# Unstable Baroclinic Modes Damped by Ekman Dissipation

BRIAN F. FARRELL

*Department of Earth and Planetary Sciences, Harvard University, Cambridge, Massachusetts*

(Manuscript received 26 April 1988, in final form 17 August 1988)

ABSTRACT

Inclusion of Ekman damping in baroclinic models severely limits the range of unstable wavenumbers as well as the growth rate of the instabilities that remain. In contrast, there is much less reduction by the same dissipation of the transient growth of perturbations chosen to resemble those associated with observations of the initial stages of cyclogenesis. It is shown here that the Charney problem with Ekman dissipation included provides a realistic model of damped instability, that the growth rates of the unstable waves are small compared both with observed deepening rates and with deepening rates for initial value problems, and that vertical discretization is likely to produce spurious instabilities in damped models.

## 1. Introduction

Motions in the atmosphere occur on a wide variety of time and space scales ranging from the planetary stationary and retrograde Rossby modes with characteristic times of many days, to the dissipating turbulent structures with scales of millimeters and seconds. It is generally accepted that the Navier–Stokes equations properly modified to include the effects of rotation and stratification provide a valid theoretical framework to account for the complete spectrum of these motions (assuming that external forcings are accounted for). Unfortunately, the generality of these equations discourages their direct use to obtain theoretical understanding of specific phenomena such as the process of cyclogenesis because a myriad of presumably extraneous processes are inextricably linked with the object of study. Experience has shown that progress is possible if an approximation to the full equations is made such that the motions of interest are accurately described and unwanted phenomena eliminated. A highly successful example of this program is the quasi-geostrophic approximation used by Charney (1947) and Eady (1949) in studies of baroclinic waves in midlatitudes. They recognized that a significant subset of phenomena of synoptic scale is associated with the evolution of the vertical component of vorticity and that a simple principle, the conservation following the geostrophic wind of the quasi-geostrophic potential vorticity, adequately described these dynamics. It is the nature of this program for gaining theoretical insight that many processes are left out, in this case including an accurate description of frontogenesis.

*Corresponding author address:* Dr. Brian Farrell, Dept. of Earth and Planetary Sciences, Harvard University, Pierce Hall, 29 Oxford Street, Cambridge, MA 02138.



Among the phenomena supported by this quasi-geostrophic dynamics are the baroclinic waves referred to as midlatitude cyclones. It is generally accepted that the origin of these waves is also properly described by the theory—they arise from the conversion by wave motion of available potential energy associated with the horizontal temperature gradient and through the thermal wind relation with the vertical shear of the jet. A distinction can be made between two mechanisms for development. The first involves the growth of unstable normal modes as in the original theory (Charney 1947; Eady 1949) and the second relies on transient dynamics involving an interaction of the modes and the continuous spectrum with the mean flow (Farrell 1985).

Because quasi-geostrophy supports only a subset of the full dynamics, the effects of motions not in this subset on the included waves cannot be explicitly accounted for by the theory but instead must be parameterized in some way. An example of such a parameterized process is dissipation by eddy viscosity. It was assumed by Charney and Eliassen (1949) that the interior of the troposphere could be regarded as inviscid on time scales appropriate to planetary-scale wave dynamics and that the primary dissipative mechanism is associated with transport of momentum by small-scale eddies between the surface and a region near the ground identified with the Ekman layer. An important advantage of this parameterization of dissipation is that it leaves the inviscid dynamics of the interior intact and accounts for damping with a simple modification of the quasi-geostrophic boundary condition. This parameterization has been widely accepted as appropriate at the level of approximation of quasi-geostrophic dynamics applied to the synoptic scale. Here we show that the Charney model, for which analytic solution is possible, captures the main qualitative and quantitative effects of Ekman damping, that the growth of unstable waves is slow in comparison to observed deepening rates, and that failure to resolve the increasingly singular steering level as damping increases can result in qualitative and quantitative errors in numerical models.

## 2. The Charney model with Ekman damping

The Charney model (Charney 1947) is the most realistic of the canonical examples of baroclinic instability, as it includes the effects of density stratification and the first-order influence of spherical geometry which is the linearized variation of the Coriolis parameter. In this model the density scale height is constant:

$$H^{-1} = \frac{1}{\rho_s}\frac{\partial \rho_s}{\partial z},$$

the shear of the wind is constant $U(z) = \Lambda z$, the Coriolis parameter is $f = f_0 + \beta y$, with $U$ being the zonal wind in the $x$ direction, $y$ the meridional coordinate, and $z$ measured vertically upward.

The nondimensional perturbation quasi-geostrophic potential vorticity equation in the scaled streamfunction:

$$\Phi = \psi(\tilde{z}, \tilde{t})e^{\tilde{z}/2}e^{i(\tilde{k}\tilde{x}+\tilde{l}\tilde{y})}$$

is (Pedlosky 1987)

$$\left(\frac{\partial}{\partial \tilde{t}} + i\tilde{k}\tilde{z}\right)\left[\psi_{\tilde{z}\tilde{z}} - \left(\tilde{\alpha}^2 + \frac{1}{4}\right)\psi\right] + i\tilde{k}(\tilde{\beta} + 1)\psi = 0, \tag{1}$$

with boundary conditions given by vertical velocity induced through Ekman convergence at the lower boundary and boundedness aloft:

$$\frac{\partial}{\partial \tilde{t}}\left(\psi_{\tilde{z}} + \frac{\psi}{2}\right) - i\tilde{k}(1 - \Gamma)\psi = 0, \quad \tilde{z} = 0 \tag{2a}$$

$$\psi \text{ bounded}, \quad \tilde{z} \rightarrow \infty. \tag{2b}$$

The nondimensional quantities are

$$\tilde{t} = t\Lambda\sqrt{\epsilon}$$

$$\tilde{k} = \frac{kH}{\sqrt{\epsilon}}$$

$$\tilde{z} = \frac{z}{H},$$

where $\tilde{\alpha} = \sqrt{\tilde{k}^2 + \tilde{l}^2}$ is the total horizontal wavenumber, and $\epsilon = f_0{}^2/N^2$ is the square ratio of the Coriolis parameter to the Brunt–Väisälä frequency. The problem is characterized by the nondimensional beta parameter $\tilde{\beta} = \beta H/\Lambda\epsilon$ and the Ekman parameter:

$$\Gamma = \frac{iN}{\Lambda H}\left(\frac{\nu}{2f_0}\right)^{1/2}\frac{\tilde{\alpha}^2}{\tilde{k}},$$

where $\nu$ is the vertical eddy viscosity coefficient. Tildes are omitted in sequel.

With the assumption of the ansatz $\psi = \psi(z)e^{-ikct}$, (1) and (2) comprise a boundary eigenvalue problem for the complex phase speed $c$. This problem can be solved exactly by transforming variables so that (1) assumes a standard form of the confluent hypergeometric equation, and iterating on the transformed boundary conditions (2) to obtain the eigenvalue. This solution is well known in the case of the undamped problem (Pedlosky 1987; Kuo 1979) and is easily extended to include Ekman damping (Card and Barcilon 1982).

Values of parameters are chosen to model the midlatitude troposphere: $f_0 = 10^{-4}\ \text{s}^{-1}$, $N = 10^{-2}\ \text{s}^{-1}$, $H = 10$ km, $\Lambda = 3\ \text{m s}^{-1}\ \text{km}^{-1}$, and $\beta = 1.6 \times 10^{-11}\ \text{m}^{-1}\ \text{s}^{-1}$. This results in $\beta = 0.53$. The meridional wavenumber $l = 2.0$ corresponds to a 3100 km wavelength, typical of midlatitude cyclogenesis and one unit of nondimensional time is 9.3 h. The dispersion relation is shown in Fig. 1 both for the undamped case and with damping corresponding to $\nu = 2.5\ \text{m}^2\ \text{s}^{-1}$.

Inspection of Fig. 1 reveals two remarkable influences of Ekman damping on baroclinic instability in this model problem. The first is the rapid decrease in the growth rate of the unstable wave with increasing vertical diffusion. At $\nu = 2.5\ \text{m}^2\ \text{s}^{-1}$ only 28% of the undamped instability remains and at $\nu \approx 4.5\ \text{m}^2\ \text{s}^{-1}$ it disappears entirely. Even the larger of these values is modest compared with those typical of midlatitudes leading to the conclusion that baroclinic instability may be strongly inhibited by Ekman damping (Card and Barcilon 1982; Farrell 1985). This is examined further by obtaining the growth rate of the unstable mode with $k = l = 2.0$ as a function of

$$\Gamma_0 = \frac{N}{\Lambda H}\left(\frac{\nu}{2f_0}\right)^{1/2}.$$

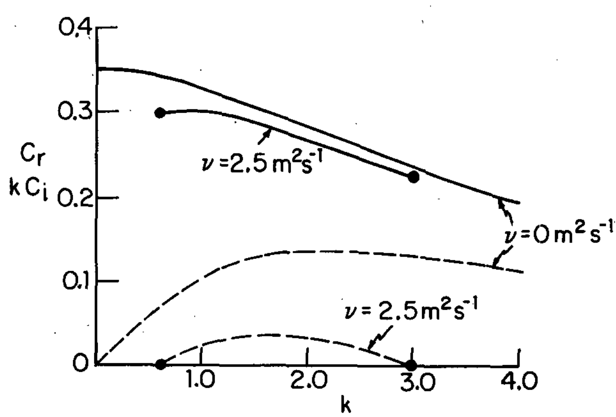


FIG. 1. Dispersion relation for the Charney model with $\beta = 0.53$ corresponding to midlatitude parameter values and wavenumber $l = 2.0$ for which the dimensional wavelength is 3100 km. Phase speeds, $c_r$, are indicated with solid lines, and growth rates $kc_i$ with dashed lines as a function of the zonal wavenumber, $k$, for the undamped case $\nu = 0$ and for a modest value of the vertical diffusion coefficient, $\nu = 2.5\ \text{m}^2\ \text{s}^{-1}$. (Adapted from Farrell 1985).

The result is shown by the solid line in Fig. 2; the unstable mode is seen to disappear entirely at a value of $\Gamma_0$ corresponding to $\nu = 4.5\ \mathrm{m^2\,s^{-1}}$.

Recently, Valdes and Hoskins (1988) examined the effect of Ekman and alternative formulations of surface drag on baroclinic instability of the observed zonally averaged winter and summer flow in a hemispheric primitive equation model. They found maximum instability for $\nu = 10\ \mathrm{m^2\,s^{-1}}$ at wavenumber 7 corresponding at 45° lat to a wavelength of 4040 km and state that the reduction in growth rate for the more realistic flow is much less marked than suggested by the Charney model. In order to test this the calculation of growth rate as a function of $\Gamma_0$ was repeated using $k = 1 = 1.5$ corresponding to a wavelength of 4200 km. The results are shown in Fig. 3. Taking account of the increase in jet strength in their model from 30 to 40 m $\mathrm{s^{-1}}$, the Ekman parameter that results from $\nu = 10\ \mathrm{m^2\,s^{-1}}$ is $\Gamma_0 = 0.056$. Referring to Fig. 3, the reduction in growth rate in the Charney model due to the inclusion of this much dissipation is from $kc_i = 0.170$ to $kc_i = 0.059$. Expressed as dimensional growth times, the Charney model gives 1.7 days per $e$-fold for the undamped and 4.9 days for the damped while the hemispheric primitive equation model gives 2.4 days and 5.6 days, respectively. Considering the simplicity of the Charney model this would appear to be rather good agreement.

It is clear that there are instabilities in the parameter space of the damped Charney model (Card and Barcilon 1982; Farrell 1985); it is also clear that there are instabilities in the parameter space of more realistic models as affirmed by Valdes and Hoskins (1988). What is at issue is the physical significance of long wavelength modes with time scales for growth of the order of a week in the presence of transient development on much shorter time scales (Farrell 1985; Valdes and Hoskins 1988). Further consideration of spatial and temporal scales of observed baroclinic development compared to the prediction of normal mode instability is deferred to section 4.

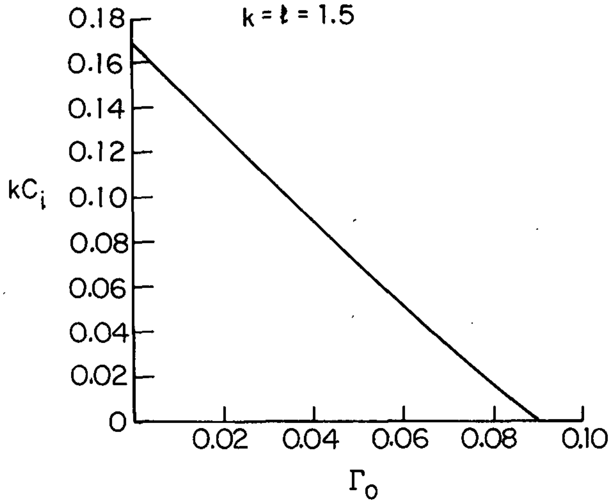


FIG. 3. Growth rate from the analytic solution of the $k = 1 = 1.5$ mode as a function of the Ekman parameter $\Gamma_0$.

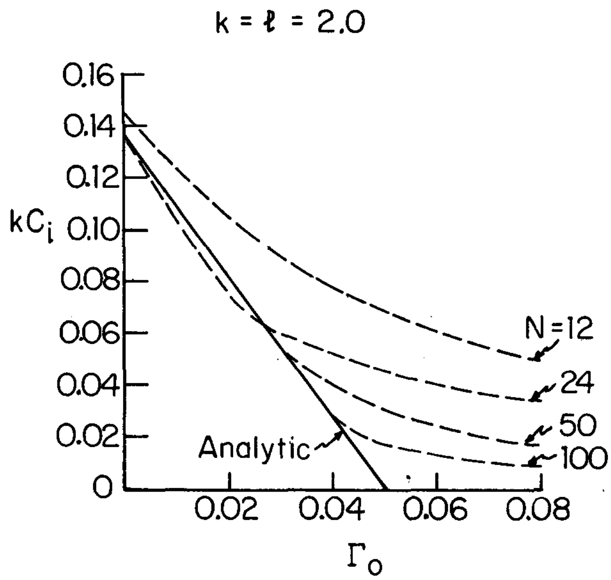


FIG. 2. Growth rate of the $k = 1 = 2.0$ mode as a function of the Ekman parameter $\Gamma_0$. The exact solution follows the solid line. Finite difference approximations to the dispersion relation are shown with dashed lines for the indicated number of collocation points: $\Gamma_0 = 0.05$ corresponds to $\nu = 4.5\ \mathrm{m^2\,s^{-1}}$ for the midlatitude parameter values used.

The second remarkable aspect of Fig. 1 is that the phase speed of the damped wave is little affected by the value of $\Gamma_0$, so that while the imaginary part of $c$ is approaching zero with increasing $\Gamma_0$, the real part of $c$ remains well inside the flow. This causes (1) to become singular at the steering level where the phase speed of the wave equals the zonal wind speed and the first term in parentheses vanishes. There are no analogous singularities in the undamped model because all steering levels are associated with nonzero imaginary part of the phase speed, as required by the critical layer instability argument of Bretherton (1966).

## 3. Numerical solution of the Ekman damped Charney model

Analytic solutions are available for only a restricted class of problems in baroclinic instability and it is therefore necessary to resort to numerical models if even modest increases in realism are sought. For this reason most studies of instability have made use of either finite difference or Galerkin methods to obtain the dispersion relation. In the former method, and assuming the ansatz $\psi = \psi(z)e^{-ikct}$, Eq. (1) becomes

$$(z - c)\left[\psi_{zz} - \left(\alpha^2 + \frac{1}{4}\right)\psi\right] + (\beta + 1)\psi = 0, \quad \text{(3a)}$$

the lower boundary is assumed to support an Ekman layer:

$$c\left[\psi_z + \frac{\psi}{2}\right] + (1 - \Gamma)\psi = 0, \quad z = 0, \qquad (3b)$$

while the upper boundary is free:

$$(z - c)\left[\psi_z + \frac{\psi}{2}\right] - \psi = 0, \quad z = z_t. \qquad (3c)$$

With centered difference expressions for the derivatives, (3) is interpreted as a discrete eigenvalue problem for which the unstable modes correspond to eigenvalues with $c_i > 0$, (see Appendix). This standard method for obtaining approximate solutions to such problems has been extensively applied to baroclinic models. Green (1960) pointed out in an early paper the importance of maintaining sufficient resolution when solving for modes with small growth rates.

Because there is an interior singularity forming as $\Gamma_0$ increases, the resolution requirements for accurate assessment of stabilization by Ekman damping of a baroclinic instability are likely to be severe. Advantage is taken of the exact solution of the Charney problem to provide a comparison with the approximate solutions obtained with resolution varying from 100 to 12 collocation points on the interval $0.0 \leqslant z \leqslant 2.0$. The use of a lid for the upper boundary does not affect the solution at this wavenumber.

Results of this experiment are shown by the dashed lines in Fig. 2. Notice that even the least resolved of these verifies the undamped instability to good accuracy. However, as the damping is increased and the singularity develops, the numerical approximations fail in the worst way by maintaining a fictitious instability.

The effect of vertical resolution on the growth rate of unstable waves has attracted attention recently, beginning with Staley (1986) who found spurious short wave instabilities and slow convergence with increasing number of vertical grid points in a matrix eigenvalue study of mean flows with varying shears and static stabilities. This problem was taken up by Arakawa and Moorthi (1988) and Bell and White (1988), who attribute the difficulties encountered in solving the eigenproblem at short wavelength to the inability of the vertical discretization to properly resolve the singularity that develops as growth rate decreases and $c_i \rightarrow 0$ with decreasing wavelength. All of these authors find (as is also found here) that in the undamped problem the maximum growth rate modes are well represented even by a very small number of grid points because of the large $c_i$ of these modes which makes them very smooth. The origin of the unresolved singularity in their cases can be traced to the decrease in growth rate with increasing wavenumber and therefore only the shortwaves are affected. It has been shown here that inclusion of modest Ekman damping decreases the growth rate of the longer most unstable waves also—to the point that very similar problems occur, including spurious and exaggerated instabilities.

## 4. Discussion and conclusion

The error revealed by the curves in Fig. 2 is consistent with a lack of resolution near the critical level. The nondimensionalization allows increments of resolution in height to be compared directly to increments in phase speed, and from Fig. 2 an approximate resolution requirement of one grid point per $2c_i$ is indicated. While the error decreases with increasing resolution it does not disappear even with 100 points. If the model is being used to do a forecast, it would suffer an *e*-fold increase of the unphysical mode in $t \approx 2/k\delta$ units of nondimensional time, where $\delta$ is the grid point spacing. This may be tolerable if the forecast is short or if observations are continually being folded in to prevent the unlimited growth of the erroneous mode. However, if it were being used as a climate model, to obtain statistics or to assess the stability or instability of a flow, then the result could be qualitatively in error because the growth rate of a subset of wavenumbers is spurious in the linear limit.

A useful distinction is drawn by Arakawa and Moorthi (1988) between computational instability in the sense in which it is ordinarily understood and these spurious and exaggerated modes that are properly interpreted as the false realizations of physical instability. Unfortunately, while true computational instability is likely to be obvious because it is not related to plausible phenomena, these instabilities have the structural and behavioral aspects of the real thing and are more likely to be accepted by the investigator as physical (even if he realizes that his resolution is not fine enough to lend them credence).

One can have assurance that in the linear problem, the asymptotic limit $t \rightarrow \infty$ is dominated by the most rapidly growing instability and argue that so long as at least that one remains unstable in the presence of damping the nonlinear flow might be unaffected by the existence of other spurious instabilities. That this is not likely to be true is demonstrated by an experiment using a nonlinear primitive equation model both with and without employing a method to suppress spurious linear instabilities (Arakawa and Moorthi 1988). They term the impact of the spurious instabilities on the initial spinup "drastic" (see their Figs. 21 and 22). It remains to be seen if a fully developed turbulent flow would be affected and this cannot be determined from their 15-day integrations. It is clear that a flow with only stable modes and no stochastic forcing would have very different statistics compared with a similar flow supporting computational modes. Longer-term integrations to explore these issues are of great interest.

There is a tendency in numerical models to sacrifice vertical resolution in favor of horizontal in an attempt to reduce phase speed errors. The implication of these results is that the consequences of failure to resolve in the vertical can be that the model is dominated by unphysical instabilities when surface drag is included.

Generally, it is extremely difficult to analyze the exact correspondence between a differential equation and a numerical method for obtaining an approximate solution (e.g., Arnason et al. 1967). Error in the example above is consistent with failure to resolve the eigenfunction singularity and can be regarded as indicative of what is apt to happen under such circumstances (Bell and White 1988; Arakawa and Moorthi 1988).

Finally, there remains the question of the physical role of the weakly unstable long waves. Cyclogenesis and energetic baroclinic events occur on a broad range of space scales including polar lows with wavelengths of 1000–1500 km (Reed 1979) and midlatitude cyclones in the 2500–4000 km range (Sanders and Gyakum 1980). Sanders and Gyakum take as typical of rapidly deepening cyclones an example in which wavelength shortened from 3400 to 2200 km over the course of development. The unstable modes of Valdes and Hoskins (1988) lie at the upper extreme of the cyclone wavelength range but the real difficulty lies in the time scales associated with their growth. It is well known that the typical period of deepening of observed cyclones is between 12 and 48 h. Sanders (1986) obtains 24–36 h for rapidly intensifying lows while Roebber (1984) finds 24 h for all lows in a 1-yr period of NMC surface analyses, covering about one-third of the Northern Hemisphere and 45 h for explosive deepeners. Recalling that the *e*-fold time for the most unstable wave found by Valdes and Hoskins (1988) is 133 h, the unstable wave would, if started off with perfect normal mode form, deepen by a factor of 1.3 in 36 h, sufficient to drop a 5 mb depression to 6.5 mb. It is suggested that the initial value problem with time scale of precisely the observed 24–48 h and with much more rapid growth rates (Farrell 1985; Valdes and Hoskins 1988) is more closely in accord with cyclogenesis as observed.

*Acknowledgments.* This work was supported by NSF Grant ATM-8712995.

## APPENDIX

### Numerical Method

The equation and boundary conditions (3) are expressed at $N$ interior levels between the surfaces at $z = 0$ and $z = z_t$. The points are equally spaced at intervals $\delta \equiv z_t/(N-1)$ so that $z_n = (n-1)\delta$. In addition there are fictitious exterior points at $z_0 = -\delta$ and $z_{N+1} = z_t + \delta$. Using centered differences, (3) becomes

$$(z_n - c)\left\{\psi_{n+1} - \left[2 + \left(\alpha^2 + \frac{1}{4}\right)\delta^2\right]\psi_n + \psi_{n-1}\right\} + (\beta + 1)\delta^2\psi_n = 0, \quad \text{(A3a)}$$

$$c[\psi_{n+1} + \delta\psi_n - \psi_{n-1}] + 2(1 - \Gamma)\delta\psi_n = 0, \quad n = 1, \quad \text{(A3b)}$$

$$(z_n - c)[\psi_{n+1} + \delta\psi_n - \psi_{n-1}] - 2\delta\psi_n = 0, \quad n = N. \quad \text{(A3c)}$$

The exterior points are eliminated between (A3a) and (A3b, c) leaving a well-posed eigenvalue problem for the eigenvalue, $c$. This method converges rapidly with increasing $N$ for smooth eigenfunctions associated with eigenvalues with large imaginary part of the phase speed such as are found near $\Gamma_0 = 0$. For example, normalized error in growth rate for the undamped mode in Fig. 2 decreases from $6.1 \times 10^{-2}$ for $N = 12$ to $4.4 \times 10^{-5}$ for $N = 100$. However, the unresolved near singularity at the interior steering level destroys this rapid convergence for larger $\Gamma_0$, as can also be seen in Fig. 2.

## REFERENCES


Arakawa, A., and S. Moorthi, 1988: Baroclinic instability in vertically discrete systems. *J. Atmos. Sci.*, **45**, 1688–1707.

Arnason, G., P. S. Brown and E. A. Newburg, 1967: A case study of the validity of finite difference approximations in solving dynamic stability problems. *J. Atmos. Sci.*, **40**, 2075–2079.

Bell, M. J., and A. A. White, 1988: Spurious stability and instability in *N*-level quasi-geostrophic models. *J. Atmos. Sci.*, **45**, 1731–1738.

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

Card, P. A., and A. Barcilon, 1982: The Charney stability problem with a lower Ekman layer. *J. Atmos. Sci.*, **39**, 2128–2137.

Charney, J. G., 1947: The dynamics of long waves in a baroclinic westerly current. *J. Meteor.*, **4**, 135–162.

——, and A. Eliassen, 1949: A numerical method for predicting the perturbations of the middle latitude westerlies. *Tellus*, **1**, 38–54.

Farrell, B., 1985: Transient growth of damped baroclinic waves. *J. Atmos. Sci.*, **42**, 2718–2727.

Green, J. S. A., 1960: A problem in baroclinic stability. *Quart. J. Roy. Met. Soc.*, **86**, 237–251.

Kuo, H. L., 1979: Baroclinic instabilities of linear and jet profiles in the atmosphere. *J. Atmos. Sci.*, **36**, 2360–2378.

Pedlosky, J., 1987: *Geophysical Fluid Dynamics.* Springer–Verlag, 710 pp.

Reed, R. J., 1979: Cyclogenesis in polar air streams. *Mon. Wea. Rev.*, **B107**, 38–52.

Roebber, P. J., 1984: Statistical analysis and updated climatology of explosively cyclones. *Mon. Wea. Rev.*, **112**, 1577–1589.

Sanders, F., 1986: Explosive cyclogenesis in the West-Central North Atlantic Ocean, 1981–84. Part I: composite structure and mean behavior. *Mon. Wea. Rev.*, **114**, 1781–1794.

——, and J. R. Gyakum, 1980: Synoptic-dynamic climatology of the "bomb." *Mon. Wea. Rev.*, **108**, 1589–1606.

Staley, D. O., 1986: Baroclinic and barotropic instability spectra as functions of *N* in *N*-level models. *J. Atmos. Sci.*, **43**, 1817–1832.

Valdes, P. J., and B. J. Hoskins, 1988: Baroclinic instability of the zonally averaged flow with boundary layer damping. *J. Atmos. Sci.*, **45**, 1584–1593.