Eulerian Mean, Contour Integral, and Finite-Amplitude Wave Activity Diagnostics Applied to a Single-Layer Model of the Winter Stratosphere

JOHN THUBURN

Centre for Global Atmospheric Modelling, Department of Meteorology, University of Reading, Reading, United Kingdom

VINCENT LAGNEAU*

Ecole Polytechnique, Palaiseau, France

(Manuscript received 1 December 1997, in final form 13 March 1998)

ABSTRACT

A shallow water model is used to simulate a case of planetary wave breaking in the lower winter stratosphere. The simulation is diagnosed in terms of zonal mean mass and zonal momentum budgets, and also in terms of potential vorticity (PV) contour mass and circulation budgets. The time evolution of the PV contour diagnostics depends only on nonconservative processes such as diabatic heating, friction, and irreversible small-scale mixing; transient but essentially reversible events such as a temporary displacement of the vortex from the pole are effectively filtered out. The PV contour diagnostics show unambiguously and quantitatively aspects of the evolution such as shrinking of the vortex and sharpening of the vortex edge. The cross-contour mass flux gives a radically different view of meridional transport from that given by the mass-weighted Eulerian mean poleward velocity, both in terms of its qualitative behavior and in terms of the physical mechanisms that cause it. The PV contour diagnostics can be used to define a balanced, zonally symmetric state, whose evolution can be compared directly with that of the Eulerian zonal mean state. A new expression is presented for finite-amplitude wave activity in terms of the PV contour diagnostics. Wave activity diagnostics for the wave-breaking simulation are shown. There are large differences between the zonal mean wave activity flux and its small-amplitude approximation, the Eliassen–Palm (EP) flux; some of the implications for interpreting EP flux diagnostics in the stratosphere are discussed.

1. Introduction

A global view of the middle-atmosphere circulation has traditionally been represented by Eulerian zonal means of temperature, zonal wind, and chemical tracers, and terms in their budgets, including the mean meridional circulation. A more Lagrangian view of the mean meridional circulation and a clearer picture of wave driving of the mean meridional circulation are obtained by using transformed Eulerian mean (TEM) or isentropic coordinate mean diagnostics rather than straightforward pressure coordinate Eulerian means (e.g., Andrews and McIntyre 1976; Dunkerton 1978; Andrews 1983; Andrews et al. 1987). However, there are at least three ways in which all of these Eulerian zonal mean diagnostics fail to capture interesting aspects of the behavior of the middle atmosphere, particularly in the winter stratosphere.

First, features such as the edge of the winter polar vortex, characterized by sharp gradients of potential vorticity (PV) and long-lived chemical tracers, can be smeared out by Eulerian zonal averaging when the vortex is displaced from the pole or distorted from zonal symmetry. Sharp gradients of PV and tracers that are sometimes observed in the subtropics can be smeared out in a similar way.

Second, Eulerian zonal means cannot give a completely Lagrangian view of transport. For example, the air mass within the winter polar vortex is kinematically isolated from lower-latitude air for long periods, and this has important implications for ozone chemistry. However, this isolation of the vortex air is obscured by Eulerian zonal mean diagnostics. Furthermore, long-term, irreversible transport in the middle atmosphere is associated with nonconservative processes like diabatic heating, friction, and irreversible PV mixing. This association is captured only indirectly and approximately by Eulerian mean diagnostics, for example, through the approximate long-term balance between dissipation of

* Current affiliation: CEA, Laboratoire de Geosciences, Experimentation et Modelisation, St Paul lez Durance, France.

Corresponding author address: Dr. John Thuburn, CGAM, Department of Meteorology, University of Reading, 2 Earley Gate, Whiteknights, Reading RG6 2AU United Kingdom.
E-mail: swsthubn@met.rdg.ac.uk

eddies and the divergence of the Eliassen–Palm flux (EP flux) $\nabla \cdot \mathbf{F}$ in the generalized Eliassen–Palm theorem, and the eddy driving of the mean circulation expressed by $\nabla \cdot \mathbf{F}$ in the zonal momentum budget (e.g., Andrews and McIntyre 1976; Andrews et al. 1987).

Third, long-term, irreversible changes to the middle-atmosphere circulation, such as the distribution of angular momentum, are associated with nonconservative processes like diabatic heating, friction, and irreversible PV mixing. As in the case of the meridional transport, this association is captured only indirectly and approximately by Eulerian mean diagnostics. Eulerian mean diagnostics emphasize short-term, transient changes to the circulation, particularly at high latitudes. For example, a transient displacement of the vortex from the pole would appear as a deceleration of the zonal mean zonal wind $\overline{u}$, a warming of the zonal mean high-latitude temperature $\overline{T}$, and an increase in wave activity. Any long-term, irreversible changes to the circulation associated with such an event would be masked by the transient changes in $\overline{u}$ and $\overline{T}$, which tend to be of much larger amplitude.

To overcome these limitations of Eulerian mean diagnostics, several authors have proposed calculating mean diagnostics with respect to quasi-Lagrangian or vortex-following coordinate systems. These include a coordinate system based on the area within contours of PV or long-lived tracer (Butchart and Remsberg 1986; Lary et al. 1995), a coordinate system based on a smooth matching to a key vortex-edge PV contour (Norton 1994), and a coordinate system based on the mass within PV contours (Nakamura 1995). The latter coordinate system is particularly useful because it gives rise to a modified Lagrangian mean viewpoint (McIntyre 1980b; Dunkerton et al. 1981). The Lagrangian nature of these diagnostics comes from the close approximation of PV contours to material contours. The modified Lagrangian mean formalism in turn is closely related to the generalized Lagrangian mean formalism (Andrews and McIntyre 1978b; McIntyre 1980a,b; Dunkerton et al. 1981), and the two become equivalent when there is no diabatic heating, friction, or irreversible PV mixing. Consideration of certain integrals within and around PV contours leads to powerful nontransport and nonacceleration theorems that relate meridional transport and changes in the circulation to diabatic heating, friction, and irreversible PV mixing (O'Neill and Pope 1993a,b; Nakamura 1995; see also Hoskins 1991). Diagnostics calculated in the area-based coordinate system are quantitatively very similar to those in the mass-based coordinate system (Butchart and Remsberg 1986; Nakamura 1995), but the area-based coordinate system does not lead to such useful exact theoretical results as the nontransport and nonacceleration theorems.

The first main purpose of this paper is to compare the kinds of information obtained from Eulerian zonal mean diagnostics (section 3) and contour integral diagnostics (section 4) of a shallow water simulation of stratospheric wave breaking. Experimental details are given in section 2. In particular, we show that the contour integral diagnostics can be used to define a balanced, zonally symmetric state, which permits a direct comparison with Eulerian zonal mean diagnostics. The contour integral diagnostics also permit the cross-contour mass flux to be diagnosed; this leads to a view of meridional transport radically different from that given by the Eulerian mean meridional velocity.

An important component of a global view of the middle-atmosphere circulation is the behavior of waves and eddies and their effect on the mean flow. Conservation laws for wave activity or pseudomomentum help to quantify this behavior. Such a conservation law, or generalized Eliassen–Palm theorem, was derived by Andrews and McIntyre (1976, 1978a) for small-amplitude waves. A very general finite-amplitude wave action conservation law was derived by Andrews and McIntyre (1978c) using the generalized Lagrangian mean formalism. However, the terms in their result depend on parcel displacements, making it difficult to apply in practice. Essentially Eulerian finite-amplitude generalized Eliassen–Palm theorems applying under nonacceleration conditions have been derived by Edmon et al. (1980) for quasigeostrophic flow and by Andrews (1983) for flow governed by the hydrostatic primitive equations. More recently, Killworth and McIntyre (1985), McIntyre and Shepherd (1987), and Haynes (1988) have derived essentially Eulerian finite-amplitude generalized Eliassen–Palm theorems, including wave transience, forcing, and dissipation, for incompressible two-dimensional flow, quasigeostrophic flow, and flow governed by the hydrostatic primitive equations. McIntyre and Shepherd (1987) showed how the existence of wave activity conservation laws is related to symmetries of the basic-state flow and how such conservation laws can be systematically derived from conservation laws for energy or impulse and suitable Casimir invariants. Finite-amplitude wave activity conservation laws have been applied to the study of wave reflection from nonlinear critical layers by Killworth and McIntyre (1985) and Brunet and Haynes (1996), and baroclinic wave life cycles by Magnusdottir and Haynes (1996).

Wave activity diagnostics are complementary to the contour integral diagnostics in the following sense: the wave activity is generally dominated by a term that quantifies the spatial rearrangement of PV relative to some reference state, whereas the contour integral diagnostics give integral measures of the mass-PV distribution while effectively filtering out all details of the particular spatial distribution of PV.

The second main purpose of this paper is to investigate the behavior of nonlinear wave activity diagnostics in our shallow water simulation of stratospheric wave breaking (section 6). A new expression for the wave activity that does not involve integrals of integrals is presented. The zonal mean northward flux of wave

activity is compared with its small-amplitude analog, the EP flux; differences between them, and implications for interpreting EP fluxes in the stratosphere, are discussed.

2. Experimental details

Single-layer simulations of stratospheric wave breaking have been carried out previously using a nondivergent-model (Juckes and McIntyre 1987), shallow water models (e.g., Juckes 1989; Norton 1994; Bates and Li 1997), and an equivalent barotropic model (Salby et al. 1990). The model used in this study is the shallow water model of Thuburn (1997), which integrates the equations in the form

$$h_t + \nabla \cdot (\mathbf{v}h) = H, \tag{1}$$

$$(hQ)_t + \nabla \cdot (\mathbf{v}hQ) = -\nabla \cdot (\hat{\mathbf{k}} \times \mathbf{X}), \tag{2}$$

$$\delta_t = -\nabla \cdot \{hQ\hat{\mathbf{k}} \times \mathbf{v} + \nabla(g(h + h_0) + \mathbf{v}^2/2)\} + \nabla \cdot \mathbf{X}, \tag{3}$$

where the prognostic variables are h the fluid depth, Q the PV, and δ the divergence. The velocity $\mathbf{v}$ is diagnosed from the prognostic variables at each time step. The term H is a mass source, the shallow water analog of radiative forcing, and $\mathbf{X}$ may be friction or an externally imposed force per unit mass. The remaining terms are h_0 the lower boundary height, $h_0 + h$ the free surface height, g the acceleration due to gravity, and $\hat{\mathbf{k}}$ the unit vertical vector. In the shallow water system Q is related to the relative vorticity $\zeta = \hat{\mathbf{k}} \cdot \nabla \times \mathbf{v}$ and the depth by

$$Q = \zeta_a/h, \tag{4}$$

where $\zeta_a = \zeta + f$ is the absolute vorticity, f being the Coriolis parameter. Equations (1)–(3) are integrated on a hexagonal–icosahedral grid that gives uniform coverage of the sphere with no polar singularity. A quasi-third-order shape-preserving advection scheme is used to advect h and Q. This scheme is dissipative enough on its own to soak up the cascade of enstrophy to unresolved scales and so prevent the buildup of grid-scale noise without the need for any extra scale-selective dissipation terms (Thuburn 1995). The use of PV as a prognostic variable, combined with the use of schemes designed to avoid numerically generated noise, facilitates the calculation and interpretation of the PV-based diagnostics described in the following sections. In this study the model was run with a total of 40 962 grid boxes, giving an average grid box spacing of about 120 km.

The experiment simulates a cyclonic polar vortex disturbed by eddies forced through zonal asymmetries in the lower boundary height. The zonally symmetric part of the lower boundary height $\overline{h_0}$ is given by

$$\overline{h_0} = 2500 \text{ m} \times \left[1 - \exp\left(-0.69\frac{(\phi - 90°)^2}{(30°)^2}\right)\right], \tag{5}$$

where ϕ is latitude in degrees. The zonally asymmetric part of the lower boundary height h_0' is very similar to that used by Norton (1994) except that only a zonal wavenumber 1 component is used:

$$h_0' = h_{00}A(t)B(\phi)C(\lambda), \tag{6}$$

where

$$A(t) = \begin{cases} \dfrac{1}{2}\left(1 - \cos\left(\dfrac{\pi t}{4 \text{ days}}\right)\right) & \\ \qquad 0 \text{ days} \le t < 4 \text{ days} & \\ 1 \qquad 4 \text{ days} \le t < 16 \text{ days} & \\ \dfrac{1}{2}\left(1 + \cos\left(\dfrac{\pi(t - 16 \text{ days})}{4 \text{ days}}\right)\right) & \\ \qquad 16 \text{ days} \le t < 20 \text{ days}, & \end{cases} \tag{7}$$

$$B(\phi) = \begin{cases} \left(\dfrac{\cot^2\phi}{\cot^2\phi_0}\right)\exp\left(1 - \dfrac{\cot^2\phi}{\cot^2\phi_0}\right) & \phi > 0 \\ 0 & \phi \le 0, \end{cases} \tag{8}$$

where $\varphi_0 = 45°$, and

$$C(\lambda) = \sin\lambda. \tag{9}$$

For times beyond 20 days $A(t)$ repeats with a period of 20 days. The amplitude h_{00} is taken to be 720 m.

The mean circulation is forced, and eddies are damped, by relaxing the depth toward a ''radiative equilibrium'' depth $h_{\text{rad}} = 4000$ m on a timescale, τ_{rad},

$$H = (h_{\text{rad}} - h)/\tau_{\text{rad}}. \tag{10}$$

The explicit friction $\mathbf{X}$ is zero. In addition to this explicit heating and friction the model's advection scheme dissipates, that is, numerically mixes, small-scale Q (and h) features as their scale shrinks to the grid scale; thus there are implicit fluxes of mass and PV giving extra, implicit, contributions to H and $\mathbf{X}$. This will be discussed further in section 4.

The initial state for the main experiment was obtained by spinning up the model from rest ($\delta = 0$, $h = 4000$ m $- \overline{h_0}$, $Q = f/h$) for 20 days using only the zonally symmetric part of the lower boundary (5) and a short radiative timescale, $\tau_{\text{rad}} = 1$ day. After 20 days the model is very close to a steady state; h is everywhere within 7 m of h_{rad} and there is a balanced westerly jet peaking at about 37 m s^{-1} at 60°N.

The main experiment was run from the spun-up state for 60 days using the full lower boundary (5) and (6) with a radiative timescale, $\tau_{\text{rad}} = 20$ days, representative of the lower stratosphere. Figures 1, 2, and 3 show a selection of PV maps from interesting periods of the

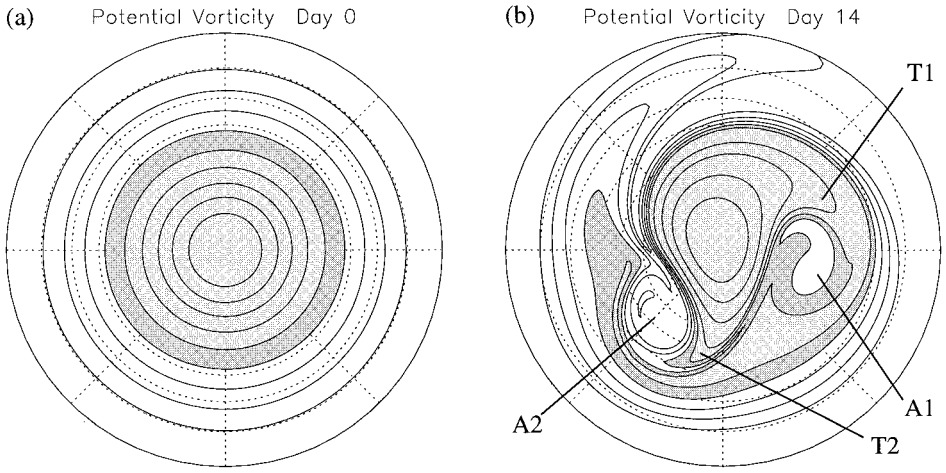

FIG. 1. Northern Hemisphere PV at (a) day 0 and (b) day 14. Contours are from 5.23×10^{-9} m^{-1} s^{-1} to 40.27×10^{-9} m^{-1} s^{-1} in steps of 4.38×10^{-9} m^{-1} s^{-1}. The PV greater than 22.75×10^{-9} m^{-1} s^{-1} is indicated by light shading. The PV in the range 18.37×10^{-9} m^{-1} s^{-1} to 22.75×10^{-9} m^{-1} s^{-1} is indicated by dark shading. The labelled features are discussed in the text.

simulation. Each of these periods will be discussed further in terms of other diagnostics in the following sections.

Figure 1a shows the initial, zonally symmetric locations of nine PV contours. Contour integral diagnostics for the same contours will be presented in section 4. The shading is intended to help identify the contours. Figure 1b shows PV at day 14. It shows a pair of anticyclones (A1, A2) that formed simultaneously, the westward member of the pair (A2) being more intense. A tongue of high PV air (T1, T2) wraps around each anticyclone. Over the next 10 days A1 weakens in situ while A2 weakens and migrates westward.

Between days 26 and 36 a third anticyclone (A3) and a fourth anticyclone (A4) form in roughly the same locations as A1 and A2. Some high PV air is advected around A3 and moves westward to the south of A3 and A4, eventually wrapping up into a cyclone, C1. Subsequently, a large tongue of high PV air (T3) is pulled from the main vortex around A3.

Figure 2 shows a period of particularly dramatic evolution over the next few days. Anticyclone A3 migrates eastward. On day 40 another tongue of high PV air (T4) is pulled from the main vortex, this time around A4. Meanwhile, a small, intense blob of low PV air (A6) at the tip of a thin tongue moves into high latitudes. On day 42 yet another tongue (T5) of high PV air is pulled from the main vortex, this time around A6, and A6 begins to merge with A4. On day 44 the merging of A6 with A4 continues. At the same time T4 wraps up into two cyclones C2 and C3, while the tongue of low PV air that was originally attached to A6 wraps up into two small anticyclones A7 and A8.

Over the next few days many of these features weaken

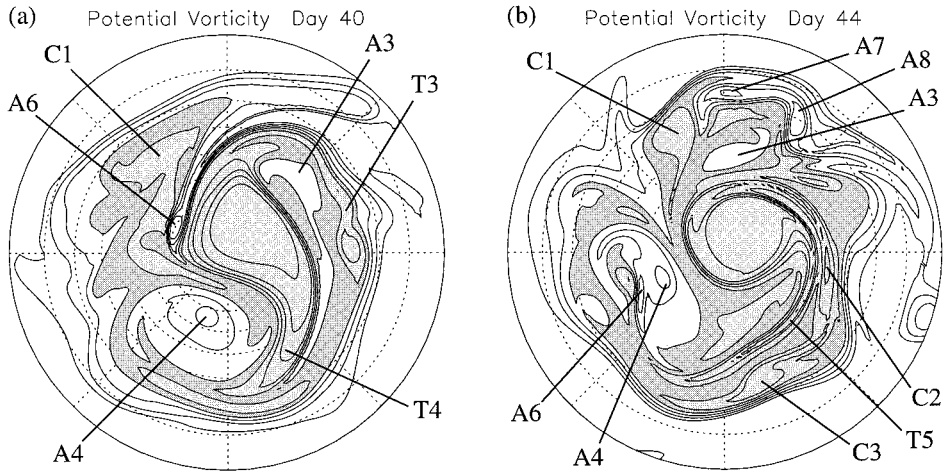

FIG. 2. As in Fig. 1 but at (a) day 40 and (b) day 44.

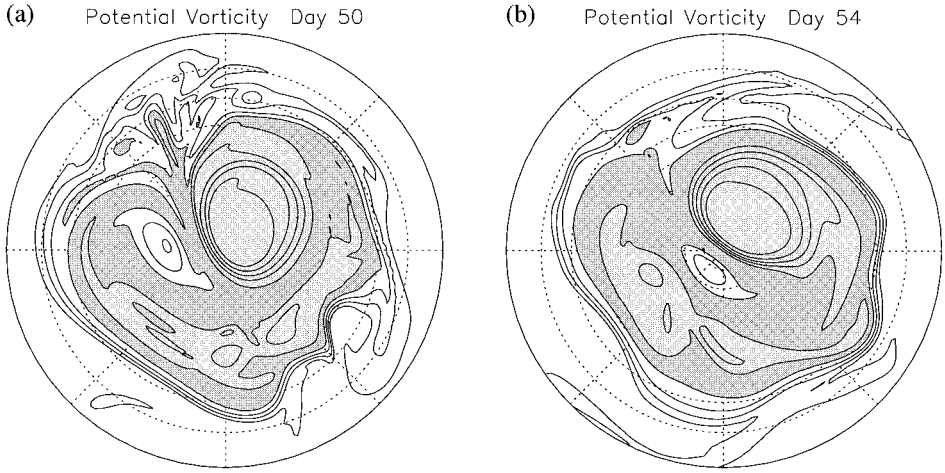

FIG. 3. As in Fig. 1 but at (a) day 50 and (b) day 54.

or are sheared out. Figure 3 shows that by day 50 a single large anticyclone, A4/A6, remains, being fed by tongues of low PV air from low latitudes. The region outside the main vortex where mixing due to wave breaking tends to homogenize PV is often referred to as the ''surf zone.'' The mature surf zone contains some cyclonic debris: the remains of the tongues of high PV air pulled from the main vortex. Sharp PV gradients at the edge of the main vortex and in the subtropics are clearly visible. The main vortex is now reduced in size and is easily displaced from the pole by the large anticyclone.

3. Eulerian zonal mean diagnostics

In this section we examine Eulerian zonal mean diagnostics of the experiment described in section 2. Many of the diagnostics presented in this section and in section 6 require quantities not predicted directly by the numerical model. These diagnostics were calculated by diagnosing the wind components normal and tangential to gridbox edges on the model grid, converting these to eastward and northward components, and interpolating wind, depth, and PV to a regular 129 latitude $\times$ 256 longitude grid suitable for taking zonal averages. Zonal mean diagnostics were computed every 6 h.

The shallow water model is analogous to a layer of the three-dimensional atmosphere bounded by a pair of isentropic surfaces. It is therefore appropriate to examine mass-weighted Eulerian mean diagnostics for the shallow water model, analogous to density-weighted isentropic coordinate Eulerian mean diagnostics for the three-dimensional atmosphere. We first establish some notation. Define an overbar to indicate a zonal mean, $\overline{Y} \equiv (1/2\pi)\int_0^{2\pi} Y\, d\lambda$, and prime to indicate a departure from the zonal mean, $Y' = Y - \overline{Y}$. Also define overbar star to indicate a mass-weighted zonal mean, $\overline{Y}^* \equiv \overline{hY}/\overline{h}$, and circumflex to indicate a departure from the mass-weighted zonal mean, $\hat{Y} = Y - \overline{Y}^*$.

Figure 4a shows $\overline{u}$ for the 60 days of the simulation. From about day 5 the mean flow begins to differ significantly from its initial profile. The peak mean winds at about 60°N weaken at first but then strengthen between days 10 and 20. After weakening to a minimum at day 30 the strong mean winds are reestablished until about day 50, but at a higher latitude and more intermittently than earlier in the simulation. Finally, beyond day 50 the high-latitude mean zonal wind weakens dramatically and eventually reverses. At lower latitudes $\overline{u}$ also weakens from about day 5 onward, leading to easterlies at about 30°N between days 12 and 25 and between days 32 and 52. These low-latitude mean easterlies occur when the high-latitude westerlies are strongest. In the Tropics, mean easterlies centered at about 10°S gradually strengthen throughout the simulation.

Figure 5a shows $\overline{h}$ for the 60 days of the simulation. Its behavior is consistent with gradient wind balance between $\overline{u}$ and $\overline{h}$,

$$\left(f + \frac{\overline{u}\tan\phi}{a}\right)\overline{u} \approx -g\frac{(\overline{h} + \overline{h_0})_\phi}{a}, \qquad (11)$$

which holds to an excellent approximation. Mid- to high-latitude maxima in $\overline{h}$ occur at the times when $\overline{u}$ has its strongest high-latitude maxima and subtropical easterlies. The polar value of $\overline{h}$ dramatically increases during the final 10 days when $\overline{u}$ is dramatically weakened.

Figure 6a shows $\overline{Q}$ for the 60 days of the simulation. There are weakened gradients of $\overline{Q}$ in the latitude band 30°–50°N between days 10 and 25 and between days 30 and 60, and it is tempting to identify these with the surf zone. Equatorward of this at around 20°N, from about day 30 onward, the gradient of $\overline{Q}$ is intensified. There is also some indication of the formation of a sharp PV gradient to the north of the surf zone at the edge of the vortex. However, this Eulerian mean diagnostic does not distinguish whether any of these features are due to

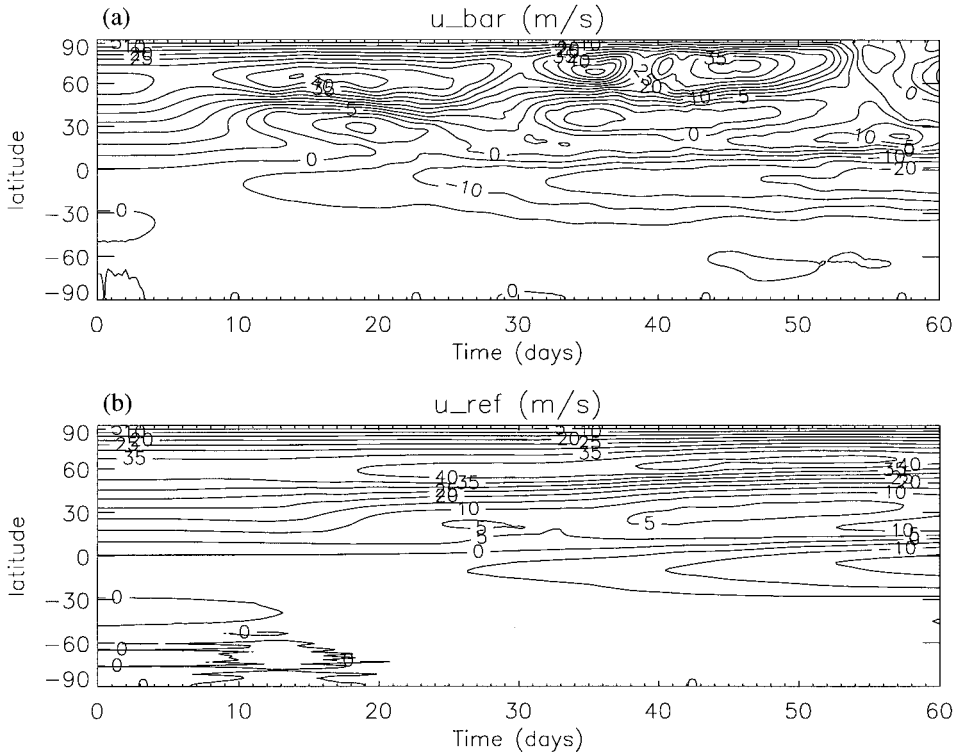

FIG. 4. The terms (a) $\overline{u}$ and (b) u_{ref} vs time and latitude, contour interval 5 m s^{-1}.

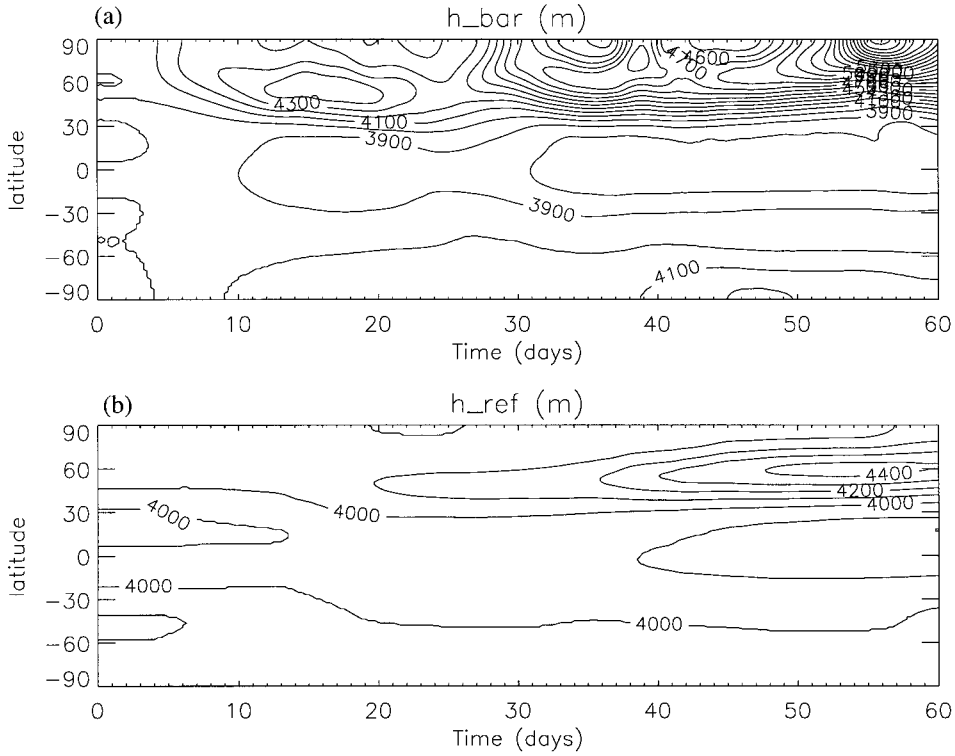

FIG. 5. The terms (a) $\overline{h}$ and (b) h_{ref} vs time and latitude, contour interval 100 m.

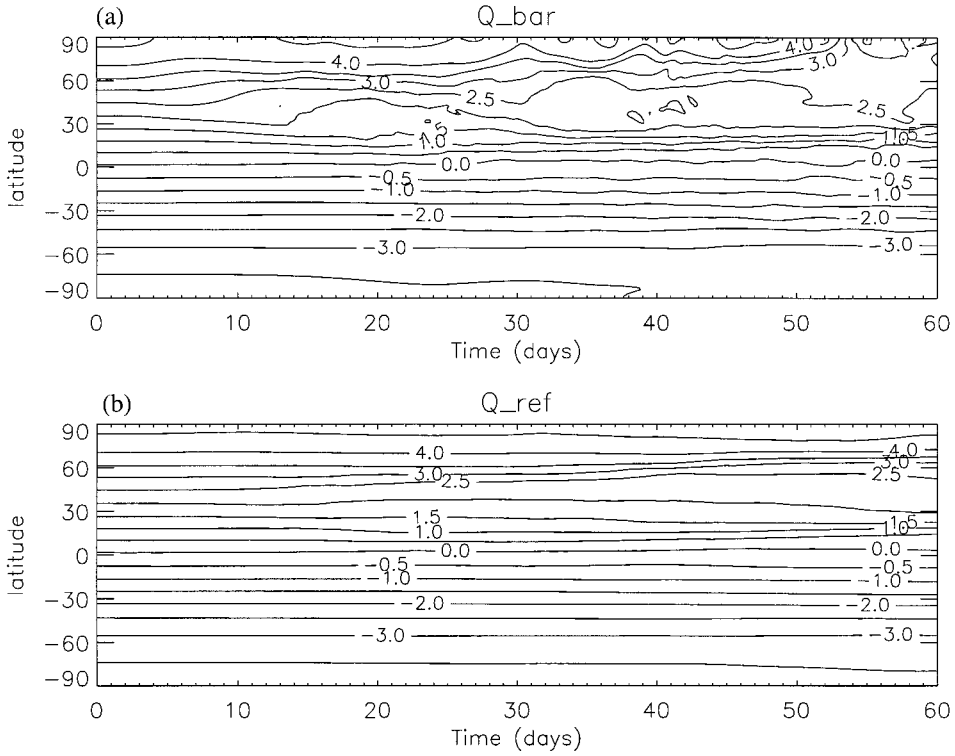

FIG. 6. The terms (a) $\overline{Q}$ and (b) Q_{ref} vs time and latitude, contour interval 0.5×10^{-9} m^{-1} s^{-1}.

genuine small-scale, irreversible PV mixing or simply the presence of large-scale eddies plus the effect of zonal averaging.

The zonal mean mass and angular momentum budgets are now examined. The Eulerian zonal mean of the mass continuity equation (1) can be written

$$\overline{h}_t + \frac{1}{a\cos\phi}(\overline{h}\overline{v}^*\cos\phi)_\phi + \overline{H}. \tag{12}$$

When averaged over a long enough time for $\overline{h}_t$ to be negligible, (12) reduces to a balance between the divergence of the mean northward mass flux and the radiative mass source.

Figure 7 shows $\overline{v}^*$ and $\overline{H}$ for the 60 days of the simulation. These quantities are the shallow water analog of the three-dimensional isentropic coordinate mean meridional circulation, with positive $\overline{H}$ corresponding to positive diabatic heating. Three passes of a ¼:½:¼ time filter have been applied to the $\overline{v}^*$ field to suppress a short timescale signal in the later part of the simulation apparently associated with a global-scale gravity wave. Large positive values of $\overline{v}^*$ occur roughly when the lower boundary wave forcing is switched on, except for a transient event at high latitudes beginning around day 37. This will be seen below to be consistent with the angular momentum budget. This northward mass flux leads to increasing $\overline{h}$ at high latitudes and decreasing $\overline{h}$ in the Tropics (Fig. 5) and hence to the pattern of $\overline{H}$ shown in Fig. 7b. In this simulation, the mass budget on all timescales is dominated by the approximate balance $\overline{h}_t + [1/(a\cos\phi)](\overline{hv}^*\cos\phi)_\phi \approx 0$. In particular, these two terms show equal and opposite values of several hundred meters per day for events of duration 5 days or less. Because the radiative timescale τ_{rad} is long, and the timescale for adjustment to radiative forcing on length scale $L\tau_{\text{adj}} \approx (1 + g\overline{h}/f^2L^2)\tau_{\text{rad}}$ is even longer, we do not see the approximate balance $[1/(a\cos\phi)](\overline{hv}^*\cos\phi)_\phi \approx \overline{H}$, even when averaged over the full 60 days of the simulation. However, we would expect to see this approximate balance over longer timescales or when using a shorter τ_{rad} representative of higher stratospheric altitudes.

The zonal mean zonal momentum equation may be written in two ways. The first way is most easily derived by taking the zonal mean of the flux form of the zonal momentum equation, combined with (12), and is

$$\overline{u}_t - \overline{v}^*\overline{\zeta_a} = \frac{1}{\overline{h}\cos\phi}\nabla\cdot\mathbf{F} - \frac{g\overline{hh_{0\lambda}}}{a\overline{h}\cos\phi} - \frac{1}{\overline{h}}(\overline{h'u'})_t + \frac{1}{\overline{h}}\overline{u'H'} + \overline{X}^*, \tag{13}$$

where $\nabla\cdot\mathbf{F} = [1/(a\cos\phi)](F^{(\phi)}\cos\phi)_\phi$ and $F^{(\phi)} = -\overline{(hv)'u'}\cos\phi$ is the shallow water version of the EP flux. In practice, the third and fourth terms on the right-hand side of (13) are very small compared with the others, while the explicit contribution to the friction X is zero in the experiment described here. Thus, when

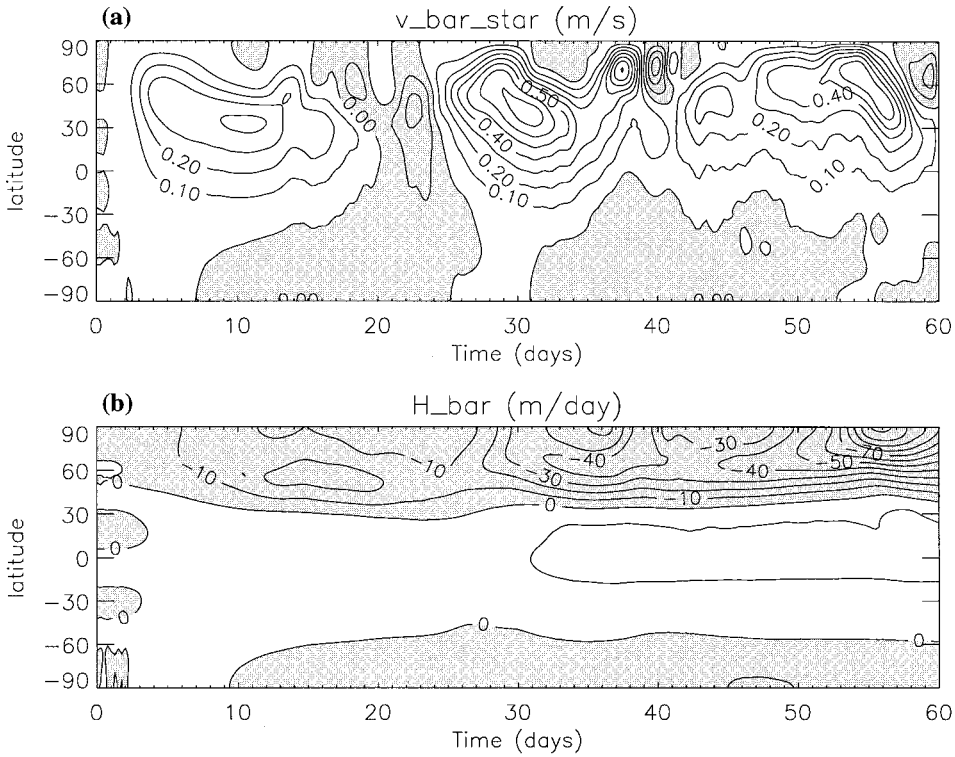

FIG. 7. Time–latitude plots of (a) $\overline{v}^*$, contour interval 0.1 m s^{-1}, and (b) $\overline{H}$, contour interval 10 m day^{-1}. Negative values are shaded.

averaged over a long enough period for $\overline{u}_t$ to be negligible, the dominant balance in (13) is between advection of absolute vorticity by the mean meridional flow and the net convergence of the eddy transport of angular momentum, where the orographic forcing term is interpreted as the convergence of the vertical eddy transport of angular momentum.

The second way to write the zonal mean zonal momentum equation is most easily derived by taking the zonal mean of the advective form of the zonal momentum equation and is

$$\overline{u}_t - \overline{v}^*\overline{\zeta_a} = \overline{h}\,\overline{\hat{v}\hat{Q}}^* + \overline{X}. \tag{14}$$

Comparison of (13) with (14), whose left-hand sides are the same, highlights the approximate relationship between eddy transport of angular momentum and PV rearrangement (a relationship that is exact in the quasigeostrophic limit). Equations (12), (13), and (14) are shallow water analogs of Eqs. (3.9.4), (3.9.7a), and (3.9.9) of Andrews et al. (1987) for a three-dimensional atmosphere in isentropic coordinates.

Figure 8 shows the dominant terms in the zonal mean angular momentum budget (13). On short timescales of an order of 5 days the balance is principally $\overline{u}_t \approx [1/(\overline{h}\cos\phi)]\nabla\cdot\mathbf{F}$, with some contribution from $\overline{v}^*\overline{\zeta_a}$. This is particularly true for the dramatic deceleration between days 50 and 55. On longer timescales there is a three-way balance $-\overline{v}^*\overline{\zeta_a} \approx [1/(\overline{h}\cos\phi)]\nabla\cdot\mathbf{F} - [g\overline{h}\overline{h}_{0\lambda}/(a\overline{h}\cos\phi)]$. Of the remaining terms in the budget, $(1/\overline{h})(\overline{h'u'})_t$ is less than 2 m s^{-1} day^{-1} virtually everywhere, while $(1/\overline{h})\overline{u'H'}$ is a further order of magnitude smaller. Figure 9a shows the eddy PV flux $\overline{h\hat{v}\hat{Q}}^*$; comparison with Fig. 9b confirms that to an excellent approximation $\overline{h\hat{v}\hat{Q}}^* \approx [1/(\overline{h}\cos\phi)]\nabla\cdot\mathbf{F} - [g\overline{h}\overline{h}_{0\lambda}/(a\overline{h}\cos\phi)]$.

An interesting feature of Fig. 8d is that the lower boundary wave forcing almost always acts to reduce the zonal mean angular momentum, an exception being the short period between days 20 and 25. This means that the trough in h usually occurs to the west of the trough in h_0; this is the shallow water analog of the westward tilt with height that occurs for upward propagating planetary waves in three-dimensional flow. The reduction in the lower boundary forcing amplitude between days 16 and 24 allows the wave structure to change temporarily into a configuration that increases the zonal mean angular momentum of the flow. During this short period the lower boundary forcing acts as a sink of wave activity, rather than a source (see section 6).

4. Contour integral diagnostics

We examine two related contour integral diagnostics. The first is the mass within the PV $= Q$ contour:

$$\mathcal{M}(Q) = \int_{\mathrm{PV}>Q} h\, dA. \tag{15}$$

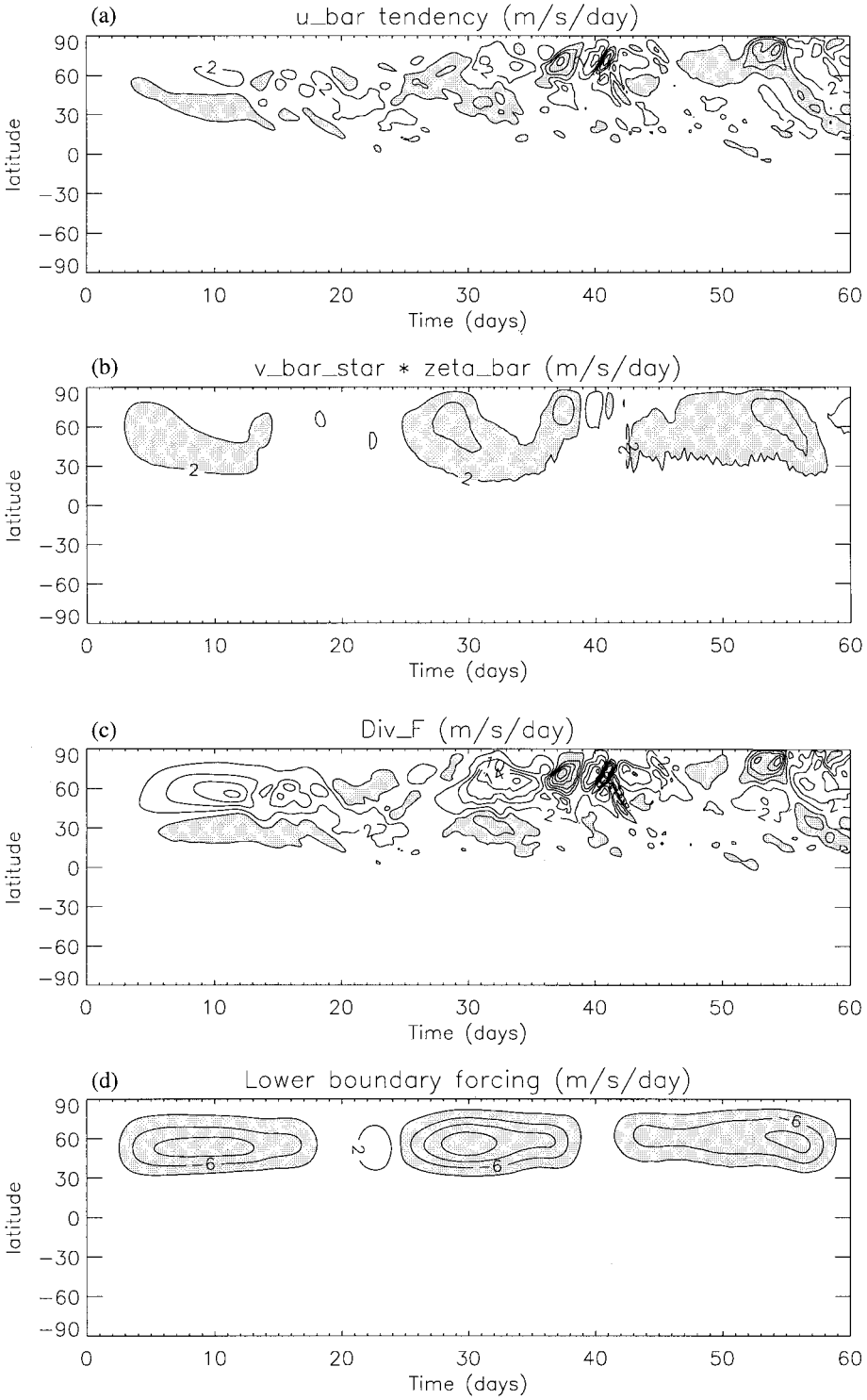

FIG. 8. Time–latitude plots of terms in the $\overline{u}$ budget. (a) $\overline{u}_t$, (b) $\overline{v}^*\overline{\zeta_a}$, (c) $[1/(\overline{h}\cos\phi)]\nabla\cdot\mathbf{F}$, (d) $-[g\overline{hh}_{0\lambda}/(a\overline{h}\cos\phi)]$. The contour interval is 4 m s^{-1} day^{-1} with contours chosen to include ± 2 m s^{-1} day^{-1}. Values less than -2 m s^{-1} day^{-1} are shaded, except in (b) where values greater than 2 m s^{-1} day^{-1} are shaded.

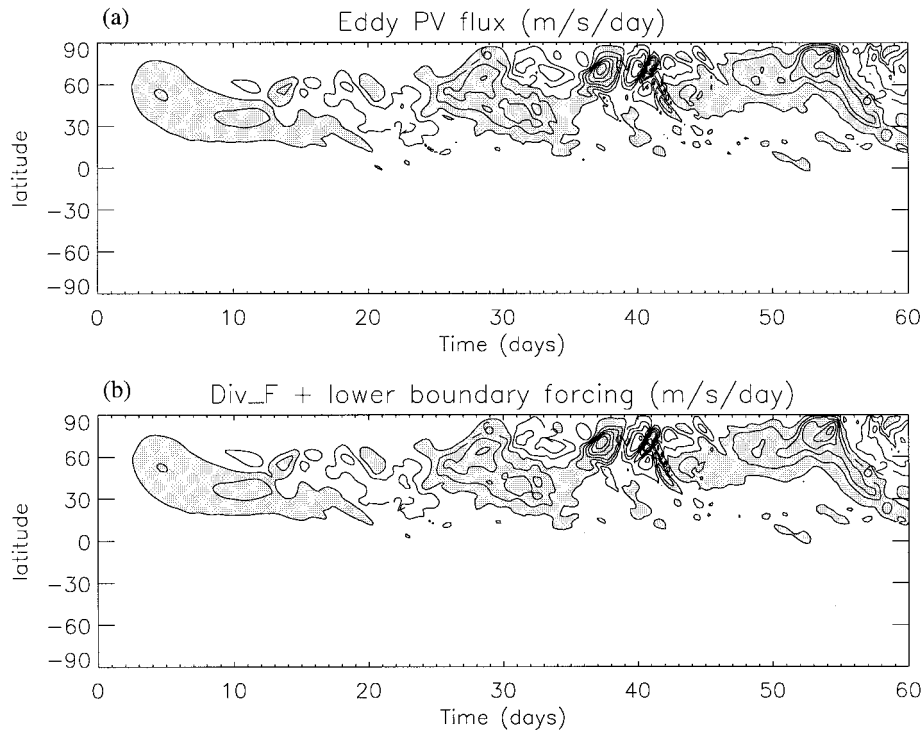

FIG. 9. As in Fig. 8 but for (a) $\overline{h}\,\overline{\hat{v}\hat{Q}}^{*}$ and (b) $[1/(\overline{h}\cos\phi)]\nabla\cdot\mathbf{F} - [g\overline{hh_{0\lambda}}/(a\overline{h}\cos\phi)]$.

Throughout this paper we define the ''inside'' of a PV contour to be the region of higher Q, even in the Southern Hemisphere where the ''outside'' may be a small region and the inside may cover most of the globe. Also, strictly, $\mathcal{M}(Q)$ as defined by (15) is the *volume* of fluid within the PV $= Q$ contour. However, we can take the fluid density to be unity, and, since the mass of fluid (per unit potential temperature) is the relevant generalization to three-dimensional compressible flow, we will use the term *mass* throughout this paper.

The evolution of $\mathcal{M}$ with time at fixed Q is given by

$$\frac{\partial}{\partial t}\mathcal{M} = \mathcal{V}(Q) + \int_{\mathrm{PV}>Q} H\, dA, \tag{16}$$

where $\mathcal{V}(Q)$ is the net mass flux across the PV $= Q$ contour toward higher PV values. An expression for $\mathcal{V}$ is

$$\mathcal{V}(Q) = \oint_{\mathrm{PV}=Q} h\mathbf{v} - (\mathbf{v}_c)\cdot\mathbf{n}\, dl, \tag{17}$$

where $\mathbf{v}_c$ is the velocity of the PV contour and $\mathbf{n} \equiv \nabla Q/|\nabla Q|$ is the unit normal to the contour in the direction of higher PV. The velocity of the PV contour is

$$\mathbf{v}_c = -\frac{Q_t}{|\nabla Q|}\mathbf{n} = \frac{\mathbf{n}}{|\nabla Q|}\left[\mathbf{v}\cdot\nabla Q + \frac{QH}{h} + \frac{1}{h}\nabla\cdot(\hat{\mathbf{k}}\times\mathbf{X})\right]. \tag{18}$$

Substituting (18) in (17) gives

$$\mathcal{V}(Q) = -\oint_{\mathrm{PV}=Q} \frac{QH + \nabla\cdot(\hat{\mathbf{k}}\times\mathbf{X})}{|\nabla Q|}\, dl. \tag{19}$$

Thus changes in $\mathcal{M}$ can occur only through diabatic or frictional effects on or within the PV contour, and cross-contour mass transport $\mathcal{V}$ can occur only through diabatic or frictional effects on the PV contour. Recall that in our numerical model H and $\mathbf{X}$ include implicit contributions from numerical mixing. In adiabatic, frictionless flow $\mathcal{M}$ is constant and $\mathcal{V}$ is zero. Nakamura (1995) calls this result a nontransport theorem.

A second quantity of interest is the absolute circulation around the PV $= Q$ contour:

$$\mathcal{C}(Q) = \oint_{\mathrm{PV}=Q} \mathbf{v}_{\mathrm{abs}}\cdot d\mathbf{l} = \int_{\mathrm{PV}=\tilde{Q}>Q} h\tilde{Q}\, dA, \tag{20}$$

where $\mathbf{v}_{\mathrm{abs}}$ is the absolute velocity including the contribution from the rotating earth, and the integral is taken counterclockwise around the contour, bearing in mind the definition of inside given above. The evolution of $\mathcal{C}$ with time at fixed Q is given by

$$\frac{\partial}{\partial t}\mathcal{C} = \oint_{\mathrm{PV}=Q} hQ(\mathbf{v}-\mathbf{v}_c)\cdot\mathbf{n}\, dl + \int_{\mathrm{PV}>Q} \nabla\cdot(\hat{\mathbf{k}}\times\mathbf{X})\, dA$$

$$= Q\mathcal{V}(Q) + \oint_{\mathrm{PV}=Q} \mathbf{X}\cdot d\mathbf{l}. \tag{21}$$

Alternative expressions for $(\partial C/\partial t)$ (valid for three-dimensional flow) have been presented by O'Neill and Pope (1993a,b) and Nakamura (1995). Equation (21) bears some formal similarity to Eq. (14), but the separation of the PV flux into a mean part ($\overline{v}^{*}\overline{\zeta}_a \equiv \overline{h}\,\overline{v}^{*}\overline{Q}^{*}$) and an eddy part in (14) does not occur in (21). Changes in $\mathcal{C}$ can occur only through diabatic or frictional effects (including real or numerical mixing) on the PV $= Q$ contour, so that in adiabatic frictionless flow $\mathcal{C}$ is constant. Nakamura (1995) calls this result a nonacceleration theorem.

Note that if the region of PV $> Q$ is composed of several distinct islands then $\mathcal{M}(Q)$ is equal to the sum of the contributions from each of the islands. Moreover, except when islands merge or split, the contribution from each island satisfies an equation like (16), where the integrals defining $\mathcal{M}$, $\mathcal{V}$, and the heating term are restricted to the island in question. Similar comments apply to $\mathcal{C}(Q)$.

On any given PV contour, $\mathcal{M}$ and $\mathcal{C}$ are independent quantities. However, their derivatives in PV space are related by

$$Q\frac{\partial}{\partial Q}\mathcal{M} = \frac{\partial}{\partial Q}\mathcal{C}, \tag{22}$$

which implies that if we know the complete distribution of either $\mathcal{M}$ or $\mathcal{C}$ as a function of Q then we can deduce the complete distribution of the other, using the boundary condition $\mathcal{M} = \mathcal{C} = 0$ where Q takes its maximum value. Equation (22) follows from

$$\frac{\partial}{\partial Q}\mathcal{M} = \oint_{\mathrm{PV}=Q} \frac{h}{|\nabla Q|}\, dl \tag{23}$$

and

$$\frac{\partial}{\partial Q}\mathcal{C} = -\oint_{\mathrm{PV}=Q} \frac{hQ}{|\nabla Q|}\, dl. \tag{24}$$

Equation (22) has been noted independently by D. G. Andrews (1997, personal communication) for the three-dimensional primitive equation case. It will be used to simplify the expression for wave activity in section 6.

The information contained in $\mathcal{C}$ can be expressed through an alternative quantity, $\mathcal{P}$, having the same dimensions as $\mathcal{M}$ and defined by

$$\mathcal{P}(Q) = \mathcal{C}(Q)/Q. \tag{25}$$

From (20) and (25) we obtain $\mathcal{P}(Q) = \int_{\mathrm{PV}=\tilde{Q}>Q} h_{\mathrm{dil}}\, dA$, where $h_{\mathrm{dil}} = h\tilde{Q}/Q$ is the depth that a fluid parcel would have if its PV, initially equal to $\tilde{Q}$, were diluted to equal the contour value Q by radiatively increasing h. Thus, $\mathcal{P}(Q)$ can be interpreted as the total mass that would lie within the PV $= Q$ contour if all the PV within the contour were diluted to equal the contour value Q by radiatively increasing h. For this reason we call $\mathcal{P}$ the "potential mass." From (21) the evolution of $\mathcal{P}$ with time at constant Q is given by

$$\frac{\partial}{\partial t}\mathcal{P} = \mathcal{V}(Q) + \frac{1}{Q}\oint_{\mathrm{PV}=Q} \mathbf{X}\cdot d\mathbf{l}. \tag{26}$$

This equation enables changes in the circulation to be related to changes in the mass distribution. For example, eliminating $\mathcal{V}$ from (16) and (26) gives

$$\frac{\partial}{\partial t}(\mathcal{P} - \mathcal{M}) = \frac{1}{Q}\oint_{\mathrm{PV}=Q} \mathbf{X}\cdot d\mathbf{l} - \int_{\mathrm{PV}>Q} H\, dA. \tag{27}$$

Then, over a long enough time average for the time derivative to be negligible, for example, over an annual cycle,

$$\frac{1}{Q}\oint_{\mathrm{PV}=Q} \mathbf{X}\cdot d\mathbf{l} \approx \int_{\mathrm{PV}>Q} H\, dA. \tag{28}$$

Thus, there must be a long-term balance between friction on a PV contour and heating within that contour. [Compare the contour integral result of Hoskins (1991) relating friction and heating on a contour of constant potential temperature at or near the ground.]

A further result involving $\mathcal{P}$ is the following. If an island positive PV anomaly initially within a PV $= Q$ contour is dissipated by radiation ($\mathbf{X} = 0$) so that all air initially within the contour finally has PV $< Q$ leaving $\mathcal{M}(Q) = \mathcal{P}(Q) = 0$, then the total amount of mass that must cross the PV contour is, from the time integral of (26), equal to the potential mass of the island $\mathcal{P}(Q)$ rather than the actual mass $\mathcal{M}(Q)$. A disadvantage of $\mathcal{P}$ as a diagnostic is that it becomes infinite as $Q \to 0$ near the equator.

The quantities

$$\mathcal{M}(Q) \approx \sum_{Q_i>Q} h_i \Delta A_i, \qquad \mathcal{C}(Q) \approx \sum_{Q_i>Q} h_i Q_i \Delta A_i, \quad \text{and}$$

$$\int_{\mathrm{PV}>Q} H\, dA \approx \frac{1}{\tau_{\mathrm{rad}}} \sum_{Q_i>Q} (h_{\mathrm{rad}} - h_i)\Delta A_i$$

were calculated for about 150 PV contours every 6 h for the experiment described in section 2. The calculation simply involves counting only those grid boxes for which PV is greater than the contour value. For coarser-resolution model experiments (or analyzed data) it may be necessary to estimate contributions from fractional grid boxes to avoid noisy diagnostics (A. Gregory 1997, personal communication). However, for the high model resolution used here, counting whole grid boxes was found to be adequate.

Figure 10 shows $\mathcal{M}$ and $\mathcal{C}$ as functions of time for a number of selected Northern Hemisphere PV contours. The contours are the same as those shown in the synoptic PV maps (Figs. 1–3). The timescale for the evolution of $\mathcal{M}$ and $\mathcal{C}$, that is, for the action of irreversible mixing and diabatic and frictional processes, is much longer than the timescale for the evolution of $\overline{h}$ and $\overline{u}$. Nevertheless, we can identify the features discussed in section 3 above: the sharpening of the PV gradient at the edge of the vortex, and indeed a shrinking of the vortex;

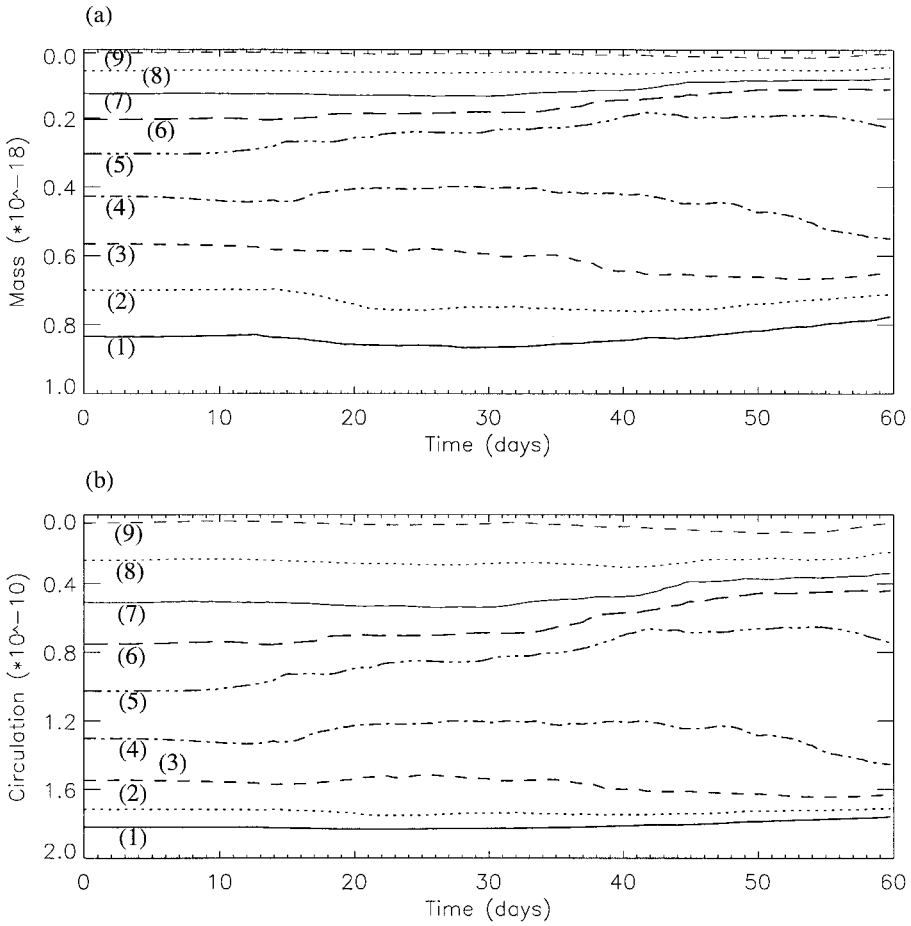

FIG. 10. Contour integrals (a) $\mathcal{M}(Q)$ and (b) $\mathcal{C}(Q)$ vs time for PV contours from (1) 5.23×10^{-9} m^{-1} s^{-1} to (9) 40.27×10^{-9} m^{-1} s^{-1} in steps of 4.38×10^{-9} m^{-1} s^{-1}. The contours are the same as those plotted in Figs. 1–3.

the formation of a surf zone with a weakened PV gradient; and the strengthening of the PV gradient in the subtropics on the equatorward flank of the surf zone. The region between contours (4) and (5) corresponds to the dark shaded region in Figs. 1–3. The mass between these two contours increases by more than 160% over the course of the integration. The behavior of the two diagnostics $\mathcal{M}$ and $\mathcal{C}$ is qualitatively very similar, as we would expect from (22). The relationship (22) was checked and found to hold very well for the diagnostics obtained numerically in this experiment.

In principle, the cross-contour mass flux $\mathcal{V}$ could be calculated explicitly from (19) if H and $\mathbf{X}$ were known. However, as noted in section 2, there are unknown contributions to H and $\mathbf{X}$ from numerical mixing. Therefore, a more practical way to estimate $\mathcal{V}$ is as a residual in (16). The second term on the right-hand side of (16) can be calculated by integrating the explicit heating rate within the PV $= Q$ contour; this term is not affected by the numerical mixing. The $(\partial\mathcal{M}/\partial t)$ term is easily estimated from successive values of $\mathcal{M}$ 6 h apart.

Alternatively, we could attempt to estimate $\mathcal{V}$ as a residual in (26). In this case we must neglect the unknown term $(1/Q)\oint_{\mathrm{PV}=Q}\mathbf{X}\cdot d\mathbf{l}$ even though it includes an implicit, and possibly significant, contribution from numerical mixing. The difference between $\mathcal{V}$ estimated from (16) and $\mathcal{V}$ estimated from (26) gives an estimate of the neglected term. In fact these differences turn out to be rather small, and both estimates show the same main features in time series of $\mathcal{V}$. Note that it is also possible to decompose $\mathcal{V}$ into a resolved-scale radiative contribution, which could be calculated explicitly using the resolved terms in (19), and a contribution from unresolved mixing processes, which could be calculated as a residual; but we have not done this here.

One caveat should be borne in mind when interpreting these results. The scale at which small-scale PV features are dissipated in the model depends on the model resolution. Therefore, as a feature becomes thinner under straining, the timing of its dissipation will depend on model resolution. Thus the timing of the transport events discussed below will be sensitive to model resolution. The degree of this sensitivity, which we have not yet investigated, should give valuable information on cas-

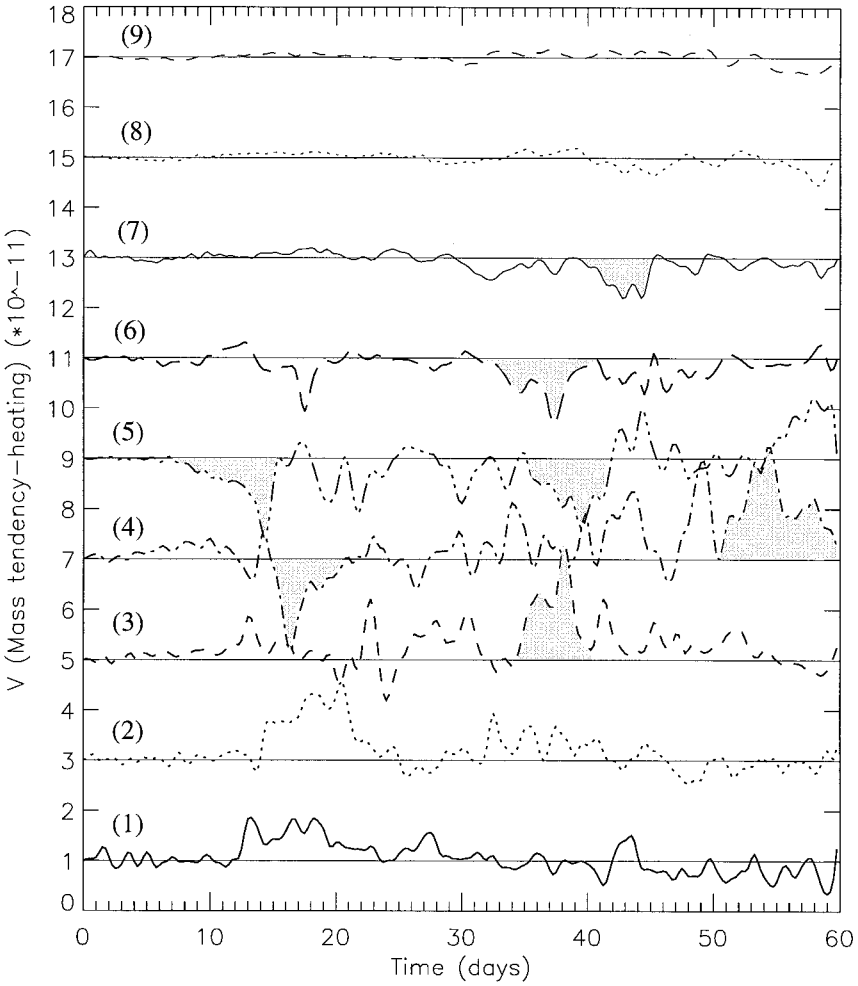

FIG. 11. Cross-PV-contour mass flux $\mathcal{V}(Q)$ for the same PV contours as in Figs. 1–3 and 10. The zero for each curve, indicated by a solid straight line, is offset along the $\mathcal{V}$ axis for clarity. Features discussed in the text are highlighted by shading.

cade rates and chemical mixdown times (e.g., Juckes and McIntyre 1987; Haynes and Anglade 1997; Thuburn and Tan 1997).

Figure 11 shows $\mathcal{V}$ diagnosed as a residual in (16) for the same PV contours shown in Figs. 1–3 and 10. It shows numerous events of both signs with typical duration 2–5 days. Most of these transport events appear to be associated with synoptic events in which small-scale PV features are formed and rapidly strained. Numerical mixing of these features then gives rise to significant cross-contour mass fluxes. In the real atmosphere we would also expect such features to be rapidly dissipated or assimilated into the background flow, either by radiation (e.g., Haynes and Ward 1993), or by small-scale mixing processes like breaking inertio-gravity waves (e.g., Dewan 1981), or, ultimately, by molecular diffusion. Generally, tongues of high PV air pulled from the main vortex are associated with negative $\mathcal{V}$, while tongues of low PV air feeding into higher-latitude anticyclones are associated with positive $\mathcal{V}$. The cross-contour transport near the vortex edge is predominantly one-sided toward lower PV, that is, out of the vortex. Similarly, the cross-contour transport in the subtropics is predominantly toward higher PV. Thus, there is net entrainment of air into the surf zone from both sides.

Several of the events in Fig. 11 correspond to the features in Figs. 1, 2, and 3: the pair of negative $\mathcal{V}$ events at days 14 and 16 [contours (4) and (5)] correspond to the pair of high PV tongues T1, T2 that are wrapped around the anticyclones (Fig. 1); the succession of negative $\mathcal{V}$ events from days 36 to 44 [contours (6), (5), and (7)] correspond to the three high PV tongues T3, T4, T5 pulled from the main vortex, while the positive $\mathcal{V}$ event around day 38 [contour (3)] corresponds to the injection of anticyclone A6 into high latitudes (Fig. 2); the large positive $\mathcal{V}$ event around day 54 [contour (4)] corresponds to the low PV air feeding into the remaining large anticyclone (Fig. 3).

It is interesting to contrast the two measures of the meridional circulation, $\mathcal{V}$ (Fig. 11) and $\overline{v}^*$ (Fig. 7). Variations in $\mathcal{V}$ occur on a 2–5-day timescale associated with synoptic events as just described, whereas variations in $\overline{v}^*$ occur mainly on the 20-day timescale of the lower boundary forcing. These differences reflect the different physical mechanisms driving $\mathcal{V}$ and $\overline{v}^*$: $\mathcal{V}$ is driven mainly by the rapid generation and straining by eddies of small-scale PV features, whereas $\overline{v}^*$ results from the continual adjustment toward balance in response to the angular momentum sink at the lower boundary. Also, $\overline{v}^*$ is predominantly poleward at all latitudes, except for a reversal around day 22 and the events around day 40, whereas $\mathcal{V}$ is predominantly into the surf zone from both sides.

Although $\mathcal{V}$ and $\overline{v}^*$ give such apparently different pictures, they are both measures of the Lagrangian mean mass circulation and therefore it must be possible to reconcile them, at least over long enough timescales. For example, in the above experiment, near the vortex edge $\overline{v}^*$ implies poleward mass flow relative to latitude circles, whereas $\mathcal{V}$ implies equatorward mass flow relative to PV contours. These are consistent because the PV contours themselves move poleward on average during the experiment. On the timescale of the annual cycle, over which PV contours return to roughly their original latitudes, there should be only small differences between the mass fluxes implied by $\mathcal{V}$ and $\overline{v}^*$, associated with slightly different average heating contributions to the two budgets.

The wave activity diagnostics examined below in section 6 are calculated relative to a zonally symmetric, though possibly time-dependent, reference state. There is considerable freedom in choosing this reference state, but it is convenient to define the reference state to be a PV-conserving mass rearrangement of the actual state. This rearrangement property is equivalent to the requirement that $\mathcal{M}_{\rm ref}(Q)$ for the reference state be the same as $\mathcal{M}(Q)$ for the actual state. However, the zonal symmetry and rearrangement properties are still not sufficient to define the reference state uniquely. We can use the remaining freedom in the choice of reference state to demand that it also be balanced (with the same zonal mean lower boundary height $\overline{h_0}$ as the actual state) and therefore, in the absence of diabatic heating and friction, steady. Hence, determining the reference state amounts to finding a zonally symmetric, balanced state with a given distribution $\mathcal{M}(Q)$ or, equivalently, $\mathcal{C}(Q)$. Using the relations

$$\mathcal{M}_\phi = -2\pi a^2 \overline{h} \cos\phi \tag{29}$$

and

$$\mathcal{C} = (\overline{u} + \Omega a \cos\phi) 2\pi a \cos\phi \tag{30}$$

for a zonally symmetric state, the balance condition,

$$\left(f + \frac{\overline{u}\tan\phi}{a}\right)\overline{u} + g\frac{(\overline{h} + \overline{h_0})_\phi}{a} = 0, \tag{31}$$

can be converted into a nonlinear boundary value problem for $Q_{\rm ref}(\phi)$:

$$\frac{\mathcal{C}(Q_{\rm ref})^2 \sin\phi}{(2\pi)^2 a^3 \cos^3\phi} - \left(\frac{g}{2\pi a^3}\right)\left(\frac{\mathcal{M}(Q_{\rm ref})_\phi}{\cos\phi}\right)_\phi = \Omega^2 a \sin\phi\cos\phi - \frac{g}{a}\overline{h_0}_\phi, \tag{32}$$

with $Q_{\rm ref} = Q_{\rm max}$ at the North Pole and $Q_{\rm ref} = Q_{\rm min}$ at the South Pole. In fact, the nonlinearity here is quite weak, because $\mathcal{C}$ is dominated by the contribution from the earth's rotation, so that (32) can be solved efficiently by linearizing and using about three iterations of a tridiagonal linear solver (appendix A). Once $Q_{\rm ref}(\phi)$ has been found, $h_{\rm ref}(\phi)$ and $u_{\rm ref}(\phi)$ can be recovered from (29) and (30). We have been unable so far to find a proof that the solution of (32) is unique. In practice, however, we found no evidence of multiple solutions.

Although the reference state is defined principally for the purpose of calculating wave activity diagnostics, it is also interesting in its own right. It contains the same information as $\mathcal{M}(Q)$ or $\mathcal{C}(Q)$, and so it evolves only through diabatic heating, friction, and irreversible PV mixing, but it can be directly compared with Eulerian zonal mean diagnostics. Figures 4, 5, and 6 compare $u_{\rm ref}$, $h_{\rm ref}$, and $Q_{\rm ref}$ with $\overline{u}$, $\overline{h}$, and $\overline{Q}$. The timescale for the evolution of the reference state, like the timescale for the evolution of $\mathcal{M}$ and $\mathcal{C}$, is much longer than the timescale for the evolution of $\overline{u}$ and $\overline{h}$. The jet in the reference state becomes narrower, strengthens, and moves poleward, as does the zonal mean jet, but there are many differences of detail between the two. In both cases the subtropical zonal wind at about 30°N decreases over the first 20 days, but $\overline{u}$ reverses and becomes easterly whereas $u_{\rm ref}$ does not. Both $\overline{u}$ and $u_{\rm ref}$ show a gradual buildup of tropical easterlies, though $u_{\rm ref}$ is somewhat weaker. The reference depth $h_{\rm ref}$ gradually increases in mid- and high latitudes, but much less than $\overline{h}$. Similarly, the tropical decrease in $h_{\rm ref}$ is less than the tropical decrease in $\overline{h}$. The reference PV $Q_{\rm ref}$ shows the sharpening of the PV gradient at the vortex edge, the formation of a surf zone with a weakened PV gradient, and the strengthening PV gradient in the subtropics on the equatorward flank of the surf zone. In this case we can be sure that these features arise through irreversible PV mixing and diabatic heating. In contrast, $\overline{Q}$ appears to show a significantly wider surf zone; this must be at least partly an artifact of zonally averaging over a range of PV values when large-scale eddies are present.

5. Balance and stationary values of energy

An interesting property of the reference state defined in section 4 is that its energy is a stationary value. Specifically, its energy is unchanged to first order under small-amplitude PV-conserving mass rearrangements. A PV-conserving mass rearrangement is one that preserves

the mass distribution in PV space $\mathcal{M}(Q)$. In appendix B we show that, for flow governed by the shallow water equations, states with stationary values of energy under rearrangements that preserve $\mathcal{M}(Q)$ are balanced, steady states.

Several similar results have been derived previously. For example, it has been known since the time of Kelvin (e.g., Thomson 1910; see also Burton 1989) that in two-dimensional incompressible flow the state that maximizes the energy under vorticity-conserving parcel rearrangements is a steady state. Cullen et al. (1987) showed that, for a three-dimensional Boussinesq fluid, stationary values of energy under incompressible rearrangements conserving parcel potential temperature and absolute momentum correspond to states of hydrostatic and geostrophic balance. Vallis (1992) showed that, for shallow water flow, stationary values of the energy integral for the linearized equations subject to a given spatial distribution of the linearized PV correspond to states of geostrophic balance. He also showed that stationary values of the energy integral for the full equations subject to a given spatial distribution of the full PV correspond to states in a related type of balance, which he called ''extreme balance.'' A systematic survey of the kinds of balance obtained for stationary values of different forms of the energy integral and subject to different constraints would be of some theoretical interest. Possible applications of these ideas to weather and climate forecasting are discussed by Cullen (1997).

6. Finite-amplitude wave activity diagnostics

We now show how finite-amplitude wave activity diagnostics can be calculated with the aid of the contour integral diagnostics already discussed. We present wave activity diagnostics for our stratospheric wave-breaking simulation and compare the full wave activity flux with the EP flux, which is often used as an approximate diagnostic of wave activity flux.

Brunet and Haynes (1996) presented a finite-amplitude wave activity conservation law, derived using the method of Haynes (1988), for perturbations to a zonally symmetric reference state for shallow water flow on the sphere

$$\frac{\partial}{\partial t}\mathcal{A} + \boldsymbol{\nabla}\cdot\mathcal{F} = \mathcal{B} + \mathcal{D}, \tag{33}$$

where

$$\mathcal{A} = -\frac{h}{2\pi a}\int_{Q_{\mathrm{ref}}}^{Q}(Q-\tilde{Q})\frac{d}{d\tilde{Q}}m_{\mathrm{ref}}(\tilde{Q})\,d\tilde{Q} - h_e u_e\cos\phi, \tag{34}$$

(here we have chosen the opposite sign convention to Brunet and Haynes); $\mathcal{F}$ has components $\mathcal{F}^{(\lambda)}$ and $\mathcal{F}^{(\phi)}$ given by

$$\mathcal{F}^{(\lambda)} = u\mathcal{A} + \frac{1}{2}(v_e^2 - u_e^2)h_{\mathrm{ref}}\cos\phi - \frac{1}{2}gh_e^2\cos\phi, \tag{35}$$

$$\mathcal{F}^{(\phi)} = v\mathcal{A} - u_e v_e h_{\mathrm{ref}}\cos\phi; \tag{36}$$

$\mathcal{B}$ represents lower boundary forcing

$$\mathcal{B} = \frac{g}{a}h_e h_{0\lambda}; \tag{37}$$

and $\mathcal{D}$ represents sources and sinks due to irreversible mixing and diabatic and frictional processes. Subscript e indicates a perturbation from the reference state, and all terms in (33) are second order in perturbation quantities. In (34), $m_{\mathrm{ref}}(Q)$ is the mass in the reference state north of latitude $\phi_{\mathrm{ref}}(Q)$, where $\phi_{\mathrm{ref}}(Q)$ is the function inverse to $Q_{\mathrm{ref}}(\phi)$. We will assume that $Q_{\mathrm{ref}}(\phi)$ is monotonic, in which case $\phi_{\mathrm{ref}}(Q)$ and $m_{\mathrm{ref}}(Q)$ are single valued, and, moreover, $m_{\mathrm{ref}}(Q) = \mathcal{M}_{\mathrm{ref}}(Q)$.

Using (22) applied to the reference state allows the integration in (34) to be carried out analytically, giving

$$\mathcal{A} = -\frac{h}{2\pi a}[Q(\mathcal{M}_{\mathrm{ref}}(Q) - \mathcal{M}_{\mathrm{ref}}(Q_{\mathrm{ref}})) - (\mathcal{C}_{\mathrm{ref}}(Q) - \mathcal{C}_{\mathrm{ref}}(Q_{\mathrm{ref}}))] - h_e u_e\cos\phi. \tag{38}$$

For some purposes this expression may be more convenient than (34), since it does not involve integrals of integrals (remember that $\mathcal{M}$ and $\mathcal{C}$ themselves are integrals). Now it is convenient to choose the reference state to be a PV-conserving mass rearrangement of the actual state, so that $\mathcal{M}_{\mathrm{ref}}(Q) = \mathcal{M}(Q)$ and $\mathcal{C}_{\mathrm{ref}}(Q) = \mathcal{C}(Q)$. Then (38) reduces to

$$\mathcal{A} = -\frac{h}{2\pi a}[Q(\mathcal{M}(Q) - \mathcal{M}(Q_{\mathrm{ref}})) - (\mathcal{C}(Q) - \mathcal{C}(Q_{\mathrm{ref}}))] - h_e u_e\cos\phi. \tag{39}$$

This choice has the advantage that, even though the maximum and minimum values of Q may change with time, for example, through diabatic heating, Q and Q_{ref} have the same range of values so that the terms in (39) are always defined.

By construction, the reference state defined in section 4 and appendix A has monotonic $Q_{\mathrm{ref}}(\phi)$. Therefore we can make the step from (34) to (38) and (39). If $Q_{\mathrm{ref}}(\phi)$ were not monotonic then $m_{\mathrm{ref}}(Q)$ would be multivalued and Lagrangian information about parcels' initial positions would be needed to resolve the ambiguity in (34) (Killworth and McIntyre 1985; McIntyre and Shepherd 1987).

The expression for $\mathcal{A}$ may be thought of as consisting of a PV rearrangement or Rossby wave term, $\mathcal{A}_R = -(h/2\pi a)[Q(\mathcal{M}(Q) - \mathcal{M}(Q_{\mathrm{ref}})) - (\mathcal{C}(Q) - \mathcal{C}(Q_{\mathrm{ref}}))]$, and a gravity wave term, $\mathcal{A}_G = -h_e u_e\cos\phi$. When $m_{\mathrm{ref}}(Q)$ is single valued, so that dm_{ref}/dQ is negative definite, $\mathcal{A}_R$ is positive definite. The term $\mathcal{A}_G$, however, may be positive or negative. This is related to the fact that a breaking or dissipating Rossby wave generally exerts a westward force on the background flow, whereas breaking

or dissipating gravity waves can exert either eastward or westward forces.

The expression for the northward component of the wave activity flux $\mathcal{F}^{(\phi)}$ is made up of two terms. One term closely resembles the northward EP flux. This term dominates when the wave amplitude is small and can be thought of as representing wavelike propagation of wave activity. The other term represents advection of wave activity and can be important when the wave amplitude is large.

The reference states used in the wave activity calculations of Brunet and Haynes (1996) and Magnusdottir and Haynes (1996) were independent of time. If the reference state is time dependent, as it is here, then a wave activity conservation law still holds but the term $\mathcal{D}$ includes contributions from the time dependence of the reference state (Haynes 1988). However, our reference state is constructed so that its time dependence arises only through irreversible PV mixing and diabatic and frictional processes. Hence, $\mathcal{D}$ differs from zero only through irreversible PV mixing and diabatic and frictional processes, through their effect both on the reference state and directly on the eddies. The explicit expression for $\mathcal{D}$ is rather complicated and will not be considered here.

Figure 12 shows synoptic maps of wave activity on days 4, 14, and 44. Note that the contour interval in Fig. 12a is much smaller than in Figs. 12b and 12c. On day 4 the waves are still very weak and the prognostic fields have an essentially linear zonal wavenumber 1 pattern. This is reflected as a zonal wavenumber 2 pattern in the wave activity, which, as noted above, is second order in disturbance quantities for small-amplitude disturbances. Even at this early stage the wave activity is dominated by the PV rearrangement term, although the negative values of wave activity must come from the gravity wave term. On day 14 the wave activity has a qualitatively very different appearance. It is dominated even more than on day 4 by the PV rearrangement term, which has a roughly quadratic dependence on Q_e, and so it picks out and emphasizes the PV anomaly associated with the stronger of the two anticyclones A2. Similarly, on day 44 the wave activity picks out and emphasizes the PV anomalies associated with the anticyclones A4 and A6 and the tongues of high PV air T4 and T5.

At first glance the wave activity maps appear to convey little or no new information compared with maps of PV itself. However, they do suggest that in strongly nonlinear situations such as these the advection of wave activity (as indicated by the $\mathbf{v}\mathcal{A}$ term) is likely to be at least as important as the wavelike propagation mechanism (the other terms making up $\mathcal{F}$). This suggestion is confirmed by Figs. 13a and 13b, which compare the zonal mean northward wave activity flux with the northward EP flux. The patterns of positive and negative values of the two quantities are broadly similar, but there are differences in detail and in the numerical magnitudes that become more marked as time goes on. In particular, the displacement of the main vortex from the pole and the dramatic deceleration in $\overline{u}$ between days 50 and 55 is accompanied by an enormous poleward flux of wave activity at high latitudes but only a relatively weak poleward EP flux.

Figure 13c shows the zonal mean of the wave activity itself during the 60 days of the simulation. There is a gradual northward shift in the maximum wave activity, but little wave activity reaches polar latitudes until after day 50. The two periods of large wave activity, from about day 8 to about day 22 and from about day 28 onward correspond to the periods when $\overline{Q}$ has a weakened midlatitude gradient (Fig. 6); this is further confirmation that this weak $\overline{Q}$ gradient is associated with a zonal average over large-scale eddies rather than with small-scale PV mixing. The poleward penetration of the small, intense anticyclone A6 followed by its return equatorward around the larger anticyclone A4 (Fig. 2) is clearly visible in Fig. 13c as a narrow maximum in wave activity between days 39 and 42 moving poleward then equatorward.

7. Discussion

We have diagnosed a shallow water simulation of stratospheric wave breaking using Eulerian zonal means and using integrals within PV contours, and we have compared the kinds of information provided by the two kind of diagnostics. The contour integral diagnostics give a clear picture of the formation of sharp PV gradients at the vortex edge and in the subtropics and of PV mixing in the surf zone in between. This picture is much less ambiguous than that given by Eulerian zonal mean PV, for example, which does not distinguish between the presence of large-scale eddies on the one hand and genuine small-scale mixing of PV on the other. Also, the contour integral diagnostics evolve only through diabatic heating, friction, and irreversible PV mixing; they effectively filter out transient, reversible behavior. In contrast, the evolution of Eulerian zonal mean wind and depth is dominated by transient, reversible behavior, at least over timescales of a few days.

Comparison of the mass budget and the circulation budget for a PV contour implies a long-term balance between friction on the contour and heating within the contour: see Eq. (28) above. A possible application of (28), or its three-dimensional generalization, is to the diagnosis of middle-atmosphere wave drag from radiative heating calculations using analyzed temperature data; a major stumbling block in the development of gravity wave drag parameterizations for middle-atmosphere models has been the lack of reliable estimates of the actual gravity wave drag in the real atmosphere.

The contour integral diagnostics have been used to define a balanced, zonally symmetric reference state, whose evolution can be compared directly with that of $\overline{u}$, $\overline{h}$, and $\overline{Q}$. The evolution of the reference state, like

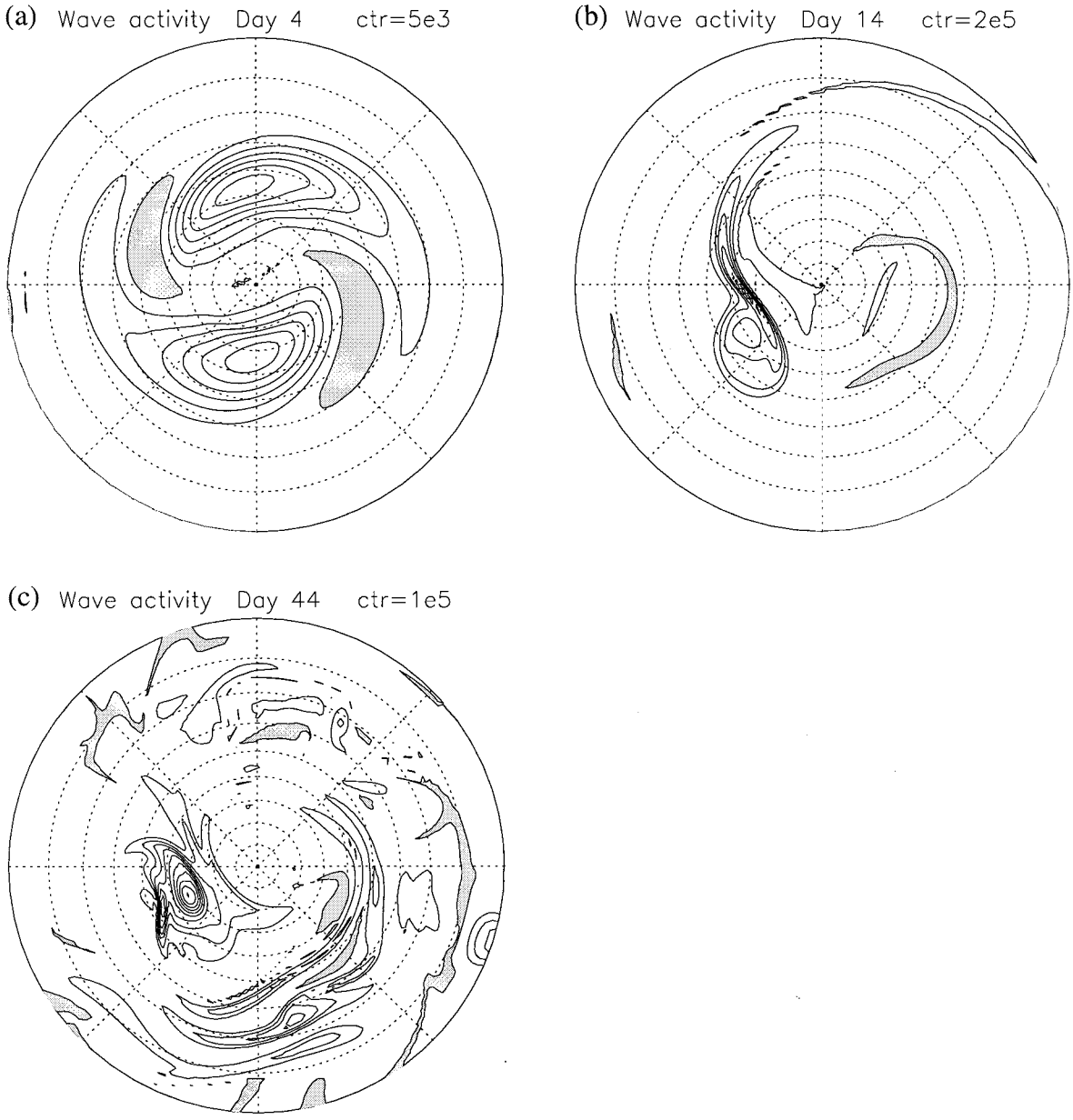

FIG. 12. Northern Hemisphere wave activity. (a) Day 4, contour interval 5×10^3 m^2 s^{-1}, (b) day 14, contour interval 2×10^5 m^2 s^{-1}, (c) day 44, contour interval 1×10^5 m^2 s^{-1}. Negative values are shaded.

that of the contour integrals on which it is based, occurs only through diabatic heating, friction, and irreversible mixing, and so filters out transient, reversible behavior. Note that although the reference state has the same $\mathcal{M}(Q)$ as the actual state, in general it has different total angular momentum. Thus, the reference state cannot be reached from the actual state by an angular momentum–conserving rearrangement. The difference in total angular momentum between the reference state and the actual state is equal to the global integral of the wave activity. This follows from the fact that the zonal mean of the lower boundary source of wave activity (37) is equal to the negative of the zonal mean lower boundary source of angular momentum. We can therefore imagine an adiabatic and frictionless, that is, PV conserving, thought experiment, beginning from the actual disturbed state, in which the lower boundary is specified so as to quickly remove all wave activity and thus force the flow to evolve to the corresponding reference state. The change in angular momentum during the thought experiment is equal and opposite to the change in wave activity. Note also that the diagnostic $\mathcal{C}(Q)$ is a measure of circulation, not of angular momentum, and that circulation and angular momentum diagnostics contain the

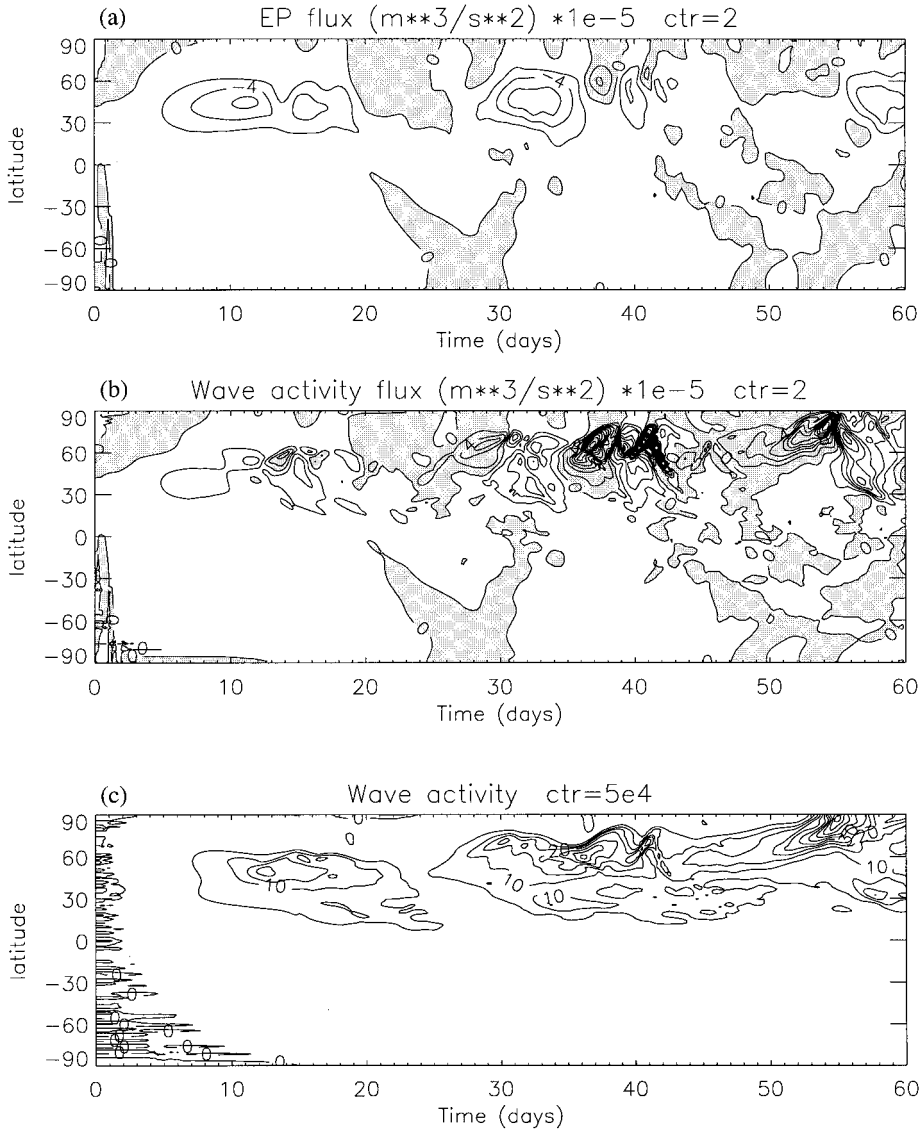

FIG. 13. Time–latitude plots of (a) EP flux and (b) zonal mean northward wave activity flux, contour interval 2×10^{-5} m^3 s^{-2}, positive values shaded, and (c) zonal mean wave activity, contour interval 5×10^4 m^2 s^{-1}.

same information only when PV contours are zonal. Thus, $\mathcal{C}(Q)$ is a quite different kind of diagnostic from $\overline{u}$ (or the generalized Lagrangian mean or modified Lagrangian mean of u) and, in particular, the nonacceleration theorem discussed by Nakamura (1995) and in section 4 above is a quite different kind of result from the nonacceleration theorems of Edmon et al. (1980) and Andrews (1983), for example.

The contour integrals can be used to diagnose the cross-PV-contour mass transport $\mathcal{V}$. This gives a picture of the meridional transport radically different from that given by the Eulerian mean $\overline{v}^*$. The cross-contour transport $\mathcal{V}$ is episodic and is directed predominantly into the surf zone from both sides, whereas $\overline{v}^*$ is generally persistent (to the extent that the lower boundary wave forcing is persistent) and is directed predominantly poleward. The cross-contour transport is intimately associated with PV mixing, and, therefore, with chemical mixing too. There is no such direct association between mixing and $\overline{v}^*$. The episodic nature of the mixing may have implications for estimates of chemical mixdown time in the stratospheric surf zone (Juckes and McIntyre 1987; Edouard et al. 1996; Haynes and Anglade 1997; Thuburn and Tan 1997).

The cross-contour mass transport can be used to attempt to quantify transport across various atmospheric ''transport barriers.'' For example, attempts to diagnose horizontal transport out of the winter stratospheric vor-

tex using analyzed winds have led to a wide range of estimated values (see, e.g., Sobel et al. 1997 and references therein). The extent to which the winter vortex acts like a ''containment vessel'' on the one hand versus a ''flowing processor'' on the other has an important bearing on the question of stratospheric ozone depletion. The results of section 4 suggest that about 45% of the air within contour (6) and about 35% of the air within contour (7) is mixed into the surf zone during the 60-day integration. As with other methods of estimating such transport, there is an ambiguity because the net transport is a residual of some outward and some inward transport. However, the fact that the transport occurs in a series of events, virtually all of which are outward, gives some (but not complete) confidence that virtually all of the transport is indeed outward. Advantages of using a model to estimate the transport are that errors due to noisy data are greatly reduced and that the PV and wind fields are guaranteed to be consistent. A disadvantage, of course, is that the simulation discussed here is highly idealized and may have transport characteristics significantly different from those of the real atmosphere. One possible approach to applying the cross-contour mass flux diagnostic to the quantification of transport in the real atmosphere might be to force a shallow water model (or three-dimensional model) using a lower boundary condition based on analyzed geopotential or Montgomery potential fields.

Another question to which these diagnostics might be applied is that of transport across the midlatitude tropopause. For example, many studies assume, implicitly or explicitly, that all such transport occurs in association with synoptic events such as tropopause folds and cutoff lows, and ignore the contribution from large-scale radiative ''seepage,'' even though the relative contributions from the two kinds of process are not known. If the tropopause is defined to be a PV surface, or a family of PV contours, $Q(\theta)$, on different surfaces of potential temperature θ, then it may be possible, using a combination of explicit radiative calculations and PV contour mass budget residuals, to estimate both the radiative seepage contribution and the contribution from irreversible mixing associated with synoptic events.

We have pointed out a relation between the two contour integrals $\mathcal{M}(Q)$ and $\mathcal{C}(Q)$: see Eq. (22) above. Note that although (22) appears to relate information about the mass and wind fields, it is not a statement of dynamical balance (e.g., geostrophic or gradient wind); it is an exact but purely kinematic relation. Indeed, the absence of any relation corresponding to geostrophic or gradient wind balance is one of the limitations of the contour integral diagnostics we have considered. (Although the reference state derived from the contour integrals is balanced, this is a condition we have chosen to impose on an artificially constructed state, not a property of the contour integrals themselves reflecting the actual fluid dynamics.) The relation (22) has been used to simplify the expression for the wave activity of finite-amplitude disturbances to the zonally symmetric reference state.

We have examined the finite-amplitude wave activity for our simulation and compared the wave activity flux with the EP flux. The EP flux is often used as a diagnostic of wave propagation, even though its use is strictly valid only for small-amplitude waves. There are large differences between the EP flux and the wave activity flux; in particular, typical magnitudes of the wave activity flux are 3–4 times as great as typical magnitudes of the EP flux, emphasizing the importance of the advection term $\mathbf{v}\mathcal{A}$ in the wave activity flux when the flow is dominated by large-amplitude interacting vortices. In view of these results, it is perhaps surprising how successful EP fluxes appear to have been as diagnostics of stratospheric sudden warmings, for example. One point to note is that $\boldsymbol{\nabla} \cdot \mathbf{F}$, and not $\boldsymbol{\nabla} \cdot \mathcal{F}$, appears in the TEM and isentropic coordinate mean zonal momentum equations and so is the relevant diagnostic for quantifying eddy driving of the zonal mean circulation. Thus, for example, we should indeed expect upward and poleward focusing of EP flux vectors to be associated with deceleration of the polar night jet. However, the success of the EP flux as a diagnostic goes beyond this. For example, concepts such as refractive indices and critical lines appear to be qualitatively useful predictors of the direction of the EP flux (e.g., Butchart et al. 1982; Palmer and Hsu 1983), even though neither the Wentzel–Kramers–Brillouin–Jeffreys (WKBJ) theory relating refractive indices and critical lines to wave propagation nor the relation between wave propagation and EP flux are valid for the strongly nonlinear flows that occur during stratospheric sudden warmings. Part of the explanation for the success of the EP flux as a diagnostic may be that the very largest scale wave components, zonal wavenumbers 1 and 2, may behave quasilinearly, even when smaller-scale motions are strongly nonlinear. Another part of the explanation may be that there is some profound reason, as yet not understood, why $\overline{v\mathcal{A}}$ should nearly always have the same sign as $F^{(\phi)}$, so that $F^{(\phi)}$ nearly always has the same sign as $\overline{\mathcal{F}^{(\phi)}}$ (Fig. 13). Another consideration is that results showing EP flux not to be a useful diagnostic may be regarded as less interesting and so be less likely to be published. Indeed, in our experience the EP flux often fluctuates rapidly and can be a poor indicator of the propagation of wave activity, while linear, WKBJ concepts such as refractive index can often be a poor predictor of the direction of the EP flux. Unfortunately, short of direct numerical simulation of the three-dimensional flow, there is currently no alternative, nonlinear theory available that could predict the amplitudes and trajectories of vortices and high and low PV tongues in and around the stratospheric surf zone.

Finally, we note that virtually all of the theoretical results and diagnostics presented in this paper can be extended straightforwardly to three-dimensional flow governed by the hydrostatic primitive equations. The

analogs of $\mathcal{M}$ and $\mathcal{C}$ involve integrals within PV contours on isentropic surfaces with h replaced by the isentropic density $\sigma = -(1/g)(\partial p/\partial\theta)$ (e.g., O'Neill and Pope 1993a,b; Nakamura 1995). The cross-contour mass flux $\mathcal{V}$ may still be diagnosed as a residual in the budget of $\mathcal{M}$. A zonally symmetric reference state may be constructed from $\mathcal{M}(\theta, Q)$ assuming hydrostatic and gradient wind balance. And the relation (22) between $\mathcal{M}$ and $\mathcal{C}$ still holds, allowing Haynes's (1988) expression for wave activity to be simplified. Furthermore, the model used in this study has been extended into a three-dimensional isentropic-coordinate PV-based primitive equation model (A. Gregory 1997, personal communication). In the near future we hope to be able to apply these kinds of diagnostics to three-dimensional simulations of the stratosphere using the new model.

Acknowledgments. We are grateful to R. J. Douglas and to A. R. Gregory for helpful discussions on the mathematical theory of rearrangements and on the use of the contour integral diagnostics, respectively, and to G. Magnusdottir and A. O'Neill for valuable comments on earlier drafts of this paper. JT was funded by the U.K. Natural Environment Research Council through the U.K. Universities Global Atmospheric Modelling Programme.

APPENDIX A

Efficient Calculation of the Reference State

We assume that, to begin with, $\mathcal{M}$ and $\mathcal{C}$ have been tabulated as functions of Q, as described in section 4. Equation (32) can be written

$$k^{(1)}\mathcal{C}^2 - k^{(2)}\left(\frac{\mathcal{M}_\phi}{\cos\phi}\right)_\phi = k^{(3)} \tag{A1}$$

where

$$k^{(1)} = \frac{\sin\phi}{(2\pi)^2 a^3 \cos^3\phi}, \tag{A2}$$

$$k^{(2)} = \frac{g}{2\pi a^3}, \tag{A3}$$

and

$$k^{(3)} = \Omega^2 a \sin\phi\cos\phi - \frac{g}{a}\overline{h_0}^{\phi}. \tag{A4}$$

This problem is to be solved subject to the boundary conditions that $\mathcal{M}$ at the North Pole should equal zero and at the South Pole should equal the total mass of fluid. Discretizing in latitude gives

$$k_j^{(1)}\mathcal{C}_j^2 - \frac{k^{(2)}}{\Delta\phi^2}\left(\frac{\mathcal{M}_{j+1} - \mathcal{M}_j}{\cos\phi_{j+1/2}} - \frac{\mathcal{M}_j - \mathcal{M}_{j-1}}{\cos\phi_{j-1/2}}\right) = k_j^{(3)}. \tag{A5}$$

The results presented in this paper used 129 regularly spaced latitudes. If Eq. (A5) were linear then we would have a tridiagonal system that could be solved directly by Gaussian elimination. In fact the nonlinearity is quite weak, so we can linearize (A5) and iterate toward the solution, solving a tridiagonal linear system at each iteration. Letting superscipt n stand for the number of iterations and using

$$(\mathcal{C}_j^{n+1})^2 \approx (\mathcal{C}_j^n)^2 + 2\mathcal{C}_j^n(\mathcal{M}_j^{n+1} - \mathcal{M}_j^n)\left(\frac{d}{d\mathcal{M}}\mathcal{C}\right)_j^n$$
$$= (\mathcal{C}_j^n)^2 + 2\mathcal{C}_j^n(\mathcal{M}_j^{n+1} - \mathcal{M}_j^n)Q_j^n, \tag{A6}$$

[from (22)], (A5) is replaced by

$$2k_j^{(1)}\mathcal{C}_j^n Q_j^n \mathcal{M}_j^{n+1} - \frac{k^{(2)}}{\Delta\phi^2}\left(\frac{\mathcal{M}_{j+1}^{n+1} - \mathcal{M}_j^{n+1}}{\cos\phi_{j+1/2}} - \frac{\mathcal{M}_j^{n+1} - \mathcal{M}_{j-1}^{n+1}}{\cos\phi_{j-1/2}}\right)$$
$$= k_j^{(3)} - k_j^{(1)}(\mathcal{C}_j^n)^2 + 2k_j^{(1)}\mathcal{C}_j^n Q_j^n \mathcal{M}_j^n. \tag{A7}$$

This is now a tridiagonal linear system for $\mathcal{M}^{n+1}$ in terms of known quantities. An initial guess $\mathcal{M}(\phi) = \frac{1}{2}(1 - \sin\phi)\mathcal{M}_{\max}$ is made, corresponding to uniform fluid depth. Given a set of values, $\mathcal{M}_j^n$, after n iterations, the corresponding values of Q_j^n are found by searching and interpolating the tabulated $\mathcal{M}(Q)$, and then the corresponding values of $\mathcal{C}_j^n$ are found by looking up and interpolating the tabulated $\mathcal{C}(Q)$. This determines the right-hand side and all of the coefficients in (A7), enabling $\mathcal{M}_j^{n+1}$ to be found. For practical purposes the calculation converges in about three iterations. The fact that $\mathcal{M}$ is a monotonic function of Q guarantees that the Q_{ref} calculated in this way is a monotonic function of ϕ as long as the fluid depth remains positive.

APPENDIX B

Stationary Values of Energy under PV-Conserving Mass Rearrangements

Consider the problem of finding stationary values of of the energy

$$E = \int_{\text{sphere}} \frac{1}{2}(h|\mathbf{v}|^2 + g(h + h_0)^2)\, dA \tag{B1}$$

subject to the constraint of given $\mathcal{M}(Q)$ for all Q. This problem can be solved by seeking stationary values of another quantity

$$S = E - \int_{Q_{\min}}^{Q_{\max}} \mu(Q)\mathcal{M}(Q)\, dQ, \tag{B2}$$

where $\mu(Q)$ is a continuous set of Lagrange multipliers, which may be determined later by imposing the constraint. We seek states that give $\delta S = 0$ for independent small variations δu, δv, and δh. First note that

$$\delta\mathcal{M}(Q) = \int_{\text{PV}>Q} \delta h\, dA + \oint_{\text{PV}=Q} \frac{\delta Q}{|\nabla Q|} h\, dl \tag{B3}$$

and

$$\delta Q = \frac{1}{h}\left[\frac{(\delta v)_\lambda}{a\cos\phi} - \frac{(\delta u\cos\phi)_\phi}{a\cos\phi} - Q\delta h\right]. \quad \text{(B4)}$$

Taking the variation of (B2), retaining only first order terms, and substituting from (B3) and (B4) gives

$$\delta S = \int_{\text{sphere}} \left[hu\delta u + hv\delta v + \left(\frac{1}{2}\mathbf{v}^2 + g(h + h_0)\right)\delta h\right] dA - \int_{Q_{\min}}^{Q_{\max}} \mu(Q)\left\{\int_{\text{PV}>Q} \delta h\, dA + \oint_{\text{PV}=Q} \frac{1}{|\nabla Q|}\left[\frac{(\delta v)_\lambda}{a\cos\phi} - \frac{(\delta u\cos\phi)_\phi}{a\cos\phi} - Q\delta h\right] dl\right\} dQ. \quad \text{(B5)}$$

Now use

$$\int_{Q_{\min}}^{Q_{\max}} \left(\oint_{\text{PV}=Q} \frac{1}{|\nabla Q|}(\,\cdot\,)\, dl\right) dQ = \int_{\text{sphere}} (\,\cdot\,)\, dA \quad \text{(B6)}$$

and reorder the first double integral in (B5),

$$\int_{Q_{\min}}^{Q_{\max}} \mu(Q)\left\{\int_{\text{PV}>Q} \delta h\, dA\right\} dQ = \int_{\text{sphere}} \delta h\left\{\int_{Q_{\min}}^{Q} \mu(\tilde{Q})\, d\tilde{Q}\right\} dA, \quad \text{(B7)}$$

to simplify (B5):

$$\delta S = \int_{\text{sphere}} \left\{hu\delta u + hv\delta v + \left(\frac{1}{2}|\mathbf{v}|^2 + g(h + h_0)\right)\delta h - \mu(Q)\left[\frac{(\delta v)_\lambda}{a\cos\phi} - \frac{(\delta u\cos\phi)_\phi}{a\cos\phi} - Q\delta h\right] - \delta h\int_{Q_{\min}}^{Q} \mu(\tilde{Q})\, d\tilde{Q}\right\} dA. \quad \text{(B8)}$$

Then integrating by parts and using periodicity in λ and $\cos\phi = 0$ at the poles gives

$$\delta S = \int_{\text{sphere}} \left\{\left[\frac{1}{2}|\mathbf{v}|^2 + g(h + h_0) + Q\mu(Q) - \int_{Q_{\min}}^{Q} \mu(\tilde{Q})\, d\tilde{Q}\right]\delta h + \left[hu - \frac{(\mu(Q))_\phi}{a}\right]\delta u + \left[hv + \frac{(\mu(Q))_\lambda}{a\cos\phi}\right]\delta v\right\} dA. \quad \text{(B9)}$$

Hence in order for δS to vanish for arbitrary δu, δv, and δh we require

$$\frac{1}{2}|\mathbf{v}|^2 + g(h + h_0) + Q\mu(Q) - \int_{Q_{\min}}^{Q} \mu(\tilde{Q})\, d\tilde{Q} = 0, \quad \text{(B10)}$$

$$hu - \frac{(\mu(Q))_\phi}{a} = 0, \quad \text{(B11)}$$

and

$$hv + \frac{(\mu(Q))_\lambda}{a\cos\phi} = 0. \quad \text{(B12)}$$

It is now clear from (B11) and (B12) that the Lagrange multiplier $\mu(Q)$ plays the role of a mass streamfunction, so that $\nabla\cdot(h\mathbf{v}) = 0$, implying [from (1) with $H = 0$],

$$h_t = 0. \quad \text{(B13)}$$

Furthermore, the mass stream function is a function only of Q, so that $h\mathbf{v}\cdot\nabla Q = 0$, implying [from (2) with $\mathbf{X} = 0$ and (B13)]

$$Q_t = 0. \quad \text{(B14)}$$

Finally, taking the gradient of (B10) gives

$$\nabla\left(\frac{1}{2}|\mathbf{v}|^2 + g(h + h_0)\right) + Q\nabla\mu = 0, \quad \text{(B15)}$$

and substituting from (B11) and (B12) and taking the divergence gives

$$\nabla\cdot\left[\nabla\left(\frac{1}{2}|\mathbf{v}|^2 + g(h + h_0)\right) + hQ\hat{\mathbf{k}}\times\mathbf{v}\right] = 0, \quad \text{(B16)}$$

implying [from (3) with $\mathbf{X} = 0$]

$$\delta_t = 0. \quad \text{(B17)}$$

Together (B13), (B14), and (B17) imply that the flow is steady and balanced.

REFERENCES

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

——, and M. E. McIntyre, 1976: Planetary waves in horizontal and vertical shear: The generalized Eliassen–Palm relation and the mean zonal acceleration. *J. Atmos. Sci.,* **33,** 2031–2048.

——, and ——, 1978a: Generalized Eliassen–Palm and Charney–Drazin theorems for waves on axisymmetric mean flows in compressible atmospheres. *J. Atmos. Sci.,* **35,** 175–185.

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

——, and ——, 1978c: On wave-action and its relatives. *J. Fluid Mech.,* **89,** 647–664.

——, J. R. Holton, and C. B. Leovy, 1987: *Middle Atmosphere Dynamics.* Academic Press, 489 pp.

Bates, J. R., and Y. Li, 1997: Simulation of stratospheric vortex erosion using three different global shallow water numerical models.

Numerical Methods for Atmospheric and Oceanic Modelling—The André Robert Memorial Volume, C. Lin, R. Laprise, and H. Ritchie, Eds., NRC Research Press, 55–74.

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

Burton, G. R., 1989: Rearrangements of functions, saddle points and uncountable families of steady configurations for a vortex. *Acta Math.,* **163,** 291–309.

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

——, S. A. Clough, T. N. Palmer, and P. J. Trevelyan, 1982: Simulations of an observed stratospheric warming with quasigeostrophic refractive index as a model diagnostic. *Quart. J. Roy. Meteor. Soc.,* **108,** 475–502.

Cullen, M. J. P., 1997: New mathematical developments in atmosphere and ocean dynamics, and their application to computer simulation. U.K. Meteorological Office Forecasting Research Division Scientific Paper 48, 61 pp. [Available from U.K. Meteorological Office, London Road, Bracknell, Berkshire RG12 2SZ, United Kingdom.]

——, J. Norbury, R. J. Purser, and G. J. Shutts, 1987: Modelling the quasi-equilibrium dynamics of the atmosphere. *Quart. J. Roy. Meteor. Soc.,* **113,** 735–757.

Dewan, E. M., 1981: Turbulent vertical transport due to thin intermittent mixing layers in the stratosphere and other stable fluids. *Science*, **211,** 1041–1042.

Dunkerton, T. J., 1978: On the mean meridional mass motions of the stratosphere and mesosphere. *J. Atmos. Sci.,* **35,** 2325–2333.

——, C.-P. F. Hsu, and M. E. McIntyre, 1981: Some Eulerian and Lagrangian diagnostics for a model stratospheric warming. *J. Atmos. Sci.,* **38,** 819–843.

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

Edouard, S., B. Legras, and V. Zeitlin, 1996: The effect of dynamical mixing in a simple model of the ozone hole. *J. Geophys. Res.,* **101,** 16 711–16 788.

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 W. E. Ward, 1993: The effect of realistic radiative transfer on potential vorticity structures, including the influence of background shear and strain. *J. Atmos. Sci.,* **50,** 3431–3453.

——, and J. Anglade, 1997: The vertical-scale cascade of atmospheric tracers due to large-scale differential advection. *J. Atmos. Sci.,* **54,** 1121–1136.

Hoskins, B. J., 1991: Towards a PV-θ view of the general circulation. *Tellus*, **43AB,** 27–35.

Juckes, M. N., 1989: A shallow water model of the winter stratosphere. *J. Atmos. Sci.,* **46,** 2934–2955.

——, and M. E. McIntyre, 1987: A high resolution, one-layer model of breaking planetary waves in the stratosphere. *Nature*, **328,** 590–596.

Killworth, P. D., and M. E. McIntyre, 1985: Do Rossby-wave critical layers absorb, reflect, or over-reflect? *J. Fluid Mech.,* **161,** 449–492.

Lary, D. J., M. P. Chipperfield, W. A. Norton, and L. P. Riishøgaard, 1995: Three-dimensional tracer initialization and general diagnostics using equivalent PV latitude-potential temperature coordinates. *Quart. J. Roy. Meteor. Soc.,* **121,** 187–210.

Magnusdottir, G., and P. H. Haynes, 1996: Wave activity diagnostics applied to baroclinic wave life cycles. *J. Atmos. Sci.,* **53,** 2317–2353.

McIntyre, M. E., 1980a: Towards a Lagrangian-mean description of stratospheric circulations and chemical transport. *Philos. Trans. Roy. Soc. London,* **A296,** 129–148.

——, 1980b: An introduction to the generalized Lagrangian-mean description of wave, mean-flow interaction. *Pure Appl. Geophys.,* **118,** 152–176.

——, 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.

Nakamura, N., 1995: Modified Lagrangian-mean diagnostics of the stratospheric polar vortices. Part I: Formulation and analysis of GFDL SKYHI GCM. *J. Atmos. Sci.,* **52,** 2096–2108.

Norton, W. A., 1994: Breaking Rossby waves in a model stratosphere diagnosed by a vortex-following coordinate system and a technique for advecting material contours. *J. Atmos. Sci.,* **51,** 654–673.

O'Neill, A., and V. D. Pope, 1993a: The coupling between radiation and dynamics in the stratosphere. *Adv. Space Res.,* **13,** 351–358.

——, and ——, 1993b: The effects of resolution on diagnostic calculations of coupling between radiation and dynamics. *Adv. Space Res.,* **13,** 359–362.

Palmer, T. N., and C.-P. F. Hsu, 1983: Stratospheric sudden coolings and the role of nonlinear wave interactions in preconditioning the circumpolar flow. *J. Atmos. Sci.,* **40,** 909–928.

Salby, M. L., D. O'Sullivan, R. R. Garcia, and P. Callaghan, 1990: Air motions accompanying the development of a planetary wave critical layer. *J. Atmos. Sci.,* **47,** 1179–1204.

Sobel, A. H., R. A. Plumb, and D. W. Waugh, 1997: Methods of calculating transport across the polar vortex edge. *J. Atmos. Sci.,* **54,** 2241–2260.

Thomson, W., 1910: Maximum and minimum energy in vortex motion. *Mathematical and Physical Papers.* Vol. 4. Cambridge University Press, 172–183.

Thuburn, J., 1995: Dissipation and cascades to small scales in numerical models using a shape-preserving advection scheme. *Mon. Wea. Rev.,* **123,** 1888–1903.

——, 1997: A PV-based shallow water model on a hexagonal-icosahedral grid. *Mon. Wea. Rev.,* **125,** 2328–2347.

——, and D. G. H. Tan, 1997: A parameterization of mixdown time for atmospheric chemicals. *J. Geophys. Res.,* **102,** 13 037–13 049.

Vallis, G. K., 1992: Mechanisms and parameterizations of geostrophic adjustment and a variational approach to balanced flow. *J. Atmos. Sci.,* **49,** 1144–1160.