Two-level primitive equation baroclinic instability on an f-plane

By KLAUS FRAEDRICH* and THOMAS FRISIUS[†]

Meteorologisches Institut, Universität Hamburg, Germany

SUMMARY

Baroclinic instability of a two-level primitive equation (PE) model is represented by a dynamical system consisting of the four interacting fields of barotropic and baroclinic vorticity, divergence and temperature. The instability mechanism is illustrated in terms of a positive feedback loop with a succession of tendencies. It occurs at wavelengths longer than the short wave cut-off, for which vortex stretching prevails over vertical temperature advection. Quasi-geostrophy (QG) reduces the positive feedback loop to the interaction of the barotropic and baroclinic vorticity fields which are diagnostically linked to divergence and temperature by the omega-equation and geostrophy, respectively.

KEYWORDS: Baroclinic instability   Primitive equations   Vorticity dynamics

1. INTRODUCTION

In the wake of the early numerical forecasting trial by Charney et al. (1950) based on the equivalent barotropic model, Phillips (1951, see also 1954) introduced the first two-level version of the baroclinic instability problem complementing Charney's (1947) and Eady's (1949) preceding analyses of vertically continuous atmospheres. Employing the quasi-geostrophic (QG) approximation, Phillips stated that it is not necessary to make the quasi-geostrophic assumption. And he comments that he considers the quasi-geostrophic framework, "because the nonlinear model uses the geostrophic assumption, so that it was not considered necessary or even advisable to use this refinement in studying the perturbation behaviour." Since those days, numerical weather prediction has passed its infancy and the quasi-geostrophic approximation, being replaced by the primitive equations (PE), has fallen out of favour for modelling the atmospheric circulation (see James 1994). Still, almost all textbooks present the quasi-geostrophic analysis of baroclinic instability and the description of its mechanism following basically these seminal papers; they discard Phillips' comment to provide the necessary update analyzing baroclinic instability in the PE environment. There is indeed a number of studies dealing with primitive equation baroclinic instability (e. g. Wiin-Nielsen 1963, Árnason 1963, Stone 1966). However, it is not clear from these studies how baroclinic instability operates in a PE model in distinction to the QG model.

Therefore, it is the purpose of this note to present a baroclinic instability analysis using the primitive equations in the framework of a two-level model. The advantage of this approach is that (i) the influence of the quasi-geostrophic approximation can be simply demonstrated, (ii) the prognostic divergence equation needs to be included, and (iii) inertia-gravity waves occur explicitly (section 2). The instability analysis (section 3) is pursued in the conventional normal mode sense yielding asymptotic growth rates. Finally, a conceptual interpretation is introduced to demonstrate the underlying mechanism of baroclinic instability in terms of positive and negative feedbacks (section 4). This, we think, provides a novel interpretation of the process induced by a positive feedback loop.

* Corresponding author: Meteorologisches Institut, Universität Hamburg, Bundesstraße 55, 20146 Hamburg, Germany
[†] Present affiliation: Institut für Meteorologie und Geophysik, J.W. Goethe-Universität, Robert Mayer Str. 1, 60325 Frankfurt a. M., Germany

2. Two-level model based on the primitive equations

As the starting point we consider the adiabatic and inviscid PE in isobaric coordinates on an f-plane

$$\left(\frac{\partial}{\partial t} + \mathbf{v} \cdot \nabla + \omega \frac{\partial}{\partial p}\right) \mathbf{v} + f\mathbf{k} \times \mathbf{v} = -g\nabla z \ , \tag{1a}$$

$$\left(\frac{\partial}{\partial t} + \mathbf{v} \cdot \nabla + \omega \frac{\partial}{\partial p}\right) \theta - S\omega = 0 \ , \tag{1b}$$

$$\nabla \cdot \mathbf{v} + \frac{\partial \omega}{\partial p} = 0 \ , \tag{1c}$$

$$\frac{\partial z}{\partial p} = -\frac{h\theta}{g} \ . \tag{1d}$$

The prognostic and diagnostic variables of the PE are the horizontal velocity vector $\mathbf{v} = (u, v)$, the potential temperature deviation θ from a horizontal mean reference potential temperature $\theta_R(p)$, the geopotential height deviation z and the pressure velocity ω. The time coordinate is denoted by t, and pressure p is taken as the vertical coordinate while the Nabla operator $\nabla = (\partial/\partial x, \partial/\partial y)$ is only applied to the two horizontal coordinates x and y. The Coriolis parameter f is assumed to be constant. Other symbols are the unit vector in vertical direction $\mathbf{k}$, the static stability $S = -\partial \theta_R/\partial p$ of the reference state, the gravitational acceleration g, the gas constant R, and $h(p) = R/p \ (p/1000\text{hPa})^{R/c_P}$.

Following Charney and Phillips (1953) we may use approximate boundary conditions for the pressure velocity

$$\omega = 0 \quad \text{at} \quad p = 0 \ \text{and} \ p_0 \ . \tag{2}$$

Here, $p_0 = 1000\text{hPa}$ is the pressure level that coincides approximately with the surface. Simplifying the 3-dimensional QG equations to a two-level QG model (Charney and Phillips 1953) the atmosphere is divided into 5 levels (0 to 4) where geopotential height is evaluated at odd levels and potential temperature at even levels. Similarly, we can simplify the 3-dimensional PE to a two-level PE model with horizontal velocity at odd and pressure velocity at even levels (see Fig. 1). Wiin-Nielsen (1963) studied baroclinic instability in a very similar two-level PE model; the only difference is that he applied the temperature equation at odd levels to allow variations in static stability. However, this difference is not relevant when the static stability of the basic state is assumed to be constant. The two-level PE model is

$$\left(\frac{\partial}{\partial t} + \mathbf{v}_1 \cdot \nabla\right) \mathbf{v}_1 + \frac{1}{2}\omega_2 \frac{\mathbf{v}_3 - \mathbf{v}_1}{\Delta p} + f\mathbf{k} \times \mathbf{v}_1 = -g\nabla z_1 \ , \tag{3a}$$

$$\left(\frac{\partial}{\partial t} + \mathbf{v}_3 \cdot \nabla\right) \mathbf{v}_3 + \frac{1}{2}\omega_2 \frac{\mathbf{v}_3 - \mathbf{v}_1}{\Delta p} + f\mathbf{k} \times \mathbf{v}_3 = -g\nabla z_3 \ , \tag{3b}$$

$$\left(\frac{\partial}{\partial t} + \frac{1}{2}(\mathbf{v}_1 + \mathbf{v}_3) \cdot \nabla\right) \theta_2 + \frac{1}{2}\omega_2 \frac{\theta_4 - \theta_0}{\Delta p} - S\omega_2 = 0 \ , \tag{3c}$$

$$\nabla \cdot \mathbf{v}_1 + \frac{\omega_2}{\Delta p} = 0 \ , \tag{3d}$$

$$\nabla \cdot \mathbf{v}_3 - \frac{\omega_2}{\Delta p} = 0 \ , \tag{3e}$$

Pressure	Levels		Variables
0 hPa	0	————————————————	$T_0, \ \omega_0 = 0$
250 hPa	1	– – – – – – – – – – – – –	$\mathbf{v}_1, \ z_1$
500 hPa	2	————————————————	$T_2, \ \omega_2$
750 hPa	3	– – – – – – – – – – – – –	$\mathbf{v}_3, \ z_3$
1000 hPa	4	————————————————	$T_4, \ \omega_4 = 0$

Figure 1.   Vertical structure of the two-level PE model

$$z_1 - z_3 = \frac{h_2 \Delta p \ \theta_2}{g} \ , \qquad (3f)$$

where $\Delta p = 500 \text{hPa}$ denotes the pressure difference between odd or even numbered levels. Only the vertical advection terms in the momentum equations (IIIa) and (IIIb), which are not present in the QG model, are represented by uncentred differences while the QG model (subsection 4c) uses centred differences throughout. It follows from (3d) and (3e) that the barotropic (vertically averaged) wind is nondivergent. This is a consequence of the boundary conditions (2) which are equivalent to impose two rigid horizontal boundaries on an incompressible fluid.

3.   BAROCLINIC INSTABILITY

Baroclinic instability in the two-level PE system is obtained from perturbations of an unstable Eady-type basic flow

$$\bar{u}_1 = U_T, \ \ \bar{u}_3 = -U_T, \ \ \bar{v}_1 = \bar{v}_3 = \bar{\omega}_1 = \bar{\omega}_3 = 0, \ \ \overline{\theta}_0 = \overline{\theta}_2 = \overline{\theta}_4 = -\frac{2f}{h_2 \Delta p} U_T \ y \ , \qquad (4)$$

where U_T is the thermal wind. For the baroclinic instability problem the PE are linearized about this basic state, where primes denote the deviations from the basic state and $\mathbf{i}$ is the unit vector in the zonal direction.

$$\left(\frac{\partial}{\partial t} + U_T \frac{\partial}{\partial x} \right) \mathbf{v}_1' - \frac{U_T}{\Delta p} \omega_2' \mathbf{i} + f \mathbf{k} \times \mathbf{v}_1' = -g \nabla z_1' \ , \qquad (5a)$$

$$\left(\frac{\partial}{\partial t} - U_T \frac{\partial}{\partial x} \right) \mathbf{v}_3' - \frac{U_T}{\Delta p} \omega_2' \mathbf{i} + f \mathbf{k} \times \mathbf{v}_3' = -g \nabla z_3' \ , \qquad (5b)$$

$$\frac{\partial \theta_2'}{\partial t} - \frac{f U_T}{h_2 \Delta p} (v_1' + v_3') - S \omega_2' = 0 \ , \qquad (5c)$$

$$\nabla \cdot \mathbf{v}_1' + \frac{\omega_2'}{\Delta p} = 0 \ , \qquad (5d)$$

$$\nabla \cdot \mathbf{v}_3' - \frac{\omega_2'}{\Delta p} = 0 \ , \qquad (5e)$$

$$z_1' - z_3' = \frac{h_2 \Delta p \; \theta_2'}{g} \; . \tag{5f}$$

For further treatment it is useful to denote the vertical average of a variable G by $G_M \equiv (G_1 + G_3)/2$ and the half vertical difference of G by $G_T \equiv (G_1 - G_3)/2$. Then, taking sum and difference of the horizontal momentum equations (5a) and (5b) leads to:

$$\frac{\partial \mathbf{v}_M'}{\partial t} + U_T \frac{\partial \mathbf{v}_T'}{\partial x} - \frac{U_T}{\Delta p} \omega_2' \mathbf{i} + f \mathbf{k} \times \mathbf{v}_M' = -g \nabla z_M' \; , \tag{6a}$$

$$\frac{\partial \mathbf{v}_T'}{\partial t} + U_T \frac{\partial \mathbf{v}_M'}{\partial x} + f \mathbf{k} \times \mathbf{v}_T' = -g \nabla z_T' \; . \tag{6b}$$

These equations can be replaced by vorticity and divergence equations by applying curl and divergence. Finally, a closed set is obtained when potential temperature and pressure velocity are eliminated using the hydrostatic and continuity equation, respectively. Introducing the streamfunctions ψ_1, ψ_3, the horizontal divergencies D_1, D_3 with $D_1 + D_3 = 0$, and the reciprocal value of the radius of deformation

$$\kappa = f/(h_2 \Delta p^2 S)^{\frac{1}{2}} \tag{7}$$

leads to the following four equations for barotropic vorticity $\nabla^2 \psi_M' = \nabla^2 (\psi_1' + \psi_3')/2$, baroclinic vorticity $\nabla^2 \psi_T' = \nabla^2 (\psi_1' - \psi_3')/2$, baroclinic divergence $D_T' = -\omega_2'/\Delta p$ and potential temperature in terms of thickness $z_T' = h_2 \Delta p \; \theta_2'/(2g)$ or geostrophic vorticity $g/f \nabla^2 z_T'$.

$$\frac{\partial \nabla^2 \psi_M'}{\partial t} = \underbrace{-U_T \frac{\partial \nabla^2 \psi_T'}{\partial x}}_{(VI)} \underbrace{+ U_T \frac{\partial D_T'}{\partial y}}_{(0)} \; , \tag{8a}$$

$$\frac{\partial \nabla^2 \psi_T'}{\partial t} = \underbrace{-U_T \frac{\partial \nabla^2 \psi_M'}{\partial x}}_{(II)} \underbrace{- f D_T'}_{(V)} \; , \tag{8b}$$

$$\frac{\partial D_T'}{\partial t} = \underbrace{-g \nabla^2 z_T'}_{(IIIa)} \underbrace{+ f \nabla^2 \psi_T'}_{(IIIb)} \; , \tag{8c}$$

$$\frac{g}{f} \frac{\partial \nabla^2 z_T'}{\partial t} = \underbrace{U_T \frac{\partial \nabla^2 \psi_M'}{\partial x}}_{(I)} \underbrace{- \frac{f}{2\kappa^2} \nabla^2 D_T'}_{(IV)} \; . \tag{8d}$$

The tendencies are in the order as labelled in Eqs.(8): (I) the meridional advection of basic state temperature by the barotropic vorticity disturbance, $U_T \partial \nabla^2 \psi_M'/\partial x$, (II) the zonal advection of barotropic perturbation vorticity by the basic flow, $-U_T \partial \nabla^2 \psi_M'/\partial x$; (IIIa) the divergence of pressure gradient force, $-g \nabla^2 z_T'$, (IIIb) the divergence of Coriolis force, $f \nabla^2 \psi_T'$, (IV) the vertical advection of basic state temperature by the divergent flow, $-f/(2\kappa^2) \nabla^2 D_T'$, (V) the vortex stretching by the divergent flow, $-f D_T'$, and, finally, (VI) the zonal advection of baroclinic perturbation vorticity by the basic flow, $-U_T \partial \nabla^2 \psi_T'/\partial x$. The vorticity tilting term $-U_T \partial D_T'/\partial y$ (labelled with 0) vanishes if the perturbation is y-independent (see below). The divergent flow may be replaced by the velocity potential χ_T' introducing $D_T' = \nabla^2 \chi_T'$

In order to make the system nondimensional the following scales are introduced: The horizontal length scale $L = 1/(2^{\frac{1}{2}}\kappa)$, the time scale $\tau = 1/(2^{\frac{1}{2}}\kappa U_T)$, and the height scale $H = U_T f/(2^{\frac{1}{2}} g\kappa)$. Thus, the nondimensional system is characterized only by the Rossby number Ro$= U_T 2^{\frac{1}{2}}\kappa/f$ and becomes

$$\frac{\partial \tilde{\nabla}^2 \Psi}{\partial \tilde{t}} = -\frac{\partial \tilde{\nabla}^2 \Theta}{\partial \tilde{x}} + \frac{\partial D}{\partial \tilde{y}} \ , \tag{9a}$$

$$\frac{\partial \tilde{\nabla}^2 \Theta}{\partial \tilde{t}} = -\frac{\partial \tilde{\nabla}^2 \Psi}{\partial \tilde{x}} - \frac{1}{\text{Ro}}D \ , \tag{9b}$$

$$\frac{\partial D}{\partial \tilde{t}} = \frac{1}{\text{Ro}}\tilde{\nabla}^2 \Theta - \frac{1}{\text{Ro}}\tilde{\nabla}^2 Z \ , \tag{9c}$$

$$\frac{\partial \tilde{\nabla}^2 Z}{\partial \tilde{t}} = \frac{\partial \tilde{\nabla}^2 \Psi}{\partial \tilde{x}} - \frac{1}{\text{Ro}}\tilde{\nabla}^2 D \ , \tag{9d}$$

where $(\tilde{..})$ denotes nondimensional coordinates and $\tilde{\nabla} = (\partial/\partial \tilde{x}, \partial/\partial \tilde{y})$. The variables Ψ, Θ, D and Z are the nondimensional counterparts of ψ'_M, ψ'_T, D'_T and z'_T, respectively. Assuming y-independent perturbations of the form, $(\Psi, \Theta, D, Z) = (\hat{\Psi}, \hat{\Theta}, \hat{D}, \hat{Z}) \exp(i\tilde{k}\tilde{x} - i\tilde{\nu}\tilde{t})$ (y-dependent perturbations in a channel are discussed in the appendix), the system (9) reduces to the following matrix equation

$$\begin{pmatrix} i\tilde{\nu} & -i\tilde{k} & 0 & 0 \\ -i\tilde{k} & i\tilde{\nu} & \dfrac{1}{\text{Ro}\tilde{k}^2} & 0 \\ 0 & -\dfrac{\tilde{k}^2}{\text{Ro}} & i\tilde{\nu} & \dfrac{\tilde{k}^2}{\text{Ro}} \\ i\tilde{k} & 0 & -\dfrac{1}{\text{Ro}} & i\tilde{\nu} \end{pmatrix} \begin{pmatrix} \hat{\Psi} \\ \hat{\Theta} \\ \hat{D} \\ \hat{Z} \end{pmatrix} = \mathbf{0} \ . \tag{10}$$

and leads to a normal mode solution fulfilling the characteristic equation

$$\tilde{\nu}^4 - \left(\frac{1}{\text{Ro}^2}(1 + \tilde{k}^2) + \tilde{k}^2\right)\tilde{\nu}^2 - \frac{1}{\text{Ro}^2}(\tilde{k}^2 - \tilde{k}^4) = 0 \ . \tag{11}$$

There are four roots, two of which are east- and westward travelling inertia-gravity waves with the frequencies

$$\tilde{\nu}_{1,2} = \pm \left(\frac{1}{2}(\frac{1}{\text{Ro}^2}(1 + \tilde{k}^2) + \tilde{k}^2) + \left(\frac{1}{4}(\frac{1}{\text{Ro}^2}(1 + \tilde{k}^2) + \tilde{k}^2)^2 + \frac{1}{\text{Ro}^2}(\tilde{k}^2 - \tilde{k}^4)\right)^{\frac{1}{2}}\right)^{\frac{1}{2}} \ . \tag{12}$$

The other two roots are referred to as baroclinic waves which travel east- and westward for $\tilde{k} > 1$. Otherwise, the eigenvalues become purely imaginary leading to stationary exponentially growing and decaying waves. The frequencies of the baroclinic waves are

$$\tilde{\nu}_{3,4} = \pm \left(\frac{1}{2}(\frac{1}{\text{Ro}^2}(1 + \tilde{k}^2) + \tilde{k}^2) - \left(\frac{1}{4}(\frac{1}{\text{Ro}^2}(1 + \tilde{k}^2) + \tilde{k}^2)^2 + \frac{1}{\text{Ro}^2}(\tilde{k}^2 - \tilde{k}^4)\right)^{\frac{1}{2}}\right)^{\frac{1}{2}} \ . \tag{13}$$

This formula demonstrates clearly that instability occurs for all $\tilde{k} < 1$ whatever the value of Ro. Fig. 2 shows the roots as function of zonal wavenumber and Rossby number. Generally, the frequency (real part, Fig. 2a) increases with increasing zonal wavenumber and decreasing Rossby number and the frequencies of the inertia-gravity waves are larger than of the baroclinic waves. In the limit Ro $\to 0$ the dimensionless frequency $\tilde{\nu}$ of the inertia-gravity waves goes to infinity while the frequency of the baroclinic waves approaches the QG value, i. e.

$$\lim_{\text{Ro} \to 0} \tilde{\nu}_{1,2} = \pm \infty \ , \tag{14a}$$

$$\lim_{\text{Ro} \to 0} \tilde{\nu}_{3,4} = \pm \tilde{k} \left(\frac{\tilde{k}^2 - 1}{\tilde{k}^2 + 1} \right)^{\frac{1}{2}} \ . \tag{14b}$$

The latter result can be verified by the l'Hopital rule. Note that the limit Ro $\to 0$ can only be realized when U_T goes to zero and the timescale τ goes to infinity. Therefore, in dimensional form, the frequencies of the inertia-gravity waves remain finite while the frequency of baroclinic waves approaches zero for Ro $\to 0$. Nevertheless, there is a large frequency gap between inertia-gravity waves and baroclinic waves for low Rossby numbers which justifies the use of filtered models. Indeed, the limit (14) are exactly in agreement with the result of a two-level QG model which is easily obtained by neglecting the time derivative of divergence in (9c), $\partial D / \partial t = 0$. The growth rate (imaginary part, Fig. 2b) is only nonzero for the baroclinic waves; its maximum growth occurs for low Rossby numbers near $\tilde{k}_m = (2^{\frac{1}{2}} - 1)^{\frac{1}{2}}$. With increasing Rossby number the growth rate decreases and the zonal wavenumber of maximum growth shifts to smaller values. The latter result can also be verified by the analytical formula for the zonal wavenumber associated with maximum growth

$$\tilde{k}_m = \left| \frac{\left(2^{\frac{1}{2}} (1 + \text{Ro}^2)^{\frac{3}{2}} - (1 + 3\text{Ro}^2) \right)^{\frac{1}{2}}}{1 - \text{Ro}^2} \right| \ . \tag{15}$$

4. MECHANISM OF BAROCLINIC INSTABILITY

The mechanism of baroclinic instability is conceptually described considering the interaction of the four fields of barotropic vorticity $\nabla^2 \psi'_M$, baroclinic vorticity $\nabla^2 \psi'_T$, geostrophic vorticity (or temperature) $g/f \nabla^2 z'_T$ and divergence D'_T. The wave-structure of the baroclinic instability wave is analyzed first.

(a) Wave-structure

The two-level PE model predicts baroclinic instability for $\tilde{k} < 1$, for which the growing baroclinic wave attains the following structure (Fig. 3)

$$\Psi = A e^{\tilde{\sigma} t} \sin(\tilde{k} \tilde{x}), \tag{16a}$$

$$\Theta = \frac{\tilde{\sigma}}{\tilde{k}} A e^{\tilde{\sigma} t} \cos(\tilde{k} \tilde{x}) \ , \tag{16b}$$

$$D = \text{Ro} \tilde{k} (\tilde{k}^2 + \tilde{\sigma}^2) A e^{\tilde{\sigma} t} \cos(\tilde{k} \tilde{x}), \tag{16c}$$
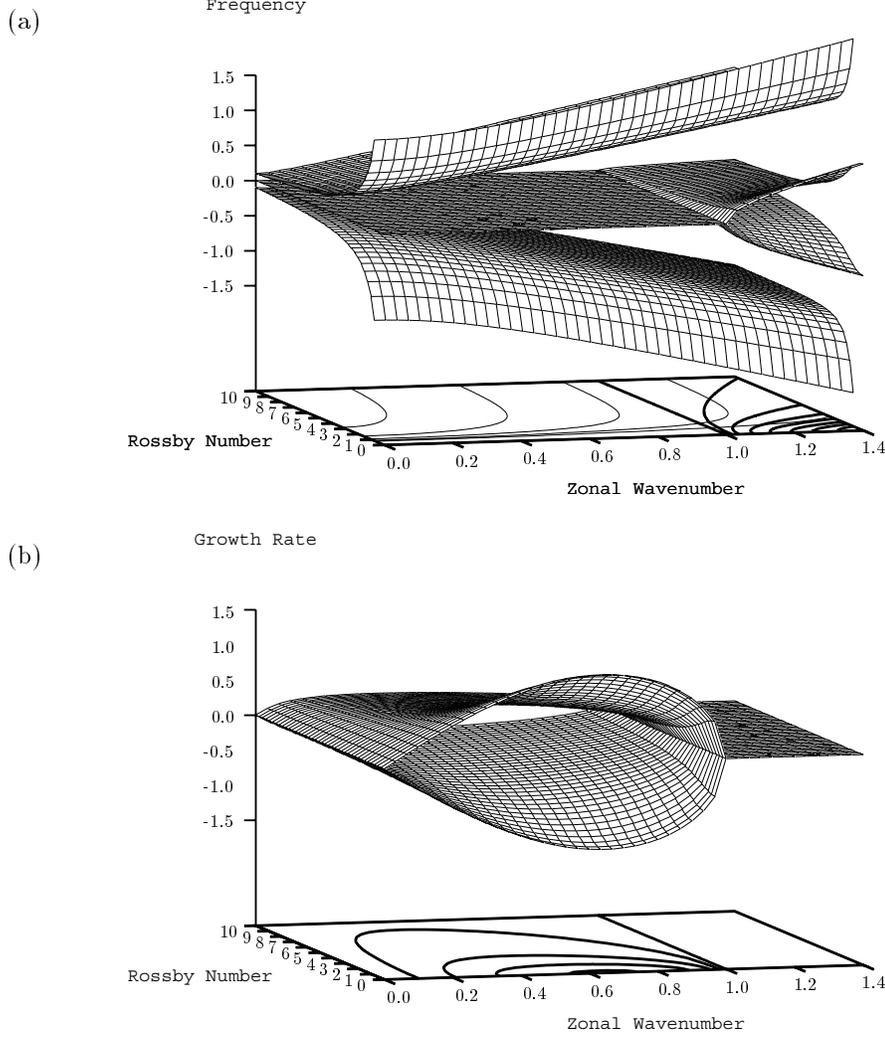
(a)

Frequency

(b)

Growth Rate

Figure 2. (a) The four non-dimensional frequencies $\tilde{\nu}_{1,2,3,4}$ displayed as surfaces in the $\tilde{k}$-Ro-Re$(\tilde{\nu})$ space. The upper and lower surfaces are associated with the inertia-gravity waves while the two middle surfaces are connected with the baroclinic waves. Isolines of the inertia-gravity waves (thin lines with contour interval 0.3) and of the baroclinic waves (bold lines with contourinterval 0.1) are plotted at the bottom (Rossby number, zonal wavenumber)-plane. (b) The imaginary part of the baroclinic wave frequencies (nondimensional) displayed as surfaces in the $\tilde{k}$-Ro-Im$(\tilde{\nu})$ space. At the bottom isolines of the growth rate (contour interval 0.1) are plotted.

$$Z = \frac{\tilde{\sigma}}{\tilde{k}}(1 + \mathrm{Ro}^2(\tilde{k}^2 + \tilde{\sigma}^2))Ae^{\sigma t}\cos(\tilde{k}\tilde{x}) \ , \tag{16d}$$

where $\tilde{\sigma} = \mathrm{Im}(\tilde{\nu}_3)$ is the nondimensional growth rate and A an arbitrary amplitude factor. For small Rossby numbers the structure of the growing baroclinic wave of the QG model is similar to the PE model. The only differences are that (i) the growth rate is given by the imaginary part of the approximated frequency in (14) and (ii) the Ro^2-term in (16d) is neglected. The latter approximation brings the nondivergent flow into exact geostrophic balance.
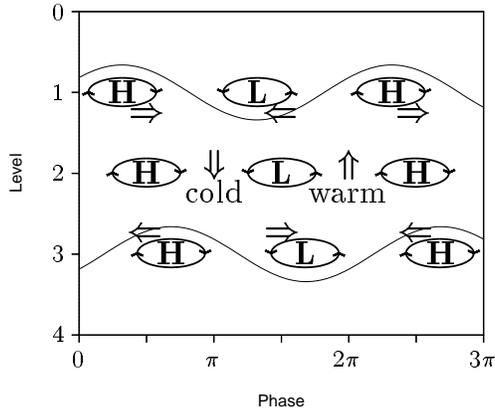
Figure 3.   Schematic baroclinic wave structure in the unstable case: The symbols H and L are placed at the positions of the ridges and troughs, the open arrows indicate the ageostrophic circulation and the vortex symbols the extrema of vorticity.

(b)   Baroclinic instability mechanism in the PE model

The mechanism of baroclinic instability is illustrated as a positive feedback loop which consists of a succession of the seven tendencies of the four normal mode components (equation set 8). This leads to the instability or positive feedback loop which can be visualized by considering the effect of one tendency after the other (Fig. 4):

Tendencies (I) and (II). The barotropic vorticity field $\nabla^2 \psi'_M$ (upper box) induces a warm (cold) anomaly east (west) of the barotropic cyclone (mid-left box) by meridional advection of basic state temperature $U_T \partial \nabla^2 \psi'_M / \partial x$. Simultaneously, the zonal advection of the barotropic vorticity tilts the barotropic vorticity wave in the direction of the vertical shear by the basic flow, $-U_T \partial \nabla^2 \psi'_M / \partial x$. This tilt leads to a baroclinic vorticity field (8b) which is shown in Fig. 4 but of opposite tilt (mid-right box). The generations of geostrophic and baroclinic vorticity, $g/f \nabla^2 z'_T$ and $\nabla^2 \psi'_T$, by these tendencies are identical in their absolute values, but they have opposite signs. Thus, the wave is forced out of geostrophic balance. Which sign (of the generation terms) will finally dominate depends only on the magnitudes of the geostrophy restoring mechanisms (see tendencies 4 and 5, below).

Tendencies (IIIa) and (IIIb). The divergence of the pressure gradient force $-g\nabla^2 z'_T$ of the temperature field and the divergence of the Coriolis force $f\nabla^2 \psi'_T$ of the baroclinic vorticity field accelerate the secondary circulation with midlevel rising (sinking) motion at the position of the warm (cold) anomaly.

Tendencies (IV) and (V). The vertical advection of basic state temperature by the divergence field $-f/2\kappa^2 \nabla^2 D'_T$ and the vortex stretching $-f D'_T$ of the divergence field (lower box) act to restore geostrophic balance by diminishing the tendencies caused by the barotropic vorticity field. The occurrence of instability depends on which of these tendencies dominates. If vertical temperature advection dominates, the temperature field will change its sign so that the baroclinic vorticity field remains with a vertical tilt in the direction of the vertical shear. In this case instability cannot take place. If, on the other hand, vortex stretching dominates, the baroclinic vorticity field will change its sign so that it attains the structure shown in Fig. 4. This causes baroclinic instability for sufficiently small wavenumbers.

Tendency (VI). The horizontal advection of the baroclinic vorticity by the basic

flow $-U_T \partial \nabla^2 \psi'_T / \partial x$, will turn the trough axis of the baroclinic vorticity field towards the vertical direction. This tendency strengthens the barotropic vorticity so that the instability feedback loop is closed.

Short wave cut-off: The criterion for instability can easily be deduced from this analysis since the difference of the tendencies (V) and (IV) gives

$$f(1 - \frac{1}{2\kappa^2}\nabla^2)D'_T \geq 0. \tag{17}$$

That is, instability occurs only for waves with a wavenumber smaller than $(2\kappa^2)^{\frac{1}{2}}$ when vortex stretching prevails over the vertical temperature advection. This is in agreement with the analytical result in section 3.

(c) *Baroclinic instability mechanism in the QG model*

For low Rossby number only slight changes in the divergent secondary circulation are necessary to restore geostrophy. This justifies the QG approximation to neglect the time derivative of baroclinic divergence in (8c) leading to the geostrophic balance of the nondivergent baroclinic flow. That is, the tendencies (IIIa) and (IIIb) cancel:

$$g/f\nabla^2 z'_T = \nabla^2 \psi'_T . \tag{18}$$

Consequently, the baroclinic vorticity and the temperature fields cannot change independently, and both act as a single field (Fig. 5, bottom box). Furthermore, time differencing of the geostrophic balance equation leads to a balance between changes in geostrophic and baroclinic vorticity. Since these vorticities are only affected by divergence and barotropic vorticity one obtains the following diagnostic relationship combining (8b) and (8d), where the tendencies (I) and (II) cancel:

$$\left(1 - \frac{1}{2\kappa^2}\nabla^2\right)D'_T = -\frac{2}{f}U_T\frac{\partial \nabla^2 \psi'_M}{\partial x} . \tag{19}$$

This is the well-known Omega equation. Note that ω does not occur explicitly since it has been replaced by the baroclinic divergence using the continuity equation. Consequently, the divergence and the barotropic vorticity cannot change independently (Fig. 5, top box). Hence, the baroclinic instability in the QG model can be reduced to the interaction of the barotropic and baroclinic vorticity fields

$$\frac{\partial}{\partial t}\nabla^2 \psi'_M = -U_T\frac{\partial}{\partial x}\nabla^2 \psi'_T \tag{20}$$

$$\frac{\partial}{\partial t}\left(\nabla^2 \psi'_T - 2\kappa^2 \psi'_T\right) = -U_T\frac{\partial}{\partial x}\left(\nabla^2 \psi'_M + 2\kappa^2 \psi'_M\right) \tag{21}$$

The baroclinic instability mechanism in the QG model is illustrated in Fig. 5:

Tendencies (I), (II) and (IIIa) and (IIIb) (not shown). With geostrophic and baroclinic vorticity being geostrophically balanced, both the meridional advection of the basic state temperature by the barotropic vorticity field (I) and the zonal advection of barotropic vorticity by the mean flow (II) cancel one another without contributing to the joint baroclinic-geostrophic vorticity field. Since the divergent flow is determined by the Omega equation, the tendencies (3a and b), that is the pressure gradient and Coriolis force, need not be considered in the QG system.

Tendencies (IV) and (V). As both temperature and baroclinic vorticity fields are geostrophically linked in the QG system, it is the competition between the vertical temperature advection, $-f/2\kappa^2\nabla^2 D'_T$, and the vortex stretching, $-fD'_T$, which determines
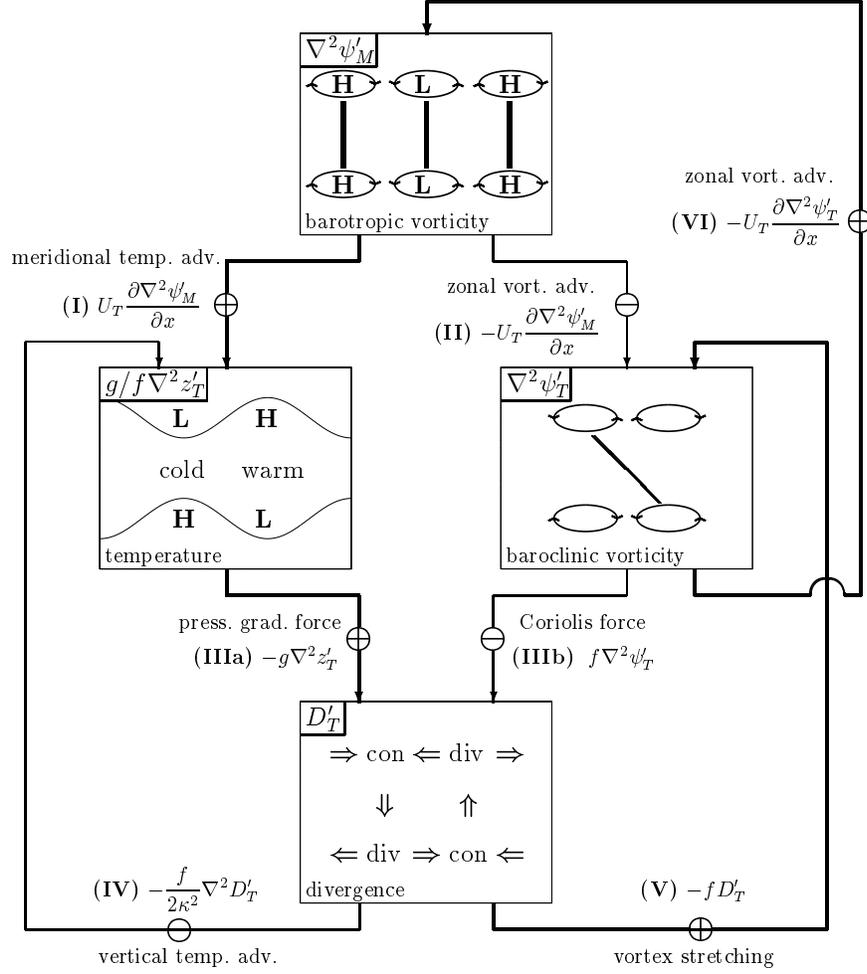
Figure 4. The baroclinic instability mechanism in a two-level PE model. The baroclinic wave is decomposed into the four components: barotropic vorticity $\nabla^2\psi'_M$, baroclinic vorticity $\nabla^2\psi'_T$, baroclinic geostrophic vorticity $g/f\nabla^2 z'_T$, and divergence D'_T. These components (represented as vertical-zonal cross-section in boxes) are linked by the seven tendencies: (I) meridional temperature advection, (II) zonal advection of barotropic vorticity, (IIIa) divergence of the pressure gradient force, (IIIb) divergence of the Coriolis force, (IV) vertical temperature advection, (V) vorticity stretching and (VI) zonal advection of baroclinic vorticity. The arrows indicate tendencies from the originating fields to the fields being effected. The $\oplus$ ($\ominus$) symbol indicates a tendency that supports (does not support) the baroclinic wave structure in the unstable case (Fig. 3).

whether baroclinic instability occurs. Thus instability occurs for waves with a wavenumber smaller than $(2\kappa^2)^{\frac{1}{2}}$ or, which is equivalent, if vortex stretching dominates over vertical temperature advection. For larger wavenumbers, these tendencies do not support sinking cold and rising warm air (divergence and temperature fields) and a westward tilted baroclinic vorticity, so that upper and lower levels become uncoupled.

Tendency (VI). The horizontal advection of the baroclinic vorticity by the basic flow, $-U_T\partial\nabla^2\psi'_T/\partial x$, supports the barotropic vorticity field by turning the westward tilted trough axis of the baroclinic vorticity field vertically, which closes the positive feedback
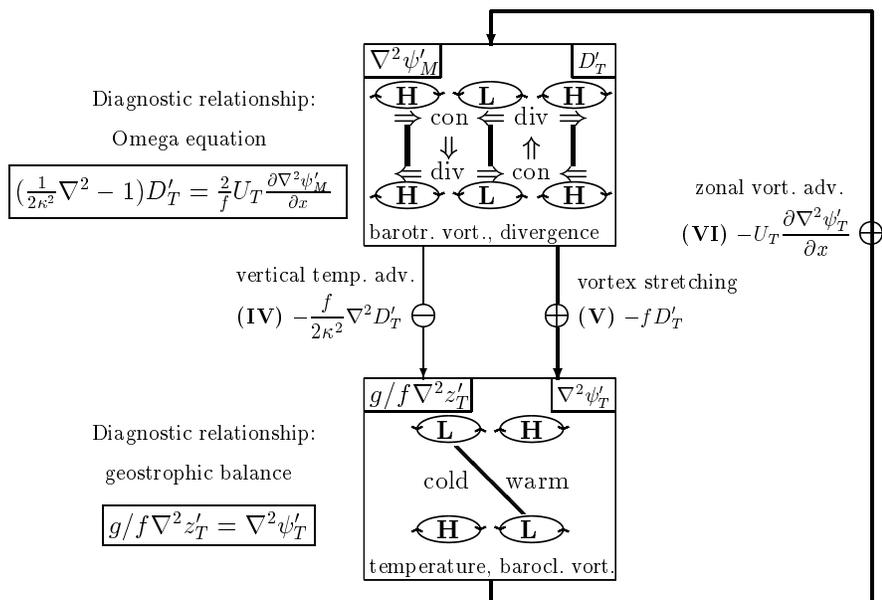
loop of the QG-system.

Figure 5. The baroclinic instability mechanism in a two-level QG model. The equation for the divergence is replaced by the diagnostic Omega equation balancing the barotropic vorticity and divergence fields (upper box; see top and bottom boxes of Figure 4); baroclinic vorticity is in geostrophic balance with the temperature (lower box; see left and right boxes of Figure 4). The tendencies (I) and (II), that is the meridional temperature and the zonal vorticity advection, as well as (IIIa) and (IIIb), that is the pressure gradient and the Coriolis force, cancel so that vortex stretching (V) and zonal vorticity advection (VI) comprise the positive feedback loop.

A final comment on the explanations of QG two-level instability and stability is in order. Holton's (1992) textbook has a paragraph and Bretherton (1966) uses potential vorticity arguments. To these two-level QG based explanations of baroclinic instability we add (i) the embedding in a PE environment, where tendencies replace the diagnostic Omega-equation (generating the secondary circulation to induce vortex stretching) and (ii) a visualisation of the meteorological fields (which interact through tendencies to generate a positive feedback loop). But two-level models do not capture short wave instabilities that occur when an interior potential vorticity gradient is present (White and Bell 1988); the general mechanism of QG baroclinic instability in continuous or multi-level models has been discussed by many authors (Green 1960, McIntyre 1970, Hide 1970).

5. CONCLUSION

Although the primitive equations (PE) are being used as the dynamical core of numerical weather prediction and climate models, present day textbooks describe the most important mid-latitude mechanism of baroclinic instability in terms of quasi-geostrophic (QG) theory. This is due to its simplicity which, utilising the two-level version, leads to a dynamical system of two linear equations describing the evolution of the barotropic and baroclinic vorticity field. However, an associated PE model of minimum complexity

requires only two further fields, divergence and temperature. This allows an immediate description of the instability mechanism in terms of the tendencies driving the evolution of the four fields. Thus the instability mechanism including short-wave cut-off can be illustrated as a positive feedback loop following the sequence of tendencies. The QG version of the instability mechanism emerges directly from the PE set, because divergence and temperature are diagnostically (and not prognostically) linked to the barotropic and baroclinic vorticity by the omega equation and geostrophy, respectively.

In this sense the baroclinic instability mechanism is presented here as a positive feedback loop in a linear dynamical system spanned by four variables representing atmospheric fields. This feedback loop consists of a succession of tendencies changing the field amplitudes, illustrates the underlying physical processes involved, and reveals the influence of the QG assumption on the more complete PE dynamics.

<div style="text-align:center">

APPENDIX A

Meridional boundaries, vorticity tilting and momentum flux

</div>

The solutions obtained in section 3 are y-independent, although the meridional extent of observed synoptic-scale disturbances is limited. Meridionally bounded solutions of the system (9) can only be obtained by introducing artificial sidewalls limiting the baroclinic zone. This effect is examined in the following.

The sidewalls may lie at $\tilde{y} = 0$ and $\tilde{y} = \pi/\tilde{l}$, where $\tilde{l}$ can be interpreted as the nondimensional minimum meridional wavenumber of a mode inside of the channel. Zero meridional velocity at these boundaries leads to

$$\frac{\partial \Psi}{\partial \tilde{x}} = \frac{\partial \Theta}{\partial \tilde{x}} = 0 \quad \text{at } \tilde{y} = 0, \ \frac{\pi}{\tilde{l}} \ . \tag{A.1}$$

Following Lorenz (1963) the Eqs. (9) subjected to the boundary conditions (A.1) may be transformed into spectral form by inserting the Fourier expansion

$$(\Psi, \Theta, D, Z) = \sum_{n=1}^{\infty} (\hat{\Psi}_n, \hat{\Theta}_n, \hat{D}_n, \hat{Z}_n) \sin(n\tilde{l}\tilde{y}) \exp(i\tilde{k}\tilde{x} - i\tilde{\nu}\tilde{t}) \ . \tag{A.2}$$

Multiplying the resulting equations with $2\tilde{l}/\pi \sin m\tilde{l}\tilde{y}$ and integrating over the meridional domain from $\tilde{y} = 0$ to $\tilde{y} = \pi/\tilde{l}$ yields

$$i\tilde{\nu}(\tilde{k}^2 + m^2\tilde{l}^2)\hat{\Psi}_m = i\tilde{k}(\tilde{k}^2 + m^2\tilde{l}^2)\hat{\Theta}_m + \frac{2\tilde{l}}{\pi}\sum_{n=1}^{\infty} \frac{(1-(-1)^{m+n})mn}{m^2 - n^2}\hat{D}_n \ , \tag{A.3}$$

$$i\tilde{\nu}(\tilde{k}^2 + m^2\tilde{l}^2)\hat{\Theta}_m = i\tilde{k}(\tilde{k}^2 + m^2\tilde{l}^2)\hat{\Psi}_m - \frac{1}{\text{Ro}}\hat{D}_m \ , \tag{A.4}$$

$$i\tilde{\nu}\hat{D}_m = \frac{1}{\text{Ro}}(\tilde{k}^2 + m^2\tilde{l}^2)\hat{\Theta}_m - \frac{1}{\text{Ro}}(\tilde{k}^2 + m^2\tilde{l}^2)\hat{Z}_m \ , \tag{A.5}$$

$$i\tilde{\nu}\hat{Z}_m = -i\tilde{k}\hat{\Psi}_m + \frac{1}{\text{Ro}}\hat{D}_m \ . \tag{A.6}$$

Generally, it is not possible to obtain analytic expressions for the eigenvalues since an infinite number of spectral coefficients is incorporated in the vorticity tilting term (in A.3).

First, the vorticity tilting term is neglected. Then the coupling between different meridional wavenumbers vanishes and the growth rate of the fastest growing mode (in the case of instability) takes the form

$$\tilde{\sigma} = \left(\left(\left(\frac{1}{4} \left(\frac{1}{\text{Ro}^2} (1 + \mu^2) + \tilde{k}^2 \right)^2 + \frac{\tilde{k}^2}{\text{Ro}^2} (1 - \mu^2) \right)^{\frac{1}{2}} - \frac{1}{2} \left(\frac{1}{\text{Ro}^2} (1 + \mu^2) + \tilde{k}^2 \right) \right)^{\frac{1}{2}} , \quad (A.7)$$

where $\mu^2 = \tilde{k}^2 + \tilde{l}^2$. Applying the quasigeostrophic approximation the growth rate simplifies to

$$\tilde{\sigma} = \tilde{k} \left(\frac{1 - \mu^2}{1 + \mu^2} \right)^{\frac{1}{2}} \qquad (A.8)$$

The fastest growing normal mode consists of monopole streamfunction cells which act to reduce the meridional temperature gradient by a northward temperature flux. The meridional momentum flux of this normal mode vanishes, because the phase of the streamfunction wave is meridionally independent.

Secondly, if the vorticity tilting term $\partial D / \partial \tilde{y}$ is not neglected, a meridional momentum flux appears. This can be understood by the fact that the major part of the divergence wave is shifted by a 90 degree westward-phase lag with respect to the barotropic streamfunction wave. The differentiation with respect to $\tilde{y}$ introduces a dipole pattern, which tilts the barotropic streamfunction eddies toward NW-SE direction. The consequence is a southward momentum flux.
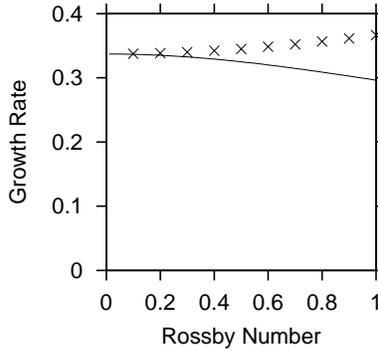

Figure A.1. Growth rate of the fastest growing normal mode as a function of Rossby number for $\tilde{k} = 0.6$ and $\tilde{l} = 0.4$. The crosses indicate the values obtained from a numerical analysis of the system (A.3)-(A.6) and the solid line displays the analytical result (A.7).

This can be substantiated by a numerical analysis of the spectral equations (A.3)-(A.6). For this analysis the Fourier expansion is truncated at $M = 100$ where the meridional and the zonal wavenumber are assumed to be $\tilde{l} = 0.4$ and $\tilde{k} = 0.6$, respectively. The value for $\tilde{k}$ is roughly associated with the maximal growth rate in the QG model and the aspect ratio $a = \tilde{k}/\tilde{l} = 1.5$ is reasonable for extratropical synoptic eddies (e. g. James 1994). The eigenvalues have been calculated by using standard routines of the Fortran NAG library. In Fig. A-1 the growth rate of the fastest growing normal mode is plotted against Rossby number Ro. For comparison the figure shows also the analytical result (A.7) which has been achieved by ignoring the vorticity tilting term. There is a close agreement between both growth rates when the Rossby number is low (Ro < 0.1). For larger Rossby numbers numerically calculated growth rates increase with increasing Rossby number while the analytically determined values decrease. Consequently, the
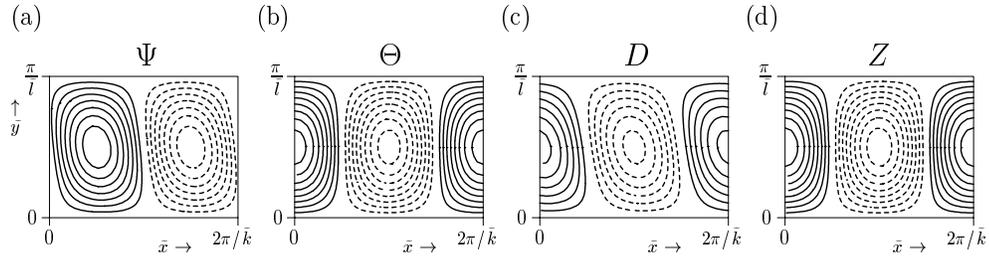
(a)   (b)   (c)   (d)

Figure A.2.   Horizontal fields of the fastest growing normal mode for Ro=0.2: (a) barotropic streamfunction (contour interval 0.1), (b) baroclinic streamfunction (contour interval 0.05), (c) divergence (contour interval 0.01), (d) thickness (contour interval 0.05). Negative contours are dashed.

vorticity tilting term acts to destabilize the basic flow further. The horizontal fields of the fastest growing normal mode for Ro=0.2 are displayed in Fig. A.2. A clear tilt of the barotropic streamfunction eddies towards the NW-SE direction appears indicating a southward momentum flux. The divergence field attains a similar structure while the baroclinic streamfunction and the temperature anomalies do not have a substantial tilt. There is an almost linear increase of the barotropic streamfunction tilt with increasing Rossby number (not shown). This is, because the tilting term scaled by Ro (see Eq. 16c) is a leading order correction to the QG model.

The introduction of a momentum flux may be important for the nonlinear behaviour of the baroclinic wave. When nonlinearities become large in the course of baroclinic wave development, the southward momentum flux of the disturbance generates a barotropic zonal mean flow with cyclonic shear. This is possibly one reason (besides other non-geostropic effects, see Snyder et al. 1991) why cyclones usually dominate the behaviour in the occluding stage of a baroclinic wave life cycle.

References

Árnason, G.	1963	The stability of nongeostrophic perturbations in a baroclinic zonal flow. *Tellus*, **15**, 205–229
Bell, M. J. and White, A. A.	1988	Spurious stability and instability in *N*-level quasi-geostrophic models. *J. Atmos. Sci.*, **45**, 1731–1738
Bretherton, F. P.	1966	Baroclinic instability and the short wavelength cut-off in terms of potential vorticity. *Q. J. R. Meteorol. Soc.*, **92**, 335–345
Charney, J. G., Fjørtoft, R. and von Neumann, J.	1950	Numerical integration of the barotropic vorticity equation. *Tellus*, **2**, 237–254
Charney, J. G.	1947	The dynamics of long waves in a baroclinic westerly current. *J. Meteor.*, **5**, 135–162
Charney, J. G. and Phillips, N. A.	1953	Numerical integration of the quasi-geostrophic equations for barotropic and simple baroclinic flows. *J. Meteor.*, **10**, 71–99
Eady, E. T.	1949	Long waves and cyclone waves. *Tellus*, **1**, 33–52
Green, J. S. A.	1960	A problem in baroclinic stability. *Q. J. R. Meteorol. Soc.*, **86**, 237–251

Hide, R.	1970	Some laboratory experiments on free thermal convection in a rotating fluid subject to a horizontal temperature gradient and their relation to the theory of the global atmospheric circulation (Appendix D), *The global circulation of the atmosphere* (ed. G. A. Corby), London; Royal Meteorological Society, 196–221
Holton J. R.	1992	*An introduction to dynamic meteorology*, 3rd edition, Academic Press, San Diego, USA
James I. N.	1994	*Introduction to circulating atmospheres*, Cambridge University Press, Cambridge, UK
Lorenz, E. N.	1963	The mechanics of vacillation. *J. Atmos. Sci.*, **20**, 448–464
McIntyre, M. E.	1970	On the non-separable baroclinic parallel flow instability problem. *J. Fluid Mech.*, **40**, 273–306
Phillips, N. A.	1951	A simple three-dimensional model for the study of large-scale extratropical flow patterns. *J. Meteor.*, **8**, 381–394
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
Snyder, C. H., Skamarock, W. C. and Rotunno, R.	1991	A comparison of primitive-equation and semigeostrophic simulations of baroclinic waves. *J. Atmos. Sci.*, **48**, 2179–2194
Stone, P. H.	1966	On non-geostrophic baroclinic instability. *J. Atmos. Sci.*, **23**, 390–400
Wiin-Nielsen, A.	1963	On baroclinic instability in filtered an non-filtered numerical prediction models. *Tellus*, **15**, 1–19