

The counter-propagating Rossby-wave perspective on baroclinic instability. Part IV: Nonlinear life cycles

By J. METHVEN^{1*}, B. J. HOSKINS¹, E. HEIFETZ² and C. H. BISHOP³

¹*University of Reading, UK*

²*Tel-Aviv University, Israel*

³*Naval Research Laboratories/UCAR, Monterey, USA*

(Submitted 16 February 2004; revised 28 September 2004)

SUMMARY

Pairs of counter-propagating Rossby waves (CRWs) can be used to describe baroclinic instability in linearized primitive-equation dynamics, employing simple propagation and interaction mechanisms at only two locations in the meridional plane—the CRW ‘home-bases’. Here, it is shown how some CRW properties are remarkably robust as a growing baroclinic wave develops nonlinearly. For example, the phase difference between upper-level and lower-level waves in potential-vorticity contours, defined initially at the home-bases of the CRWs, remains almost constant throughout baroclinic wave life cycles, despite the occurrence of frontogenesis and Rossby-wave breaking. As the lower wave saturates nonlinearly the whole baroclinic wave changes phase speed from that of the normal mode to that of the self-induced phase speed of the upper CRW. On zonal jets without surface meridional shear, this must always act to slow the baroclinic wave. The direction of wave breaking when a basic state has surface meridional shear can be anticipated because the displacement structures of CRWs tend to be coherent along surfaces of constant basic-state angular velocity, \bar{U} . This results in up-gradient horizontal momentum fluxes for baroclinically growing disturbances. The momentum flux acts to shift the jet meridionally in the direction of the increasing surface \bar{U} , so that the upper CRW breaks in the same direction as occurred at low levels.

KEYWORDS: Home-bases Self-induced phase speed Wave breaking Zonal propagation

1. INTRODUCTION

A major limitation of linear instability theory is that it only describes small-amplitude waves. Clearly weather systems involve strong nonlinearities, such as frontogenesis and Rossby-wave breaking, resulting in the stretching and folding of potential-vorticity (PV) contours on isentropic surfaces. It is of considerable interest whether any predictions by linear theory are robust as waves saturate and break nonlinearly. In addition, the direction of wave breaking in baroclinic wave life cycles has been shown to be strongly dependent on surface meridional shear in the basic state (e.g. Simmons and Hoskins 1980; Davies *et al.* 1991; Thorncroft *et al.* 1993). Balasubramanian and Garner (1997) (hereafter BG97) proposed that the most important determinant of the direction of breaking is the distribution of normal-mode (NM) horizontal momentum flux, the associated modification of the zonal flow and its positive feedback on the momentum flux—the nonlinear barotropic governor effect (James 1987; Nakamura 1993).

As discussed by Thorncroft *et al.* (1993) (hereafter THM), the nature of wave breaking has a profound influence on wave-mean flow interaction. They studied two baroclinic wave life cycles. LC1 develops from the fastest growing NM on a zonal jet without surface meridional shear (state Z1) resembling the mid-latitude jet in a storm-track region. Cyclonic PV spirals form on isentropic surfaces intersecting the tropopause poleward of the jet axis while, on the equatorward side, breaking results in large PV filaments that are wrapped around each anticyclone (see Fig. 2). These PV filaments stretch and thin exponentially, therefore having less effect on the wind field and dissipating more quickly as time progresses. As wave activity dissipates, it leaves an

* Corresponding address: Department of Meteorology, University of Reading, PO Box 243, Earley Gate, Reading RG6 6BB, UK. e-mail: J.Methven@reading.ac.uk

almost zonal state with a stronger more barotropic jet. The second jet, Z2, is constructed by adding barotropic shear that is cyclonic in the mid-latitudes to Z1. Wave breaking in the LC2 experiment is entirely cyclonic and results in the formation of a large-scale spiral in PV at all levels (see Fig. 3). In this case, the large-scale cyclones that are formed dissipate extremely slowly, since the mixing of PV filaments within a spiral does not influence its circulation and zonalization of the flow does not occur.

Heifetz *et al.* (2004a) (hereafter Part I) showed how the mechanism of baroclinic instability on general zonal jets can be viewed as arising from the interaction between pairs of counter-propagating Rossby waves (CRWs). Each CRW pair can be constructed from a growing NM and its decaying complex conjugate. Methven *et al.* (2005) (hereafter Part III) generalized CRW theory to describe linearized primitive-equation dynamics on a rotating sphere. Equations were obtained which describe the evolution of the CRWs' amplitudes and phases. The same CRW equations can be obtained by two methods. In the first method, two 'home-bases' are chosen and the CRW propagation and interaction is described exactly by considering only the PV equation at these home-bases (ϕ_1, θ_1) and (ϕ_2, θ_2). This method is appealing because it quantifies the 'PV-thinking' mechanism for baroclinic growth of Bretherton (1966) and Hoskins *et al.* (1985). However, the choice of home-bases is subjective. In the second method, two orthogonality constraints derived from global wave-activity conservation laws are used to obtain CRWs objectively from a growing NM. The disadvantage is that the coefficients in the CRW equations then relate to global integrals rather than the properties of the basic state at two home-bases. However, it was shown that the two methods give very similar results for realistic jets if the choice of home-bases is motivated by the structure of the orthogonal CRWs. Consequently, the 'PV-thinking' mechanism applies to the orthogonal CRWs to a good approximation. The reader should refer to section 2 of Part III for the primitive-equation definitions of CRWs and their 'home-base winds' (U_1 and U_2), 'self-induced phase speeds' (c_1^1 and c_2^2), interaction strength and the CRW evolution equations.

In Part III, the CRW structures associated with baroclinically growing disturbances on the two zonal jets, Z1 and Z2, were analysed. It was shown how the slant* of the PV structures of CRWs expected in an environment with barotropic shear determines the horizontal momentum flux of the NM.

Here the LC1 and LC2 nonlinear baroclinic wave life cycles developing from the zonal wave-number 6 NMs on these two basic states are re-examined in the light of CRW theory. In particular, the phase speeds of wave crests and troughs of the PV contours corresponding to the upper and lower CRWs are investigated throughout the life cycle as frontogenesis happens at the ground followed by wave breaking at upper levels.

2. LOCATING THE HOME-BASES OF ORTHOGONAL CRWS FROM LINEAR THEORY

The natural reference locations for the upper-level and lower-level waves in a baroclinic life cycle are given by the home-bases determined from the orthogonal CRWs. During the small-amplitude stage of exponential growth, it is expected that the phase difference between PV waves at these reference locations will approximately equal the CRW phase difference calculated by the orthogonality method and they will both move with the phase speed of the NM.

The home-base latitude of each CRW can be identified with the latitude of its maximum meridional wind (v) because v has much less vertical variation than the

* The term 'slant' is used exclusively to describe structures in the y - z plane, as opposed to 'vertical tilt' in the x - z plane and 'horizontal tilt' in the x - y plane.

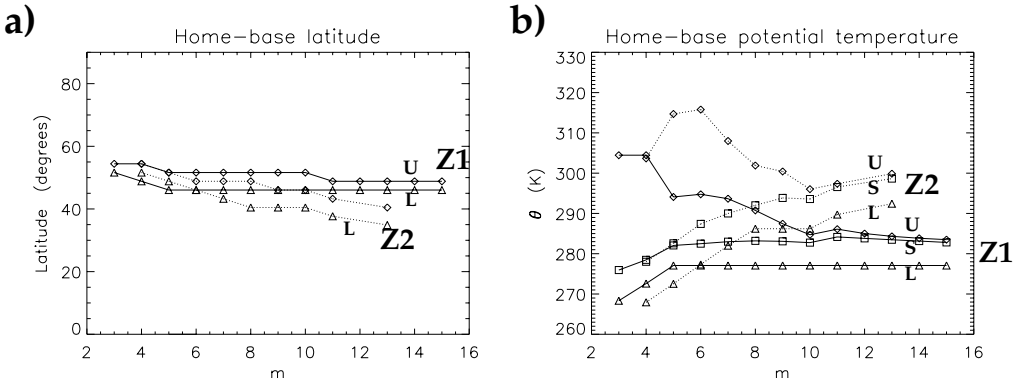


Figure 1. (a) Home-base latitudes, identified using the maximum v for each counter-propagating Rossby wave (CRW), versus zonal wave-number m . Triangles mark the latitude of the lower CRW, ϕ_1 , and diamonds mark the upper CRW, ϕ_2 . Solid lines are used for zonal jet Z1 and dotted lines for Z2. (b) Home-base potential temperatures, defined by $\bar{\theta}(\phi_1)$ at the ground for the lower CRW (L) and $\bar{\theta}(\phi_2)$ at the height where the zonal flow matches the home-wind of the upper CRW (U). The steering level (S) at latitude $(\phi_1 + \phi_2)/2$ is located where the zonal flow matches the phase speed of the NM (square symbols).

PV structure and less slant. The results for the fastest growing modes (referred to as branch 1 modes in Part III) on both jets is shown in Fig. 1(a). For the Z1 jet ($m > 4$), the lower CRW is anchored to the ground at the latitude of the maximum $\bar{\theta}$ -gradient (46°N). However, the Z2 modes clearly shift equatorwards with wave number*. For the Z1 modes, the v anomalies of the upper CRW are centred on the same latitude as the lower CRW at every level (Figs. 3(a) and (c) of Part III) and the difference in home-base latitude only occurs because both v structures slant polewards in the upper troposphere along isentropic surfaces. For the Z2 modes, the v anomalies of the upper CRW lie poleward of those of the lower CRW at all levels (Figs. 10(a) and (c) of Part III).

The basic-state potential temperature ($\bar{\theta}$) associated with the home-base of each CRW is shown in Fig. 1(b). For the lower CRW, θ_1 corresponds to the surface $\bar{\theta}$ at the latitude shown in Fig. 1(a). For the upper CRW, θ_2 is the value of $\bar{\theta}$ where the zonal angular velocity \bar{U} matches its home-base wind U_2 (defined by global integral (21) in Part III) searching upwards at latitude ϕ_2 . If $\bar{U} < U_2$ at all heights then $\bar{\theta}$ at the level of maximum wind is selected. Similarly, the value of $\bar{\theta}$ at the steering level is found by searching upwards for $\bar{U} = c_r$ at the latitude $(\phi_1 + \phi_2)/2$.

For the Z1 jet, the approach with increasing wave number of the upper CRW towards the ground (decreasing θ_2) is clear (Fig. 1(b)). Moreover, at high wave numbers the upper CRW's home-base converges onto the steering level, as anticipated from CRW arguments for Charney-mode behaviour (see Section 4(c) of Heifetz *et al.* 2004b—hereafter Part II). The NM phase speed varies weakly with wavelength because, to some extent, the increase in the self-induced eastward propagation rate of the lower CRW offsets the increase in westward propagation rate of the upper CRW. The CRWs do not shift meridionally, so that the steering level remains in the range $282 \leq \theta \leq 285$ K, or only just above the 800 hPa surface.

For the Z2 jet, the equatorward shift of the lower CRW with wave number (Fig. 1(a)) results in an increase in θ_1 (Fig. 1(b)). Since the NM phase speed hardly varies with

* The fastest growing NMs on jet Z2 with wave-numbers 2, 3 and 12 are from branch 2 rather than branch 1 (see Part III). CRWs were only calculated for the fastest-growing NM at each wave number so that the branch 1 CRWs for $m = 2, 3, 12$ were not found and are, therefore, absent from Fig. 1.

wave number and the tropospheric static stability varies little with latitude, the θ spacing between the steering level and the lower CRW is almost constant. For wave-number 6 the steering level at 46°N is at 288 K (Fig. 1(b)), close to 700 hPa.

3. NONLINEAR DEVELOPMENT OF BAROCLINIC WAVES

CRW structures are obtained from a linear stability analysis, where the assumption is that meridional displacements are small relative to the scale of variations in the basic state. Here, it is investigated whether the development can still be described by the zonal propagation and mutual interaction of two CRWs as the baroclinic wave attains large amplitude and wave breaking occurs.

(a) *Tracking the crests and troughs of home-base reference contours*

The approach taken here is to track the nonlinear evolution of material contours whose Lagrangian coordinates are given by the home-bases of the upper and lower CRWs from the linear analysis. The θ_1 contour corresponding to the lower CRW's home-base (Fig. 1(b)) is tracked on the lowest model level throughout the baroclinic wave life cycle. The home-base of the upper CRW is associated with the θ_2 isentropic surface at the latitude ϕ_2 . The basic-state PV at this location is found and this PV contour, P_2 , is tracked on the θ_2 surface. During the early stages of growth, both θ and PV are conserved and these two contours are material contours. As the lower and upper waves saturate nonlinearly, the associated fronts collapse to scales close to the model resolution and the hyper-diffusion (included for numerical stability) dissipates temperature, divergence and vorticity. As a result, θ and PV are no longer conserved. However, by tracking the displacement of the same θ and PV contours, the wave activity relative to the slowly evolving modified Lagrangian-mean zonally-symmetric state (McIntyre 1980) is followed, as described exactly by the nonlinear extension to the conservation of pseudomomentum by Haynes (1988).

The baroclinic-wave life cycles were initiated with the fastest-growing NMs at wave-number $m = 6$. The primitive-equation spectral-transform model of Hoskins and Simmons (1975) was run at T341L30 resolution imposing six-fold zonal and hemispheric symmetry (results from these high-resolution life cycles have been shown previously by Methven and Hoskins (1998, 1999) and Shapiro *et al.* (1999)). Figure 1(b) shows that the home-base of the lower CRW is at the same potential temperature (277 K) for both the LC1 and LC2 experiments. The contour $\theta_1 = 280$ K was tracked on the lowest model level. In the LC1 case, the home-base contour of the upper CRW was identified with $P_2 = 1$ PVU* on the isentropic surface $\theta_2 = 300$ K. In the LC2 case, the contour $P_2 = 2$ PVU was tracked on the isentropic surface $\theta_2 = 320$ K. In both cases the isentropic surface chosen lies a few Kelvin above the home-base value identified in Fig. 1(b).

Figure 2 shows the evolution of LC1 after the weather system has reached large amplitude. The shading marks the upper CRW's reference contour (P_2) and the bold line marks the lower CRW (θ_1). By day 6.25, strong cold and warm fronts have formed and occlusion is occurring: the northern tip of the warm sector is just about to be pinched off (at point O) leaving a sharp crest in the θ_1 contour. The upper-level positive PV anomaly clearly lies to the west of the warm sector (the lower CRW's positive boundary PV anomaly). Cyclonic turning is seen north of 46°N in both waves. By day 7.25, the upper CRW has clearly started breaking anticyclonically south of 46°N so that the trough

* Potential vorticity unit. 1 PVU = 10^{-6} K kg⁻¹m²s⁻¹.

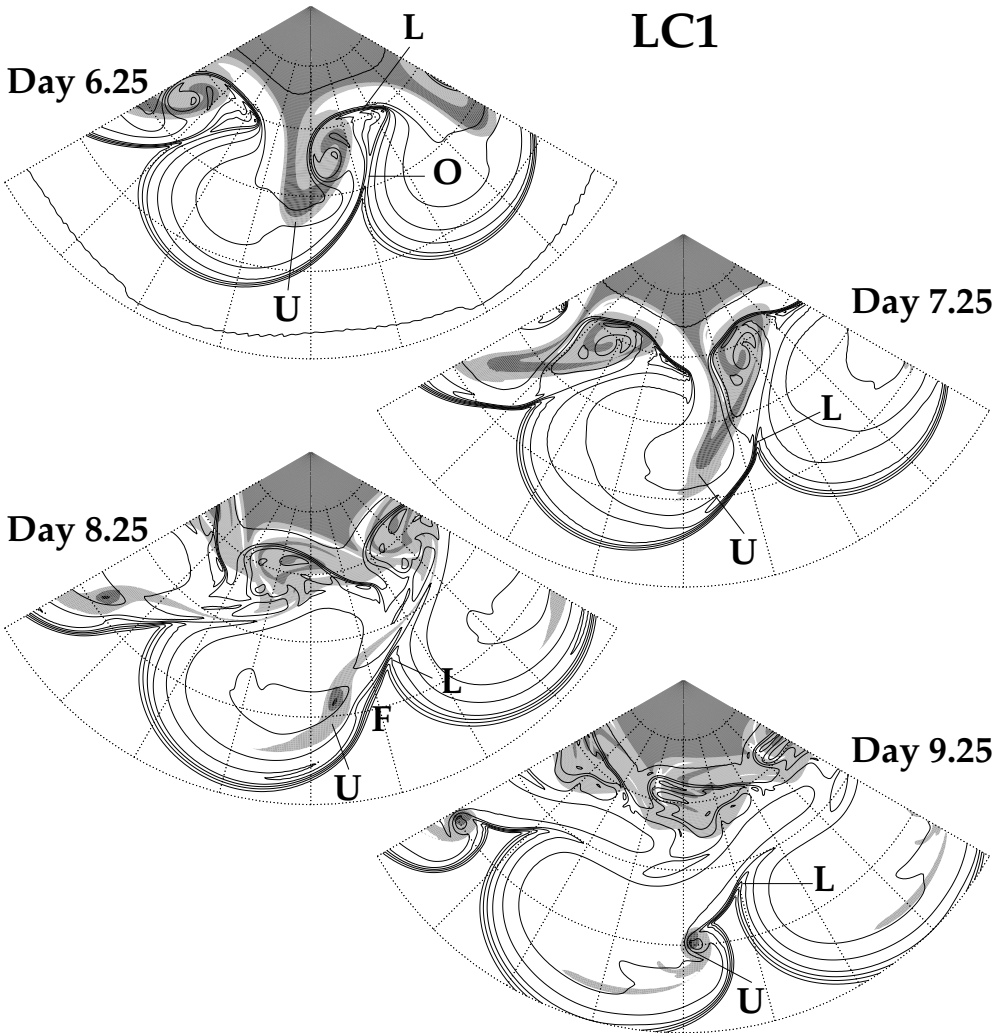


Figure 2. LC1 baroclinic wave evolution at lower and upper levels. The contours show θ on the lowest model level. The reference contour $\theta_1 = 280$ K is in bold and the contour interval is 5 K. The shading marks $PV > 1$ PVU and is darker for $PV > 3$ PVU on the 300 K surface, outlining the upper CRW. The stereographic projection extends from the north pole to 15°N and the grid spacing is 15° .

aligns with the cold front, retaining its displacement to the west. Poleward of 46°N the tropopause wraps into a cyclonic spiral.

The PV filament formed by anticyclonic wave breaking is associated with a deep tropopause fold. By day 8.25, vortex roll-up has clearly begun far from the polar vortex where its far-field strain is weaker. At point F, the wave is also evident in surface θ , although it is weaker. Thorncroft and Hoskins (1990) described this secondary cyclogenesis as arising from upstream development on the surface cold front, the first stages of which are suggested by the kink south of L on day 7.25. At the higher resolution used here, the upper-level vortex roll-up is stronger and is evident earlier, suggesting that the interaction between the tropopause fold and trailing cold front is initially dominated by the upper-level PV wave. As the secondary cyclone matures,

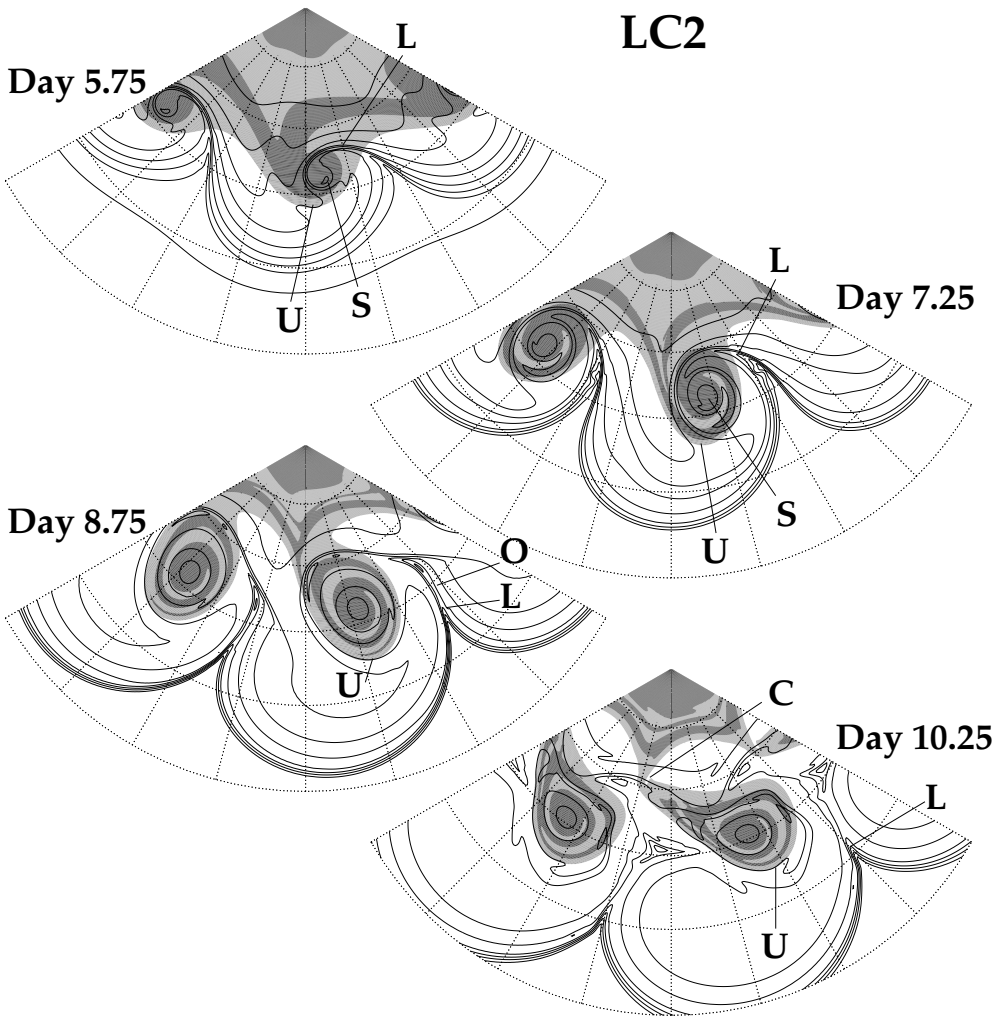


Figure 3. LC2 baroclinic wave evolution at lower and upper levels (contouring as for Fig. 2). The shading is for $PV > 2$ PVU and is darker where $PV > 5$ PVU on the 320 K surface.

a cut-off vortex forms at upper levels (point U at day 9.25), and at the ground some warm air is drawn cyclonically into a seclusion below and just to the east of the cut-off vortex centre.

When cyclonic barotropic shear is added to the basic state in the mid-latitudes, the surface warm sector wraps cyclonically, as seen for LC2 in Fig. 3. The cold front merges with the stronger warm front forming the characteristic bent back λ -structure which wraps cyclonically, forming a warm cut-off (S) below the upper wave trough (similar examples from the atmosphere have been shown by Shapiro *et al.* (1999) alongside model results). Such cut-offs are referred to as warm seclusions (Bjerknes and Solberg 1922; Shapiro and Keyser 1990). The seclusion is pinched off at day 6.00 leaving a warm sector which extends far polewards (to point L on day 7.25). This feature stretches and thins until its northern tip is pinched off by occlusion (feature O) at day 7.75. Meanwhile, the upper wave breaks cyclonically forming a large-scale cyclone overlying the warm seclusion (S)—a warm-core vortex through the depth of the troposphere.

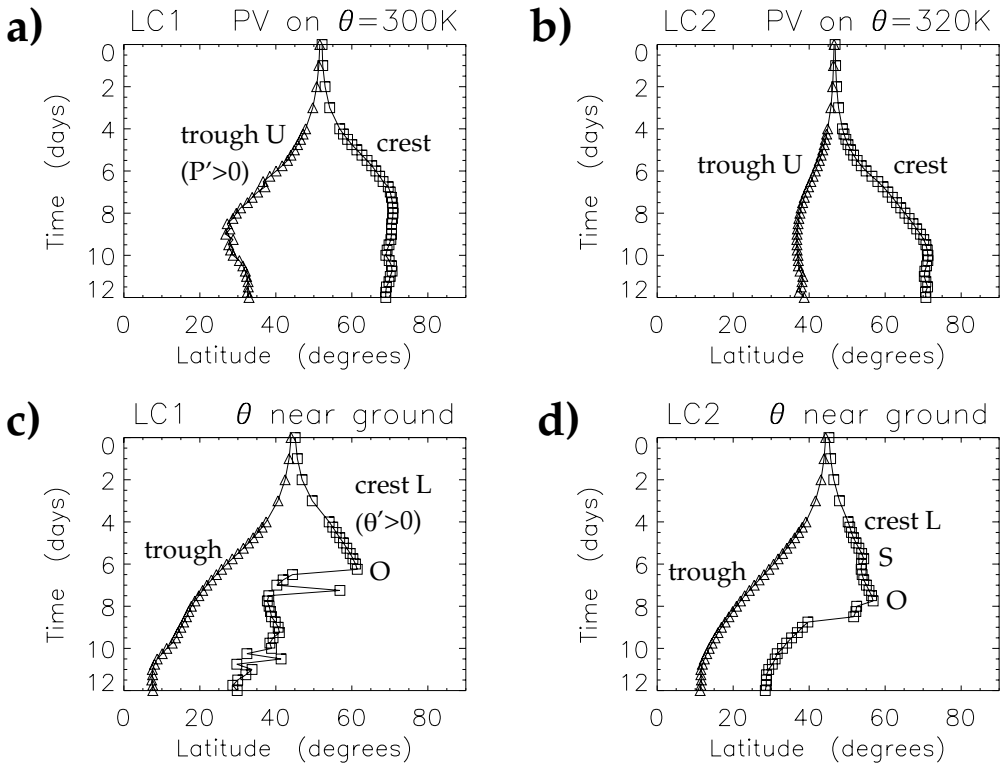


Figure 4. The amplitudes of upper and lower CRWs versus time, shown using the maximum and minimum latitudes for points along contours coinciding with their home-bases at day 0. For LC1, the 1 PVU contour on the 300 K surface describes the upper wave. For LC2, the 2 PVU contour on 320 K is used. In both cases, the 280 K contour describes the surface temperature wave of the lower CRW.

On lower isentropic surfaces the tropopause trough wraps into a large-scale cyclonic PV spiral (shown by Methven and Hoskins 1998). Eventually, the upper-level cyclone is pinched off from the polar vortex at point C. Although it is stretched to the north-west during the cut-off process, afterwards it regains axisymmetry.

Figure 4 shows the maximum and minimum latitude of points along each home-base contour throughout the two life cycles. These points are referred to as the crest and trough, respectively. During the first four days the growth of wave amplitude is nearly symmetric about its initial latitude and exponential, as expected for linear instability. Between four and six days, the poleward extension of the lower θ -wave saturates during frontogenesis as the warm sector occludes (see Figs. 5 and 8 of THM for the early evolution of surface θ in LC1 and LC2).

In LC1, the sharp drop in maximum latitude (Fig. 4(c)) after day 6.25 occurs as the θ_1 contour occludes—the pinched-off warm air is not tracked. After this time the poleward tip of the wave (labelled L in Fig. 2) remains at roughly the same latitude while the southern edge of the cold front continues to extend equatorwards, only halting at day 11. At upper levels (Fig. 4(a)), the wave amplitude saturates later, at about day 7. The equatorward extension continues for a further two days as the PV filament wraps around the anticyclone. After day 7.5, vortex roll-up occurs and an upper-level cut-off forms, partly through interaction with surface θ . The triangles in Fig. 4(a) mark the equatorward edge of this persistent vortex (marked by U in Fig. 2) even after the PV

filament connecting it to the PV spiral in the north has thinned and dissipated. After day 6.5 the 2 PVU contour is tracked, so that the cut-off is followed to day 12 rather than the dynamically insignificant PV filament that extends even further equatorwards (tracked by the 1 PVU contour). At day 6.5 when the switch is made, the 1.0 and 2.0 PVU contours are almost coincident because the PV gradient in the tropopause zone is extremely sharp.

Similar behaviour occurs at the early stages in LC2. Figure 4(d) shows that the poleward extension of the surface temperature wave (L) first saturates at day 5.75 when the whole warm sector wraps cyclonically, forming a warm seclusion (S). The second sharp drop in this curve occurs after day 7.75 when the warm sector occludes (feature O).

(b) *Comparing CRW theory with reference contour development*

An important aspect of the discussion of normal-mode baroclinic instability in terms of CRWs is the phase difference between the upper and lower waves. This phase difference for the nonlinear life cycles is shown in Fig. 5. It is calculated by finding the difference in longitude between the trough of the upper contour (labelled U in Figs. 2 and 3) and the crest of the surface wave (L) which correspond to the positive PV anomalies of each wave. The result is multiplied by -6 to convert to the phase difference (since $m = 6$). Initially, the phase difference is close to the phase-locked angle ϵ^+ from CRW theory (the correspondence is not exact because ϵ^+ is calculated by the orthogonality method (see Part III), so that the association with the wave phases at two home-bases is only approximate).

In LC1, even during the exponential-growth stage, the phase difference increases slightly as the crest of the lower wave travels more quickly than the upper trough due to anticyclonic wrapping. However, after day 3 the warm sector extends onto the cyclonic side of the jet (north of 46°N) and the phase difference shifts back to ϵ^+ . The rapid increase in phase difference between days 5.25 and 6.25 occurs during frontogenesis until the warm sector occludes and the θ_1 contour is pinched off (point O). The phase difference again returns to close to ϵ^+ until secondary cyclogenesis reaches large amplitude at day 8.5. Then there is a transition from measuring the longitudinal separation of the upper trough and warm sector of the primary wave (before day 8.5) to measuring the separation of the upper trough and warm sector of the secondary wave (after day 10). Note that the secondary wave has a much smaller scale than the primary wave and is not sinusoidal. It is initiated as the tropopause fold and cold front produced during the primary cyclogenesis interact and thus is not described by the interaction of two monochromatic CRWs. Nevertheless, the upper and lower features become almost phase-locked with a westward tilt in height and a phase speed that is eastward relative to the zonally averaged surface wind at the latitude of the cut-off cyclone centre ($-12^\circ \text{ day}^{-1}$). Note also the structural similarity between the secondary cyclone at day 9.25 (Fig. 2) and the primary cyclone of LC2 at day 5.75 (Fig. 3), although the whole secondary cyclone system is oriented from south-south-west to north-north-east along the axis of the original cold front.

In LC2, the cyclonic wrapping can be seen clearly in the phase difference diagnostic (Fig. 5(b)) which decreases with time as the trough of the upper PV wave (U) travels faster than the crest of the surface warm sector (L). After the seclusion is cut off (day 6.00), the phase difference returns to ϵ^+ until day 7.75, when the warm sector occludes and the contour is again pinched off. Eventually the phase difference returns to ϵ^+ , even though the wave has completely wrapped up forming a synoptic-scale cyclone with a spiral in PV at upper levels and a cut-off warm pool below.

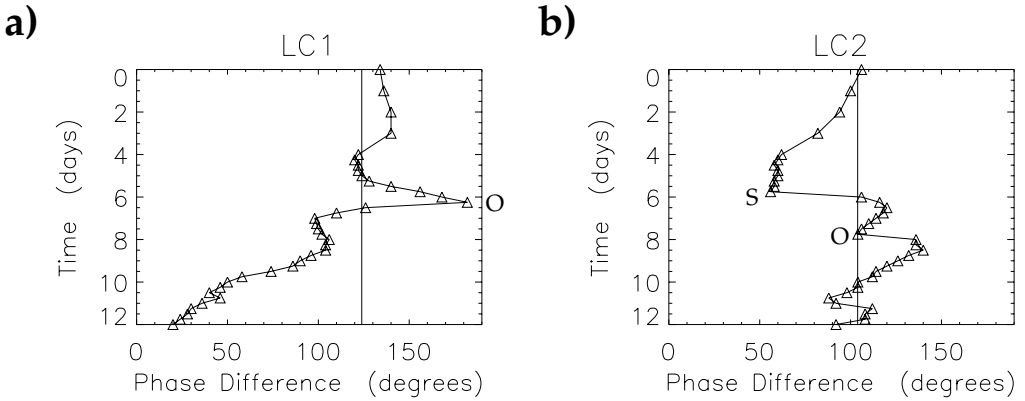


Figure 5. The phase difference between the southern tip of the upper CRW's potential vorticity trough (point U in Figs. 2 and 3) and the northern tip of the surface warm sector (L). The vertical line marks the phase for the locked configuration, ϵ^+ , determined from the normal mode using CRW theory.

The rate of change in longitude of the home-base contour crests and troughs is shown in Fig. 6 (calculated by centred difference in time). For the first three days the upper and lower waves progress at the phase speed of the NM, c_r .

Between days 2 and 4 in LC1, the lower wave slows slightly as it begins to break (Fig. 6(c)). From day 4, the poleward cusp (L) runs ahead of the synoptic-scale wave as the warm sector tilts from south-west to north-east. At day 6.5 it occludes and the diagnosed longitude of maximum poleward extension skips from the cut-off warm sector to the new cusp (see point L on day 7.25 in Fig. 2) which lies to the south-west (giving rise to the negative spike in 'phase speed'). In the meantime the synoptic-scale wave slows steadily until day 6 when its phase speed matches the self-induced phase speed of the upper CRW from linear theory, c_2^2 . This can be seen clearly from the trough of the lower wave, but it is also apparent that the crest tends towards this speed after frontal occlusion.

The upper wave deviates from NM behaviour after day 4, when its crest slows slightly relative to the PV trough. This is a signature of the cyclonic wrapping that occurs north the jet axis (46°N). Between days 4 and 6.75, the trough and crest stay in step straddling the jet axis. At this stage the upper wave speed lies between c_r and c_2^2 . Then the upper-level anticyclonic wave-breaking event occurs, and the trough slows rapidly as it is advected around the anticyclone. From day 8 onwards, the trough (U) tracks the (almost stationary) cut-off secondary cyclone.

In LC2, the cyclonic wrapping of the lower wave results in the trough running faster than the crest (Fig. 6(d)). As with LC1, the lower wave in LC2 slows after saturation from the NM phase speed to speed c_2^2 . The two spikes in the speed of the wave crest correspond to the seclusion (S) and occlusion (O) events. At upper levels, the equatorward side of the upper PV spiral (U) continues at speed c_r while the crest of the PV contour slows to speed c_2^2 at day 4 and less than this at day 6 (Fig. 6(b)). Between days 5 and 7, the speed of the upper wave obtained by averaging its crest and trough speeds is approximately c_2^2 . After this time, the crest keeps in step with the connection between the PV spiral and the polar vortex and as it pinches off (at point C in Fig. 3) it then tracks the westward propagating Rossby wave on the polar vortex formed during the cut-off process.

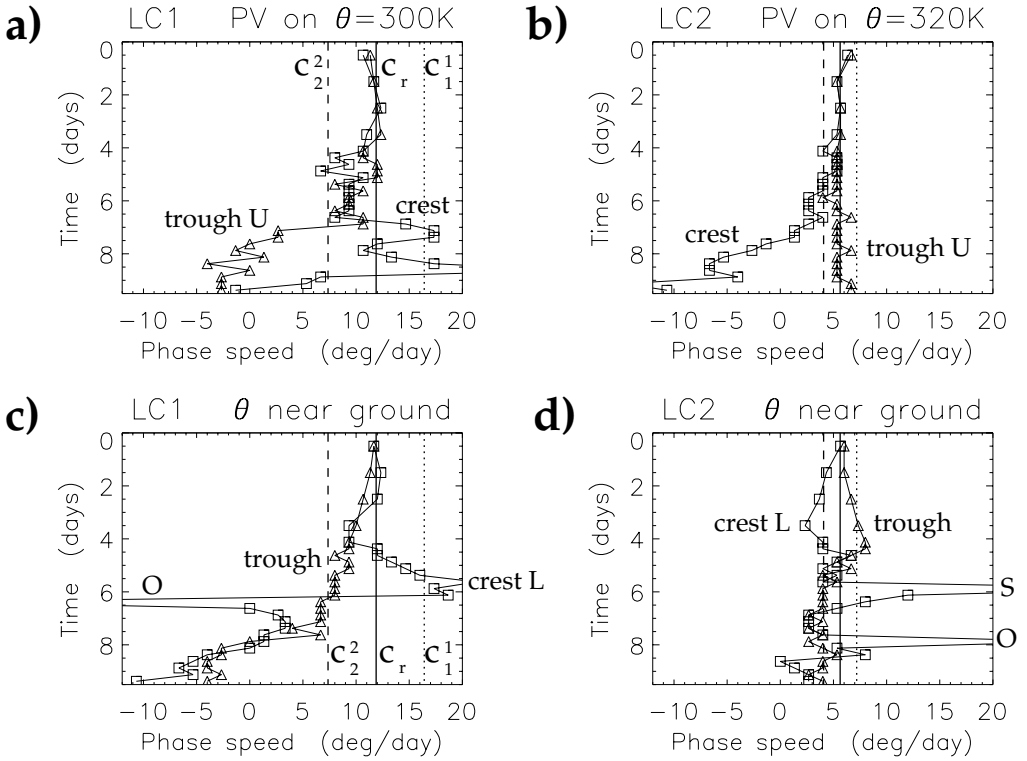


Figure 6. Phase speeds (degrees longitude per day) calculated separately for the crest (squares) and trough (triangles) of each home-base contour. The solid vertical line shows the normal-mode phase speed, c_r , and the dashed and dotted lines show, respectively, the self-induced phase speeds of the upper CRW, c_2^2 , and lower CRW, c_1^1 .

The slowing down of both upper and lower waves to a speed close to the self-induced speed of the upper CRW (c_2^2) occurs in both life cycles. This can be understood as a consequence of nonlinear wave saturation occurring first at low levels. During the small-amplitude growth phase, the CRWs are phase-locked in a ‘hindering’ configuration ($\pi/2 < \epsilon^+ < \pi$) so that the upper CRW propagates less rapidly to the west (relative to the westerly jet) as a result of meridional wind induced by the surface temperature anomalies of the lower CRW (i.e. $c_r > c_2^2$). As the amplitude of the surface temperature wave saturates (the occlusion process), the winds induced by the upper CRW’s PV become stronger at all levels relative to the winds induced by the lower CRW. Consequently, the westward propagation of the upper CRW is less hindered by interaction and the eastward propagation of the lower CRW is more hindered than before. The whole structure then has a slower phase speed. The disturbance cannot propagate slower than speed c_2^2 , which is the speed when the lower CRW has zero amplitude. The speed of the baroclinic wave can be predicted to lie in the range $c_2^2 < c < c_r$ until it breaks, perhaps decays barotropically and the development of other waves takes over (in the LC1 case through frontal cyclogenesis).

4. DISCUSSION ON THE DIRECTION OF WAVE BREAKING

Many studies have found that the direction of breaking of baroclinic waves is strongly influenced by the barotropic wind shear associated with the meridional surface

pressure gradient in the initial basic state (e.g. Simmons and Hoskins 1980; Davies *et al.* 1991; THM). Surface θ -contours would be expected to overturn in the sense given by this shear. However, at upper levels, although the baroclinic jet dominates \overline{U}_y , wave breaking still occurs in the sense given by the surface shear. In LC2, for example, where the surface shear is negative between 20°N and 50°N, PV contours wrap up cyclonically at all levels, even though the shear is anticyclonic equatorward of 40°N in the upper troposphere. Two mechanisms may be responsible for this behaviour at upper levels: (i) advection of PV is influenced by the wind anomalies induced by the lower CRW which breaks first, and (ii) the meridional shear is modified by eddy momentum fluxes.

First note that an untilted PV wave induces meridional winds that are $-\pi/2$ out of phase and, therefore, can only propagate zonally without change in amplitude or structure, even when considering large amplitude (without linearization of the PV equation). Therefore, the change in structure associated with wave breaking can only occur through winds induced by another PV anomaly or wave. As a pair of growing CRWs reach finite amplitude, such that the horizontal wave slope of PV contours is about unity, the winds induced by one CRW at the other's home-base enable wave breaking.

Davies *et al.* (1991) have shown that, in a semi-geostrophic uniform PV (Eady-like) model, baroclinic waves possess a horizontal symmetry about the jet axis when viewed in geostrophic coordinates, if the jet is symmetric. Nakamura (1993) referred to this symmetry as 'sinuous' and showed that if an NM possesses this symmetry then it is preserved when it reaches finite amplitude and develops nonlinearly. Material contours overturn cyclonically to the north of the initial jet axis and anticyclonically to the south, as expected from the shear on the flanks of the jet. Importantly, this behaviour is seen in surface θ , even though the basic-state surface wind is almost zero, indicating that the surface wave is influenced by the direction of breaking of the upper wave at the lid where the jet is strongest. Also, the geostrophic transformation into physical space tightens regions with cyclonic anomalies, which at the ground correspond to the warm sectors. These features are all observed in the LC1 surface θ field (e.g. day 6.25 in Fig. 2).

Additional barotropic shear breaks the meridional symmetry of the jet and the sinuous symmetry of the waves. However, Davies *et al.* (1991) also showed that the waves have reflectional symmetry about the baroclinic jet axis on change in sign of the barotropic shear. In other words, if waves break cyclonically on the addition of cyclonic shear, they will break anticyclonically on the addition of anticyclonic shear.

On realistic jets in models without a rigid lid, the surface θ wave breaks before the wave at tropopause level (Simmons and Hoskins 1978). Meridional parcel displacements first reach large amplitude near the latitude of maximum baroclinicity around the steering level, which is typically in the lower troposphere. For short waves ($m > 8$), the upper CRW from linear theory lies only just above the steering level (Fig. 1(b)). Since the steering level is near the ground, air-parcel displacements are expected to be more horizontal than along the slope of basic-state $\overline{\theta}$ contours (see section 3(d) of Part III), and therefore the meridional shear in \overline{U} along σ -surfaces is more relevant to wave-breaking direction than the shear along $\overline{\theta}$ -surfaces. As baroclinic waves attain large amplitude and the lower wave breaks, the wave activity shifts upwards towards the tropopause (Edmon *et al.* 1980). However, the PV anomalies, already formed by the breaking wave near the steering level and below, induce winds which encourage meridional overturning in the same sense near the tropopause.

For longer waves ($m < 7$), the upper CRW's home-base is closer to the tropopause, but the breaking direction is still determined by the horizontal shear at low levels.

In this case the steering level is much closer to the lower CRW's home-base and the horizontal wave-slope of its reference contour, θ_1 , reaches $O(1)$ long before the reference PV contour of the upper CRW, P_2 . It is also worth noting that the winds induced by surface θ anomalies are more influenced by small scales than those induced by interior PV anomalies due to a difference in the scale effect of PV inversion. This results in a greater likelihood of vortex roll-up in surface quasi-geostrophic (QG) dynamics than in barotropic vorticity dynamics (Held *et al.* 1985) and may plausibly make the overturning of surface θ contours more rapid than upper-level PV contours. The large wind anomalies induced aloft by the breaking surface θ wave encourage the upper PV wave to overturn in the same direction.

The second mechanism involves a wave-mean flow feedback, described as the 'nonlinear barotropic governor' by Nakamura (1993). In Part III it was argued that the requirement for weak variation in Rossby wave speed over the CRW structure (especially for short waves) results in a tendency for the air-parcel displacement-structure of CRWs to be coherent along \bar{U} -surfaces in the lower troposphere. If the jet is broad relative to the Rossby radius of deformation, the boundary displacement structure of the lower CRW will have narrower structure than the jet (Jukes 2000). However, in the interior its displacement structure follows the steering surface, $\bar{U} = c_r$, around the flanks of the jet. In section 4(b)(i) of Part III it was shown that this structure implies that the growing NM must have vertically integrated momentum-flux convergence at the latitudes where the boundary wave activity dominates and compensating divergence elsewhere (see Fig. 12 of Part III). The barotropic component of the jet accelerates where there is convergence. Additional cyclonic barotropic shear imposes a poleward slant on \bar{U} surfaces and in the displacement structure of CRWs. Consequently, the interior wave activity is displaced polewards relative to the boundary wave activity so that momentum flux converges above and equatorward of the boundary wave activity maximum and diverges on the poleward side. Therefore, the barotropic component of the jet becomes more cyclonic. Associated with this, the poleward slant of \bar{U} surfaces is accentuated, the CRW displacement structures follow suit and the vertically integrated momentum flux amplifies.

The above argument does not describe how the zonal-flow changes are distributed in the vertical. However, the slant implies a poleward shift of the meridional-wind anomalies of the upper CRW, relative to the lower CRW at all levels. Figure 13 of Part III illustrates how this shift gives rise to a negative NM momentum flux and convergence on the equatorward flank of the jet. This would tend to accelerate the zonal-average zonal angular velocity, \bar{U} , on this flank at all levels, although in the upper troposphere this is partly cancelled by the Coriolis acceleration associated with the indirect meridional circulation induced by the wave (McIntyre 1970). Figure 9 of Simmons and Hoskins (1977) shows that the zonal-flow acceleration is mainly in the lower troposphere for NMs on the Z1 jet. However, when they attain large amplitude their momentum fluxes are much larger in the upper troposphere (Simmons and Hoskins 1978). Therefore, the barotropic component of the Z2 jet will accelerate due to the negative momentum flux of the baroclinic wave, shifting the zonal-average jet equatorwards, expanding the region of cyclonic shear on its northern flank (\bar{U} -change shown in Fig. 3 of THM). As a result, PV contours at upper levels would tend to break in the same direction as near the ground.

5. CONCLUSIONS

The self-induced phase speeds and home-bases of the CRWs obtained from the NM using CRW theory (see Parts I and III) are pertinent to the baroclinic-wave evolution

right until the time at which the upper wave breaks and decays through dissipation. This is remarkable, given that surface fronts have formed and occluded before this stage and upper-level PV contours have stretched and folded. It is, however, consistent with semi-geostrophic baroclinic wave development (Hoskins and West 1979) in which very strong surface and tropopause-level frontogenesis occurs but the larger-scale wave characteristics are unaffected. From the CRW perspective, although surface temperature fronts form, the lower CRW structure consisting of a zonal contrast between the cold and warm sectors of the primary wave continues to be important. Although fine-scale PV filaments form on isentropic surfaces intersecting the tropopause, the inversion to obtain velocity is a smoothing operation so that the synoptic-scale contrasts between the primary-wave trough and ridge encapsulated by the upper CRW continue to dominate the induced velocity field. The interaction of the finite-amplitude CRW-like structures continues to occur as in the linear regime. Furthermore, the material conservation of θ and PV and the global conservation of pseudomomentum and pseudoenergy, vital to CRW theory, are also valid in the nonlinear regime, provided that the flow is almost adiabatic and frictionless. Therefore, the arguments concerning the rate of Rossby-wave propagation and the relative phases between the CRWs necessary for the hindering or helping of zonal propagation and mutual growth are unaltered at large amplitude until the final dissipation of the wave.

The lower wave breaks before the upper wave during the frontal occlusion process. Therefore, it saturates nonlinearly while the upper wave continues to amplify. The main effect of the nonlinear saturation and the reduced relative importance of the lower CRW is to modify the phase speed of the primary wave from that of the NM, c_r , to the self-induced phase speed of the upper CRW which is defined as $c_2^2 = U_2 - \gamma_2^2/m$, where its self-propagation rate γ_2^2/m is proportional to the background PV gradient at its home-base. While the features usually tracked by the synoptic meteorologist, such as the triple point where cold and warm fronts intersect, may travel faster or slower than the CRWs depending upon the direction of wave breaking, once occlusion has occurred the warm sector assumes the upper CRW's self-induced phase speed. It was shown in Part III that CRWs on realistic jets without a meridional slant (e.g. Z1) behave like Charney modes and must always lock in a hindering configuration with a large phase difference ($\pi/2 < \epsilon < \pi$). Consequently, the CRW theory (Eq. (24) in Part III) shows that $c_2^2 < c_r$ and low-level saturation must slow down the baroclinic wave.

In summary, two surprising results of the CRW analysis are that, when surface meridional shear is weak compared with self-propagation rates:

- the westward Rossby-wave propagation on the interior positive PV gradients is so strong that the self-induced phase speed of the upper CRW (advection plus propagation) is slower than that of the lower CRW ($c_2^2 < c_1^1$),
- when the surface wave occludes nonlinearly the system *slows* to the self-induced phase speed of the upper CRW given by linear theory.

This behaviour is consistent with, for example, the interaction of synoptic systems with blocks. The creation or reinforcement of blocks is performed by systems that, as they occlude, slow and transport large amounts of low-PV air polewards. Younger, growing systems tend to move too rapidly to do this.

THM proposed the 'saturation-propagation-saturation' picture for baroclinic-wave life cycles. In this picture, the wave first reaches large amplitude at low levels and then saturates; the occlusion of the warm sector. They then identify the Eliassen-Palm (EP) flux with the group velocity of a Rossby-wave packet and refer to the upward component, F_z , as the vertical propagation of Rossby waves. Furthermore, the

equatorward EP flux, $F_y = -\overline{u'v'}$, in LC1 and the poleward EP flux in LC2 are identified with the meridional propagation of a wave packet. They argue, more tentatively, that the direction of the group velocity can be anticipated from the refractive index for linear Rossby-wave propagation in QG theory. The meridional propagation into the anticyclonic or cyclonic flank of the upper-level jet determines the direction of wave breaking.

CRWs and finite-amplitude wave-activity diagnostics suggest a perspective that differs from the propagation aspect of the ‘saturation–propagation–saturation’ picture. As low-level saturation occurs, the amplitude of the upper CRW increases relative to the lower CRW. Concomitant with this, wave-activity density increases faster at upper levels and, since it is conserved, there must be a vertical flux of *nonlinear wave-activity density* from the ground towards the upper CRW’s home-base. Magnusdottir and Haynes (1996) have shown that the nonlinear meridional wave-activity flux \mathcal{F}_y (Haynes 1988) in LC1 and LC2 is associated mainly with the advection of wave-activity density during wave breaking, which has no counterpart in the EP flux and, therefore, cannot be related to refractive index. Moreover, CRWs retain their gross structure and zonal propagation characteristics and cannot be considered to ‘propagate’ meridionally or vertically—in this view, the wave-activity flux is not associated with a group velocity. However, the horizontal eddy tilt in the early stages of growth is determined simply by the relative positions of the CRWs (Fig. 13 of Part III) and is also robust at finite amplitude. Therefore, the horizontal-momentum flux is directed ‘up-gradient’ relative to the surface meridional shear, as for the NM, tending to accelerate the barotropic component of the zonal flow associated with the surface shear. As a result the \overline{U} -surfaces slant more and the displacement structures of the CRWs slant with them, increasing the meridional separation between boundary and interior wave-activity density and, therefore, amplifying the vertically integrated momentum flux (see Fig. 12 of Part III). As a result of the positive feedback, called the ‘nonlinear barotropic governor effect’ by Nakamura (1993), F_y gains importance relative to F_z as the wave reaches large amplitude, accounting for the apparent meridional propagation in the upper troposphere as the lower wave saturates. The zonal-flow changes act to shift the jet meridionally in the direction of increasing surface \overline{U} so that the upper CRW would tend to break in the same direction as occurred at low levels. The influence of the wind anomalies induced by the lower CRW, which reaches large amplitude and breaks first, would also encourage the upper CRW to break in the same sense. This perspective accords with BG97 who observed that, in general, NM momentum fluxes are a more reliable indicator of the direction of nonlinear wave breaking than the refractive index.

Although it has been established here that some key properties of the nonlinear development of baroclinic instability can be predicted and understood by considering the linear dynamics of CRWs, there are still major difficulties in applying the theory in a quantitative diagnosis of the atmosphere. Baroclinic waves in the extratropics are continuously growing and decaying and, at any one time, there will be several synoptic-scale weather systems of different scales at various stages in their development. The atmospheric state is never zonally symmetric so the identification of a relevant basic state is problematic. Many systems appear to be initiated by a finite-amplitude disturbance from upstream which cannot be described by a single zonal wave number—for example, the secondary cyclogenesis occurring in LC1. Even if linear dynamics gives quantitative predictions for such a situation, it is necessary to consider the continuous spectrum in initial-value problems (Farrell 1982). Since a CRW pair only describes the dynamics of a single growing NM and its CC, further work is needed to investigate interaction between elements of the discrete spectrum and the continuous

spectrum before more general conclusions about finite-time baroclinic growth can be drawn.

ACKNOWLEDGEMENTS

John Methven is grateful for an Advanced Fellowship sponsored jointly by the Natural Environment Research Council and the Environment Agency. Craig Bishop acknowledges support from National Science Foundation grants ATM-96-12502 and ATM-98-14376. Support from the Office of Naval Research grant N00014-00-1-0106 and Program Element 0601153N Project number BE-0333-0345 are also gratefully acknowledged. Many thanks to Jeff Cole, Lois Steenman-Clark and Mike Blackburn from the Centre for Global Atmospheric Modelling (CGAM), University of Reading, who parallelized the Reading spectral model and ran the high-resolution life-cycle experiments on the CRAY T3D at the Edinburgh Parallel Computing Centre. Thanks to Martin Jukes and the referees whose reviews led to improvements in the discussion.

REFERENCES

- Balasubramanian, G. and Garner, S. 1997 The role of momentum fluxes in shaping the life cycle of a baroclinic wave. *J. Atmos. Sci.*, **54**, 510–533
- Bjerknes, J. and Solberg, H. 1922 Life cycle of cyclones and the polar front theory of atmospheric circulation. *Geophys. Publ.*, **3**(1), 1–18
- Bretherton, F. P. 1966 Baroclinic instability and the short wavelength cut-off in terms of potential vorticity. *Q. J. R. Meteorol. Soc.*, **92**, 335–345
- Davies, H., Schär, C. and Wernli, H. 1991 The palette of fronts and cyclones within a baroclinic wave development. *J. Atmos. Sci.*, **48**, 1666–1689
- Edmon, H., Hoskins, B. J. and McIntyre, M. E. 1980 Eliassen–Palm cross-sections for the troposphere. *J. Atmos. Sci.*, **37**, 2600–2616
- Farrell, B. 1982 The initial growth of disturbances in a baroclinic flow. *J. Atmos. Sci.*, **39**, 1663–1686
- Haynes, P. 1988 Forced, dissipative generalisations of finite-amplitude wave-activity conservation relations for zonal and non-zonal basic flows. *J. Atmos. Sci.*, **45**, 2352–2362
- Heifetz, E., Bishop, C. H., Hoskins, B. J. and Methven, J. 2004a The counter-propagating Rossby wave perspective on baroclinic instability. Part I: Mathematical basis. *Q. J. R. Meteorol. Soc.*, **130**, 211–231
- Heifetz, E., Methven, J., Hoskins, B. J. and Bishop, C. H. 2004b The counter-propagating Rossby wave perspective on baroclinic instability. Part II: Application to the Charney model. *Q. J. R. Meteorol. Soc.*, **130**, 233–258
- Held, I., Pierrehumbert, R., Garner, S. and Swanson, K. 1995 Surface quasi-geostrophic dynamics. *J. Fluid Mech.*, **282**, 1–20
- Hoskins, B. and Simmons, A. 1975 A multi-layer spectral model and the semi-implicit method. *Q. J. R. Meteorol. Soc.*, **101**, 637–655
- Hoskins, B. and West, N. 1979 Baroclinic waves and frontogenesis. Part II: Uniform potential vorticity jet flows. *J. Atmos. Sci.*, **36**, 1663–1680
- Hoskins, B., McIntyre, M. and Robertson, A. 1985 On the use and significance of isentropic PV maps. *Q. J. R. Meteorol. Soc.*, **111**, 877–946
- James, I. N. 1987 Suppression of baroclinic instability in horizontally sheared flows. *J. Atmos. Sci.*, **44**, 3710–3720
- Jukes, M. N. 2000 Linear instability of broad baroclinic zones. *Q. J. R. Meteorol. Soc.*, **126**, 1065–1098
- Magnusdottir, G. and Haynes, P. 1996 Wave activity diagnostics applied to baroclinic wave life cycles. *J. Atmos. Sci.*, **53**, 2317–2353
- McIntyre, M. E. 1970 On the non-separable baroclinic parallel flow instability problem. *J. Fluid Mech.*, **40**, 273–306
- 1980 Towards a Lagrangian-mean description of stratospheric circulations and chemical transports. *Phil. Trans. R. Soc. London*, **A296**, 129–148
- Methven, J. and Hoskins, B. 1998 Spirals in potential vorticity. Part I: Measures of structure. *J. Atmos. Sci.*, **55**, 2053–2066
- 1999 The advection of high-resolution tracers by low-resolution winds. *J. Atmos. Sci.*, **56**, 3262–3285

- Methven, J., Heifetz, E., Hoskins, B. J. and Bishop, C. H. 2005 The counter-propagating Rossby-wave perspective on baroclinic instability. Part III: Primitive-equation disturbances on the sphere. *Q. J. R. Meteorol. Soc.*, **131**, 1393–1424
- Nakamura, N. 1993 Momentum flux, flow symmetry, and the nonlinear barotropic governor. *J. Atmos. Sci.*, **50**, 2159–2179
- Shapiro, M. and Keyser, D. 1990 'Fronts, jet streams and the tropopause'. Pp. 167–191 in *Extratropical cyclones – The Erik Palmén Memorial Volume*, American Meteorological Society, 45 Beacon Street, Boston MA02108-3693, USA
- Shapiro, M., Wernli, H., Bao, J.-W., Methven, J., Zou, X., Doyle, J., Holt, T., Donall-Grell, E. and Neiman, P. 1999 'A planetary-scale to mesoscale perspective of the life cycles of extratropical cyclones: The bridge between theory and observations'. Pp. 139–185 in *The life cycles of extratropical cyclones*, Eds. M. Shapiro and S. Grönås, American Meteorological Society, 45 Beacon Street, Boston MA02108-3693, USA
- Simmons, A. and Hoskins, B. 1977 Baroclinic instability on the sphere: Solutions with a more realistic tropopause. *J. Atmos. Sci.*, **34**, 581–588
- 1978 The life cycles of some nonlinear baroclinic waves. *J. Atmos. Sci.*, **35**, 414–432
- 1980 Barotropic influences on the growth and decay of nonlinear baroclinic waves. *J. Atmos. Sci.*, **37**, 1679–1684
- Thorncroft, C. and Hoskins, B. 1990 Frontal cyclogenesis. *J. Atmos. Sci.*, **47**, 2317–2336
- Thorncroft, C., Hoskins, B. and McIntyre, M. 1993 Two paradigms of baroclinic wave life cycle behaviour. *Q. J. R. Meteorol. Soc.*, **119**, 17–55