duffy, D. G., 1976: The application of the semigeostrophic equations to the frontal instability problem. *J. Atmos. Sci.*, **33,** 2322–2337.

Eliassen, A., 1983: The Charney–Stern theorem on barotropic-baroclinic instability. *Pure Appl. Geophys.*, **121,** 563–572.

Fantini, M., 1990: Nongeostrophic corrections to the eigensolutions of a moist baroclinic instability problem. *J. Atmos. Sci.*, **47,** 1277–1287.

Gall, R., 1977: Some nonquasi-geostrophic effects in linear baroclinic waves. *Mon. Wea. Rev.*, **105,** 1039–1051.

Heckley, W. A., and B. J. Hoskins, 1982: Baroclinic waves and frontogenesis in a nonuniform potential vorticity semigeostrophic model. *J. Atmos. Sci.*, **39,** 1999–2016.

Holt, M. W., and G. J. Shutts, 1990: An analytical model of the growth of a frontal discontinuity. *Quart. J. Roy. Meteor. Soc.*, **116,** 269–286.

Hoskins, B. J., 1975: The geostrophic momentum approximation and the semigeostrophic equations. *J. Atmos. Sci.*, **32,** 233–242.

——, 1976: Baroclinic waves and frontogenesis. Part I: Introduction and Eady waves. *Quart. J. Roy. Meteor. Soc.*, **102,** 103–122.

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

——, and N. V. West, 1979: Baroclinic waves and frontogenesis. Part II: Uniform potential vorticity jet flows—cold and warm fronts. *J. Atmos. Sci.*, **36,** 1663–1680.

Joly, A., and A. J. Thorpe, 1990: Frontal instability generated by tropospheric potential vorticity anomalies. *Quart. J. Roy. Meteor. Soc.*, **116,** 525–560.

Keyser, D., B. D. Schmidt and D. G. Duffy, 1989: A technique for representing three-dimensional vertical circulations in baroclinic disturbances. *Mon. Wea. Rev.*, **117,** 2463–2494.

Klemp, J. B., and R. B. Wilhelmson, 1978: The simulation of three-dimensional convective storm dynamics. *J. Atmos. Sci.*, **35,** 1070–1096.

Koclas, P., A Staniforth and H. Warn, 1986: A variable-resolution finite-element model of frontogenesis. *Mon. Wea. Rev.*, **114,** 1340–1353.

Magnusdottir, G., and W. H. Schubert, 1991: Semigeostrophic theory on the hemisphere. *J. Atmos. Sci.*, **48,** 1449–1456.

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

Monin, A. S., and A. M. Yaglom, 1971: *Statistical Fluid Mechanics: Mechanics of Turbulence.* Vol. 1, The MIT Press, 769 pp.

Moore, G. W. K., and W. R. Peltier, 1988: Frontal cyclogenesis and the geostrophic momentum approximation. *Geophys. Fluid Dyn.*, **45,** 183–197.

Mudrick, S. E., 1974: A numerical study of frontogenesis. *J. Atmos. Sci.*, **31,** 869–892.

Pedlosky, J., 1979: *Geophysical Fluid Dynamics.* Springer Verlag, 624 pp.

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.

Robert, A. J., 1966: The integration of a low order spectral form of the primitive meteorological equations. *J. Meteor. Soc. Japan,* **44,** 237–245.

Simmons, A. J., and B. J. Hoskins, 1976: Baroclinic instability on the sphere: Normal modes of the primitive and quasi-geostrophic equations. *J. Atmos. Sci.*, **33,** 1454–1477.

Shapiro, M. A., and D. Keyser, 1990: Fronts, jet streams, and the tropopause. Chapter 10, *Extratropical Cyclones, The Erik Palmén Memorial Volume,* C. W. Newton and E. O. Holopainen, Eds., Amer. Meteor. Soc., 167–189.

Shutts, G. J., 1989: Planetary semigeostrophic equations derived from Hamilton's principle. *J. Fluid Mech.*, **208,** 545–573.

Smolarkiewicz, P. K., 1984: A fully multidimensional positive definite advective transport algorithm with small implicit diffusion. *J. Comput. Phys.*, **54,** 325–362.

Takayabu, I., 1986: Roles of the horizontal advection on the occlusion of a cyclone developing in the baroclinic westerly jet. *J. Meteor. Soc. Japan,* **64,** 329–345.

Thorncroft, C. D., and B. J. Hoskins, 1990: Frontal cyclogenesis. *J. Atmos. Sci.*, **47,** 2317–2336.

Volkert, H., and C. H. Bishop, 1990: The semi-geostrophic Eady problem as a testbed for numerical simulations of frontogenesis. *Tellus,* **42A,** 202–207.

Xu, Q., 1990: Cold and warm frontal circulations in an idealized moist semigeostrophic baroclinic wave. *J. Atmos. Sci.*, **47,** 2337–2352.