Baroclinic Instability and Loss of Balance

M. JEROEN MOLEMAKER AND JAMES C. MCWILLIAMS

Institute of Geophysics and Planetary Physics, University of California, Los Angeles, Los Angeles, California

IRAD YAVNEH

Department of Computer Science, Technion, Haifa, Israel

(Manuscript received 24 July 2003, in final form 4 March 2005)

ABSTRACT

Under the influences of stable density stratification and the earth's rotation, large-scale flows in the ocean and atmosphere have a mainly balanced dynamics—sometimes called the slow manifold—in the sense that there are diagnostic hydrostatic and gradient-wind momentum balances that constrain the fluid acceleration. The nonlinear balance equations are a widely successful, approximate model for this regime, and mathematically explicit limits of their time integrability have been identified. It is hypothesized that these limits are indicative, at least approximately, of the transition from the larger-scale regime of inverse energy cascades by anisotropic flows to the smaller-scale regime of forward energy cascade to dissipation by more nearly isotropic flows and intermittently breaking inertia–gravity waves. This paper analyzes the particular example of an unbalanced instability of a balanced, horizontally uniform, vertically sheared current, as it occurs within the Boussinesq equations. This ageostrophic, anticyclonic, baroclinic instability is investigated with an emphasis on how it relates to the breakdown of balance in the neighborhood of loss of balanced integrability and on how its properties compare with other examples of ageostrophic anticyclonic instability of rotating, stratified, horizontally sheared currents. It is also compared with the more familiar types of instability for a vertically sheared current: balanced (geostrophic) baroclinic instability, centrifugal instability, and Kelvin–Helmholtz instability.

1. Introduction

The problem of baroclinic instability for a vertically sheared, rotating, stratified flow is a classical one in geophysical fluid dynamics. Following early papers by Charney (1947), Eady (1949), and Phillips (1954)—all of which differed in their representation of the vertical structure and boundary conditions for the mean stratification and currents—baroclinic instability has been the topic of an enormous number of papers, including some, following Stone (1966), that examine the regime of strong flow with finite values for the Rossby and Richardson numbers. Nevertheless, this problem is one that can be looked at freshly from the perspective of the limits of balanced dynamics (McWilliams et al. 1998) in order to investigate the conundrum of how energy dissipation occurs for large-scale winds and currents.

Corresponding author address: Dr. Jim McWilliams, IGPP/UCLA, Los Angeles, CA 90095-1567.
E-mail: jcm@atmos.ucla.edu

The general circulation of the ocean is forced by surface fluxes of heat, water, and momentum primarily at large space and long time scales. The circulation has comparably large and long scales, as well as important smaller mesoscale flows in equatorial and lateral boundary currents, and its principal instability modes also occur on the mesoscale. All of these circulation elements approximately satisfy geostrophic, hydrostatic, and incompressible dynamical momentum and mass balances, at least in the interior region. How does the energy dissipation occur for the general circulation in an equilibrium balance with the energy generation by surface fluxes? The general circulation is well known to be unstable to barotropic and baroclinic instabilities, both of which are instability types associated with an inflection point in the velocity profile (cf. Drazin and Howard 1966). Furthermore, both instability types can occur within the dynamical balance constraints listed above. Yet the nonlinear dynamics of balanced flows, often called geostrophic turbulence (Charney 1971; McWilliams et al. 1994), is generally understood to pro-

duce an inverse energy cascade toward larger scales both horizontally and vertically, hence away from the small scales where dissipation by molecular viscosity can occur. Thus inflection-point instabilities do not obviously provide an effective route to dissipation for the general circulation.

Some of the requisite dissipation undoubtedly occurs within the turbulent boundary layer near the surface and bottom. Some dissipation also occurs through creation of internal gravity waves by flow over topography, with subsequent wave propagation into the interior and a wave-dynamical cascade (sometimes involving breaking) down to dissipation at small scales. Each of these routes to dissipation involves an extraction of energy from the circulation near the vertical boundaries, although the bulk of the energy resides in the vertical interior, and these could only provide sufficient dissipation if there were a large-scale mechanism for efficient energy flux toward the boundaries. A possible alternative, more local route to dissipation is directly through the interior, turbulent cascade dynamics of the circulation. In oceanic general circulation models, the local route to dissipation is implied by the ad hoc use of eddy viscosities to parameterize this cascade. Our conceptual view of the interior mechanism is the following: large- and mesoscale circulations typically satisfy a balanced dynamics (as defined below), which have little interaction with the inertia–gravity wave field; balanced turbulent cascades are very inefficient in energy dissipation; there are explicitly specifiable limits to the regime of balanced dynamics that are violated sometimes for the circulation; violations of these limits lead to energy transfer to unbalanced motions; unbalanced turbulent cascades are much more efficient in their dissipation. In this view, the important bottleneck in the local route to dissipation is loss of balance and its evolutionary consequences. The purpose of this article is to examine an initiation mechanism for the local route to dissipation, without here trying to assess the relative contributions among the various possible routes.

We previously investigated the relevance of the conditions for loss of balance to some other types of flow instabilities that may also be relevant to the initiation of a local energy cascade by unbalanced motions. In McWilliams et al. (1998) we show that there is an exact correspondence of these conditions with the onset of gravitational and centrifugal instabilities (associated with change of sign of vertical density gradient and potential vorticity, respectively) for parallel and axisymmetric flows. In McWilliams and Yavneh (1998), we show that there is an apparent correspondence between one of these conditions (associated with the difference between absolute vorticity and strain rate, $A - S$; see below) and the onset of elliptical instability for a barotropic flow in an infinite domain. In Molemaker et al. (2000) and Yavneh et al. (2001), we show for Taylor–Couette flow that there is a close correspondence of the potential vorticity condition with the onset of centrifugal instability and there is a looser correspondence for another type of instability with the $A - S$ condition. The looseness of the correspondence in this latter case is in the sense that the growth rate of the unstable mode is exponentially weakening, with exponent $\propto 1/\mathrm{Ro}$ (where $\mathrm{Ro} > 0$ is the Rossby number), as this condition is met on the path toward the quasigeostrophic limit, $\mathrm{Ro} \to 0$. In McWilliams et al. (2001), we summarize the results for these problems plus that for another flow, a barotropic boundary current. Our conclusion is that they all have in common an ageostrophic instability for anticyclonic flows, in at least a loose correspondence with the $A - S$ condition, and this type of instability is distinct from the more familiar gravitational, centrifugal, and inflection-point types.

In this paper we continue this line of investigation by calculating solutions for an anticyclonic, ageostrophic type of baroclinic instability as another avenue for initiating an energy transfer from balanced to unbalanced motions. In support of this latter distinction, we devise a new method for analyzing the degree of balance of the unstable modes.

2. Formulation and methods

We pose the problem as the linear instability of a mean horizontal current $V(z)$ in geostrophic balance for a rotating, stably stratified fluid with Coriolis frequency $f(y) > 0$ and Brunt–Väisälä frequency $N(z) > 0$. To focus on our primary issue, we assume here f, N, and dV/dz are spatially uniform, although effects due to their variations are not unimportant for realistic, large-scale currents. Similarly we ignore horizontal variations in V, except in some final remarks. Since our focus is on possible departures from balanced dynamics, we will use the Boussinesq equations as the fundamental fluid dynamics of an incompressible fluid. The problem considered is therefore a nongeostrophic, nonhydrostatic generalization of Eady (1949) and a nonhydrostatic generalization of Stone (1966, 1970), whose formulation we will initially follow. Thus, we will be able to assess the degree and importance of loss of balance, with respect to both the geostrophic and hydrostatic relations.

The dimensional velocity and buoyancy fields in the basic state are defined by

$$\tilde{V}(\tilde{z}) = V_0\left(\frac{\tilde{z}}{H} - 0.5\right) \quad \text{and} \tag{1a}$$

$$\tilde{B}(\tilde{x}, \tilde{z}) = N^2\tilde{z} + \frac{fV_0}{H}\tilde{x}. \tag{1b}$$

Tildes are used here to denote dimensional variables. Here, $\tilde{B}$ is the buoyancy (i.e., $-g/\rho_o$ times the departure in density from its background value ρ_o, with g the gravitational acceleration) in thermal-wind balance with the horizontally homogeneous vertically sheared current $\tilde{V}(\tilde{z})$. The domain is unbounded horizontally and has planar vertical boundaries at $\tilde{z} = 0$ and H. In the inviscid Boussinesq equations, the Ertel potential vorticity Π is a Lagrangian, invariant, and as we shall see, its distribution is often relevant to possible flow instabilities. For the basic state, its value in the interior is a constant,

$$\tilde{\Pi} = fN^2 - \frac{\partial\tilde{V}}{\partial\tilde{z}}\frac{\partial\tilde{B}}{\partial\tilde{x}} = fN^2\left[1 - \left(\frac{\tilde{V}}{NH}\right)^2\right] \le fN^2. \tag{2}$$

The final inequality shows that this baroclinic shear flow is an anticyclonic one, independent of the sign of $\tilde{V}$ and f.

If (u, v, w, p, b) are perturbation fields, their governing nondimensional equations, linearized about the basic state (1), are the following:

$$\mathrm{Ro}\left(\frac{\partial u}{\partial t} + V\frac{\partial u}{\partial y}\right) - v = -\frac{\partial p}{\partial x}, \tag{3a}$$

$$\mathrm{Ro}\left(\frac{\partial v}{\partial t} + V\frac{\partial v}{\partial y} + \epsilon\, w\frac{\partial V}{\partial z}\right) + u = -\frac{\partial p}{\partial y}, \tag{3b}$$

$$F^2\lambda^2\left(\frac{\partial w}{\partial t} + V\frac{\partial w}{\partial y}\right) = -\frac{\partial p}{\partial z} + b, \tag{3c}$$

$$\frac{\partial u}{\partial x} + \frac{\partial v}{\partial y} + \epsilon\frac{\partial w}{\partial z} = 0, \quad \text{and} \tag{3d}$$

$$\frac{\partial b}{\partial t} + V\frac{\partial b}{\partial y} + u\frac{\partial B}{\partial x} + w = 0, \tag{3e}$$

together with the boundary conditions, $w = 0$ at $z = 0$ and 1. Here u, v, and w are cross-stream, streamwise, and vertical perturbation velocity, and p is perturbation pressure. To obtain the set of nondimensional equations in (3), the following scales have been used for nondimensionalization of the basic state and perturbation quantities: horizontal length L, vertical length H (the domain height), horizontal velocity V_0, time L/V_0, dynamic pressure $P = \rho_o f V_0 L$, background buoyancy gradient N^2, perturbation buoyancy fV_0L/H, and vertical velocity $w = fV_0^2/N^2H$. As a result several nondimensional parameters appear in (3): Rossby number $\mathrm{Ro} = V_0/fL$, Froude number $F = V_0/NH$, aspect ratio $\lambda = H/L$, and $\epsilon = fLV_0/N^2H^2 = F^2/\mathrm{Ro}$. This nondimensionalization, which follows that in McWilliams (1985), is designed to expose the generally weak deviations of the flow from balance when these parameters are not large. The nondimensional basic-state variables are

$$V = z - 0.5 \quad \text{and} \quad B = x. \tag{4}$$

We will solve for normal-mode perturbations in the form,

$$(u, v, w, p, b) = [\hat{u}(z), \hat{v}(z), \hat{w}(z), \hat{p}(z), \hat{b}(z)] \times \exp(\sigma t)\exp[i(kx + ly)], \tag{5}$$

where $(\hat{u}, \hat{v}, \hat{w}, \hat{p}, \hat{b})$ are complex functions of z only, and k and l are cross-stream and streamwise wavenumbers. To compare our results with previous work we will need a translation into previously used nondimensional parameters. Since there is no horizontal length scale associated with the basic state (1), we are free to choose it such that $\lambda = H/L = f/N$ (i.e., L is chosen as a gravest baroclinic deformation radius), understanding that the actual length scale of unstable perturbation modes will be internally determined by the solution. This choice implies that

$$\mathrm{Ro}^2 = \frac{V_0^2}{F^2L^2} = \frac{V_0^2}{N^2H^2} = F^2 = \mathrm{Ri}^{-1}, \tag{6}$$

where $\mathrm{Ri} = N^2/(d\tilde{V}/d\tilde{z})^2 = (NH/V_0)^2$ is the Richardson number. Hence $F = \mathrm{Ro}$, and $\epsilon = F^2/\mathrm{Ro} = \mathrm{Ro}$. The only two remaining independent nondimensional parameters are Ro (or Ri) and the ratio between background stratification and rotation $\lambda = f/N$.

The following set of nondimensional equations is thus obtained for the normal modes:

$$\mathrm{Ro}(\sigma\hat{u} + ilV\hat{u}) - \hat{v} = -ik\hat{p}, \tag{7a}$$

$$\mathrm{Ro}(\sigma\hat{v} + ilV\hat{v} + \mathrm{Ro}\hat{w}) + \hat{u} = -il\hat{p}, \tag{7b}$$

$$\mathrm{Ro}^2\lambda^2(\sigma\hat{w} + ilV\hat{w}) = -\frac{\partial\hat{p}}{\partial z} + b, \tag{7c}$$

$$ik\hat{u} + il\hat{v} + \mathrm{Ro}\frac{\partial\hat{w}}{\partial z} = 0, \quad \text{and} \tag{7d}$$

$$\sigma\hat{b} + ilV\hat{b} + \hat{u} + \hat{w} = 0, \tag{7e}$$

plus the vertical boundary conditions $\hat{w} = 0$ at $z = 0$ and 1. In the limit of strong stratification (the usual situation in the oceanic interior), λ^2 is small, and the vertical momentum equation reduces to hydrostatic

balance. (Effects of weaker stratification will be investigated below.)

The system (7) is discretized in the vertical using standard finite differences on a staggered grid (appendix A). This leads to a generalized eigenvalue problem that may be quite large depending on the number of points used in the grid. Previously, authors have reported difficulties in obtaining positive, but small eigenvalues (i.e., weak instabilities) for this problem (cf. Stone 1970). This is due to the nearly singular behavior of the coefficients of the equations when the growth rate (i.e., the real part of σ) becomes very small. Since the resulting matrix is sparse and we are only interested in a few, largest eigenvalues, we use a Krylov subspace method, Jacobi Davidson Q–Z (Sleijpen and van der Vorst 1996) to solve for the most unstable eigenmodes. Although a direct method (e.g., Q–Z; Golub and Van Loan 1996) solves for all eigenvalues and eigenmodes, the computation time for this algorithm increases as the cube of the order of the matrix and therefore becomes unusable for large grids. Krylov methods do not scale with a fixed rate of the matrix order, but instead depend in a nonanalytical way on the size, condition number, and eigenvalue spectrum of the matrix. For many problems, a Krylov method is usable for matrix sizes where direct methods have long become impractical.

The largest resolution we have used in the vertical is 9192 grid points. The solution of a single eigenvalue problem of this size for the four largest eigenvalues takes approximately 5 cpu seconds on a Cray J90.

3. Unstable modes

Following Stone (1970), Fig. 1 shows the perturbation growth rate as a function of the streamwise wavenumber l for several values of the cross-stream wavenumber k. There are two distinct types of unstable mode for each k, which we will show in section 4 are appropriately called geostrophic and ageostrophic types. Note that, for these values of Ro and λ, the geostrophic instability has a somewhat larger growth rate and occurs on a larger horizontal scale, relative to the ageostrophic instability. The geostrophic instability mode is the most familiar one from previous analyses: it remains strong as Ro $\to 0$, and it is associated with an inflection point (i.e., a change of sign of the horizontal gradient of the basic state potential vorticity, which involves the vertical boundary conditions in this uniform shear flow; see below).

There are modest differences between Fig. 1 and Fig. 2 in Stone (1970) as a result of our inclusion of nonhydrostatic effects for finite λ. The nonhydrostatic effects on the geostrophic modes are small; for example, for $\lambda = 1$ the maximum growth rate is reduced by only 5%. For the ageostrophic modes, however, the effect is

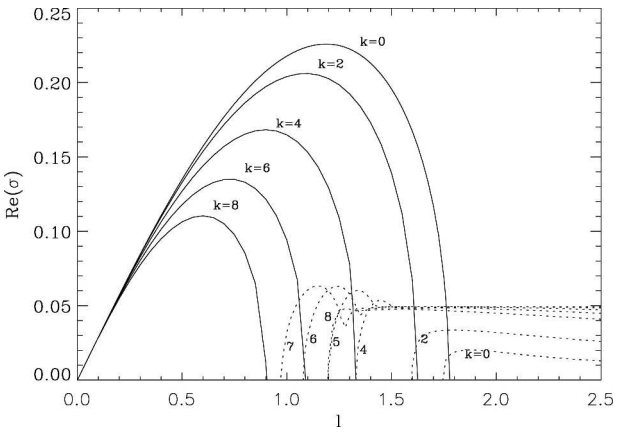

FIG. 1. Growth rate, Re(σ), for Ro = $\mathrm{Ri}^{-1/2} = 1$ and $\lambda = 1$ as a function of streamwise wavenumber l for several values of cross-stream wavenumber k. Solid and dashed lines indicate, respectively, the geostrophic and ageostrophic types of instability.

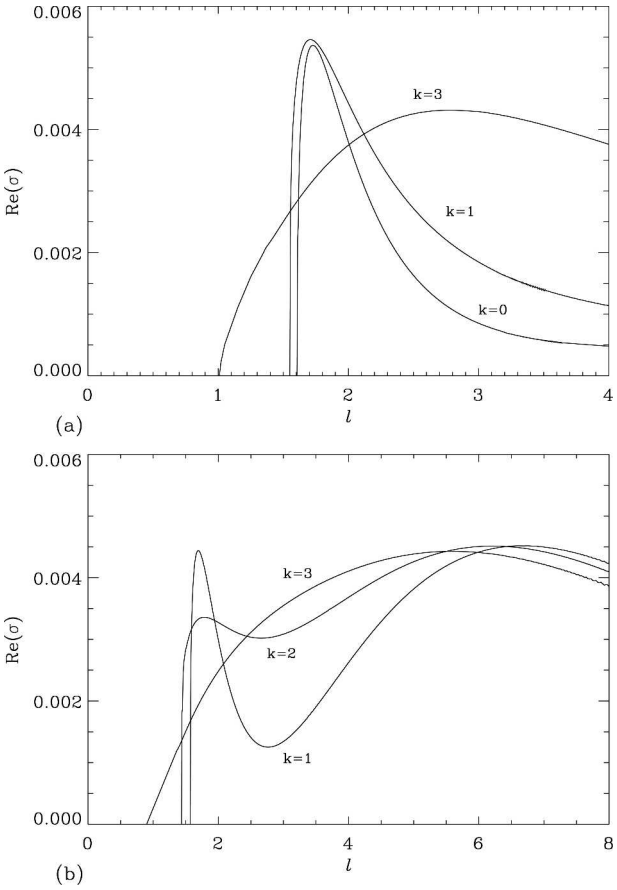

FIG. 2. Growth rate, Re(σ), of the ageostrophic instability mode for Ri = 2 (Ro = $1/\sqrt{2}$) as a function of streamwise wavenumber l for several values of cross-stream wavenumber k: (a) $\lambda = 1.0$ and (b) $\lambda = 0.0$ (the hydrostatic limit).

somewhat more significant; for $\lambda = 1$ the ageostrophic mode that becomes first unstable for increasing l is now $k = 7$; in contrast, in the hydrostatic limit, the first mode to become unstable for increasing l is the one with the largest k (which implies an inconsistency with the hydrostatic approximation).

For values of Ro < 1 (Ri > 1), the ageostrophic instability has smaller growth rates. The numerical method used in Stone (1970) was limited in its ability to obtain positive eigenvalues, because of the aforementioned nearly singular behavior of the coefficients. By using much higher grid resolutions, we are able to track the ageostrophic instability behavior, even for extremely small growth rates. In Fig. 2, the growth rates for ageostrophic instability are shown again as a function the horizontal wavenumbers, but now for Ro = 0.71 (Ri = 2) and for two different values of λ. For $\lambda = 1$ it is noteworthy that, contrary to previous characterizations of this type of instability as small scale, the wavenumbers for which maximum growth is obtained are in fact comparable to the gravest deformation radius. However, keep in mind that $\lambda = 1$ implies that $L = H$ and $N = f$, which for geophysical regimes corresponds to rather weak stratification and small horizontal scale. In contrast, in the hydrostatic limit $\lambda = 0$ (Fig. 2b), a shift to much larger wavenumbers occurs for maximum growth (n.b., the change in abscissa scale), implying perturbation scales are small relative to the deformation radius. Thus, in this regard the ageostrophic instability behaves differently for small and large aspect ratios and/or stratification/rotation ratios.

Since the breakdown of balanced dynamics is at least sometimes likely to be associated with inertia–gravity wave generation, it is informative to study the behavior of these instabilities with varying stable stratification N. In Fig. 3 the geostrophic and ageostrophic modal growth rates are shown as a function of stratification, optimized in each case over the horizontal wavenumbers. Figures 3a and 3b show a different parametric route toward $N/f = 0$. In Fig. 3a, the Rossby number is constant at Ro = 1. This implies that for decreasing values of N and for constant f the unstable length scale L decreases since we have chosen $L/H = N/f$. This, in turn, implies that V_0 decreases, since the Rossby number is constant. The time scale L/V_0 is therefore constant for this parametric route. In this particular limit, the growth rate decreases as $N^{1/2}$ in the limit of small N for the geostrophic mode [as it does in the Eady (1949) analysis for Ro = 0]. The growth rate of the ageostrophic mode decreases as N in the small N limit implying that there is a (inertia) gravity wave component in the ageostrophic mode. This confirms the hypothesis of, among others, Nakamura (1988) that the ageostrophic instability is produced by a resonance between a boundary mode and a sheared inertia–gravity mode. For large N, the growth rates of both geostrophic and ageostrophic modes attain constant values.

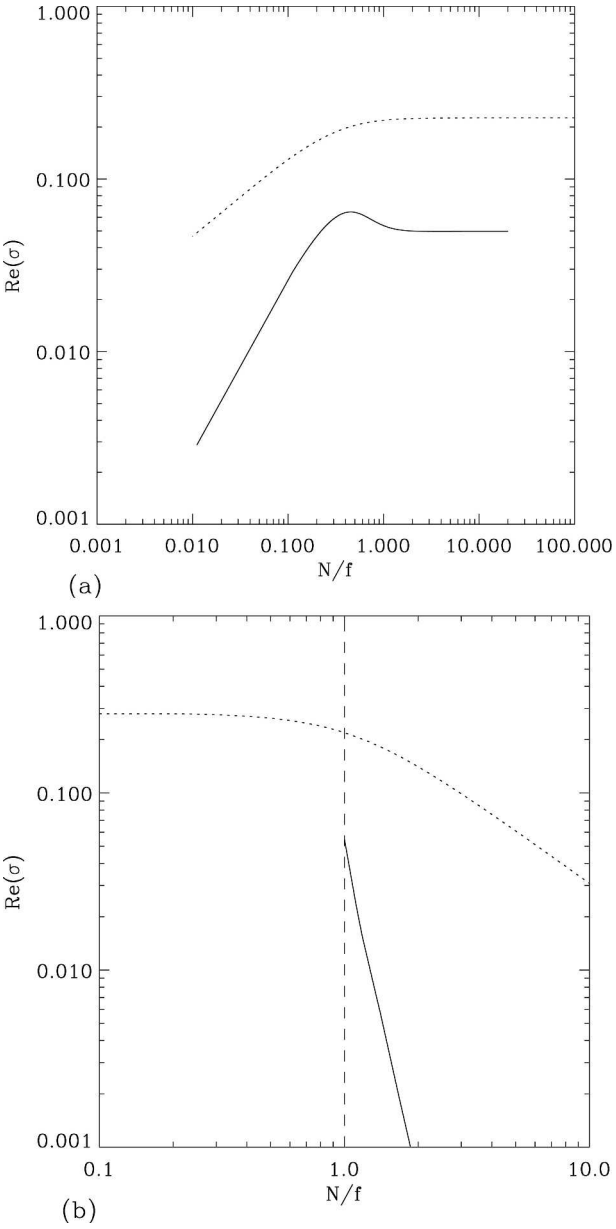

FIG. 3. Growth rate of geostrophic (dots) and ageostrophic (solid) modes as a function of vertical stratification $N/f \equiv \lambda$ (optimized over horizontal wavenumbers): (a) for fixed Ro = 1 and (b) for fixed V_0 and f [this corresponds to varying the Rossby number as Ro = $(N/f)^{-1}$]. In (b) the dashed vertical line indicates Ro = 1, and centrifugal instabilities may occur for $N/f < 1$ (Ro $>$ 1) in (b) to its left. Note that the growth rates coincide at $N/f = 1$ in both panels.

In Fig. 3b, the stratification N/f is varied while keeping V_0 and f constant. Because of our choice of length scale L, this implies that Ro varies as N^{-1}. This choice also leads to a different dimensional time scale. For ease of comparison with Fig. 3a, the growth rates in Fig. 3b have been rescaled using a constant time scale $1/f$. The dependence of Ro on N for this case does mean that for smaller values of N, the flow becomes increasingly unstable, and at $N/f = 1$ (Ro = 1) centrifugal modes become unstable. For $N/f < 1$ (and in this case, Ro > 1), the ageostrophic mode becomes harder to identify separately since it joins with the centrifugal modes (Stone 1970), and we have not attempted to do so. The geostrophic mode, on the other hand, becomes independent of vertical stratification for small values of N. In the limit of $N \gg 1$, the ageostrophic growth rate decays quite rapidly with N, while the geostrophic growth rate varies as $N^{-1/2}$ [again as in the Eady (1949) quasigeostrophic analysis].

4. Loss of balance

The three conditions for loss of integrability of the balance equations in isentropic coordinates (Gent and McWilliams 1984) are that sign changes occur for any of the buoyancy stratification (N^2); the dimensional absolute vorticity,

$$A = f + \zeta^{(z)} \equiv f + v_X - u_Y$$

(where the horizontal derivatives denoted by capital letters are in isentropic coordinates), and the difference between A and the dimensional horizontal strain rate,

$$S \equiv \sqrt{(u_X - v_Y)^2 + (v_X + u_Y)^2}$$

(McWilliams et al. 1998). For the basic state (1), we can write the dimensional isentropic horizontal derivatives as

$$\partial_X = \partial_x - \frac{fV_0^2}{N^2H^2}\,\partial_z \quad \text{and} \quad \partial_Y = \partial_y;$$

therefore, the latter two conditions for loss of balance may be expressed as sign changes in

$$A = f\left(1 - \frac{1}{\mathrm{Ri}}\right) \quad \text{or} \quad A - S = f\left(1 - \frac{2}{\mathrm{Ri}}\right). \tag{8}$$

Again we see that this basic-state flow is anticyclonic since $A/f < 1$ for all Ri. Both A and $A - S$ monotonically decrease with decreasing Ri.

Thus, we can make the following regime categorization in terms of $\mathrm{Ri} = 1/\sqrt{\mathrm{Ro}}$ and in relation to these conditions for loss of balance and known instability types.

- The quasigeostrophic limit occurs for Ro $\to 0$, Ri $\to \infty$. It has a geostrophic baroclinic instability (Charney 1947; Eady 1949), whose inflection point in this particular flow configuration occurs at the vertical boundaries.
- The $A - S$ condition is satisfied if $\mathrm{Ro} > 1/\sqrt{2}$ ($\mathrm{Ri} < 2$).
- The A condition is satisfied if Ro > 1 (Ri < 1). It has a symmetric centrifugal instability (Stone 1966; Hoskins 1974).
- The classical (nonrotating) condition for the onset of Kelvin–Helmholtz instability (Miles 1961; Howard 1961) is satisfied if Ro > 2 (Ri $< \frac{1}{4}$).
- The classical inviscid condition for the onset of gravitational instability is satisfied if $N^2 < 0$ or (Ri < 0).

Our attention is therefore drawn to the intermediate range of $1 > \mathrm{Ro} > 0$ ($1 < \mathrm{Ri} < \infty$) where previous results give less complete guidance. Note that this regime is the neighborhood of the $A - S = 0$ (Ri = 2) condition.

Growth rates for both geostrophic and ageostrophic modes are shown in Fig. 4a. In the quasigeostrophic limit (Ro = 0), only the geostrophic mode has a positive growth rate. Some nongeostrophic influences are present in the geostrophic mode as Ro increases, but they are small; for example, the growth rate is reduced from $\sigma = 0.30$ at Ro = 0 to $\sigma = 0.22$ at Ro = 1. On the other hand, the ageostrophic mode attains a significant growth rate near Ro = 1, but it decreases very rapidly for Ro $\to 0$. In Fig. 4b the growth rate of the ageostrophic mode is shown as a function of the $A - S$ indicator for loss of balance. Although there is a very rapid decrease in σ in the vicinity of $A - S = 0$ (Ro = $1/\sqrt{2}$), this point is not a sharp boundary between stable and unstable behavior. For Ro > 1 (Ri < 1), centrifugal instability does occur for both symmetric (zero streamwise wavenumber) and nonsymmetric (nonzero streamwise wavenumber) perturbations. These results are fully consistent with the analyses of this problem in Stone (1966, 1970) and Nakamura (1988).

5. Diagnosis of balance

To assess quantitatively the departure from the balanced manifold in both the geostrophic and ageostrophic baroclinic instabilities, we must specify the balance equations for the present situation of linearization around the basic state (1). To derive the balance equations, the horizontal flow $\mathbf{u}_h = (u, v)$ is decomposed into horizontally nondivergent and vertically irrotational components. This is accomplished by defining a streamfunction Ψ and a velocity potential χ:

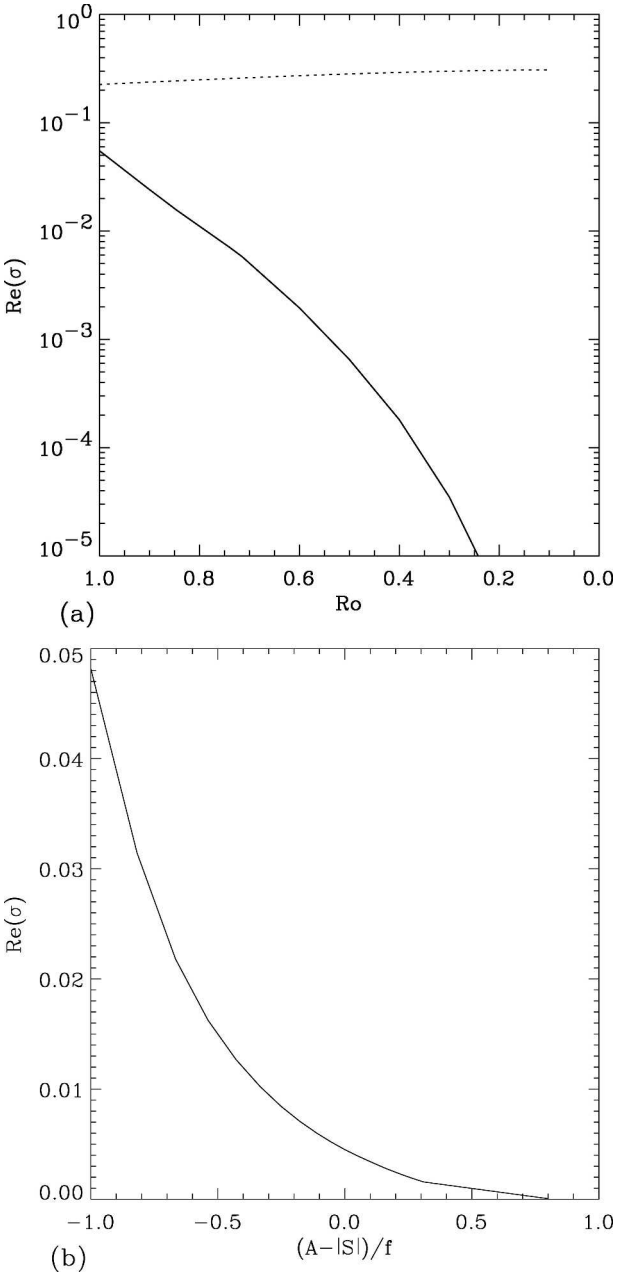

FIG. 4. Growth rates maximized over both horizontal wavenumbers for the hydrostatic regime, $\lambda = 0.01$: (a) geostrophic (dots) and ageostrophic (solid) modes as a function of Ro; (b) only the ageostrophic mode as a function of the third condition for breakdown of the balance equations, $(A - S)/f$.

$$u = -\Psi_y + \mathrm{Ro}\chi_x \quad \text{and} \quad v = \Psi_x + \mathrm{Ro}\chi_y. \tag{9}$$

The vertical component of the vorticity is $\zeta = v_x - u_y = \nabla_h^2\Psi$, and the divergence of the horizontal flow is $\delta = u_x + v_y = \mathrm{Ro}\nabla_h^2\chi$. Taking the curl of the horizontal momentum balance in (3), we obtain the vorticity equation

$$\mathrm{Ro}\left(\frac{\partial\zeta}{\partial t} + V\frac{\partial\zeta}{\partial y} + \mathrm{Ro}\frac{\partial w}{\partial x}V_z\right) + \mathrm{Ro}\nabla_h^2\chi = 0. \tag{10}$$

Alternatively, taking the divergence of the horizontal momentum balance in (3) leads to

$$\mathrm{Ro}\left(\mathrm{Ro}\frac{\partial\nabla_h^2\chi}{\partial t} + \mathrm{Ro}V\frac{\partial\nabla_h^2\chi}{\partial y} + \mathrm{Ro}\frac{\partial w}{\partial y}V_z\right) - \zeta = -\nabla^2 p. \tag{11}$$

Making the appropriate approximation (i.e., dropping all terms in the previous two equations at higher than the leading two orders in Ro) leads us to the balance equations in height coordinates (cf. Lorenz 1960; McWilliams 1985):

$$\frac{\partial\zeta}{\partial t} + V\frac{\partial\zeta}{\partial y} + \mathrm{Ro}\frac{\partial w}{\partial x}V_z + \nabla_h^2\chi = 0, \tag{12a}$$

$$\zeta = \nabla^2 p, \tag{12b}$$

$$\frac{\partial p}{\partial z} = b, \tag{12c}$$

$$\nabla_h^2\chi + \frac{\partial w}{\partial z} = 0, \quad \text{and} \tag{12d}$$

$$\frac{\partial b}{\partial t} + V\frac{\partial b}{\partial y} + u + w = 0. \tag{12e}$$

For normal-mode solutions (with $\zeta = -K^2\Psi$ and $K^2 = k^2 + l^2$), these simplify to

$$-(\sigma + ilV)K^2\hat{\Psi} + ikV_z\,\mathrm{Ro}\hat{w} - K^2\hat{\chi} = 0, \tag{13a}$$

$$\hat{\Psi} = \hat{p}, \tag{13b}$$

$$\partial_z\hat{p} = \hat{b}, \tag{13c}$$

$$K^2\hat{\chi} = \partial_z\hat{w}, \quad \text{and} \tag{13d}$$

$$(\sigma + ilV)\hat{b} - il\hat{\Psi} + ik\mathrm{Ro}\hat{\chi} + \hat{w} = 0. \tag{13e}$$

To cancel the explicit time dependence in the vorticity and buoyancy relations, we combine (13a) and (13e) to form the so-called ω equation. After also using (13b)–(13d) for substitutions, the ω equation is

$$-il(1 + V_z)\hat{\Psi} + ik\mathrm{Ro}(1 + V_z)\hat{\chi} - \partial_z\hat{\chi} + \hat{w} = 0. \tag{14}$$

Equation (14) plus (13b)–(13d) form a set of constraints that have to be satisfied by the balanced part of the modes, from now on indicated by a subscript "bal."

We can now project the solution obtained with the full Boussinesq equations onto this manifold, given the

definition of an inner product (see appendix C). This projection operation reduces the dimensionality of the space of Boussinesq solutions by splitting it into two orthogonal subspaces: "bal" and "un," where the latter subscript denotes the unbalanced component, equal by definition to the total field minus the balanced part. The definition of the inner product is dependent on the particular measure of imbalance that we choose to minimize in fitting the balanced component. The projection is accomplished by fitting a balanced solution component $\mathbf{X}_b$ to the full solution $\mathbf{X}$ by minimizing a cost function schematically of the form

$$\mathcal{F} = \int_0^1 F(\mathbf{X} - \mathbf{X}_b)\, dz = \begin{pmatrix} \hat{\Psi} - \hat{\Psi}_{\text{bal}} \\ \hat{\chi} - \hat{\chi}_{\text{bal}} \\ \hat{w} - \hat{w}_{\text{bal}} \\ \hat{p} - \hat{p}_{\text{bal}} \\ \hat{b} - \hat{b}_{\text{bal}} \end{pmatrix} \mathbf{D} \begin{pmatrix} \hat{\Psi} - \hat{\Psi}_{\text{bal}} \\ \hat{\chi} - \hat{\chi}_{\text{bal}} \\ \hat{w} - \hat{w}_{\text{bal}} \\ \hat{p} - \hat{p}_{\text{bal}} \\ \hat{b} - \hat{b}_{\text{bal}} \end{pmatrix}^*, \tag{15}$$

where $\mathbf{D}$ is a Hermitian, positive, semidefinite matrix, and * denotes a complex conjugate. The unbalanced solution component is the residual; for example,

$$\mathbf{X}_{\text{un}} = \mathbf{X} - \mathbf{X}_{\text{bal}}.$$

By minimizing $\mathcal{F}$, we determine the largest part of a particular mode of the Boussinesq equations that satisfies the (subset) of the balance equations defined by (13b)–(13d) and (14). The coefficients of the matrix $\mathbf{D}$ depend on the choice of the quantity that we wish to measure the closeness of the fit by. Below we present results for alternative choices of the fitting norm (i.e., several $\mathbf{D}$ choices; see appendix C) and examine the sensitivity of the fitted amplitude to these choices. To present the results of the partitioning in balanced and unbalanced parts we choose a norm, defined by the same inner product that defines the cost function $\mathcal{F}$. In general, a different measure for norm and cost function could be chosen, but at least initially it seems natural to use the same measure for both.

Potential vorticity (see appendix B) often has been used as the representative field for the balanced component of both natural and simulated flows (Hoskins et al. 1985). In Olsson and Cotton (1997), for example, the potential vorticity is fit exactly at each point in space, but only a less accurate balanced approximation is used in "inverting" the potential vorticity for the various balanced flow fields (specifically, the terms representing the vertical shear of horizontal divergent winds and the horizontal shear of the vertical velocity are neglected).

However, for this particular problem, potential vorticity is not a natural choice for the cost function. The basic state has a constant potential vorticity, and, since potential vorticity is conserved by fluid elements in this inviscid, adiabatic flow, all perturbations are characterized by a pointwise zero potential vorticity. The use of a generalized potential vorticity such as suggested by Bretherton (1966) for quasigeostrophic flows and Schneider et al. (2003) for more general flows was also considered. In this case, the interior potential vorticity is supplemented by sheet of potential vorticity just inside the horizontal boundaries at $z = 0$ and 1. However, for this very specific problem, even the use of a generalized potential vorticity did not lead to meaningful estimates of degree of balance of the considered eigenmodes. Since only the boundary values of fields appear effectively in the cost function, the amount of underdetermindness of the used set of balanced equations lead to that the balanced part of the eigenmodes could always be fitted such that it contained all of the generalized potential vorticity, which would lead to the misleading result that all eigenmodes are fully balanced. This does not mean that, for such a fit, the unbalanced part of the solution is absolutely small, let us say in a l_2 sense, but merely that the extended potential vorticity (PV) content of the unbalanced part is small relative to the extended PV content of the balanced solution, which is dominated by the surface PV. This makes the choice of (extended) PV as a basis for the cost function a poor one and we therefore did not use it in the remainder of the work.

Another difference in the present approach is that most previous ones were restricted to the analysis of a particular snapshot in time. Here, because of the harmonic nature of the perturbations in time and the linearity of the perturbation equations, we are able to incorporate the evolution of the perturbations into the analysis. Using quadratic forms like ff^* amounts to fitting the balanced component of the unstable eigenmodes over a full oscillation period. As illustration consider the integral,

$$\int\int\int\int |f|^2\, dx\, dy\, dz\, dt, \tag{16}$$

where $f = \hat{f}(z) \exp(\sigma t) \exp[i(kx + ly)]$ is harmonic in x, y, and t if we neglect the real growth rate σ_{re}. The integral, which is over one full period in x, y, and t, (16) can be evaluated analytically:

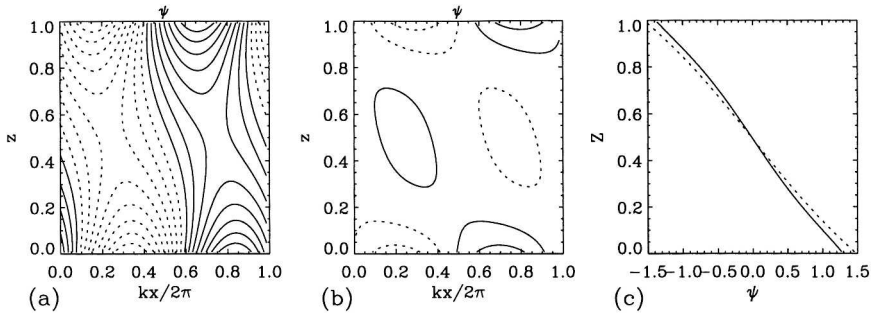

FIG. 5. (a) Streamfunction, $\Psi(x, z)$, for the geostrophic instability (Ro = 1). (b) The unbalanced component, $\Psi_{un}(x, z)$, based on a total energy norm in the cost function. (c) Vertical profiles of Ψ (solid) and Ψ_{bal} (dots) at $x = 0$.

$$\int\int\int\int \hat{f} \exp[i(kx + ly + \sigma_{im}t)]^2 \, dx \, dy \, dz \, dt = \frac{\pi^3}{2} \int (f_{re}^2 + f_{im}^2) \, dz, \tag{17}$$

where f_{re} and f_{im} are the real and imaginary parts of $\hat{f}$ respectively. It is now obvious that (17) is equal to ff^*, multiplied by the constant $\pi^3/2^*$.

In Fig. 5 the perturbation eigenmode streamfunction and its balanced fit are shown for the geostrophic mode at Ro = 1. Figure 5a shows a y–z slice of Ψ calculated with the Boussinesq equations. In Fig. 5b, the unbalanced part of this solution is shown, based on a total energy norm. In Fig. 5c, the vertical profiles of the real part of the total and the balanced component of $\hat{\Psi}$ are shown. As expected, the unbalanced part is relatively small (~10%). This indicates a high degree of balance for the geostrophic instability, even at the rather large value of Ro = 1.

In Fig. 6 the streamfunction is shown for the ageostrophic mode for Ro = 1. From Fig. 6c it is clear that throughout the lower part of the domain, the solution is mostly balanced, but in the upper part a critical layer is evident, and the solution is highly unbalanced in its vicinity. Critical layers or "inertia critical levels" (Jones 1967) are levels where the Doppler-shifted wave frequency equals plus or minus the Coriolis frequency. Mathematically they result in singularities of the equations for neutral waves. The position of these critical layers depends on the wavenumbers and eigenvalue and therefore cannot be determined a priori. A discussion of the physical characteristics of critical layers can be found, for example, in Kitchen and McIntyre (1980). The occurrence of a critical layer becomes even more evident as Ro decreases and the growth rate becomes very small (see Fig. 7).

Figure 8 shows the degree of balance of the strongest geostrophic and ageostrophic unstable modes as a function of Ro. The degree of unbalance of the geostrophic mode (shown with a solid line) is small for Ro = 1 (7%) and rapidly decreases for Ro $\rightarrow$ 0. In contrast, the degree of unbalance of the ageostrophic mode (dotted line) is large and remains relatively constant throughout the range of Ro.

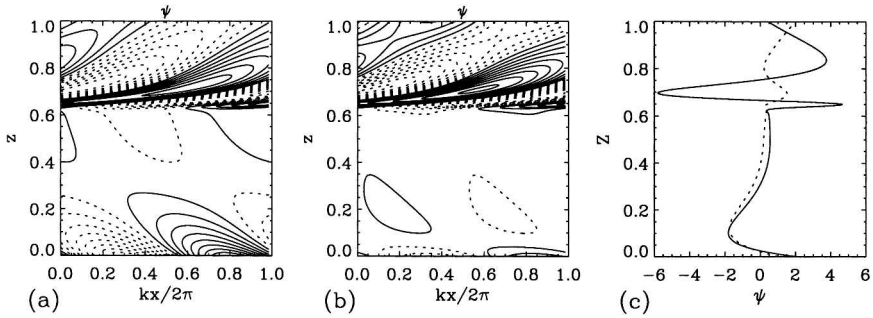

FIG. 6. (a) Streamfunction $\Psi(x, z)$ for the strongest ageostrophic instability at Ro = 1 (optimized over horizontal wave numbers). (b) Unbalanced part of the streamfunction; notice that near the critical layer the streamfunction is almost completely unbalanced. (c) Vertical profiles of Ψ (solid line) and Ψ_{bal} at $x = 0$.

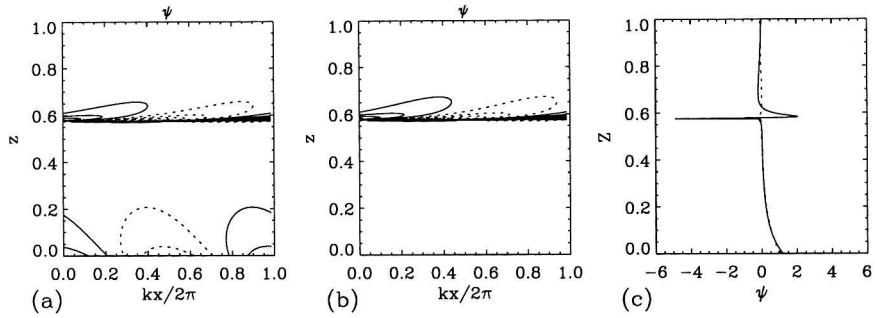

FIG. 7. As in Fig. 6, but for Ro = $1/\sqrt{2}$ (Ri = 2).

From the point of view of loss of balance, we thus obtain the following picture. The baroclinically unstable shear flow may lose energy into unbalanced motions in two ways. First, the geostrophic baroclinic instability, which has a nearly constant growth rate with Ro, becomes increasingly unbalanced for larger Ro, albeit modestly so. Second, the ageostrophic instability, which has a nearly constant and large degree of unbalance, attains a significant growth rate when Ro becomes $O(1)$. To test the generality of these conclusions, we have repeated the results of the last section using the kinetic energy as a fitting norm (appendix C). These computations lead to results that are only slightly different from the ones using total energy, and the estimates of degree of balance remain the same within a few percent.

6. Discussion

We have reexamined the classical geophysical fluid-dynamical problem of baroclinic instability in a uniformly sheared parallel flow to emphasize its dependence on the Rossby number Ro. In particular we have investigated both geostrophic and ageostrophic baroclinic instability modes for their degree of diagnostic momentum balance and for their relationship to the conditions for loss of time integrability of the balance equations. Our principal conclusions are that the geostrophic mode is nearly balanced for all values of Ro examined; the ageostrophic mode is highly unbalanced for all Ro; and the occurrence of significant ageostrophic growth rates (i.e., comparable to the geostrophic ones) occurs in the neighborhood of the $A - S = 0$ and $A = 0$ conditions for loss of balance. The ageostrophic mode is thus clearly distinct from the more familiar geostrophic mode. Although it does not exhibit any critical transition at a finite Ro value, it does achieve a significant growth rate even before the onset of centrifugal instabilities at $A = 0$. Its growth rate becomes vanishingly small, at nearly an exponential rate, as Ro $\rightarrow$ 0. Since the mean baroclinic flow is inherently anticyclonic (i.e., its potential vorticity is reduced relative to the stratified resting state), we view this solution to the baroclinic instability problem as consistent with the paradigm that is emerging from previously discovered examples of unbalanced, anticyclonic, ageostrophic instabilities that arise at intermediate values of Ro between 0 and 1.

We have introduced a method to analyze the degree of balance of solutions of the Boussinesq equations and thus to decompose the flow into balanced and unbalanced components. This method appears to be satisfactory for the particular linear eigenvalue problem examined here, insofar as it allows us to draw rather sharp conclusions about the mostly balanced geostrophic and mostly unbalanced ageostrophic modes. However, the method involves fitting a balanced component—de-

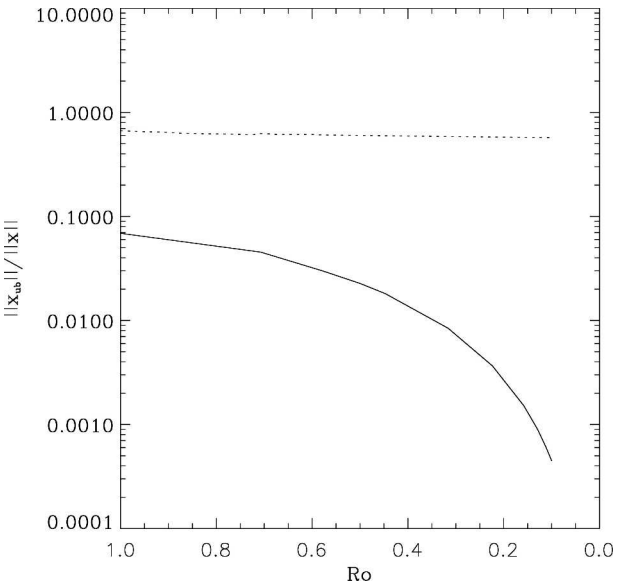

FIG. 8. Ratio of norm of the unbalanced part of the solution to the norm of the full solution, using total energy as the cost function, for geostrophic (solid line) and ageostrophic (dots) modes.

fined entirely by diagnostic constraints derived from the balance equations (including the ω equation)—to a Boussinesq solution whose time dependence is not necessarily consistent with the time evolution implied by the balance equations. Because the linear modal solutions have a factorable harmonic time dependence, our balanced fits are uniformly valid over the oscillation cycle. In more general, nonlinear evolutionary regimes, this procedure may require modification to reconcile the balanced and Boussinesq tendencies. Also, the norms used here to measure the degree of unbalance were different types of energy norms. In general, we expect that the use of potential vorticity will also prove to be useful in providing insight in the degree of balance of a particular flow. Only the singular nature of the Eady problem, with zero interior gradient on the potential vorticity of the basic state leads to the fact that, for this problem, potential vorticity measures do not give meaningful answers for the solution decomposition.

As always in discussions of balanced dynamics, the particular definition of balance is an issue since many alternatives have been proposed. For example, the balance equations defined in isentropic coordinates (Gent and McWilliams 1984) might be used as an alternative in section 5, since it is the basis for the conditions for loss of balance in section 4 and it has a potential vorticity that is an exact Lagrangian invariant [even though not one that is equal to the perturbation Ertel potential vorticity at relative $O(\varepsilon\lambda^2)$]. However, since the distinction between balanced and unbalanced parts is so clear in the geostrophic and ageostrophic instabilities (see Fig. 8), we have chosen not to pursue these alternatives here.

At finite Ro values the ageostrophic instability provides a mechanism through which the balanced mean flow can transfer energy into the unbalanced manifold. This consequently can be the first step of kinetic energy entering into a forward energy cascade en route to eventual dissipation. What still remains to be investigated is the nonlinear evolution along this path, since we have only looked here at the linear instability problem. The linear solution does not allow us to estimate how large the energy flux from balanced to unbalanced flow components may become in the equilibrated regimes. Nevertheless, this study provides a partial answer to the question of how a highly balanced, large-scale circulation may dissipate energy through a local forward energy cascade into unbalanced motions.

Acknowledgments. JCM and MJM received support from the National Science Foundation Grant OCE 96-33681 and the Office of Naval Research Grant N00014-98-1-0165.

APPENDIX A

Discretized Equations

The Boussinesq equations are discretized in the vertical using a staggered grid. All variables are defined at cell centers with exception of w, which is defined at cell boundaries. This results in a set of $5 \times l$ coupled algebraic equations for $5 \times l$ unknowns. Formally, one would need integral conditions for p and b, since one can add an arbitrary constant to either variable without dynamical effect. However, for eigenvalue purposes, this merely results in two extra neutral modes, characterized by a constant eigenmode and zero eigenfrequency:

$$Ro(\sigma u_i + ilV_i u_i) - v_i + ikp_i = 0, \tag{A1a}$$

$$\mathrm{Ro}[\sigma v_i + ilV_i v_i + \mathrm{Ro}V_z(w_i - w_{i-1})/2] + u_i + ilp_i = 0, \tag{A1b}$$

$$\mathrm{Ro}^2\lambda^2(\sigma w_i + ilV_i w_i) + (p_{i+1} - p_i)/dz - (b_{i+1} - b_i)/2 = 0, \tag{A1c}$$

$$iku_i + ilv_i + \mathrm{Ro}(w_i - w_{i-1})/dz = 0, \quad \text{and} \tag{A1d}$$

$$\sigma b_i + ilVb_i + u_i + (w_i + w_{i-1})/2 = 0. \tag{A1e}$$

All equations are defined for $i \in [1, n]$, where n is the number of grid points, with the exception of (A1c), which is defined for $i \in [1, n - 1]$ and is supplemented with the boundary condition $w_n = 0$. Here w_0 is not considered an unknown of this system and is set to $w_0 = 0$ at all times. We can now take a linear combination of these equations and neglect small terms to arrive at a (discretized) system of balance equations that is consistent with the discretized Boussinesq equations:

$$\sigma K^2\Psi_i + ilV_iK^2\Psi_i - ik\mathrm{Ro}V_z(w_i + w_{i-1})/2 + K^2\chi_i = 0, \tag{A2a}$$

$$\Psi_i = p_i, \tag{A2b}$$

$$2(p_{i+1} - p_i)/dz = b_{i+1} + b_i, \tag{A2c}$$

$$dzK^2\chi_i = w_i - w_{i-1}, \quad \text{and} \tag{A2d}$$

$$\sigma b_i + ilV_i b_i + (-il\Psi_i + ik\mathrm{Ro}\chi_i) + (w_i + w_{i-1})/2 = 0. \tag{A2e}$$

Analogous to the continuous derivation, previously discussed, we now take a linear combination of above equations to eliminate the eigenvalue σ and arrive at a discrete *omega* equation in a way that is consistent with the discretization,

$$-il(V_z + 1)(\Psi_{i+1} + \Psi_i) + ik\mathrm{Ro}(V_z + 1)(\chi_{i+1} + \chi_i) - 2(\chi_{i+1} - \chi_i)/dz + (w_{i+1} + 2w_i + w_{i-1})/2 = 0. \quad \text{(A3)}$$

APPENDIX B

Ertel Potential Vorticity

The dimensional Ertel potential vorticity is defined (e.g., Pedlosky 1987, p. 39) by

$$\Pi = (w_y - v_z)b_x + (u_z - w_x)b_y + (v_x - u_y + f)(N^2 + b_z). \quad \text{(B1)}$$

Using the nondimensionalizing scales defined following (3) plus fN^2 for Π itself, we arrive at a nondimensional potential vorticity,

$$\Pi = (\mathrm{Ro}^2\lambda^2 w_y - \mathrm{Ro}v_z)b_x + (\mathrm{Ro}u_z - \mathrm{Ro}^2\lambda^2 w_x)b_y + (\mathrm{Ro}v_x - \mathrm{Ro}u_y + 1)(\mathrm{Ro}^{-1} + b_z). \quad \text{(B2)}$$

For the linear perturbations under consideration, this reduces to a perturbation potential vorticity,

$$\Pi = \mathrm{Ro}^2\lambda^2 w_y - \mathrm{Ro}v_z + \mathrm{Ro}b_x + (v_x - u_y) + b_z, \quad \text{(B3)}$$

and a mean-state potential vorticity, $\Pi = \mathrm{Ro}^{-1} - \mathrm{Ro}$. For the linear eigenmodes, the perturbation form is

$$\Pi = il\mathrm{Ro}^2\lambda^2 w - \mathrm{Ro}\partial_z v + ik\mathrm{Ro}b + (ikv - ilu) + \partial_z b. \quad \text{(B4)}$$

In terms of (Ψ, χ, w, p, b), this is

$$\Pi = -(K^2 + ik\mathrm{Ro}\partial_z)\Psi - il\mathrm{Ro}^2\partial_z\chi + il\mathrm{Ro}^2\lambda^2 w + (ik\mathrm{Ro} + \partial_z)b. \quad \text{(B5)}$$

Since the basic state of the Eady problem possesses a PV that is identical zero everywhere we have sought to use the extended potential vorticity as originally proposed by Bretherton (1966). He proposed that the interior potential vorticity is supplemented by sheet of potential vorticity just inside the horizontal boundaries at $z = 0$ and 1. Following Bretherton (1966), the amount of surface PV, which results from a varying temperature at the boundary, is

$$\Pi_{\mathrm{surf}} = fb\delta(z - z_s). \quad \text{(B6)}$$

At the time, Bretherton was working with a quasigeostrophic flow and recently an extended PV has been proposed for more general flows (Schneider et al. 2003). For inviscid and adiabatic Boussinesq flow, the surface potential vorticity becomes

$$\Pi_{\mathrm{surf}} = (\omega \cdot \mathbf{n})b\delta(z - z_s) = (v_x - u_y + f)b_{z=1} - (v_x - u_y + f)b_{z=0}. \quad \text{(B7)}$$

However, since the basic state depends linearly on x, this expression would introduce a secular term if used in a cost function for the (periodic) perturbations. We were not able to find a way to deal in an appropriate manner with that and therefore opted to use the original expression from Bretherton. In either case, as described in the text, the fit to the set of balanced constraint is underdetermined in such a way that we are able to fit any perturbation completely to the surface PV. This leads to a characterization that, falsely, would indicate that all perturbations are completely balanced. So finally, as discussed in the text, none of the considered definitions of PV proved to be useful for this particular problem. We do envision that for more general flows, the use of PV to characterize the degree of balance of a flow will prove to be a useful and practical quantity.

APPENDIX C

Projection onto the Balanced Manifold

The solutions to the set of balanced constraints, (13b,c) and (14), define a manifold in function space. Minimization of the cost function $\mathcal{L}$ may also be seen as a projection operation given a suitably defined inner product. Define the balanced manifold with

$$\mathbf{B}\mathbf{x}_{\mathrm{bal}} = 0, \quad \text{(C1)}$$

where $\mathbf{B}$ represents the set of (discretized) balanced constraints and $\mathbf{x}$ is the solution vector. Now introduce the inner product,

$$(\mathbf{x}, \mathbf{y})_D = \mathbf{x}^{\mathrm{T}}\mathbf{D}\mathbf{y}^*, \quad \text{(C2)}$$

where $\mathbf{D}$ is a positive semidefinite, Hermitian matrix and * indicates a complex conjugate. Given a solution $\mathbf{x}$ of the Boussinesq equations, minimize the distance to a solution to the balanced constraints $\mathbf{x}_{\mathrm{bal}}$ or minimize $(\mathbf{e}, \mathbf{e})_D$, where $\mathbf{e} = \mathbf{x} - \mathbf{x}_{\mathrm{bal}}$; $\mathbf{e}$ must then be orthogonal to $\mathbf{x}_{\mathrm{bal}}$; hence

$$\mathbf{x}_{\mathrm{bal}}^{\mathrm{T}}\mathbf{D}\mathbf{e}^* = 0. \quad \text{(C3)}$$

This can be solved if we let $\mathbf{D}\mathbf{e}^* = \mathbf{B}^{\mathrm{T}}\boldsymbol{\lambda}$ since this yields

$$\mathbf{x}_{\mathrm{bal}}^{\mathrm{T}}\mathbf{D}\mathbf{e}^* = \mathbf{x}_{\mathrm{bal}}^{\mathrm{T}}\mathbf{B}^{\mathrm{T}}\boldsymbol{\lambda} = (\mathbf{B}\mathbf{x}_{\mathrm{bal}})^{\mathrm{T}}\boldsymbol{\lambda} = 0. \quad \text{(C4)}$$

We therefore find the following system to be solved:

$$\mathbf{D}^*\mathbf{x} - \mathbf{D}^*\mathbf{x}_{\mathrm{bal}} = \mathbf{B}^{\mathrm{H}}\boldsymbol{\lambda}^* \quad \text{and} \quad \text{(C5a)}$$

$$\mathbf{B}\mathbf{x}_{\mathrm{bal}} = 0, \quad \text{(C5b)}$$

for which a solution can be found by solving

$$\mathbf{B}\mathbf{D}^{*-1}\mathbf{B}^{T^*}\boldsymbol{\lambda}^* = \mathbf{B}\mathbf{x}. \tag{C6}$$

Different choices of $\mathbf{D}$ will correspond to specific cost functions $\mathcal{L}$. For instance, if we are considering the energy as a measure of unbalance, then

$$\begin{aligned} E &= 0.5(uu^* + vv^* + \mathrm{Ro}^2\lambda^2 ww^* + bb^*) \\ &= 0.5(K^2\Psi\Psi^* + \mathrm{Ro}^2K^2\chi\chi^* + \mathrm{Ro}^2\lambda^2 ww^* + bb^*), \end{aligned} \tag{C7}$$

and the cost function is

$$\mathcal{L} = \sum_{i=1}^{i=n} dz\, E[x(i) - x_{\mathrm{bal}}(i)]. \tag{C8}$$

In this case, $\mathbf{D}$ is simply

$$\mathbf{D} = \begin{bmatrix} K^2 & 0 & 0 & 0 & 0 \\ 0 & \mathrm{Ro}^2K^2 & 0 & 0 & 0 \\ 0 & 0 & \mathrm{Ro}^2\lambda^2 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix}. \tag{C9}$$

Since the pressure p does not appear in these cost functions, $\mathbf{D}$ cannot be inverted in this form. To overcome this we merely need to use (13b) to eliminate p from $\mathbf{D}$ and the system of balanced constraints.

REFERENCES

Bretherton, F. P., 1966: Critical layer instability in baroclinic flows. *Quart. J. Roy. Meteor. Soc.,* **92,** 325–334.

Charney, J. G., 1947: The dynamics of long waves in a baroclinic westerly current. *J. Meteor.,* **4,** 135–163.

——, 1971: Geostrophic turbulence. *J. Atmos. Sci.,* **28,** 1087–1095.

Drazin, P. G., and L. N. Howard, 1966: Hydrodynamic stability of parallel flow of inviscid fluid. *Adv. Appl. Mech.,* **9,** 1–89.

Eady, E. T., 1949: Long waves and cyclone waves. *Tellus,* **1,** 33–52.

Gent, P. R., and J. C. McWilliams, 1984: Balanced models in isentropic coordinates and the shallow water equations. *Tellus,* **36A,** 166–171.

Golub, G. H., and C. F. Van Loan, 1996: *Matrix Computations.* Johns Hopkins Studies in Mathematical Sciences, 694 pp.

Hoskins, B. J., 1974: The role of potential vorticity in symmetric stability and instability. *Quart. J. Roy. Meteor. Soc.,* **100,** 480–482.

——, M. E. McIntyre, and A. W. Robertson, 1985: On the use and significance of isentropic potential vorticity maps. *Quart. J. Roy. Meteor. Soc.,* **111,** 877–946.

Howard, L. N., 1961: Note on a paper of John W. Miles. *J. Fluid Mech.,* **10,** 509–512.

Jones, W. L., 1967: Propagation of internal gravity waves in fluids with shear and rotation. *J. Fluid Mech.,* **30,** 439–448.

Kitchen, E. H., and M. E. McIntyre, 1980: On whether inertia-gravity waves are absorbed of reflected when their intrinsic frequency is Doppler-shifted towards *f. J. Meteor. Soc. Japan,* **58,** 118–126.

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

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

——, and I. Yavneh, 1998: Fluctuation growth and instability associated with a singularity of the Balance Equations. *Phys. Fluids,* **10,** 2587–2596.

——, J. B. Weiss, and I. Yavneh, 1994: Anisotropy and coherent structures in planetary turbulence. *Science,* **264,** 410–413.

——, I. Yavneh, M. J. P. Cullen, and P. R. Gent, 1998: The breakdown of large-scale flows in rotating, stratified fluids. *Phys. Fluids,* **10,** 3178–3184.

——, M. J. Molemaker, and I. Yavneh, 2001: From stirring to mixing of momentum: Cascades from balanced flows to dissipation in the oceanic interior. *'Aha Huliko'a Hawaiian Winter Workshop,* Honolulu, HI, University of Hawaii at Manoa, 59–66.

Miles, J. W., 1961: On the stability of heterogeneous shear flows. *J. Fluid Mech.,* **10,** 496–508.

Molemaker, M. J., J. C. McWilliams, and I. Yavneh, 2000: Instability and equilibration of centrifugally-stable stratified Taylor-Couette flow. *Phys. Rev. Lett.,* **86,** 5270–5273.

Nakamura, N., 1988: Scale selection of baroclinic instability—Effects of stratification and nongeostrophy. *J. Atmos. Sci.,* **45,** 3253–3267.

Olsson, P. Q., and W. R. Cotton, 1997: Balanced and unbalanced circulations in a primitive equation simulation of a midlatitude MCC. Part II: Analysis of balance. *J. Atmos. Sci.,* **54,** 479–497.

Pedlosky, J., 1987: *Geophysical Fluid Dynamics.* 2d ed. Springer, 710 pp.

Phillips, N. A., 1954: Energy transformations and meridional circulations associated with simple baroclinic waves in a two-level, quasi-geostrophic model. *Tellus,* **6,** 273–286.

Schneider, T., I. M. Held, and S. T. Garner, 2003: Boundary effects in potential vorticity dynamics. *J. Atmos. Sci.,* **60,** 1024–1040.

Sleijpen, G. L. G., and H. A. van der Vorst, 1996: A Jacobi-Davidson iteration method for linear eigenvalue problems. *SIAM J. Matrix Anal. Appl.,* **17,** 401–425.

Stone, P. H., 1966: On non-geostrophic baroclinic instability. *J. Atmos. Sci.,* **23,** 390–400.

——, 1970: On non-geostrophic baroclinic instability: Part II. *J. Atmos. Sci.,* **27,** 721–726.

Yavneh, I., J. C. McWilliams, and M. J. Molemaker, 2001: Non-axisymmetric instability of centrifugally stable, stratified Taylor-Couette flow. *J. Fluid Mech.,* **448,** 1–21.