Frontal instability generated by tropospheric potential vorticity anomalies

By ALAIN JOLY* and ALAN J. THORPE

Department of Meteorology, University of Reading

SUMMARY

The development of secondary disturbances on a two-dimensional front is described using a linear stability analysis. A two-stage process is envisaged in which, first, strong frontogenesis produces condensation at the front. This leads to the generation of a lower tropospheric zone of high potential vorticity (PV) lying along the front. As frontogenesis weakens, the second stage is entered in which an essentially steady front is susceptible to the growth of linear normal modes along the front owing to the existence of the PV anomaly or, equivalently, the frontal rainband. Thus for a typical band of width 160 km a frontal wave of some 800 km is produced with a growth rate of about 1 day^{-1}. The dynamics of these modes is examined by considering both the realistic frontal flow as well as idealized strips of PV anomalies. The idealized case is potentially applicable to a variety of meteorological situations such as upper jet streaks. Useful insight is gained from an analysis of the energetics of the modes. This shows that for narrow frontal PV anomalies the waves' main energy source is the basic kinetic energy and further they have a negative vertical heat flux. Recent observations suggest that frontal waves are indeed characterized by such a heat flux. As most frontal zones involve significant condensation, and therefore are likely to exhibit a lower tropospheric PV maximum, it is stressed that studies of frontal waves need to include this aspect of the frontal structure.

1. ON FRONTAL WAVES

Many of the cyclones typical of mid-latitude weather are observed to form and grow in regions of already existing air-mass contrasts. So many, in fact, that it led the team of scientists around Vilhelm Bjerknes in Bergen in 1920 to propose that *all* cyclones were generated in 'frontal' zones. Despite the different view of cyclogenesis developed after the Second World War, observations of all kinds, including satellite imagery, often show precisely this process. The long, smoothly curved cold front which is part of a mature, large-scale weather system appears to become distorted by a single or several smaller depressions. In a few cases, the deepening of one of these ripples is so large and sudden that it becomes a major storm. This explosive behaviour will be illustrated below. Such events are still poorly predicted, even when forecasters are helped by present day sophisticated data assimilation and forecasting systems. This may be partly the result of inadequate observations, but clearly a better understanding of the dynamics involved could provide useful rules.

The surface maps reproduced in Figs. 1(a) and (b) typify the situation studied in this paper. In the case of Fig. 1(a) three waves are developing on a front more than 2500 km long, quasi two-dimensional, connected to the rather deep low which had previously moved from Labrador to Iceland. This front is likely to have a marked signature in low-level vorticity. Also noticeable is the low-level southerly jet before the

---

* Present affiliation: Centre de Recherche en Météorologie Dynamique, 2 av. Rapp, F-75340 Paris 7, France.
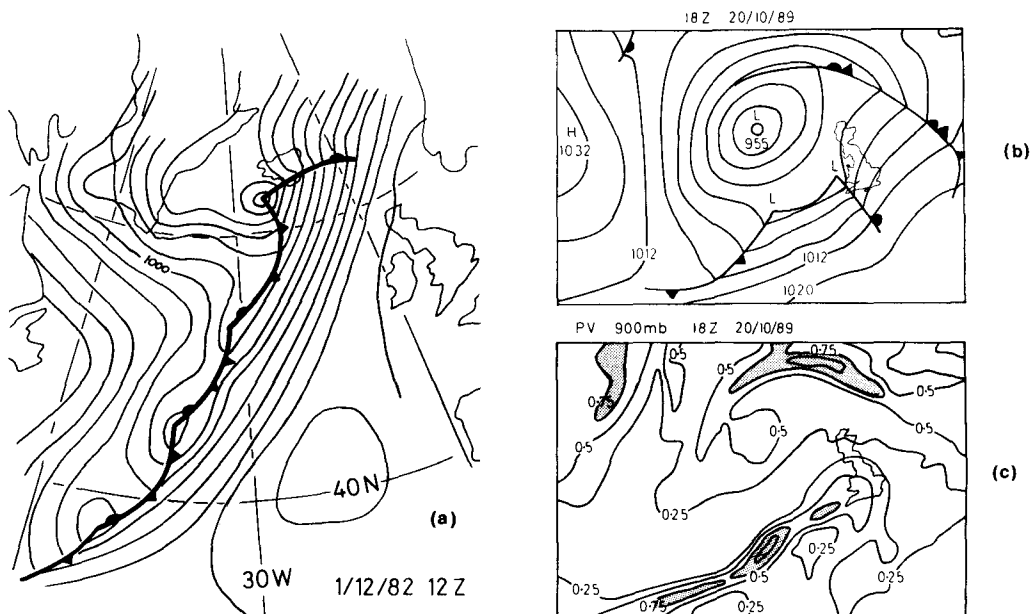
Figure 1. (a) Analysis of mean sea level pressure over the North Atlantic, 1 Dec. 1982 at 12 h UTC produced by the Meteorological Office. Interval between isobars, 4 mb. (b) Analysis of mean sea level pressure for 18 h UTC on 20 Oct. 1989. Interval between isobars 8 mb. (c) Analysis of potential vorticity at 18 h UTC on 20 Oct. 1989 at the 900 mb level. Contours every 0·25 PVU with values greater than 0·75 PVU in dotted shading.

frontal zone, with surface winds reaching 25 m s^{-1} in places. Supporting this analysis, the satellite picture taken three hours later shows the change in the cloud pattern induced by a growing wave. Turning to upper-air fields from the ECMWF analysis (not shown), the waves are developing in a region of strong baroclinicity in the lower troposphere. They bear no signature at higher levels. This suggests that the waves are not deep at this stage. On the other hand, the thermal contrast extends throughout the troposphere, with a rather constant magnitude.

Surprisingly enough, very few well documented case studies have been published in the post-war literature. However, the shallowness of the frontal waves at their early stage is suggested by Browning and Harrold (1969).

Looking again at Fig. 1, we clearly see that the typical *wavelength* of frontal waves is of the order of 1000 km or less; a figure also quoted for observations made a century ago.

In order to gain an idea of the time scale, we consider the time development of another frontal wave. In the latter stages, it became the disastrous 'Fastnet storm' of August 1979. Concentrating on the early stages, we estimate a deepening of one/two mb to five/six mb in twelve hours, an e-folding time of the order of less than half a day. Again, classical baroclinic instability models with uniform potential vorticity and baroclinicity predict e-folding times of the order of two days and wavelength of 3000 km, so that they do not provide a good explanation of the rapid development of frontal waves. The presence of several waves along the front was also noted in this case study. The subsequent growth within the wave of a zone of severe winds is not the subject of this paper: we are interested only in its original appearance.

To summarize the observational evidence, then, frontal waves exist with horizontal wavelengths of 1000 km or less, and grow with time scales of less than one day. They

form in strong, rather deep, baroclinic zones, but seem to be shallow structures in their early stages.

Looking for a theoretical description of such waves, we reconsider the early ideas on frontal cyclogenesis. Bjerknes (1919) and Bjerknes and Solberg (1922) proposed that cyclones were occurring along a quasi-permanent 'frontal' surface, viewed as a narrow boundary separating cold polar air from warm tropical air, with a well-defined slope known from the work of Margules (1906). Bjerknes was convinced they were growing by converting the available potential energy stored in the large temperature gradient into kinetic energy, following another idea of Margules (1905).

Solberg (1928) was the first to attempt a mathematical justification of the underlying ideas. He considered a system of two barotropic layers of different density moving at different zonal speed, and looked for normal modes growing on the sloping interface between fluids. Of the two modes, one had a wavelength of 2000 km but was moving at $25 \, m \, s^{-1}$, the other had a slower phase-speed and a wavelength of 1000 km. The contribution of the kinetic energy of the two parallel currents to both was significant. Further attempts to elaborate this simple frontal instability have been made by Kotschin (1932), Eliasen (1960), Orlanski (1968) and more recently by Sinton and Mechoso (1984).

As their prediction tallied with upper-air observations, baroclinic instability models, such as Charney's (1947) or Eady's (1949) superseded the Norwegian idea of frontal instability. Instead, a zonal, jet-like current is shown to be unstable to waves with a wavelength comparable to upper-air observations. Hoskins and Bretherton (1972) explained how fronts form in such waves. Frontogenesis in fully three-dimensional baroclinic waves is further elucidated by Hoskins and West (1979), within the framework of semi-geostrophic theory. The key parameter in semi-geostrophic analysis is the potential vorticity of the fluid. In all these studies, it is assumed that tropospheric potential vorticity is uniform. But an explanation viewing frontal waves as dry baroclinic disturbances of a uniform flow does not provide a good model: the observed short scales are not predicted by this theory in reasonable conditions.

The concept of potential vorticity leads us directly to a rather different line of thinking on frontal waves: namely, that they could be the indirect product of latent-heat release during the stage of frontogenesis. The idea was originally introduced, in a finite amplitude perspective, by Kleinschmidt (Eliassen and Kleinschmidt 1957): 'the pure wave, taken as a thoroughly dry adiabatic process, cannot account for the formation of the cyclone. The latter requires the creation of an air mass of increased potential vorticity'. He then demonstrates that, indeed, the heating in the frontal ascent could create, below its maximum, a region of higher potential vorticity.

Kleinschmidt also sketches out the invertibility principle (see Hoskins et al. 1985): given the distribution of mass and potential vorticity in some region of a rotating fluid, temperature and wind distribution can be recovered, within the limits of some balance assumption. A positive anomaly of potential vorticity, or equivalently, a positive anomaly of potential temperature on a horizontal boundary, induces cyclonic motion, increases the static stability in the anomaly and decreases it some distance away. Scales depend on the geometry of the anomaly. Kleinschmidt imagined such an anomaly rolling itself up into a new cyclone along the front.

A non-uniform potential vorticity distribution, however, can also lead to a dynamical instability, as realized by Charney and Stern (1962). They pointed out that even if potential temperature is uniform at boundaries, the fluid may remain unstable provided potential vorticity has a local extreme inside the fluid. Such is the case for the jet around the polar vortex in the middle atmosphere. This second source of instability is known as internal baroclinic instability. James and Hoskins (1985) simulated internal baroclinic

normal modes and showed that their evolution is similar to 'boundary' baroclinic modes' life cycle, but for the absence of fronts.

Probably the most systematic study of baroclinic instability combining non-uniform potential vorticity and potential temperature on the boundaries, has been published by Killworth (1980). His analysis is based on asymptotic expansions of the quasi-geostrophic equations. The length and time scales of the unstable mode are predicted whenever possible as well as the main source of energy feeding the disturbance.

In these studies, the non-uniformity of potential vorticity results from an *a priori* choice of the basic current and static stability. A more general approach consists in taking a given distribution of potential vorticity and *deducing* the basic state: that is the essence of the invertibility principle.

Thorpe and Emanuel (1985) included a treatment of condensation in the deformation-forced two-dimensional frontogenesis model, within the framework of semi-geostrophic theory. They found that latent heat release produced a 'local, line symmetric extremum in potential vorticity'. They then proposed that frontal waves of the kind shown on Fig. 1 may be caused by the presence of this along-front anomaly, which allows, according to Charney and Stern, for an internal baroclinic instability to take place. The purpose of this paper is to assess the correctness of this hypothesis. Emanuel *et al.* (1987) and Joly and Thorpe (1989) confirmed the building of a potential vorticity anomaly of large amplitude in the horizontal shear model of frontogenesis.

Returning to the example given in Fig. 1(b) we see that in the analysis of potential vorticity, Fig. 1(c), the waving frontal zone is characterized by a narrow strip of large values. (The average value at 900 mb in the atmosphere is about 0·25 PVU (potential vorticity unit).) The upper tropospheric analysis for this case (not shown) shows that the upper baroclinic zone on the tropopause is essentially two-dimensional parallel to the mean position of the surface front. Evidently the waves apparent at the surface and at 900 mb are relatively shallow. In Fig. 2 another example is given of a frontal wave event on 15 October 1987. This is the early stages of the development of the Great Storm (see Berrisford and Hoskins 1988 for further details). In Fig. 2 the surface synoptic map is shown together with the potential vorticity distribution at 850 mb and a vertical cross-section through the front. It can be seen that the disturbed front co-exists with a narrow strip of high potential vorticity in the lower troposphere. The hypothesis in this paper is that the early development of such frontal waves is related to the existence of that strip. The subsequent explosive development of one of these disturbances most likely occurred owing to significant reinforcement from an upper-level disturbance. It is suggested that relatively weak frontal waves are commonplace but those that subsequently develop rapidly are more unusual and involve dynamics different from that outlined here.

The proposed process responsible for frontal instability is best described in two stages. First, deformation in a *growing* baroclinic wave shapes a quasi two-dimensional front, while at the same time, condensation of water vapour causes the potential vorticity to increase locally, below the maximum of heating. Somehow, deformation inhibits the small-scale instability, forcing the high potential vorticity to extend all along the front. Then, as the cyclone reaches a mature stage, deformation weakens, and in the limit, the front becomes steady: frontal waves start their own development.

Frontal waves originating from low-level potential vorticity anomalies are best analysed within the framework of semi-geostropic theory. The scaling assumption as well as the basic equations are recalled in section 2. A classical numerical technique is applied to these equations in order to extract normal modes: this is briefly described in the same section. Section 3 is dedicated to idealized internal baroclinic instability. It presents results on the stability of strips of anomalous potential vorticity. Section 4 shows the
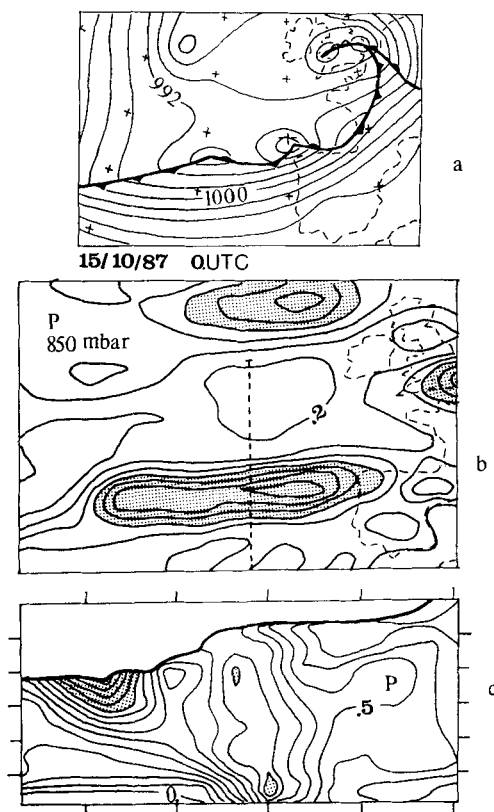
Figure 2. The situation on 15 October 1987, 00 h UTC: frontal waves of small amplitude are developing over the eastern Atlantic Ocean. One of these waves will amplify, and become the so-called Great Storm, but this paper is concerned only with the early stages such as shown here. (a) Sea level pressure showing the synoptic system, from the analysis of the Metorological Office. (b) Potential vorticity at 850 mb re-analysed by the French global NWP EMERAUDE with all available data. Contour interval 0·1 PVU, areas with P larger than 0·5 PVU shaded. (c) Vertical cross-section of potential vorticity from the same analysis. Same contour interval as (b), but shading highlights regions where P is larger than 0·9 PVU. Tick marks are set every 2 km vertically and 500 km horizontally. The bold solid line is the tropopause, above which contours of P are quasi-horizontal. Notice the deep and narrow anomaly corresponding to the front and the low-level maximum.

possibility of instabilities along fronts which have been modified by condensation, and discusses their structure. Section 5 contains some conclusions.

2. PHYSICAL ASSUMPTIONS, MATHEMATICAL AND NUMERICAL FORMULATION

Here a set of equations is derived to study the stability of flows defined by an initial two-dimensional distribution of potential vorticity and distributions of potential temperature on the upper and lower boundaries.

The flow is assumed hydrostatic and inviscid. A further assumption is made on the basis of scales. As seen in the observed examples, the inverse time scale of the wave amplification is $T^{-1} \approx 2 \times 10^{-5} \text{ s}^{-1}$. With a wavelength of 1500 km, and a warm conveyor belt along the front with a mean speed of 15 m s^{-1}, there is an advective time scale slightly smaller than $T^{-1} \approx 5 \times 10^{-5} \text{ s}^{-1}$. Even for extreme events, the acceleration is smaller than the time scale of inertial forces, f (10^{-4} s^{-1} at mid-latitudes). Therefore, it is assumed that the accelerations of individual air parcels even in the vicinity of a front are

dominated by the effects of rotation. In other words, a Rossby number, defined in a Lagrangian sense, is small:

$$Ro_L = \frac{D|\mathbf{V}|}{Dt} \Big/ (f|\mathbf{V}|) \ll 1. \tag{1}$$

The consequences of this assumption were examined in detail by Hoskins (1975). The first of these are the equations of motion resulting from the geostrophic momentum approximation (Eliassen, 1948):

$$Du_g/Dt - fv_a = 0 \qquad Dv_g/Dt + fu_a = 0. \tag{2}$$

Here, (u_g, v_g) are components of the geostrophic velocity $(-f^{-1}\partial\phi/\partial y, f^{-1}\partial\phi/\partial x)$, ϕ the geopotential, (u_a, v_a) stands for the ageostrophic horizontal wind, and f is the Coriolis parameter. D/Dt is the Lagrangian or material derivative. Recall that advection by the full wind is retained. To Eq. (2) a continuity equation is added,

$$\frac{\partial u}{\partial x} + \frac{\partial v}{\partial y} + \frac{1}{r}\frac{\partial rw}{\partial z^*} = 0 \tag{3}$$

together with a thermodynamic equation,

$$D\theta/Dt = S \tag{4}$$

where $z^* = c_p H_s/R[1 - (p/p_{00})^{R/c_p}]$, a modified pressure coordinate introduced by Hoskins and Bretherton (1972), r a density factor depending on z^* only, w the (pressure) vertical velocity (rate of change of z^*), θ potential temperature, S a diabatic source term and R is the gas constant for unit mass of air. Hoskins (1975) has shown that this set conserves energy (when S is zero), and, provided a modified vorticity is used, it conserves potential vorticity:

$$P = \frac{1}{r}\zeta_g.\nabla\theta. \tag{5}$$

ζ_g is the sum of the curl of the geostrophic wind (consistent with the hydrostatic approximation) the earth rotation, and a Jacobian vector, which, according to Hoskins and Draghici (1977) remains a small term, consistent with (1). The dynamics contained in (2) and (3) are described in an equation for P:

$$\frac{DP}{Dt} = \frac{1}{r}\zeta_g.\nabla S. \tag{6}$$

The geostrophic momentum approximation has proved very useful to study large-scale baroclinic instability, frontogenesis, frontal structure in dry and moist air as well as sea breezes (Cullen et al 1987), flow over mountains (Blumen and Gross 1987) and other meso-scale phenomena. However, it has limitations. In adiabatic flows, with uniform potential vorticity, the Richardson number, which can be scaled by ζ^{-1}, becomes locally very small as frontogenesis proceeds (Hoskins 1982), and turbulent motion sets in (because of the onset of Kelvin–Helmholtz and Rayleigh small-scale instabilities). Results with values of vorticity substantially larger than $10f$ should not be considered very realistic; eventually mixing and adjustment to gravity cannot be neglected. This constraint is more stringent than the smallness of the accelerations of air parcels. In diabatic flows, however, it may not be so, as outlined by Emanuel (1985), as then potential vorticity does not remain uniform. It decreases when both frontogenesis and condensation operate. In the limit, we may find regions in the fluid where it becomes

negative. From the work of Hoskins (1974) and Bennetts and Hoskins (1979) we know that this is equivalent to the onset of symmetric instability, which is not well described by filtered equations. Emanuel estimates that the geostrophic momentum approximation breaks down when P becomes about ten times smaller than its typical tropospheric value of 0·3 PVU where vorticity is large. Ageostrophic accelerations become significant beyond this point. These are definitely extreme values, which do not occur in the examples shown here, and are rarely met in general. This supports, the relevance of the geostrophic momentum approximation in this case.

Thus, throughout this paper, we consider flows which satisfy the smallness of the Rossby number as defined in (1). The resulting equations still retain the full trajectories of fluid parcels. The high degree of non-linearity implied can be reduced by introducing geostrophic coordinates (Hoskins and Bretherton 1972; Hoskins 1975). We define

$$X = x + v_g/f \qquad Y = y - u_g/f \qquad Z = z^* \qquad T = t. \tag{7}$$

In transformed space, displacements are geostrophic ($DX/Dt = u_g$, $DY/Dt = v_g$), and most of the mathematical simplicity of the quasi-geostrophic system is recovered. The Jacobian of the change of coordinate, J, is the vertical component of absolute vorticity ζ scaled by f. The geostrophic flow is completely determined by the knowledge of a potential $\Phi = \phi + 1/2(u_g^2 + v_g^2)$. This function Φ is a stream-function of the geostrophic components (u_g, v_g, θ).

We choose the X-axis as the cross-front direction. The domain is unbounded following the along-front axis Y, bounded by rigid horizontal surfaces at $Z = 0$ and $Z = H$. Another assumption is implicit in this model: the curvature of the front is neglected, again following Hoskins and Bretherton (1972). The radius of curvature has to be much larger than $|V|/f$, which is the case for typical long fronts linked to baroclinic waves.

In order to describe a two-dimensional basic state, we assume that a strictly two-dimensional distribution of potential vorticity $\overline{P}(X, Z)$ is given, together with a distribution of $\overline{\theta}$ on the horizontal boundaries $Z = 0$ and $Z = H$.

The invertibility principle is embodied in the inversion equation:

$$r\overline{P}\frac{\partial^2\overline{\Phi}}{\partial X^2} + \frac{\theta_0 f^3}{g}\frac{\partial^2\overline{\Phi}}{\partial Z^2} = f^2 r\overline{P} \tag{8}$$

with

$$\overline{\theta}(Z = 0 \quad \text{and} \quad Z = H) = \frac{\theta_0}{g}\frac{\partial\overline{\Phi}}{\partial Z}$$

as vertical boundary conditions. g is the gravitational acceleration.

Equation (8) is written in a way suitable for its numerical solution even with steep changes in the function $\overline{P}(X, Z)$. The geostrophic flow ($\overline{v}_g(X, Z), \overline{\theta}(X, Z), \overline{J}(X, Z)$) is deduced from a given distribution of potential vorticity $\overline{P}$ by solving (8) subject to the boundary conditions. This is conceptually quite different from the usual approach, where potential vorticity can be deduced *from* an *a priori* distribution of wind (say). Here, $\overline{P}(X, Z)$ is either specified (as in section 3) or is the result of a time-dependent integration of the moist semi-geostrophic equations simulating moist frontogenesis (as in section 4). As the basic state is assumed stationary or steady, there is no geostrophic large-scale forcing, and consequently no ageostrophic motion.

All current two-dimensional distributions of boundary potential temperature or potential vorticity (jet-flow, deformation front, horizontal shear front) can be set up in a modified way allowing for periodic lateral boundary conditions. In some cases, it is important to choose a half-period large enough so that it does not affect the result.

However, this is the only limitation implied by such a choice: consequently, periodic lateral boundaries were used throughout this study, with period L.

An infinitely small three-dimensional perturbation is added to the basic flow representing a two-dimensional steady front, $P'(X, Y, Z, T)$ and $\theta'(X, Y, Z = 0$ and $Z = H, T)$, and its possible growth is assessed. Linearizing the equations for potential vorticity (6) and potential temperature (4) with the vertical boundary conditions $w = 0$ at $Z = 0$ and $Z = H$, we find:

$$\left(\frac{\partial}{\partial T} + \bar{v}_g \frac{\partial}{\partial Y}\right) P' = -u'_g \frac{\partial \bar{P}}{\partial X} - w' \frac{\partial \bar{P}}{\partial Z} + \frac{\bar{f}J}{r} \frac{\partial S'}{\partial Z}$$

$$\left(\frac{\partial}{\partial T} + \bar{v}_g \frac{\partial}{\partial Y}\right) \theta' = -u'_g \frac{\partial \bar{\theta}}{\partial X} \qquad \text{on } Z = 0 \qquad \text{and } Z = H.$$

$\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad$ (9)

S' is the perturbation diabatic source term. Following Emanuel et al. (1987), the diabatic source term consistent with conservation of equivalent potential temperature is

$$S = \frac{r}{f}\left(P - \frac{\theta}{\theta_e}\frac{\Gamma_m}{\Gamma_d} P_e\right) \max\{0, w^*\}$$

$\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad$ (10)

where P_e is the moist potential vorticity and Γ_m/Γ_d is the ratio of the saturated to the dry adiabatic lapse rate. For simplicity, the term involving P_e is treated as a constant ηP_{00}, where P_{00} is a reference potential vorticity and η a number which, observations suggest, is very near zero in frontal ascents. The effective vertical velocity w^* is defined as w/J.

A choice is offered because (10) is not differentiable. Applying the linearization strictly, the sign of the *basic* vertical velocity determines S': only air parcels in a basic-state ascent experience condensation. The alternative is to test the sign of $w^{*'}$ instead: latent heat would be released in air parcels moved by the perturbation ascent. Stability problems making use of this alternative introduce another step in the complexity of the method of solution, because of matching conditions where $w^{*'}$ changes sign. They have been studied, for instance, analytically by Emanuel et al. (1987) or numerically by Joly and Thorpe (1989) in the case of moist baroclinic instability, and by a number of authors in the study of convection. Here, we will keep the equations separable in Y by setting $S' = 0$. Including a non-zero S' may modify the results by causing an asymmetry between ascent and descent so that, for example, regions of ascent will have an effective potential-vorticity anomaly smaller than those of descent. It would therefore reduce their horizontal scale and enhance the growth.

The perturbation vertical velocity is recovered diagnostically from the linearized w-equation (see Hoskins and Draghici 1977 for a derivation of the full equation):

$$\left(\frac{\partial^2}{\partial X^2} + \frac{\partial^2}{\partial Y^2}\right)\left(\frac{g}{\theta_0}\frac{\bar{P}}{f} rw^{*'}\right) + f^2 \frac{\partial}{\partial Z}\left(\frac{1}{r}\frac{\partial}{\partial Z} rw^{*'}\right) = F'_Q$$

$$F'_Q = -2\frac{g}{\theta_0}\left[\frac{\partial}{\partial X}\left(\frac{\partial \theta'}{\partial Y}\frac{\partial \bar{v}_g}{\partial X} - \frac{\partial v'_g}{\partial Y}\frac{\partial \bar{\theta}}{\partial X}\right) + \frac{\partial}{\partial Y}\left(\frac{\partial u'_g}{\partial Y}\frac{\partial \bar{\theta}}{\partial X}\right)\right]$$

$\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad$ (11)

F'_Q is the linearized divergence of the Q vector.

Then the Jacobian of the change to geostrophic coordinates $J = \zeta/f$ is linearized into

$$J' = \frac{\bar{J}^2}{f^2}\left(\frac{\partial^2}{\partial X^2} + \frac{1}{\bar{J}}\frac{\partial^2}{\partial Y^2}\right)\Phi'$$

and the linearized inversion equation allowing recovery of $(u'_g, v'_g, \theta', J')$ is

$$r\bar{P}\left(\frac{\partial^2\Phi'}{\partial X^2} + \frac{1}{\bar{J}}\frac{\partial^2\Phi'}{\partial Y^2}\right) + \frac{\theta_0}{g}f^3\frac{\partial^2\Phi'}{\partial Z^2} = \frac{f^2}{\bar{J}}rP'. \tag{12}$$

Note that this form results from the straightforward linearization of the full semi-geostrophic equations and includes a weight $1/\bar{J}$ in the Laplacian which is often set to unity for flows with small relative vorticity. Here it is retained in full. Boundary conditions are provided by the knowledge of θ'. Equations (9)–(12) are similar to those derived by Hoskins (1976). They look like their quasi-geostrophic counterpart except that they are expressed in a transformed space and that *potential vorticity*, now a full function of space, replaces the buoyancy frequency $N^2(z^*)$.

Equations (9)–(12) form a complete set. Interior equations can be combined into a single one for Φ', which is given for reference:

$$\left(\frac{\partial}{\partial T} + \bar{v}_g\frac{\partial}{\partial Y}\right)\left\{\overline{PJ}\left[\frac{1}{f^2}\left(\frac{\partial^2}{\partial X^2} + \frac{1}{\bar{J}}\frac{\partial^2}{\partial Y^2}\right)\Phi' + \frac{\theta_0}{g}\frac{f}{r}\frac{\partial}{\partial Z}\left(\frac{1}{\bar{P}}\frac{\partial\Phi'}{\partial Z}\right)\right]\right\} = \frac{1}{f}\left(\frac{\partial\bar{P}}{\partial X}\right)_{\bar{\theta}}\frac{\partial\Phi'}{\partial Y}$$

$$\left(\frac{\partial}{\partial T} + \bar{v}_g\frac{\partial}{\partial Y}\right)\frac{\partial\Phi'}{\partial Z} = \frac{g}{f\theta_0}\frac{\partial\bar{\theta}}{\partial X}\frac{\partial\Phi'}{\partial Y} \qquad \text{on } Z = 0 \qquad \text{and } Z = H. \tag{13}$$

From this equation, the arguments leading to the Charney and Stern (1962) theorem can be repeated. Necessary conditions for instability are that either $(\partial\bar{P}/\partial X)_{\bar{\theta}}$ changes sign within the domain and/or $\partial\bar{\theta}/\partial X$ keeps the same sign on both vertical boundaries, or $\partial\bar{\theta}/\partial X$ has both signs on one of the boundaries.

As the basic front is independent of the along-front coordinate Y and of time T, an unstable perturbation can be expressed in terms of normal modes of the form:

$$\Phi' = \hat{\phi}(X, Z)\, e^{ilY}\, e^{\sigma T} \tag{14}$$

where l is the along-front wavenumber of the mode, and σ its complex growth rate. Unstable solutions correspond to a positive real part of σ. Equations (9)–(12) can be transformed (or alternatively, (13)). Furthermore, the choice of periodic lateral boundaries implies that

$$\hat{\phi}(X, Z) = \sum_{m=-\infty}^{+\infty} \phi_m(Z)\exp\left(im\frac{2\pi}{L}X\right).$$

It should be noted that, $\hat{\phi}$ being complex, ϕ_{-m} is NOT the conjugate of ϕ_m.

In general, the eigenvalue problem formed by the transformed equations can only be solved numerically. One way is via discretizing (13). While a suitable spectral expansion exists in the horizontal, it should be noted that this is not true in the vertical. The eigenvalue σ is involved in the boundary condition: this is one of the technical difficulties with (13). However, for a finite subset of a spectral basis, finite elements or finite differences, equations (13) can be cast in matrix form. The eigenvalues are recovered by a standard algorithm. Such a technique was developed for this study, but the size of the matrices required for analysing frontal structure with enough precision is prohibitive. It was used in order to discriminate roughly between different modes obtained with a second, much more accurate, method.

This other method consists of treating the original set (9)–(12) as a time-dependent problem. An arbitrary initial condition is chosen, and the system is integrated in time until exponential growth is achieved. The solution is then the eigenvector corresponding to the dominant eigenvalue. The latter is recovered by comparing successive realizations of the eigenvector. This is the so-called 'initial value approach', originally introduced

into meteorology by Brown (1969). Details can be found in Joly and Thorpe (1989). It was checked that results are insensitive to changes in the initial conditions.

The method was assessed by comparison with analytical results and performed well. This included a verification against the analysis of the stability of the horizontal shear front with uniform potential vorticity (Joly and Thorpe (to be published)). This problem is weakly non-separable, and thus is a good test of a spectral code. In the next section, some basic properties of internal baroclinic instability are analysed.

3. THE STABILITY OF SIMPLE POTENTIAL VORTICITY DISTRIBUTIONS

The Charney and Stern (1962) theorem suggests that, in the absence of gradients of potential temperature on horizontal boundaries, a basic state having $(\partial \bar{P}/\partial X)_{\bar{\theta}}$ of both signs within the fluid might still be unstable to three-dimensional perturbations. There must be an extremum of potential vorticity in the fluid interior. This section considers the simplest form of this situation, where the stability of an arbitrary anomaly of potential vorticity with a single sign is examined to gain dynamical insight. The presence of an isolated anomaly of potential vorticity can be assigned to many possible sources: radiative transfer, surface fluxes, condensation, evaporation, gradients in the distribution of currents, temperature, and salt in the ocean. For the sake of consistency with the remaining sections, typical atmospheric constants are used to provide dimensional orders of magnitude, but results are quite general.

A simple anomaly arises from the following definition:

$$r\delta\bar{P} = \rho_0 P_0' \exp - \frac{(X - L_1)^2}{L_2^2} \exp - \frac{(Z - H_1)^2}{H_2^2}. \tag{15}$$

It describes a local extremum of amplitude P_0' at location $X = L_1$ and $Z = H_1$, with characteristic scales L_2 in the horizontal and H_2 in the vertical. In the following experiments, L_1 is the middle of a domain of width L, bounded by 'walls', or periodic, such that $L_2 < L/2$. In this way, the lateral boundaries do not affect the solution in the region of the anomaly. At the horizontal boundaries $Z = 0$ and $Z = H$, the potential temperature is uniform on these surfaces.

The basic parameters are: a background potential vorticity $P_{00} = 0.267$ PVU, with vertical extent $H = 10$ km, a scale height $H_s = 8434$ m (or $\theta_0 = 288$ K), $f = 10^{-4}$ s^{-1}, $H_1 = 5$ km, $H_2 = 870$ m and $P_0' = 1$ PVU.

A first example of such a distribution and of the resulting geostrophic flow is shown in physical space in Fig. 3, where $L_2 = 310$ km and $L = 2000$ km. As discussed by Hoskins *et al.* (1985), the induced flow is cyclonic in the region of the potential vorticity anomaly, with maximum vorticity $1.9f$. Reducing the horizontal scale L_2 increases the response. For $L_2 = 50$ km, vorticity becomes $3.2f$, while the isentropes are barely disturbed.

Negative anomalies can also be considered, although their amplitude must be limited: the invertibility principle is meaningful provided potential vorticity remains positive everywhere. As expected, a negative anomaly generates an anticyclonic circulation, with a reduced static stability at the anomaly, increased near the vertical boundaries (not shown here). The response, however, is not symmetric: a negative anomaly has more impact than a positive one having the same magnitude. Vorticity is increased by $0.2f$ with the positive anomaly of $+0.25$ PVU and reduced by more than $0.5f$ with a negative one.

In the absence of the anomaly, the background potential vorticity and basic static stability are related by

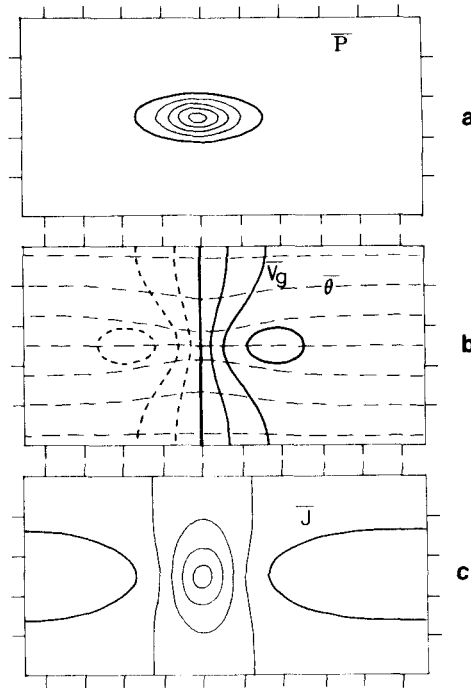$$P_{00} = f/\rho_0 \, \partial\bar{\theta}/\partial Z.$$

Figure 3.   Vertical cross-section of a positive potential vorticity anomaly and the induced basic flow. Tick marks horizontally every 200 km, 2 km vertically. (a) Potential vorticity $\bar{P}$, contour interval 0·2 PVU. Outer line is 0·4 PVU. The scale L_2 is 310 km. (b) Solid and dotted lines, along-strip geostrophic wind $\bar{v}_g$, contour interval 4 m s^{-1}, negative values dotted. Dashed lines, potential temperature $\bar{\theta}$, contour interval 4 K. (c) Non-dimensional absolute vorticity J, contour interval 0·2, heavy line 1.

It is useful to rescale P_{00} as

$$N_0^2 = \frac{g}{\theta_0} \frac{\rho_0 P_{00}}{f}$$

for future reference. With an anomaly of magnitude δP, we have

$$P + \delta P = \frac{\zeta}{\rho_0} \left(\frac{\partial\theta}{\partial z^*}\right)_{fx+V_g} = \frac{\zeta}{\rho_0} \frac{\partial\theta}{\partial Z}.$$

Let α be a parameter measuring the partition of the anomaly impact on vorticity and (slantwise) static stability:

$$\alpha = \frac{f^2}{\rho_0 P_{00}} \frac{\partial\theta/\partial Z}{\zeta}.$$

Our results suggest that a wide anomaly strongly influences static stability, so that α is large, whereas a narrow anomaly is expressed as a large change in vorticity, or a small value of α.

Thermal wind balance is another strong constraint on the flow, allowing α to be related to the scales of the anomaly. After one differentiation with respect to X and another one with respect to Z, this constraint reads:

$$f \frac{\partial^2}{\partial Z^2}\left(\frac{\partial v_g}{\partial X}\right) = \frac{g}{g_0} \frac{\partial^2}{\partial X^2}\left(\frac{\partial\theta}{\partial Z}\right).$$

Now,

$$\frac{\partial v_g}{\partial X} = f - \frac{f^2}{\zeta}$$

so that

$$\frac{\partial^2}{\partial Z^2}\left(\frac{\partial v_g}{\partial X}\right) = \frac{f^2}{\zeta^2}\frac{\partial^2 \zeta}{\partial Z^2} - 2\frac{f^2}{\zeta^3}\left(\frac{\partial \zeta}{\partial Z}\right)^2.$$

In terms of scales, letting L be the horizontal scale (proportional to L_2) and H the vertical scale (combination of H_2, H and fL/N_0), those two terms have the same magnitude, so that thermal wind balance imposes that

$$\frac{f^3}{\zeta H^2} \sim \frac{g}{\theta_0}\frac{\partial\theta/\partial Z}{L^2}$$

and so

$$\alpha \sim \frac{f^2}{N_0^2}\cdot\frac{f^2}{\zeta^2}\cdot\frac{L^2}{H^2}$$

in *geostrophic* space, or using L_p as the horizontal scale in physical space:

$$\alpha \sim \frac{f^2}{N_0^2}\frac{L_p^2}{H^2}$$

and the response of the fluid to a potential vorticity anomaly is summarized by the magnitude of vorticity:

$$\zeta \sim N_0\frac{H}{L_p}\left(1 + \frac{\delta P}{P_{00}}\right)^{1/2}. \tag{16}$$

This equation embodies the two physical constraints in the invertibility problem: the product of vorticity and slantwise static stability is the potential vorticity, and the flow is in thermal wind balance. It explains why, as the horizontal scale is reduced, vorticity increases at the expense of static stability, and also why positive and negative anomalies do not have the same effect. These properties affect the nature of a possible instability. It is clear, for example, that in the presence of narrow anomalies, there is no significant potential energy as opposed to kinetic energy available for conversion into the growth of a three-dimensional perturbation.

The stability problem (9)–(12) is solved for the same two basic states discussed earlier. Inspection of Fig. 4 immediately reveals that the geometry of the potential vorticity anomaly can strongly influence the scale of the fastest growing mode. The growth-rate curves have well-defined maxima having a finite growth rate at a finite non-zero along-strip wavenumber l. But the scale spanned by varying the width of the anomaly *is* dramatically increased. When the anomaly has a half-width of 310 km, the most unstable mode has an along-strip wavelength of 3600 km with an e-folding time of 1·8 days. If the half-width is reduced to 50 km, the e-folding time is reduced to 1·2 days, and the wavelength to 570 km. The phase speed of these modes is zero.

It is of course not possible to undertake a survey of the whole range of free parameters defining the anomaly, its half-width L_2, half-depth H_2, its amplitude P_0' or its position on the vertical H_1. Choosing a reference point corresponding to the 'narrow' anomaly:

$$L_2 = 50\,\text{km}, \qquad P_0' = 1\,\text{PVU}, \qquad H_2 = 870\,\text{m}, \qquad H_1 = 5000\,\text{m}\,(=H/2),$$
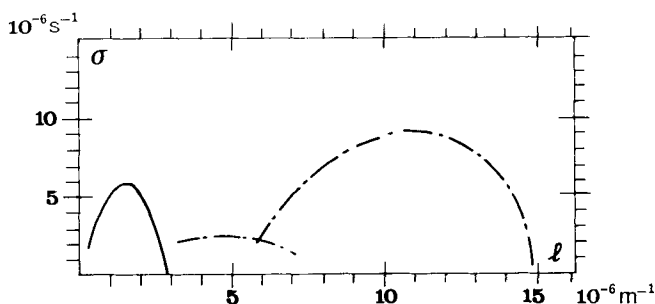
Figure 4.   Real part of the growth rate of a normal-mode perturbation to the flow sketched on Fig. 3 as a function of its along-strip wavenumber l. Solid line: 'wide' anomaly ($L_2 = 310$ km). Dash-dotted line: 'narrow' anomaly ($L_2 = 50$ km). The corresponding imaginary part, or mode frequency, is zero.

one parameter at a time is changed, and, against its variations, the maximum growth rate and the corresponding most unstable wavenumber is plotted.

Changing the width L_2 strongly influences the wavenumber. In fact, there is a good linear relationship between L_2 and $2\pi/l = L_y$, the wavelength. Multiplying L_2 by ten gives an excellent approximation to the most unstable wavelength. The growth rate is, like l, a decreasing function of the width. Here also the two are related linearly, but reducing L_2 by a factor 2 increases the growth rate by a factor 1·3 (Fig. 5).

The growth rate is primarily influenced by the amplitude of the anomaly, with a curve not unlike $P_0'^{1/2}$ in shape. However, much like the basic state itself, the response is not symmetric with respect to zero anomaly. Unstable modes are influenced by wind shear and by horizontal gradients of θ, and, as just seen, the shear is stronger for a negative anomaly than for a positive one, thus the conversion might be expected to be more efficient and the growth quicker for a negative anomaly. The wavenumber of the most unstable mode increases linearly with the amplitude, with a different slope for negative and positive anomalies. This effect is, however, limited; the slope is weak (Fig. 6).

Turning to the depth of the potential vorticity anomaly, we see that as it becomes deeper the growth rate increases substantially, reaching 0·8 day^{-1} for a half-depth of 2·5 km. The most unstable wavenumber levels off more quickly than the growth rate, implying that the vertical extent of the domain dominates the scale selection before H_2 limits the maximum growth rate (Fig. 7).

To gain deeper insight, consideration of energetics is useful in linear theory. Frontal waves have two obvious energy sources: the large kinetic energy in the frontal jets, with large shear to enhance conversion, and the equally large temperature gradient, implying a large source of potential energy. Margules and Bjerknes thought that the atmospheric
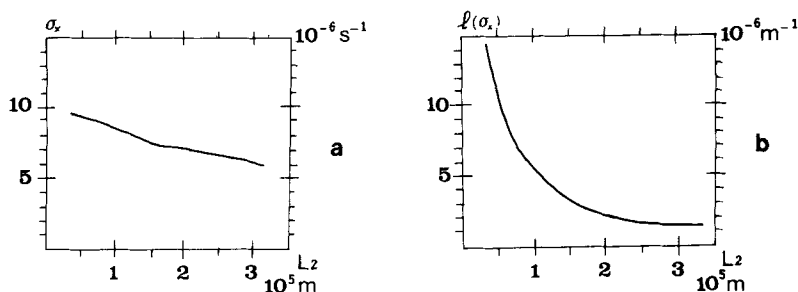

Figure 5.   (a) Maximum growth rate σ as a function of the half-width L_2. Other parameters as for the 'narrow' anomaly ($L_2 = 50$ km). (b) Wavenumber l of the most unstable normal mode corresponding to (a).
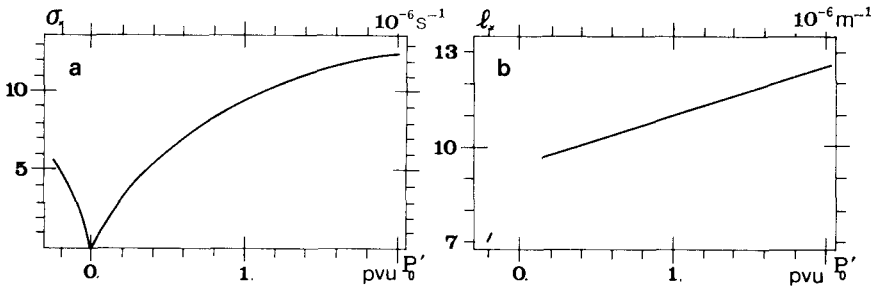
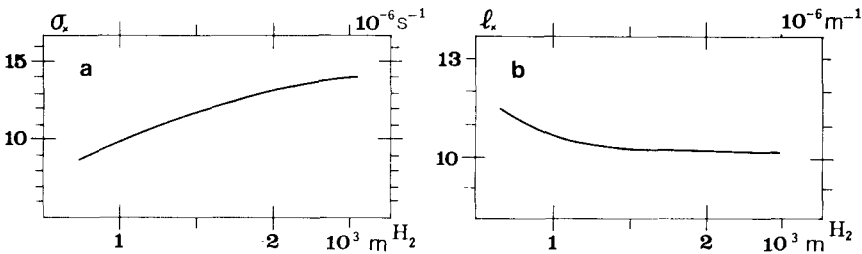Figure 6. As in Fig. 5 but as a function of the anomaly amplitude P_0'.

Figure 7. As in Fig. 5 but as a function of the half-depth H_2.

waves were mostly driven by the latter, but when Solberg attempted the study of the two-fluid problem, he found a larger contribution from the basic-state kinetic energy. Baroclinic instability theory justified the early intuition of the famous pioneers. A simple two-dimensional baroclinic wave converts potential energy from the mean flow to the wave, and then, along with the direct circulation, it is converted to the wave kinetic energy. Despite the various energy sources, the basic instability mechanism relies on the potential vorticity structure.

A way of labelling the instabilities stems from the energy equations shown in the appendix. A mode for which the Reynolds stress term is zero, and the energy source is the mean potential energy, will be called a 'baroclinic mode'. Conversely, a mode with zero horizontal heat-flux conversion, growing solely from the main kinetic energy will be named a 'barotropic mode'. The internal conversion can have both signs: it is positive in 'pure' baroclinic waves, negative in symmetric instability.

A good summary of most of the mode structure is provided by the mean–wave-energy correlations. Figure 8 shows the central domain, to be compared with Fig. 3. It shows the extent of the mode in the (x, z^*) plane having the same scales as the basic-state features. In the region where the isentropes are significantly bent, away from the maximum anomaly, some potential energy is retrieved. From Eq. (A4), this implies that the phase relation between u_g' and θ' is different from that in the mid-anomaly vertical plane. The other two non-negligible terms show that the internal conversion term is negative everywhere, completely unlike the external baroclinic waves, and the negative correlation between u_g' and v_g' in a region of positive horizontal shear $\partial \bar{v}_g / \partial x$ implies a gain for the perturbation. This term in fact constitutes *the main energy supply of these waves*. It implies a noticeable horizontal tilt of the pressure wave.

As for pure barotropic instability (Kuo 1949), the horizontal Reynolds stress is dominant. The barotropic nature of internal baroclinic waves is clearly apparent on the full–mean-energy conversions, which are shown for both horizontal widths (Fig. 9). Note also the characteristic reversed internal conversion, implying indirect vertical circulation in the perturbation, similar to what is observed in the linear normal modes of symmetric
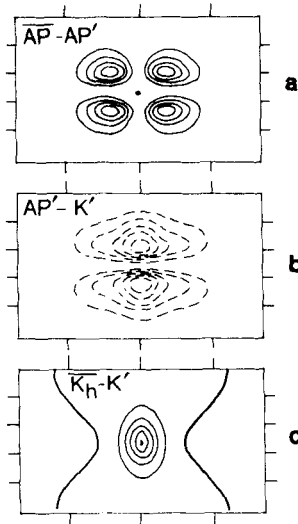
Figure 8.   Energy conversions averaged over one wavelength in the along-strip direction Y in the flow induced by the most unstable normal mode of Fig. 3 for the 'wide' anomaly ($l = 1.75\ 10^{-6}\ \mathrm{m}^{-1}$). Pictures are shown in a vertical cross-section (x, z) plane, in the physical space defined by the basic state. Only part of the domain is shown. Here one unit is $10^{-5}\ \mathrm{J\ kg^{-1}\ s^{-1}}$. See Fig. 9 caption for definitions of the symbols. Contour intervals are (a) 0·6 unit for the conversion from basic available potential energy $\overline{AP}$, (b) 0·1 unit for the internal conversion and (c) 5 units for the conversion from basic kinetic energy $\overline{K}$.
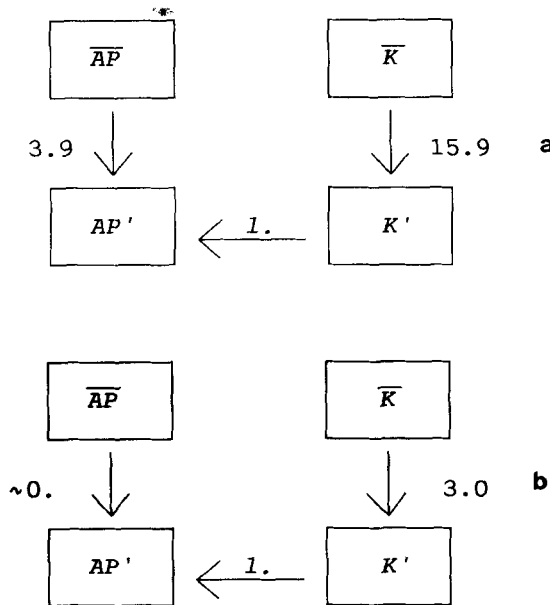

Figure 9.   Energy box diagram, showing the rates of conversion as well as their signs, normalized by the internal conversion based on the correlation $\langle w'\theta'\rangle$. AP stands for a reservoir of available potential energy, K for kinetic energy. The overbar denotes a basic-state property, the prime a perturbation quantity. The conversion from $\overline{AP}$ to AP' is based on $\langle \partial_x\overline{\theta}.u'_g\theta'\rangle$, the conversion from $\overline{K}$ to K' is dominated by $\langle \partial_x v_g.u'_g v'_g\rangle$. (a) The most unstable mode on the wide anomaly, otherwise shown on Fig. 8. (b) The most unstable mode growing on the narrow anomaly ($l = 11 \times 10^{-6}\ \mathrm{m}^{-1}$).

instability (Emanuel, 1979). However, all these terms are small, as are the relative amplitudes of the perturbations of θ' and w', strongly suggesting the relevance of a barotropic approximation to these problems. For both widths, the contribution of the vertical Reynolds stress, excluded from quasi-geostrophic energetics is negligible. For 1 unit of internal conversion, it is -0.02 and -0.05 unit respectively for $L_2 = 310$ km and $L_2 = 50$ km.

Thus the horizontal scale of an unstable perturbation can be much reduced compared to classical baroclinic instability theory. Modes with wave-lengths smaller than 1000 km are possible, provided the anomaly is not wider than 200 km, has a large amplitude and is reasonably deep. The growth rate can reach e-folding times of about a day under the same circumstances. The structure of the resulting unstable perturbations is dominated by their horizontal tilt (X-ward with decreasing Y if the anomaly is positive) allowing them to grow from the kinetic energy of the basic flow. Another typical aspect is the negative correlation between w' and θ' along the maximum of the anomaly. For relatively narrow anomalies (which do not affect the isentropes), it implies in the mean a negative heat-flux. However, as the anomaly widens (the basic static stability is changed significantly), conversion from the available potential energy is possible, and may reverse the mean internal conversion rate.

Thorpe and Emanuel (1985) have shown that latent-heat release in fronts considerably reduces the width of the ascent, and thus of the potential vorticity anomaly it generates between the maximum of heating and the surface. Through this property, it appears possible to destabilize fronts on the scale of 1000 km or less, generating frontal waves. The simplest model of this instability is just as described as long as the dynamics are not significantly affected by the presence of large gradients of potential temperature on the boundaries; a point to be pursued in the next section, dedicated to steady fronts.

4. STEADY FRONTS WITH CONDENSATION GENERATED POTENTIAL VORTICITY ANOMALIES

Suppose that the source of potential energy for the growing baroclinic wave in which the front forms is exhausted, so that deformation and shear cease to act on the flow. A steady front remains with no ageostrophic motion, but with geostrophic flow along the front consistent with a low-level potential vorticity maximum as the tracer of latent heating during the active stage. Such is the situation studied in this section. Evidence for such lower tropospheric potential vorticity anomalies and their origin in latent heating has been presented by Hoskins and Berrisford (1988) and Ferris (1989). The FRONTS 87 data set also indicates the presence of such structures.

(a) *Potential vorticity anomalies from time-dependent solutions of the moist semi-geostrophic equations*

Several fronts modified by latent-heat release were studied by making use of the following procedure. The initial state is the most unstable dry Eady mode with the following constants:

$$P_{00} = 0.3 \text{ PVU}, \qquad H = 8 \text{ km}, \qquad H_s = 8 \text{ km}$$

or

$$\theta_0 = 273.3 \text{ K}, \qquad L = 3247 \text{ km or } k = 2\pi/L = 1.94 \, 10^{-6} \text{ m}^{-1}.$$

The source of energy leading to the growth of the basic wave and to frontogenesis is provided by a constant value of baroclinicity $\overline{\partial_Y \theta} = -1$ K/100 km. When some limiting

value of maximum vorticity—about $9f$—is reached, we assume that this gradient vanishes: the solution obtained becomes the steady basic state for the stability analysis. The reader is referred to Emanuel *et al.* (1987) for a discussion of the realism of the structure of these fronts. Each particular front is identified by three parameters: the initial amplitude of the Eady mode, the moist slantwise static stability parameter η, which, according to (10) determines the diabatic source S, and the duration of the time integration. According to the results of section 3, the width, L_2^*, and depth, H_2^*, of the anomaly are amongst the most determinant factors for the stability calculation. They were monitored empirically for the final state. Values of L_2^* and H_2^* were estimated for values of η ranging from 0·1 to 0·01 by decreasing steps of 0·01, and for each η four amplitudes of the initial potential temperature wave were tested, from $\delta\theta = 0·5$ K to $\delta\theta = 2$ K in steps of 0·5 K. About 40 fronts thus generated were examined in some detail. Some idea of the variability of the results is conveyed by the two fronts described by Figs. 10 and 11.
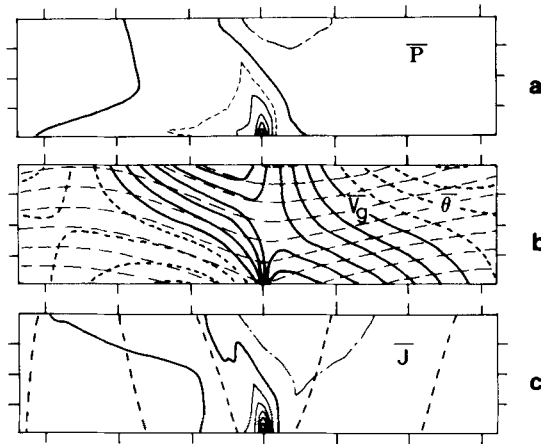

Figure 10. Vertical cross-section of a steady horizontal shear-produced front. It is a solution of the moist semi-geostrophic time-dependent equations. Moist slantwise stability parameter $\eta = 0·07$. This front is hereafter referred to as front 1. Cross-front wavelength 3247 km. Ticks horizontally every 500 km, 2 km vertically. Contour intervals: (a) Potential vorticity $\bar{P}$, 0·2 PVU; reference 0·3 PVU. The dash-dotted contour is 0·2 PVU. The dashed contour, 0·4 PVU. (b) Along-front wind $\bar{v}_g$, 4 m s^{-1}. Negative contours dashed. Potential temperature $\bar{\theta}$, 4 K. (c) Non-dimensional absolute vorticity $\bar{J}$, 0·5f, reference f. Heavy dashed lines are contours of absolute momentum $M = fX$. Contour interval 70 m s^{-1}.
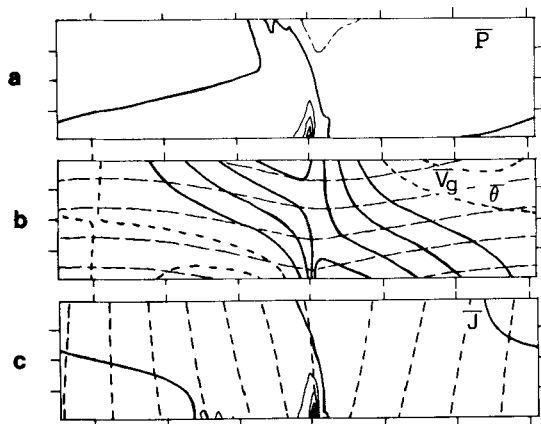

Figure 11. As in Fig. 10, except that the moist slantwise stability parameter η was 0·02 during frontogenesis. This is front 2. Differences from contouring in Fig. 10 are (a) Potential vorticity reference is 0·4 PVU, (b) contour interval for θ is 6 K.

The overall structures of these fronts are rather similar. Yet, as will be shown next, their response to a three-dimensional small perturbation is substantially different. The potential vorticity anomaly, with an amplitude always larger than 1 PVU is maximum at the surface. This is due to the assumption of air saturated in water vapour. Condensation begins as soon as the air ascends. There is a distinct asymmetry in the distributions. Also note how widespread the negative anomaly aloft is, with a very small amplitude. Despite this, the conservation of the total potential vorticity is exact in this model. There is also an obvious correlation between the potential vorticity anomaly and the vorticity itself.

The small differences between the two frontal structures are as follows. The potential temperature field of the first front (Fig. 10, hereafter named 'front 1') compared to the other one (Fig. 11 which will be referred to as 'front 2') has isentropes less affected by frontogenesis. The change in potential temperature is small compared to the dry model having the same magnitude of vorticity. The frontal gradients in the cold air are 1·8 K/ 100 km in front 1 and 0·8 K/100 km in front 2. Another difference of importance is that the width of the anomaly is reduced.

The properties of these fronts are summarized in Table 1. Despite the difference in the initial amplitude, with a corresponding difference in duration before frontal collapse, it is the decrease in the moist slantwise static stability η which is responsible for the rapid increase in vorticity. Also, potential temperature is given less time to change. Both the reduction of L_2^* from about 120 km to 80 km, and the increase of H_2^* from 1000 m to 1400 m is now quantified. At the same time, the potential vorticity has about the same amplitude in the two fronts. The maximum anomaly of θ, however, is nearly halved. Note that, together with the stronger warm-side low-level jet, the positive anomaly of potential temperature (with extremes also at the surface) is larger than the negative anomaly behind. This other loss of symmetry compared to the dry solution is not without consequences for the stability problem. The minimum Richardson numbers are 2·6 and 4·2 respectively. The frontal troughs are about 15 and 8 mb lower than the reference surface pressure (1000 mb).

(b)   Growth rate curves

Bearing in mind the main properties of these fronts, we now examine the results of the stability calculation. On both fronts, two modes could be easily identified.

In Joly and Thorpe (to be published) it is shown analytically that for uniform potential vorticity, the Eady initial condition is unstable to a long-wave mode (3000 km wavelength) only, irrespective of the strength of the frontal zone.

With front 1, the horizontal scale of frontal waves is reached. The second mode dominates the solution, with an e-folding time of one day, a phase-speed along the front of 10 m s^{-1} at a wavelength of 1700 km (Fig. 12). The speed is typical of the basic flow toward the north in the vicinity of the potential-vorticity anomaly. The most unstable mode of front 2 also has a similar e-folding time, a phase speed of 9 m s^{-1}, and a wavelength of less than 800 km (Fig. 13). The anomaly scales in Table 1 suggest these results are exactly in line with the previous calculations. However, we can now relate the scales to the properties of the growing baroclinic wave in which the front formed. A very low slantwise moist stability, for example, which increases the precipitation rate over a narrow frontal ascent zone, leads to the smallest-scale waves. This is a result of some importance in forecasting frontal waves.

Whilst a detailed dynamical understanding of these modes relies on the study of their structure, a discrimination between boundary baroclinic instability and interior instability is useful. It leads to an assessment of the interaction between the two mechanisms.
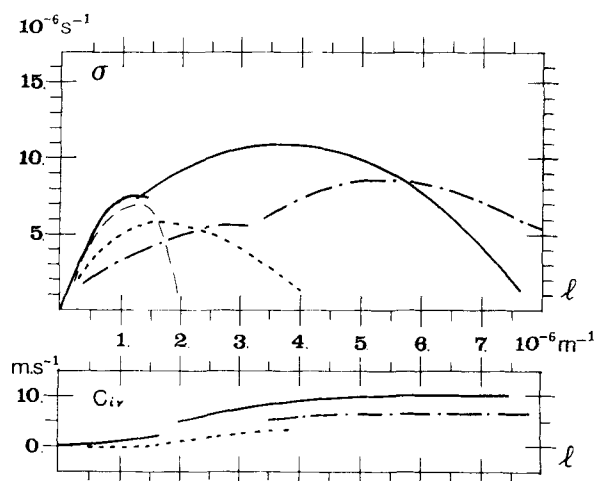
Figure 12. Growth rate σ and along-front phase speed c_{iy} as functions of the along-front wavenumber l for normal-mode perturbations growing along front 1 (Fig. 10) (bold solid line). Other superimposed curves correspond to: front 1 with boundary anomalies of potential temperature θ removed (bold dash-dotted line); front 1 with interior anomaly of potential vorticity removed (bold dotted line); the uniform potential vorticity horizontal shear front having the same mean anomaly of potential temperature θ of 5·9 K as front 1 (light dashed line).
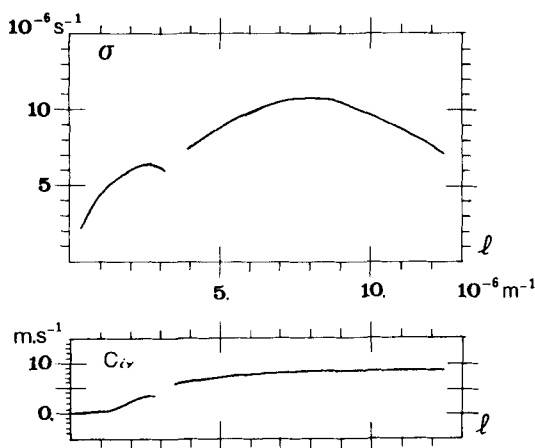

Figure 13. As in Fig. 12 but for front 2 (Fig. 11).

If the potential temperature is artificially set to constant values on horizontal boundaries in front 1, the basic state retains the signature of moist processes in the interior potential vorticity. The situation resembles those of section 3, except with a more complicated anomaly. The cyclonic circulation induced is asymmetric, with along-front wind in the positive branch of more than 12 m s^{-1}, against -5 m s^{-1} in the opposite side. Vorticity still reaches $3f$. The phase tilts of the original wave have nearly disappeared. The e-folding time here is also increased, to 1·4 days (instead of 1). The most unstable wavenumber, on the other hand, is notably increased, from $3·5 \times 10^{-6} \text{m}^{-1}$ to $5·5 \times 10^{-6} \text{m}^{-1}$, reducing the wavelength to 1000 km. The instability of the simplified basic state is quite weak where, in the full problem, it is most intense (Fig. 12).

Consider now the contribution of boundary baroclinic instability, by replacing the anomaly by a uniform potential vorticity. Replacing also the actual distribution of θ by that of the 'horizontal shear front' having the same mean amplitude leads to the thin

TABLE 1. A SUMMARY OF THE MAIN PROPERTIES OF THE FRONTS PRODUCED BY
SOLUTION OF A TIME INTEGRATION OF THE MOIST SEMI-GEOSTROPHIC EQUATIONS

		Front 1	Front 2
Initial	$\delta\bar{\theta}(t = 0)$ K	1·5	0·5
conditions	η	0·07	0·02
Time-integration duration (hours)		30·6	40·2
Potential-vorticity	δP_{max} PVU	1·5	1·2
anomaly	δP_{min} PVU	−0·14	−0·15
generated by	L_2^* km	116	76
condensation	H_2^* m	1011	1409
Maximum vorticity in units of f		9·1	7·1
Surface potential-	$\delta\theta_{max}$ K	6·9	4·2
temperature anomaly	$\delta\theta_{min}$ K	−4·9	−2.8

The initial conditions are the amplitude of the dry Eady mode in terms of surface potential temperature $\delta\bar{\theta}(t = 0)$, and the constant moist static stability parameter η. The potential-vorticity scales L_2^* and H_2^* are to be compared with the Gaussian scales of section 3. The other parameters are self explanatory.

dashed curves of Fig. 12. By the 'horizontal shear front' we mean the geostrophic flow associated with a dry Eady mode; such fronts develop through the horizontal shear mechanism. It is the same curve as the analytical solution of Joly and Thorpe (1990a), but with a reduced amplitude, and consequently a reduced growth rate. However, the warm air is slightly warmer with the actual distribution of $\bar{\theta}$, and the cold side less cold. The heavy dashed curve of Fig. 12 shows that this is sufficient to remove the short wave cut-off of the dry case, and destabilize the larger wavenumbers. An asymmetric surface distribution of potential temperature is sufficient to reduce (slightly) the most unstable wavelength along the front. This property, which leaves a rather simple problem because the interior equation keeps its Laplace form, was recently exploited by Schär and Davies (1990). They proposed this asymmetry as the primary mechanism for generating frontal waves. However, as can be seen here, the growth rate is quite small (an e-folding time of two days), and also, the change in linear wavelength selection is also marginal compared to the full problem.

The point made here is that both boundary and interior baroclinic instability mechanisms are important to create the frontal wave mode, and are strongly interacting with one another. These mechanisms are unified using the ideas of potential vorticity distributions which include boundary thermal gradients.

(c) Mean energetics of the modes

Further properties of the growth rate curves will be discussed later. The remainder of this subsection is dedicated to the detailed identification of the modes. The most direct way is by means of the energetics.

The mean cycles are considered first. Figure 14 shows the energy cycle of the first mode of each front. They are all typical of classical baroclinic waves, with energy going from the basic-state available potential source to perturbation kinetic. As the wavenumber increases, the relative contribution of the basic-state kinetic energy also increases, and even becomes dominant. We can identify the first mode, occurring on the large scale, as a modified classical baroclinic instability or Eady-type mode.

Figure 15 pictures the cycles corresponding to the smaller wavelength modes. They are the most unstable. In front 1, the internal conversion, although positive, appears to be an order of magnitude smaller than the two main conversions. Indeed, the internal term is of the same order as the usually negligible vertical Reynolds stress component $\langle -w'v_g'\partial\bar{v}_g/\partial Z\rangle$, which is −1·3 times the internal conversion. The contribution from the
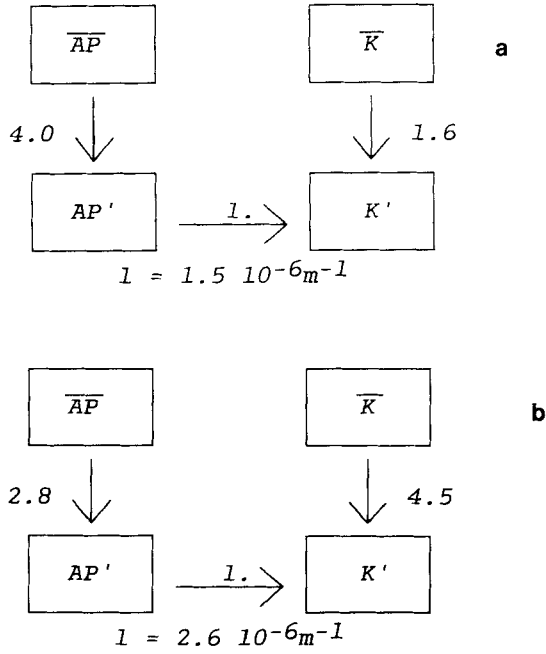
Figure 14. Mean energy cycle of the first mode (most unstable smallest wavenumber l) growing along (a) front 1, (b) front 2. Conventions as in section 3, Fig. 9. Note the unit is different in each figure.
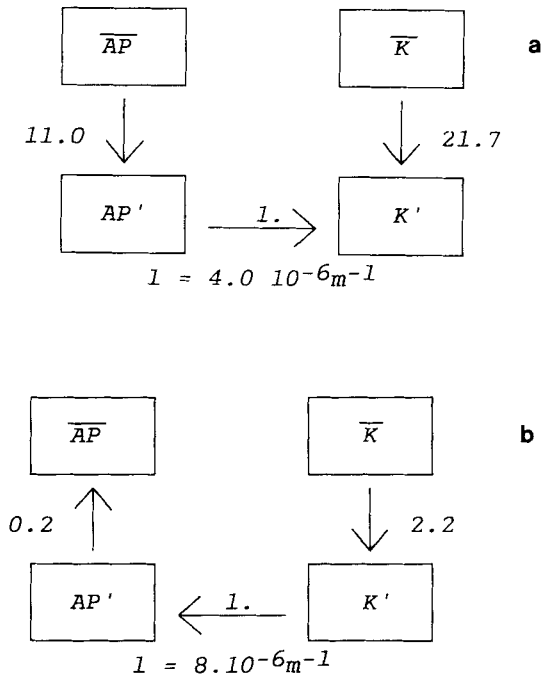

Figure 15. As in Fig. 14, but for the second mode (most unstable largest wavenumber l), for both fronts.

horizontal Reynolds stress $\langle -u'_g v'_g \partial \bar{v}_g / \partial X \rangle$ is 23 times the internal conversion, twice as large as the available energy conversion. The barotropic conversion dominates the dynamics of this mode, where, on the whole, little or no kinetic energy is gained from the basic state baroclinicity. This is even more true with front 2, whose most unstable mode has the same energy cycle as the simple anomalies of section 3. Kinetic energy grows from the barotropic conversion, and the vertical heat flux is negative. In this case, the mode even returns available potential energy to the basic state.

Note that, in front 1, the sign of the internal conversion $\langle w'\theta' g/\theta_0 \rangle$ is also negative for the second mode from l slightly larger than $4 \times 10^{-6}\,\mathrm{m}^{-1}$. In that section of the curve, we also find the characteristic energy cycle of internal baroclinic modes (Fig. 16).
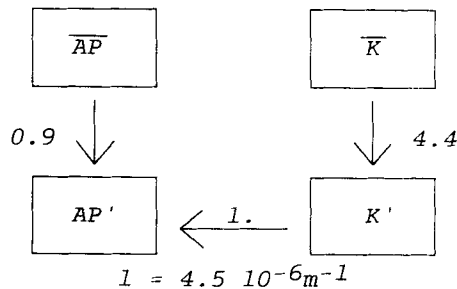

$$1 = 4.5 \; 10^{-6} m^{-1}$$

Figure 16. As in Fig. 15(a), the second mode of front 2 at a larger wavenumber l (with a smaller growth rate). Compare with Fig. 15(b).

(d) Structure of the modes

Cross-sections along the front in geostrophic space will be shown near the location of the maximum potential vorticity. They follow the absolute momentum surface fX crossing the anomaly. In physical space, they are slanted, generally toward the cold air. The cross-section along front 1 reveals that, despite its shorter scale, this frontal wave has strong classical baroclinic features. It is also a deep mode, and u'_g, θ' and w' have the same tilts with height as an Eady wave, except in a thin layer near the surface. Note that the negative correlation between u'_g and v'_g extends over three quarters of the total depth. This structure is somewhat surprising, especially the deepness of the fields on that reduced horizontal scale. It shows how efficient the interaction between the two basic mechanisms of instability can be to produce relatively large-scale features (Fig. 17). Figure 18 shows horizontal sections in physical space (defined by the basic state alone) of mass and temperature at the surface, and the mid-level vertical velocity. A bi-modal overall aspect is present, as in the 'dry' large scale mode. The extreme values are very near the front itself, unlike the 'dry' mode. Vertical velocity aloft is largest coldward of the surface front. This was true of the 'dry' mode. Thus, in the case of front 1, the normal mode also looks like a 'boundary-baroclinic' mode, with reduced horizontal scales, both along and across the front. However, its typical depth is still H: the maximum amplitude of Φ' at $Z = H$ is two-thirds of the amplitude at the surface.

The structure of the instability in front 2 is somewhat different. Along the front, the mode now fills three-eights of the total depth only. The wind components have their largest amplitude at the surface. Largest amplitudes for θ' and w', however, occur at $Z = 1\,\mathrm{km}$. The correlations between each pair of fields are similar to those of the 'pure internal' mode. There is a change, in that these fields are also tilted vertically (Fig. 19). Viewed on horizontal sections, the amplitude is now maximum along the front (no double structure) in both Φ' and θ'. The horizontal tilt is as expected from a perturbation which
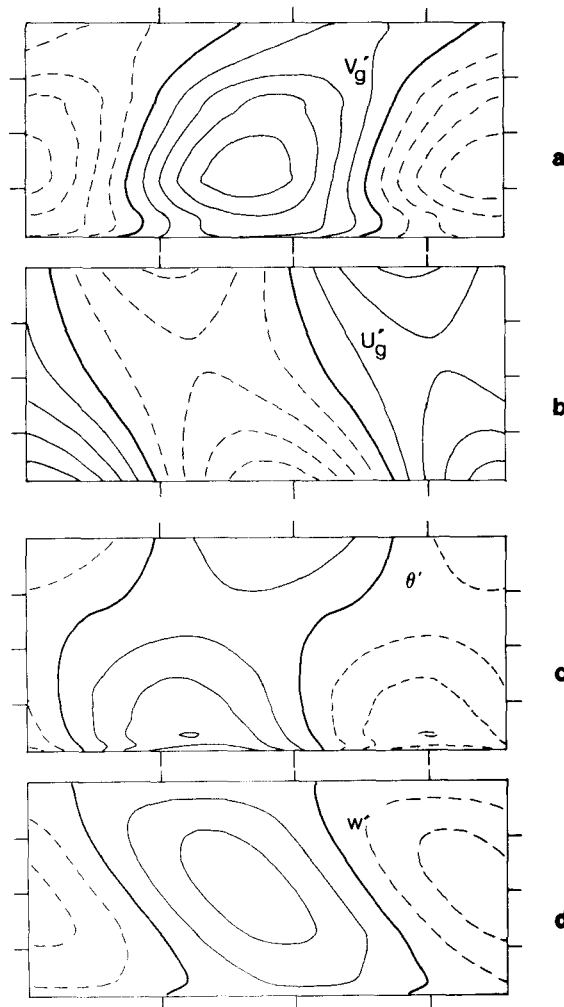
Figure 17. Vertical cross-section of the flow associated with the most unstable normal mode of front 1, with $l = 3.5 \, 10^{-6} \, \text{m}^{-1}$, in the along-front direction Y. The picture is drawn in geostrophic space ($X = 1688 \, \text{km}$). The mode structure alone is shown. Tick marks are set every 500 km horizontally, 2 km vertically. Negative contours are dashed and the zero line is thickened. (a) Geostrophic wind v'_g in the plane. (b) Geostrophic wind u'_g across the plane. (c) Potential temperature perturbation θ'. (d) Vertical velocity w'. Amplitudes are arbitrary. A suggestion of the relative magnitudes is given by the following contour intervals: (a) $0.6 \, \text{m s}^{-1}$, (b) $0.9 \, \text{m s}^{-1}$, (c) $0.3 \, \text{K}$, (d) $0.07 \, \text{cm s}^{-1}$.

is growing mostly from the existing kinetic energy at the front. The vertical velocity at the height of two kilometres still has a double structure, but with a definite maximum on the warm side of the front (the side of the anomaly). The negative correlation between θ' and w' is obvious even at this level. The potential vorticity perturbation is nearly out of phase by $\pi/2$ with the pressure perturbation, so that $u'_g P'$ is maximized, and negative (the flux of potential vorticity is toward the cold air). The pattern of P' and Φ' is such as to disrupt the bandedness of high potential vorticity very efficiently. Above the surface, the pattern of P' is consistent with the double structure of w'. The position of the perturbation potential vorticity close to the mean potential vorticity maximum indicates that frontal waves will be characterized by noticeable positive potential vorticity anomalies with weaker, or zero, negative anomalies. Concentrated at the front, the anomaly of potential vorticity constrains the flow to be nearly parallel to it (Fig. 20).
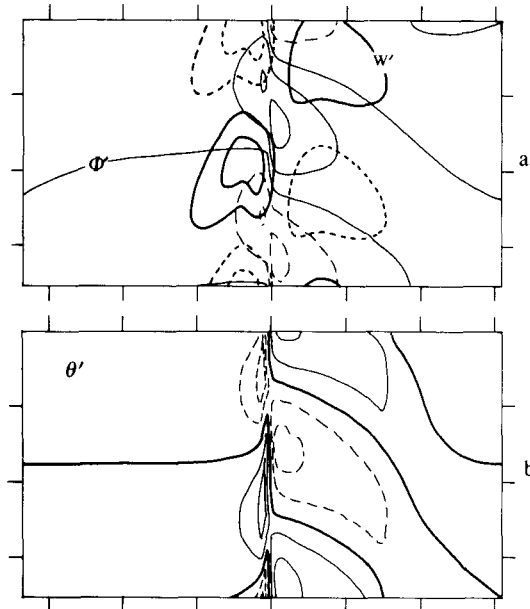
Figure 18. Horizontal cross-section of the flow associated with the most unstable mode of front 1 ($l =$ $3 \cdot 5 \, 10^{-6} \, \text{m}^{-1}$). Picture drawn in physical space as defined by the basic state, although the latter is not superimposed. Tick marks are set every 500 km. Negative contours are dashed. (a) Thin contours, geostrophic potential perturbation Φ' at the surface. Thick lines, vertical velocity w' at mid-level ($Z = 4108$ m). (b) Potential temperature perturbation θ' at the surface. Amplitudes are arbitrary. An idea of the relative magnitudes of perturbations is given by the following contour intervals: (a) Φ', 60 J kg^{-1}; w', 0.14 cm s^{-1}. (b) θ', 0.4 K.

For this mode, a section across the front, near a centre of low pressure at the surface is also shown on Fig. 21. The basic potential vorticity together with isentropes are recalled as only the frontal zone is pictured. This kind of frontal wave appears as a 'warm core' type of cyclone. However, these sections indicate that the 'heating' is really caused by parcels displaced vertically downward.

To summarize all these properties, we now consider the energy conversions averaged over one wavelength along the front. They provide the best representation of the scales of the waves in the across-front plane, with an idea of the amplitude function. The sign can be directly related to phase shifts and sometimes to phase tilts in the along-front plane.

From the selected cross-sections, front 1 did not appear too different from the organization of a large-scale 'dry' baroclinic wave. Differences now appear more neatly in this representation (Fig. 22). While the depth is still large, the maximum amplitude clearly is at the surface front. We observe a shallow mode with a deep but weak vertical extension. Maximum amplitude follows the frontal slope. In the vicinity of the front, a new characteristic appears. Instead of very narrow bands of slightly negative correlations in the $\overline{AP}$ to AP' and AP' to K' fields, we observe wider zones of significantly negative values. The internal conversion $\langle w' \theta' g / \theta_0 \rangle$ has even its maximum amplitude negative in the lower front. We already noted this feature in the pure internal baroclinic mode.

Finally, front 2 is similar in terms of energy conversions but more confined near the lower boundary.

(e)   Discussion of the negative vertical heat flux

Despite the reduction in baroclinicity from front 1 to front 2, the maximum growth rate is always of the same magnitude, having an e-folding time of one day. This forces
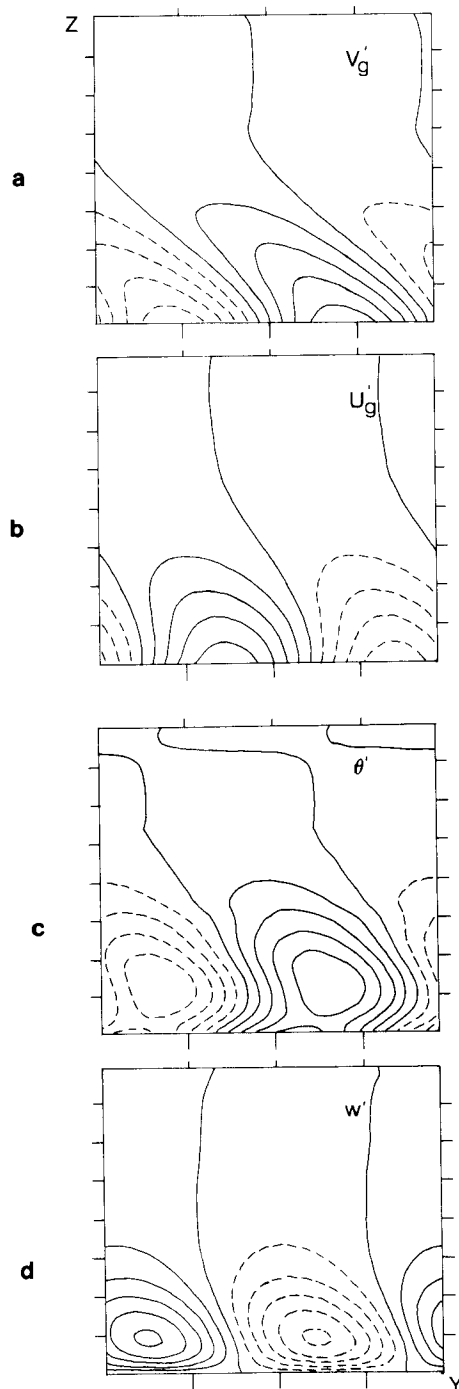
Figure 19. As in Fig. 17, but for the most unstable mode of front 2 ($l = 8 \cdot 10^{-6}$ m, $X = 1916$ km). Ticks are set every 200 km horizontally, 1 km vertically. Suggested contour intervals: (a) $0 \cdot 8$ m s^{-1}, (b) $0 \cdot 8$ m s^{-1}, (c) $0 \cdot 1$ K, (d) $0 \cdot 08$ cm s^{-1}.
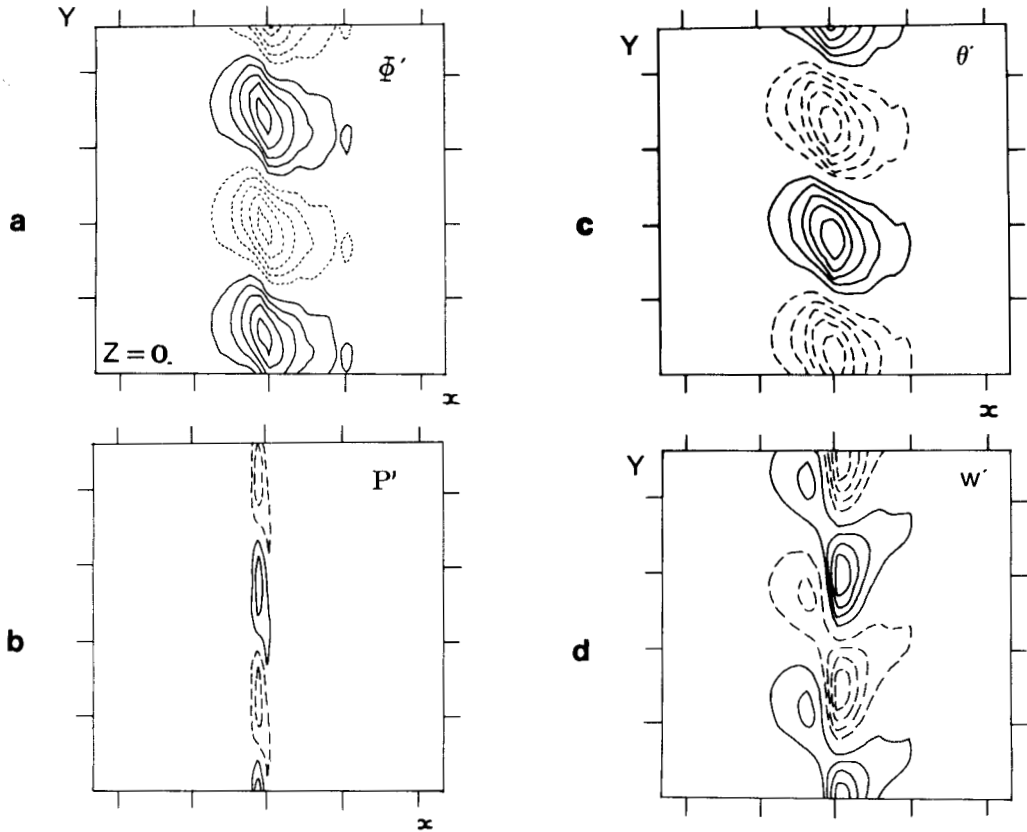
Figure 20. Horizontal cross-section of the flow associated with the most unstable mode of front 2 ($l = 8.10^{-6}\,\mathrm{m}^{-1}$). Pictures drawn in physical space as defined by the basic state, although the latter is not superimposed. Tick marks are set every 200 km. Only 1/4 of the total width in x is shown, the frontal zone itself. Negative contours are dashed. Zero contour omitted. (a) Geostrophic potential perturbation at the surface Φ'. (b) Potential vorticity perturbation P' at the surface. (c) Potential temperature perturbation θ' at $Z = 2051$ m. (d) Vertical velocity w' at the same level. Amplitudes are arbitrary. Suggested contour intervals are: (a) $1\,\mathrm{J\,kg^{-1}}$, (b) 0.05 PVU, (c) 0.01 K, (d) $0.007\,\mathrm{cm\,s^{-1}}$.

the energetics of the mode, as the scales are reduced, to rely more on the mechanism of barotropic instability. Thus, in frontal waves (waves on the scale 1500 km and less), the contribution from the basic state kinetic energy constitutes the main source of energy. It gives to the structure a horizontal tilt with respect to the front.

Also there is a shift between the perturbation fields of θ' and w' leading to a stronger negative correlation between the two in the frontal zone itself, to a negative vertical heat flux at the front, and ultimately, to a negative mean internal conversion. This feature is present in 'pure internal modes', and we now attempt to explain why the frontal waves have this structure.

All our results show that, in the kinetic energy equation, the vertical component of Reynolds stress is negligible. Therefore, it is omitted from this analysis. The simplified energy equations can be written as:

$$\frac{\partial \langle K'_g \rangle}{\partial T} = \int_0^H \int_0^L \left(-\frac{\partial \bar{v}_g}{\partial X} \langle u'_g v'_g \rangle_Y + \frac{g}{\theta_0} \langle w'\theta' \rangle_Y \right) r \frac{dXdZ}{\bar{J}}$$

$$\frac{\partial \langle AP' \rangle}{\partial T} = \int \int \left(-\frac{g}{\theta_0} \frac{\partial \bar{\theta}/\partial X}{\partial \bar{\theta}/\partial Z} \langle u'_g \theta' \rangle_Y - \frac{g}{\theta_0} \langle w'\theta' \rangle_Y \right) r \frac{dXdZ}{\bar{J}}$$
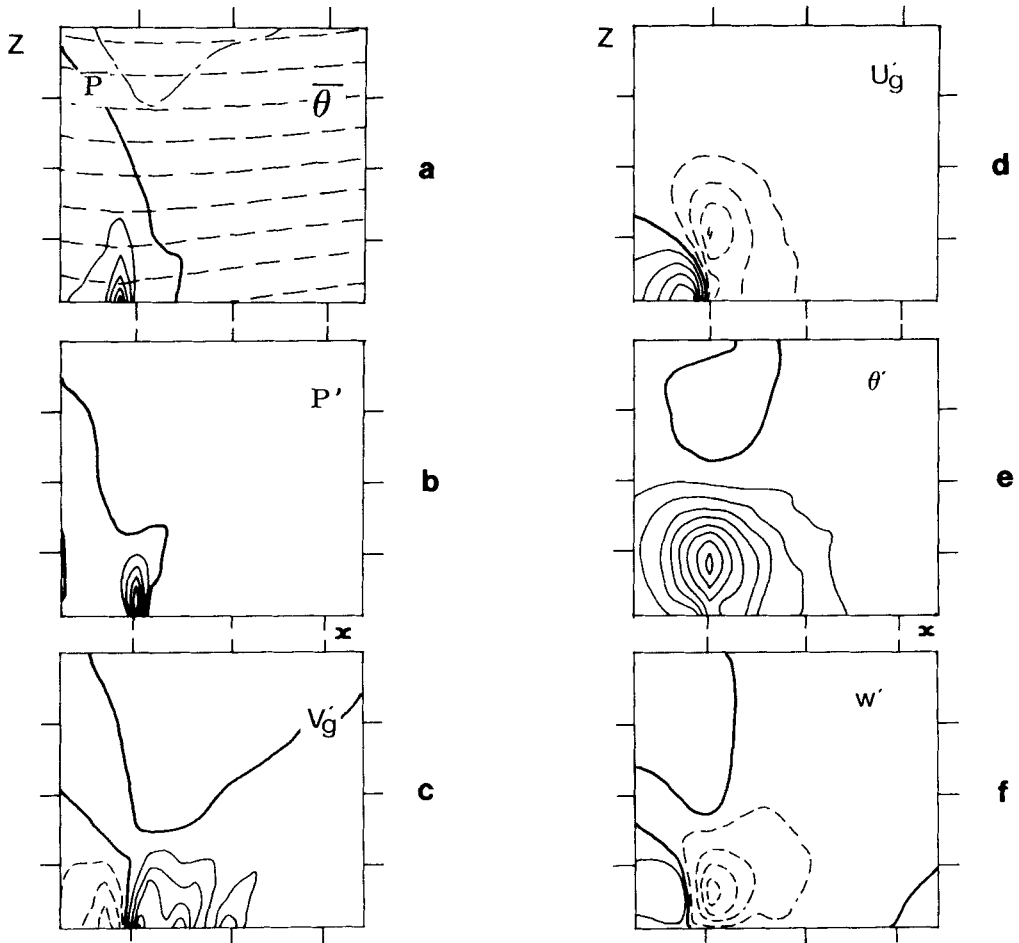
Figure 21.   Vertical cross-section of the flow associated with the most unstable normal mode of front 2, in the cross-front x direction. Pictures drawn in physical space as defined by the basic state alone. Only 1/5 of the total width is shown, the frontal zone. Tick marks are set every 200 km horizontally, 2 km vertically. (a) Recalling the basic state from Fig. 11. Potential vorticity $\overline{P}$, solid lines and dash-dotted line (0·2 PVU). Bolder contour 0·3 PVU, interval, 0·2 PVU. Potential temperature $\overline{\theta}$, dashed lines, contour interval 3 K. Other pictures are sections across a low in the surface potential Φ' (see Fig. 20). Negative contours dashed, zero bolder. (b) Perturbation potential vorticity P'. (c) Geostrophic wind across the plane v'_g. (d) Geostrophic wind in the plane u'_g. (e) Perturbation potential temperature θ'. (f) Vertical velocity w'. Contours are arbitrary. A suggestion of relative magnitudes is given by the following intervals. (b) 0·2 PVU, (c) 0·7 m s^{-1}, (d) 0·3 m s^{-1}, (e) 0·08 K, (f) 0·12 cm s^{-1}, zero added (first thin contours: 0·06 cm s^{-1}).

with

$$\langle K'_g \rangle + \langle AP' \rangle = \int \int \left(\tfrac{1}{2}\langle u'^2_g + v'^2_g \rangle_Y + \frac{g}{\theta_0}\left(\frac{\partial \overline{\theta}}{\partial Z}\right)^{-1} \tfrac{1}{2}\langle \theta'^2 \rangle_Y \right) r \frac{dXdZ}{J}$$

where

$$\langle \ldots \rangle_Y = \frac{1}{L_y} \int_0^{L_y} \ldots dY$$

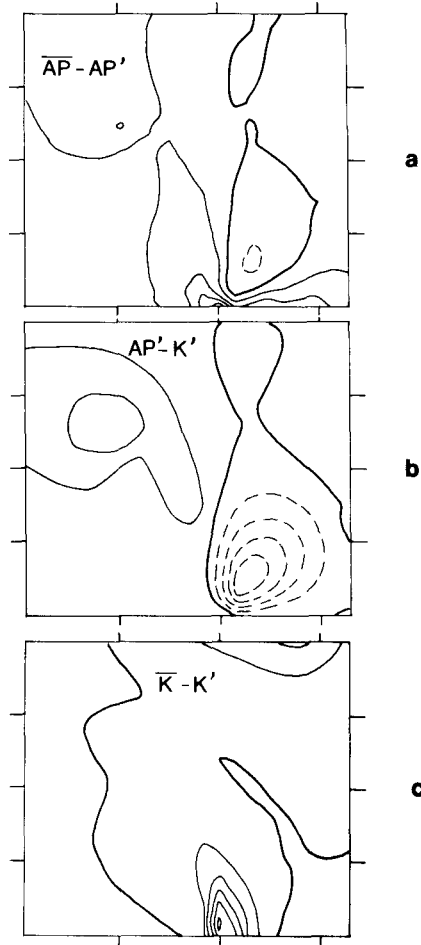(see the Appendix for derivation). Useful functions of X and Z describing the basic state are

Figure 22. Energy conversions over one wavelength in Y in the flow induced by the most unstable normal mode of front 1 ($l = 3.5 \times 10^{-6}\,\mathrm{m}^{-1}$), in a vertical cross-section in the cross-front plane (x, z^*). Pictures are drawn in physical space as defined by the basic state. Only 1/5 of the total width is shown, corresponding to the frontal zone. Ticks horizontally every 200 km, 2 km vertically. Negative contours are dashed, zero contour bolder. (a) Potential energy conversion. (b) Internal conversion. (c) Kinetic energy conversion. Amplitudes are arbitrary. Suggested relative magnitude: (a) 0.4 unit, (b) 0.2 unit, (c) 3 units (Note!). Typical unit $10^{-4}\,\mathrm{J\,kg^{-1}\,s^{-1}}$.

$$N^2 = \frac{g}{\theta_0}\frac{\partial \overline{\theta}}{\partial Z} = \frac{g}{\theta_0}\frac{r\overline{P}}{f\overline{J}}$$

and

$$-\frac{\partial \overline{\theta}/\partial X}{\partial \overline{\theta}/\partial Z} = -\frac{g}{\theta_0}\frac{\partial \overline{\theta}/\partial X}{N^2} = \left(\frac{dZ}{dX}\right)_{\overline{\theta}}. \tag{17}$$

The latter is the slope of the basic state isentropes in geostrophic space.

By definition, the energy of an unstable perturbation increases, and to relate this information to the mode structure, we have to express the correlations. The basic element of the solution is the geostrophic potential, which can be written in two ways,

$$\Phi' = A(X, Z)\cos(lY + \lambda(X, Z)) = (\phi_r(X, Z) + i\phi_i(X, Z))\,\mathrm{e}^{ilY} \tag{18}$$

from which u'_g, v'_g and θ' can easily be derived. For example,

$$\theta' = \frac{\theta_0}{g}\left(\frac{\partial A}{\partial Z}\cos(lY + \lambda) - A\frac{\partial \lambda}{\partial Z}\sin(lY + \lambda)\right) = (\theta_r + i\theta_i)\,e^{ilY}$$

or

$$\theta' = \Theta\cos(lY + \lambda + \mu)$$

$$\Theta = \frac{\theta_0}{g}\left\{\left(\frac{\partial A}{\partial Z}\right)^2 + \left(A\frac{\partial \lambda}{\partial Z}\right)^2\right\} = (\theta_r^2 + \theta_i^2)^{1/2}$$

$$\mathrm{tg}\,\mu = \frac{A\,\partial\lambda/\partial Z}{\partial A/\partial Z} = \frac{\theta_i}{\theta_r}.$$

Taking the Y-transformed equation for θ' leads to the determination of $\hat{w}$. (For definition of $\hat{\ }$ see Eq. 14):

$$\frac{\theta_0}{g}N^2\hat{w} = \left\{(\sigma_i + l\bar{v}_g)\theta_i - \sigma_r\theta_r - \frac{l}{f}\frac{\partial\bar\theta}{\partial X}\phi_i\right\} + i\left\{-(\sigma_i + l\bar{v}_g)\theta_r - \sigma_r\theta_i + \frac{l}{f}\frac{\partial\bar\theta}{\partial X}\phi_r\right\}. \quad (19)$$

This expression can then be combined with θ' to establish the general expression for any balanced two-dimensional steady state of the mean-wave vertical heat flux:

$$\langle w'\theta'\rangle_Y = \frac{1}{2}\left\{-\frac{g}{\theta_0}\sigma_r\frac{\theta_r^2 + \theta_i^2}{N^2} + \frac{l}{f}\frac{\partial\bar\theta/\partial X}{N^2}\left(\phi_r\frac{\partial\phi_i}{\partial Z} - \phi_i\frac{\partial\phi_r}{\partial Z}\right)\right\}.$$

It can easily be shown that the last bracketed term is in fact simply related to the phase function $\lambda(X, Z)$. Indeed,

$$\phi_r\frac{\partial\phi_i}{\partial Z} - \phi_i\frac{\partial\phi_r}{\partial Z} = A^2\frac{\partial\lambda}{\partial Z}.$$

The same kind of calculation can also be done for the other fluxes. Finally, it is found that:

$$\langle u'_g\theta'\rangle_Y = -\frac{\theta_0}{g}\frac{l}{f}\frac{A^2}{2}\frac{\partial\lambda}{\partial Z}$$

$$\langle u'_g v'_g\rangle_Y = -\frac{l}{f}\frac{A^2}{2}\frac{\partial\lambda}{\partial X} \quad (20)$$

$$\langle w'\theta'\rangle_Y = -\frac{g}{\theta_0}\frac{\sigma_r}{2}\frac{\Theta^2}{N^2} + \frac{l}{f}\frac{A^2}{2}\frac{\partial\bar\theta/\partial X}{N^2}\frac{\partial\lambda}{\partial Z}.$$

These expressions directly relate energetics properties to the shape of the potential perturbation Φ', which is also similar in u'_g. A number of rather general conclusions can be made from these properties.

Substituting into the equation for potential energy the fluxes given by (20), or simply making use of the definition of $\langle AP'\rangle$ and of the normal mode form, the rate of change of available potential energy is

$$\frac{\partial\langle AP'\rangle}{\partial T} = \int\int\frac{g^2}{\theta_0^2}\frac{\sigma_r}{2}\frac{\Theta^2}{N^2}r\frac{dX\,dZ}{\bar{J}}. \quad (21)$$

For an unstable mode, the integrand is positive definite. Comparing expressions of the possible sources, $(dZ/dX)_\theta \langle u'_g \theta' \rangle_Y$ and $\langle w' \theta' \rangle_Y$, it is seen that the energy extracted from baroclinicity (first source) is entirely converted to kinetic energy by internal conversion, the latter being also written as

$$\langle w' \theta' \rangle_Y = -\frac{g}{\theta_0} \frac{\sigma_r}{2} \frac{\Theta^2}{N^2} + \left(\frac{dZ}{dX}\right)_{\bar\theta} \langle u'_g \theta' \rangle_Y.$$

This leads directly to consideration of the internal conversion of vertical heat flux of the modes. As clearly shown by the various expressions used, the internal conversion is made up of two terms. The first one is negative definite in an unstable mode and notice that according to (21), it is the opposite of the rate of change of perturbation potential energy. The second is proportional to the baroclinicity. Thus, in all regions where baroclinicity is negligible, internal conversion is negative. If this region is also where the mode amplitude is largest, it is significantly negative.

A typical example of such a zone is an extremum in potential vorticity. Where potential vorticity is extreme, isentropes are quasi-horizontal. In isolated anomalies, the instability is precisely trapped in the same region, hence the sign of the internal conversion in those modes. It also explains the systematic observation, in those modes, of the region of negative internal conversion in moist fronts, at the location of the anomaly. When the scales selected confine the mode to that particular zone, it explains why the mean energy cycle exhibits the reversed internal conversion. In regions where the barotropic conversion $\langle u'_g v'_g \partial \bar v_g / \partial X \rangle$ is the actual source of energy, this negative term represents the necessary contribution to the budget of internal energy associated with quasi-barotropic instability in a stratified fluid. It maintains the vertical stability despite the horizontal redistribution of air parcels.

We consider the local negative heat flux as one of the more unexpected conclusions of this work, inevitably accompanying the identification of frontal instability with horizontal shear instability. Indeed, it can be seen as an observational test of the validity of the theory developed here. Detailed energy budgets of developing frontal waves are scarce in the literature, as mentioned in the introduction. Yet, it has been done by Matsumoto et al. (1970). These authors studied waves growing along the Baiu or Mei-Yu front, between China and Japan. Here, both the large evaporation and heavy precipitation make likely the existence of large values of potential vorticity along this front. The development of waves is seen frequently. The authors mention the shallowness of the young waves and they also estimate the energy conversions inside a finite domain. The budgets revealed that some of the waves under scrutiny have a mean internal conversion which is negative. The authors support this result by mentioning evidence of an 'indirect circulation' in the waves. A similar study involving more cases was recently reported (Zhao 1988). A number of the waves also have a mean negative heat flux. Accordingly, the waves derived energy from the momentum correlations, a property constituting another, more classical, observational test. These independent experimental results directly support the present analysis.

(f) Sensitivity to resolution and to the formulation

Two points need further attention. The first is numerical resolution. Are these relatively narrow potential vorticity anomalies well enough resolved? Fronts 1 and 2 were re-computed and the stability analysis done again with 66 horizontal Fourier modes. This can be done at a reasonable cost. The prediction of phase speed is identical. It is nearly so for growth rate and wavenumber with front 1, where the agreement is excellent. The energy cycle of the most unstable mode reveals also the excellent convergence toward

the most unstable mode (not shown). All conversions are enhanced in the new resolution, but only slightly (5% in the barotropic term).

Front 2, with its extremely narrow anomaly, shows more sensitivity. The growth rate is decreased by 18%, and the most unstable wavenumber by 13%. Moreover, the energy cycle of the most unstable mode (not shown) is also very similar to the one with 33 modes. The barotropic input is reduced to 1·9 times the (negative) internal conversion instead of 2·2. Other figures and signs are the same as in Fig. 15.

With 33 horizontal modes again, the analysis of the stability of both fronts was done again with the length scaling along Y, $1/\bar{J}$, set to unity in Eq. (12). The sensitivity to this particular point is larger than the sensitivity to resolution. As expected from dimensional arguments, the most unstable wavenumber is now reduced to respectively $3 \times 10^{-6} \, \mathrm{m}^{-1}$ and $5 \times 10^{-6} \, \mathrm{m}^{-1}$ (instead of $4 \cdot 5 \times 10^{-6} \, \mathrm{m}^{-1}$ and $8 \times 10^{-6} \, \mathrm{m}^{-1}$). But the growth rate is also decreased, down to an e-folding time of two days for front 2. When the basic-state vorticity is large this term in the semi-geostrophic approximation appears rather important.

5. CONCLUSIONS

Steady fronts whose structure comprises a potential vorticity anomaly generated earlier by latent-heat release have been considered. Several configurations were studied including anomalies resulting from different magnitudes of moist potential vorticity. Anomalies with cross-frontal width smaller than 200 km produced a distinct mode with wavelengths smaller than 1500 km. The frontal boundary gradient of θ was responsible for a large-scale mode (5000 km). It was shown, however, that an asymmetric distribution of potential temperature on the surface with uniform potential vorticity was enough to destabilize marginally smaller wavelengths. But realistic frontal waves appear to require a full representation of the potential vorticity anomaly. The maximum growth rate achieved has an e-folding time of about one day.

It should be borne in mind that, in order to keep the calculation two-dimensional, further condensation was not considered in the perturbation flow; the growth rates are thus underestimated. For a discussion of the stability problem in an evolving front including latent-heat release in the perturbation see Joly and Thorpe (to be published).

The scale of the most unstable mode depends mostly on the cross-frontal width of the anomaly: a narrow anomaly (150 km wide) causes frontal waves with a wavelength in the range 700–900 km. This kind of result is of practical concern. This small width might explain why numerical weather prediction systems have difficulty in forecasting frontal waves. A narrow anomaly results from the most intense frontal convection and precipitation, with very low moist potential vorticity, occurring over a long time. This idea should be tested in a detailed analysis of a frontal case study. The scales involved make it clear that more than conventional data should be included. Data sets such as FRONTS 87 are being studied for such cases.

The structure of the normal modes and the physical nature of the instability have been investigated. An advantage of filtered models lies in the possibility of relating the shape of the modes to energy conversions. The large-scale waves deepening in uniform potential vorticity flows are baroclinic in nature. Energy is converted from the frontal gradient of potential temperature. There is, however, a small contribution from the kinetic energy of the front. Frontal waves on the scale 1000–1500 km can interact with such baroclinic structure. Although the barotropic instability mechanism becomes dominant in the energy cycle, the normal modes have a vertical scale near H, the depth of the troposphere. Perturbation kinetic energy is supplied also from potential energy.

Smaller scale waves are driven by barotropic instability. They start as shallow structures trapped near the surface, with a depth of 3 or 4 km. All waves are characterised by a locally negative vertical heat flux at the front. Observations confirm the existence of frontal waves with negative internal conversion, as predicted here.

Frontal waves have recently received a renewed interest. A simplified version of the present theory, neglecting the interior anomaly of potential vorticity, is proposed by Schär and Davies (1990). Moore and Peltier (1987) describe frontal waves in a uniform potential vorticity flow using fundamentally ageostropic modes. Their linear analysis produces structures which are shallower than ours, with significantly smaller growth rates (e-folding time: two days). Thorncroft (1988) showed that explosively deepening cyclones were the result of the interaction of the upper-tropospheric cut-off and the large surface gradients of potential temperature at a cold front during the last stages of a normal-mode baroclinic wave life-cycle on the sphere. Thus, a finite amplitude non-linear mechanism is advanced, instead of linear amplification of normal modes. However, it seems clear that some of the observed frontal waves may be the result of such a process, but by no means all of them, such as in our initial example. Here we suggest that quasi-two-dimensional fronts which have active moist processes are likely to exhibit the spontaneous development of frontal waves. The mechanism is as outlined in this paper. On occasion one of these waves amplifies rapidly and this development may rely on the finite amplitude feedback described in the work of Thorncroft (1988). The modification of the potential-vorticity distribution in the troposphere by latent-heat release may play a significant role in that feedback as well as in the original instability mechanism.

Finally, the relation to Kleinschmidt's original work (Eliassen and Kleinschmidt 1957) seems clearer if it is realised that he studied growing but finite-amplitude waves. The horizontal extension of the 'producing mass of higher potential vorticity' along the front is not considered by Kleinschmidt. There is in fact no contradiction between his intuition and our results. The distribution of potential vorticity associated with the mode links a positive perturbation to the low (see Fig. 21). On the other hand, the actual triggering of frontal waves is not examined by Kleinschmidt.

<center>APPENDIX</center>

<center>*Energy equation in the linear semi-geostrophic model*</center>

Linearizing the geostrophic momentum equations (2)–(4) gives, in the special case of a steady or stationary frontal flow:

$$\left(\frac{\partial}{\partial T} + \bar{v}_g \frac{\partial'}{\partial Y}\right) u'_g - f v^*_a = 0$$

$$\left(\frac{\partial'}{\partial T} + \bar{v}_g \frac{\partial}{\partial Y}\right) v'_g + u'_g \frac{\partial \bar{v}_g}{\partial X} + f u^{*'}_a = 0$$

$$\left(\frac{\partial}{\partial T} + \bar{v}_g \frac{\partial}{\partial Y}\right) \theta' + u'_g \frac{\partial \bar{\theta}}{\partial X} + w' \frac{r\bar{P}}{f\bar{J}} = 0 \tag{A1}$$

where the following notations from Hoskins and Draghici (1977) were used:

$$u^*_a = u_a + \frac{w}{f} \frac{\partial v_g}{\partial Z} \qquad v^*_a = v_a - \frac{w}{f} \frac{\partial u_g}{\partial Z}.$$

Defining the perturbation kinetic energy as

$$K'_g = 1/2(u'^2_g + v'^2_g),$$

we can write its rate of change as

$$\left(\frac{\partial}{\partial T} + \bar{v}_g \frac{\partial}{\partial Y}\right) K'_g + u'_g v'_g \frac{\partial \bar{v}_g}{\partial X} - f u'_g v^{*'}_a + f v'_g u^{*'}_a = 0.$$

The total kinetic energy in a wavelength of the mode perturbation is:

$$\langle K'_g \rangle = \int_0^L \int_0^{L_v} \int_0^H K'_g r\, dx\, dy\, dz^*. \tag{A2}$$

The wavelengths are not affected by the change of coordinate. The volume element, however, becomes

$$dx\, dy\, dz^* = 1/\bar{J}\, dX\, dY\, dZ.$$

Thus,

$$\frac{\partial \langle K'_g \rangle}{\partial T} + \int\int\int \frac{\partial}{\partial Y}\left(\frac{\bar{v}_g K'_g}{\bar{J}}\right) r\, dX dY dZ + \int\int\int u'_g v'_g \frac{\partial \bar{v}_g}{\partial X} r \frac{dX dY dZ}{\bar{J}} -$$

$$- \int\int\int f u'_g v^{*'}_a r \frac{dX dY dZ}{\bar{J}} + \int\int\int f v'_g u^{*'}_a r \frac{dX dY dZ}{\bar{J}} = 0.$$

From (A1), we linearize u^*_a and v^*_a:

$$u^{*'}_a = u'_a + \frac{w'}{f} \frac{\partial \bar{v}_g}{\partial Z} = u' - u'_g + \frac{w'}{f} \frac{\partial \bar{v}_g}{\partial Z}$$

$$v^{*'}_a = v'_a = v' - v'_g.$$

Substituting in the last two integrals, we find another conversion term from the basic state kinetic energy:

$$\int\int\int v'_g w' \frac{\partial \bar{v}_g}{\partial Z} r \frac{dX dY dZ}{\bar{J}}$$

and the remaining integral is

$$C = \int_0^L \int_0^{L_y} \int_0^H (-fu_g'v' + fv_g'u')r\,dxdydz^*$$

which we express in physical space. Replacing u_g' and v_g' by their definition, and integrating by parts, we find

$$C = \int\int [\phi'v']_0^{L_y} r\,dxdz^* + \int\int [\phi'u']_0^L r\,dxdz^* - \int\int\int \phi'\left(\frac{\partial ru'}{\partial x} + \frac{\partial rv'}{\partial y}\right) dxdydz^*.$$

The integrated terms vanish in all the configurations of boundary conditions discussed in section 2, and the continuity equation implies that

$$C = +\int\int\int \phi' \frac{\partial rw'}{\partial z^*}\,dxdydz^*.$$

Another integration by parts, making use of the boundary conditions on w' at $Z = 0$, H, and the hydrostatic equation gives

$$C = \int\int\int -\frac{g}{\theta_0} w'\theta'\,dxdydz^*.$$

The kinetic energy equation may finally be written as:

$$\frac{\partial \langle K_g' \rangle}{\partial T} = -\int\int\int \left(\frac{\partial \bar{v}_g}{\partial X} u_g'v_g' + \frac{\partial \bar{v}_g}{\partial Z} w'v_g'\right)r\,\frac{dXdYdZ}{\bar{J}} \qquad (A3)$$

$$+ \int\int\int \frac{g}{\theta_0} w'\theta'r\,\frac{dXdYdZ}{\bar{J}}.$$

The first integral corresponds to the horizontal and vertical Reynolds stress terms, and expresses the conversion from the mean kinetic energy $\overline{K}$ to the wave kinetic energy K_g'. Multiplying throughout the thermodynamic equation by $(g/\theta_0)(f/r)(\bar{J}/\bar{P})\theta'$ and integrating in a similar fashion gives

$$\frac{\partial \langle AP' \rangle}{\partial T} = -\int\int\int \frac{g}{\theta_0} \frac{f\bar{J}}{r\bar{P}} \frac{\partial \bar{\theta}}{\partial X} u_g'\theta'r\,\frac{dXdYdZ}{\bar{J}} - \int\int\int \frac{g}{\theta_0} w'\theta'r\,\frac{dXdYdZ}{\bar{J}} \qquad (A4)$$

where the wave total potential enegy is defined as

$$\langle AP' \rangle = \int_0^L \int_0^{L_y} \int_0^H \frac{g}{\theta_0} \frac{\bar{\xi}}{r\bar{P}} \frac{1}{2} \theta'^2 r'\,dxdydz^*. \qquad (A5)$$

From Eq. (A4), we see that the last integral in (A3) and (A4), involving the correlation between w' and θ', represents the internal conversion between the wave potential energy $\langle AP' \rangle$ and the wave kinetic energy $\langle K' \rangle$ showing the important energetical interpretation of the direct circulation. The other integral in (A4), a weighted mean horizontal heat-flux, is responsible for the conversion of mean available potential energy $\overline{AP}$ into wave available potential energy AP'. It is the main energy supply in simple baroclinic waves.

These equations are slightly approximate because in the transformation of the conversion C, we have loosely exchanged the order of integrations, whereas strictly speaking, the vertical integration and horizontal ones are functions of one another, in that the limits $Y = 0$ and $Y = L_y$ depend on z^* if $y = 0$ and $y = L_y$ do not.

The energy equations in semi-geostrophic linear theory look very similar to the quasi-geostrophic energy equations. In the kinetic energy equations, the vertical momen-

tum flux is one addition. We have found it to be negligible in most cases. In the wave potential energy definition and equation, the weighting factor is proportional to $\bar{\zeta}/\bar{P}$, instead of to the static stability. Notice that all the integrals involved could be defined in physical space. We will evaluate them in geostrophic space, but interpret the result in physical space.

REFERENCES

Bennetts, D. A. and Hoskins, B. J.	1979	Conditional symmetric instability—a possible explanation for frontal rainbands. *Q. J. R. Meteorol. Soc.*, **105**, 945–962
Bjerknes, J.	1919	On the structure of moving cyclones. *Geofys. Publ.* **1**, 1
Bjerknes, J. and Solberg, H.	1922	Life cycle of cyclones and the polar front theory of atmospheric circulation. *ibid.* **3**, 1
Blumen, W. and Gross, B. D.	1987	Semigeostrophic flow over topography in a stratified rotating atmosphere. Part I: steady three-dimensional solutions over finite ridges. *J. Atmos. Sci.*, **44**, 3007–3019
Brown, J. A.	1969	A numerical investigation of hydrodynamic instability and energy conversions in the quasi-geostrophic atmosphere. Part I. *ibid.*, **26**, 352–365
Browning, K. A. and Harrold, T. W.	1969	Air motion and precipitation growth in a wave depression. *Q. J. R. Meteorol. Soc.*, **95**, 288–309
Charney, J. G.	1947	The dynamics of long waves in a baroclinic westerly current. *J. Meteorol.*, **4**, 135–162
Charney, J. G. and Stern, M. E.	1962	On the stability of internal baroclinic jets in a rotating atmosphere. *J. Atmos. Sci.*, **19**, 159–162
Cullen, M. J. P., Norbury, J., Purser, R. J. and Shutts, G. J.	1987	Modelling the quasi-equilibrium dynamics of the atmosphere. *Q. J. R. Meteorol. Soc.*, **113**, 735–757
Eady, E. T.	1949	Long-waves and cyclone waves. *Tellus*, **1**, (3), 33–52
Eliasen, E.	1960	On the initial development of frontal waves. *Pub. Det. Danske Meteorol. Inst.*, N° **13**, 107
Eliassen, A.	1948	The quasi-static equations of motion. *Geofys. Publ.*, **17**, 3
Eliassen, A. and Kleinschmidt, E.	1957	*Dynamic Meteorology. Handbuch der physik*, vol. **48**, Springer-Verlag, Berlin
Emanuel, K. A.	1979	Inertial instability and mesoscale convective systems. Part I: linear theory of inertial instability in rotating viscous fluids. *J. Atmos. Sci.*, **36**, 2425–2449
	1985	Frontal circulations in the presence of small moist symmetric stability. *ibid.*, **42**, 1062–1071
Emanuel, K. A., Fantini, M. and Thorpe, A. J.	1987	Baroclinic instability in an environment of small stability to slantwise moist convection. Part I: two-dimensional models. *ibid.* **44**, 1559–1573
Ferris, P. D.	1989	'Frontal structure in a mesoscale model'. PhD thesis, University of Reading
Hoskins, B. J.	1974	The role of potential vorticity in symmetric stability and instability. *Q. J. R. Meteorol. Soc.*, **100**, 480–482
	1975	The geostrophic momentum approximation and the semi-geostrophic equations. *J. Atmos. Sci.*, **32**, 233–242
	1976	Baroclinic waves and frontogenesis. Part I: introduction and Eady waves. *Q. J. R. Meteorol. Soc.*, **102**, 103–122
	1982	The mathematical theory of frontogenesis. *Ann. Rev. Fluid Mech.*, **14**, 131–151
Hoskins, B. J. and Berrisford, P.	1988	The storm of 15–16 October 1987. A potential vorticity perspective. *Weather*, **43**, 122–129
Hoskins, B. J. and Bretherton, F. P.	1972	Atmospheric frontogenesis models: mathematical formulation and solution. *J. Atmos. Sci.*, **29**, 11–37
Hoskins, B. J. and Draghici, I.	1977	The forcing of ageostrophic motion according to the semi-geostrophic equations and in an isentropic coordinate model. *ibid.*, **34**, 1859–1867
Hoskins, B. J. and West, N. V.	1979	Baroclinic waves and frontogenesis. Part II: Uniform potential vorticity jet flows—cold and warm fronts. *ibid.*, **36**, 1663–1680
Hoskins, B. J., McIntyre, M. E. and Robertson, R. W.	1985	On the use and significance of isentropic potential vorticity maps. *Q. J. R. Meteorol. Soc.*, **111**, 877–946

James, I. N. and Hoskins, B. J. — 1985 — Some comparisons of atmospheric internal and boundary baroclinic instability. *J. Atmos. Sci.*, **42**, 2142–2155

Joly, A. and Thorpe, A. J. — 1989 — Warm and occluded fronts in two-dimensional moist baroclinic instability. *Q. J. R. Meteorol. Soc.*, **115**, 513–534

Killworth, P. D. — 1980 — Barotropic and baroclinic instability in rotating stratified fluids. *Dyn. Atmos. Ocean*, **4**, 143–184

Kotschin, N, — 1932 — Über die Stabilität von Margulesschen Diskontinuitätsflachen. *Beitr. Phys. Atmos.*, **5**, 129–164

Kuo, H. L. — 1949 — Dynamic instability of two-dimensional non-divergent flow in a barotropic atmosphere. *J. Meteorol.*, **6**, 105–122

Margules, M. — 1905 — Ueber die Energie der Stürme. *Jahrbuch der K.-K. Zentralanstalt für Meteorologie und Erdmagnetismus*, Year 1903. New series vol. **40** (Full series vol. **48**), Vienna 1905, Appendix pp. 1–26

— 1906 — Ueber Temperaturschichtung in stationär bewegter und in ruhender Luft. Hann-Vol., *Met. Zeit.*, 243–254

Matsumoto, S., Yoshizumi, S. and Takeuchi, M. — 1970 — On the structure of the 'Baiu-Front' and the associated intermediate scale disturbances in the low atmosphere. *J. Met. Soc. Japan*, **48**, 479–491

Moore, G. K. W. and Peltier, W. R. — 1987 — Cyclogenesis in frontal zones. *J. Atmos. Sci.*, **44**, 384–409

Orlanski, I. — 1968 — Instability of frontal waves. *ibid.*, **25**, 178–200

Schär, C. and Davies, H. C. — 1990 — An instability of mature cold fronts. *ibid.*, (in press)

Sinton, D. M. and Mechoso, C. R. — 1984 — Nonlinear evolution of frontal waves. *ibid.*, **41**, 3501–3517

Solberg, H. — 1928 — Integrationen der atmosphärischen Störungsgleichungen. *Geofys. Publ.*, **5**, 9

Thorncroft, C. — 1988 — 'Frontal waves'. PhD Thesis, Department of Meteorology, University of Reading

Thorpe, A. J. and Emanuel, K. A. — 1985 — Frontogenesis in the presence of small stability to slantwise convection. *J. Atmos. Sci.*, **42**, 1809–1824

Zhao, S. — 1988 — The energetics of cyclogenesis on Mei-Yu (Baiu) front. Pp. 205–208 in preprint. *Palmén Memorial Symposium on extratropical cyclones*. Helsinki. American Meteorological Society