THEORIES OF CYCLOGENESIS

Brian Hoskins
Department of Meteorology, University of Reading,
Reading, U.K.

1. INTRODUCTION

Extra-tropical cyclogenesis has been the topic of numerous theoretical discussions in the last fifty years. In many ways the best modern perspective is provided by "potential vorticity thinking". However since this is a viewpoint to be addressed in other talks, the approach here will be a more classical one.

As our starting point we take the parcel energetics argument of Eady (1949) which was elaborated on by Green (1960). Consider the situation shown in Fig. 1 in which there is an atmosphere in thermal wind balance

$$N^2 = \frac{g}{\theta_o}\frac{\partial\bar{\theta}}{\partial z} \quad \text{and} \quad f\frac{d\bar{u}}{dz} = -\frac{g}{\theta_o}\frac{\partial\theta}{\partial y} \qquad \text{both constant.}$$

The slope of the isentropes $\alpha_\theta = -\frac{\partial\bar{\theta}}{\partial y} / \frac{\partial\bar{\theta}}{\partial z}$ is also constant. If we now assume that a ring of fluid (P) is displaced polewards along a trajectory sloping at an angle α then the reduction in basic state potential energy and thus the energy available for the perturbation may be shown to be (with the hydrostatic approximation)

$$\delta E \propto \frac{1}{2}\,\delta y^2\,\frac{\alpha(\alpha_\theta-\alpha)}{\alpha_\theta}\,f\,\frac{d\bar{u}}{dz}$$

This is positive for trajectory slopes between the horizontal and that of the basic isentropes. For such angles the parcel is warm compared with its surroundings and is rising. The maximum value is obtained for $\alpha=\alpha_\theta/2$ and is

$$\frac{1}{2}\,\delta y^2\,\frac{1}{4}\,\alpha_\theta\,f\,\frac{d\bar{u}}{dz}\,. \tag{1.1}$$

However a rapidly rotating atmosphere has angular momentum constraints which usually act against such a displacement. The Coriolis force will imply a westerly wind along the ring and a horizontal acceleration back towards its initial position. This will not occur only if there is inertial instability or neutrality in the chosen direction. The first direction in the "wedge of instability" for which this can occur is along the isentropes so that the condition for this direct conversion of energy by overturning is (in the Northern Hemisphere)

$$\zeta_\theta < 0 \qquad , \text{ where } \zeta_\theta = f + \frac{\partial v}{\partial x}\Big|_\theta - \frac{\partial u}{\partial y}\Big|_\theta ,$$

$$\text{or} \quad P < 0 \qquad \text{where the potential vorticity } P = \zeta_\theta \left(-\frac{1}{g}\frac{\partial \theta}{\partial p}\right) \tag{1.2}$$

or $n = fy - u$ surfaces more horizontal than θ surfaces.

(In the x direction the relevant "absolute momentum" surfaces would be $m = fx + v = \text{const}$).

The instability that occurs in this situation is usually referred to as symmetric (baroclinic) instability. The large-scale atmosphere does not in general satisfy these criteria and so release of potential energy by zonally symmetric overturning is inhibited by the angular momentum constraint. This constraint can be broken by allowing variation in the zonal (x) direction. Suppose now that some parcels are displaced polewards as P in Fig. 1 and at different longitudes other parcels are displaced southwards along similar sloping trajectories. In this sloping plane the motions will look like the continuous vectors shown in Fig. 2, with parcels curving in a clockwise sense. However this tendency for trajectories to curve can be opposed by pressure gradients corresponding to a simple pressure pattern of the form shown. The meridional motion will then be in balance between the Coriolis and pressure gradient forces i.e. there will be geostrophic balance.

A vertical section across the system must then look like the schematic shown in Fig. 3 with the warm air moving polewards and upwards east of the low pressure region and the cold air moving equatorwards and downwards to the west. From hydrostatic balance, the warm air being less dense, corresponds to a relatively small change in pressure with height. This implies a tendency to the low pressure at low levels to be shifted towards the warm air region whilst at upper levels it is shifted away. Thus the low (and high) pressure region tilts westwards with height.

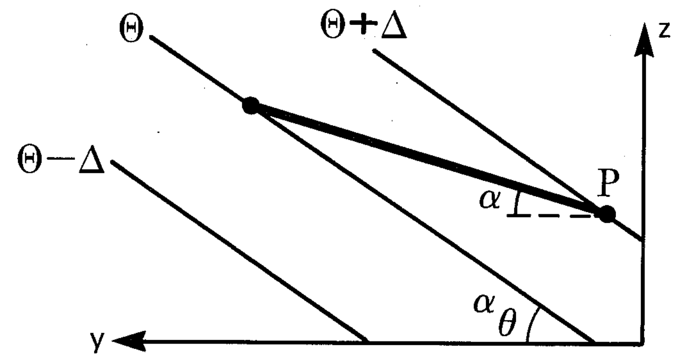

Fig. 1 Basic isentropes at an angle α_θ and a displacement at an angle α.

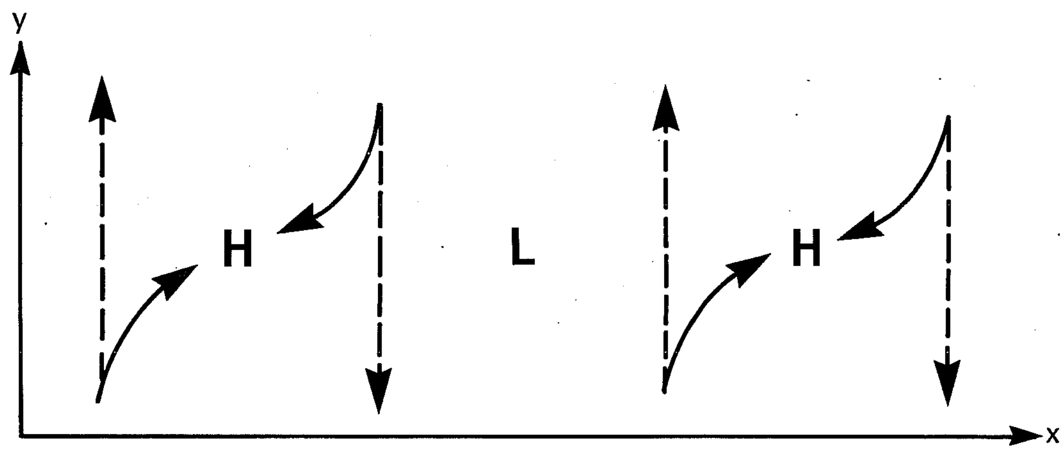

Fig. 2 The motions as seen in the plane of the trajectories. The continuous vectors are those that would occur in the absence of the pressure field indicated by L and H, and the dashed vectors are the actual trajectories.

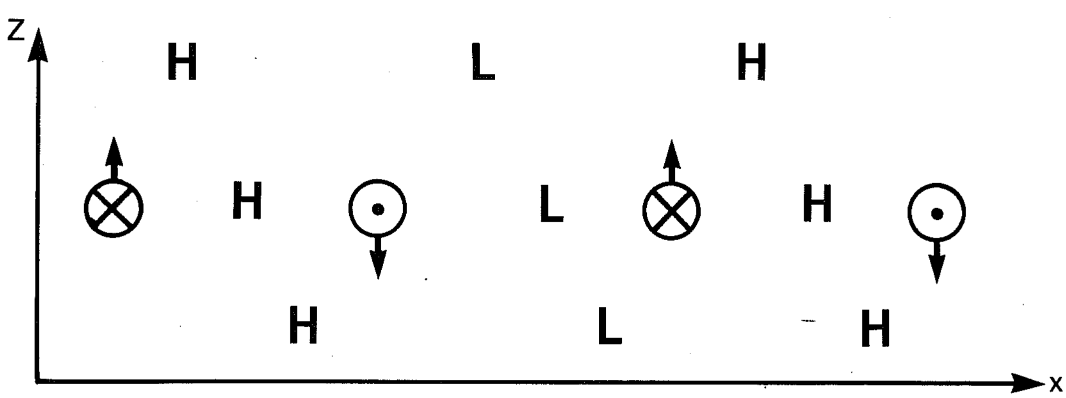

Fig. 3 A longitude-height section showing poleward (⊗), equatorward (⊙) and vertical motion, and the longitudinal variation of pressure for a wave that converts potential to kinetic energy.

From the energetics argument above we can even obtain an idea of maximum possible growth rates by assuming that all the energy can be realised by the perturbation in the kinetic energy of the N-S motion. From (1.1),

$$\frac{1}{2}\,(\delta\dot{y})^2_{max} = \frac{1}{2}\,\delta y^2\,(\alpha_\theta\,\frac{1}{4}\,f\,\frac{d\bar{u}}{dz})$$

$$= \frac{1}{2}\,\delta y^2\,(\frac{1}{2}\,f\,\frac{d\bar{u}}{dz}\,/N)^2$$

Thus for such a structure,

$$\sigma_{max} = 0.5\,f\,\frac{d\bar{u}}{dz}\,/N\,. \qquad (1.3)$$

Taking reasonable parameter values

$$f = 10^{-4}\ s^{-1},\ N = 10^{-2}\ s^{-1}\ \text{and}\ \frac{d\bar{u}}{dz} = 3\ ms^{-1}/km\ (\frac{\partial\bar{\theta}}{\partial y} \simeq -\ 1K/100\ km)$$

this suggests a maximum possible growth factor in 1 day of about 3.7.

2. NORMAL MODE INSTABILITY

2.1 Linear quasi-geostrophic theory

The quasi-geostrophic equations when linearised about a basic westerly flow in (y,z) may be written

$$\left(\frac{\partial}{\partial t} + \bar{u}\frac{\partial}{\partial x}\right) q' + v'\frac{\partial \bar{q}}{\partial y} = 0$$

$$\text{with } \left(\frac{\partial}{\partial t} + \bar{u}\frac{\partial}{\partial x}\right) \theta' + v'\frac{\partial \bar{\theta}}{\partial y} = 0 \qquad \text{on } z = 0,\ z_T\ , \tag{2.1}$$

where

$$q' = \frac{\partial^2 \psi'}{\partial x^2} + \frac{\partial^2 \psi'}{\partial y^2} + \frac{1}{\rho}\frac{\partial}{\partial z}\left[\rho \frac{f_o^2}{N^2}\frac{\partial^2 \psi'}{\partial z}\right] \tag{2.2}$$

and

$$\theta' = \frac{f\theta_o}{g}\frac{\partial \psi'}{\partial z},$$

and Z_T is the level of the upper boundary, if any. The basic flow appears through its advection and in the terms

$$\frac{\partial \bar{q}}{\partial y} = \beta - \frac{\partial^2 \bar{u}}{\partial y^2} - \frac{1}{\rho}\frac{\partial}{\partial z}\left(\rho \frac{f_o^2}{N^2}\frac{\partial \bar{u}}{\partial z}\right) = \beta - \frac{\partial^2 \bar{u}}{\partial y^2} - \frac{1}{\rho}\frac{\partial}{\partial z}(f_o \rho \alpha_\theta)$$

$$\frac{\partial \bar{\theta}}{\partial y} = -\frac{f\theta_o}{y}\frac{\partial \bar{u}}{\partial z}\ .$$

It was shown by Charney and Stern (1962) that growing normal mode solutions of the form $\psi'(y,z)\, e^{\sigma t} \cos k(x-ct)$ (with $\sigma > 0$) are possible only if

$$\frac{\partial \bar{q}}{\partial y},\ \left.\frac{\partial \bar{\theta}}{\partial y}\right|_{z=0},\ -\left.\frac{\partial \bar{\theta}}{\partial y}\right|_{z=z_T} \text{ are somewhere positive and negative} \tag{2.3}$$

Further, they must be positively correlated with $\bar{u}$, though this restriction is not usually of practical importance in the atmosphere.

We note that a natural ratio of scales such that the contributions for the vertical and horizontal derivative terms in (2.2) are comparable is

$$\frac{H}{L} \sim \frac{f_o}{N} \sim \frac{1}{100} \tag{2.4}$$

Further, if $\bar{u}$ is a function of z only, then we may seek solutions which are sinusoidal in y with wavenumber ℓ. The structure of (2.1) implies that

$$\sigma = kF(K), \tag{2.5}$$

where $K = (k^2 + \ell^2)^{\frac{1}{2}}$ is the total wavenumber.

In the Eady (1949) model, the basic state and approximations are:

$$\frac{d\bar{u}}{dz} = \text{cst.},\ f = \text{cst.},\ N = \text{cst.},\ \text{boundaries at } z = 0,\ z_T.$$

The Charney-Stern conditions (2.3) are satisfied by the corresponding negative $\partial\bar{\theta}/\partial y$ at both boundaries. The growth rate curve is shown in Fig. 4. For no y dependence, the maximum growth rate is at

$$k \simeq 1.6\ f/(NH) \text{ and has } \sigma \simeq 0.31\ f\ \frac{d\bar{u}}{dz}/N\ . \tag{2.6}$$

The wavenumber is consistent with the ratio of scales (2.4) and the growth rate shows the same parameter dependence as (1.3) and does not fall far short of this upper bound. For the values used in Section 1 and z_T = 10 km, the maximum growth factor is 2.3 per day, the wavelength 4000 km and the phase speed equal to the flow at 5 km. The most unstable mode has a structure similar to that deduced in Section 1. It has maximum v' amplitude at both boundaries. The short waves are stable because for $H \sim fL/N \ll z_T$ they do not "feel" the upper boundary.

The Charney (1947) model differs in having

$$f = f_o + \beta y,\ \rho = \rho_o e^{-Z/H} \text{ and no upper boundary.}$$

The Charney-Stern conditions (2.3) are satisfied by the negative $\partial\bar{\theta}/\partial y$ at z=0 and the positive $\partial\bar{q}/\partial y = \beta + \frac{f^2}{N^2}\frac{1}{H}\frac{\partial\bar{u}}{\partial z}$ in the interior. A typical growth rate curve for ℓ = const. is sketched in Fig. 5. The long waves are stabilised by the β-effect but the short waves, even though they are shallow now "feel" the positive interior $\partial\bar{q}/\partial y$. The growth rates, wavelengths, phase speed and structure are generally similar to those of the Eady modes, though lacking the

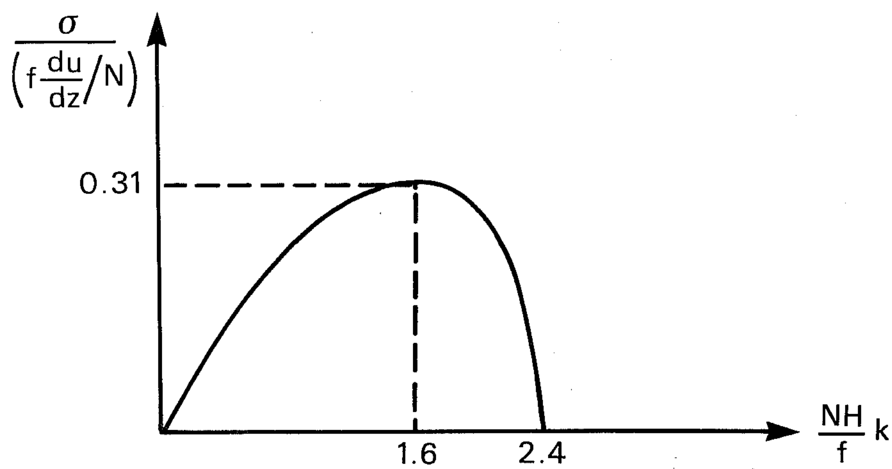

Fig. 4 Schematic growth rate curve for the Eady modes with zero latitudinal variation. This is the most unstable wave for each value of k.

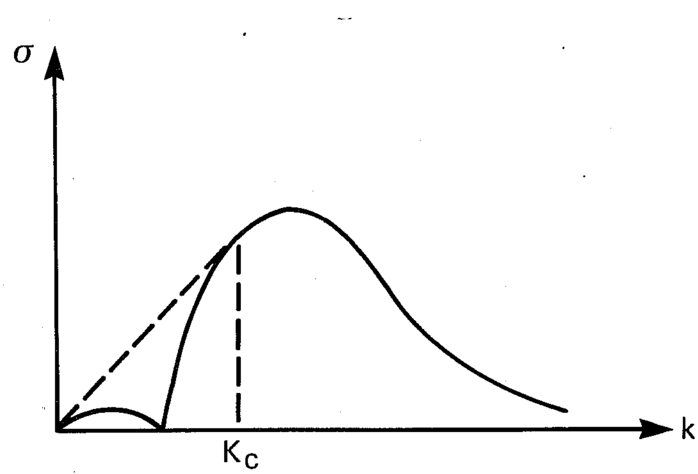

Fig. 5 Schematic growth rate curve for the Charney modes with zero latitudinal variation, for typical parameter values. For $k > K_c$ this is the most unstable mode for each k. For $k < K_c$ the most unstable mode has latitudinal variation such that the total wavenumber is K_c and the growth rate as shown by a heavy dashed line.

upper v' maximum. The importance of the long wavelength stabilisation should not be over-stated. If ℓ is allowed to increase, then from (2.5) and following Hoskins and Revell (1981), the most unstable long waves have a fixed total wavenumber K_c and $\sigma = kF(K_c)$. This modification is shown in Fig. 5.

2.2 Linear and nonlinear normal modes using primitive equations on the sphere

Numerical stability calculations have been performed by e.g. Simmons and Hoskins (1977) for more realistic zonal flows, including westerly wind maxima on sloping tropopauses, and using spherical geometry and the primitive equations. The growth rate curves are generally similar to the modified Charney curve. There is generally a growth rate maximum near zonal wavenumber 8 (zonal wavelength ~ 4000 km). For a $47 ms^{-1}$ jet, this growth rate is equivalent to a factor of 3.1 per day.

The phase speeds are similar to the 700 mb flow. Wavenumber 6 and longer tend to have a maximum in v' at the tropopause comparable to the surface maximum. The large change in static stability and reversal in shear there imply a very large value in $\partial\bar{q}/\partial y$ and a behaviour akin to the lid in the Eady model. The shorter waves are shallower and their structure is similar to that of the short Charney modes. All the modes have a structure similar to that deduced in Section 1.

Of course mid-latitude cyclones are of practical interest when they are of large amplitude, with perturbation velocities comparable to that of the ambient flow in which they are embedded. We are therefore interested in the behaviour of normal modes after they have grown to finite amplitude. Primitive equation models including only horizontal diffusion terms have been initialized with zonal flows plus small amplitude normal modes (typically a surface pressure wave of amplitude 1 mb). As described in detail in Simmons and Hoskins (1978, 1980) and reviewed in Hoskins (1983), after a period of essentially linear behaviour frontal structures and the occlusion process occur. The low-level θ contours are virtually expelled from the latitudinal band occupied by the baroclinic wave. However for wavenumber 6 and longer, upper tropospheric growth, which may be envisaged as an upward propagation of wave activity (Edmon et al., 1980), continues for a day or so. The energy then peaks and decays rapidly. This may be viewed as an equatorward propagation of wave activity and a "wave-breaking" in the region where the flow speed relative to the waves is small (Held and Hoskins, 1985). Shorter waves (wavenumber 9 or greater) which do not "feel" the tropopause, tend not to experience the large upper tropospheric growth phase and achieve much smaller energy levels in their non-linear development.

3. OTHER DYNAMICAL CONSIDERATIONS

Underlying the work of many, but of Eady (1949) in particular was the idea of natural selection. From an initial state, the normal mode components would grow according to their particular growth rates. In time the most unstable normal mode would dominate and so this mode has often tended to figure largely in the discussion. In reality the growth rate spectra e.g. Figs 4 and 5 are generally not very peaked so that even if the initial conditions are in the linear regime, non-linear effects such as the higher energy levels attained by the longer wavelengths mentioned above, become crucial.

Farrell (1982, 84, 85) has recently made an important contribution in stressing another reservation about the concentration on normal modes. Pedlosky (1964) and others have shown that the representation of arbitrary initial conditions requires a so-called continuous spectrum in addition to the normal modes. Although the long-term amplitude of this continuous spectrum decays as t^{-2}, in the short-term it can be very important. For example there can be transient growth in a baroclinic system with growth rate approaching (1.3) even in cases where the Charney-Stern criteria (2.3) are not satisfied. Of course the energetics argument leading to (1.3) did not embody these criteria.

Farrell (1985) went further and proposed that for reasonable surface friction the atmosphere is in fact not baroclinically unstable. Using a more complete linear model, Valdes and Hoskins (1988) find this not to be the case, though the shallow short waves are stabilised. However, the point still remains that in a certain sense the initial perturbation for an individual weather system is the debris from previous weather systems and has a non normal mode structure, and that for a period of a day or so, local transient growth at rates well above that of the most unstable normal mode is possible. Normal mode theory and the Charney-Stern criteria should not be taken too seriously. Its value is in providing well posed problems in which the structures hypothesised to exist in the energetics arguments are indeed found, which show the details of these structures, and their dependence on parameters in the problem. They further provide the only general initial conditions for non-linear integrations, in the sense that the prescribed amplitude is irrelevant provided that it is in the regime of linear theory.

One example of an initial value problem is the study of Simmons and Hoskins (1979) in which a localised perturbation was included in an unstable flow. An orderly growth of baroclinic instability occurred in space and time. The longitude of the development of the downstream waves moved at a speed approximately that of the 300 mb jet and these waves had amplitude more biased to upper tropospheric levels than the most unstable normal modes. Upstream the waves continued to develop at almost the same longitude. The spatial fringes of this development were found to be well represented by a sum of normal modes. An interpretation of the spatial growth is given by an extension of the normal mode theory in which the dispersion relation is considered in complex wavenumber space. In particular, Merkine (1977) has differentiated between convective instability in which the growth from an initial perturbation is all downstream from that perturbation and absolute instability in which there is growth at the point.

Pierrehumbert (1984) has pointed out that this distinction is very important in the context of a longitudinally varying flow. In the case of only convective instability, normal modes will tend to be global even if there is a region of enhanced instability. However if there is absolute instability, there can be local modes in such a region i.e. a self-sustaining storm-track region. He interpreted some of the normal mode calculations of Frederiksen (1983 and refs) in this context.

4. DIABATIC PROCESSES

In the previous section we mentioned the stabilising effect of surface friction. There does not yet appear to be any definitive work on the effects of boundary fluxes of heat. However observational studies of rapidly developing systems such as that of Reed and Albright (1986) have stressed the very large heat fluxes into the warm air ahead of the system. When the cold continental air behind a depression moves over a warm ocean there can be enormous heat fluxes into the atmosphere. In this case their effect is presumably in the direction of damping growth.

Latent heat release in the large-scale ascent of the warm air can be approached in a number of ways. In the parcel energetics argument of Section 1 we may anticipate that the relevant isentropes are now the moist isentropes and that the ascent will tend to be steeper and at half their angle with the horizontal. From (1.1) we expect that the growth rates will be correspondingly increased. In a similar qualitative manner one may anticipate that the Eady growth rate (2.6) should be increased by replacing N by a reduced effective moist N_W. Both these discussions gloss over the fact that one can expect little direct impact on the descending dry cold air.

Hoskins (1982) has discussed how as a frontal region becomes strong, though P is approximately conserved, the stability to "symmetric" motion along isentropes becomes smaller. When the rising air becomes saturated and the relevant surfaces are θ_e then it is possible that these surfaces are more vertical than the absolute momentum (m or n) surfaces. This has been described by Bennet and Hoskins (1979) as conditional symmetric instability and was raised as a possibility for the origin of frontal rainbands. Thus although energy could not be released in this manner in the ambient flow and longitudinal variation was necessary to overcome the angular momentum constraint, at this stage such direct energy release is a possibility. Emanuel (1983) has found that the saturated warm air in an active depression is usually observed to be neutral to this slantwise ascent, suggesting that the adjustment by symmetric instability type motions is relatively fast and efficient. This led Emanuel et al. (1987) to their very interesting study of baroclinic instability in which moist processes were parametrized by setting the "moist" potential vorticity based on θ_e to zero in the rising air. They found a shrinkage of the longitudinal extend of the region of rising motion which was relatively larger in magnitude. The maximum growth rate in the Eady problem can be as much as twice that in the original dry model.

5. FINAL COMMENTS

The theories discussed above, as well as potential vorticity thinking, all influence our view of the development of extra-tropical cyclones. However the observed developments clearly involve a longitudinally varying ambient flow, small and finite amplitude perturbations of various structures, nonlinear dynamics, and boundary layer and free atmosphere frictional and diabatic processes.

References

Bennets, D.A. and B.J. Hoskins, 1979: Conditional symmetric instability - a possible explanation for frontal rainbands. Quart.J.Roy.Meteor.Soc., 105, 945-962.

Charney, J.G., 1947: The dynamics of long waves in a baroclinic westerly current. J.Meteor., 4, 125-162.

Charney, J.G. and M.E. Stern, 1962: On the stability of internal baroclinic jets in a rotating atmosphere. J.Atmos.Sci., 19, 159-172.

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

Edmon, H.J., B.J. Hoskins and M.E. McIntyre, 1980: Eliassen-Palm cross sections for the troposphere. J.Atmos.Sci., 37, 2600-2616. See also Errata, J.Atmos.Sci., 38.

Emanuel, K., 1983: On assessing local conditional symmetric instability from atmospheric soundings. Mon.Wea.Rev., 111, 2016-2033.

Emanuel, K., M. Fantini and A.J. Thorpe, 1987: Baroclinic instability in an environment of small stability to slantwise moist convection. Part I: Two-dimensional models. J.Atmos.Sci., 44, 1559-1573.

Farrell, B., 1982: The initial growth of disturbances in a baroclinic flow. J.Atmos.Sci., 39, 1663-1686.

Farrell, B., 1984: Modal and nonmodal baroclinic waves. J.Atmos.Sci., 41, 668-673.

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

Frederiksen, J.S., 1983: Disturbances and eddy fluxes in Northern Hemisphere flows: instability of three-dimensional January and July flows. J.Atmos.Sci., 40, 836-855.

Green, J.S.A., 1960: A problem in baroclinic instability. Quart.J.Roy.Meteor.Soc., 86, 237-251.

Held, I.M. and B.J. Hoskins, 1985: Large-scale eddies an the General Circulation of the Troposphere. Adv. Geophys., 28, 3-31.

Hoskins, B.J., 1982: The mathematical theory of frontogenesis. Ann.Rev. Fluid Mech., 14, 141-151.

Hoskins, B.J., 1983: Dynamical processes in the atmosphere and the use of models. Quart.J.Roy.Meteor.Soc., 109, 1-21.

Hoskins, B.J. and M.J. Revell, 1981: The most unstable long wavelength baroclinic instability modes. J.Atmos.Sci., 38, 1498-1503.

Merkine, L., 1977: Convective and absolute instability of baroclinic eddies. Geophys.Astrophys.Fluid Dyn., 9,. 129-157.

Pedlosky, J., 1964: An initial value problem in the theory of baroclinic instability. Tellus, 16, 12-17.

Pierrehumbert, R.T., 1984: Local and global instability of zonally varying flow. J.Atmos.Sci., 41, 2141-2162.

Reed, R.J., and M.D. Albright, 1986: A case study of explosive cyclogenesis in the eastern Pacific. Mon.Wea.Rev., 114, 2297-2319.

Simmons, A.J., and B.J. Hoskins, 1977: Baroclinic instability on the sphere: solutions with a more realistic tropopause. J.Atmos.Sci., 34, 581-588.

Simmons, A.J., and B.J. Hoskins, 1978: The life cycles of some nonlinear baroclinic waves. J.Atmos.Sci., 35, 414-432.

Simmons, A.J., and B.J. Hoskins, 1979: The downstream and upstream development of unstable baroclinic waves. J.Atmos.Sci., 36, 1239-1254.

Simmons, A.J., and B.J. Hoskins, 1980: Baroclinic influences on the growth and decay of nonlinear baroclinic waves. J.Atmos.Sci., 37, 1679-1684.

Valdes, P.J., and B.J. Hoskins, 1988: Baroclinic instability of the zonally averaged flow with boundary layer damping. J.Atmos.Sci. (to appear).