The DOI for this manuscript is doi: 10.1175/JAS-D-12-0228.1

The final published version of this manuscript will replace the preliminary version at the above DOI once it is available.

If you would like to cite this EOR in a separate work, please use the following full citation:

Methven, J., 2013: Wave Activity for Large Amplitude Disturbances Described by the Primitive Equations on the Sphere. J. Atmos. Sci. doi:10.1175/JAS-D-12-0228.1, in press.

Wave Activity for Large Amplitude Disturbances Described by the Primitive Equations on the Sphere

JOHN METHVEN, *

Department of Meteorology, University of Reading, UK

* *Corresponding author address:* John Methven, Department of Meteorology, University of Reading, Earley Gate, Reading, RG6 6BB, UK.

E-mail: J.Methven@reading.ac.uk

ABSTRACT

Pseudomomentum and pseudoenergy are both measures of wave activity for disturbances in a fluid, relative to a notional background state. Together they give information on the propagation, growth and decay of disturbances. Wave activity conservation laws are most readily derived for the primitive equations on the sphere by using isentropic coordinates. However, the intersection of isentropic surfaces with the ground (and associated potential temperature anomalies) are a crucial aspect of baroclinic wave evolution. A new expression is derived for pseudoenergy that is valid for large amplitude disturbances spanning isentropic layers that may intersect the ground. The pseudoenergy of small amplitude disturbances is also obtained by linearising about a zonally symmetric background state. The new expression generalises previous pseudoenergy results for quasi-geostrophic disturbances on the β-plane and complements existing large amplitude results for pseudomomentum.

The pseudomomentum and pseudoenergy diagnostics are applied to an extended winter from ERA-Interim data. The time series identify distinct phenomena such as a baroclinic wave life cycle where the wave activity in boundary potential temperature saturates nonlinearly almost two days before the peak in wave activity near the tropopause. The coherent zonal propagation speed of disturbances at tropopause level, including distinct eastward, westward and stationary phases, is shown to be dictated by the ratio of total hemispheric pseudoenergy to pseudomomentum. Variations in the lower boundary contribution to pseudoenergy dominate changes in propagation speed; phases of westward progression are associated with stronger boundary potential temperature perturbations.

1. Introduction

Wave activity is a measure of the amplitude of the difference between any flow and a suitable background flow. It is defined to be second order in disturbance quantities so that it represents an amplitude and it is also globally conserved for adiabatic and frictionless flows. Wave activity is the basis of most wave-mean flow interaction theory (Bühler, 2009) and has led to important concepts such as the non-acceleration theorem of Charney and Stern (1961), expressing the inability of steady, conservative waves to alter the zonal mean zonal flow, and its many generalisations subsequently (Andrews et al., 1987). Wave activity theorems are also central to the theory of wave instability on shear flows (Bretherton, 1966b).

Solomon and Nakamura (2012) described three different forms of wave activity and their relationship. The first type are Eulerian measures of wave activity, evaluated at each point in physical coordinates based on deviations of the full flow from a background state. If the background is defined using the Eulerian zonal mean of the full flow, as in Charney and Stern (1961), the global conservation law is not respected exactly at large amplitude. However, McIntyre and Shepherd (1987) formulated a general recipe to construct Eulerian measures of wave activity that are conserved exactly at large-amplitude when measured relative to a zonally symmetric background state that is a solution of the governing fluid equations. It is possible to specify a wave activity density and flux at every point in physical space using their method. The second type are Lagrangian measures based on averaging quantities over selected material volumes and using their centre of mass as a coordinate. The resulting Generalised Lagrangian Mean theory, first obtained by Andrews and McIntyre (1978), has an exact wave activity conservation law but becomes problematic as material surfaces are increasingly distorted by stretching and folding associated with chaotic advection. The third type, introduced as A^* by Nakamura and Solomon (2010), replaces material contours with potential vorticity (PV) contours and uses these to calculate deviations from a Modified Lagrangian Mean (MLM) background state, as defined by McIntyre (1980). The MLM background state is the zonally symmetric re-arrangement of the full flow obtained by preserving the mass and circulation of volumes sandwiched between two isentropic surfaces where PV exceeds some value Q (for all θ and Q). The *equivalent latitude* of any wavy PV contour is defined as the latitude occupied by the corresponding PV contour in the MLM state. The wave activity A^* is defined in equivalent latitude space and has an exact conservation law like the GLM wave activity. However, since non-conservative processes eventually limit the finescales in the PV distribution it is possible to evaluate A^* for chaotic flows where it would eventually not be possible to follow the material contours necessary to calculate the GLM wave activity. A^* satisfies a non-acceleration theorem for the Eulerian zonal mean flow. However, wave activity density cannot be evaluated at every location in physical space - it is defined in the PV-θ coordinates of the MLM background state.

Other forms of wave activity for large-amplitude disturbances have been derived previously by considering different background states. For example, Tanaka et al. (2004) have formulated a wave activity (pseudomomentum) flux which is valid for large-amplitude disturbances to the primitive equations and makes an attractive separation between the vertical flux associated with form drag over corrugated isentropic surfaces and those associated with eddy diabatic mixing. This theory makes use of the Eulerian zonal mean of pressure on isentropic surfaces as a vertical coordinate and the background state is defined in terms of the mass-weighted isentropic zonal mean state (Iwasaki, 1989).

The approach taken here will be to develop the theory of Eulerian wave activity measures, but evaluate disturbances relative to the MLM background state. The MLM state is in balance and an exact solution to the primitive equations without eddy forcing. As will be seen below, these wave activity measures also relate to the displacement of PV contours from their position in the background state but the disturbances are evaluated in physical space rather than equivalent latitude.

A crucial aspect in the definition of wave activity density, A, is that it should obey a local conservation law:

$$\frac{\partial A}{\partial t} + \frac{1}{a\cos\phi}\frac{\partial F^{(\lambda)}}{\partial \lambda} + \frac{1}{a\cos\phi}\frac{\partial}{\partial \phi}\left(F^{(\phi)}\cos\phi\right) + \frac{\partial F^{(\theta)}}{\partial \theta} = S \tag{1}$$

where $(F^{(\lambda)}, F^{(\phi)}, F^{(\theta)})$ are the components of wave activity flux in isentropic spherical coordinates (λ is longitude, ϕ is latitude, θ is potential temperature and a is Earth's radius). S denotes non-conservative effects including diabatic and frictional processes. The global integral of wave activity is conserved if $S = 0$ and there is no flux across the boundaries of the integration domain.

Wave activity conservation laws relate to conserved properties of the full flow, for example energy or zonal angular momentum. However, these properties are not conserved by the perturbation alone because there is in general "exchange" between the background state and perturbation. In addition to the usual invariants such as energy and angular momentum, any function of θ and potential vorticity (PV) is globally conserved for the full flow, since these two quantities are conserved following all fluid parcels if the flow is adiabatic and frictionless. This family of additional invariants are called Casimirs. A systematic approach to finding wave activity conservation laws (McIntyre and Shepherd, 1987) is to combine energy or angular momentum with a Casimir that is chosen to obtain a disturbance quantity that is at least second order and globally conserved.

The definition of the background state is vital to the existence of a wave activity conservation law at finite perturbation amplitude. It is essential to describe the background state as a function of PV and θ in order to use the Casimir method. If the background state is also zonally symmetric, the pseudomomentum conservation law is obtained by the angular momentum-Casimir method. If the background state is steady (time symmetric) the pseudoenergy conservation law is obtained by the energy-Casimir method.

Bretherton (1966b) was the first to point out that growth of normal mode disturbances on a shear flow requires that the normal mode structure has zero global pseudomomentum (otherwise its pseudomomentum would increase with mode amplitude). This arises from cancellation between positive wave activity focussed where the background state meridional PV gradient is positive and negative wave activity where the PV gradient is negative. In the case of baroclinic instability, the negative wave activity is associated with potential temperature perturbations along the lower boundary. Bretherton (1966a) described baroclinic growth in a 2-layer quasi-geostrophic model in terms of counter-propagating Rossby waves which have equal and opposite pseudomomentum. This result has been generalised to any zonal jet (Heifetz et al., 2004) and the primitive equations on the sphere (Methven et al., 2005a). The phase propagation of the Rossby wave components depends on the ratio of their pseudoenergy to pseudomomentum, taking into account the boundary terms. However, these theories consider only small amplitude waves. New theory is needed for large amplitude disturbances, taking into account potential temperature perturbations along the lower boundary.

Brunet (1994) was the first to use the ratio of pseudoenergy and pseudomomentum to define the phase speed of structures obtained from the statistics of atmospheric analysis data. The technique he developed obtains Empirical Normal Modes as structures emerging from an eigen value decomposition of the data using pseudomomentum as a norm of disturbances. His initial work applied a shallow water form of wave activity to PV data on the 315K surface. Zadra et al. (2002) extended this technique to data on 16 isentropic levels using the full primitive equation wave activity. In both cases, the boundary terms in pseudoenergy and pseudomomentum were neglected and a small amplitude form of pseudoenergy was used. The primary purpose of this paper is to consider the ramifications of wave activity conservation for the zonal propagation of disturbances when including new theory relating to large amplitude disturbances with boundary wave activity.

The novel theoretical results of this paper relate to pseudoenergy and terms associated with the intersection of isentropic layers with the ground. However, the methodology is illustrated by deriving pseudomomentum results (which have already been published in similar forms). Section 2a applies the Casimir technique to derive pseudomomentum valid for large amplitude disturbances described by the primitive equations on the sphere. The result is essentially the same as Haynes (1988) but including a method to simplify the evaluation of wave activity using mass and circulation integrals, introduced in the shallow water context by Thuburn and Lagneau (1999). Section 2b considers the problem of evaluating the pseudomomentum integral for isentropic layers that intersect the ground. The presentation is brief, following Magnusdottir and Haynes (1996). Section 2c illustrates the procedure to derive wave activity in the limit of small disturbance amplitude. The Haynes (1988) result for pseudoenergy density valid at large amplitude is re-derived in Section 3a, as a necessary step towards the new result for integral pseudoenergy in Section 3b. The small amplitude limit of pseudoenergy is derived in Section 3c.

Many studies involving wave activity have been theoretical, applied to idealised models or applied to atmospheric data with approximations (such as small amplitude or quasi-geostrophic expressions). Nakamura and Solomon (2011) is the first study applying wave activity calculations valid at large amplitude to study wave-mean flow interaction throughout the atmosphere (from the ground to stratopause) using atmospheric analyses. They used the A^* measure of pseudomomentum rather than the "Casimir type" evaluated in physical space. Here, the large amplitude expressions for pseudoenergy (energy-Casimir) and pseudomomentum (zonal angular momentum-Casimir) are applied to re-analysis data in Section 4. Conclusions are obtained regarding the link between the integral conservation properties and the coherent zonal propagation of disturbances at tropopause level.

2. Pseudomomentum conservation

a. Pseudomomentum density for large amplitude disturbances

Specific zonal angular momentum (divided by the Earth's radius, a) at a point on the sphere rotating at rate Ω is:

$$Z = (u + a\Omega \cos\phi) \cos\phi \tag{2}$$

Following McIntyre and Shepherd (1987) and Haynes (1988), the pseudo-(angular)momentum density is defined by:

$$P(\lambda, \phi, \theta, t) = -r(Z + C) + r_o(Z_o + C_o). \tag{3}$$

$C(q, \theta)$ is called a Casimir density and is a function of PV and potential temperature alone. Ertel PV under the hydrostatic approximation is given by $q = \zeta/r$ where r is the pseudo-density in isentropic coordinates and ζ is the vertical component of absolute vorticity evaluated taking derivatives of velocity components along isentropic surfaces. The notation C_o means $C(q_o, \theta)$ where $q_o(\phi, \theta, t)$ denotes the background state PV and the perturbation $q_e = q - q_o$ is defined as the difference between the full PV and the background state at a point on a given isentropic surface[1]. Since Z and C are globally conserved, so is P and it must obey a conservation law where A is replaced by P in (1). The aim is to choose C so that P is second order in disturbance quantities.

Taylor expansion of the Casimir density can be written:

$$C = C_o + \left(\frac{\partial C}{\partial q}\right)_o q_e + C_2(q_o, q_e, \theta) \tag{4}$$

where $\left(\frac{\partial C}{\partial q}\right)_o$ means the functional derivative of the Casimir at constant θ, evaluated at the background state PV value q_o. C_2 is the residual which would include the second and all higher order terms in a series expansion. An exact integral form for C_2 is given later. Writing (3) in terms of background state and perturbation quantities:

$$\begin{aligned} P &= -rC_2 - r_e u_e \cos\phi \\ &\quad -r_o u_e \cos\phi - \left(\frac{\partial C}{\partial q}\right)_o \zeta_e - r_e \left\{ Z_o + C_o - q_o \left(\frac{\partial C}{\partial q}\right)_o \right\} \end{aligned} \tag{5}$$

where the identity $rq_e = \zeta_e - r_e q_o$ has been used. Expressing $\zeta_e = (1/a\cos\phi)\partial v_e/\partial\lambda -$

[1]Note that P is positive where the meridional PV gradient is positive – see (22). Haynes (1988) and Magnusdottir and Haynes (1996) used the opposite sign for P.

$(1/a\cos\phi)\partial(u_e\cos\phi)/\partial\phi$ and rearranging gives:

$$
\begin{aligned}
P &= -rC_2 - r_e u_e \cos\phi \\
&\quad -\frac{1}{a\cos\phi}\frac{\partial}{\partial\lambda}\left\{v_e\left(\frac{\partial C}{\partial q}\right)_o\right\} + \frac{1}{a\cos\phi}\frac{\partial}{\partial\phi}\left\{u_e\left(\frac{\partial C}{\partial q}\right)_o\cos\phi\right\} \\
&\quad -u_e\left\{r_o\cos\phi + \frac{1}{a}\frac{\partial}{\partial\phi}\left(\frac{\partial C}{\partial q}\right)_o\right\} - r_e\left\{Z_o + C_o - q_o\left(\frac{\partial C}{\partial q}\right)_o\right\}
\end{aligned}
\tag{6}
$$

The top line is second order (or higher) and the next line is expressed as a horizontal divergence. Therefore in order to make the global integral of P a second order quantity the terms in the last line must be zero, giving two relations defining the Casimir density:

$$
\frac{\partial}{\partial\phi}\left(\frac{\partial C}{\partial q}\right)_o = -ar_o\cos\phi \tag{7}
$$

$$
Z_o + C_o - q_o\left(\frac{\partial C}{\partial q}\right)_o = 0 \tag{8}
$$

Integration of the first equality gives:

$$
\begin{aligned}
\left(\frac{\partial C}{\partial q}\right)_o &= -a\int_0^{\phi} r_o(\tilde{\phi},\theta,t)\cos\tilde{\phi}\,d\tilde{\phi} \\
&= \frac{1}{2\pi a}\left\{\mathcal{M}(Q,\theta) - \mathcal{M}_s(\theta)\right\}
\end{aligned}
\tag{9}
$$

where $\mathcal{M}(Q,\theta)$ is the integral of mass across an isentropic layer

$$
\mathcal{M}(Q,\theta) = 2\pi a^2\int_{\phi(Q)}^{\pi/2} r_o\cos\tilde{\phi}\,d\tilde{\phi} \tag{10}
$$

and $\mathcal{M}_s(\theta)$ is total mass of the isentropic shell in the Northern Hemisphere. Here it is assumed that the background state is zonally symmetric with PV varying monotonically along isentropic surfaces, so that each latitude, ϕ, maps to a unique PV value $Q = q_o(\phi,\theta)$.

The wave activity is simpler to evaluate if the background state is identified with the modified Lagrangian mean (McIntyre, 1980). The MLM state is defined as an adiabatic rearrangement of the 3D flow (at any instant) to obtain a zonally symmetric state with the same mass and circulation integrals as evaluated from the 3D state:

$$
\mathcal{M}(Q,\theta) = \int\!\!\int_{q\geq Q} ra^2 d\lambda\cos\phi\,d\phi\ ;\quad \mathcal{C}(Q,\theta) = \int\!\!\int_{q\geq Q} rqa^2 d\lambda\cos\phi\,d\phi \tag{11}
$$

where the double integral spans high PV regions enclosed by the disturbed contours defined by $q = Q$. For adiabatic, frictionless flow both these integrals are conserved (for all Q, θ) owing to mass continuity and Kelvin's circulation theorem. This in turn implies that the equivalent latitudes[2] of the PV contours defining the MLM state cannot change: the state is steady. The final step is to note an explicit expression for C_2 valid at arbitrary perturbation amplitude:

$$C_2(q_o, q_e, \theta) = \int_0^{q_e} (q_e - \tilde{q}) \frac{\partial^2 C}{\partial \tilde{q}^2}(q_o + \tilde{q}, \theta)\, d\tilde{q} = C - C_o - q_e \left(\frac{\partial C}{\partial q}\right)_o \tag{12}$$

which can be verified using integration by parts. Thuburn and Lagneau (1999) simplified this expression by performing the integration over PV values analytically:

$$\begin{aligned} 2\pi a C_2 &= \int_0^{q_e} (q_e - \tilde{q}) \frac{\partial \mathcal{M}}{\partial \tilde{q}}(q_o + \tilde{q}, \theta)\, d\tilde{q} \\ &= \int_{q_o}^{q_o+q_e} (q_e + q_o - \eta) \frac{\partial \mathcal{M}}{\partial \eta}(\eta, \theta)\, d\eta \\ &= q \left[\mathcal{M}\right]_{q_o}^{q} - \left[\mathcal{C}\right]_{q_o}^{q} \end{aligned} \tag{13}$$

where the first step uses (9), the second step changes integration variable to $\eta = q_o + \tilde{q}$ and the last step uses the result:

$$Q \frac{\partial \mathcal{M}}{\partial q}(Q) = \frac{\partial \mathcal{C}}{\partial q}(Q) \tag{14}$$

relating the variation of mass and circulation with PV value along isentropic surfaces.

b. *Pseudomomentum including boundary terms*

If an isentropic layer does not intersect the ground, the integral of pseudomomentum over the global shell amounts to the integral of $-rC_2 - r_e u_e \cos\phi$ because the flux divergence terms on the second line of (6) integrate to zero and the third line is identically zero from the Casimir definition. However, care must be taken to include boundary terms in the wave activity when isentropic layers intersect the ground. Define $\mathcal{D}$ to be the domain where the

[2] Equivalent latitude is here defined as the latitude of the PV contour in the zonally symmetric background state with value Q.

isentropic layer of the full flow is above ground and $\mathcal{D}_o$ the domain where the layer in the background state is above ground. They will differ due to displacements of θ contours along the ground in the wavy state, as illustrated in Fig. 2 of Magnusdottir and Haynes (1996).

It is useful to partition space into several subdomains dependent on the locations of the lower boundary in the 3D and background (2D) states. $\mathcal{D} \cap \mathcal{D}_o$ is the intersection of regions $\mathcal{D}$ and $\mathcal{D}_o$. In general its boundary is not zonally symmetric. Let $\bar{\mathcal{D}}$ denote an area bounded to the south on each isentropic surface by the maximum latitude at which the disturbed surface intersects the lower boundary ($\bar{\mathcal{D}}$ must be a subset of $\mathcal{D} \cap \mathcal{D}_o$ where the full and background states are above ground at every longitude). Let $\mathcal{D} \setminus (\mathcal{D} \cap \mathcal{D}_o)$ denote the portion of the isentropic layer of the full flow that lies outside the intersection domain. The background state quantities are not defined here. Similarly $\mathcal{D}_o \setminus (\mathcal{D} \cap \mathcal{D}_o)$ is the portion of the background state outside the intersection domain. The global integral of pseudomomentum density (5) can then be written:

$$\begin{aligned}
\mathcal{P} \quad = \quad & \int_{\bar{\mathcal{D}}} \{-rC_2 - r_e u_e \cos\phi\} \; a^2 \cos\phi \, d\lambda \, d\phi \, d\theta \\
& - \int_{\partial\bar{\mathcal{D}}} \left(\frac{\partial C}{\partial q}\right)_o u_e \cos\phi \, a \, d\lambda \, d\theta \\
& + \int_{(\mathcal{D}\cap\mathcal{D}_o)\setminus\bar{\mathcal{D}}} \left\{-rC_2 - (r_o + r_e) u_e \cos\phi - \left(\frac{\partial C}{\partial q}\right)_o \zeta_e\right\} \; a^2 \cos\phi \, d\lambda \, d\phi \, d\theta \\
& - \int_{\mathcal{D}\setminus(\mathcal{D}\cap\mathcal{D}_o)} r(Z + C) \; a^2 \cos\phi \, d\lambda \, d\phi \, d\theta + \int_{\mathcal{D}_o\setminus(\mathcal{D}\cap\mathcal{D}_o)} r_o(Z_o + C_o) \; a^2 \cos\phi \, d\lambda \, d\phi \, d\theta.
\end{aligned} \qquad (15)$$

The first line is the "interior pseudomomentum" split into a "Rossby wave term" (related to displacing PV contours) and a "gravity wave term" (which is typically much smaller on baroclinic eddy scales). The second line comes from Gauss' theorem applied to the flux divergence term in (6) and noting that v_e integrates to zero around a latitude circle. It will be denoted $\mathcal{P}_b$ for boundary integral. The third line will be denoted $\mathcal{P}_d$ for within the domain of intersection and the fourth line $\mathcal{P}_e$ for exterior to the intersection domain.

$\mathcal{P}_b$ and $\mathcal{P}_d$ are evaluated using (9) and the values of the mass integrals obtained from the disturbed 3-D state. In order to evaluate the $\mathcal{P}_e$ term, (8) is used to express Casimir density in terms of mass and circulation integrals:

$$\begin{aligned} C(Q,\theta) &= -Z(q_o = Q, \theta) + \frac{Q}{2\pi a}\left\{\mathcal{M}(Q,\theta) - \mathcal{M}_s(\theta)\right\} \\ &= \frac{1}{2\pi a}\left(-\mathcal{C}(Q,\theta) + Q\left\{\mathcal{M}(Q,\theta) - \mathcal{M}_s(\theta)\right\}\right). \end{aligned} \tag{16}$$

where Stokes' theorem was used to relate the angular momentum around the zonally symmetric contour $q_o = Q$ to the circulation integral, $\mathcal{C}(Q,\theta)$.

c. Pseudomomentum in the small amplitude limit

In the limit of small perturbation amplitude, the expression for pseudomomentum density (15) can be simplified. This is especially important for the boundary terms because as the perturbations to the intersection of isentropic shells with the ground become smaller, $\mathcal{D} \to \mathcal{D}_o$ and the integrals $\mathcal{P}_d$ and $\mathcal{P}_e$ cannot be evaluated by numerical integration. Nevertheless, their contribution is important to the pseudomomentum of normal modes (Heifetz et al., 2004).

Firstly, consider the "Rossby wave term" $-rC_2$. In (12) we can assume that the second derivative of C is constant across the range of the perturbation so that integration over PV values gives:

$$\begin{aligned} P_w &= -r_o\left(\frac{\partial^2 C}{\partial q^2}\right)_o \frac{q_e^2}{2} \\ &= -r_o \frac{\partial}{\partial \phi}\left(\frac{\partial C}{\partial q}\right)_o \frac{\partial \phi}{\partial q_o}\frac{q_e^2}{2} \\ &= \frac{r_o^2 \cos\phi_o}{q_{o_y}}\frac{q_e^2}{2} \end{aligned} \tag{17}$$

where $y = a\phi$ and $q_{o_y} = \partial q_o/\partial y$ is the background state meridional PV gradient. The "gravity wave term" is unaltered at small amplitude. The integral over the intersection region, $\mathcal{P}_d$, can be incorporated into the interior integral if the boundary integral is taken around $\partial(\mathcal{D} \cap \mathcal{D}_o)$.

By definition the mass enclosed by the background state PV contour everywhere coincident with the intersection of the isentropic layer with the ground ($q_{b_o} = Q$) is $\mathcal{M}(Q,\theta) =$

$\mathcal{M}_s(\theta)$ giving $(\partial C/\partial q)_o = 0$ at the boundary $\partial\mathcal{D}_o$ from (9). Therefore there is no contribution to the boundary integral $\mathcal{P}_b$ wherever $\partial(\mathcal{D}\cap\mathcal{D}_o)$ is coincident with $\partial\mathcal{D}_o$. This occurs if the boundary θ contour of the perturbed state lies south of the contour for the background state ($\phi_{b_e} = \phi_b - \phi_{b_o} < 0$). Furthermore, the derivative can be written:

$$\left(\frac{\partial C}{\partial q}\right)_o = \frac{\partial}{\partial\phi}\left(\frac{\partial C}{\partial q}\right)(\phi_b - \phi_{b_o}) = -r_o\cos\phi\, a\phi_{b_e} \tag{18}$$

using (7) for the last step. This can be substituted into the integral $\mathcal{P}_b$ where $\phi_{b_e} > 0$.

The final integrals are the exterior terms, $\mathcal{P}_e$. In region $\mathcal{D}\setminus(\mathcal{D}\cap\mathcal{D}_o)$ the perturbed isentropic shell lies south of the background shell so that $\phi_{b_e} < 0$. Using (2), (4), (8) and (18), then dropping second order terms in the integrand, the first $\mathcal{P}_e$ term becomes:

$$\begin{aligned}
\mathcal{P}_{e_1} &\approx -\int_{\mathcal{D}\setminus(\mathcal{D}\cap\mathcal{D}_o)} \{(r_o+r_e)(Z_o+C_o) + r_o(Z - Z_o + C - C_o)\}\, a^2\cos\phi\, d\lambda\, d\phi\, d\theta \qquad (19)\\
&\approx -\int_{\mathcal{D}\setminus(\mathcal{D}\cap\mathcal{D}_o)} \left\{(r_o+r_e)q_o\left(\frac{\partial C}{\partial q}\right)_o + r_o\left(u_e\cos\phi + q_e\left(\frac{\partial C}{\partial q}\right)_o + C_2\right)\right\} a^2\cos\phi\, d\lambda\, d\phi\, d\theta\\
&\approx -\iint\int_{\phi_{b_e}}^{0} \{-r_o^2 q_o a\phi' + r_o u_e\}\, a^2\cos^2\phi_{b_o}\, d\phi'\, d\lambda\, d\theta\\
&\approx \iint\left\{-r_o^2 q_o\frac{a^2\phi_{b_e}^2}{2} + r_o u_e a\phi_{b_e}\right\} a\cos^2\phi_{b_o}\, d\lambda\, d\theta
\end{aligned}$$

In region $\mathcal{D}_o\setminus(\mathcal{D}\cap\mathcal{D}_o)$ the perturbed isentropic shell lies north of the background shell so that $\phi_{b_e} > 0$. The second $\mathcal{P}_e$ term becomes:

$$\begin{aligned}
\mathcal{P}_{e_2} &\approx \int_{\mathcal{D}_o\setminus(\mathcal{D}\cap\mathcal{D}_o)} r_o q_o\left(\frac{\partial C}{\partial q}\right)_o a^2\cos\phi\, d\lambda\, d\phi\, d\theta \qquad (20)\\
&\approx \iint\int_0^{\phi_{b_e}} -r_o^2 q_o a\phi' a^2\cos^2\phi_{b_o}\, d\phi'\, d\lambda\, d\theta\\
&\approx \iint -r_o^2 q_o\frac{a^2\phi_{b_e}^2}{2} a\cos^2\phi_{b_o}\, d\lambda\, d\theta
\end{aligned}$$

Note that the $\phi_{b_e}^2$ and $u_e\phi_{b_e}$ terms from $\mathcal{P}_e$ and $\mathcal{P}_b$ appear in both domains where $\phi_{b_e} > 0$ and $\phi_{b_e} < 0$ and can therefore be integrated globally. It is useful to write all boundary terms as delta-function contributions to the global integral:

$$\begin{aligned}
\mathcal{P}_b + \mathcal{P}_e &= \iint\left\{-r_o^2 q_o r_o\frac{a^2\phi_{b_e}^2}{2} + r_o u_e a\phi_{b_e}\right\}\cos^2\phi_{b_o} a\, d\lambda\, d\theta \qquad (21)\\
&= \iiint\left\{r_o^2 q_o r_o\frac{y_{b_e}^2}{2} - r_o u_e y_{b_e}\right\}\cos\phi_{b_o}\frac{\partial\theta_o}{\partial y}\delta(\theta-\theta_{b_o})a^2\cos\phi\, d\lambda\, d\phi\, d\theta
\end{aligned}$$

where the integral over θ values along the boundary was transformed to an integral over latitude using $d\theta = -\left.\partial\theta/\partial\phi\right|_b d\phi$ and then the delta-function $\delta(\theta - \theta_{b_o})$ was introduced to pick out the boundary from a 3-D integral re-introducing θ as the vertical coordinate. Gathering all terms, the expression for the pseudomomentum density of small amplitude waves is:

$$P = \frac{r_o^2 \cos\phi_o}{q_{o_y}}\frac{q_e^2}{2} - r_e u_e \cos\phi + \left\{ r_o^2 q_o \frac{y_{b_e}^2}{2} - r_o u_e y_{b_e} \right\} \cos\phi_{b_o} \frac{\partial\theta_{b_o}}{\partial y}\delta(\theta - \theta_{b_o}). \tag{22}$$

Note that the interior terms were first derived for the primitive equations for small amplitude disturbances by Andrews (1983b). Equivalent boundary terms were derived by Magnusdottir and Haynes (1996) and presented in this form by Methven et al. (2005a). Often the assumption of PV conservation is used to relate small amplitude meridional air parcel displacements along isentropic surfaces, η, to PV anomalies using $\eta = -q_e/q_{o_y}$. In this case the Rossby wave term P_w can be written in the familiar form, $r_o^2 \cos\phi\, q_{o_y} \frac{1}{2}\eta^2$.

3. Pseudoenergy conservation

a. Pseudoenergy density for large amplitude disturbances

Following Haynes (1988), the pseudoenergy density can be defined by:

$$H(\lambda, \phi, \theta, t) = r(E + B) - r_o(E_o + B_o). \tag{23}$$

where specific energy is defined as:

$$E = \frac{1}{2}\left(u^2 + v^2\right) + h(p, \theta) \tag{24}$$

and h is the specific enthalpy. As before, the Casimir density (written B to distinguish it from the Casimir C used for pseudomomentum) can be expanded in terms of PV perturbations following (4) and similarly the enthalpy function can be expanded in the pressure perturbation defined with reference to a given isentropic surface:

$$
\begin{aligned}
h &= h_o + \left(\frac{\partial h}{\partial p}\right)_o p_e + h_2(p_o, p_e, \theta) \\
&= h_o + \left(\frac{\partial h}{\partial p}\right)_o p_e + \int_0^{p_e} (p_e - \tilde{p}) \left.\frac{\partial^2 h}{\partial \tilde{p}^2}\right|_\theta (p_o + \tilde{p}, \theta)\, d\tilde{p}
\end{aligned} \tag{25}
$$

Writing (23) in terms of background state and perturbation quantities and using $rq_e = \zeta_e - r_e q_o$ and $r_e = -(1/g)\partial p_e/\partial\theta$ obtains:

$$
\begin{aligned}
H &= \frac{r}{2}(u_e^2 + v_e^2) + rB_2 + rh_2 + r(u_o u_e + v_o v_e) \\
&\quad + r_e\left\{E_o + B_o - q_o\left(\frac{\partial B}{\partial q}\right)_o\right\} + \left(\frac{\partial B}{\partial q}\right)_o \zeta_e - \left(\frac{\partial h}{\partial p}\right)_o \frac{1}{2g}\frac{\partial (p_e^2)}{\partial \theta} + r_o\left(\frac{\partial h}{\partial p}\right)_o p_e
\end{aligned} \tag{26}
$$

The first order $r_o(u_o u_e + v_o v_e)$ and ζ_e terms can be transformed into a horizontal flux divergence as for pseudomomentum. The final first order p_e term requires more attention. For the particular case of an ideal gas:

$$
h = c_p T = \theta c_p \left(\frac{p}{p_{oo}}\right)^\kappa \tag{27}
$$

where c_p is the specific heat capacity, R is the specific gas constant, $\kappa = R/c_p = 2/7$ and p_{oo} is a constant reference pressure. The enthalpy derivatives can then be evaluated explicitly:

$$
\left.\frac{\partial h}{\partial p}\right|_\theta = \frac{\kappa h}{p}; \quad \left.\frac{\partial^2 h}{\partial p^2}\right|_\theta = \frac{\kappa(\kappa-1)h}{p^2}; \quad \left.\frac{\partial}{\partial \theta}\right|_{\lambda,\phi} \left.\frac{\partial h}{\partial p}\right|_\theta = \left\{\frac{\kappa}{p\theta} - \kappa(\kappa-1)\frac{gr}{p^2}\right\} h \tag{28}
$$

Using the definition of pseudodensity, $r_o = \rho_o \partial z_o/\partial\theta$, and the ideal gas law, $p = \rho R T$, yields:

$$
r_o \left(\frac{\partial h}{\partial p}\right)_o p_e = \frac{\partial z_o}{\partial \theta} p_e = \frac{\partial}{\partial \theta}(z_o p_e) + g z_o r_e \tag{29}
$$

Manipulating the above expressions gives the result:

$$
\begin{aligned}
H &= \frac{r}{2}(u_e^2 + v_e^2) + rB_2 + \left\{rh_2 + \frac{p_e^2}{2g}\frac{\partial}{\partial\theta}\left(\frac{\partial h}{\partial p}\right)_o\right\} + r_e(u_o u_e + v_o v_e) \\
&\quad + \frac{1}{a\cos\phi}\frac{\partial}{\partial\lambda}\left\{v_e\left(\frac{\partial B}{\partial q}\right)_o\right\} - \frac{1}{a\cos\phi}\frac{\partial}{\partial\phi}\left\{u_e\left(\frac{\partial B}{\partial q}\right)_o \cos\phi\right\} \\
&\quad + \frac{\partial}{\partial\theta}\left\{z_o p_e - \frac{p_e^2}{2g}\left(\frac{\partial h}{\partial p}\right)_o\right\} \\
&\quad + u_e\left\{r_o u_o + \frac{1}{a}\frac{\partial}{\partial\phi}\left(\frac{\partial B}{\partial q}\right)_o\right\} + v_e\left\{r_o v_o - \frac{1}{a\cos\phi}\frac{\partial}{\partial\lambda}\left(\frac{\partial B}{\partial q}\right)_o\right\} \\
&\quad + r_e\left\{E_o + B_o + g z_o - q_o\left(\frac{\partial B}{\partial q}\right)_o\right\}.
\end{aligned} \tag{30}
$$

The top line contains four second order terms: perturbation kinetic energy, a "Rossby wave term" (rB_2 involving PV contour displacements), available potential energy (APE) and a "gravity wave term" involving correlations between perturbation density and velocity. The second and third lines are expressed as a flux divergence. However, the last two lines are first order and must be eliminated by defining the energy-Casimir using the relations:

$$\left(\frac{\partial B}{\partial q}\right)_o = \Psi \tag{31}$$

$$E_o + B_o + gz_o - q_o\left(\frac{\partial B}{\partial q}\right)_o = 0 \tag{32}$$

where the background state mass streamfunction is defined by $r_o u_o = -(1/a)\partial\Psi/\partial\phi$ and $r_o v_o = (1/a\cos\phi)\partial\Psi/\partial\lambda$. Note that u_o and v_o must be rotational when the background state flow is adiabatic (Haynes, 1988). It can also be shown that the second equality is always satisfied if the Casimir is defined using the streamfunction.

b. *Pseudoenergy including boundary terms*

The procedure from Section 2b is used to produce a new expression for integral pseudoenergy including boundary terms where isentropic surfaces intersect the ground:

$$\begin{aligned}
\mathcal{H} = & \int_{\bar{\mathcal{D}}} \left\{ \frac{r}{2}(u_e^2 + v_e^2) + rB_2 + rh_2 + \frac{p_e^2}{2g}\frac{\partial}{\partial\theta}\left(\frac{\partial h}{\partial p}\right) + r_e(u_o u_e + v_o v_e) \right\} a^2 \cos\phi \, d\lambda \, d\phi \, d\theta \\
& + \int_{\partial\bar{\mathcal{D}}} \left(\frac{\partial B}{\partial q}\right)_o u_e \cos\phi \, a \, d\lambda \, d\theta \\
& + \int_{(\mathcal{D}\cap\mathcal{D}_o)\setminus\bar{\mathcal{D}}} \left\{ \frac{r}{2}(u_e^2 + v_e^2) + rB_2 + rh_2 + \frac{p_e^2}{2g}\frac{\partial}{\partial\theta}\left(\frac{\partial h}{\partial p}\right)_o \right. \\
& \qquad \left. + (r_o + r_e)(u_o u_e + v_o v_e) + \left(\frac{\partial B}{\partial q}\right)_o \zeta_e \right\} a^2 \cos\phi \, d\lambda \, d\phi \, d\theta \\
& + \int_{\mathcal{D}\cap\mathcal{D}_o} \left[z_o p_e - \frac{p_e^2}{2g}\left(\frac{\partial h}{\partial p}\right)_o \right]_{bot}^{top} a^2 \cos\phi \, d\lambda \, d\phi \\
& + \int_{\mathcal{D}\setminus(\mathcal{D}\cap\mathcal{D}_o)} r(E+B) \, a^2 \cos\phi \, d\lambda \, d\phi \, d\theta - \int_{\mathcal{D}_o\setminus(\mathcal{D}\cap\mathcal{D}_o)} r_o(E_o + B_o) \, a^2 \cos\phi \, d\lambda \, d\phi \, d\theta.
\end{aligned} \tag{33}$$

The last line involves integration of $E + B$ over the portions of the 3D or background (2D) state that lie outside the intersection domain $\mathcal{D} \cap \mathcal{D}_o$. Preliminary work evaluating the terms from atmospheric analyses has found that the two integrals typically are large with a high degree of cancellation in their sum ($\mathcal{H}_e$). In Section 3c it will be shown that together they reduce to a second order boundary term in the small amplitude limit.

The fourth integral, $\mathcal{H}_t$, is taken across the bottom and top boundaries of the intersection domain and arises from integrating the vertical flux divergence in (30). Note that the $z_o p_e$ term is typically much smaller because on the boundaries the zonal average pressure of the 3D state is close to the pressure of the background MLM state, so that $z_o p_e$ integrates almost to zero around a latitude circle.

The top line is the interior pseudoenergy which can be partitioned into kinetic energy ($H_k = \frac{r}{2}(u_e^2 + v_e^2)$), a "Rossby wave term" ($H_w = rB_2$), the available potential energy ($H_a = rh_2 + \frac{p_e^2}{2g}\frac{\partial}{\partial\theta}\left(\frac{\partial h}{\partial p}\right)_o$) and the "gravity wave term" ($H_g = r_e(u_o u_e + v_o v_e)$). The second integral, $\mathcal{H}_b$, arises from the horizontal flux divergence terms in (30) evaluated around the boundary of the zonally symmetric inner region $\bar{\mathcal{D}}$. The third integral, $\mathcal{H}_d$, spans the part of the intersection region $\mathcal{D} \cap \mathcal{D}_o$ lying outside $\bar{\mathcal{D}}$ (derived from (26)).

Mass streamfunction $\Psi(Q, \theta)$ is found by integrating $r_o u_o$ polewards along isentropic surfaces from the equator (assigning $\Psi = 0$ there) or the lower boundary if the isentropic surface intersects it. In order that the boundary terms are all quadratic, $E_o + B_o = 0$ at the boundary $\mathcal{D}_o$. Using the condition (32) implies that:

$$\Psi_b = \left.\frac{g z_o}{q_o}\right|_b \tag{34}$$

The energy-Casimir density is found by integrating (31) with respect to PV along isentropic surfaces using the boundary condition $B_b = -E_{o_b}$. B_2 can be calculated from its definition (4) given $B(Q, \theta)$ and $\Psi(Q, \theta)$.

c. *Pseudoenergy in small amplitude limit*

Now consider the pseudoenergy terms in the limit of small amplitude perturbations to a steady, zonally symmetric basic state. The second order energy-Casimir term becomes:

$$B_2 \approx \left(\frac{\partial^2 B}{\partial q^2}\right)_o \frac{q_e^2}{2} = \left(\frac{\partial \Psi}{\partial q}\right)_o \frac{q_e^2}{2} = \frac{\partial \Psi}{\partial \phi}\frac{\partial \phi}{\partial q_o}\frac{q_e^2}{2} = -\frac{r_o u_o}{q_{o_y}}\frac{q_e^2}{2} \tag{35}$$

The available potential energy reduces at small amplitude to:

$$\begin{aligned} rh_2 + \frac{p_e^2}{2g}\frac{\partial}{\partial \theta}\left(\frac{\partial h}{\partial p}\right)_o &\approx r_o \left(\frac{\partial^2 h}{\partial p^2}\right)_o \frac{p_e^2}{2} + \frac{p_e^2}{2g}\frac{\partial}{\partial \theta}\left(\frac{\partial h}{\partial p}\right)_o \\ &\approx \frac{\kappa h_o}{g p_o \theta}\frac{p_e^2}{2} \end{aligned} \tag{36}$$

where the last step uses (28). In the absence of background state orography, $\Psi_b = 0$ from (34) and for small amplitude perturbations:

$$\Psi \approx \frac{\partial \Psi}{\partial \phi}(\phi_b - \phi_{b_o}) = -a r_o u_o \phi_{b_e} \tag{37}$$

Following the same technique as for the pseudomomentum boundary terms (19) and (20), one obtains the pseudoenergy density for small amplitude perturbations:

$$\begin{aligned} H &= \frac{r_o}{2}\left(u_e^2 + v_e^2\right) + \frac{\kappa h_o}{g p_o \theta}\frac{p_e^2}{2} - \frac{r_o^2 u_o}{q_{o_y}}\frac{q_e^2}{2} + r_e u_o u_e \\ &\quad + \left\{-r_o^2 q_o u_o \frac{y_{b_e}^2}{2} + r_o u_o u_e y_{b_e}\right\}\frac{\partial \theta_{b_o}}{\partial y}\delta(\theta - \theta_{b_o}) \end{aligned} \tag{38}$$

using $v_o = 0$ for the zonally symmetric state. Note that Methven et al. (2005a) presented (without showing a derivation) a similar small amplitude expression for pseudoenergy, but the boundary and APE terms have been corrected here[3]. The "Rossby wave terms" in the interior (H_w) and on the boundary are equal to the Rossby wave terms in pseudomomentum density multiplied by $-u_o/\cos\phi$ (i.e., the zonal angular velocity of the background state). Zadra et al. (2002) describe H_w as the "Doppler term" because the ratio of interior integral pseudoenergy to pseudomomentum can be written as an intrinsic phase speed, $-(KE +$

[3]The errors in Eqn.(6) of Methven et al. (2005a) do not affect the results obtained in that paper since only the "Rossby wave terms" (H_w) were required in the calculations.

$APE)/\mathcal{P}$, plus a Doppler mean wind, $-\mathcal{H}_w/\mathcal{P}$. This pseudoenergy expression entends the results of Andrews (1983a) for small amplitude quasi-geostrophic disturbances in height or pressure coordinates on a β-plane.

4. Application to atmospheric analyses

The pseudomomentum (15) and pseudoenergy (33) diagnostics for large amplitude disturbances were applied to atmospheric data obtained from the ECMWF Re-Analysis Interim dataset (ERA-I). The results are illustrated for the extended northern hemisphere winter 00UT 1 November 2009 to 00UT 1 April 2010 using re-analyses every 6 hours.

a. Calculating perturbation variables from the ERA-Interim dataset

The calculations are performed on the full resolution output of the ECMWF IFS model used for the re-analysis. The core output is spectral data at T255 resolution on 60 model levels (hybrid-pressure η-coordinate) for vorticity, divergence, temperature as well as surface pressure. The data is first transformed to u, v and θ on a linear Gaussian grid (512 longitudes $\times$ 256 latitudes) on model levels. The variables u and θ are interpolated horizontally to the mid-points of the grid in longitude and v and θ are interpolated to the mid-points in μ (sine of latitude). Linear vertical interpolation is then used to find u, v and p in a C-grid pattern in the horizontal (with p at the cell centres) on 131 isentropic (θ) levels from 218 to 2979K (equally spaced in θ upto 320K, blending to equal spacing in pseudo-height above 400K). Relative vorticity at cell centres is found by finite difference on the C-grid along isentropic surfaces. Geopotential on the top isentropic surface is found by integrating the hydrostatic equation upwards in η-coordinates from the ground. Pressure and geopotential there define Montgomery potential, M. The hydrostatic relation in isentropic coordinates is then integrated downwards to find M on every isentropic surface. The pseudo-density in isentropic coordinates, r, is found from derivatives of M with respect to θ using centred finite difference. The wind components are then readily obtained from derivatives of M by finite difference along isentropic surfaces.

This particular numerical method was used because it is consistent with the technique used to obtain the modified Lagrangian mean (MLM) background state through PV inversion. The first step in obtaining the MLM state is to calculate the mass and circulation integrals (11) within the contours of a discretised set of PV values, Q_k, on a set of potential temperature surfaces, θ_m, from the full 3-D state. The MLM state is defined as a zonally symmetric adiabatic re-arrangement of the 3-D state that contains the same mass and circulation within every PV contour. The procedure to satisfy the mass and circulation constraints simultaneously starts by calculating a first guess zonally symmetric state. Given the PV of this state, $q_o(\phi, \theta)$, and circulation integrals, $\mathcal{C}(q_o = Q, \theta)$, the lower boundary potential temperature and upper boundary pressure, it is possible to obtain M through inverting an almost elliptic equation. Winds and density are found from horizontal and vertical derivatives of M. Mass and circulation integrals are then calculated for this 2-D state. They will differ from those of the 3-D state, but the latitudes of the background state PV contours on isentropic surfaces and θ contours on the lower boundary are adjusted and the state is inverted again. The process is iterated until the mass and circulation integrals converge on those of the original 3-D data. The details of this procedure will be described in a separate article where the properties of the background state will be explored.

The most important aspects for this article are that the MLM state is a zonally symmetric solution of the primitive equations which would be steady if PV and θ were materially conserved (i.e., the flow were adiabatic and frictionless). It is a suitable state to partition perturbations from the full flow (i.e., $q_e = q - q_o$) that will obey the wave activity conservation laws derived here, even at large amplitude.

b. Pseudomomentum results

Figure 1 shows the integral over the northern hemisphere of the various terms comprising pseudomomentum and pseudoenergy, divided by the total mass of atmosphere in the hemisphere. Both the "boundary" terms in pseudomomentum, $\mathcal{P}_b$ and $\mathcal{P}_e$, are negative while all the other terms are positive. The "gravity wave term", $\mathcal{P}_g$, is the smallest, as is the "gravity wave term" in pseudoenergy, $\mathcal{H}_g$. Although the balanced flow associated with Rossby waves can contribute to the "gravity wave term", gravity and Kelvin waves have no influence on PV contours (unless they break and dissipate). The implication of the small magnitude of $\mathcal{P}_g$ relative to all other terms is that Rossby wave activity dominates.

The interior pseudomomentum has been partitioned into three. $\mathcal{P}_d$ is associated with the volume $(\mathcal{D} \cap \mathcal{D}_o) \setminus \bar{\mathcal{D}}$ which is just above ground in both the full and background states but lies within the range of latitudes where the full state intersects the ground. $\mathcal{P}_{trop}$ represents wave activity above $\mathcal{P}_d$ to the 400K isentropic surface and $\mathcal{P}_{strat}$ is the integral of all wave activity above 400K to the top isentropic boundary of the analysis domain (3043K; pressure $\approx 10 - 20\,\mathrm{Pa}$) which lies in the mesosphere. Clearly pseudomomentum is dominated by the troposphere. Interestingly, $\mathcal{P}_d$ is approximately equal and opposite to $\mathcal{P}_b + \mathcal{P}_e$. Baroclinic waves have negative boundary wave activity associated with a surface potential temperature wave and positive interior wave activity associated with an upper wave in PV along isentropic surfaces. In the small-amplitude limit these two counter-propagating Rossby wave (CRW) components describe the evolution and mechanism for baroclinic instability (Methven et al., 2005a). It is the case that any growing normal mode must have exactly zero total pseudomomentum (otherwise the disturbance could not grow without violating global conservation of pseudomomentum). The near-cancellation observed in the analyses is suggestive that $\mathcal{P}_d$ and the boundary terms are dominated by baroclinic wave activity.

However, it is also clear that there is much more interior pseudomomentum in the $\mathcal{P}_{trop}$ term. This must be related to wave activity in the upper troposphere and lower stratosphere that is in excess of that required for baroclinic normal mode growth. There are many possible interpretations of this result that merit further investigation.

One is simply that some finite amplitude wave activity at tropopause level persists without recourse to modal baroclinic growth. This could perhaps occur through continuous excitation of transient wave growth by baroclinic or barotropic mechanisms (Farrell, 1982) associated with existing perturbations on the tropopause. Rivest and Farrell (1992) introduced "quasi-modes" as particular combinations of continuous spectrum modes which have similar zonal phase speeds. They showed that the decay rate of quasi-modes is related to the spread in frequencies of the contributing modes. De Vries et al. (2009) showed how such non-modal growth on any zonal shear flow can readily be interpreted in terms of Rossby wave components, even in situations where the PV gradient is continuous. If somehow upper level Rossby waves are continually forced, they would cause low level Rossby waves (associated with boundary potential temperature perturbations and vorticity) to grow as they moved along. However, the weaker amplitude in the boundary wave activity at all times indicates that they do not have sufficient time to phase-lock and grow in concert with the upper waves (modally) before the waves decay, by damping or transience.

Even if starting with modal growth, the nonlinear saturation of baroclinic waves also occurs faster at low levels than at the tropopause. Thorncroft et al. (1993) outlined a "saturation-propagation-saturation" mechanism involving lower wave nonlinear saturation in amplitude, the vertical propagation of a Rossby wave packet resulting in continued upper wave amplification and eventually nonlinear saturation there by Rossby wave breaking. Methven et al. (2005b) replaced the vertical propagation element of the paradigm with the interpretation that the lower and upper counter-propagating Rossby wave properties do not change, except that the lower CRW amplitude ceases to grow, due to nonlinear wave breaking limiting its meridional extent, while the upper CRW continues to grow through the same baroclinic growth mechanism. Thus nonlinear baroclinic wave behaviour may explain to some extent the dominance of interior pseudomomentum.

Planetary wave activity, including stationary waves, also make a large contribution since the background used to partition disturbances from the full atmospheric state is defined as zonally symmetric. Evidence for planetary and near-stationary waves will be presented in Section 4d.

Note that there are two clear maxima in stratospheric pseudomomentum at 45 and 92 days (15 Dec 2009 and 31 Jan 2010). These correspond to the beginning of a minor and major stratospheric sudden warming event respectively and are related to large-amplitude planetary wave activity and nonlinear wave breaking.

c. *Pseudoenergy results*

In the pseudoenergy time series, the interior PV displacement term, $\mathcal{H}_w$, and "gravity wave term", $\mathcal{H}_g$, are both negative, but with the $\mathcal{H}_w$ term which is associated with Rossby waves being much larger. The small-amplitude limits of pseudoenergy and pseudomomentum show that the corresponding density $H_w = -P_w u_o / \cos\phi$. Since the flow at tropopause level is mainly westerly ($u_o > 0$), this explains the strong anti-correlation between $\mathcal{H}_w$ and $\mathcal{P}_{trop}$. It is also clear that $\mathcal{H}_w$ has a larger fractional variation than $\mathcal{P}_{trop}$.

The interior (domain $\bar{\mathcal{D}}$) disturbance kinetic energy and available potential energy are positive definite and exhibit variability, although not as marked as in $\mathcal{H}_w$. The sum $\mathcal{H}_d + \mathcal{H}_b + \mathcal{H}_t$ is generally positive and smaller than the interior energy terms. The $\mathcal{H}_b$ term has the smallest magnitude of the three and $\mathcal{H}_t$ is always positive over this period. $\mathcal{H}_d$ can be both positive or negative and is more variable. The most variable term is the "exterior term" $\mathcal{H}_e$ related meridional displacements of potential temperature contours along the lower boundary. It is positive throughout the winter shown but smaller and even negative in November and March. It also exhibits a stronger diurnal cycle than the other terms which is related to a diurnal cycle in the isentropic density field of the tropical lower troposphere of the background state. The diurnal cycle will not be explored here, but is shown so that no time filtering is applied to the re-analysis data.

d. *Interpretation of wave activity*

The peak hemispheric pseudomomentum, KE, APE and $\mathcal{H}_w$ are all associated with one event between days 84-90 (23-29 Jan 2010) which makes an interesting case study. The signature of this event is visible first in the growth of APE from 23 Jan 2010. At the same time a weak dip develops in the boundary pseudomomentum term $\mathcal{P}_e$. These are associated with the growth in meridional displacements of potential temperature contours along the lower boundary (i.e., a lower CRW). $\mathcal{P}_e$ reaches a minimum first at 12UT 25 Jan 2010 followed by a peak in APE 12 hours later. The interior KE and $\mathcal{P}_{trop}$ peak at 06UT 27 Jan 2010, coincident with a distinct minimum in $\mathcal{H}_w$.

Figure 2a shows a snapshot of PV anomalies on the 311K surface at 12UT 26 January 2010 between the peak in disturbance APE and KE. The field shown is $r_o(q-q_o) = r_o q_e$ which has units of s^{-1} and is closely related to quasigeostrophic PV. To a reasonable approximation the magnitude of these anomalies scales in proportion to the balanced winds that would be obtained by PV inversion. Although the Ertel PV, q, is approximately conserved moving with air parcels on isentropic surfaces, clearly the PV anomalies are not and depend on the displacement of PV contours relative to their latitudes in the background state. The striking feature is a PV wave with zonal wavenumber 8. It has large amplitude so that positive PV anomalies are displaced to the south of the background state tropopause location ($\approx$ 50N) and negative PV anomalies are displaced to the north. The wave is much more distinct around the latitude of the positive (cyclonic) PV anomalies. Animations reveal that the wave grew at all longitudes simultaneously and strongly resembles a baroclinic wave life cycle. The hemispheric wave activity diagnostics show that it developed through mutual interaction between lower boundary potential temperature and tropopause level PV waves, saturated first a low levels and peaked several days later coinciding with the maximum in disturbance KE, as described in Thorncroft et al. (1993) and Methven et al. (2005b). It is a beautiful example of the relevance of baroclinic instability to the atmosphere. However, it is also clear that this disturbance occurred on a backdrop of much greater wave activity throughout the hemisphere. As mentioned earlier, it is possible that a large portion of the other wave activity is associated with stationary waves.

The relationship between pseudoenergy and zonal pseudomomentum contains information regarding zonal propagation. In the case of neutral sinusoidal modes, $c = -\mathcal{H}/\mathcal{P}$ equals the phase speed of the mode. For disturbances of more general large amplitude structure, Brunet (1994) argued that c can be taken as a definition of "coherent propagation speed". The physical interpretation is that c is the speed of the frame of reference from which the disturbance appears most steady (i.e., moving with the disturbance). In the case of growing normal modes, both $\mathcal{H}$ and $\mathcal{P}$ are zero and this formula cannot work. However, Heifetz et al. (2004) showed that the problem is solved by decomposing the growing normal mode into two untilted counter-propagating Rossby wave structures with equal and opposite pseudomomentum and non-zero pseudoenergy. In this case, the phase speed of the growing normal mode is given by the average self-propagation speed of the two components $(-\mathcal{H}_1/\mathcal{P}_1 - \mathcal{H}_2/\mathcal{P}_2)/2$. When presented with the analysed atmospheric flow featuring large amplitude breaking Rossby waves it is not known precisely how to partition into suitable Rossby wave components. However, Brunet (1994) pioneered the method of Empirical Normal Mode (ENM) decomposition based on obtaining eigenstructures from data that are orthogonal with respect to a pseudomomentum norm, in a similar fashion to the CRW theory. He discussed the Haynes (1988) expressions for pseudomomentum and pseudoenergy in his theory, but in his analysis of PV on the 315K isentropic surface he used expressions appropriate for the shallow water equations to avoid the need to integrate wave activity in the vertical. Zadra et al. (2002) applied the ENM technique to analysis data using the Haynes (1988) wave activity on 16 isentropic levels spanning 270 to 450K, but treating 850hPa as the lower boundary of the data. They presented results for zonal wavenumbers 1, 5 and 9 and inferred that all the modes had eastward phase speeds in the range 4-15 $\mathrm{m\,s^{-1}}$. However, their analysis neglected the effects of boundary wave activity.

Here, boundary terms will be included. Since the boundary pseudomomentum is negative, it is important to note that the total pseudomomentum is always positive due to the dominance of the interior tropospheric term. The relevance of $c = -\mathcal{H}/\mathcal{P}$ to observed wave behaviour around the mid-latitudes will be investigated, where $\mathcal{H}$ and $\mathcal{P}$ are the total pseudoenergy and pseudomomentum. Figure 3a shows the speed c evaluated throughout the extended winter. The ratio is converted from $\mathrm{m\,s^{-1}}$ to degrees longitude per day by assuming that the reference frame moves as a solid body rotation in the zonal direction and that c relates to the speed at 50°N which is the approximate tropopause location and centre of wave activity throughout DJF (not shown here). There is clearly variability on long timescales. For example, between day 95 and 120 the value is particularly steady oscillating about zero (dominated by the diurnal cycle mentioned earlier). In this period we might expect the dominance of stationary wave activity. Figure 3b shows a longitude-time (Hovmoeller) plot of meridional wind on the 311K surface averaged at each instant across the mid-latitude band 45-60N (where it intersects the tropopause). Clearly, days 95-120 are indeed relatively stationary with three especially strong ridges (flanked by strong $v > 0$ to the west and $v < 0$ to the east) at 50E, 240E and 340E (approximately the Urals, Rockies and East Atlantic). Figure 2b shows the PV anomaly pattern at 12UT 9 Feb 2010 (near the beginning of this period). Meridionally oriented ridges of low PV air are seen extending from the subtropics into the polar regions in the vicinity of 240E and 340E and it is these features and the elongated troughs between them that were relatively steady for almost a month.

The first half of November 2009 (to day 18) and the last portion of March 2010 (from day 128) are characterised by positive (eastward) propagation speed c and it is clear from the Hovmoeller plot that these periods have a succession of eastward moving troughs and ridges. The faster disturbances appear to be moving at approximately $20^{\circ} day^{-1}$ which is consistent with c.

There is a long period with predominantly negative c from days 23-70, implying net westward propagation. The Hovmoeller plot reveals that this is associated with a planetary wave pattern retrogressing with eastward synoptic activity superimposed. The planetary wave is dominated by zonal wavenumber 2. Initially it propagates slowly westwards at $\sim -5°day^{-1}$. The ridge moving from 240E to 210E (westwards from the Rockies) is particularly prominent until day 40. The same wavenumber 2 pattern then continues to propagate westwards at a faster pace $\sim -10°day^{-1}$ until day 70. The observed propagation of meridional wind patterns is consistent with the time series of c in the top panel. Note that the $c < 0$ period is interupted by a strong event of $c \approx 0$ around day 50. This appears to be associated with a stronger packet of eastward synoptic wave activity superposed on the wavenumber 2 disturbance. It could also be related to wave activity at different latitudes or levels in the atmosphere. Note that the "baroclinic wave life cycle" event on days 87-89 is also clear as a spike of eastward propagation during an otherwise near-stationary period.

5. Conclusions

Expressions for two measures of wave activity, pseudomomentum and pseudoenergy, have been derived that are valid for large amplitude disturbances described by the primitive equations on the sphere. Account is taken of the intersection of isentropic layers with the ground and the movement of the intersection. The result for pseudomomentum (15) was obtained previously by Magnusdottir and Haynes (1996), but the pseudoenergy expression (33) has not been shown before. A new expression for pseudoenergy (38) is also obtained in the limit of small disturbances from a zonally symmetric background state.

In order to evaluate pseudomomentum and pseudoenergy from analysis or numerical model data, it is first necessary to define and calculate a background state. Disturbances are naturally defined as deviations between the full 3-D state and the background. In order for the global wave activity conservation laws to apply, it is essential that the background state is itself a solution of the primitive equations. It was shown that pseudomomentum is easier to evaluate if the zonally symmetric modified Lagrangian mean state is used as the background. Methven (2010) presented some preliminary results obtaining the modified Lagrangian mean state from meteorological analyses and the same technique has been used here (detailed paper in preparation). Nakamura and Solomon (2011) have obtained a similar modified Lagrangian mean PV distribution from global data, but prescribing the Eulerian zonal mean potential temperature as the lower boundary. They did not obtain the associated density field by inverting background state PV which would be necessary to define interior or boundary wave activity as presented here.

It was shown using ERA-Interim atmospheric data that the "coherent propagation speed" measure c, obtained from hemispheric integrals of pseudoenergy and pseudomomentum, does reflect the key characteristics of disturbance propagation seen at tropopause level. The wave activity diagnostics then enable a dissection of the aspects of the atmospheric flow that are most important to the propagation. The two periods of particularly strong westward propagation (days 36-46 and 55-70) were associated with the highest values of the "lower boundary term" in pseudoenergy $\mathcal{H}_e$ and also lower magnitude (and therefore more positive) PV displacement term $\mathcal{H}_w$. Although possessing synoptic and longer-timescale variability, the pseudomomentum is much less variable than the pseudoenergy. However, in these two westward periods $\mathcal{P}_e$ was stronger (more negative) and $\mathcal{P}_w$ was slightly weaker (less positive). This indicates a stronger disturbance in potential temperature in the lower troposphere and slightly less activity at tropopause level.

These results differ markedly from Brunet (1994) who identifies westward modes (from Empirical Normal Mode decomposition) as those where interior disturbance energy is greater than the magnitude of the Doppler term in pseudoenergy, $(KE + APE) > |\mathcal{H}_w|$. In the season studied here disturbance energy is always smaller, $(KE + APE) < |\mathcal{H}_w|$. It is the boundary term in pseudoenergy, $\mathcal{H}_e$, that makes to total pseudoenergy positive and therefore the coherent zonal propagation speed c $(= -\mathcal{H}/\mathcal{P})$ negative. The boundary wave activity terms were also neglected in Zadra et al. (2002), which likely explains why they deduced that quasi-modes at all zonal wavenumbers were associated with positive (eastward) phase speeds. They also used the zonal and time mean of the analyses to define the background state, even though on its own it is not a solution to the governing equations. This would tend to reduce the pseudomomentum density on isentropic surfaces where they intersect the tropopause because the zonal mean state has a much smaller meridional PV gradient than the MLM state (recall the linearised form $P_w = r_o^2 \cos\phi\, q_{o_y} \frac{1}{2}\eta^2$). However, the difference in zonal mean and MLM zonal flow is likely to have the greatest influence on zonal phase speed through the Doppler term $-\mathcal{H}_w/\mathcal{P}$ which in the small amplitude limit reduces to $-u_o/\cos\phi$.

An event was also identified from the wave activity diagnostics resembling a baroclinic wave life cycle and the evolution of PV at this time reveals that there was indeed the almost simultaneous growth and decay of a zonal wavenumber 8 disturbance. Wave activity growth in boundary potential temperature and APE were first to peak (nonlinear saturation) with upper level PV disturbance and KE peaking 1-2 days later. The later stage of the life cycle has the opposite signature (more negative $\mathcal{H}_w$, $\mathcal{H}_e$ and more positive $\mathcal{P}_w$ and KE) relative to the "westward propagation phases".

In the extended winter studied, the month-long stationary period and the periods of westward coherent zonal propagation ($c < 0$) were dominated by a zonal wavenumber two pattern at tropopause level. At the same time there were clearly eastward propagating disturbances with shorter wavelengths (synoptic scale baroclinic waves). The Empirical Normal Mode decomposition technique of Brunet (1994) presents a means to partition cleanly the total pseudoenergy and pseudomomentum between different wavenumber components and estimate their characteristic phase speeds. It would be necessary to extend the analysis of Zadra et al. (2002) to include the boundary wave activity terms and re-examine the dominant modes or quasi-modes that describe the observed atmospheric behaviour.

The boundary term in pseudoenergy, $\mathcal{H}_e$, was shown to have much stronger variation over the season than the other terms in pseudoenergy or pseudomomentum. This is a very interesting aspect because it has an influence on the net propagation speed around the hemisphere, even at tropopause level. Further investigation into the phenomena responsible for this variation and its characteristics in other years could yield insight into why this particular winter was characterised by blocked flow and persistent weather patterns bringing especially cold conditions in northern Europe, North America and the Far East of Asia. For example, greater zonal asymmetry in lower boundary potential temperature (perhaps enhancement of land-sea contrast) would be reflected in greater $\mathcal{H}_e$, increasing the likelihood for stationarity or slow westward propagation. If the cold surface conditions intensify under the stationary weather systems, this raises the possibility of a positive feedback mechanism on the lower tropospheric temperature pattern via its effects on zonal wave propagation.

Acknowledgments.

Thanks to Paul Berrisford, part of the National Centres of Atmospheric Science (NCAS) Climate Directorate, for his advice on isentropic coordinates and tenacity with our work on background states. Thanks to the Dynamical Processes Group in the Department of Meteorology at the University of Reading, especially Brian Hoskins, Mike Blackburn, Tom Frame, Hylke de Vries and Tim Woollings, for exciting discussions on Rossby waves, wave breaking and relation to persistent weather patterns. This work was supported by grant NE/D011507 from the Natural Environment Research Council in the UK.

REFERENCES

Andrews, D., 1983a: A conservation law for small-amplitude quasi-geostrophic disturbances on a zonally asymmetric flow. *J. Atmos. Sci.*, **40**, 85–90.

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

Andrews, D., J. Holton, and C. Leovy, 1987: *Middle atmosphere dynamics*. Academic Press, 489 pp.

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

Bretherton, F., 1966a: Baroclinic instability and the short wavelength cut-off in terms of potential vorticity. *Q. J. R. Meteorol. Soc.*, **92**, 335–345.

Bretherton, F., 1966b: Critical layer instability in baroclinic flows. *Q. J. R. Meteorol. Soc.*, **92**, 325–334.

Brunet, G., 1994: Empirical normal mode analysis of atmospheric data. *J. Atmos. Sci.*, **51**, 932–952.

Bühler, O., 2009: *Waves and mean flows*. Cambridge University Press, 341pp.

Charney, J. and M. Stern, 1961: Propagation of planetary-scale disturbances from the lower into the upper atmosphere. *J. Geophys. Res.*, **66**, 83–110.

De Vries, H., J. Methven, T. Frame, and B. Hoskins, 2009: An interpretation of baroclinic initial value problems: Results for simple basic states with nonzero interior PV gradients. *J. Atmos. Sci.*, **66**, 864–882.

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

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

Heifetz, E., C. Bishop, B. Hoskins, and J. Methven, 2004: The counter-propagating Rossby wave perspective on baroclinic instability. Part I: Mathematical basis. *Q. J. R. Meteorol. Soc.*, **130**, 211–231.

Iwasaki, T., 1989: A diagnostic formulation for wave-mean flow interactions and Lagrangian mean circulation in a hybrid vertical coordinate of pressure and isentropes. *J. Meteorol. Soc. Japan*, **67**, 293–312.

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

McIntyre, M., 1980: Towards a Lagrangian-mean description of stratospheric circulations and chemical transports. *Phil. Trans. R. Soc. Lond.*, **A296**, 129–148.

McIntyre, M. and T. 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.

Methven, J., 2010: Diagnostics of the extratropics. *ECMWF Seminar 2009 on Diagnosis of Forecasting and Data Assimilation Systems*, ECMWF, 25–51.

Methven, J., E. Heifetz, B. Hoskins, and C. Bishop, 2005a: The counter-propagating Rossby wave perspective on baroclinic instability. Part III: Primitive equation disturbances on the sphere. *Q. J. R. Meteorol. Soc.*, **131**, 1393–1424.

Methven, J., E. Heifetz, B. Hoskins, and C. Bishop, 2005b: The counter-propagating Rossby wave perspective on baroclinic instability. Part IV: Nonlinear life cycles. *Q. J. R. Meteorol. Soc.*, **131**, 1425–1440.

Nakamura, N. and A. Solomon, 2010: Finite-amplitude wave activity and mean flow adjustments in the atmospheric general circulation. Part I: Quasigeostrophic theory and analysis. *J. Atmos. Sci.*, **67**, 3967–3983.

Nakamura, N. and A. Solomon, 2011: Finite-amplitude wave activity and mean flow adjustments in the atmospheric general circulation. Part II: Analysis in the isentropic coordinate. *J. Atmos. Sci.*, **68**, 2783–2799.

Rivest, C. and B. Farrell, 1992: Upper-tropospheric synoptic-scale waves. Part II: Maintenance and excitation of quasi-modes. *J. Atmos. Sci.*, **49**, 2120–2138.

Solomon, A. and N. Nakamura, 2012: An exact Lagrangian-mean wave activity for finite-amplitude disturbances to barotropic flow on a sphere. *J. Fluid Mech.*, **693**, 69–92.

Tanaka, D., T. Iwasaki, S. Uno, M. Ujiie, and K. Miyazaki, 2004: Eliassen-Palm flux diagnosis based on isentropic representation. *J. Atmos. Sci.*, **61**, 2370–2383.

Thorncroft, C., B. Hoskins, and M. McIntyre, 1993: Two paradigms of baroclinic wave life cycle behaviour. *Q. J. R. Meteorol. Soc.*, **119**, 17–55.

Thuburn, J. and V. Lagneau, 1999: Eulerian mean, contour integral and finite-amplitude wave activity diagnostics applied to a single layer model of the winter stratosphere. *J. Atmos. Sci.*, **56**, 689–710.

Zadra, A., G. Brunet, and J. Derome, 2002: An empirical normal mode diagnostic algorithm applied to NCEP Re-analyses. *J. Atmos. Sci.*, **59**, 2811–2829.

List of Figures

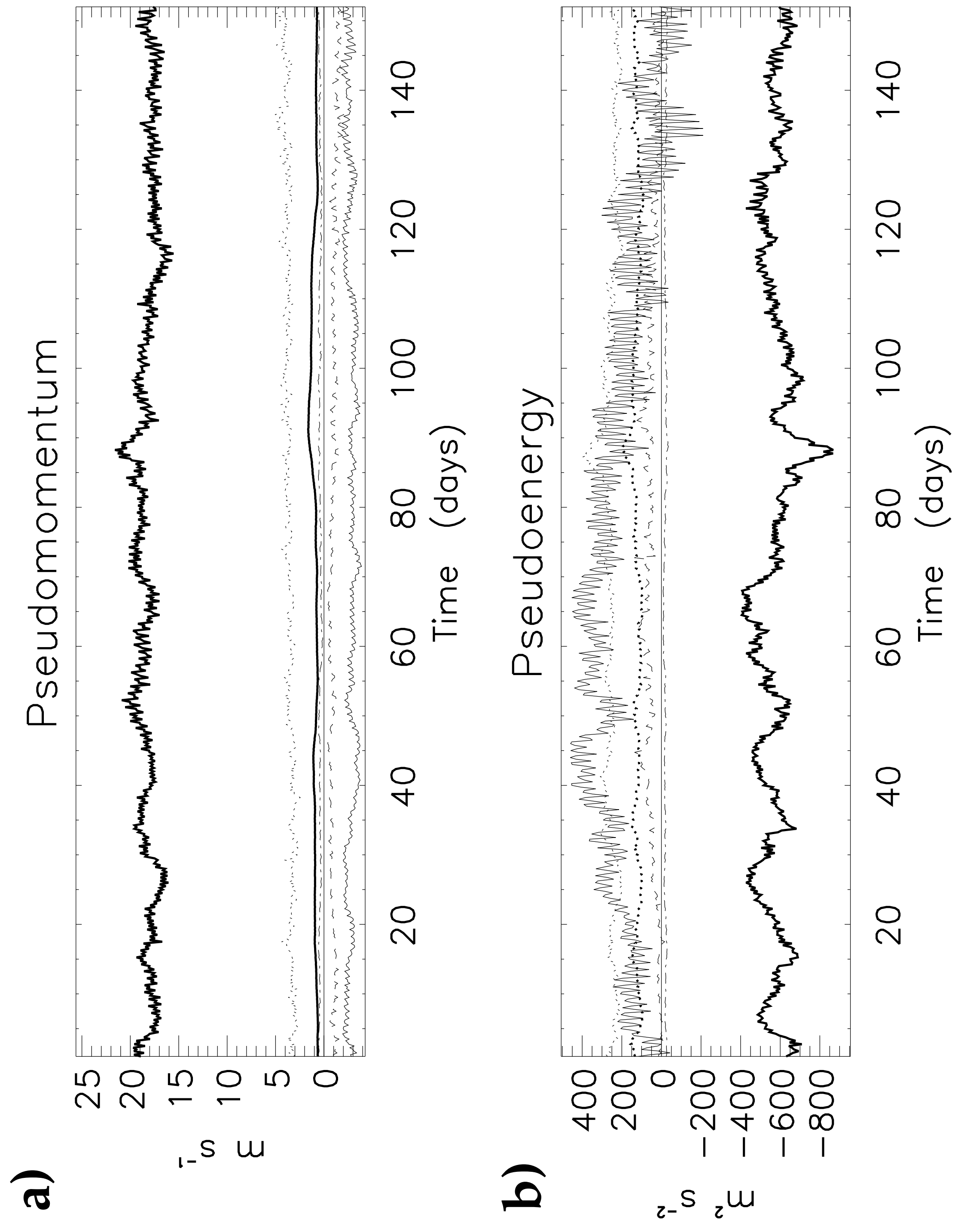

FIG. 1. Time series of pseudomomentum and pseudoenergy terms integrated over the northern hemisphere and divided by atmospheric mass. Using 6-hourly ERA-Interim data from 00UT 1 Nov 2009 until 00UT 1 Apr 2010. **(a)** $\mathcal{P}_{trop}$ and $\mathcal{P}_{strat}$ (bold, solid), $\mathcal{P}_e$ (thin, solid), $\mathcal{P}_d$ (dotted), $\mathcal{P}_b$ (dashed) and $\mathcal{P}_g$ (dash-dot). **(b)** $\mathcal{H}_w$ (bold, solid), $\mathcal{H}_e$ (thin, solid), KE (bold, dotted), APE (thin, dotted), $\mathcal{H}_d + \mathcal{H}_b + \mathcal{H}_t$ (dashed) and $\mathcal{H}_g$ (dash-dot).

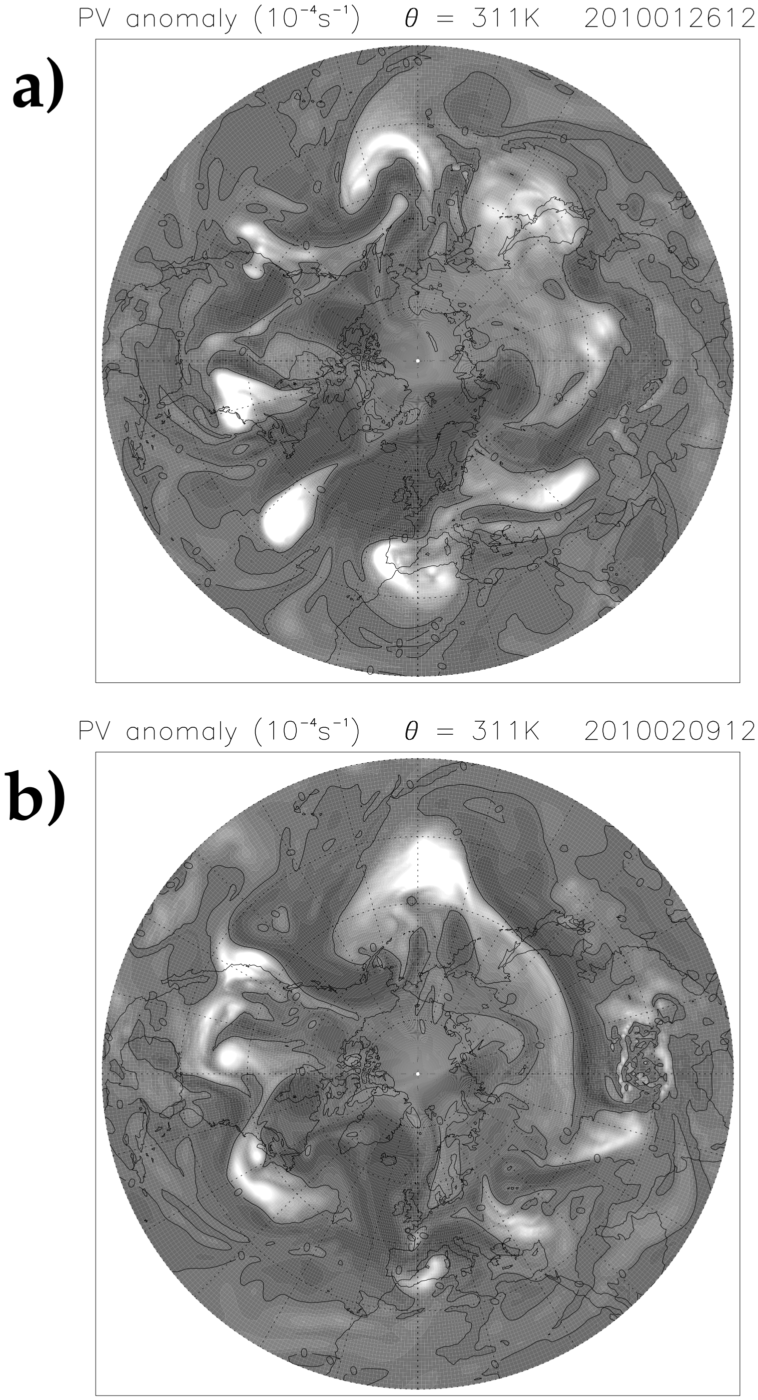

Fig. 2. Snapshots of PV anomalies on the 311K isentropic surface. Anomalies defined as $r_o(q - q_o)$ where q_o is the Ertel PV of the zonally symmetric background state. **(a)** 12UT 26 Jan 2010 one day before the northern hemisphere disturbance KE peaked, associated with the mature phase of a baroclinic wave life cycle with zonal wavenumber 8. **(b)** 12UT 9 Feb 2010 during the stationary phase characterised by strong ridges at 50, 240 and 340E where low PV air reached polar latitudes. Polar stereographic projection from North Pole to 15°N with 0°E at the bottom of each plot. Contour interval $0.25 \times 10^{-4}\mathrm{s}^{-1}$. White positive.

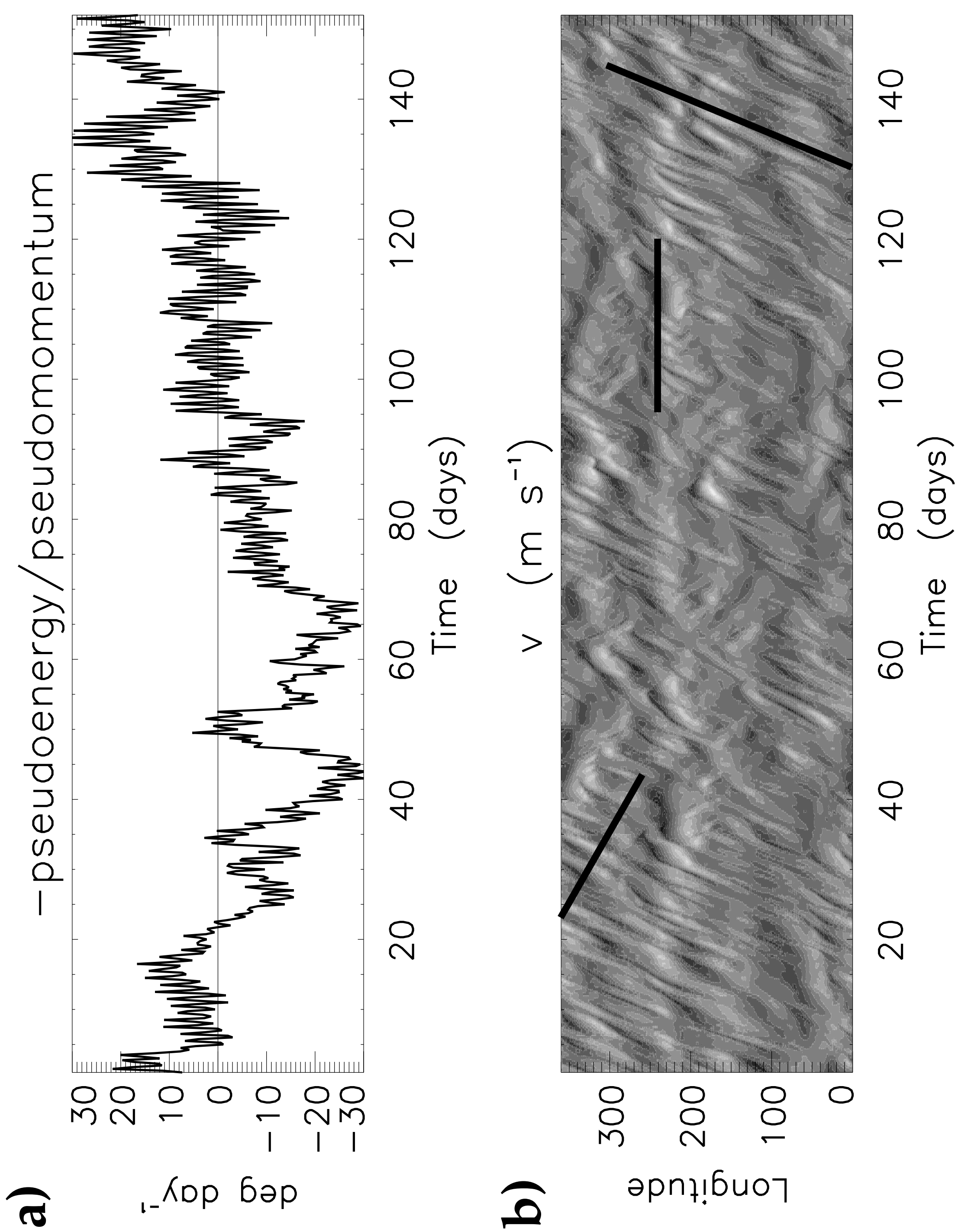

FIG. 3. **(a)** Time series of minus the ratio of total pseudoenergy to pseudomomentum which can be interpreted as a form of coherent zonal propagation speed (see text). **(b)** Longitude-time plot of mid-latitude meridional wind on the 311K surface (averaged over 45-60N). Shading from black to white over range -70 to +70 $\mathrm{m\,s^{-1}}$ every 10 $\mathrm{m\,s^{-1}}$. The bold lines indicate translation speeds of $-5°day^{-1}$, the stationary phase and $+20°day^{-1}$.