Wave Activity Diagnostics Applied to Baroclinic Wave Life Cycles

GUDRUN MAGNUSDOTTIR* AND PETER H. HAYNES

*Centre for Atmospheric Science, Department of Applied Mathematics and Theoretical Physics,
University of Cambridge, Cambridge, United Kingdom*

ABSTRACT

Wave activity diagnostics are calculated for four different baroclinic wave life cycles, including the LC1 and LC2 cases studied by Thorncroft, Hoskins, and McIntyre. The wave activity is a measure of the disturbance relative to some zonally symmetric, time-independent basic state, which need not be the initial zonally averaged state and which satisfies a finite-amplitude conservation relation. The wave activity density and fluxes may be calculated in terms of Eulerian variables provided that the potential vorticity is a monotonic function of latitude on isentropic surfaces in the basic state. The LC1 and LC2 experiments used initial states in which the potential vorticity (PV) did not satisfy this monotonicity condition. Therefore two approaches are taken. The first is to define a basic state that is not the initial state and use this to calculate the wave activity diagnostics. The second is to carry out new LC1- and LC2-type experiments on initial states in which the monotonicity condition is satisfied. New basic states are generated by PV rearrangement and inversion.

The results allow quantification of the difference between LC1- and LC2-type life cycles. They also show that LC1- and LC2-type behavior occurs for different initial states other than those used by Thorncroft, Hoskins, and McIntyre and that the classification is therefore robust in terms of the potential vorticity field and wave activity diagnostics. If one were to consider only eddy kinetic energy, the distinction is no longer clear. In fact, in the evolution of eddy kinetic energy the modified LC1-type life cycle resembles LC2 and the modified LC2 more than it resembles LC1.

The results also shed new light on the role of wave propagation in baroclinic life cycles. In particular, it is found that during the later stages of the life cycle the pattern of equatorward wave activity flux that has often been interpreted as associated with equatorward wave propagation in the subtropical upper troposphere is in fact associated primarily with advective transport of wave activity.

New finite-amplitude expressions are presented for the wave activity associated with potential temperature gradients on the lower boundary. Problems with using PV rearrangement techniques are discussed.

1. Introduction

Eliassen–Palm flux (hereafter EP flux) cross sections have been widely used as a diagnostic for baroclinic wave life cycles (e.g., Edmon et al. 1980; Thorncroft et al. 1993, hereafter THM). The EP flux pattern has two important implications. First, its divergence is a measure of the wave forcing of the zonal-mean flow and can be interpreted as an eddy-induced force per unit mass or, equivalently, in many cases of interest, as the latitudinal eddy potential vorticity flux.[1] Second, it indicates wave propagation, being the flux of a conservable wave property, known as the Eliassen–Palm wave activity.

The appearance of the EP flux cross sections prompted Edmon et al. (1980) to distinguish four distinct stages to baroclinic life cycles, namely (i) linear growth, (ii) low-level nonlinear saturation, (iii) upward and equatorward propagation of wave activity, and (iv) low-latitude nonlinear saturation. The wave activity flux interpretation of the EP flux was particularly important in making these distinctions (Held and Hoskins 1985). However, one drawback in using the EP flux as a wave activity flux is that the conservation relation for EP wave activity holds only for small-amplitude waves, whereas the baroclinic wave life cycles grow far into the nonlinear regime. Thus, even for adiabatic and frictionless dynamics there still is a source/sink term in the conservation relation for EP wave activity associated with nonlinear effects.

There is now a well-defined method for constructing wave activity conservation relations valid for finite-amplitude disturbances; that is there are no source/sink terms provided that the dynamics is conservative. This

---

[1] The equivalence is well known for quasigeostrophic dynamics. A similar relation has been found for semigeostrophic dynamics (Magnusdottir and Schubert 1991). Andrews et al. (1987, section 3.9) and McIntyre and Norton (1990) discuss other more general cases.

---

* Current affiliation: Department of Earth System Science, University of California, Irvine, California.

Corresponding author address: Dr. Gudrun Magnusdottir, Earth System Science, University of California, Irvine, CA 92717-3100.
E-mail: gudrun@uci.edu

method is based on the energy-Casimir or momentum-Casimir methods pioneered by Arnol'd to prove stability theorems for Hamiltonian systems. Its usefulness for constructing wave activity conservation theorems was first shown by McIntyre and Shepherd (1987). The explicit forms of such conservation relations for the primitive equations were first derived by Haynes (1988, hereafter H88). Unfortunately, the wave activities so constructed have not yet been shown to have any simple relation to forcing of the mean flow, outside the small-amplitude regime. Nonetheless, they may still be used as a quantitative diagnostic for the waves themselves.

One of the attractions of such conservation relations is that the quantities appearing in them, that is, the wave activity density and the wave activity flux components, are expressed in terms of Eulerian dynamical variables and do not require knowledge of particle displacements. This is to be contrasted with the generalized Lagrangian mean formulation of Andrews and McIntyre (1978). The absence of particle-displacement information relies on the fact that there are sufficient materially conserved quantities and that such quantities are monotonic functions of the relevant geometric coordinate. For the primitive equations the relevant quantities are potential vorticity (hereafter PV) and potential temperature. Thus for the primitive equations the wave activity quantities may be expressed in terms of Eulerian variables only if both the potential temperature is a monotonic function of height and (more problematically) the PV is a monotonic function of latitude.

Quantification of the waves necessarily requires a decomposition of the flow into a basic state part and a wave part. For the finite-amplitude wave activity diagnostics the basic state is simply a prenominated flow that is a self-consistent steady solution of the equations of motion. Wave activity conservation follows if this basic state has some invariance, for example in time in which case the wave activity is referred to as "pseudoenergy" or in a particular direction in which case the relevant wave activity is referred to as "pseudomomentum." If the basic state is not a self-consistent unforced solution of the equations of motion, then the wave activity conservation relation contains source/sink terms, even in the absence of nonconservative effects acting on the waves (H88). In this paper we shall concentrate on the use of the wave activity whose conservation arises from the longitudinal invariance of the basic-state flow. This is generally called the angular pseudomomentum since its conservation arises from the invariance of the basic state to rotation about the earth's axis. We shall call it pseudomomentum for short. The steadiness of the basic flow is in contrast to the basic state for the generalized Eliassen–Palm relation, for example, which is the instantaneous longitudinally averaged state and is therefore time dependent. This aspect of the definition of the basic state has some disadvantages; that is, the difference between the actual flow and the basic-state flow may become very large as the system evolves in time so that it may no longer seem sensible to regard the difference between the actual and basic flows as being due to waves. However, it also has some potential advantages in that it allows flexibility in the choice of the basic state.

These new wave activity diagnostics have already been used in a number of different contexts to study the evolution of wavy flows. For example, Scinocca and Peltier (1994) have presented such diagnostics for mountain wave simulations, using the expressions for finite-amplitude pseudoenergy density and flux derived by Scinocca and Shepherd (1992). Brunet and Haynes (1996) have recently used pseudomomentum diagnostics for the shallow-water equations to study reflection of Rossby wave trains from low-latitude wave breaking regions.

Here we apply such diagnostics to baroclinic life cycle simulations, concentrating in particular on the two different types of life cycle behavior noted by THM. The only difference in initial conditions for the two life cycle experiments, LC1 and LC2, is the addition of a barotropic component to the wind so as to increase the cyclonic shear in midlatitudes for the case LC2 compared to LC1. However, the subsequent development for the two cases is quite different, as was documented in THM. In the case of LC1 the picture that emerges is that of Rossby wave activity propagating upward into the jet after the onset of low-level saturation, followed by a second saturation event in the upper troposphere associated with Rossby wave breaking on the equatorial side of the jet. This last stage of the life cycle is analogous to an absorbing Rossby wave critical layer. This is exactly the stage where LC2 differs markedly from LC1. In the case of LC2, the wave breaking that occurs takes place on the cyclonic side of the jet. These two life cycles were originally studied by Simmons and Hoskins (1980), who used the term "anomalous" to describe LC2.

One complication that arose in the application of wave activity diagnostics to these flows was that the initial states used by THM for their LC1 and LC2 experiments do not satisfy the monotonicity condition on PV. For example, Fig. 1a shows PV variation on a number of different isentropic surfaces in the LC1 basic state. Figure 1b shows the same for the LC2 basic state. In both cases there is a PV maximum at latitudes between 50° and 60°N, depending on the particular isentropic surface. This is interesting in itself in that such life cycles, particularly LC1, have been considered over the past decade or so as paradigms for nonlinear baroclinic instability. One might argue that a more straightforward example of baroclinic instability would be having an initial state in which PV was a monotonic function of latitude on interior isentropic surfaces, with the instability arising purely from the interaction between the positive latitudinal gradient of PV (on isentropic surfaces) in the interior and the negative potential temperature gradient at the ground as is the case in the
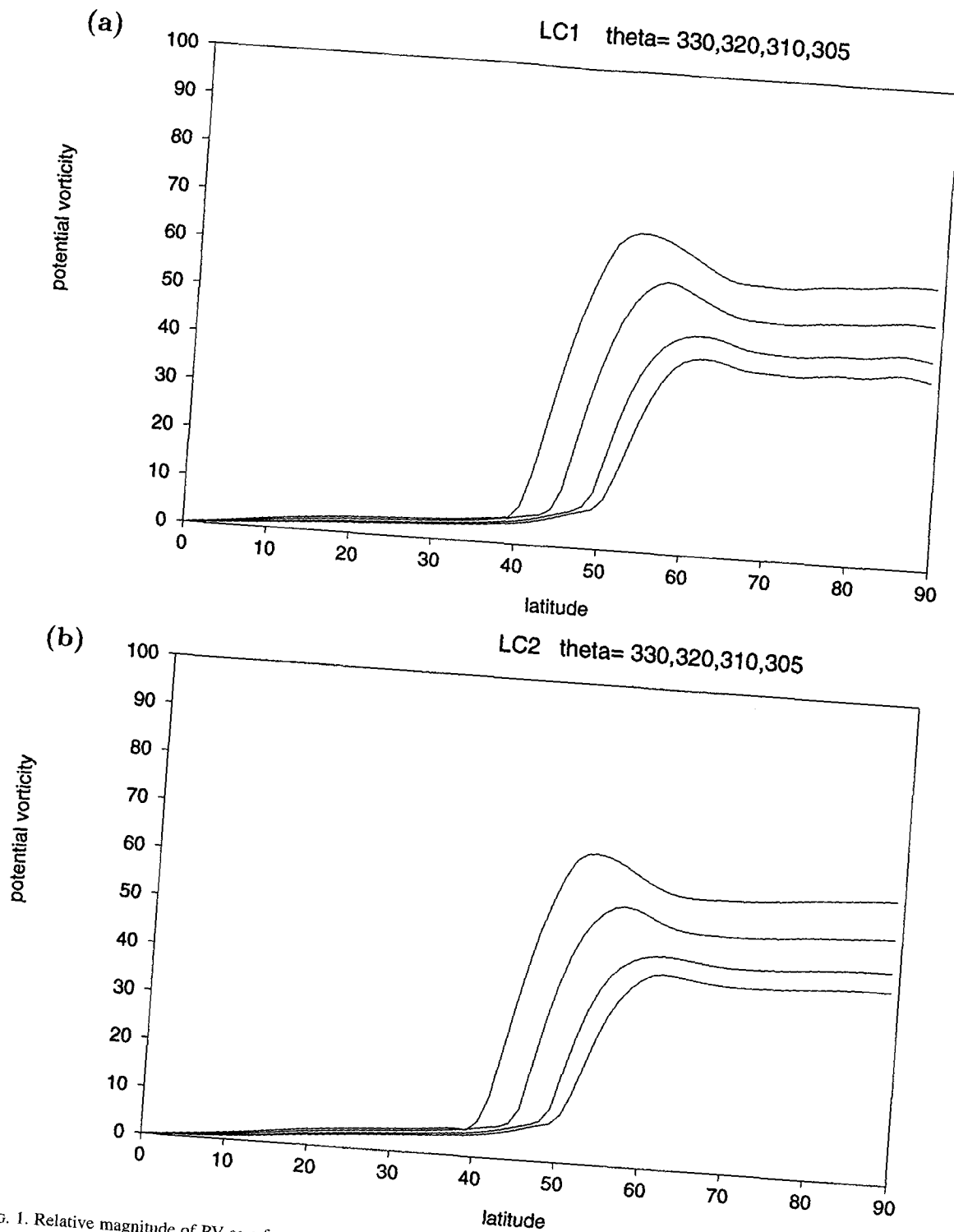
FIG. 1. Relative magnitude of PV as a function of latitude on isentropic surfaces 305 K (lowest PV), 310 K (next lowest PV), 320 K (next highest PV), and 330 K (highest PV) for the initial state of (a) LC1 and (b) LC2.

Charney problem, for instance. So it is also of some interest to determine whether the interior PV maximum on isentropic surfaces plays any substantial role in the evolution of the baroclinic disturbances. For example, it is possible that such a maximum might allow an internal barotropic instability to play an important role at

some stage during the life cycle, with little role for the surface potential temperature gradient.

Even when the basic-state PV distribution is non-monotonic, it is still possible to define a pseudomomentum by using Lagrangian information (McIntyre and Shepherd 1987). In principle, such information could be extracted from the numerical simulation. However, it is to be expected that the resulting pseudomomentum will not be single signed. It therefore loses its appeal as a diagnostic tool since there is always the possibility of the simultaneous appearance of balancing amounts of positive and negative pseudomomentum within the flow. The nonmonotonic PV distribution therefore means that for practical purposes it is not possible to take the initial states of LC1 and LC2 as the basic states, at least not without restriction. In fact it is possible to use the Eulerian expressions to evaluate wave activity in certain parts of the flow, that is, in those parts of the flow where fluid parcels originate from and stay within a span of latitudes with a monotonic distribution of PV. Some such diagnostics will be presented in sections 4 and 6. They provide a useful quantification of the differences between LC1 and LC2. More generally, it is necessary to generate alternative zonally symmetric states in which PV is everywhere a monotonic function of latitude. These symmetric states may be used for two purposes: either as basic states from which to calculate the wave activity diagnostics for the LC1 and LC2 runs, or as initial states for alternative baroclinic life cycle simulations. The first approach is reported in section 4, the second in section 5, and both are considered in section 6.

How are alternative states to be generated? Since the requirement is for PV to be a monotonic function of latitude on isentropic surfaces, the most straightforward way to meet this requirement is to specify the PV distribution on such surfaces. However, there are a number of technical problems with this approach associated with the fact that, in this study, the PV is not a model prognostic variable and potential temperature surfaces are not model coordinate surfaces. These issues and details of the rearrangement are discussed in section 3 and appendixes A and B.

2. Wave activity diagnostics, including lower-boundary contributions

The derivation of finite-amplitude wave activity conservation relations for the primitive equations, including that for angular pseudomomentum, can be found in H88. For completeness, we shall write down the expressions used here. As noted above, any application of pseudomomentum diagnostics requires the definition of a zonally symmetric, steady basic state that is a solution of the equations of motion. We shall let $(\cdot)_0$ denote a basic-state variable and $(\cdot)_e$ denote the deviation of that variable from the basic state, such that $(\cdot) = (\cdot)_0 + (\cdot)_e$. The independent variables are the usual spherical coordinates (λ, ϕ), longitude and latitude in the

horizontal, and potential temperature θ in the vertical; a is the radius of the earth. The form of the conservation law for pseudomomentum under adiabatic and frictionless conditions may be written as

$$\frac{\partial A}{\partial t} + \nabla \cdot \mathbf{F} = \frac{\partial A}{\partial t} + \frac{1}{a \cos\phi} \frac{\partial F^{(\lambda)}}{\partial \lambda}$$

$$+ \frac{1}{a \cos\phi} \frac{\partial (F^{(\phi)} \cos\phi)}{\partial \phi} + \frac{\partial F^{(\theta)}}{\partial \theta} = 0, \quad (2.1)$$

where the pseudomomentum density A is given by

$$A = -\sigma_e u_e \cos\phi$$

$$+ \sigma \int_0^{P_e} (P_e - \tilde{P}) \frac{\partial}{\partial \tilde{P}} m_0(P_0 + \tilde{P}, \theta) d\tilde{P} \quad (2.2)$$

and the components of the flux are given by

$$F^{(\lambda)} = uA - \frac{1}{2} \sigma_0 (u_e^2 - v_e^2) \cos\phi - \tilde{\tau}(p_e, p_0, \theta) \cos\phi,$$

$$(2.3a)$$

$$F^{(\phi)} = vA - \sigma_0 v_e u_e \cos\phi, \quad (2.3b)$$

$$F^{(\theta)} = g^{-1} a^{-1} p_e M_{e\lambda}. \quad (2.3c)$$

In the above, (u, v) are horizontal components of the velocity, $\sigma = -g^{-1} \partial p / \partial \theta$ is the pseudodensity, P is PV, and $M = \Pi\theta + gz$ is the Mongomery potential, where z is geometric height, $\Pi = c_p(p/p_s)^\kappa$ the Exner function, and p_s is a reference pressure. The function $m_0(\cdot)$ is defined by

$$m_0(P_0(\phi, \theta)) = \int_0^\phi \sigma_0(\check{\phi}, \theta) a \cos\check{\phi} d\check{\phi} \quad (2.4)$$

and is single valued when, for each θ, PV is a monotonic function of latitude. The corresponding m_0 represents a measure of position in the latitudinal direction based on the basic-state PV field on each isentropic surface. The function $\tilde{\tau}$ is defined by

$$\tilde{\tau}(p_e, p_0, \theta)$$

$$= g^{-1} \int_0^{p_e} \frac{\kappa\tilde{p}}{(p_0 + \tilde{p})} \Pi(p_0 + \tilde{p}, \theta) d\tilde{p}. \quad (2.5)$$

Note that the sign convention is different from that of H88 so that under quasigeostrophic scaling A agrees with the usual definition of Eliassen–Palm wave activity.

In the context of baroclinic instability, it is relevant to recall that the total wave activity in the interior is not necessarily conserved but that there is a reservoir of wave activity on the boundary, which may be freely exchanged with the interior of the flow. Indeed, in many examples of baroclinic instability, the simplest being the Charney problem, there is a flux of positive wave activity out of the boundary into the interior, and the growing disturbance is characterized by increasing positive wave activity in the interior and increasing negative wave activity on the boundary. Expressions

for the wave activity density on the lower boundary under quasigeostrophic scaling are well known and involve the boundary potential temperature distribution. See McIntyre and Shepherd (1987) and Shepherd (1989) for details, including the finite-amplitude case. Here we extend the H88 analysis to calculate the finite-amplitude contribution to the pseudomomentum from the boundary potential temperature gradients in the primitive equations. We give an expression only for the pseudomomentum density, that is, for the contribution to the pseudomomentum from any finite part of the boundary. It should also be possible to calculate contributions to the pseudomomentum flux, both into the domain from the boundary and along the boundary. However, to our knowledge this has not yet been done even for the quasigeostrophic case.

As in H88, we consider the sum of two global invariants, the angular momentum $\mathcal{M}$ about the earth's axis plus a suitable Casimir $\mathcal{C}$, being the integral over the fluid domain of the density multiplied by a function of PV and potential temperature. The boundary should be allowed to intersect potential temperature surfaces. Note that the boundary of the fluid domain is therefore free to move in three-dimensional (λ, ϕ, θ) space. Since both $\mathcal{M}$ and $\mathcal{C}$ are constants of the motion, it follows that the quantity $\mathcal{A} = -(\mathcal{M} + \mathcal{C} - \mathcal{M}_0 - \mathcal{C}_0)$, where $\mathcal{M}_0$ and $\mathcal{C}_0$ are, respectively, the angular momentum and the Casimir evaluated for the predefined basic state, is also a constant. The sign convention has been chosen, as earlier, to agree with the normal conventions of quasigeostrophic theory. In terms of dynamical variables for isentropic coordinates, $\mathcal{M} + \mathcal{C}$ is given by

$$\mathcal{M} + \mathcal{C} = \int_{\mathcal{D}} \{\sigma(u \cos\phi + \Omega a \cos^2\phi$$
$$+ C(P, \theta))\} a^2 \cos\phi d\theta d\phi d\lambda, \quad (2.6)$$

where C is the function defining the Casimir and $\mathcal{D}$ is the fluid domain. Thus $\mathcal{A}$ is given by

$$\mathcal{A} = -(\mathcal{M} + \mathcal{C} - \mathcal{M}_0 - \mathcal{C}_0)$$

$$= -\int_{\mathcal{D} \cap \mathcal{D}_0} \{\sigma(u \cos\phi + \Omega a \cos^2\phi + C(P, \theta))\}$$
$$\times a^2 \cos\phi d\theta d\phi d\lambda$$

$$+ \int_{\mathcal{D} \cap \mathcal{D}_0} \{\sigma_0(u_0 \cos\phi + \Omega a \cos^2\phi$$
$$+ C(P_0, \theta))\} a^2 \cos\phi d\theta d\phi d\lambda$$

$$- \int_{\mathcal{D}\backslash(\mathcal{D} \cap \mathcal{D}_0)} \{\sigma(u \cos\phi + \Omega a \cos^2\phi$$
$$+ C(P, \theta))\} a^2 \cos\phi d\theta d\phi d\lambda$$

$$+ \int_{\mathcal{D}_0\backslash(\mathcal{D} \cap \mathcal{D}_0)} \{\sigma_0(u_0 \cos\phi + \Omega a \cos^2\phi$$
$$+ C(P_0, \theta))\} a^2 \cos\phi d\theta d\phi d\lambda, \quad (2.7)$$

where $\mathcal{D}_0$ is the fluid domain in the basic state. It is useful to introduce the notation $\phi_b(\lambda, \theta)$ as the latitudinal position of the boundary for specified values of λ and θ and $\phi_{b0}(\theta)$ as the basic-state position of the boundary. Thus the boundary of $\mathcal{D}_0$ is defined by $\phi = \phi_{b0}(\theta)$, and the boundary of $\mathcal{D}$ by $\phi = \phi_b(\lambda, \theta)$. Note that the decomposition of the domain of integration is required because disturbed flow quantities are defined only in $\mathcal{D}$ and basic-state quantities in $\mathcal{D}_0$. Figure 2a shows a schematic representation of the projection of these different regions onto an isentropic surface. Figure 2b shows the corresponding slice through the domain at a fixed value of λ.

Similar considerations of the change in the fluid domain from the basic state to the disturbed state are necessary in the semigeostrophic case, when the boundaries appear as freely moving in geostrophic coordinates (Kushner and Shepherd 1995a,b).

It remains to show that, taking account of variations in the fluid domain, the function C may be chosen so that $\mathcal{A}$ is second order in wave quantities. Combining the first two integrals, the integrand may be written

$$-\{\sigma(u \cos\phi + \Omega a \cos^2\phi + C(P, \theta))\}$$

$$+ \{\sigma_0(u_0 \cos\phi + \Omega a \cos^2\phi + C(P_0, \theta))\}$$

$$= -\frac{1}{a \cos\phi} \frac{\partial}{\partial\lambda} (v_e C_P(P_0, \theta))$$

$$+ \frac{1}{a \cos\phi} \frac{\partial}{\partial\phi} (u_e C_P(P_0, \theta) \cos\phi)$$

$$- \sigma_e \{C(P_0, \theta) + (u_0 + a\Omega \cos\phi) \cos\phi$$

$$- P_0 C_P(P_0, \theta)\}$$

$$- u_e \left\{\sigma_0 \cos\phi + \frac{1}{a} \frac{\partial}{\partial\phi} [C_P(P_0, \theta)]\right\}$$

$$- \sigma_e u_e \cos\phi - \sigma C_2(P_0, P_e, \theta). \quad (2.8)$$

The first two terms are in the form of a divergence and may be written as a surface integral over the boundary of $\mathcal{D} \cap \mathcal{D}_0$. The last two terms are second order in disturbance quantities, with $C_2(P_0, P_e, \theta)$ being defined as $C(P, \theta) - C(P_0, \theta) - P_e C_P(P_0, \theta)$. It follows that the integral over $\mathcal{D} \cap \mathcal{D}_0$, neglecting the boundary terms for the time being, is second order in wave quantities if

$$\sigma_0 \cos\phi + \frac{1}{a} \frac{\partial}{\partial\phi} [C_P(P_0, \theta)] = 0 \quad (2.9a)$$

and

$$C(P_0, \theta) + (u_0 + a\Omega \cos\phi) \cos\phi$$
$$- P_0 C_P(P_0, \theta) = 0. \quad (2.9b)$$

These are just the conditions derived in H88 and they lead to the expression (2.2). As was noted there, the

two conditions are compatible, with (2.9a) being proportional to the ϕ derivative of (2.9b).

Returning to the boundary terms, we note that each is generally first order in disturbance quantities. The boundary integral arising from the divergence in (2.8) is generally first order because the integrand is proportional to terms such as u_e and v_e. The integrals over regions $\mathcal{D}\backslash(\mathcal{D}\cap\mathcal{D}_0)$ and $\mathcal{D}_0\backslash(\mathcal{D}\cap\mathcal{D}_0)$ are first order because the volume of such regions is proportional to the displacement of the boundary of the fluid domain. However, noting the equality in (2.9b), it follows that each of these terms is second order if

$$(u_0\cos\phi + \Omega a\cos^2\phi + C(P_0,\theta))|_{\phi=\phi_{b0}(\theta)}$$

$$= P_0 C_P(P_0(\phi_{b0}(\theta),\theta),\theta) = 0, \quad (2.10)$$

that is, if $C_P(P_0,\theta)$ vanishes on $\partial\mathcal{D}_0$, the boundary of $\mathcal{D}_0$.

Thus $\mathcal{A}$ is a second-order wave quantity provided that C satisfies the conditions in (2.9b) and (2.10), and it may be evaluated in the form

$$\mathcal{A} = \int_{\mathcal{D}\cap\mathcal{D}_0} Aa^2\cos\phi\, d\theta d\phi d\lambda$$

$$- \int_{\partial(\mathcal{D}\cap\mathcal{D}_0)} C_P(P_0,\theta)(v_e d\phi + u_e\cos\phi d\lambda)ad\theta$$

$$- \int_{\mathcal{D}\backslash(\mathcal{D}\cap\mathcal{D}_0)} \{\sigma[u\cos\phi + \Omega a\cos^2\phi$$

$$+ C(P,\theta)]\}a^2\cos\phi d\theta d\phi d\lambda$$

$$+ \int_{\mathcal{D}_0\backslash(\mathcal{D}\cap\mathcal{D}_0)} \{\sigma_0[u_0\cos\phi + \Omega a\cos^2\phi$$

$$+ C(P_0,\theta)]\}a^2\cos\phi d\theta d\phi d\lambda, \quad (2.11)$$

where the first integral is a contribution from the interior of the fluid domain and the second, third, and fourth integrals may be regarded as representing the boundary contribution. We are free to impose the condition (2.10) on C, as an initial condition on the differential equation (2.9b) providing that $\partial\mathcal{D}_0$ corresponds to only one value of ϕ on each θ surface. Problems would arise only if for each part of the fluid domain intersecting a θ surface there was more than one value of $\phi_{b0}(\theta)$. This is not the case if the value of θ at the surface, $\theta_{s0}(\phi)$, is a monotonic function of ϕ or has at most one maximum. Consider the case relevant to the real atmosphere, where $\theta_{s0}(\phi)$ has a low-latitude maximum, θ_s^{max}. Then for $\theta < \theta_s^{max}$ the intersection of the θ surface with the fluid domain consists of two distinct areas, each with a different low-latitude boundary, or else one area with a low-latitude boundary. Whether there are one or two such areas, the ordinary differential equation (2.9b) may be integrated from the low-latitude boundary, applying (2.10) as initial condition, to the pole, for each area independently. For $\theta \geq \theta_s^{max}$ the fluid domain covers the whole sphere
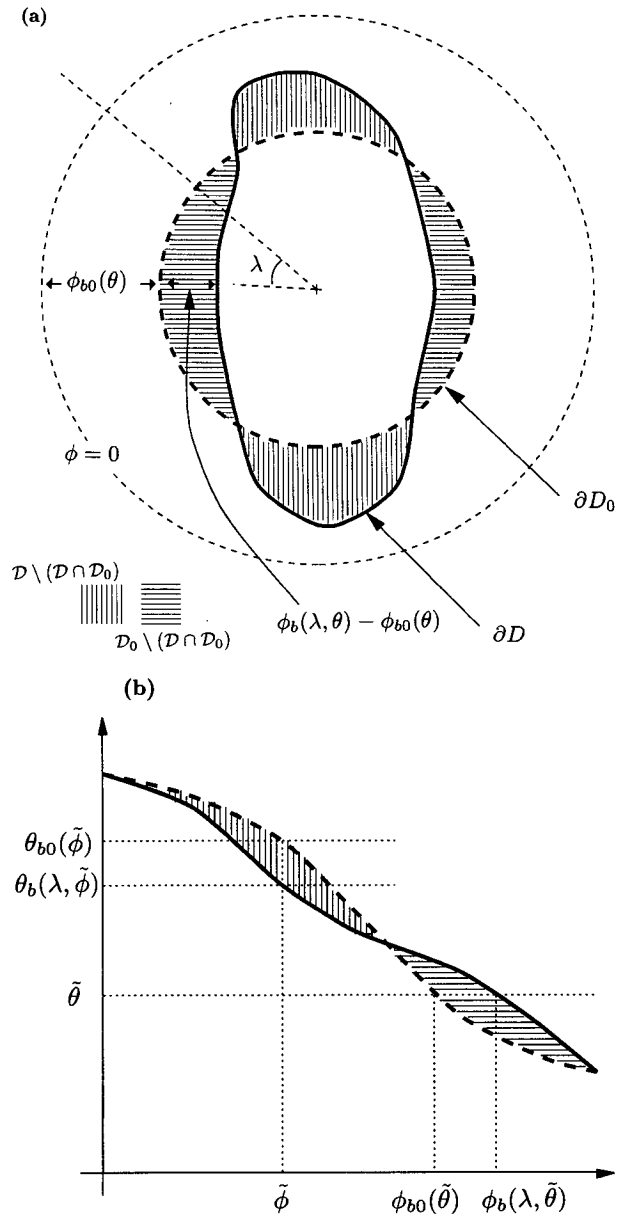

FIG. 2. (a) Schematic diagram of the fluid domain projected onto an isentropic surface in the basic state $\mathcal{D}_0$ and the disturbed state $\mathcal{D}$. $\partial\mathcal{D}$ (full contour) is the boundary of $\mathcal{D}$, and $\partial\mathcal{D}_0$ (dashed contour) is the boundary of $\mathcal{D}_0$; $\phi_{b0}(\theta)$ indicates the latitude of intersection of the fluid domain in the basic state with the surface of the earth; and $\phi_b(\lambda,\theta)$ indicates the latitude of intersection of the fluid domain in the disturbed state with the surface. As indicated in the lower-left-hand corner, the vertically hatched region is $\mathcal{D}\backslash(\mathcal{D}\cap\mathcal{D}_0)$, that is, the region containing points that are within the fluid domain in the disturbed state but outside the fluid domain in the undisturbed state. The horizontally hatched region is $\mathcal{D}_0\backslash(\mathcal{D}\cap\mathcal{D}_0)$, that is, the region containing points that are within the fluid domain in the undisturbed state but outside the fluid domain in the disturbed state. (b) Schematic diagram of a slice through the fluid domain at a fixed longitude, λ. The abscissa is potential temperature, and the ordinate is latitude. Again, the full contour indicates the disturbed state; the dashed contour indicates the basic state. $\phi_{b0}(\theta)$ and $\phi_b(\lambda,\theta)$ have the same meaning as in (a). $\theta_{b0}(\phi)$ indicates the basic state's potential temperature at latitude ϕ, and $\theta_b(\lambda,\phi)$ is the disturbed state's surface surface potential temperature at longitude λ and latitude ϕ.

and the condition (2.10) is not needed. In that case the second-order property holds irrespective of the initial condition applied to (2.9b). For the case where the lower boundary is a θ surface, $\mathcal{D} = \mathcal{D}_0$, so the third and fourth terms vanish, while the integration on each θ surface is over the whole sphere, so the second term, a boundary integral, vanishes. It is only when there are potential temperature variations on the lower boundary that these terms give rise to a boundary contribution to the wave activity.

It is interesting to evaluate the boundary contribution in the limit of small wave amplitude. To fix ideas we assume the "Northern Hemisphere" configuration, two different views of which are shown in Figs. 2a and 2b, where $\mathcal{D}_0$ is the region $\phi > \phi_{b0}(\theta)$ and $\mathcal{D}$ is the region $\phi > \phi_b(\lambda, \theta)$. Note that $\partial(\mathcal{D} \cap \mathcal{D}_0)$ coincides in part with $\partial\mathcal{D}$ and in part with $\partial\mathcal{D}_0$, depending on whether $\phi_{be}(\lambda, \theta)$ is greater than or less than zero. On $\partial\mathcal{D}_0$, $C_P(P_0, \theta)$ is zero by definition. Thus there is a contribution to the first boundary term in (2.11) only when $\phi_{be} > 0$. There $C_P(P_0, \theta)$ may be approximated by $C_{PP}(P_0, \theta)P_{0\phi}\phi_{be}$ and hence by $-a\sigma_0\phi_{be}\cos\phi_{b0}$, using (2.9a). Furthermore the $v_e d\phi$ term may be neglected at leading order, so that the integral is over λ and θ, with the leading-order integrand $-a\sigma_0\cos^2\phi_{b0}u_e\phi_{be}$.

Regarding the second and third boundary terms, the integrands may be approximated using linearized expressions and then simplified using (2.9a), (2.9b), and (2.10) to give

$$[\sigma_0 u_e + (\phi - \phi_{b0})\sigma_0 P_0 C_{PP}(P_0, \theta)P_{0\phi}]a^2 \cos\phi_{b0}$$

(2.12a)

and

$$[(\phi - \phi_{b0})\sigma_0 P_0 C_{PP}(P_0, \theta)P_{0\phi}]a^2 \cos\phi_{b0}, \quad (2.12b)$$

respectively, where the second is evaluated only for λ such that $\phi_{be} < 0$ and the third for λ such that $\phi_{be} > 0$. The regions are defined, respectively, by $\phi_{be} = \phi - \phi_{b0} < 0$ and $0 < \phi - \phi_{b0} = \phi_{be}$. From the above with further use of (2.9a) it follows that the boundary contribution is, to second order in disturbance quantities,

$$-\int d\theta \int \sigma_0 \left(u_e\phi_{be} + \frac{1}{2} a\sigma_0 P_0 \phi_{be}^2 \right) a^2 \cos^2\phi_{b0} d\lambda,$$

(2.13)

where quantities inside the integral are to be evaluated for $\phi = \phi_{b0}(\theta)$ and the θ integral is over the range of θ appearing on the boundary.

An alternative route to the linearized expression for the boundary contributions to the wave activity is via the linearized form of the EP relation in isentropic coordinates, given by Andrews (1987). Here straightforward integration over the fluid domain gives a contribution to the rate of change of wave activity from the flux out of the lower boundary. This flux may be shown

to be equal to minus the rate of change of the integral of a second-order wave quantity over the lower boundary and the latter quantity, which is identical to (2.13), may therefore be interpreted as a boundary contribution to the total wave activity, subject, of course, to the usual caveats about nonuniqueness. Details are given in appendix C.

It is well known that within quasigeostrophic theory the pseudomomentum includes a contribution involving the boundary potential temperature distribution. As noted earlier, this term plays a crucial role in Charney-type baroclinic instabilities. The "boundary" terms appearing in (2.11) are the exact analogs of this contribution. The linearized form of these terms, given in (2.13), makes the connection more explicit. Note that it is the second term that dominates under quasigeostrophic scaling and that the only wave quantity induced is ϕ_{be}, determined by the boundary configuration in θ coordinates, or, equivalently, by the θ distribution (as a function of λ and ϕ) on the boundary. Note also that this term is negative as expected by analogy with quasigeostrophic results.

For completeness we note that it is possible to proceed from the finite-amplitude boundary contribution in (2.11) to the finite-amplitude expression for the boundary contribution in quasigeostrophic theory given by Shepherd (1989). First, note that under quasigeostrophic scaling the integrand in the second term of (2.11) is $O(\text{Rossby number})$ smaller than the integrands in the third and fourth terms. Furthermore, at leading order in Rossby number, the integrands in both third and fourth terms may be approximated by $P_0(\phi, \theta)C_P(P_0, \theta)$. Denoting the ϕ-independent leading-order contribution to σ_0 as $\bar{\sigma}$, it follows that the leading-order approximation to the integrand is

$$\bar{\sigma} a^3 f_0(\sin\phi - \sin\phi_b(\theta)) \cos\phi, \quad (2.14a)$$

where f_0 is the Coriolis parameter, evaluated for a suitable latitude. Note that at leading order under quasigeostrophic scaling, $\bar{\sigma}$ may be treated as constant over the boundary, since the boundary itself must be closely approximated by a θ surface. The leading-order approximation to the boundary terms in (2.11) is therefore

$$\int \cos\phi d\phi d\lambda \int_{\theta_b(\lambda,\phi)}^{\theta_{b0}(\phi)} \bar{\sigma} a^3 f_0\{\sin\phi - \sin\phi_b(\theta)\} d\theta,$$

(2.14b)

where $\theta_b(\lambda, \phi)$ is the distribution of θ on the lower boundary, as a function of λ and ϕ, and $\theta_{b0}(\phi)$ is the basic-state value. Integrating by parts, and noting that $\phi_{b0}(\theta_{b0}(\phi)) = \phi$, it follows that the boundary contribution is

$$\iint d\phi d\lambda \bar{\sigma} a^3 f_0 \cos\phi$$

$$\times \int_{\theta_{b0}(\phi)}^{\theta_b(\lambda,\phi)} (\theta_b - \tilde{\theta}) \frac{d}{d\tilde{\theta}} \{\sin\phi_{b0}(\tilde{\theta})\} d\tilde{\theta}. \quad (2.15)$$

Note that this is negative if $\phi_b(\theta)$ is a decreasing function and $f_0 > 0$, as has been assumed above. Under the usual quasigeostrophic restriction that the flow is confined to a small range of latitudes, the above may be shown equal to the finite-amplitude expression given by Shepherd (1989), taking account of an extra factor of $\cos\phi$ since here it is explicitly the angular pseudomomentum, rather than the (linear) pseudomomentum on a β plane, that has been considered.

A final note concerns the numerical calculation of the boundary contribution to the wave activity. The second term in (2.11) is, as noted, second order in wave quantities but requires integration over the boundary of $\mathcal{D} \cap \mathcal{D}_0$, which may be highly contorted. Since it is computationally more straightforward to perform a volume integral, the following strategy is proposed: $\mathcal{D} \cap \mathcal{D}_0$ in Fig. 2a is divided into two parts, $\bar{\mathcal{D}}$ and ($\mathcal{D} \cap \mathcal{D}_0) \setminus \bar{\mathcal{D}}$, where $\bar{\mathcal{D}}$ is the largest longitudinally symmetric region that lies entirely within $\mathcal{D} \cap \mathcal{D}_0$. Then the first two terms in (2.11) may be rewritten as the integral of A over $\bar{\mathcal{D}}$ plus a boundary integral over $\partial\bar{\mathcal{D}}$ plus the integral of the left-hand side of (2.8) over ($\mathcal{D} \cap \mathcal{D}_0) \setminus \bar{\mathcal{D}}$. The resulting expression for the total wave activity is

$$\mathcal{A} = \int_{\bar{\mathcal{D}}} A a^2 \cos\phi\, d\theta d\phi d\lambda$$

$$- \int_{\partial\bar{\mathcal{D}}} a C_P(P_0, \theta) u_e \cos\phi\, d\lambda d\theta$$

$$- \int_{(\mathcal{D} \cap \mathcal{D}_0)\setminus\bar{\mathcal{D}}} \{ \sigma(u\cos\phi + \Omega a \cos^2\phi + C(P,\theta))$$

$$- \sigma_0(u_0\cos\phi + \Omega a \cos^2\phi + C(P_0,\theta)) \}$$

$$\times\, a^2 \cos\phi\, d\theta d\phi d\lambda$$

$$- \int_{\mathcal{D}\setminus(\mathcal{D} \cap \mathcal{D}_0)} \sigma(u\cos\phi + \Omega a \cos^2\phi$$

$$+ C(P,\theta))a^2 \cos\phi\, d\theta d\phi d\lambda$$

$$+ \int_{\mathcal{D}_0\setminus(\mathcal{D} \cap \mathcal{D}_0)} \{ \sigma_0(u_0\cos\phi + \Omega a \cos^2\phi$$

$$+ C(P_0, \theta)) \} a^2 \cos\phi\, d\theta d\phi d\lambda. \quad (2.16)$$

While the integrand in the third term is first order in wave quantities, it is being integrated over a region ($\mathcal{D} \cap \mathcal{D}_0) \setminus \bar{\mathcal{D}}$ that is first order. Therefore in this alternative representation of the wave activity the various terms are all second order, and calculation of them is likely to be well conditioned. Furthermore, since the difference between $\mathcal{D} \cap \mathcal{D}_0$ and $\bar{\mathcal{D}}$ is first order, it seems valid to regard the integral over $\bar{\mathcal{D}}$ as the only "interior" term and the remaining integrals as boundary terms.

3. Generating alternate basic states

If the wave activity densities and fluxes are to be evaluated in terms of Eulerian flow quantities, then PV

in the basic state must vary monotonically in latitude on each isentropic surface [see McIntyre and Shepherd (1987) for a further discussion]. As we have already noted, perhaps the most natural choice for the basic state, that of the zonally symmetric initial state, does not satisfy this requirement for LC1 and LC2. The solution to this problem is, in principle, simple, namely to choose an alternative basic state as close as possible to the initial state but with a monotonic PV distribution.

The first approach taken was of simply applying a heuristic correction to the velocity field and then using the model-balancing routine to obtain the mass field. However it proved difficult to find such a correction that was large enough to make the PV monotonic but small enough for the new basic state to be useful. A second, more systematic approach was to specify the PV field in each isentropic layer, so that meeting the monotonicity condition was guaranteed. The specified PV field was then inverted—assuming gradient wind balance and taking account of spherical geometry and the nonisentropic lower boundary—to deduce the velocity and pressure distributions. Details of the inversion calculation are given in appendix A.

Within the constraints of the monotonicity requirement, there are many different possible specifications of the PV field. Clearly, one possibility would have been to solve the variational problem for the PV field that, within the monotonicity requirement, minimizes some measure of the initial wave activity (that of the initial state with respect to the alternate basic state). The chosen approach was to "rearrange" the PV field in each layer so that it was monotonic. By "PV rearrangement" we mean that fluid elements in each isentropic layer are redistributed in latitude, with each element preserving its value of PV. We considered two different approaches to such rearrangement, both of which are discussed in detail in appendix B. First, we required that the area of each element, as well as its PV value, was preserved during the rearrangement. This approach is ad hoc but convenient because the rearrangement defines the new PV field directly as a function of ϕ and θ. However, when the resulting PV field was inverted, it turned out that mass had been lost; the resulting surface pressure was too low. This is perfectly acceptable in principle, but in practice the corresponding basic state is too far from the initial state to be of use. Second, we required that the mass of each element, as well as its PV value, was preserved during the rearrangement. This second approach might be argued as less ad hoc because the resulting alternate basic state is, in principle, accessible from the initial state through purely adiabatic and frictionless evolution. However, it is also technically more complicated than the first because the rearrangement defines the new PV field in each isentropic layer as a function of mass, not of ϕ. The position of each element is unknown and must be solved for as part of the inversion. Thus, the rearrangement became closely linked with the inversion process, updating the density of fluid elements, and therefore

their location, as we iterated toward a solution. This second approach gave satisfactory results; however, if one were to require more flexibility, for example, wanting to impose a certain minimum isentropic gradient to the PV distribution, it would seem natural to relax the requirement of PV conservation for each element but to continue to impose that the mass-weighted integral over isentropic layer be constant; that is, $\Delta(\int \sigma P dV)$ = 0. This is equivalent to the requirement that the total circulation around the boundary of each isentropic layer remain constant. In this case one might think of the PV rearrangement in going from basic state to initial state or vice versa as being accomplished by notional diabatic and frictional processes. Potential vorticity would have been diluted in some regions and concentrated in others; however the total "PV-substance," or the total mass-weighted PV (Haynes and McIntyre 1990), would have been conserved along with the total mass in that layer.

As it turned out, the PV distribution obtained with the second method, that of PV and mass conservation, had gradients close to zero in some regions of the flow. In principle, this poses no problem; that is, the basic state need not have a strictly monotonic PV distribution, it needs only PV gradients that do not change sign on isentropic surfaces. Practical problems arise because the model is formulated in sigma coordinates with discrete vertical levels, whereas the idea of rearrangement of PV in isentropic layers requires that the inversion algorithm be formulated in isentropic coordinates. The PV field on isentropic surfaces is extracted from the sigma coordinate model and inserted into the inversion algorithm. The PV is rearranged and inverted to give the corresponding wind and temperature fields, which must then be reinserted into the sigma coordinate model. The first important point is that, even if there is no PV rearrangement, the wind and temperature fields inserted into the model are not the same as those in the original state. The second important point is that, while such wind and temperature fields have been derived from a PV inversion algorithm, there is no guarantee that they are in balance in the sigma coordinate model. Both of these inconsistencies arise, in our view, because the finite vertical resolution of the sigma coordinate model implies that there is no simple expression of the model dynamics in isentropic coordinates. It is possible that the idea of PV rearrangement in isentropic layers may be formulated in such a way that it is exactly consistent with the finite vertical resolution sigma coordinate model. However, since our aim is the practical one of generating new and useful basic states, we have chosen not to pursue this possibility. We emphasize that the usefulness of any ideas of PV rearrangement, which has been suggested in a number of contexts (e.g., McIntyre 1982) as an approach to numerical experimentation, depends on such inconsistencies not being too great at practical model resolution. Of course, one of the great attractions of a model formulated in isen-

tropic coordinates would be that such inconsistencies would not arise.

The practical approach taken, bearing in mind the above, was as follows. First the PV was interpolated from the sigma levels of the numerical model used for the life cycle integrations to the potential temperature surfaces of the inverter. Two inversions were then performed, one for this PV field and one for a rearranged PV field. The difference between these two inversions represents the differences in the various balanced fields due to the rearrangement. Rather than reinterpolating the rearranged balanced winds and temperatures back onto sigma surfaces and then reinserting them into the model, the two inversions were used to calculate the difference in the wind field on sigma surfaces due to the rearrangement, and this difference was then added to the original winds in the sigma model.

When PV of the resulting state was computed and then interpolated to isentropic surfaces, the resulting PV distribution was no longer monotonic everywhere. The deviations were, however, minor and were dealt with in a consistent manner.

4. Results for the LC1 and LC2 life cycles

The model used throughout this study is that used by THM, that is, the baroclinic sigma-coordinate spectral model developed by Hoskins and Simmons (1975), at horizontal resolution of T95 restricted to zonal wavenumbers with wave-6 symmetry and with 15 sigma levels in the vertical; at sigma = 0.967, 0.887, 0.784, 0.674, 0.569, 0.477, 0.400, 0.338, 0.287, 0.241, 0.197, 0.152, 0.106, 0.060, and 0.018. The model is adiabatic and frictionless except for ∇^6 hyperdiffusion that is added to the vorticity, divergence, and temperature tendency equations with a decay rate of $(4\text{ h})^{-1}$ for the smallest horizontal scales. For the LC1 and LC2 runs we used exactly the same zonally symmetric initial states as THM with a 0.001-mb noise added to the surface pressure field at zonal wavenumber 6. Figures 3a,b show the initial states for LC1 and LC2, respectively. The only difference in initial states for the two life cycles is the addition of a barotropic component to the wind that is 10 m s^{-1} westerly at 20°N, 10 m s^{-1} easterly at 50°N, and varies linearly between the two latitudes. THM found that development of the life cycle was not sensitive to the structure and amplitude of the disturbance except in timing of various features of the life cycle. Thus we used small-amplitude noise rather than the fastest-growing normal mode as our initial disturbance. Consequently, the timing of various events in our life cycle simulations is delayed from that of THM by about 11–12 days.

As noted earlier, because of the nonmonotonic distribution of PV in the initial state of both LC1 and LC2, the initial states could not be used as basic states for the purposes of calculating wave activity diagnostics everywhere in the flow. Two approaches were taken to calculating the pseudomomentum flux and density.
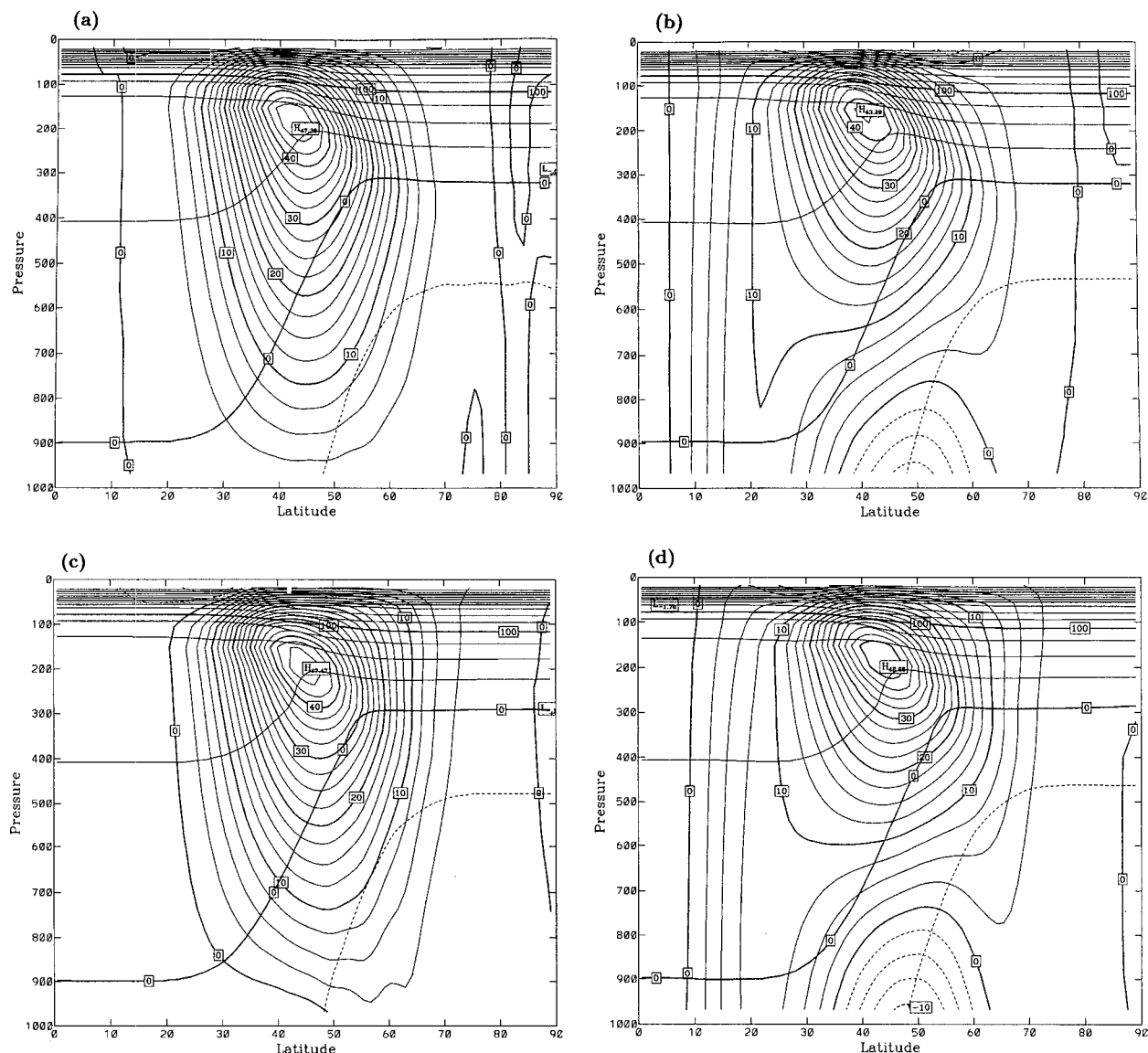
FIG. 3. (a) Initial winds (contour interval 2.5 m s^{-1}) and potential temperatures (contour interval is 25 K; 300 K has been subtracted) for LC1, Z1. (b) As in (a) except for LC2, Z2. (c) As in (a) except for the modified LC1, Z1R. (d) As in (a) except for the modified LC2, Z2R.

First, we used the initial state as the basic state and were therefore able to calculate pseudomomentum flux and density reliably only equatorward of the latitude $\phi_{max}(\theta)$, at which there is a maximum in the initial PV on each isentropic surface. This is with the additional proviso that the fluid in the disturbed state equatorward of $\phi_{max}(\theta)$ did not arrive from latitudes poleward of $\phi_{max}(\theta)$ in the initial state. Let us call the basic states corresponding to LC1 and LC2 with the above restrictions Z1 and Z2, respectively. Second, for each of LC1 and LC2 we used not the initial state but another dynamically consistent state (close to it in a sense explained in section 3) as the basic state for calculating wave activity diagnostics for LC1 and LC2. Let us call these alternative basic states for LC1 and LC2 Z1R

and Z2R, respectively. Figure 3c shows the state Z1R, winds, and potential temperature on sigma levels, and Figure 3d shows the state Z2R. Note that the differences between Z1 and Z1R, for instance, are considerably smaller than the differences between Z1 and Z2, but they are still quite visible. In particular the jet in Z1R has been shifted slightly poleward relative to that in Z1. The same is true for Z2 and Z2R. Wave activities were then calculated using these alternative states as basic states. Z1R and Z2R are generated from Z1 and Z2, respectively, as explained in section 3. Details of how the PV of Z1 and Z2 is rearranged are given in appendix B and details of how the PV of Z1R and Z2R is inverted are discussed in appendix A.

Here we shall concentrate on three of the four stages of life cycle development mentioned earlier:

1) linear growth,
2) upward and equatorward propagation of wave activity, and
3) low-latitude nonlinear saturation.

a. Wave activity diagnostics for LC1

Figures 4a–h show the zonally averaged meridional-plane (latitude and potential temperature) pseudomomentum flux and the zonally averaged pseudomomentum density referring to basic-state Z1 for days 17, 20, 22, and 23. Figures 5a–h show the same fields on the same days when referring to basic state Z1R. We have marked the location of the tropopause as a bold line superimposed on the contours of pseudomomentum density in Fig. 4b. Since the basic-state PV gradients for Z1 are not monotonic, in Fig. 4 we have only computed wave activity in areas equatorward of the latitude $\phi_{max}(\theta)$, at which there is a maximum in the initial PV on each isentropic surface. There is only minor transfer of fluid parcels across $\phi_{max}(\theta)$ as evidenced, for example, by the evolution of PV on isentropic surfaces. This is further reinforced by the similarity of the diagnostics in Figs. 4 and 5, when the wave has grown enough for the difference between basic and initial states to be of secondary importance. Note that Figs. 4 and 5 show only the region above the 300-K θ surface, since θ surfaces lower than this intersect the ground. Referring to Figs. 3a–d, it may be seen that this effectively concentrates attention to the area above the mid-troposphere, at least in midlatitudes.

In previous papers, extensive use has been made of EP fluxes for diagnosing life cycles. We have therefore included EP fluxes for days 17 and 20 in Figs. 6a,b. We choose to include only those two days because EP fluxes can be interpreted as the fluxes of a conserved quantity only for small-amplitude disturbances. Beyond day 20 the wave is of large amplitude, and the usefulness of the EP flux as a wave–activity flux is diminished. In order to facilitate comparison with the wave-activity fluxes, we have superposed potential temperature contours in Figs. 6a,b. Note that it is not a trivial matter to transform conventional EP fluxes to pseudomomentum fluxes since the dissection of the flow into "basic state" and "wave" parts is different in the two cases. In the former, wave quantities are defined as the difference from the basic-state value measured on pressure surfaces, while in the latter they are defined as the difference from the basic-state value on isentropic surfaces [see Andrews (1987) for a more thorough discussion]. If we made a purely geometric transformation of the EP flux pattern from using pressure as vertical coordinate to using zonal-mean potential temperature as a vertical coordinate, then the transformation law for the flux components would be

$$\mathscr{F}_{EP}^{(\phi)} = F_{EP}^{(\phi)}/(\partial\bar{\theta}/\partial p),$$

$$\mathscr{F}_{EP}^{(\theta)} = F_{EP}^{(p)} - \left(\frac{\partial p}{\partial\phi}\right)_{\bar{\theta}} F_{EP}^{(\phi)}, \qquad (4.1)$$

where $\mathscr{F}$ indicates a component of the flux in isentropic coordinates and $(\partial p/\partial\phi)_{\bar{\theta}}$ indicates that the derivative is taken holding $\bar{\theta}$ fixed. In other words, there would be an extra term added to the vertical component of the flux, whereas the horizontal component would simply be multiplied by a certain factor.

Day 17 is at a stage in the life cycle LC1 when the disturbance is still linear in the exponential growth stage. Figures 4a,b show the concentration of wave activity along the tropopause in the linear mode. As for any such growing mode, the wave activity flux pattern in Fig. 4a is convergent, leading to exponential growth in the pattern of wave activity density seen in Fig. 4b. The differences between Figs. 4a–h and 5a–h are solely due to differences in basic state. These differences are small for the fluxes in Figs. 4a and 5a and very large for the wave-activity densities in Figs. 4b and 5b. The pattern in Fig. 5b is totally dominated by contributions due to differences between the initial state and the basic state. This difference is predominantly in the zonally averaged flow and is certainly not associated with "waves." There is therefore little useful information in Fig. 5b at this time. Figures 4a and 5a are reminiscent of the upper part of a Charney-like EP flux diagram associated with a normal mode [see Fig. 2b in Edmon et al. (1980)]. However, note how the wave activity is distorted from the Charney-like picture by the presence of the tropopause—the wave activity flux extends along the tropopause. A similar distortion is also perceptible in the corresponding EP flux, which is typical of early, linear growth seen in Figure 6a.

Day 20 corresponds to the upward-propagation stage of the life cycle. The shape of the flux pattern has now changed from that in Figs. 4a and 5a. In Fig. 5d there is still some signature of the difference between the basic state and the initial state especially at upper levels at low latitude and also at high latitudes (cf. Fig. 5b), but in general the pattern is dominated by the baroclinic disturbance. In both cases, Figs. 4c–4d and 5c–5d, wave activity is now filling up the area in the troposphere along the tropopause, stretching far into low latitudes at upper levels, as shown by the pseudomomentum flux, Figs. 4c and 5c, and by the pseudomomentum density, Figs. 4d and 5d. The same signature has, of course, been identified previously in the EP fluxes and is apparent in Fig. 6b. The pseudomomentum density and flux together suggest that this propagation is intimately connected with the tropopause itself. At lower levels, note the lateral broadening of the area of upward EP flux also seen in the pseudomomentum flux.

By days 22–23 LC1 has reached its final stage of low-latitude, upper-tropospheric saturation. Note the change that has occurred in the pseudomomentum flux in Figs. 4e and 4g and Figs. 5e and 5g. On day 22 the
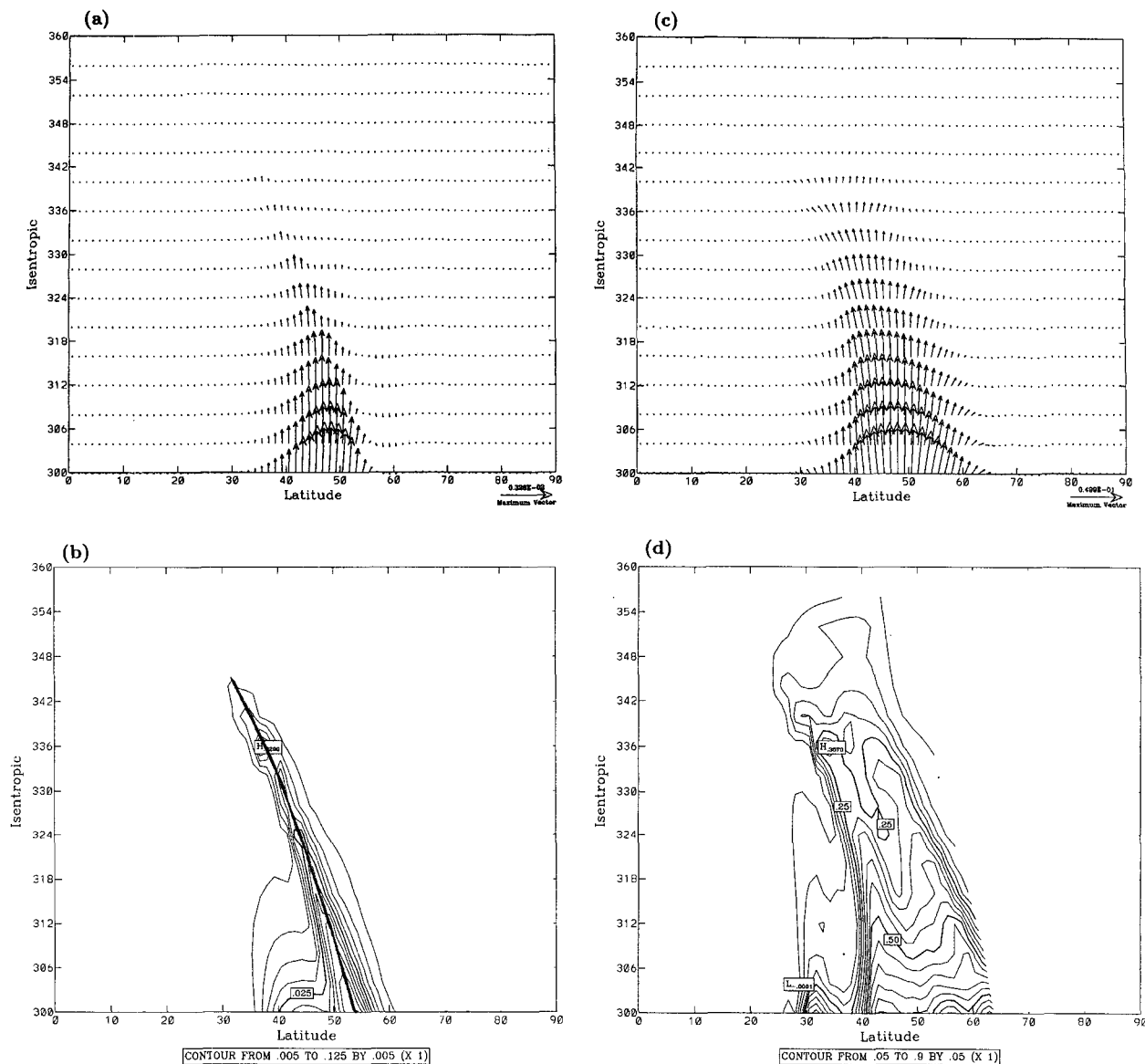
FIG. 4. (a) Zonally averaged wave activity flux for LC1 referring to basic-state Z1 on day 17. The maximum vector (indicated in the lower-left corner) has magnitude of 3.26×10^{-3}. (b) Zonally averaged wave activity density for LC1 referring to basic-state Z1 on day 17. Contour interval is 5.00×10^{-3}. The heavy contour indicates the location of the tropopause. (c) As in (a) except on day 20 and the maximum vector is of length 4.99×10^{-2}. (d) As in (b) except on day 20 and the contour interval is 5.00×10^{-2}. (e) As in (a) except on day 22 and the maximum vector length is 2.26×10^{-2}. (f) As in (b) except on day 22 and the contour interval is 0.100. (g) As in (a) except on day 23 and the maximum vector length is 2.77×10^{-2}. (h) As in (b) except on day 23 and the contour interval is 0.100.

region of flux has been shifted equatorward, and the flux tends to be more horizontal. A similar pattern often shows up in the EP fluxes and indeed did in this case. (See Fig. 15c of THM for an explicit example.) This has often been interpreted as horizontal propagation of wave activity. However we shall note below evidence that it is mostly associated with advective transport of wave activity rather than wave propagation. Given the fact that the advective contribution to the wave activity flux is not included in the EP flux and that the wave is saturating at upper levels, the agreement between EP

fluxes and wave activity fluxes at this stage would seem to be purely fortuitous. On day 23 (Figs. 4g and 5g) the wave activity flux is pointing downward, and we might interpret this as evidence of reradiation from the nonlinear flow above. Note that the maximum length of the flux arrows on day 23 represents about half the magnitude of their maximum length on day 20, when there was maximum upward propagation of wave activity. The explanation for the double maximum in wave activity density in latitude—seen in Figs. 4d, 5d, 4f, 5f, and more pronounced in Figs. 4h and 5h—will
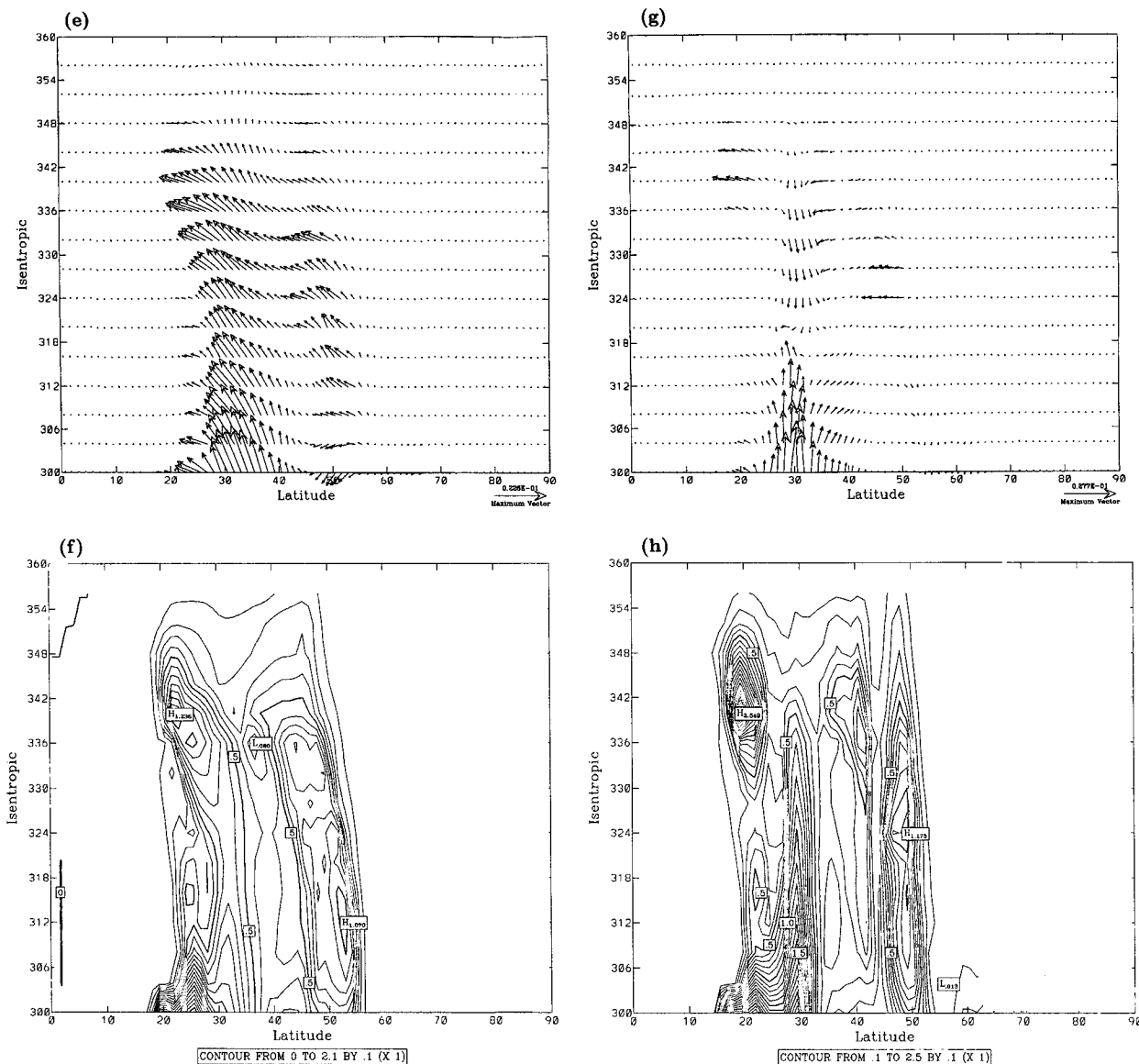
FIG. 4. (*Continued*)

become evident when we look at the field on isentropic surfaces. On day 23 the pseudomomentum density has filled the entire troposphere above the 300 K surface from about 50° to 20°N, as seen in Figs. 4h and 5h. Corresponding to the flux pictures in Figs. 4c and 5c and 4e and 5e, a pronounced maximum in pseudomomentum density at 20°N and 336–340 K is building up, as seen in Figs. 4h and 5h.

Interpretation of the wave activity densities and flux pictures is clarified if we examine the fields on an upper-level isentropic surface. Here we shall concentrate on 336 K, which is at the lower edge of the low-latitude maximum in pseudomomentum density. Figures 7a–7c show PV on the 336-K surface on days 20, 22, and 23, respectively. On day 20 there are large-scale undula-

tions in the PV contours, indicating the presence of waves of moderate amplitude. (The smaller-scale structure to the PV contour closest to the equator is almost certainly a numerical artifact, and we do not regard it of any significance.) By day 22, tongues of high PV air, indicative of Rossby wave breaking, stretch into low latitudes and on day 23 are beginning to roll up into vortices. That the pseudomomentum density is dominated by these PV features may be clearly seen in Figs. 8a, 8b, 8c corresponding to days 20, 22, and 23, respectively. The basic state in these calculations is Z1. Closer investigation of Fig. 7a reveals that the meridional contour displacements are comparable to the width of the region of concentrated gradients and suggests that the linearized expression for the pseu-
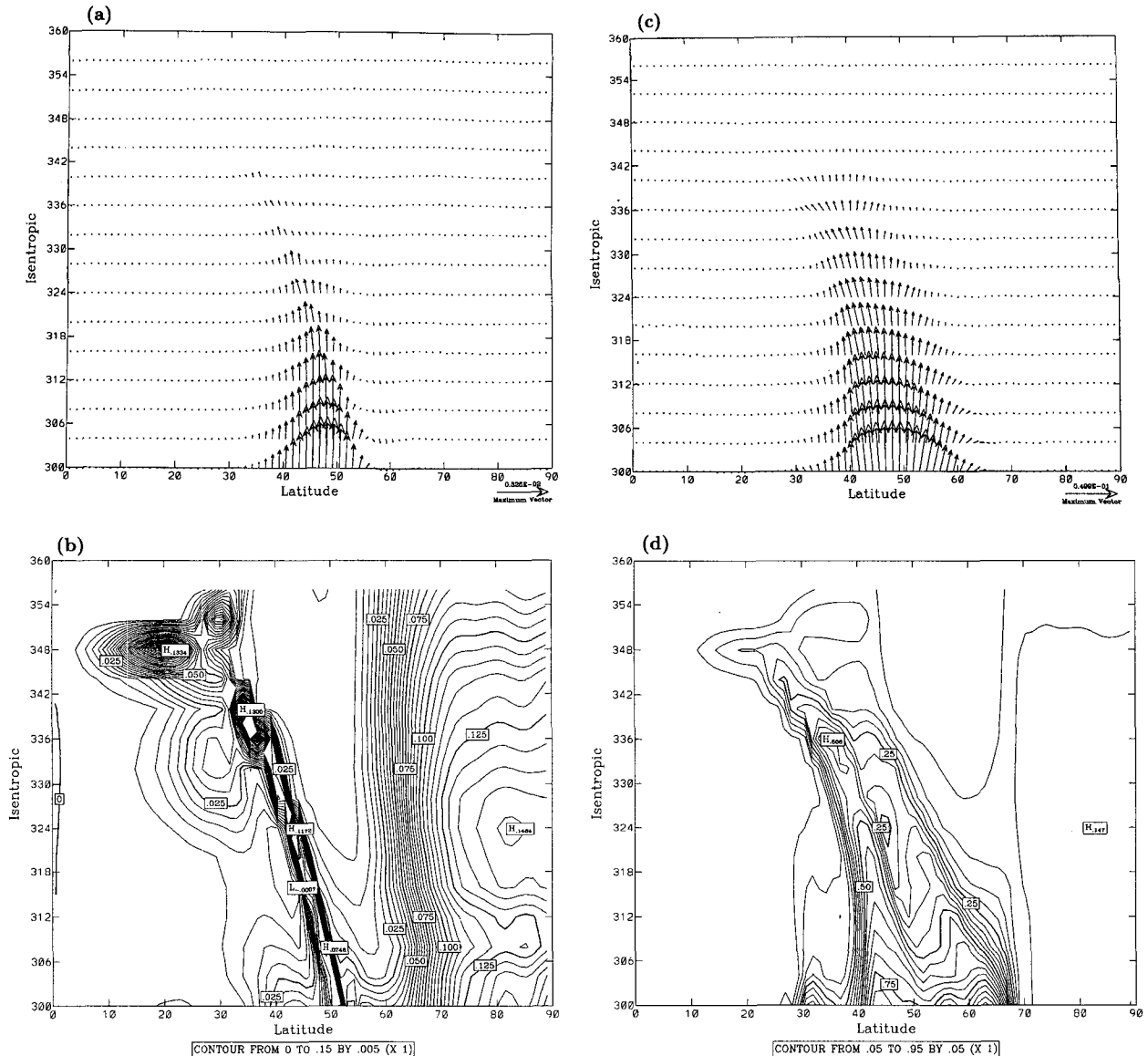
FIG. 5. (a) Zonally averaged wave activity flux for LC1 referring to basic-state Z1R on day 17. The maximum vector (indicated in the lower-left corner) has magnitude of 3.25×10^{-3}. (b) Zonally averaged wave activity density for LC1 referring to basic-state Z1R on day 17. Contour interval is 5.00×10^{-3}. (c) As in (a) except on day 20 and the maximum vector is of length 4.99×10^{-2}. (d) As in (b) except on

domomentum would not be quantitatively accurate. Indeed, comparison between Fig. 7a and the pseudomomentum density on the 336-K surface, shown in Fig. 8a, suggested that the double-peak structure in the zonally averaged pseudomomentum seen in Fig. 4d or 5d is primarily associated with the substantial meridional displacements rather than any complicated structure of a linear wave. Nevertheless, on the basis of Fig. 7a, the interpretation of this part of the life cycle, as that of upward and equatorward propagation of wave activity along the tropopause, would seem justified. Wave propagation ideas are not so clearly relevant to later stages, however. The location of the largest fluxes in Fig. 4e with the largest densities in Fig. 4f suggests that the

equatorward flux, at least, is mostly associated with advection of PV and pseudomomentum, corresponding to the first term on the right-hand sides of (2.3a) and (2.3b). Indeed, when the different contributions to the total wave activity flux on day 22 on the 336-K surface were examined, it turned out that the advective contribution was about four times greater than the nonadvective contribution. It would therefore be incorrect to associate the equatorward flux of wave activity, seen in Figs. 4e and 5e, for example, with wave propagation. In horizontally propagating waves, one would expect the nonadvective flux terms, particularly the second terms on the right-hand sides of (2.3a) and (2.3b), to be most important.
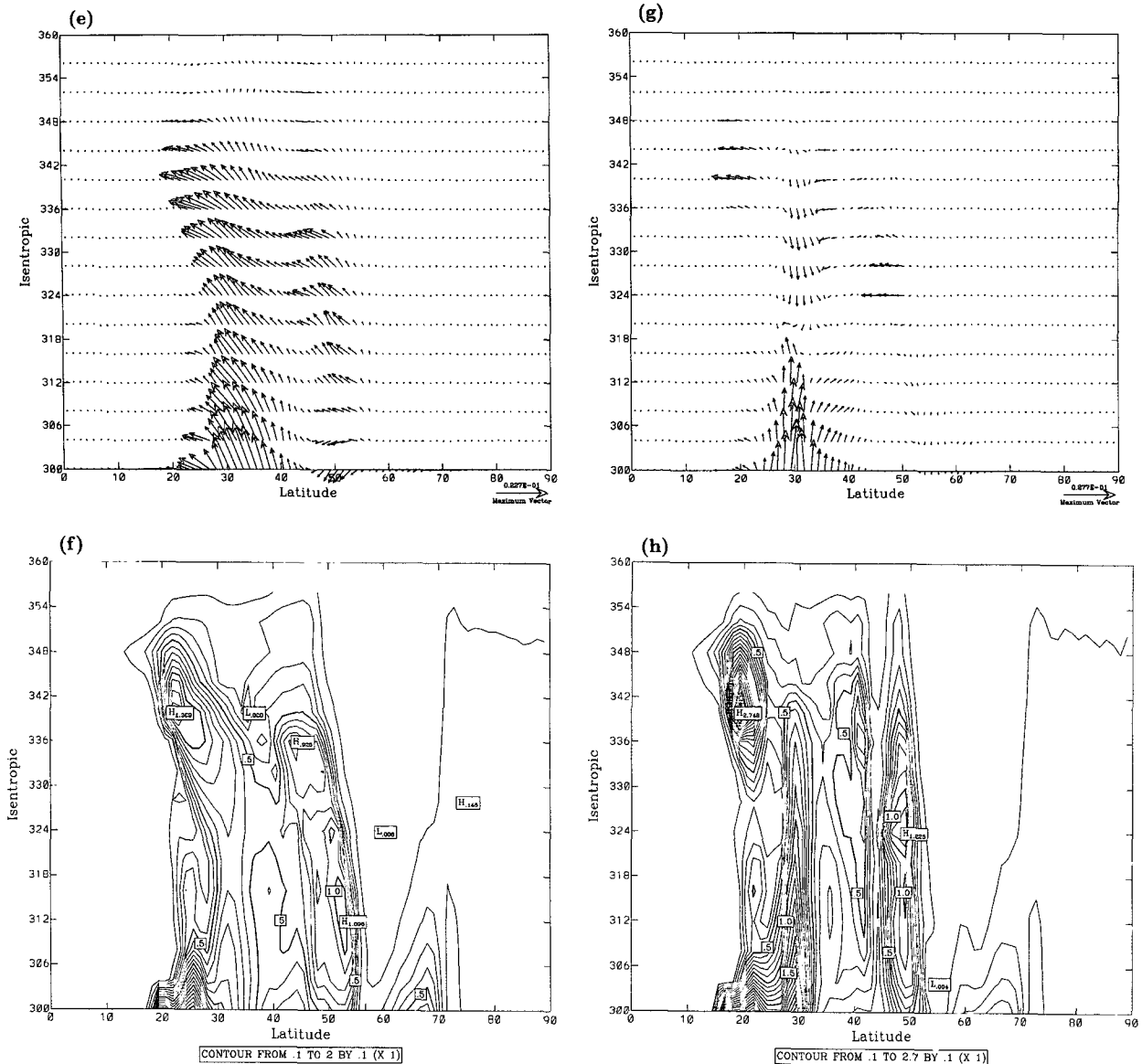
FIG. 5. (Continued) day 20 and the contour interval is 5.00×10^{-2}. (e) As in (a) except on day 22 and the maximum vector length is 2.27 $\times 10^{-2}$. (f) As in (b) except on day 22 and the contour interval is 0.100. (g) As in (a) except on day 23 and the maximum vector length is 2.77×10^{-2}. (h) As in (b) except on day 23 and the contour interval is 0.100.

While there is no advective contribution to the vertical flux, it is clear that vertical fluxes cannot always be associated with wave propagation. For example, the downward flux on day 23 has already been noted in Fig. 4g and is centered on about 30°N, as is a maximum in wave activity. Examination of Figs. 7c and 8c, and comparison with Fig. 4g, suggests that this flux is associated with the compact vortices that have formed from rolled-up PV streamers and which are also observed on isentropic surfaces lower than 336 K. The lower-level upward flux is associated with the formation of a low-level cyclone, identified as a "frontal cyclone" by Thorncroft and Hoskins (1990). Both these

vertical fluxes are likely to be associated with vertical interactions, within vortex structures in the first case, or between vortices and surface temperature gradients in the second. While such interactions are seen to give a net vertical transport of wave activity, it would seem difficult to argue that wave propagation is an appropriate description of the process.

b. Wave activity diagnostics for LC2

For the LC2 case we shall concentrate on results using Z2R as basic state. This is because most of the important flow evolution in LC2 takes place poleward
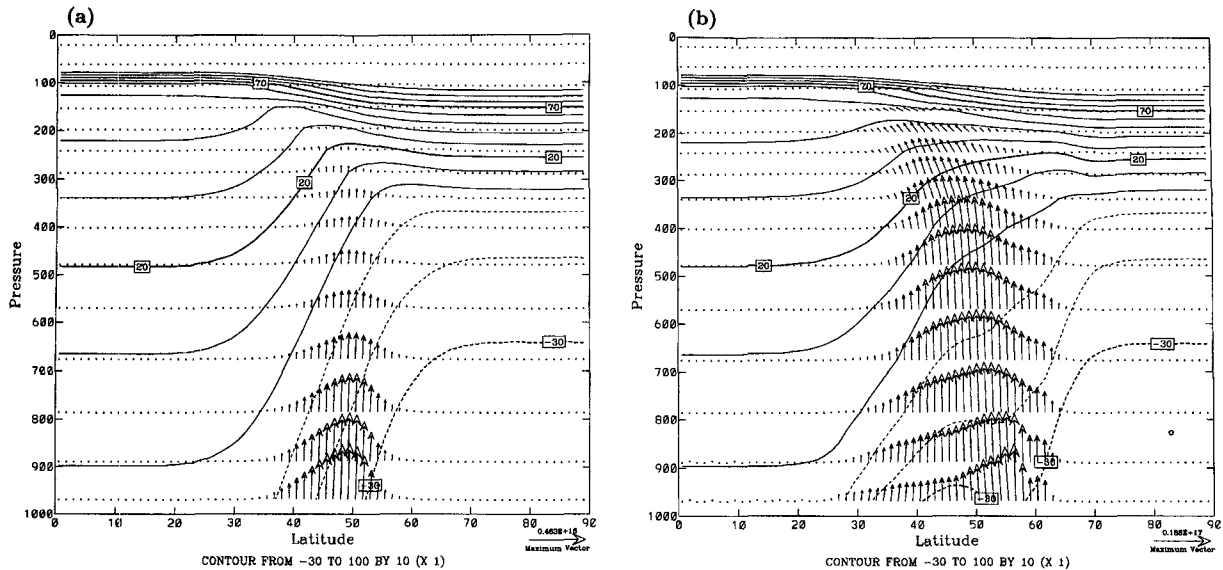
FIG. 6. Eliassen–Palm flux and potential temperature (contour interval 10 K; 300 K subtracted) for LC1 (a) on day 17, with maximum vector 4.63×10^{15} and (b) on day 20, with maximum vector 1.86×10^{16}.

of 45°N, where the conditions necessary for using Z2 as basic state are not satisfied. Figures 9a–j show the zonally averaged, meridional-plane (latitude and potential temperature) pseudomomentum flux and the zonally averaged pseudomomentum density referring to basic state Z2R for days 15, 20, 24, 25, and 26. We have marked the location of the tropopause as a bold contour on top of the diagram for pseudomomentum density in Fig. 9b. Day 15 is at a stage in the life cycle LC2 when the disturbance is still linear. However, at this stage the pseudomomentum density pattern is dominated by a contribution from the difference between initial and basic states and the contribution from the growing baroclinic wave is almost completely obscured. The flux, on the other hand, is dominated by the contribution from the growing baroclinic wave, and there are small differences from the corresponding stage of LC1 (Figs. 4a and 5a); for example, the fluxes are tilted more poleward in LC2. The EP fluxes are shown in Figs. 10a,b, for days 15 and 20, respectively. Note the slight poleward turning of the arrows in Fig. 10a as compared to the LC1 case of Fig. 6a. This is entirely consistent with THM's results.

In Figs. 9d, 9f, 9h, and 9j, the pseudomomentum density pictures, only contours below the dashed curve should be considered since at upper levels at low latitudes the signal remains dominated by the basic-state/ initial-state differences. Day 20 shows the upward propagating stage of the life cycle and the stage at which the pseudomomentum flux is at its maximum. The flux in Fig. 9c should be compared to the flux in Fig. 5c. A major difference is the substantial poleward branch of the wave activity flux in LC2. Similarly, in the wave activity density, A, it may be seen that wave activity is now filling up the poleward half of the tro-

posphere. There is still wave activity propagating along the tropopause but to a lesser extent than in LC1. Note that the EP flux in Fig. 10b has broadened laterally from day 15, but compared to Fig. 6b its horizontal component is mostly pointing poleward, and the vertical component drops off much faster with height than in Fig. 6b.

Even though some decay has set in by day 21, the life cycle is still quite active for several days. Figures 9e,f show pseudomomentum flux and density on day 24. Now, the maximum in pseudomomentum density is located at 68°N and at 305 K, compared to the late stages of LC1 when the maximum was at 20°N and 340 K. It is consistent with the notion of THM that the higher-latitude region has filled with wave activity and is therefore to some extent acting as a Rossby wave reflector. It is striking that now the flux is mostly directed downward from the upper-level nonlinear region and its maximum amplitude is about half of what it was on day 20. There is little equatorward (or poleward) flux. Indeed, comparing Figs. 9f and 9h, there appears to be some reduction in wave activity in the high-latitude region. On days 25 and 26 (Figs. 9g and 9i) the flux continues to show a complicated structure, mostly directed upward and poleward, its absolute value remaining mostly the same through time. The distribution of the corresponding pseudomomentum density (Figs. 9h and 9j) remains remarkably constant. All the above are indicative of the idea raised by THM that, whereas during the later stages of LC1 the upper subtropics acts as an absorptive cavity that drains wave activity out of the jet region, during the later stages of LC2 the wave activity tends to be trapped in a far less absorptive cavity. This is reflected in the quasi-steady coherent structure in LC2 below and slightly poleward of the jet. As
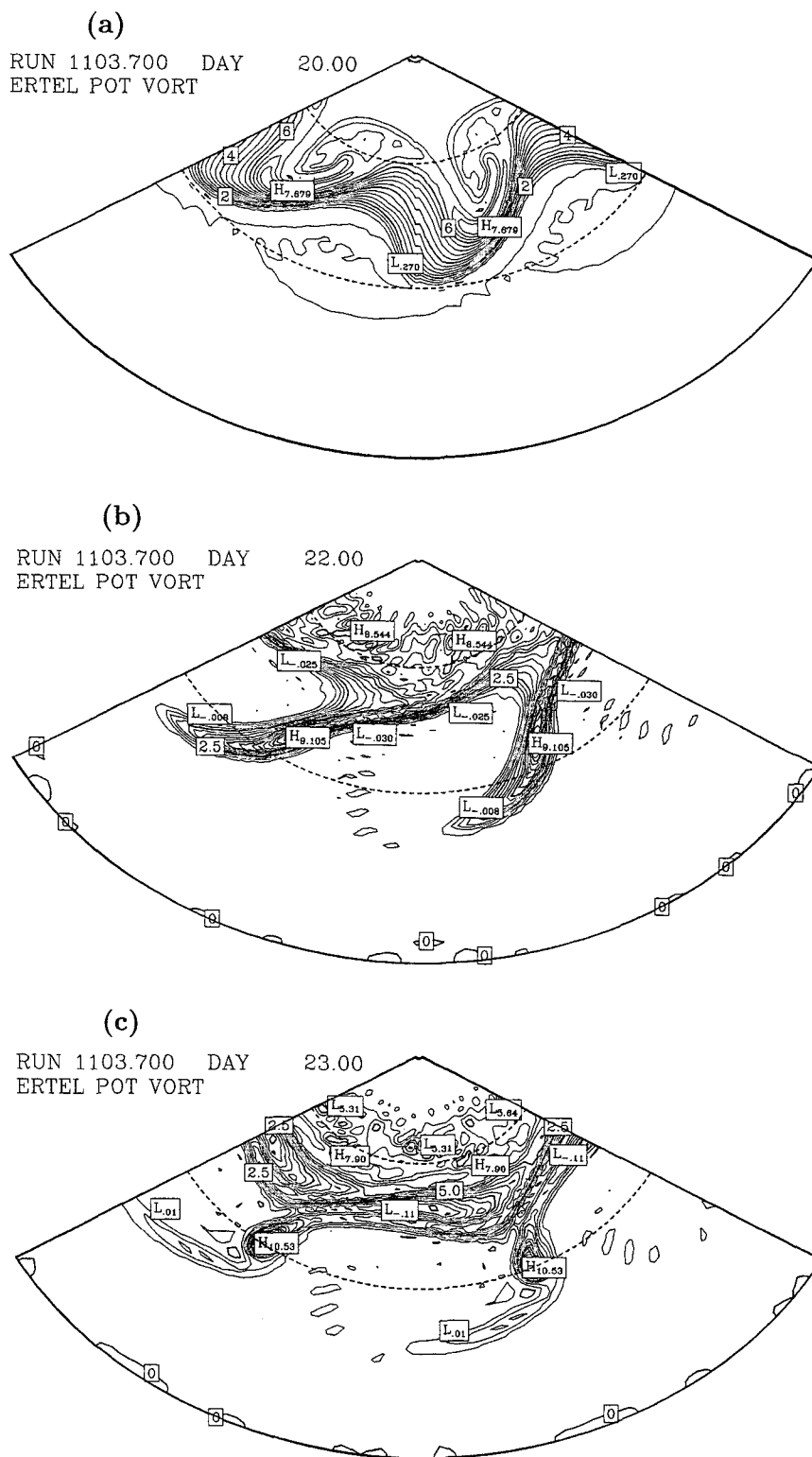
FIG. 7. Potential vorticity in PV units (10^{-6} m^2 s^{-1} K kg^{-1}) on the 336-K potential temperature surface for LC1 (a) on day 20, contour interval is 0.4, (b) on day 22, and (c) on day 23. Both (b) and (c) have contour interval of 0.5. Latitudes of 30° and 60°N are shown as dashed contours.

(a)

RUN 1103.700 DAY 20.00
pseudomomentum

(b)

RUN 1103.700 DAY 22.00
pseudomomentum

(c)

RUN 1103.700 DAY 23.00
pseudomomentum

FIG. 8. (a) Wave activity density on the 336-K surface for LC1 referring to basic state Z1 on day 20, contour interval is 0.1. (b) As in (a) except on day 22 and contour interval is 0.25. (c) As in (b) except on day 23. Latitudes 30° and 60°N are shown as dashed contours.

(a)

(c)

(b)

(d)

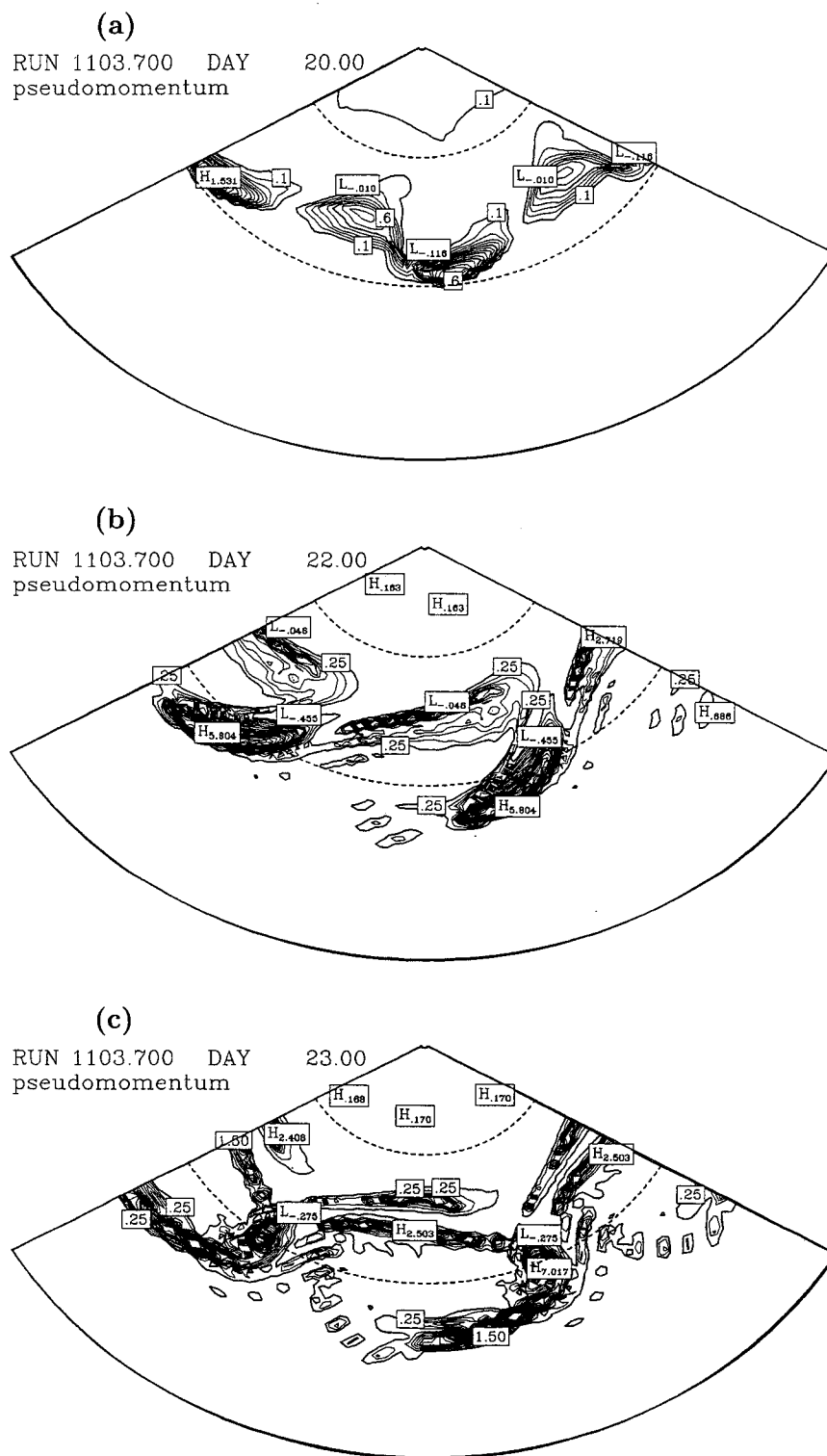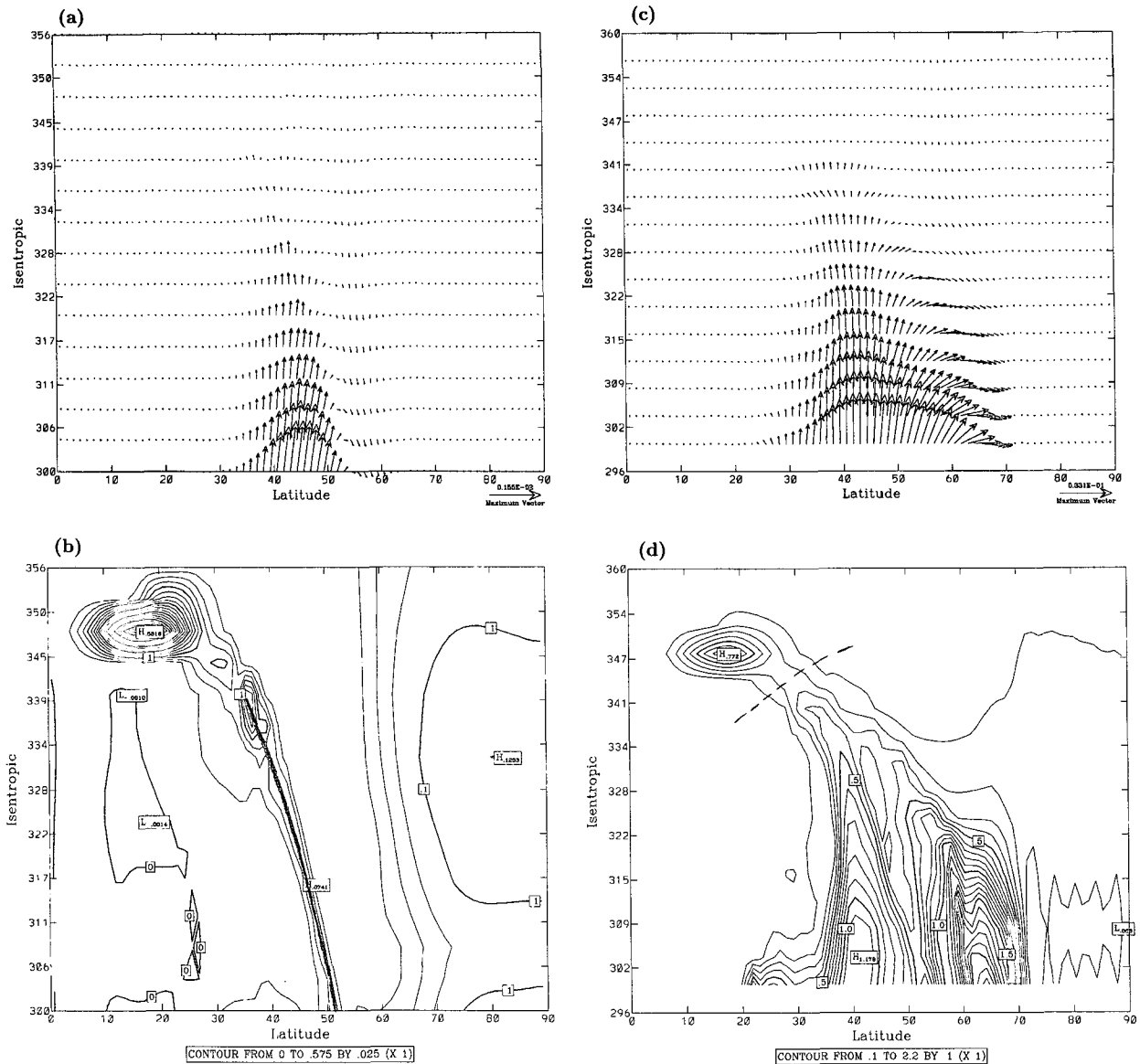FIG. 9. (a) Zonally averaged wave activity flux for LC2 referring to basic-state Z2R on day 15. The maximum vector (indicated in the lower-left corner) has magnitude of 1.55×10^{-3}. (b) Zonally averaged wave activity density for LC2 referring to basic-state Z2R on day 15. Contour interval is 2.50×10^{-2}. The heavy solid line indicates the location of the tropopause. (c) As in (a) except on day 20 and the maximum vector is of length 3.31×10^{-2}. (d) As in (b) except on day 20 and the contour interval is 0.10. The heavy dashed line separates the region where the wave activity arises from the disturbance (below) and the region where the wave activity arises from the basic state not being the initial state. (e) As in (a) except on day 24 and the maximum vector length is 1.51×10^{-2}. (f) As in (b) except on day 24 and the contour interval is 0.250. (g) As in (a) except on day 25 and the maximum vector is of length 1.05×10^{-2}. (h) As in (b) except on day 25 and the contour interval is 0.20. (i) As in (a) except on day 26 and the maximum vector length is 1.10×10^{-2}. (j) As in (h) except on day 26.

will be further discussed in section 6, there is little evidence that the wave activity associated with this structure is being maintained against dissipation by wave activity fluxes, at least not on the timescales considered here. The structure is of sufficiently large scale that the effects of small-scale dissipation are weak.

Since the maximum in pseudomomentum density is on a lower isentropic level in this case for LC2 than it was for LC1, we shall examine the fields at 312 K, on

a lower isentropic level than before. This isentropic level can be thought of as depicting upper-level structures for areas at or poleward of 60°N. Figures 11a–d show PV on the 312-K surface on days 20, 24, 25, and 26, respectively. The high-latitude vortices are strikingly persistent in time, and it is their contribution that tends to dominate the pseudomomentum. Pseudomomentum density is shown in Figs. 12a–d, for days 20, 24, 25, and 26. Comparing Figs. 11a and 12a, note that

(e)

(g)

(f)

CONTOUR FROM .25 TO 4.25 BY .25 (X 1)

(h)

CONTOUR FROM 0 TO 3.2 BY .2 (X 1)

FIG. 9. (*Continued*)

each vortex has a double structure in pseudomomentum and that this is manifested in the zonally averaged picture in Fig. 9d. The pseudomomentum density shown in Figs. 12a–d tends to follow the PV distribution. Again, this is an indication that the pseudomomentum at the mature stage in the life cycle is associated with advection of PV, corresponding to the first term in (2.3a) and (2.3b), and not to wave propagation, that is, the second term in (2.3a) and (2.3b).

5. Results for the modified LC1 and LC2 life cycles

Here we examine life cycles that were run in exactly the same way as LC1 and LC2 except that we used Z1R and Z2R as initial states instead of Z1 and Z2,

which are the initial states for the true LC1 and LC2. We shall call these modified life cycles LC1R and LC2R, respectively. Thus, the main difference between the initial states for LC1 and LC1R (and between LC2 and LC2R) is that the zonally symmetric PV has been rearranged for the latter—under constraints of PV and mass conservation—such that the latitudinal gradient of PV is not negative. We find that the resulting life cycles show the same type of behavior as the original ones, only somewhat less exaggerated. Thus there is no evidence that the interior reversals of PV on isentropic surfaces in the initial states of LC1 and LC2 were particularly important for unstable growth. The upper-level vortices of LC1R's mature phase turn out to be more vigorous than those at the same stage of LC1.
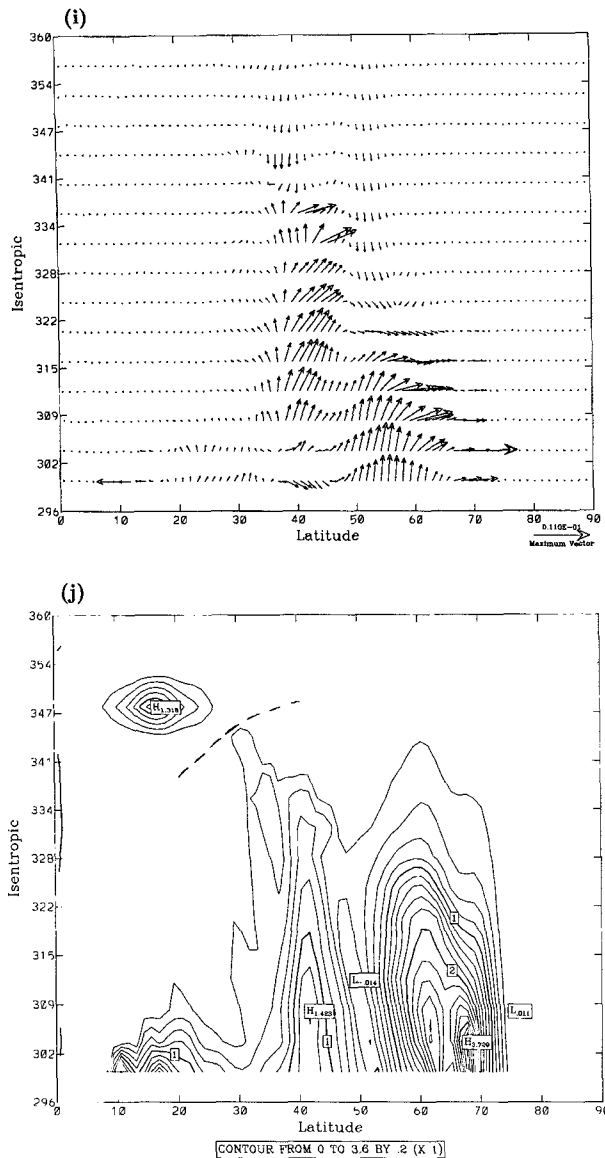
(i)

(j)

FIG. 9. (Continued)

LC2R turns out to be less active than LC2 in that the disturbance is both shallower and more confined to high latitudes than LC2.

a. Wave activity diagnostics for LC1R

We are particularly interested in examining differences between LC1R and LC1. We shall therefore concentrate on the same stages of development as we considered for LC1. In the case of LC1R it appears that these stages occur on days 17, 19, 21, and 22, that is, a day earlier than for LC1 everywhere except for the initial, linear stage. Examination of the zonally averaged pseudomomentum flux and density on those days corresponding to Figs. 4 and 5 shows structure in

LC1R's linear stage that is considerably noisier than LC1's linear stage. The PV rearrangement seems to introduce small-scale structure into the basic state around the tropopause level in midlatitudes on the lower and southern fringe of the jet maximum. Expressions of this structure can be seen in Fig. 3c, in the steepness of the 325 K isentrope just below the jet. The small-scale structure becomes less important as development takes place. As for LC1 the wave activity flux and density plots show that in the growing linear mode there is some propagation, albeit relatively weak, along the tropopause. Eliassen–Palm fluxes at this stage are very similar to those for LC1.

The upward propagating stage of LC1R on day 19 is very similar to the corresponding LC1 stage, shown in Figs. 4 and 5c,d. Similarly, the EP fluxes for LC1R are almost the same in magnitude and direction as in Fig. 6b. By day 21 we have reached the mature stage of LC1R. The flux, shown in Fig. 13a, is pointing more upward than either in Figs. 4e or 5e, depicting the corresponding mature-stage flux of LC1. Also, its magnitude is greater. Day 21.3 of LC1R and day 22 of LC1 are probably more exactly comparable. Note that LC1R shows the same equatorward excursion of wave activity as LC1. The following day the wave activity flux in Fig. 13c is mostly horizontal at upper levels, with a strong equatorward signal toward 20°N. Figure 13d shows the corresponding pseudomomentum density picture. An interesting difference here compared to the LC1 case in Figs. 4f and 5f is that LC1R seems to be active farther poleward than LC1 at this late stage. There is a sharp edge to the wave activity density picture in Fig. 5h at about 50°N, whereas this edge in Fig. 13d is at about 60°N. This farther poleward extent of LC1R can also be seen on the following day (not shown in Fig. 13) so that it does not appear to be associated with the possible time shift of evolution of LC1R compared to LC1.

Looking at PV on the 336-K isentropic surface, we again see the large lateral excursions of PV that characterized LC1. Figure 14a shows PV on day 19 and should be compared to Fig. 7a. This picture further confirms that indeed we are looking at the upward propagating phase of the life cycle since it shows that PV contours are undulating reversibly. By day 21 shown in Fig. 14b, we have tongues of high-PV air reaching into low latitudes. Comparing this figure with Fig. 7b shows that the high-PV tongues are broader and therefore take up more area in LC1R than at the corresponding time in LC1. This is further substantiated by the field in Fig. 14c that shows that the tongues have rolled up into intense vortices on day 22, more intense than at the corresponding time for LC1 shown in Fig. 7c. This difference between LC1 and LC1R is also reflected in the area-average eddy kinetic energy field as will be discussed in section 6.

The pseudomomentum density on 336 K corresponds clearly to features in the PV distribution, as can be seen in Figs. 15a–c, for days 19, 21, and 22. Comparing Figs.
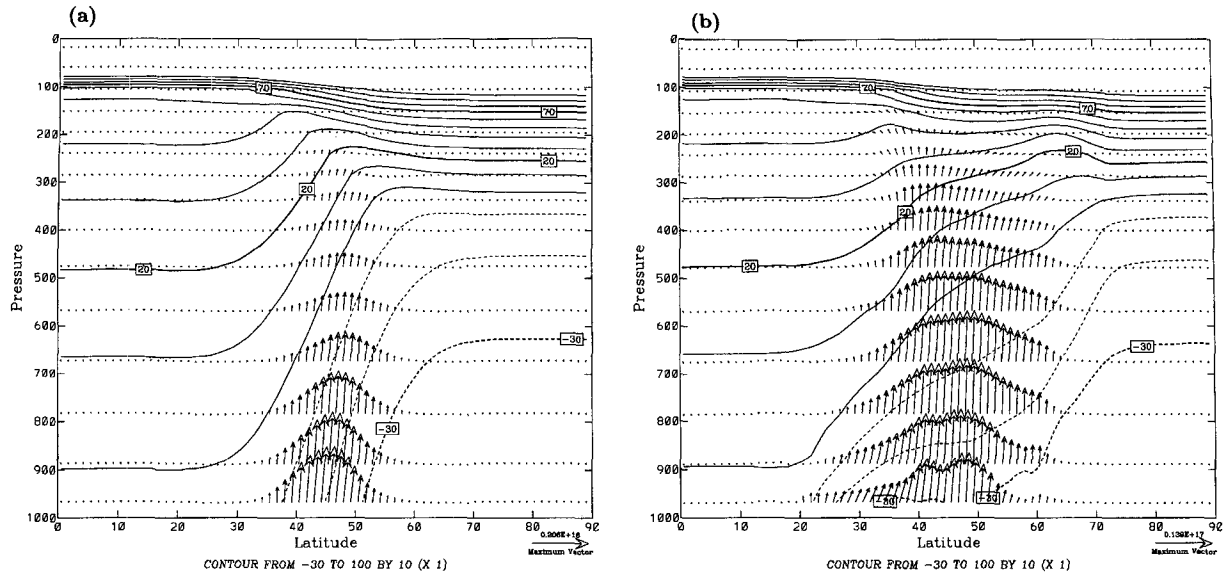
FIG. 10. EP flux and potential temperature (contour interval 10 K, with 300 K subtracted) for LC2 (a) on day 15 with maximum vector of 2.06×10^{15} and (b) on day 20 with maximum vector of 1.39×10^{16}.

15a and 8a, we see that the pseudomomentum distribution for LC1R at the upward propagating stage takes up less area and has a higher amplitude than that for LC1 at the corresponding time. The location of the maximum wave activity density corresponds to the location of the tropopause where there are concentrated PV gradients on each isentropic surface intersecting it. Again, there is reason to believe that the final stage of equatorward displacement of wave activity is associated with nonlinear advection of potential vorticity anomalies rather than with wave propagation.

b. Wave activity diagnostics for LC2R

The linear stage of LC2R's zonally averaged wave activity flux is very similar to the linear stage of LC2's zonally averaged wave activity flux shown in Fig. 9a. The corresponding zonally averaged wave activity density for LC2R does not have the anomalous maximum at low-latitude upper levels in LC2, which arises solely from the differences between the basic and initial states. However, for LC2R as for LC1R, there is some small-scale structure in the wave activity density of the linear stage concentrated at tropopause level. Eliassen–Palm fluxes at this stage are very similar to those for LC2 (Fig. 10a) except that the location of the maximum upward flux is shifted by about 5° latitude poleward.

The zonally averaged wave activity flux and density for the upward propagating stage are very similar to the corresponding stage for LC2, shown in Figs. 9c,d. The exception is that in the LC2R case the basic state is the initial state so that there is no anomalous maximum in the density at low-latitude upper levels, as can be seen in Fig. 16a. Also the wave activity in Fig. 16a is more

latitudinally confined than for the LC2 case in Fig. 9d. The EP fluxes for LC2R's upward propagating stage are very similar to those for the corresponding stage of LC2 shown in Fig. 10b.

At the mature stage, LC2R is again similar to LC2 in Figs. 9e,f. The zonally averaged wave-activity density is shown in Fig. 16b. Note that the magnitudes are considerably lower for LC2R than for LC2. Notice the upper-level, low-latitude bubble of wave activity that appears in Fig. 16b. This is indicative of some wave activity propagating through to low latitudes, an effect that is also detected for LC2 if Z2 is used as the basic state, which is possible in this low-latitude region. When LC2 is referred to basic-state Z2R, the signal in this region is dominated by basic-state/initial-state differences.

Figures 17a,b show PV on the 312-K isentropic surface for LC2R and should be compared to Figs. 11a,b, showing the same field for the same stages for LC2. Note that the rolled-up vortices are located farther poleward for LC2R than for LC2. Again, when later times were examined, the persistence of the vortices with time is striking. The pseudomomentum density on the same isentropic surface for the same times is shown in Figs. 18a,b. Compared to the LC2 case shown in Figs. 12a,b, the wave activity density is both latitudinally more confined and has lower values than for LC2. Again, the wave activity fields follow the PV distribution, explaining the double-peaked maxima in the cross sections.

6. Global diagnostics and comparisons

In order to improve in simple quantitative measure of the difference between the four life cycles, it is use-

(a)

RUN 1115.700   DAY   20.00
ERTEL POT VORT

(b)

24.00

RUN 1115.700   DAY
ERTEL POT VORT

(c)

RUN 1115.700   DAY   25.00
ERTEL POT VORT

(d)

26.00
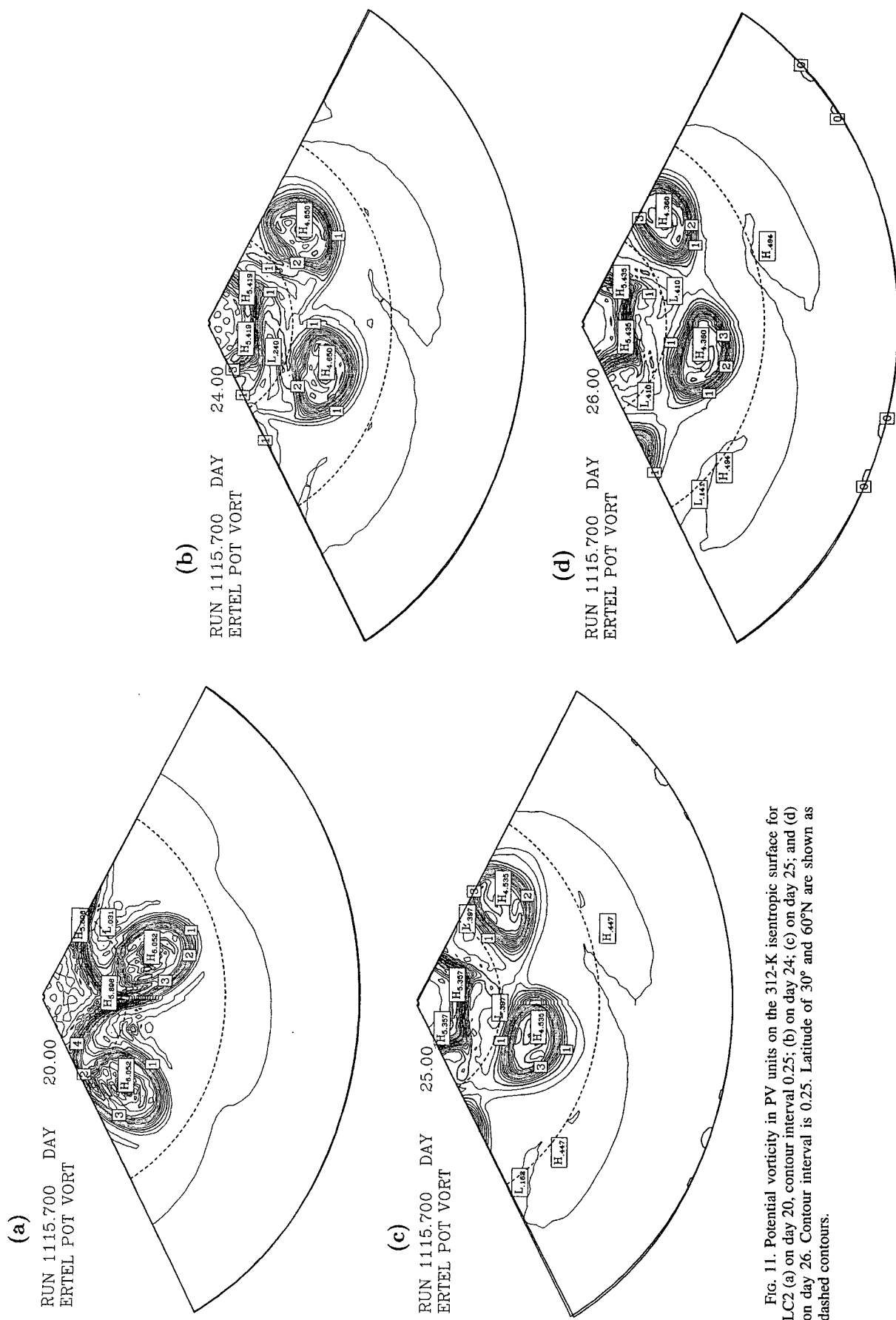
RUN 1115.700   DAY
ERTEL POT VORT

FIG. 11. Potential vorticity in PV units on the 312-K isentropic surface for LC2 (a) on day 20, contour interval 0.25; (b) on day 24; (c) on day 25; and (d) on day 26. Contour interval is 0.25. Latitude of 30° and 60°N are shown as dashed contours.
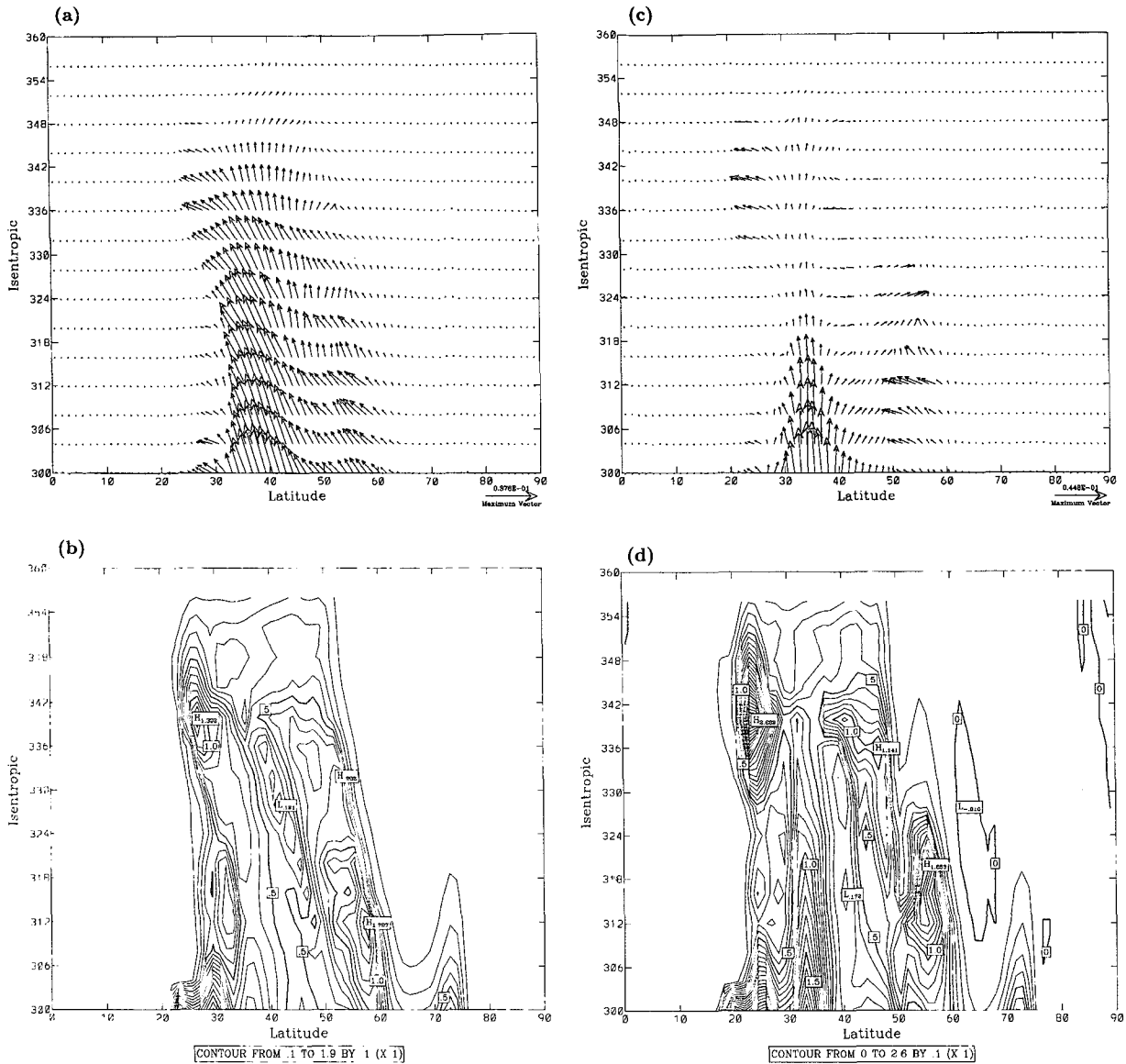
(a)

RUN 1115.700   DAY   20.00
pseudomomentum

(b)

RUN 1115.700   DAY   24.00
pseudomomentum

(c)

RUN 1115.700   DAY   25.00
pseudomomentum

(d)

RUN 1115.700   DAY   26.00
pseudomomentum

Fig. 12. (a) Wave activity density on the 312-K surface for LC2 referring to basic-state Z2R on day 20. Contour interval is 0.2. (b) Same as (a) except on day 24. (c) Same as (a) except on day 25. (d) Same as (a) except on day 26. Latitudes of 30° and 60°N are shown as dashed contours.

(a)

(c)

(b)

(d)

FIG. 13. (a) Zonally averaged wave activity flux for LC1R referring to basic-state Z1R on day 21. The maximum vector (indicated in the lower-left corner) has magnitude of 3.76×10^{-2}. (b) Zonally averaged wave activity density for LC1R referring to basic-state Z1R on day 21. Contour interval is 0.10. (c) Same as (a) except on day 22 and the maximum vector length is 4.48×10^{-2}. (d) Same as (b) except on day 22.

ful to consider global diagnostics. Figure 19a shows the time evolution of the area-averaged eddy kinetic energy (EKE) for the four life cycles. The growing phase of the life cycle is similar in all four cases in that the slope of the curve of rising EKE is about the same. There is a time shift that presumably arises because the initial disturbance, which is not itself a normal mode, projects differently onto the relevant normal mode in each case. Probably the most striking difference between the four cases is the rapid and substantial drop in EKE for LC1 early in the decay phase, which is not seen so strongly for the other three life cycles. LC1R, LC2, and LC2R all have similar values of EKE on day

28, at about 7×10^5 J m^{-2}. LC1 on the other hand has an EKE value of about 5×10^5 J m^{-2} at the same time. If the only diagnostic considered was EKE, one would be tempted to conclude that LC1R, LC2, and LC2R are all similar, with only LC1 showing different behavior. Of course, the wave activity diagnostics show this not to be the case. In particular the apparent similarity between the EKE signatures of LC1R and LC2R is contrasted by the wave activity signatures clearly showing them to be of LC1 and LC2 type, respectively.

Figure 19b shows the area-integrated vertical wave-activity flux through 304 K as a function of time. Here 304 K is an isentropic surface that remains above the
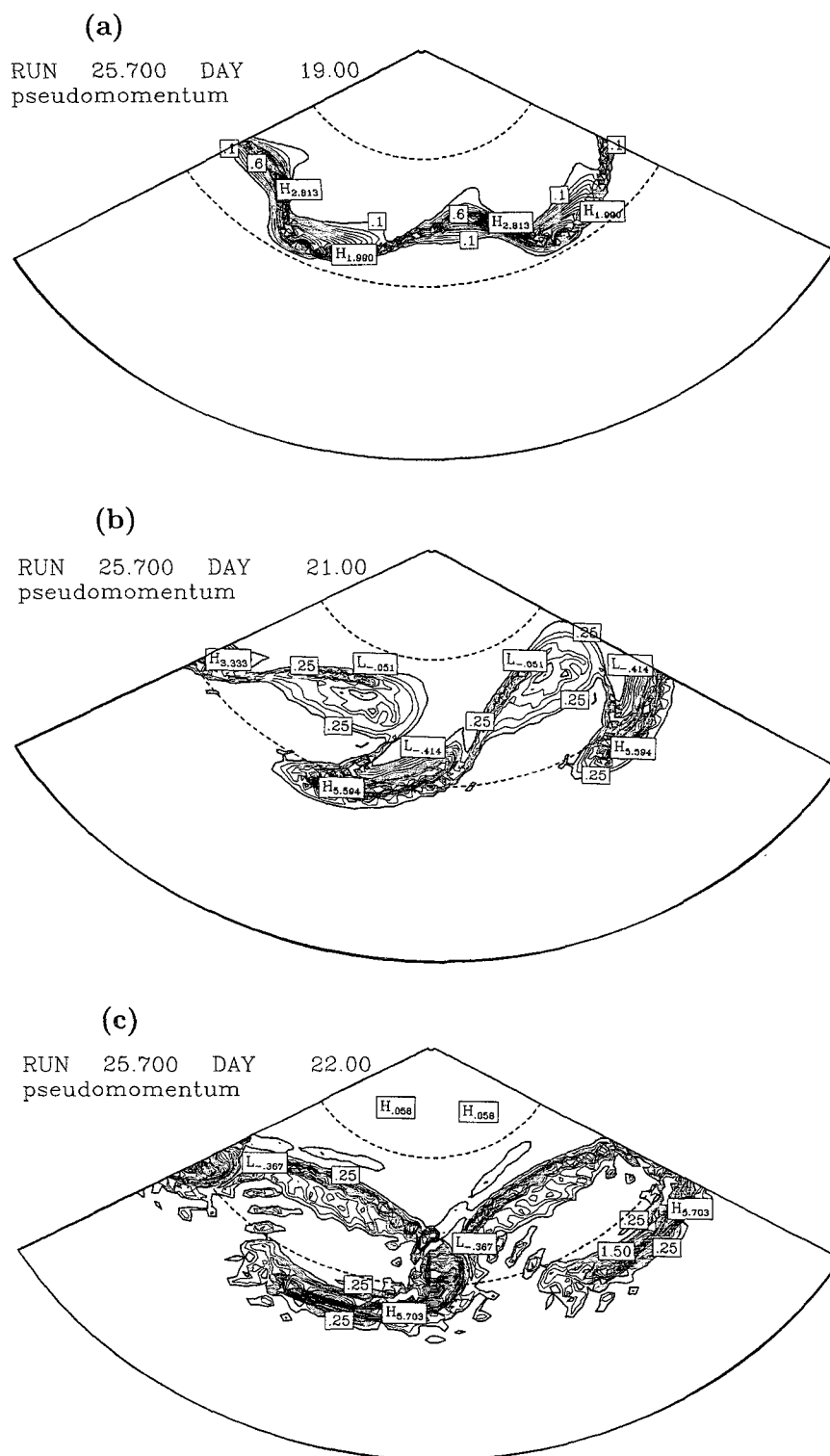
FIG. 14. Potential vorticity in PV units on the 336-K potential temperature surface for LC1R (a) on day 19, with a contour interval of 0.4 and (b) on day 21 and (c) on day 22, both with contour intervals of 0.5. Latitude of 30° and 60°N are shown as dashed contours.

(a)

RUN   25.700   DAY   19.00
pseudomomentum

(b)

RUN   25.700   DAY   21.00
pseudomomentum

(c)

RUN   25.700   DAY   22.00
pseudomomentum

FIG. 15. (a) Wave activity density on the 336-K surface for LC1R referring to basic-state Z1R on day 19. Contour interval is 0.10. (b) Same as (a) except on day 21 and contour interval is 0.25. (c) Same as (b) except on day 22. Latitudes of 30° and 60°N are shown as dashed contours.
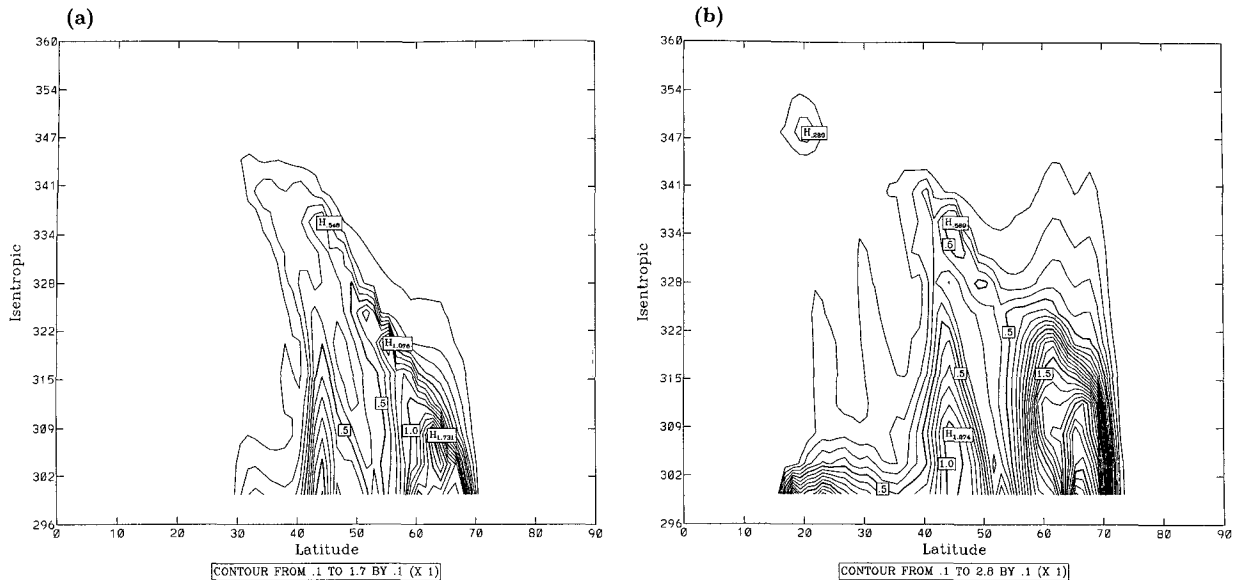
FIG. 16. (a) Zonally averaged wave activity density for LC2R referring to basic-state Z2R on day 18. Contour interval is 0.1.
(b) Same as (a) except on day 23.

earth's surface at all times for the four life cycles. Both LC2 and LC2R organize themselves earlier than LC1 and LC1R, as seen in Fig. 19b. Notice that the late stages of both LC2 and LC2R show fluctuations between positive and negative values of the total vertical flux with a period of about 3 days. At the end of the simulation, LC1R has a negative vertical wave activity flux that is about 30% in magnitude of the maximum vertical flux on day 20, whereas LC1's vertical flux is only slightly negative at the end of the simulation. This could be interpreted as a sign of the more effective absorption of wave activity in LC1 than in the other life cycles.

Figure 19c shows the cumulative (in time) vertical wave activity flux. LC2R has the lowest values for the mature phase consistent with it being the most latitudinally and vertically confined of all the life cycles, as was seen in Figs. 16–18. LC1R shows the largest cumulative value on day 26, after which it starts dropping off because of the negative vertical flux at late stages. Both LC2 and LC2R show a remarkably constant cumulative vertical flux after the initial increase up to day 22.

Figures 19b,c suggest that in later stages of LC2 and LC2R vertical fluxes are insignificant. Further calculations lead to similar conclusions regarding fluxes across control surfaces closer to the lower boundary. Therefore there seems to be no evidence that the coherent vortices seen in the later stages of LC2 and LC2R exchange wave activity to any great extent with the boundary, nor are they maintained against dissipation by a wave activity flux out of the boundary. This confirms the existence of a mid- and high-latitude Rossby wave cavity suggested by THM.

The vertical wave activity flux may be expressed in terms of Eulerian variables, whether or not the basic state has monotonic PV gradients. Furthermore, the vertical flux of wave activity is independent of zonally symmetric changes to the basic state. One could therefore use the initial states, Z1 and Z2 or equivalently Z1R and Z2R, as basic states for both LC1 and LC2 over the whole domain. However for computing wave activity density, only Z1R and Z2R are acceptable, as Z1 and Z2 both have negative PV gradients in the interior. Figure 19d shows the total integrated wave activity over the hemisphere and 304–356 K in the vertical. The vertical extent is the same as that shown in the cross sections, except that the lowest layer is omitted, since at certain stages of some of the life cycles the 300-K isentrope was found to dip underground. We adjusted for the fact that the basic states are not the initial states for LC1 and LC2 by subtracting the value on day 13 from the values on the following days. The nonadjusted values of wave activity on days 8–13 were less than 0.01% of the maximum values for both LC1 and LC2. If wave activity were exactly conserved, then one would expect the cumulative vertical flux through 304 K on a particular day (Fig. 19c) to agree with the total wave activity above that level on the same day (Fig. 19d). Comparing 19c,d, we see that in most cases the total wave activity is slightly less than the cumulative vertical flux, consistent with the fact that the simulation is not entirely dissipationless; a ∇^6 hyperdiffusion is included to limit the growth of small scales. Note in particular that this difference between the vertical flux and the wave activity is greater for the LC1-type life cycles, consistent with the more effective scale

(a)

RUN   34.700   DAY   18.00
ERTEL POT VORT   THETA=312.   PVU

(b)

RUN   34.700   DAY   23.00
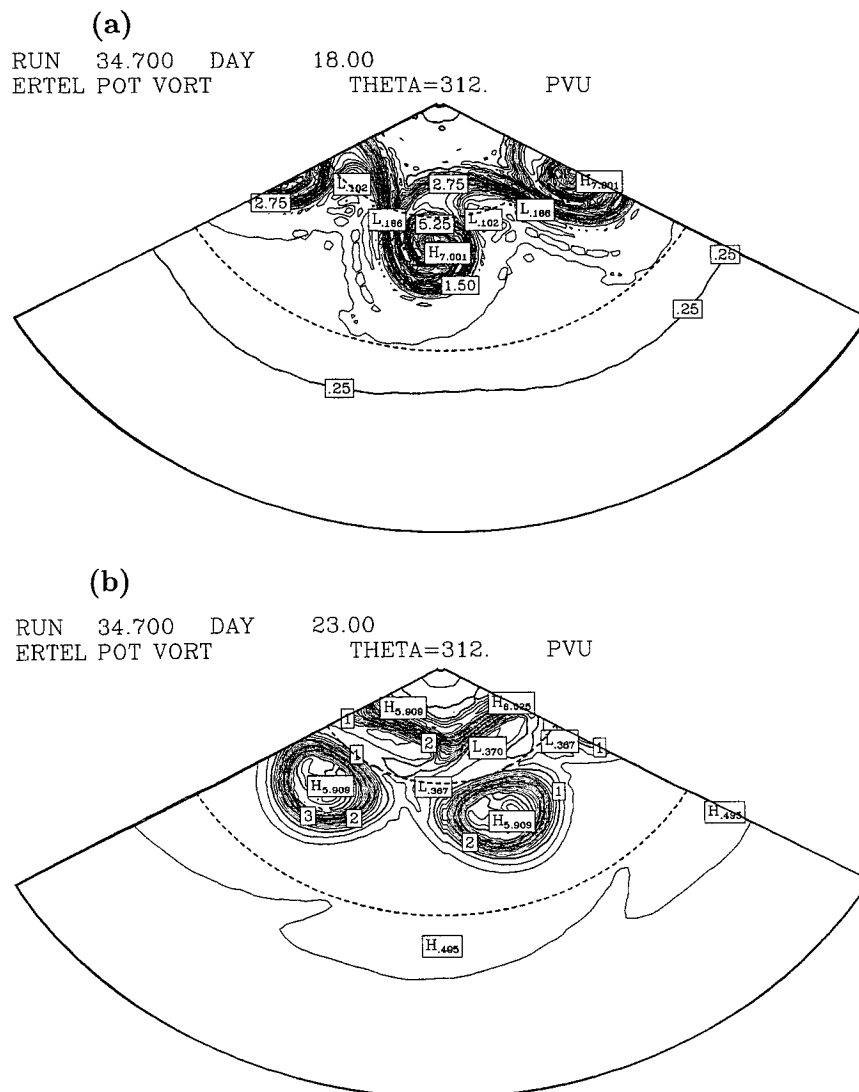ERTEL POT VORT   THETA=312.   PVU

FIG. 17. Potential vorticity on the 312-K isentropic surface for LC2R (a) on day 18 and (b) on day 23. Contour interval in both cases is 0.25 PV units. Latitudes 30° and 60°N are shown as dashed contours.

cascade and therefore more dissipation in those life cycles.

In Fig. 19e we confine the calculation of wave activity to the area equatorward of 45°N. For LC1 and LC2 we can thus use the initial states, Z1 and Z2, respectively, as basic states. This figure serves two purposes. First, it allows comparison of the amount of wave activity that has accumulated in lower latitudes for the LC1-type life cycles as compared to the LC2 types. Second, it serves to compare results for LC1/LC2 obtained using the initial state as basic state to results obtained using the rearranged state as basic state. The area equatorward of 45°N is the area where the LC1-type life cycles are particularly active in the decay phase at upper level. This is reflected in Fig. 19e. Comparing it to Fig. 19c, we see that 81% of the total wave activity of LC1 as measured in Fig. 19c is equatorward of 45°N on day 23 when the wave activity is at a maximum. When the wave activity of LC1R is at a maximum on day 22, 74% of the total wave activity is equatorward of 45°N. This is consistent with the more effective equatorward wave breaking in LC1 as compared to LC1R and also with the lack of a sharp decrease in EKE for LC1R. For the LC2-type life cycles the picture is completely different. When the wave activity of LC2 is at a maximum on day 23, only 45% of it is equatorward of 45°N. For LC2R the percentage is even smaller, 33% of the total wave activity on day 22 is equatorward of 45°N. According to Fig. 19e, using Z1R and Z2R as basic states for LC1 and LC2, respectively, tends to overestimate the wave activity by a relatively constant value.

(a)

RUN   34.700   DAY   18.00
pseudomomentum   THETA =   312.

(b)

RUN   34.700   DAY   23.00
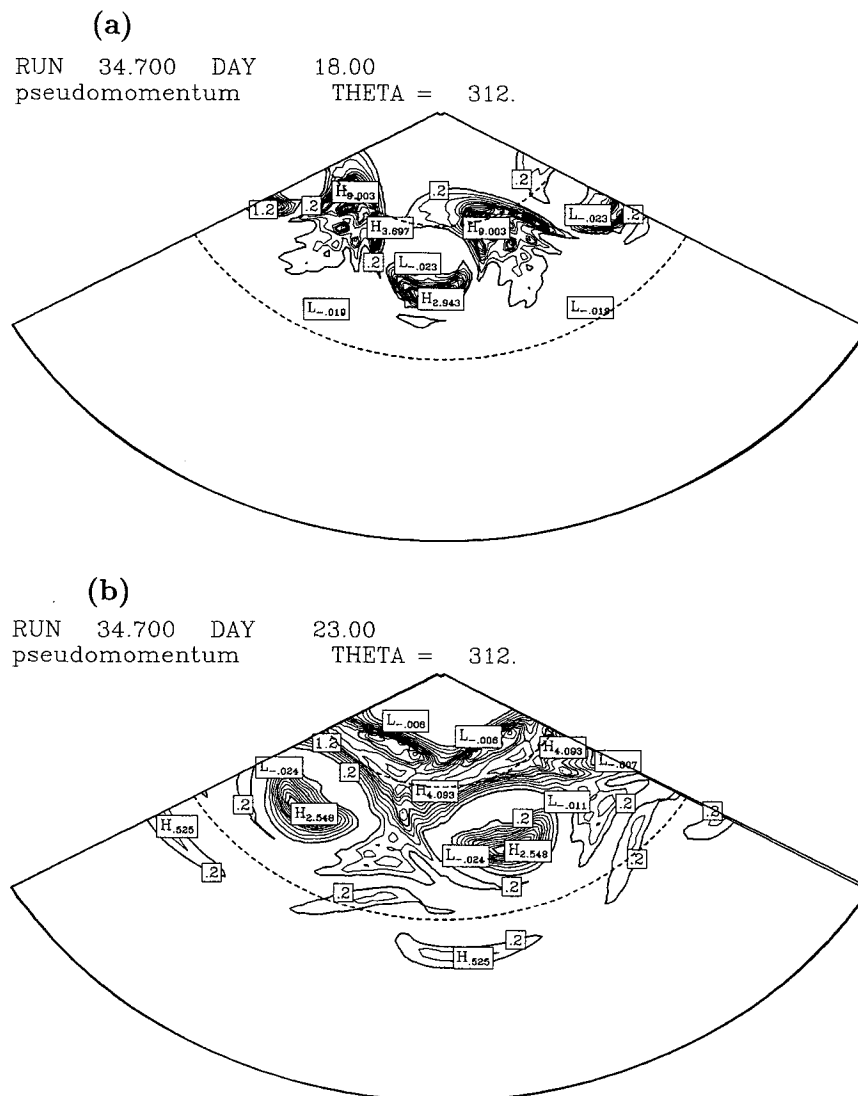pseudomomentum   THETA =   312.

FIG. 18. (a) Wave activity density on the 312-K surface for LC2R referring to basic-state Z2R on day 18. Contour interval is 0.2. (b) Same as (a) except on day 23. Latitudes of 30° and 60°N are shown as dashed contours.

Figure 19f shows the time evolution of that proportion of the total wave activity (equator to pole) that is equatorward of 45°N for LC1R and LC2R. In the normal-mode growth stage, there is not, by this measure, much difference between LC1R and LC2R. The striking difference is seen in the nonlinear stage. Interestingly, as both LC1R and LC2R approach the nonlinear stage, they show a decrease in the "low latitude" proportion of wave activity, with LC2R dropping off earlier than LC1R. (Note, e.g., from Fig. 19d that the non-linear stage is expected to be about two days earlier for LC2R than LC1R.) This corresponds to there being more rapid increase in wave activity poleward of 45°N than equatorward at about the time when there is low-level saturation in wave activity. For LC2R the equatorward proportion stays fairly constant at later times,

but LC1R shows a substantial increase to a secondary maximum at day 22 as the disturbance saturates at low latitudes. This equatorward shift in the wave activity is, of course, expected from the flux patterns shown in Figs. 13a and 13c. Figure 19f is consistent with both Figs. 19d,e and perhaps gives an even clearer picture of the difference between LC1- and LC2-type cyclones.

Finally, we consider global diagnostics including the lower boundary, wave activity terms discussed in section 2 and calculated using the approach described there. To show the typical θ-latitude configuration of the zonally averaged interior and boundary wave activities, each of the volume integral terms in (2.16) may be considered as the integral of a wave activity density over the whole θ-latitude plane. Such a density is identically zero at values of θ and latitude for which there are no points, at
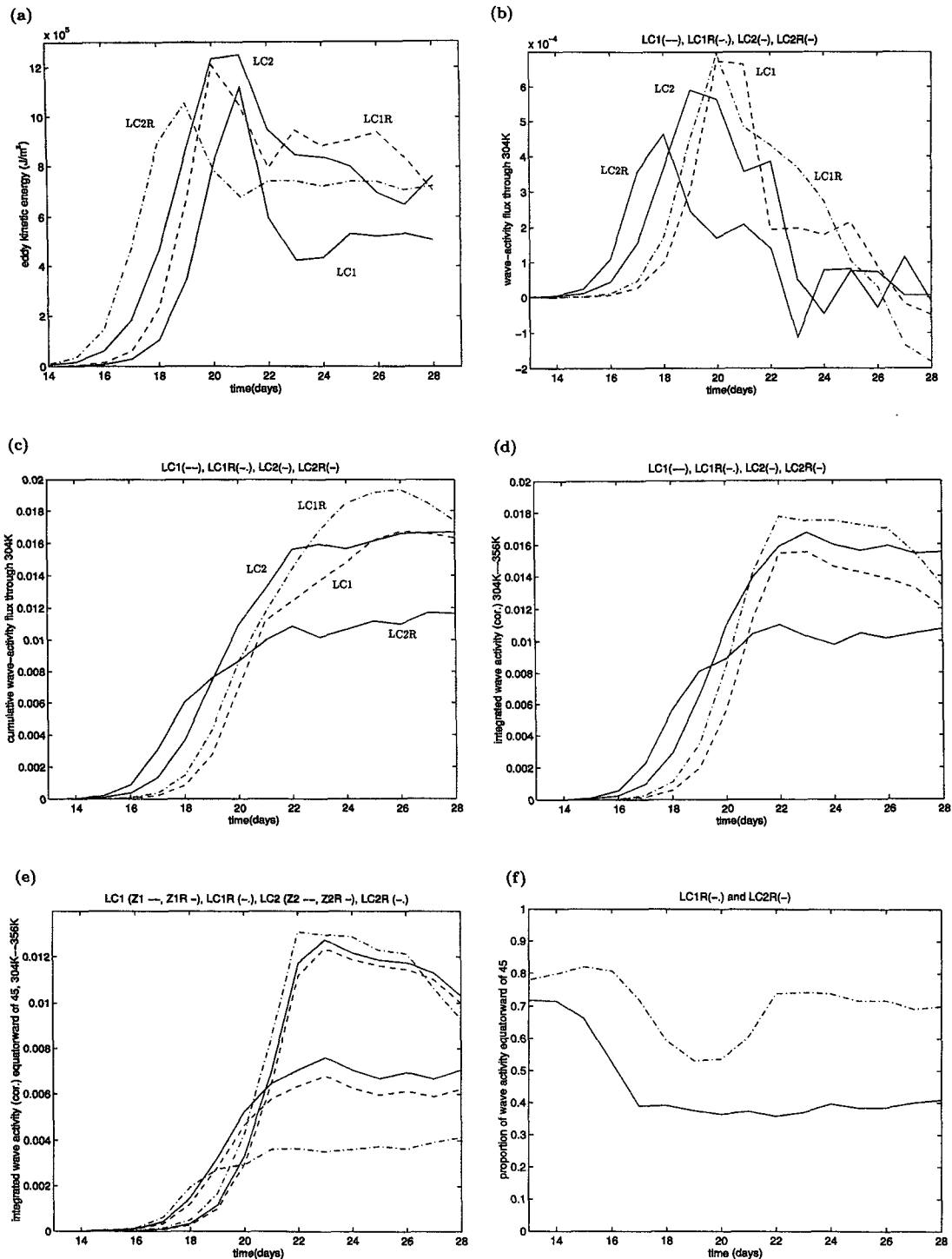
FIG. 19. (a) Area-averaged, eddy kinetic energy for the four life cycles as a function of time (in days). LC1 and LC2 are the full curves, LC1R the dashed curve, and LC2R the dashed-dotted curve. (b) Wave activity flux through the 304-K surface. LC2 and LC2R are the full curves, LC1 is dashed, and LC1R is dashed-dotted. (c) Cumulative (in time) vertical wave activity flux through 304 K as a function of time. The different life cycles are represented as before. (d) Total wave activity over the hemisphere and from 304 to 356 K. (e) Wave activity equatorward of 45°N and from 304 to 356 K. The different life cycles are represented as before. Additionally, LC1 referring to basic-state Z1 and LC2 referring to basic-state Z2 are represented by dashed curves. (f) Proportion of wave activity, integrated over levels 304–356 K, equatorward of 45°N. LC1R is indicated by the dashed–dotted curve, and LC2R is indicated by the full curve.

(a) interior and boundary wave-activity density for LC1R
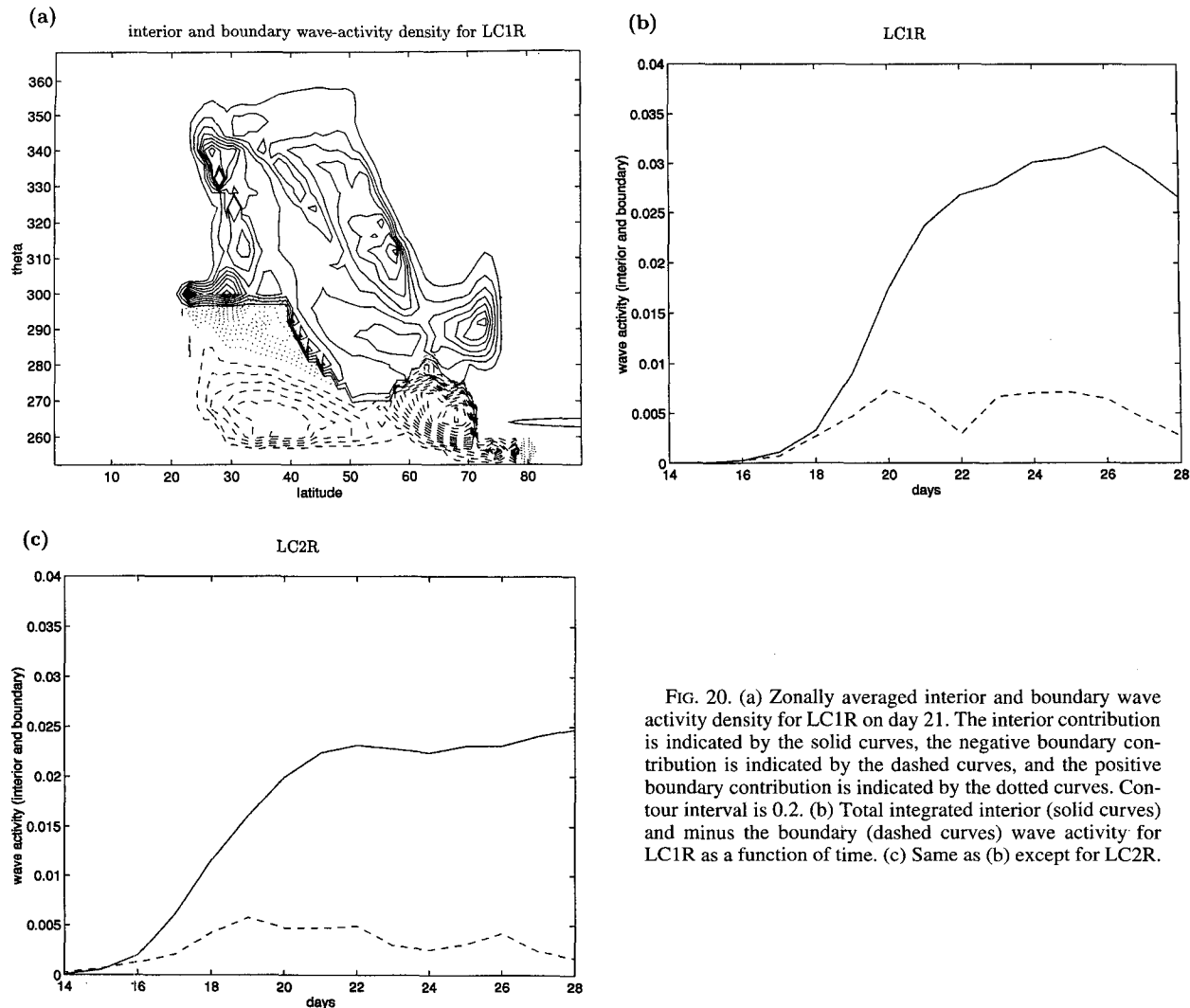
(b) LC1R

(c) LC2R

FIG. 20. (a) Zonally averaged interior and boundary wave activity density for LC1R on day 21. The interior contribution is indicated by the solid curves, the negative boundary contribution is indicated by the dashed curves, and the positive boundary contribution is indicated by the dotted curves. Contour interval is 0.2. (b) Total integrated interior (solid curves) and minus the boundary (dashed curves) wave activity for LC1R as a function of time. (c) Same as (b) except for LC2R.

any value of longitude, inside the region of integration concerned. Figure 20a is a θ-latitude cross section showing the interior wave activity density, that is, that giving rise to the first term in (2.16), and the boundary wave activity density, that is, that giving rise to the last three terms in (2.16). No account can be taken here of the boundary integral, the second term, because this is not a volume integral of a density. This term, however, has been included in the integrated boundary calculations (Figs. 20b,c), though in practice this term turns out to be only a small contribution to the total.

Figures 20b,c show the time evolution of the integrated interior and boundary wave activity for LC1R and LC2R, respectively. The integrated interior wave activity is the first term in (2.16) and minus the integrated boundary wave activity is minus the remaining terms in (2.16). The interior term is shown by the solid curve, and the boundary term by the dashed curve. In a conservative flow these two curves should match. We see here that for LC1R (Fig. 20b) they begin to diverge

substantially at about day 18 and for LC2R (Fig. 20c) at about day 16. Investigation of the low-level temperature field in the model simulation shows that at such times it is beginning to become substantially distorted. This is entirely consistent with the idea that frontogenetic processes are strongest at low levels. (Note, e.g., the surface temperature distribution on day 5 of LC1 shown in Fig. 5 of THM, or day 6 of LC2 in Fig. 8, both considerably preceding the maximum in EKE.) It is therefore to be expected that small-scale dissipative processes are playing an important role at that stage. In other words, in the upward propagation stage of both life cycles the wave activity flux out of low levels is being balanced at low levels not by further growth of (negative) wave activity, but by dissipation of (negative) wave activity.

7. Concluding remarks

In this work we have demonstrated how the differences between baroclinic wave life cycles may be

quantified using finite-amplitude, wave activity diagnostics. Such diagnostics are conceptually attractive because of their clear theoretical significance, but there are a number of aspects of their application that require care. In particular, the basic-state flow must satisfy PV monotonicity requirements. We have shown that even if the initial state does not satisfy such requirements, it may be possible to find other basic states that are sufficiently close to the initial state that the contribution to the wave activity from the growing disturbances dominates that from the basic-state/initial-state difference. Thus the wave activity diagnostics allow insight into the growth and saturation of the disturbances.

Potential vorticity rearrangement is an attractive method for defining new basic states with the required monotonicity properties, but, at least when used in conjunction with sigma coordinate models, it needs to be combined with a heuristic approach.

The nonmonotonicity of the THM basic states raised the question of the dynamical significance of such reversals in potential vorticity gradients. Numerical experiments with adjusted initial states show that the linearly growing modes clearly have the same baroclinic character whether or not the reversals are present and furthermore that the reversals are not an important part of the maintenance of the LC1/LC2 distinction. Nevertheless, there were important differences between the simulations in the original and adjusted states, particularly for LC1-type simulations, where the EKE signatures were very different. The EKE signature for the LC1R simulation was actually closer to those for the LC2 and LC2R simulations than to that for the LC1 simulation. Nonetheless, the upper-level potential vorticity structure and the latitudinal distribution of wave activity for LC1R were clearly of LC1 type. This is a reminder of the pitfalls in trying to characterize flows by a single type of diagnostic.

New insight was obtained into the late stages of evolution of all four life cycles. We found that during the late stages the propagation of wave activity to low latitudes is primarily associated with advection of wave activity and not wave propagation.

The sensitivity in the evolution to relatively small changes in the initial state also warns against too much emphasis on LC1 and LC2 as distinct types into which all nonlinear baroclinic life cycles must fall. Simulations not reported here seem to fall in the middle ground between LC1 and LC2. Instead, it seems best to regard LC1 and LC2 as useful paradigms against which different life cycle simulations may be measured.

APPENDIX A

Inverting the Balanced Zonal Flow

The equations for adiabatic, inviscid, zonally symmetric, balanced flow on the sphere can be written

$$\frac{Du}{Dt} - \left(2\Omega \sin\phi + \frac{u \tan\phi}{a}\right)v = 0, \quad (A.1)$$

$$\left(2\Omega \sin\phi + \frac{u \tan\phi}{a}\right)u + \frac{\partial M}{a\partial\phi} = 0, \quad (A.2)$$

$$\frac{\partial M}{\partial\theta} = \Pi, \quad (A.3)$$

$$\frac{D\sigma}{Dt} + \sigma \frac{\partial(v \cos\phi)}{a \cos\phi \ \partial\phi} = 0, \quad (A.4)$$

where θ, the potential temperature, is the vertical coordinate; $D/Dt = \partial/\partial t + v\partial/a\partial\phi$ the total derivative; u and v the zonal and meridional components of the wind; $\Pi = c_p(p/p_r)^\kappa$ the Exner function; $M = \theta\Pi + gz$ the Montgomery potential; and $\sigma = -\partial p/\partial\theta$ the pseudodensity.

The vorticity equation is derived by taking $-\partial \cos\phi/a \cos\phi\partial\phi$ of (A.1) to get

$$\frac{D\zeta}{Dt} + \zeta \frac{\partial(v \cos\phi)}{a \cos\phi \ \partial\phi} = 0, \quad (A.5)$$

where $\zeta = 2\Omega \sin\phi - \partial(u \cos\phi)/a \cos\phi\partial\phi$ is the absolute vorticity in isentropic coordinates. Eliminating divergence between (A.4) and (A.5), we obtain

$$\sigma \frac{DP}{Dt} = 0, \quad (A.6)$$

where P is the potential vorticity, $P = \zeta/\sigma$. Here P is the quantity we will want to invert to obtain the wind and temperature field. Using the hydrostatic equation to express σ in terms of M, we can write the definition of potential vorticity along with the equation for the gradient wind balance [(A.2)] as a coupled set of two partial differential equations:

$$2\Omega\mu - \frac{\partial(u(1 - \mu^2)^{1/2})}{a\partial\mu} + \frac{P}{\Gamma}\frac{\partial^2 M}{\partial\theta^2} = 0, \quad (A.7a)$$

$$\left(2\Omega\mu + \frac{u\mu}{a(1 - \mu^2)^{1/2}}\right)u + (1 - \mu^2)^{1/2}\frac{\partial M}{a\partial\mu} = 0, \quad (A.7b)$$

where $\mu = \sin\phi$ and $\Gamma = d\Pi/dp$. The two differential equations allow us to compute u and M given P along with the boundary conditions:

$$u = 0 \quad \text{at} \quad \mu = 1, \quad (A.7c)$$

$$u = u_{\text{eq}}(\theta) \quad \text{at} \quad \mu = 0, \quad (A.7d)$$

$$\frac{\partial M}{\partial \theta} = \Pi_T(\mu) \quad \text{at} \quad \theta = \theta_T, \quad (A.7e)$$

$$M - \theta\frac{\partial M}{\partial \theta} = 0 \quad \text{at} \quad \theta = \theta_B, \quad (A.7f)$$

where $u_{\text{eq}}(\theta)$ signifies the wind at the equator, which is specified, and θ_T and θ_B are the top and bottom isentropes, respectively. The pressure is specified at the top so that $\Pi_T(\mu)$ is known. At the bottom we assume that there is no topography such that $z = 0$.

Let us define a nondimensional vertical coordinate z, such that $z = H\ln(\theta/\theta_B)$, where $H = (\ln(\theta_T/\theta_B))^{-1}$, and let us nondimensionalize M by $c^2 = \alpha R(\theta_T - \theta_B)\lambda$, P by $2\Omega/\sigma_r$, Γ by $\Gamma_r = R\lambda/\bar{p}_B$, Π by $\Pi_0 = c_p\lambda$, p by $\bar{p}_B$, and u by $2\Omega a$. Then (A.7) becomes

$$\frac{\Gamma}{P}\left[\mu(1 - \mu^2)^{1/2} + u\mu - (1 - \mu^2)\frac{\partial u}{\partial \mu}\right]$$

$$+ (1 - \mu^2)^{1/2}\left[H^2\frac{\partial^2 M}{\partial z^2} - H\frac{\partial M}{\partial z}\right](\beta e^{z/H})^{-2} = 0,$$

$$(A.7a')$$

$$(\mu(1 - \mu^2)^{1/2} + u\mu)u + \frac{1}{\epsilon}(1 - \mu^2)\frac{\partial M}{\partial \mu} = 0,$$

$$(A.7b')$$

$$u = 0 \quad \text{at} \quad \mu = 1, \quad (A.7c')$$

$$u = u_{\text{eq}}(z) \quad \text{at} \quad \mu = 0, \quad (A.7d')$$

$$\kappa\alpha\frac{\partial M}{\partial z} = \frac{\beta}{H}e^{1/H}\Pi_T(\mu) \quad \text{at} \quad z = 1, \quad (A.7e')$$

$$M - H\frac{\partial M}{\partial z} = 0, \quad \text{at} \quad z = 0, \quad (A.7f')$$

where $\Gamma = \Pi^{(\kappa-1)/\kappa}$, $\alpha = (\bar{p}_B - \bar{p}_T)/\bar{p}_B$, $\lambda = (\bar{p}_B/p_r)^\kappa$, $p_r = 1000$ hPa, $R = 287$ J kg^{-1} K^{-1}, $\sigma_r = (\bar{p}_B - \bar{p}_T)/(\theta_T - \theta_B)$, $\epsilon = 4\Omega^2 a^2/c^2$, $\beta = \theta_B/(\theta_T - \theta_B)$, and u_{eq} has been scaled by $2\Omega a$. All of (A.7$'$) constitutes what has been called the invertibility principle (Hoskins et al. 1985). Given a potential vorticity distribution on the hemisphere along with wind at the equator and pressure at top, we can compute the wind and pressure on the hemisphere corresponding to that potential vorticity distribution. The way (A.7f$'$) is written one would think that we are restricted to constant potential temperature at the ground. This is not the case. Given the potential temperature at the ground, one simply extends the domain vertically to include parts that are below ground, the so-called

massless layer [see Andrews (1983) for a theoretical discussion and Fulton and Schubert (1991) for applications]. The bottom of the domain, θ_B, then corresponds to the highest isentrope that remains below the earth's surface everywhere. We assume that isentropes that intersect the ground continue just under the ground with pressure and geopotential equal to their surface values. Thus there is no mass trapped between isentropes that are underground, and $\sigma = 0$ in the massless layer making σ discontinuous across the ground. The Montgomery potential, M, remains continuous across the ground, varying linearly in θ in the massless layer. Thus the hydrostatic equation (A.3) holds in the massless layer ($\theta < \theta_{\text{sfc}}$). We define u for $\theta < \theta_{\text{sfc}}$ by (A.2) such that the positive root is taken. Similarly, we define v for $\theta < \theta_{\text{sfc}}$ such that (A.1) holds. Since the governing equations and definitions all apply unchanged in the massless layer, the derivation of the invertibility relation goes through in the same way as before and (A.7a$'$–f$'$) are valid for the extended region. Note that since $\sigma = 0$ for $\theta < \theta_{\text{sfc}}$ while ζ remains finite, $P \to \infty$ when $\theta < \theta_{\text{sfc}}$. Also note that in the massless layer, Γ is equal to its surface value. When solving (A.7$'$), we assume that $\Gamma/P = 0$ when $\theta < \theta_{\text{sfc}}$.

Numerical method for solving the invertibility relation

To solve the discretized version of (A.7a$'$–f$'$), we use an iterative method similar to the one used by Schubert et al. (1991) based on Newton's method and line relaxation.

Consider the grid defined by $\mu_j = j\Delta\mu$, where $j = 0$, 1, $\cdots$, J and $\Delta\mu = 1/J$, and by $z_k = k\Delta z$, where $k = -1, 0, 1, \cdots, K + 1$ and $\Delta z = 1/K$. Defining M at odd j points, u at even j points, and demanding that J always be even ensures that $j = 0$ and $j = J$ are u points where the boundary conditions (A.7c$'$–d$'$) are imposed. We define

$$x_{j,k} = \begin{cases} M_{j,k}, & \text{if } j \text{ odd} \\ u_{j,k}, & \text{if } j \text{ even}, \end{cases} \quad (A.8)$$

and use centered differences to discretize (A.7a$'$) at the M points and (A.7b$'$) at the u points. Then the discretized form of (A.7a$'$–b$'$) is

$$F_{j,k} = 0, \quad (A.9a)$$

where

$$F_{j,k} = 2\Delta\mu(1 - \mu_j^2)^{1/2}H^2(x_{j,k+1} - 2x_{j,k} + x_{j,k-1})$$

$$- \Delta\mu(1 - \mu_j^2)^{1/2}H\Delta z(x_{j,k+1} - x_{j,k-1})$$

$$+ \frac{\Gamma_{j,k}}{P_{j,k}}(\beta e^{z_k/H})^2(\Delta z)^2\{\mu_j\Delta\mu(x_{j+1,k} + x_{j-1,k})$$

$$- (1 - \mu_j^2)(x_{j+1,k} - x_{j-1,k}) + 2\mu_j\Delta\mu(1 - \mu_j^2)^{1/2}\},$$

for j odd,

$$F_{j,k} = 2\epsilon\mu_j\Delta\mu((1 - \mu_j^2)^{1/2} + x_{j,k})x_{j,k}$$

$$+ (1 - \mu_j^2)(x_{j+1,k} - x_{j-1,k}),$$

for j even. The boundary conditions are

$$x_{j,K+1} = x_{j,K-1} + \frac{2\Delta z}{\kappa\alpha}\frac{\beta}{H}e^{1/H}p_{T_j}^\kappa, \quad \text{for } j \text{ odd} \quad \text{(A.9b)}$$

$$x_{j,-1} = x_{j,1} - \frac{2\Delta z}{H}x_{j,0}, \quad \text{for } j \text{ odd} \quad \text{(A.9c)}$$

$$x_{0,k} = u_{eq_k}, \quad \text{(A.9d)}$$

$$x_{J,k} = 0. \quad \text{(A.9e)}$$

The interior equations are applied for $j = 1, \cdots, J - 1$ and $k = 0, \cdots, K$. We use ghost points to satisfy the boundary conditions at top and bottom. The boundary conditions are updated after a sweep of the entire grid. Also updated at that time is $\Gamma_{j,k} = \Pi_{j,k}^{(\kappa-1)/\kappa}$, where $\Pi_{j,k} = \kappa\alpha H(\beta e^{zk/H})^{-1}(M_{j,k+1} - M_{j,k-1})/(2\Delta z)$.

For the results of this paper, θ_T was chosen at 550 K, and θ_B was the same for both Z1 and Z2 or 253 K. We used $J = 80$ and $K = 122$. As an initial guess, we used a zero wind field and we integrated the horizontally averaged pressure (on each isentropic level) of Z1 and Z2 in the vertical to obtain an initial M field for each case, that is, for inverting the PV of Z1R and Z2R, respectively. The value of $\bar{p}_B$ is the horizontally averaged pressure on the lowest isentropic surface (θ_B) of Z1 or Z2, and $\bar{p}_T$ is the horizontally averaged pressure on θ_T of Z1 or Z2. Even at 550 K, the top of our inverter turned out to be below the highest sigma level in the model. To get a velocity on the highest sigma level, we simply extrapolated linearly from the value on θ_T, smoothing the highest values at the same time.

Suppose we have an estimate of x along a line of interior points $j = 1, \cdots, J - 1$ and along lines of surrounding points, above and below, but that this estimate does not satisfy (A.9a). Holding the values along the surrounding lines fixed, we update the entire line simultaneously so that its values satisfy (A.9a). Then we have $J - 1$ equations—the even-labeled ones are nonlinear—that can be written in vector form as $F(x) = 0$, where x is a vector consisting of the $J - 1$ values along the line (with k fixed). Then Newton's method can be written as

$$J(x^{new} - x^{old}) + F(x^{old}) = 0, \quad \text{(A.10)}$$

where J is the tridiagonal Jacobian matrix of the system. We use zebra relaxation to update x by (A.10). First, we simultaneously update the lines corresponding to k even, overwriting old values of the unknowns. Then, we simultaneously update the lines corresponding to k odd using the new values of the unknowns on neighboring lines. Having thus updated all interior points, we then update the ghost points by (A.9b,c).

APPENDIX B

Rearrangement of Zonally Symmetric Potential Vorticity

The aim is to monotonize the distribution of PV in latitude such that the largest value of PV is at the pole, with decreasing values as we go toward lower latitudes.

a. Rearranging PV by area

Consider the grid (j, k) defined in appendix A, equally spaced in $\sin\phi$ in the horizontal and $\log\theta$ in the vertical. Let us consider a particular isentropic surface, say $z_k = k\Delta z$. Now PV is defined on the odd j points, so we define area elements a_{jj} $(jj = 1, 2, \cdots, J/2)$ corresponding to those points starting from the equator; $a_1 = \sin\phi_2$, $a_2 = \sin\phi_4 - \sin\phi_6$, $a_3 = \sin\phi_6 - \sin\phi_4$ all the way to the point next to the pole where $a_{J/2} = 1 - \sin\phi_{J-2}$. Note that the total area of the hemisphere has been normalized to 1. We now use a sorting routine to reorder the P_{jj}'s such that the largest value is at $jj = J/2$, that is, at the point closest to the pole, etc., so that we arrive at a distribution of PV, P^*, that does not have a reversal in the latitudinal gradient of PV. Corresponding to shifting the position of P_{jj}, the position of a_{jj} is shifted accordingly. Hence we obtain a function $A(P^*)$ that represents the total area on that particular isentropic surface taken up by fluid elements that have PV $\geqslant P^*$. [See Butchart and Remsberg (1986) for a more detailed discussion.] Corresponding to each value of the function $A(P^*)$, there is a value of an equivalent latitude since the normalized area of the polar cap to latitude ϕ_{eq} is $A = 1 - \sin\phi_{eq}$. Thus we have a one-to-one correspondence of PV and latitude (ϕ_{eq}). The final step is then to interpolate this new PV distribution back to the original grid that is equally spaced in $\sin\phi$.

b. Rearranging PV by mass

Again, we fix the isentropic surface to k. In this case we consider elements of mass m_{jj} $(jj = 1, 2, \cdots, J/2)$ to be reordered with the P_{jj}s rather than elements of area a_{jj}. In a similar manner to the above, we build a function $M(P^*)$ that ideally should represent the normalized mass of the polar cap shell enclosed by PV $= P^*$ and lying between $k - 1/2$ and $k + 1/2$ in the vertical, where the mass is normalized by the total pole to equator mass. We would then know our reordered PV field (P^*) at the latitudes Φ_{jj}^* where

$$1 - \sin\Phi_{jj}^* = \frac{\sum_{i=1}^{jj} \sigma_i^{(0)} a_i^*}{\sum_{i=1}^{J/2} \sigma_i^{(0)} a_i},$$

and where a_i^* indicates area elements that have been reordered to correspond to the reordering of P_i^*. To get the PV field at our original grid, we would simply interpolate to the latitudes defined by

$$1 - \sin\tilde{\phi}_{jj}^{(n)} = \frac{\sum\limits_{i=1}^{jj} \sigma_i^{(n-1)} a_i}{\sum\limits_{i=1}^{J/2} \sigma_i^{(n-1)} a_i} .$$

However, we do not know the pseudodensity before the inversion so, as we iterate toward the correct pseudodensity field, progressing toward a solution, we also update the interpolation of the P^* field to the grid. We used the pseudodensity field corresponding to the original PV field as our first guess, $\sigma_i^{(0)}$. For the results of this paper we iterated five times on the σ field ($n = 5$).

APPENDIX C

Comparison with Linearized Wave Activity Conservation Relations

A linearized form of the Eliassen–Palm relation in isentropic coordinates is given by Andrews (1987) and may be shown to be equivalent to the small-amplitude limit of the finite-amplitude pseudomomentum conservation relation of H88 if basic-state quantities in the latter are replaced by zonally averaged quantities. The linearized form given by Andrews (1987) is

$$\frac{\partial \tilde{A}}{\partial t} + \frac{1}{a\cos\phi}\frac{\partial}{\partial\phi}(\cos\phi \hat{F}^{(\phi)}) + \frac{\partial \hat{F}^{(\theta)}}{\partial\theta} = 0, \quad (C.1)$$

where

$$\tilde{A} = a\cos\phi\left(\frac{1}{2}a\sigma_0^2\overline{P_e^2}/P_{0\phi} - \overline{\sigma_e u_e}\right) \quad (C.2a)$$

$$\hat{F}^{(\phi)} = -a\sigma_0\overline{u_e v_e}\cos\phi \quad (C.2b)$$

and

$$\hat{F}^{(\theta)} = g^{-1}\overline{p_e M_{e\lambda}}. \quad (C.2c)$$

The $\overline{(\cdot)}$ denotes a zonal average, but we retain the $(\cdot)_e$ and $(\cdot)_0$ notation for consistency with the rest of the paper. We consider a domain as in section 2, where $v = 0$ at lateral boundaries, for example at $\phi = \pm\pi/2$, and the fluid domain is unbounded above. The lower boundary on the other hand is defined by the curve $\theta = \theta_b(\lambda, \phi)$. Then integrating over the domain, and noting that changes in the extent of the fluid domain may be neglected at leading order, it follows that

$$\frac{d}{dt}\int_{\mathcal{D}_0} \tilde{A}a\cos\phi d\phi d\theta$$

$$= \int_{\partial\mathcal{D}_0} \hat{F}^{(\theta)}a\cos\phi d\phi - \hat{F}^{(\phi)}\cos\phi d\theta. \quad (C.3)$$

We now assume that the flux components $\hat{F}^{(\phi)}$ and $\hat{F}^{(\theta)}$ vanish on all but the bottom boundary. The aim is to write the integral over the bottom boundary as the rate of change of a quantity integrated over that boundary.

It is useful to define the geometric height on the lower boundary in the basic state as $z = z_{b0}(\phi)$. If the lower boundary distribution of θ in the basic state is given by $\theta_{b0}(\phi)$, then it must be the case that $z_0(\phi, \theta_{b0}(\phi)) = z_{b0}(\phi)$ and hence that $z_{0\phi} + \theta_{b0\phi}z_{0\theta} - z_{b0\phi} = 0$. It follows that

$$\theta_{b0\phi} = (z_{b0\phi} - z_{0\phi})/z_{0\theta}. \quad (C.4a)$$

Similarly, since $z(\lambda, \phi_b, \theta) = z_b(\phi_b, \theta)$, it follows that

$$z_e = \phi_{be}(z_{b0\phi} - z_{0\phi}). \quad (C.4b)$$

Transforming the right-hand side of (C.3) to an integral over ϕ using (C.4b) gives

$$\frac{d}{dt}\int_{\mathcal{D}_0} \tilde{A}a\cos\phi d\phi d\theta$$

$$= \int\left\{g^{-1}\overline{p_e M_{e\lambda}} + \sigma_0\overline{u_e v_e}\cos\phi \frac{(z_{b0\phi} - z_{0\phi})}{z_{0\theta}}\right\}$$

$$\times a\cos\phi d\phi. \quad (C.5)$$

Note further that the fluctuation in z at the lower boundary satisfies the linearized equation

$$z_{et} + \frac{u_0}{a\cos\phi}z_{e\lambda} + \frac{v_e}{a}(z_{0\phi} - z_{b0\phi}) = 0, \quad (C.6)$$

while the λ component of the linearized momentum equation is

$$u_{et} - \zeta_{a0}v_e + \frac{(u_0 u_e)_\lambda}{a\cos\phi} + \frac{M_{e\lambda}}{a\cos\phi} = 0, \quad (C.7)$$

where ζ_a is the absolute vorticity.

Substituting from (C.6) for v_e in the second term on the right-hand side of (C.5), and then combining with (C.7), gives

$$\frac{d}{dt}\int_{\mathcal{D}_0} \tilde{A}a\cos\phi d\phi d\theta = \int_{\partial\mathcal{D}_0}\left\{g^{-1}\overline{p_e M_{e\lambda}} + \frac{\sigma_0}{z_{0\theta}}a\cos\phi\right.$$

$$\left.\times\left[-\frac{\overline{M_{e\lambda}z_e}}{a\cos\phi} + (u_e z_e)_t + \frac{a\zeta_{a0}\left(\frac{1}{2}\overline{z_e^2}\right)_t}{z_{b0\phi} - z_{0\phi}}\right]\right\}a\cos\phi d\phi. \quad (C.8)$$

Finally, using the identity that $M_e = -p_e z_{0\theta}/\sigma_0 + gz_e$, it is possible to show that the first two terms in the integrand on the right-hand side cancel. Transforming again, using (C.4a), to θ as the independent variable for the integration, and using (C.4b) to eliminate z_e in favor of ϕ_{be}, it follows that

$$\frac{d}{dt}\int_{\mathcal{D}_0} \tilde{A}a\cos\phi d\phi d\theta + \frac{d}{dt}\int_{\partial\mathcal{D}_0} -\sigma_0 a^2\cos^2\phi_{b0}$$

$$\times\left[\overline{u_e\phi_{be}} + aP_0\sigma_0\left(\frac{1}{2}\overline{\phi_{be}^2}\right)\right]d\theta = 0. \quad (C.9)$$

The second integrand may therefore be interpreted as a boundary contribution to the wave activity, in agreement with (2.15).

REFERENCES

Andrews, D. G., 1983: A finite-amplitude Eliassen–Palm theorem in isentropic coordinates. *J. Atmos. Sci.*, **40**, 1877–1883.

——, 1987: On the interpretation of the Eliassen–Palm flux divergence. *Quart. J. Roy. Meteor. Soc.*, **113**, 323–338.

——, and M. E. McIntyre, 1978: An exact theory of nonlinear waves on a Lagrangian-mean flow. *J. Fluid Mech.*, **89**, 609–646.

——, J. R. Holton, and C. B. Leovy, 1987: *Middle Atmospheric Dynamics*. International Geophysical Series, Vol. 40, Academic Press, 489 pp.

Brunet, G., and P. H. Haynes, 1996: Low-latitude reflection of Rossby wave trains. *J. Atmos. Sci.*, **53**, 482–496.

Butchart, N., and E. E. Remsberg, 1986: The area of the stratospheric polar vortex as a diagnostic for tracer transport on an isentropic surface. *J. Atmos. Sci.*, **43**, 1319–1339.

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

Fulton, S. R., and W. H. Schubert, 1991: Surface frontogenesis in isentropic coordinates. *J. Atmos. Sci.*, **48**, 2534–2541.

Haynes, P. H., 1988: Forced, dissipative generalizations of finite-amplitude wave-activity conservation relations for zonal and nonzonal basic flows. *J. Atmos. Sci.*, **45**, 2352–2362.

——, and M. E. McIntyre, 1990: On the conservation and impermeability theorems for potential vorticity. *J. Atmos. Sci.*, **47**, 2021–2031.

Held, I. M., and B. J. Hoskins, 1985: Large-scale eddies and the general circulation of the troposphere. *Advances in Geophysics*, Vol. 28A, Academic Press, 3–31.

Hoskins, B. J., and A. J. Simmons, 1975: A multi-layer spectral model and the semi-implicit method. *Quart. J. Roy. Meteor. Soc.*, **101**, 637–655.

——, M. E. McIntyre, and A. W. Robertson, 1985: On the use and significance of isentropic potential vorticity maps. *Quart. J. Roy. Meteor. Soc.*, **111**, 877–946.

Kushner, P. J., and T. G. Shepherd, 1995a: Wave-activity conservation laws and stability theorems for semi-geostrophic dynamics. Part I: Pseudomomentum-based theory. *J. Fluid Mech.*, **290**, 67–104.

——, and ——, 1995b: Wave-activity conservation laws and stability theorems for semi-geostrophic dynamics. Part II: Pseudoenergy-based theory. *J. Fluid Mech.*, **290**, 105–129.

Magnusdottir, G., and W. H. Schubert, 1991: Semigeostrophic theory on the hemisphere. *J. Atmos. Sci.*, **48**, 1449–1456.

McIntyre, M. E., 1982: How well do we understand the dynamics of stratospheric warmings? *J. Meteor. Soc. Japan*, **60**, 37–65.

——, and T. G. Shepherd, 1987: An exact local conservation theorem for finite amplitude disturbances to non-parallel shear flows, with remarks on Hamiltonian structure and on Arnol'd's stability theorems. *J. Fluid Mech.*, **181**, 527–565.

——, and W. A. Norton, 1990: Dissipative wave-mean interactions and the transport of vorticity or potential vorticity. *J. Fluid Mech.*, **212**, 403–435; Corrigendum, **220**, 693.

Schubert, W. H., P. E. Ciesielski, D. E. Stevens, and H.-C. Kuo, 1991: Potential vorticity modeling of the ITCZ and the Hadley circulation. *J. Atmos. Sci.*, **48**, 1493–1509.

Scinocca, J. F., and T. G. Shepherd, 1992: Nonlinear wave-activity conservation laws and Hamiltonian structure for the two-dimensional anelastic equations. *J. Atmos. Sci.*, **49**, 5–27.

——, and W. R. Peltier, 1994: Finite-amplitude wave-activity diagnostics for Long's stationary solution. *J. Atmos. Sci.*, **51**, 613–622.

Shepherd, T. G., 1989: Nonlinear equilibration of baroclinic instability. Part II: Continuously stratified fluid. *J. Atmos. Sci.*, **46**, 888–907.

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

Thorncroft, C. D., and B. J. Hoskins, 1990: Frontal cyclogenesis. *J. Atmos. Sci.*, **47**, 2317–2336.

——, ——, and M. E. McIntyre, 1993: Two paradigms of baroclinic-wave life-cycle behaviour. *Quart. J. Roy. Meteor. Soc.*, **119**, 17–55.