The slowly evolving background state of the atmosphere

RMetS
Royal Meteorological Society

The slowly evolving background state of the atmosphere

John Methven[a]* and Paul Berrisford[b]
[a]*Department of Meteorology, University of Reading, UK*
[b]*NCAS–Climate, ECMWF, Reading, UK*

*Correspondence to: J. Methven, Department of Meteorology, University of Reading, Earley Gate, Reading, RG6 6BB, UK.
E-mail: J.Methven@reading.ac.uk

The theory of wave–mean flow interaction requires a partition of the atmospheric flow into a notional background state and perturbations to it. Here, a background state, known as the Modified Lagrangian Mean (MLM), is defined as the zonally symmetric state obtained by requiring that every potential vorticity (PV) contour lying within an isentropic layer encloses the same mass and circulation as in the full flow. For adiabatic and frictionless flow, these two integral properties are time-invariant and the MLM state is a steady solution of the primitive equations. The time dependence in the adiabatic flow is put into the perturbations, which can be described by a wave-activity conservation law that is exact even at large amplitude. Furthermore, the effects of non-conservative processes on wave activity can be calculated from the conservation law.

A new method to calculate the MLM state is introduced, where the position of the lower boundary is obtained as part of the solution. The results are illustrated using Northern Hemisphere ERA-Interim data. The MLM state evolves slowly, implying that the net non-conservative effects are weak. Although 'adiabatic eddy fluxes' cannot affect the MLM state, the effects of Rossby-wave breaking, PV filamentation and subsequent dissipation result in sharpening of the polar vortex edge and meridional shifts in the MLM zonal flow, both at tropopause level and on the winter stratospheric vortex. The rate of downward migration of wave activity during stratospheric sudden warmings is shown to be given by the vertical scale associated with polar vortex tilt divided by the time-scale for wave dissipation estimated from the wave-activity conservation law. Aspects of troposphere–stratosphere interaction are discussed.

The new framework is suitable to examine the climate and its interactions with disturbances, such as midlatitude storm tracks, and makes a clean partition between adiabatic and non-conservative processes.

Key Words: circulation; modified Lagrangian mean; wave activity; conservation

Received 11 August 2014; Revised 10 December 2014; Accepted 15 January 2015; Published online in Wiley Online Library

1. Introduction

The theory of wave–mean flow interaction requires a partition of the atmospheric flow into a notional background state and perturbations to it. The evolution of both components and their diagnosed 'interaction' depends upon the partition. This article defines a new background state, designed to isolate the effects of non-conservative processes on the evolution of both the background state and large-amplitude perturbations. In an adiabatic, frictionless flow, the background state would be strictly steady and all the time-dependent motion would be reflected in the perturbations, often described loosely as 'wave activity'. Therefore any time dependence in the background is associated with non-conservative processes. The non-conservative effects on wave activity can also be quantified from global data. The lower boundary of the background state is found as part of the solution, which makes the partition useful for examination of large-amplitude baroclinic waves and their two-way interaction with the background flow.

The background state is defined in terms of two fundamental integral quantities, mass and circulation, that are conserved in adiabatic, frictionless flow. Each integral quantity is treated as a functional of only two variables, potential vorticity (PV) and potential temperature, that are themselves conserved for adiabatic and frictionless flow in a material sense (following fluid parcels). When potential temperature is materially conserved, air

is constrained to move along isentropic surfaces. Since PV is also conserved, it can be used to label air parcels on each isentropic surface so that every PV contour is a material contour (moves with the air) and there is no mass flux across it. The mass and circulation enclosed by every PV contour within every isentropic layer are conserved, meaning that they are invariant in time (see Hoskins, 1991). This is true for the full three-dimensional (3D) time-dependent flow and, since the background state is defined with the same integral properties (as functionals of PV and θ), it shares the same conservation properties. The key difference from the full flow is that the background state is defined to be zonally symmetric. McIntyre (1980) described a zonally symmetric state referenced with respect to the position of PV contours as the *modified Lagrangian mean* (MLM) state.

The latitude of a PV contour, with value Q, on an isentropic surface, with potential temperature Θ, in the background state will be described as its 'equivalent latitude', $\phi_e(Q,\Theta)$.* The concept of equivalent latitude was introduced by Butchart and Remsberg (1986) and developed by Norton (1994) to examine the stratospheric polar vortex. These authors defined ϕ_e by requiring PV contours to contain the same area under rearrangement to zonal symmetry. However, since isentropic density varies on isentropic surfaces, this form of rearrangement is not conservative, since it implies a change in the mass of layers. Nakamura (1995) used the mass integrals to describe a meridional coordinate that was mapped to equivalent latitude by assuming constant isentropic density within each layer. He also derived MLM evolution equations describing changes in the MLM PV and zonal flow (his equations 2.12 and 2.13). In particular, he showed that the effects of non-conservative processes can be represented as an equivalent advection by a meridional circulation without eddy flux terms. A drawback of the Nakamura (1995) approach is that uniform density in each isentropic layer is assumed when mapping mass from (Q,Θ) coordinates into physical coordinates (ϕ,z). While this approximation works quite well in the stratosphere, it is poor crossing the tropopause and throughout the troposphere.

Here, the definition will be such that both the mass and circulation enclosed by $\phi_e(Q,\Theta)$ in the background state must equal the corresponding mass, $\mathcal{M}(Q,\Theta)$, and circulation, $\mathcal{C}(Q,\Theta)$, integrals from the full 3D state. Since the integral properties are the same and the only difference is the geometry of PV contours on each isentropic surface, such a state is described as an 'adiabatic rearrangement' of the flow. The difficulty in the procedure is that both integral constraints must be respected, determining both the MLM vorticity and isentropic density distributions and therefore the location of every PV contour in (ϕ,z) or (ϕ,p) coordinates (assuming hydrostatic and gradient wind balance on the sphere). The mass and vorticity distributions are linked through the PV inversion relation. The approach taken is to iterate towards a set of equivalent latitudes that satisfies both constraints, using PV inversion to define the partition between vorticity and isentropic density anomalies.

Rossby-wave disturbances in the atmosphere can be described in terms of meridional displacements of PV contours along isentropic surfaces from their equivalent latitudes in the zonally symmetric background state. In an adiabatic, frictionless fluid, these would be the same as air parcel displacements from a notional 'undisturbed' reference position. The set of reference locations could be interpreted as describing a Lagrangian mean state. The atmosphere is not adiabatic or inviscid and air parcels mix. These effects result in mass crossing isentropic surfaces and the PV contours on them. However, it is still possible to identify disturbances with the positions of PV contours. This MLM approach (McIntyre, 1980) has several advantages over a Lagrangian mean, where quantities are averaged over selected material volumes, 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. Solomon and Nakamura (2012) have recently reviewed the relation between Lagrangian and MLM definitions of wave activity.

A major advantage of the MLM definition is that PV can be calculated as a diagnostic from the data without a time-dependent tracer calculation. In this study, ERA-Interim analyses (Dee *et al.*, 2011) will be used to provide the atmospheric data required. The European Centre for Medium-Range Weather Forecasts (ECMWF) model cannot represent tracer structures as fine as those observed, so an inevitable side effect is that numerical mixing will transfer mass across PV contours. However, the quality of ERA-Interim data has been found to be comparable to other contemporary reanalyses and considerably better than older reanalyses. For example, Simmons *et al.* (2014) report improvement in the variability and trends of temperature and Berrisford *et al.* (2011) found better closure in the global budgets of mass, moisture, energy and angular momentum. This improvement in data quality is thought to be due to the use of 4D-Var, variational bias correction of satellite data, improved observational datasets and other improvements to the forecast model and assimilation system.

An important attribute of the new background-state definition is the treatment of the lower boundary. If the flow were adiabatic, the potential temperature would be materially conserved following the *horizontal flow* along the lower boundary. Therefore, just as the MLM framework tracks PV contours along isentropic surfaces in the fluid interior, it is natural to track θ contours along the lower boundary. In practice, there is strong mixing in the atmospheric boundary layer, so θ contours approximately follow the horizontal flow along the top of the boundary layer. A novel aspect is that the background-state lower boundary is determined as part of the solution and defines a set of equivalent latitudes for disturbances near the Earth's surface.

Wave-activity conservation laws are derived from conserved properties of the full flow: for example, energy or zonal angular momentum. In general, these properties are not conserved by the perturbation alone, because there is '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. These additional invariants are called Casimirs. McIntyre and Shepherd (1987) outlined a systematic approach to finding wave-activity conservation laws by combining 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. The MLM background state meets these criteria. Since it is zonally symmetric, a conservation law can be obtained by the zonal angular momentum–Casimir method following Haynes (1988). Many studies involving wave activity have been theoretical, applied to idealized models or atmospheric data with approximations (such as small amplitude or quasi-geostrophic expressions). The key aim here is to apply the full diagnostics in the examination of global meteorological analysis data. Preliminary results were presented in Methven (2009) and the derivation of finite-amplitude expressions for pseudomomentum, pseudoenergy and their calculation from data was presented in Methven (2013). This article describes the method for obtaining the background state that was used in those studies.

Nakamura and Solomon (2011) have also devised wave-activity calculations valid at large amplitude to study wave–mean flow

*Throughout, subscript 'e' will be used to denote quantities defined at equivalent latitudes.

interaction throughout the atmosphere using analyses. Similar to the approach here, they used the functionals $\mathcal{M}(Q,\Theta)$ and $\mathcal{C}(Q,\Theta)$ plus gradient wind and hydrostatic balance to define a zonally symmetric 'reference state'. One key difference is that they use the Eulerian zonal mean potential temperature on a pressure surface to specify the lower boundary of their background state. This is not consistent with an adiabatic rearrangement, such that the same mass and circulation are enclosed within isentropic layers that intersect the Earth's surface. In the new method presented here, a lower boundary is found that is consistent with mass and circulation. Consequently, it is possible to define wave activity associated with displacements of potential temperature contours along the lower boundary (following Magnusdottir and Haynes, 1996), which are an essential component of baroclinic wave development in the extratropics. The boundary-wave activity has also been shown to dominate variability on the large-scale, such as the periods of eastward and westward phase propagation of Rossby waves at the tropopause (Methven, 2013).

The background state is defined and the calculation method described in section 2, including the integral properties, balance relations, boundary conditions used for the PV inversion and iteration of equivalent latitudes to obtain consistency with the mass and circulation of the atmospheric data. The numerical methods are outlined in the Appendix. Section 3 summarizes the definition of pseudomomentum and its flux, following Haynes (1988). Readers who are more interested in the applications than in the details of the calculation can skip to section 4.

Application of the methods is illustrated for three problems: (i) the background state and wave activity in the troposphere, (ii) wave–mean flow interaction during sudden stratospheric warmings (SSWs) and (iii) troposphere–stratosphere interaction. A dynamical phenomenon central to all three is the *vortex erosion mechanism* (Legras and Dritschel, 1993). Rossby-wave breaking stirs PV on isentropic surfaces, resulting in stretching and folding of PV contours and the formation of PV filaments. Wave breaking tends to act preferentially outwards on the edge of a large-scale vortex, so that mass is carried away from the vortex edge in the filaments and is transferred irreversibly into the surrounding surf zone, where the fine-scale filaments mix with their surroundings. The net effect is to enhance the PV gradient of the vortex edge in the background state and to reduce the gradient in the surf zone. The vortex is typically more compact and axisymmetric as a result (Melander *et al.*, 1987). The more compact vortex and sharper PV gradient are associated with a stronger jet around the vortex. The MLM framework is ideally suited to pulling out this behaviour from the data, as illustrated in section 4 using ERA-Interim data for a six-month period (1 November 2009–1 May 2010).

SSWs are perhaps the most extreme example of the vortex erosion mechanism in the atmosphere. Section 5 focuses on the stratospheric polar vortex in the build-up and during SSWs and on the links with wave activity in the troposphere. The vertical profile of pseudomomentum is used to distinguish periods dominated by barotropic modal structures (Esler and Scott, 2005) or by the upward group propagation of Rossby-wave packets. The mechanism for the repeated downward migration of wave activity that occurs during SSWs is also investigated. The new diagnostics point to the importance of wave dissipation in the mechanism and in determining the rate of downward migration. Conclusions and the potential for future applications are described in section 6.

2. Definition and calculation of the background state

2.1. Mass and circulation integrals

The background state is defined in terms of volumes of integration bounded above and below by isentropic surfaces, with fixed separation $\Delta\theta$, and laterally by an Ertel PV contour $q = Q$, such that only higher PV is contained:

$$\mathcal{M}(Q,\Theta) = \frac{1}{\Delta\theta}\iiint_{q\geq Q} ra^2\,\mathrm{d}\lambda\,\mathrm{d}\mu\,\mathrm{d}\theta, \tag{1}$$

$$\mathcal{C}(Q,\Theta) = \frac{1}{\Delta\theta}\iiint_{q\geq Q} \zeta a^2\,\mathrm{d}\lambda\,\mathrm{d}\mu\,\mathrm{d}\theta, \tag{2}$$

where a is the Earth's radius, λ is longitude, $\mu = \sin\phi$ and ϕ is latitude. $\theta = T(p_{00}/p)^{\kappa}$ is potential temperature, where T is temperature, p_{00} is a constant reference pressure (1000 hPa), $\kappa = R/c_\mathrm{p} = 2/7$, R is the gas constant for dry air and c_p is its specific heat capacity at constant pressure. Θ is the value of θ labelling the midpoint of the isentropic layer. r is the isentropic density (9) and ζ is the vertical component of absolute isentropic vorticity (4). $\mathcal{M}$ is the mass divided by the isentropic surface separation, $\Delta\theta$. $\mathcal{C}$ is the circulation around the volume. The integration volumes include any PV cut-offs. Ertel PV is defined under the approximations of the hydrostatic primitive equations by

$$q = \frac{\zeta}{r}. \tag{3}$$

The vertical component of absolute vorticity on a sphere rotating with angular velocity Ω is given by

$$\zeta = 2\Omega\mu - \frac{\partial U}{\partial\mu} + \frac{1}{1-\mu^2}\frac{\partial V}{\partial\lambda}, \tag{4}$$

where the derivatives of the horizontal wind are taken along isentropic surfaces. Here $U = (u/a)\cos\phi$, $V = (v/a)\cos\phi$, where (u, v) is the horizontal velocity.

The ERA-Interim reanalysis data used here are obtained from the analyzed fields archived on model levels. Since the ECMWF model is pseudo-spectral, the data are taken in spherical harmonic form and transformed on to the full Gaussian grid. The model levels are defined in a terrain-following hybrid-pressure vertical coordinate, often described as η-levels (Simmons and Burridge, 1981). Pressure is calculated at every location from the surface pressure, $p_\mathrm{s}(\lambda,\mu,t)$, and the coordinate definition. Temperature and pressure together define the potential temperature $\theta(\lambda,\mu,\eta,t)$. The data are then interpolated on to a finite set of isentropic surfaces labelled θ_m. Centred finite differences in isentropic coordinates are used to calculate the relative vorticity. The method used to calculate isentropic density, r, is described in the next section. Further details on the numerical methods are given in the Appendix.

2.2. Balance relations for the zonally symmetric state

The background state is defined in terms of the integrals $\mathcal{M}(Q,\Theta)$ and $\mathcal{C}(Q,\Theta)$. The zonal flow can be associated with the background state circulation by using Stokes' theorem applied to a zonally symmetric state:

$$\mathcal{C}(Q,\Theta) = 2\pi a^2\left[U + \Omega(1-\mu^2)\right]. \tag{5}$$

A subtlety with this equation is that it is necessary to know the equivalent latitude of every PV contour $\mu(Q,\Theta)$ to find $U(Q,\Theta)$. The method starts from first-guess equivalent latitudes specified by the area enclosed by PV contours in the 3D state:

$$\mu_\mathrm{e}^1(Q,\Theta) = 1 - \frac{\mathcal{A}(Q,\Theta)}{2\pi a^2}, \tag{6}$$

where the area $\mathcal{A}(Q,\Theta)$ is defined with isentropic density set to 1 in the integrand of the mass integral (1). However, unless the isentropic density is uniform in each isentropic layer, the background-state contour $q = Q$ will not contain the same mass

or circulation as the PV contour $q = Q$ from the 3D state. The full problem requires a simultaneous solution for the isentropic density and velocity field of the background state and this is achieved through a new technique called *equivalent latitude iteration with PV inversion* (ELIPVI for short).

In isentropic coordinates, hydrostatic balance and the horizontal pressure gradient force are best described in terms of the Montgomery potential, M, defined by

$$M = c_p T + gz. \tag{7}$$

Starting from the first variation of (7), followed by hydrostatic balance in height coordinates, $\mathrm{d}p = -\rho g\,\mathrm{d}z$ and the ideal gas law, $\rho = p/(RT)$, to eliminate the density, ρ, and the definition of θ, it is possible to obtain an expression for hydrostatic balance in θ coordinates (e.g. Bleck, 1973):

$$\frac{\partial M}{\partial \theta} = c_p \frac{T}{\theta} = c_p \left(\frac{p}{p_{00}}\right)^{\kappa} \tag{8}$$

and to express the isentropic density in terms of M:

$$r = -\frac{1}{g}\frac{\partial p}{\partial \theta} = -\frac{1}{\Gamma g}\frac{\partial^2 M}{\partial \theta^2}, \tag{9}$$

where the function

$$\Gamma = \frac{RT}{p\theta} = \frac{R}{p}\left(\frac{p}{p_{00}}\right)^{\kappa} = \frac{R}{p_{00}}\left(\frac{1}{c_p}\frac{\partial M}{\partial \theta}\right)^{(\kappa-1)/\kappa}. \tag{10}$$

Substituting the definitions for absolute vorticity (4) for a zonally symmetric flow and isentropic density (9) into the PV definition (3) gives

$$2\Omega\mu - \frac{\partial U}{\partial \mu} = -\frac{q}{\Gamma g}\frac{\partial^2 M}{\partial \theta^2}. \tag{11}$$

For a steady, zonal flow, in the absence of friction, the meridional component of the momentum equation reduces exactly to gradient wind balance on the sphere:

$$\left(2\Omega \sin\phi + \frac{U \sin\phi}{\cos^2\phi}\right) Ua + \frac{\cos\phi}{a}\frac{\partial M}{\partial \phi} = 0. \tag{12}$$

Gradient wind balance (12) and Stokes' theorem (5) can be used to replace $\partial U/\partial \mu$ in (11) with an expression involving only M and the circulation $\mathcal{C}$ (which is a known function in this problem), to give

$$1 + \frac{1}{(2\Omega a \sin\phi\cos\phi)^2}\left\{\frac{\left(1+\sin^2\phi\right)^3}{2\sin^2\phi\cos\phi}\frac{\partial}{\partial\phi}\left(\frac{\cos\phi}{1+\sin^2\phi}\frac{\partial M}{\partial\phi}\right) - \frac{\left(1+3\sin^4\phi\right)}{2\sin^2\phi\cos\phi}\frac{\partial}{\partial\phi}\left(\cos\phi\frac{\partial M}{\partial\phi}\right) + \frac{q\,\mathcal{C}(q,\theta)\sin\phi}{\Gamma g \pi}\frac{\partial^2 M}{\partial\theta^2}\right\} = 0. \tag{13}$$

PV inversion amounts to solving this equation for M, given q and $\mathcal{C}(q,\theta)$. It would be linear in M and elliptic, if it were not for the function Γ, which introduces weak nonlinearity. The PV inversion (13) is discretized using second-order finite differences and boundary conditions are required around the perimeter of the domain. This particular form (13), expressed in terms of ϕ, is used because it involves only second-order meridional derivatives of M and enables the boundary conditions to be applied more easily (section 2.2.1).

Note that Nakamura and Solomon (2011) combine (5), (10) and (12) to obtain an elliptic equation relating derivatives of $\mathcal{M}(\mu_e,\theta)$ and $\mathcal{C}(\mu_e,\theta)$, which also has nonlinearity introduced by Γ. They start from a first-guess given by the integral of mass poleward of fixed latitude circles, $\overline{\mathcal{M}}$, and use it to identify a PV value, Q, with matching mass functional $\mathcal{M}(Q,\theta) = \overline{\mathcal{M}}(\mu,\theta)$. This PV value is then used to find the corresponding value of the circulation functional $\mathcal{C}(Q,\theta)$. Their method proceeds by iterating mass and circulation simultaneously as functions of (μ,θ), rather than the approach here, where the functional values are preserved and the iteration is to the equivalent latitudes. The Γ function is also updated by an outer 'nonlinear iteration' (see section A9). Therefore, although the state they obtain should have both mass and circulation integrals consistent with gradient wind and hydrostatic balance, it is not guaranteed that these integrals converge on the functionals obtained from the original 3D data (although relation (27) constrains them to be close to the observed values). Also, since they use the Eulerian zonal average of θ on a pressure surface as the lower boundary condition, the mass integrals cannot match at lower levels (except in the special circumstance in which the θ distribution on the lower boundary obtained here matches the Eulerian zonal average).

The numerical method for inverting (13) is outlined in the Appendix and involves a linear inverter (assuming Γ is known) and a 'nonlinear iteration' to recalculate Γ and the lower boundary terms from the results and feed this back into the linear inverter. Once M is known, the zonal wind and isentropic density can be calculated from (12) and (9).

2.2.1. *Boundary conditions at the Equator and North Pole*

Since the method has been implemented for the Northern Hemisphere in spherical polar coordinates, it is necessary to specify lateral boundary conditions at the North Pole and Equator. At the Pole, $\cos\phi = 0$, so, for non-singular M,

$$\cos\phi\frac{\partial M}{\partial \phi} = 0 \tag{14}$$

and from (12) it can be seen that $u = 0$. At the Equator, using $\sin\phi = 0$ in (12) implies that (14) also applies there (for any equatorial flow). To solve the second-order equation (13), boundary conditions are required on the meridional derivative of $\cos\phi\,\partial M/\partial\phi$ at the Equator. To simplify presentation, the geostrophic wind, u_g, is introduced:

$$u_g = -\frac{1}{2\Omega a \sin\phi}\frac{\partial M}{\partial \phi}. \tag{15}$$

Rearranging (12) then yields

$$u_g = u\left(1 + \frac{u}{2\Omega a\cos\phi}\right). \tag{16}$$

On applying l'Hôpital's rule to (15) as $\phi \to 0$,

$$u_g(\phi = 0) = -\frac{1}{2\Omega a}\frac{\partial^2 M}{\partial \phi^2}. \tag{17}$$

A quadratic function is fitted to $M(\phi)$ approaching the Equator by using (17) with (14) at the Equator and assuming an additional condition of uniform curvature of M from the Equator to the junction between the first and second columns of grid boxes adjacent to the Equator (denoted by the latitude label $j2$). This gives

$$\left.\frac{\partial M}{\partial \phi}\right|_{j2} = -2\Omega a u_g(0)\phi_{j2} = -2\Omega a u_{eq}\left(1 + \frac{u_{eq}}{2\Omega a}\right)\phi_{j2}, \tag{18}$$

where the second equality uses (16) and the zonal wind at the Equator, u_{eq}, is specified from the circulation integrals spanning

the Northern Hemisphere on each isentropic surface using (5). This provides a near-equatorial lateral boundary condition, so that (13) can be solved in a way that correctly constrains the first and second derivatives of M along each isentropic surface at the Equator. Once M is known polewards of this boundary condition, M on the first column of grid points from the Equator is given by

$$M_1 = M_2 - \left.\frac{\partial M}{\partial \phi}\right|_{j2} (\phi_2 - \phi_1)\,, \tag{19}$$

where M_1, M_2, ϕ_1 and ϕ_2 are the values of M and latitude at the first and second points from the Equator on the inversion grid.

2.2.2. *Boundary condition at the top boundary*

The top boundary of the inversion domain is taken to be the 3043 K isentropic surface, $\theta_{\rm top}$, and the highest isentropic level used on the inversion grid is 2979 K. These isentropic surfaces were chosen so that they remain within the range of the ERA-Interim data. Note that the top model level in ERA-Interim corresponds to a pressure of 0.1 hPa. The pressure at $\theta_{\rm top}$ is used to specify the top boundary condition. However, the pressure in the background state is not known, since the aim is to solve for it, so the top boundary pressure must be specified from the 3D data. During winter, the horizontal flow and pressure perturbations (relative to the mean) are strong at 0.1 hPa, associated with the extension of the polar vortex into the lower mesosphere. However, throughout much of winter the polar vortex is approximately circularly symmetric, even when its centre is displaced from the North Pole. If the vorticity of the vortex is circularly symmetric (at all levels), then we can also expect its pressure field to be concentric about its centre.

The approach taken is to identify the polar vortex centre as the maximum in the vertical average of PV across the top third of θ levels (mid-stratosphere and above). The native data grid is transformed into spherical coordinates, where the polar axis points through the vortex centre. The pressure from the analysis data at the top boundary is then averaged within a set of latitude bands defined in this transformed coordinate, which determines the function $p_{\rm top}(\phi)$. The goal is that this is the most consistent boundary condition, with a vortex in the zonally symmetric state that is being sought.

The function $p_{\rm top}(\phi)$ determines $\partial M/\partial\theta$ according to (8), thereby providing the top boundary condition for (13).

2.2.3. *Boundary condition at the lower boundary*

The lower boundary taken for the input to the calculation is the second terrain-following model level above the Earth's surface in the ERA-Interim data ($\eta = 0.99586$). For a typical atmospheric profile, this corresponds to 30 m above the surface. The lowest level (≈ 10 m) was not used because very stable nocturnal boundary layers, especially during the polar winter, introduce very low isentropic density and high magnitude PV anomalies (of both signs), which disrupted the attempt to calculate a balanced state. Recall that friction is not being included in the momentum equation of the background state (12) and so it could be more consistent to take a lower boundary coincident with the top of the planetary boundary layer. However, taking a higher η level for the lower boundary excludes a larger proportion of atmospheric mass from the background. The level chosen was found to obtain the most consistent results in terms of temporal variation of the lower boundary of the background state.

For isentropic surfaces intersecting the lower boundary, the total mass and circulation enclosed by the isentropic layer of depth $\Delta\theta$ (centred on the θ surface) are obtained by integration. These integrals are functions of surface θ only and will be denoted as $\mathcal{M}_{\rm s}(\Theta)$ and $\mathcal{C}_{\rm s}(\Theta)$. They are used to define the equivalent latitudes of the intersection of each isentropic surface with the lower boundary. As will be described in section 2.3, this defines the position of the lower boundary on the inverter grid, $\theta_{\rm LB}(\mu_j)$.

For the inversion of (13), the lower boundary condition and the intersection of isentropic surfaces with the surface are treated following Bleck (1973, 1974). Combining (7) with (8) gives

$$M = gz + \theta\frac{\partial M}{\partial \theta}. \tag{20}$$

A quadratic curve ($M = a + b\theta + c\theta^2$) is fitted in the vertical through the two lowest isentropic levels above the lower boundary, subject to the constraint that at the lower boundary

$$M = gz_{\rm LB} + \theta_{\rm LB}\frac{\partial M}{\partial \theta}, \tag{21}$$

where $z_{\rm LB}$ and $\theta_{\rm LB}$ are the geopotential height and potential temperature on the lower boundary at position (ϕ, t). See section A2 for the way in which $z_{\rm LB}$ is calculated from the 3D data. This provides a mixed lower boundary condition for (13).

The same quadratic fit is used to extrapolate M in the vertical to isentropic levels below the lower boundary. This enables horizontal derivatives of M to be evaluated in (13) and (12) at interior points adjacent to the intersection of isentropic surfaces with the lower boundary.

Once M is known, together with the boundary condition (21), this quadratic curve also defines pressure, $p_{\rm LB}$, and density, $\rho_{\rm LB}$, at the lower boundary of the background state. The geostrophic wind ($u_{\rm gLB}$) is given by

$$u_{\rm gLB} = -\frac{1}{2\Omega a\sin\phi}\left(\frac{1}{\rho_{\rm LB}}\frac{\partial p_{\rm LB}}{\partial\phi} + g\frac{\partial z_{\rm LB}}{\partial\phi}\right) \tag{22}$$

and (16) is then used to obtain the full zonal wind, $u_{\rm LB}$.

2.3. *Iteration of equivalent latitudes*

The MLM state is defined as the zonally symmetric state with PV $q_{\rm o}(\mu,\theta)$ and lower boundary potential temperature $\theta_{\rm os}(\mu)$ that has the same integral properties $\mathcal{M}(Q,\Theta)$, $\mathcal{C}(Q,\Theta)$, $\mathcal{M}_{\rm s}(\Theta)$, $\mathcal{C}_{\rm s}(\Theta)$ as the full 3D state. The zonal wind and isentropic density are obtained by inverting the background PV distribution using the balance relation (13) with appropriate boundary conditions.

In practice, the area, mass and circulation integrals are calculated from the 3D state for a discrete number of PV values, Q_k, and isentropic layers, $\theta_{\rm m}$. The first-guess equivalent latitudes $\mu_{\rm e}^1(Q_k,\theta_{\rm m})$ are calculated from (6). First-guess PV $q_{\rm o}^1(\mu_j,\theta_{\rm m})$ on the inverter grid, μ_j, is obtained by inverting the functional relation by interpolating PV from values Q_k at irregular latitude points $\mu_{\rm e}(Q_k)$ to the regular grid points μ_j (section A5).

The first-guess PV distribution is then inverted using (13) to obtain the Montgomery potential and hence temperature, pressure, isentropic density and zonal flow. These are used to calculate the mass and circulation integrals of the 2D state. Since density is positive-definite, the mass integrals decrease monotonically with equivalent latitude:

$$\frac{\partial\mathcal{M}}{\partial\mu} = \frac{\partial\mathcal{M}}{\partial\mathcal{A}}\frac{\partial\mathcal{A}}{\partial\mu} \approx -r_{\rm o}2\pi a^2, \tag{23}$$

where the definition of the integrals yields $\partial\mathcal{M}/\partial\mathcal{A} = r_{\rm o}$, where $r_{\rm o}$ is the isentropic density of the MLM state. The $2\pi a^2$ factor assumes that area integrals from the MLM state will approximately equal those from the 3D state, so that (6) can be used. This gives a direction for the update of equivalent latitude:

$$\delta\mu_{\rm e}^M(Q,\theta) = -\left(\mathcal{M}_{\rm 2D} - \mathcal{M}\right)\frac{\partial\mu}{\partial\mathcal{M}}\mu_{\rm e}^{1/4}. \tag{24}$$

The factor $\mu_e^{1/4}$ is included to control steps nearer the Equator, where the PV inversion is most sensitive. If the state is to be stable to symmetric perturbations, the PV must be positive in the Northern Hemisphere and therefore, from (2), the circulation also decreases monotonically with Q and equivalent latitude. Differentiation of (5) gives

$$\frac{\partial \mathcal{C}}{\partial \mu} = 2\pi a^2 \left(\frac{\partial U}{\partial \mu} - 2\Omega\mu \right), \tag{25}$$

which can be inserted into an alternative formula for the equivalent latitude update:

$$\delta\mu_e^C(Q,\theta) = -\left(\mathcal{C}_{2D} - \mathcal{C}\right) \frac{\partial \mu}{\partial \mathcal{C}} \mu_e^{1/4}. \tag{26}$$

Equal weighting is given to the mass and circulation estimates of the update and the iteration proceeds until the equivalent latitude updates across the domain fall below a chosen tolerance. The numerical details on the equivalent latitude iteration procedure can be found in section A10. Thuburn and Lagneau (1999) showed that the variation of mass with PV value along isentropic surfaces is related to circulation by

$$Q\frac{\partial \mathcal{M}}{\partial q}(Q) = \frac{\partial \mathcal{C}}{\partial q}(Q) \tag{27}$$

and this is also true for the background state when the iteration has converged. Therefore, it is not important whether mass or circulation is used to obtain the state, but convergence is found to be fastest if both are used.

A similar procedure is used to iterate on the equivalent latitudes of θ_m contours along the lower boundary. However, $\partial\mathcal{M}/\partial\mu$ along the lower boundary cannot be calculated from (23) because the derivative there is taken along isentropic surfaces. However, (25) still applies and can be used to define the shifts in surface equivalent latitudes, μ_{es}, using (26).

The new method described in this section is named *equivalent latitude iteration with PV inversion* (ELIPVI). In addition to the adiabatic rearrangement of the flow to zonal symmetry using the ELIPVI procedure outlined above, a number of additional operations are required to deal with features of atmospheric data that do not fit naturally with the concept of a zonally symmetric background state. These include orography, negative PV near the Equator and short-lived PV anomalies associated with non-conservative processes. These will be described in the Appendix with reference to the results from the background-state calculation.

3. Measuring large-amplitude wave activity relative to the background state

Wave-activity density is a measure of the amplitude of the difference between a perturbed flow and a suitable reference flow. It is defined to be second-order in disturbance quantities, in order to represent an amplitude. The definition is chosen so that the wave-activity density, P, obeys a local conservation law:

$$\frac{\partial P}{\partial t} + \frac{1}{a\cos\phi}\frac{\partial F^{(\lambda)}}{\partial \lambda} + \frac{1}{a\cos\phi}\frac{\partial}{\partial\phi}\left(F^{(\phi)}\cos\phi\right) + \frac{\partial F^{(\theta)}}{\partial\theta} = S, \tag{28}$$

where $(F^{(\lambda)}, F^{(\phi)}, F^{(\theta)})$ are the components of wave-activity flux in isentropic spherical coordinates and 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.

The definition of the reference state is vital to the existence of an exact wave-activity conservation law at finite perturbation amplitude. It is essential to describe the reference state as a function of PV and θ in order to use the Casimir method (see section 1). If the reference state is also zonally symmetric (i.e. the MLM state defined in this article), the pseudomomentum conservation law is obtained by the zonal angular momentum–Casimir method. If the reference state is steady (time-symmetric), the pseudoenergy conservation law is obtained by the energy–Casimir method. Following Haynes (1988), the pseudomomentum density (within isentropic layers that do not intersect the lower boundary) is given by

$$P(\lambda,\phi,\theta,t) = P_w - r'u'\cos\phi, \tag{29}$$

where term P_w is associated with displacements of PV contours on isentropic surfaces and is therefore called the 'Rossby-wave term' (e.g. Methven *et al.*, 2005). An exact integral form[†] is given by

$$2\pi a P_w = -rq\left[\mathcal{M}\right]_{q_o}^{q} + r\left[\mathcal{C}\right]_{q_o}^{q}, \tag{30}$$

following Thuburn and Lagneau (1999) and Methven (2013), where $q(\lambda,\phi,\theta,t)$ is the Ertel PV of the full state, $q_o(\phi,\theta,t)$ denotes the background-state PV and the perturbation $q' = q - q_o$ is defined as the difference between the full PV and the background state at a point on a given isentropic surface. $[\mathcal{M}]_{q_o}^{q}$ stands for $\mathcal{M}(q,\theta_m) - \mathcal{M}(q_o,\theta_m)$, where, at each location (λ,ϕ,θ_m), q and q_o are found from the full state and background state and used as arguments in the functional $\mathcal{M}(Q,\Theta)$. The same procedure is applied to the circulation functional $\mathcal{C}$. This expression is valid for arbitrary disturbances with any amplitude or shape in PV contours and obeys a conservation law with the form (28). At small amplitude, the term reduces to $P_w = \frac{1}{2}r_o Q_y \eta^2$, where the meridional parcel displacement $\eta = -q'/(\partial q_o/\partial y)$ and

$$Q_y = r_o\cos\phi\frac{1}{a}\frac{\partial q_o}{\partial\phi}. \tag{31}$$

Q_y is the form of meridional PV gradient that acts in place of β from the familiar quasi-geostrophic β-plane setting. P_w is therefore positive-definite, since the background meridional PV gradient is positive.

The $-r'u'\cos\phi$ term is often called the 'gravity-wave term', although it could also arise in large-scale balanced flows. In practice, the 'Rossby-wave' term P_w is greater than the 'gravity-wave' term locally and by several orders of magnitude in the integral over the Northern Hemisphere.

On isentropic surfaces that intersect the lower boundary, there is an additional 'boundary contribution' to pseudomomentum that is chiefly negative and discussed briefly in section 4.2. Methven (2013) presents the derivation of equations used to calculate the various terms in the large-amplitude pseudomomentum. The pseudomomentum flux (see Haynes, 1988 for a derivation) is given by

$$\begin{aligned} F^{(\lambda)} &= uP - \frac{r_o}{2}\left(u'^2 - v'^2\right)\cos\phi - a\cos\phi\,\tilde{\tau}(p_o,p',\theta), \\ F^{(\phi)} &= vP - r_o u'v'\cos\phi, \\ F^{(\theta)} &= \frac{p'}{ga}\frac{\partial M'}{\partial\lambda}, \end{aligned} \tag{32}$$

where the function of pressure $\tilde{\tau}$ is defined by (3.12c) of Haynes (1988). If the atmosphere is assumed to be an ideal gas and $|p'|/p_o \ll 1$, then it can be shown that $\tilde{\tau} \approx c_p\kappa/(p_o g)(p_o/p_{00})^\kappa p'^2$. Note that the meridional and vertical

[†]Note that the opposite sign has been used in Haynes (1988) and Magnusdottir and Haynes (1996).

components have the same expression as the Eliassen–Palm flux in isentropic coordinates (Andrews, 1983), but with the important addition of the meridional advection of pseudomomentum density, vP. The horizontal advection of pseudomomentum often dominates the nonlinear flux (Magnusdottir and Haynes, 1996) and has a major bearing on the interpretation of the results.

Both the pseudomomentum and the components of the flux can be calculated at every point in the 3D domain (λ, ϕ, θ) at any time. However, in all the figures shown the information will be summarized on the meridional plane (ϕ, θ) by taking zonal averages. Nevertheless, it is important that all the perturbations have been defined relative to the MLM state, not the Eulerian zonal mean state. Throughout the remainder of the article, the 'pseudomomentum density', P, will be abbreviated to 'pseudomomentum' and should not be confused with its volume integral.

4. Results illustrated for the troposphere in winter

4.1. Background-state structure

The background state is illustrated for a snapshot during the Northern Hemisphere winter in Figure 1. It is completely specified by the distribution of PV on isentropic surfaces, $q_o(\phi, \theta)$, plus the boundary conditions for PV inversion (13). These include the position of the lower boundary, $\theta_{os}(\phi)$, the pressure on the isentropic top boundary and the zonal flow at the Equator (obtained from hemispheric circulation integrals). Since the isentropic density is positive-definite, the mass integrals (1) must decrease as Q increases along an isentropic surface. When rearranged to obtain the zonally symmetric background, this means that PV must increase monotonically on every isentropic surface (such that higher value PV contours enclose less mass on their northern side). Furthermore, θ_{os} also decreases monotonically towards the pole, as indicated by the negative slope of the lower boundary.

The mass and circulation integrals used to construct the state are defined for discrete PV values, Q_k, and isentropic levels, θ_m. In the troposphere, the isentropic levels have a regular separation, $\Delta\theta$, of 1.5 K, while the level spacing is stretched above 320 K (see section A1). The inversion method requires this relatively small $\Delta\theta$ in order to cope accurately with the lower boundary, since $\Delta\theta$ determines the intervals in latitude between the locations where the isentropic surfaces intersect the lower boundary. Furthermore, the stratification is much weaker in the troposphere than the stratosphere, so small $\Delta\theta$ is required to resolve tropospheric structures in the background state and perturbations. The change in stratification across the tropopause also means that Q_k values need to be set closer together in the troposphere. Figure 1(a) shows the PV contours used in the calculation. The contour interval is regular in Ertel PV (1 PVU) above 5 PVU. The contour interval is regular in log(PV) from 0 to 1 PVU. The PV-level spacing blends between the two sets over the interval 1–5 PVU. The troposphere and lower stratosphere have similar latitude spacing between Q_k contours, although with closer spacing across the tropopause, where PV gradients are strongest, and wider spacing poleward of the tropopause on these isentropic surfaces.

PV inversion of this zonally symmetric distribution is used to obtain the Montgomery potential, M, and hence isentropic density, r_o, and zonal wind, u_o, from vertical (9) and horizontal (12) derivatives of M. Figure 1(b) shows that the isentropic density falls rapidly across the tropopause, indicating the rapid increase in static stability. Note that the 2 PVU contour of the background state closely follows the sharp gradient in isentropic density in the extratropics. At approximately 30°N, there is a subtropical tropopause jump where the 2 PVU contour is almost vertical and the gradient in density almost horizontal (for example along 350 K). Equatorward of the jump, PV contours are no longer coincident with the tropopause, which can be identified with the vertical gradient in r_o at about 360 K.

The MLM zonal flow, u_o, is strongest in the subtropical jet on the tropopause jump (Figure 1(c)). The strong vertical wind shear in the troposphere below is in balance with the mass field as described by the PV inversion relation (13). If the state is transformed into pressure coordinates, the meridional gradient of θ along pressure surfaces is in gradient thermal wind balance with $\partial u_o/\partial p$. Consequently, the θ gradient is large below the subtropical jet and the lower boundary has a steep slope in θ coordinates. At the time shown, there is a second jet at 65°N, which could be described as the tropospheric polar jet.

Some features of the background state arise from operations required to define the state from global data, in addition to the adiabatic rearrangement to zonal symmetry. For example, it does not make sense to rearrange the orographic height of the lower boundary or θ anomalies associated with elevated orography and these aspects are dealt with in section A2. Small-scale PV anomalies in the boundary layer and high-latitude lower troposphere are also not regarded as part of the background state and therefore are removed before calculating the mass and circulation integrals (section A3). Nevertheless, after finding the first-guess equivalent latitudes, there is typically a very narrow spike of high PV at the pole, associated with the rearrangement of the highest PV values to the North Pole in the monotonic background state. These high values typically occur at midlatitude fronts and are not regarded as part of the background; they are removed by truncating the spike in the first guess (section A6). Similarly, the meridional gradient in boundary θ must tend to zero approaching the Equator to obtain a balanced background state; this is achieved by modifying the first guess. Finally, any negative PV in the Northern Hemisphere background state is removed by setting PV to zero in those locations, as described in section A7.

4.2. Wave-activity amplitude in the troposphere

The evolution of perturbations, defined as the difference between the full 3D state and the modified Lagrangian mean at every point (λ, μ, θ), can be described by the pseudomomentum conservation law (28). In the absence of diabatic or frictional processes, pseudomomentum is a globally conserved property of the perturbations. The conservation relation provides a powerful way to view nonlocal transfer of wave activity throughout the atmosphere in terms of a pseudomomentum flux (32) and the local increase in pseudomomentum associated with flux convergence. The nonconservative term S is important to pseudomomentum (for example in wave dissipation) and will be examined in section 5.4.

The pseudomomentum is defined by (29), where the isentropic surface, θ_m, is above the lower boundary at all longitudes around a latitude circle in the full 3D state. The set of these latitude circles from all isentropic levels will be referred to as the *interior domain*. The structure of pseudomomentum in the meridional plane is summarized by presenting its zonal mean in Figure 1(d). The interior pseudomomentum takes the sign of the background-state meridional PV gradient in the limit of small disturbance amplitude. Since the MLM PV gradient is everywhere positive, the pseudomomentum is expected to be positive almost everywhere, as seen in the figure. The interior pseudomomentum is focused on the location of the extratropical tropopause in the background state, with the largest values just on the stratospheric side at high latitudes in the vicinity of the 'polar jet' seen in Figure 1(c). The pseudomomentum is much weaker in the vicinity of the subtropical jet at the time shown. This is consistent with the distinction between the subtropical jet and 'eddy-driven jet' further poleward (Lee and Kim, 2003).

Figure 1(e) shows the 'boundary contribution' to pseudomomentum associated with the meridional displacement of θ contours along the lower boundary. The contribution covers

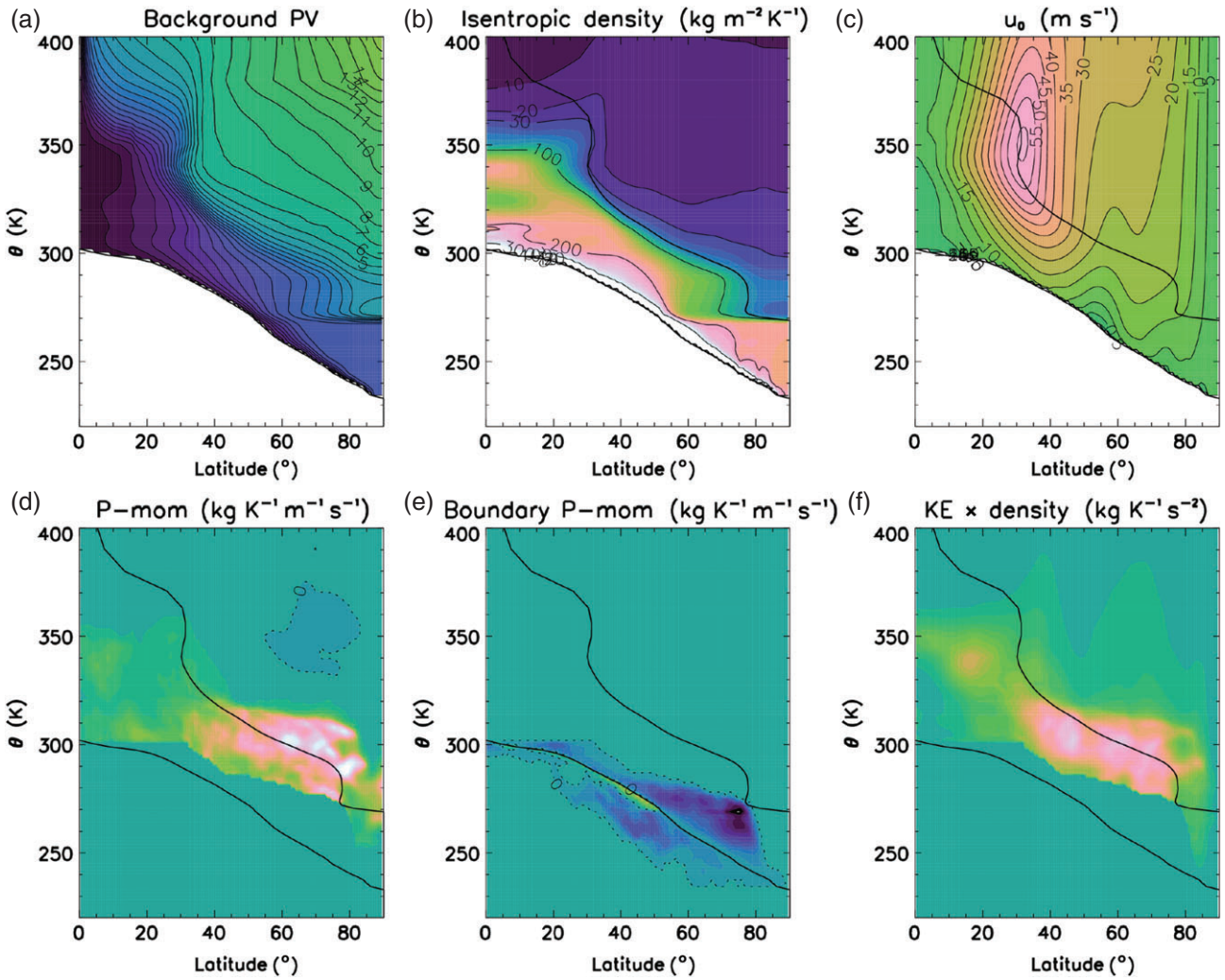

Figure 1. Partition of the atmospheric state at 1800 UTC on 22 January 2010 into the MLM background state and the perturbation from it, focusing on the troposphere. (a) Background-state Ertel PV, q_o (PVU) –the contour interval (CI) is not uniform below 5 PVU (see text). (b) MLM isentropic density, r_o (CI = 10 kg m^{-2} K^{-1}). (c) MLM zonal wind, u_o. (d) Interior pseudomomentum (CI = 5000 kg K^{-1} m^{-1} s^{-1}). (e) Boundary contribution to pseudomomentum from the *exterior domain* (CI as in (d)). (f) Perturbation kinetic energy times isentropic density (CI = 20 000 kg K^{-1} s^{-2}). In all panels, the lower bold curve is the lower boundary of the background state; the upper bold curve is its q_o = 2 PVU contour (tropopause).

much of the extratropical lower troposphere, because the range of latitudes reflects the large meridional excursion of each θ contour along the lower boundary. For example, the contour $\theta_s = 273$ K makes excursions between 25 and 75°N in the 3D state. The calculation follows the method of Magnusdottir and Haynes (1996). The non-zero values of pseudomomentum below the location of the lower boundary in the background state are associated with points that are above the lower boundary in the 3D state. Since these points are *exterior* to the background state range of θ at that latitude, Methven (2013) described this contribution as the *exterior term*, P_e. The non-zero values of 'boundary pseudomomentum' in Figure 1(e) above the background-state lower boundary are associated only with points (λ, ϕ, θ) that are 'below ground' in the full 3D state. Therefore this contribution is also described as part of the 'exterior term' P_e. In the small-amplitude limit, P_e takes the sign of absolute vorticity times the meridional θ gradient along the lower boundary of the background state (see Methven, 2013), which is negative-definite for the modified Lagrangian mean.

It is readily shown that growing normal modes must have zero global pseudomomentum, in order to permit exponential growth in meridional parcel displacements while also conserving pseudomomentum (Held, 1985). Baroclinically unstable normal modes are characterized by $P_e = -P_w$. However, Methven (2013) has shown that in atmospheric data the positive interior contribution far outweighs the negative exterior contribution in the hemispheric integral. Reasons for this are not known precisely. However, in nonlinear life cycles the boundary-wave activity saturates first as surface cyclones reach the mature phase, wrapping air masses cyclonically, while the tropopause-level pseudomomentum continues to amplify for several days. The net effect of many such baroclinic growth events is likely to result in more wave activity at tropopause level than in the boundary contribution. In addition, surface fluxes lead to a much more rapid dissipation of lower boundary thermal anomalies than the rate of dissipation of interior PV anomalies.

There are also additional terms in the large-amplitude expression for pseudomomentum that are not shown here. In addition to the 'gravity-wave term' $-r'u'\cos\phi$, which is relatively small, there is a contribution from the 'intersection domain' (denoted P_d in Methven, 2013) representing points that are above the lower boundary in the background state, but only above the lower boundary at some longitudes around the latitude circle in the full 3D state. This term is almost everywhere positive and partially cancels the negative contribution to P_e seen above the lower boundary of the background state in Figure 1(e) (it occupies the same region in (ϕ, θ) space).

Figure 1(f) shows disturbance kinetic energy (multiplied by density r) over the interior domain. Energy is not conserved for the disturbances alone. However, pseudoenergy is conserved for adiabatic, frictionless disturbances to a steady background state. Although the MLM state is not strictly steady, it will be shown that it varies very slowly. Note that the kinetic energy is calculated for disturbances defined relative to the MLM state. At the time

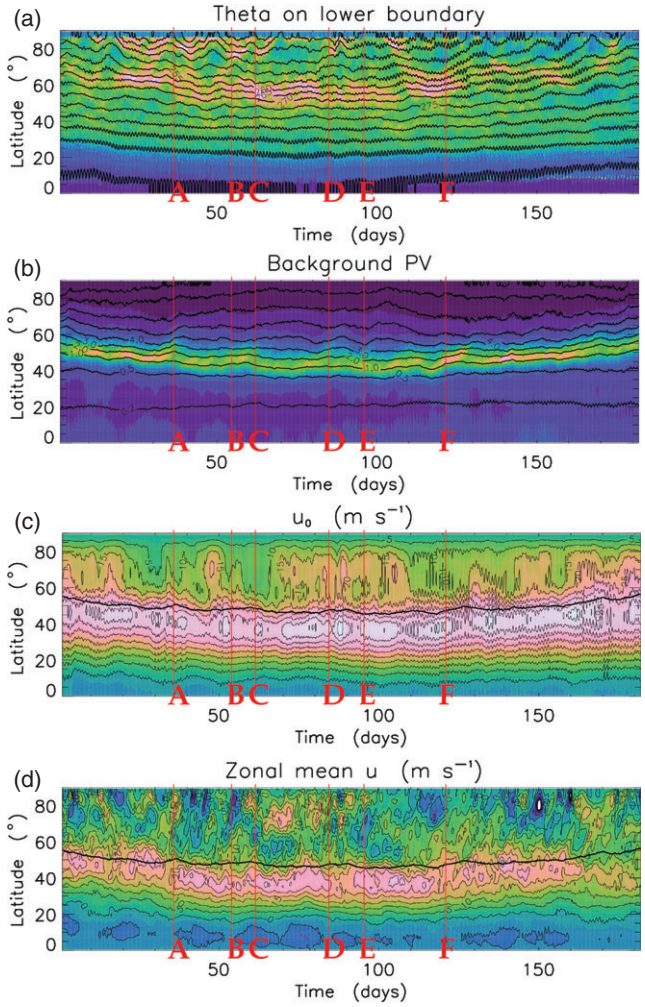

Figure 2. Evolution of the background state in the troposphere. Time runs from 0000 UTC on 1 November 2009 to 0000 UTC on 1 May 2010. (a) Potential temperature at the lower boundary of the background state (CI = 5 K) with gradient $-\partial\theta_o/\partial\phi$ in colour (CI = 0.1 K deg^{-1}). (b) PV at the 311 K surface, which crosses the tropopause in the midlatitudes with gradient Q_y in colour (CI = 10^{-11} m^{-1} s^{-1}). (c) MLM zonal wind (CI = 2.5 m s^{-1}) with the 2 PVU contour in bold. (d) Eulerian zonal mean wind (same scale). Labels A–F refer to events in pseudomomentum at the tropopause level (see Figure 3(b)).

shown, the structure of the kinetic energy is rather similar to the pseudomomentum, with the largest values near the extratropical tropopause.

4.3. *Background-state evolution in the troposphere*

The background state was obtained for analysis times (every 6 h) throughout the six-month extended winter period from 0000 UTC on 1 Nov 2009 to 0000 UTC on 1 May 2010. The troposphere is bounded by two key surfaces: the lower boundary and the tropopause. The lower boundary potential temperature $\theta_{os}(\mu)$ is defined from the equivalent latitudes where the isentropic levels, θ_m, intersect the lower boundary, $\mu_{es}(\theta_m)$. Each isentropic surface typically forms a cap over the North Pole, bounded by its intersection with the lower boundary in the 3D analysis. The tropopause poleward of about 30°N is characterized by the $q = 2$ PVU surface and the equivalent latitude of this contour, $\mu_e(Q, \Theta)$, describes its position in the background state.

The time series of potential temperature at the lower boundary of the background state, θ_{os}, is shown in Figure 2(a). The most striking feature is that the time series varies only slowly over the six-month period. The cold surface temperatures expand slowly equatorward until midwinter (event C is 1800 UTC on 31 December 2009) and then slowly retract polewards moving into spring (event F at 0000 UTC on 1 March 2010). It is important to note that no time averaging has been applied to the data. The background state has an inherently slow variation, because it can only change through non-conservative processes (diabatic, frictional or mixing effects). The implication is that the effects of these processes are weak, but nevertheless systematic on seasonal time-scales.

The small-amplitude, rapid oscillations in the contours are associated with the diurnal cycle (the analyses are every 6 h). They are most prominent in the lower troposphere at low latitudes and reflect a diurnal variation in the mass of isentropic layers in the Northern Hemisphere. The variation is associated with the large diurnal variation of near-surface potential temperature in the mountainous regions of the Northern Hemisphere that are high enough to outcrop through potential temperature surfaces (especially the Tibetan Plateau). To some extent, the effect is ameliorated by shifting the lower boundary vertically to a reference geopotential in the 3D state before calculating the mass and circulation integrals (section A2). However, there is still a substantial diurnal variation and in the rearrangement to obtain the monotonic θ_s profile of the background state, the fluctuations associated with 'hotter spots' on the Earth's surface are reflected near the Equator in the background state. They are readily eliminated by post-processing with a diurnal filter but one was not applied here.

The colour shading in Figure 2(a) reflects the meridional gradient of θ_{os} and highlights pronounced intraseasonal variability with time-scales of 5–30 days, which will be discussed in section 4.4 with relation to wave activity. Some relates to meridional motion of the θ_{os} contours, while some relates to meridional progression of the gradient polewards or equatorwards (e.g. before events A, B, C and F).

The PV on the 311 K isentropic surface is contoured in Figure 2(b) and the meridional PV gradient Q_y, defined by (31), is shown by colour shading. The isentropic density weighting in (31) is important, since it gives more weight to the tropospheric side of the tropopause, where the density is higher (Figure 1(b)). The PV gradient highlights the seasonal cycle in the position of the tropopause on this isentropic surface, associated with the expansion of the polar PV reservoir in the lowermost stratosphere moving into winter and its contraction again moving into spring. The dominant non-conservative process is long-wave radiative cooling, which has a characteristic time-scale in the upper troposphere of 35 days. Long-wave cooling from water vapour just below the tropopause results in a systematic maintenance of PV above the tropopause (e.g. Chagnon *et al.*, 2013).

The intraseasonal signals noted at the lower boundary are also seen at tropopause level. For example, in the episodes labelled A, D and F, the PV contours near the tropopause and the gradient migrate polewards. It will be shown later that this is a manifestation of wave–mean flow interaction. Recall that this state can only change through non-conservative processes and these episodes are a result of Rossby-wave breaking, filamentation of PV and consequent dissipation, enabling mass to cross PV contours equatorwards. This is the *vortex erosion mechanism* described in the Introduction. Nakamura (1995) derived an equation describing the evolution of the MLM PV q_o derived from the time dependence of mass integrals (1). Although his eq. (2.12) uses equivalent latitudes defined via a uniform isentropic density for reference, the derivation carries over to the equivalent latitude obtained here by the ELIPVI method, which is consistent with the variable isentropic density distribution obtained by preservation of both mass and circulation integrals in the rearrangement. q_o can only change through non-conservative processes.

Figure 2(c) shows the background-state zonal flow, u_o, on the 311 K isentropic surface, obtained by ELIPVI for every analysis. The bold curve marks the 2 PVU tropopause position. The flow is maximum on the equatorward side of the tropopause and will be described as the 'subtropical jet', although on θ surfaces below the jet maximum it lies in the midlatitudes (see Figure 1(c)). This jet is also slowly varying, with an obvious seasonal march. The amplitude of the subtropical jet is seen to pulse, tying in with the intraseasonal variability noted above. In the season shown, it does not vary substantially in latitude on these time-scales. However,

this season was characterized by a very persistent southern jet state across the North Atlantic in particular. In addition, there is marked variability in the 'polar jet' at 70°N in the background state, characterized by the presence or absence of the jet.

Figure 2(d) shows the Eulerian zonal mean zonal wind, $[u]$, for comparison with the MLM zonal wind. The subtropical jet is similar, although weaker in the zonal mean (the same colour scale has been used). Typically $u_o \geq [u]$, which can be anticipated by the following argument of Nakamura and Zhu (2010). For simplicity, consider an isentropic layer with uniform density and a uniform PV patch of area $\mathcal{A}(Q)$ bounded by the wavy contour, $q = Q$, and surrounded by lower PV on the isentropic surface. Now consider drawing a circle described by the equivalent latitude $\mu_e(Q)$, which by definition contains the same area $\mathcal{A}(Q)$. It must be the case that the latitude circle contains some area outside the vortex patch where the PV is lower (unless the vortex is circular and centred on the pole). Therefore, the circulation within the latitude circle must be lower than the circulation within the wavy vortex patch itself (2). When isentropic density is not uniform across a layer, mass rather than area is conserved (for adiabatic motion) and it can happen that $u_o < [u]$.

At high latitudes, $[u]$ is much more variable than u_o and flips between positive and negative at this level. There is evidence of poleward migration of $[u]$ anomalies that is not seen in u_o. Nakamura (1995) derived an equation (his 2.13) describing the evolution of the MLM zonal flow u_o that is a form of Kelvin's circulation theorem, although he used equivalent latitudes defined via a uniform isentropic density for reference. With the ϕ_e obtained by the ELIPVI method of section 2, which is consistent with rearrangement preserving circulation as well as mass, it is clear that u_o can only change through non-conservative processes and a similar evolution equation would apply. It will be shown below that the synoptic time-scale variability in $[u]$ is a reflection of adiabatic eddy processes that cannot directly influence the MLM state.

4.4. *Wave-activity evolution at tropopause level*

Fluctuations in the MLM state are often related to wave activity, but only indirectly through non-conservative effects. Figure 3(a) shows the pseudomomentum calculated relative to the MLM background state on the 311 K isentropic surface. A strong characteristic of the pseudomomentum is maxima moving polewards from the location of the tropopause (and midlatitude jet) in the background state. They move at a similar rate on many occasions throughout the season.

A time series of the meridional average of pseudomomentum poleward of 70°N at 311 K is shown by the bold curve in Figure 3(b) (multiplied by −1 to distinguish it from the stratospheric curves shown). Events A–E label the start of five marked episodes of high-latitude pseudomomentum at tropopause level. A sixth episode, labelled F, does not feature high-latitude pseudomomentum. The dates of the events are as follows: A (1200 UTC on 6 December 2009), B (0600 UTC on 24 December 2009), C (1800 UTC on 31 December 2009), D (1800 UTC on 23 January 2010), E (0000 UTC on 4 February 2010) and F (0600 UTC on 1 March 2010).

Episodes B and C exhibit clear migration of a pseudomomentum maximum from 40°N right to the North Pole over five days and many other events show a similar rate of poleward movement (10° per day or 13 m s^{-1}).

In episode D, pseudomomentum at 311 K can be seen to progress polewards from 50°N at a similar rate and then return equatorwards. Figure 2(a) of Methven (2013) shows that this event resembles a baroclinic wave with zonal wavenumber 8. PV anomalies at tropopause level grew simultaneously at all longitudes and hemispheric pseudomomentum diagnostics revealed simultaneous growth of positive interior and negative boundary pseudomomentum in the early phase, as would be expected for an unstable baroclinic normal mode. In this event, the movement of pseudomomentum is a result of the growth in meridional excursions in PV contours and its subsequent return in the decay phase of the baroclinic wave. Note that the pseudomomentum moving polewards is associated with the poleward advection of low PV air within the baroclinic wave (negative PV anomalies), while the signature of positive PV anomalies associated with equatorward displacement of PV contours only reaches 30–40°N during this episode, because the Rossby waves break cyclonically.

During the life-cycle event (D), the background-state PV contours in the range 2–5 PVU migrate polewards (Figure 3(a)). This is a result of the *vortex erosion mechanism* (see section 1). Rossby-wave breaking forms PV filaments, with positive PV extending to the south and negative PV to the north. Stretching in the time-dependent strain field of the wave means that the filaments thin until they are dissipated by mixing with their surroundings. Typically the breaking is asymmetric, with thinner PV positive filaments, and the net result is equatorwards expulsion of mass from the polar vortex into the surrounding surf zone. The background flow also accelerates slightly (Figure 2(c)) in the decay phase of the baroclinic wave, as a result of changing the PV distribution, to achieve the maximum jet strength for the season. As a result of the baroclinic wave event, the background-state θ gradient along the lower boundary is reduced over 55–65°N and enhanced slightly to the south and also near 80°N. This is typical of baroclinic wave experiments, where the boundary θ distribution is flattened as a result of stirring and the gradient is expelled to latitudes either side of the wave activity (e.g. Greenslade and Haynes, 2008).

Episodes B and C are also associated with poleward migration of enhanced θ gradients at the lower boundary of the background state, while the θ contours themselves shift equatorwards in the polar regions and the lowest temperatures ($\theta_{os} \approx 230$ K) are seen at the North Pole immediately following these two events. Close comparison of Figures 2(a) and 3(a) shows that the enhanced θ gradients lag the tropopause-level wave activity at each latitude during the poleward migration. This is intriguing and must be related to non-conservative processes in poleward-moving air masses. It is consistent with diabatic cooling of air masses as they progress polewards, such that a greater proportion of mass is enclosed in lower isentropic layers capping the North Pole. In these events, there is not a strong signature in background-state PV gradients at tropopause level (Figure 2(b)).

Two notable periods where the background-state PV contours migrate polewards (Figure 2(b)) are labelled A and F. During episode F, pseudomomentum extends equatorwards to 25°N for the first time in the season (Figure 3(a)). This activity is characterized by anticyclonic wave breaking at tropopause level on the equatorward flank of the jet. The sharpest PV gradient shifts polewards and the background-state jet initially accelerates. The lower boundary θ gradient (Figure 2(a)) is strong throughout this episode at 60°N –further poleward than throughout the winter. The jet then weakens and re-establishes itself further north following the end of the poleward migration of PV gradients (Figure 2(c)).

Episode A occurred in early December 2009 and was associated with a marked build-up of pseudomomentum and poleward extension right into the polar regions (Figure 3(a)). As shown in Methven (2013), this was associated with poleward expansion of nearly stationary ridges in the Alaskan and Scandinavian sectors. The episode was characterized by a more rapid equatorward migration of lower boundary θ contours and the strongest gradient, which then stayed in the southern position throughout winter. Note that the zonal mean flow (Figure 2(d)) is particularly negative (westward) in the polar regions after event A and the anomalies migrate with the wave activity. This behaviour illustrates that $[u]$ reflects the wave activity itself, rather than the background state. In contrast, the MLM flow, u_o, does not exhibit poleward migration events. Episodes A, B, C, D,

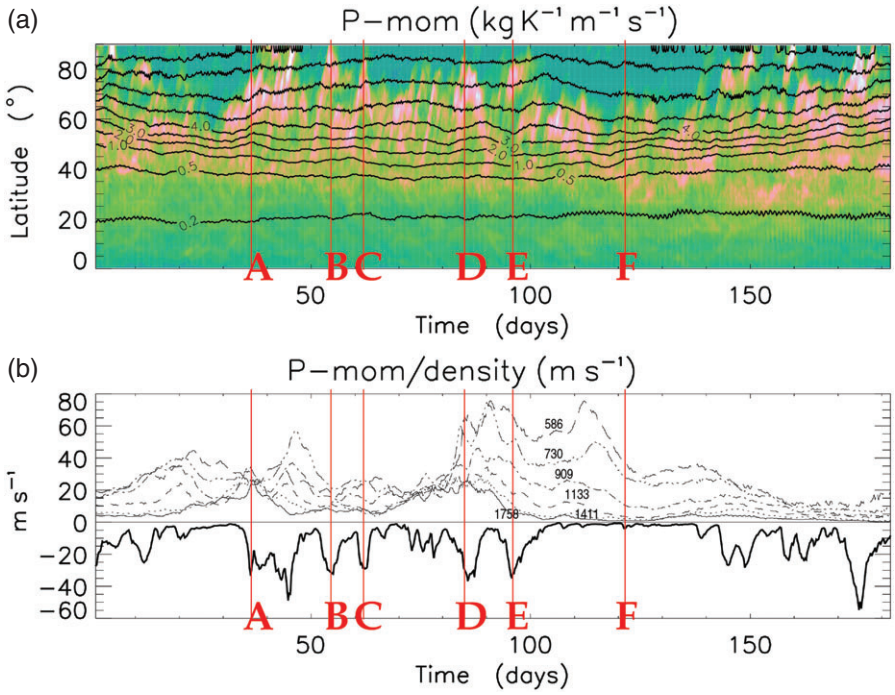

Figure 3. (a) Evolution of pseudomomentum (wave activity) on the 311 K surface (CI = 5000 kg K^{-1} m^{-1} s^{-1}) with background-state PV (PVU) overlaid in bold contours. (b) The bold line shows the meridional average of pseudomomentum (kg K^{-1} m^{-1} s^{-1}) poleward of 70°N at 311 K (multiplied by -10^{-4}). Labels A–F refer to events in wave activity at tropopause level used to relate to other time-series plots (see text). The other curves show pseudomomentum at six isentropic levels (labelled in K) spaced every 5 km in pseudoheight (see section 5). The measure shown is the meridional average of pseudomomentum poleward of 30°N, rescaled by dividing by the density $\sqrt{r_{00} r_o}$.

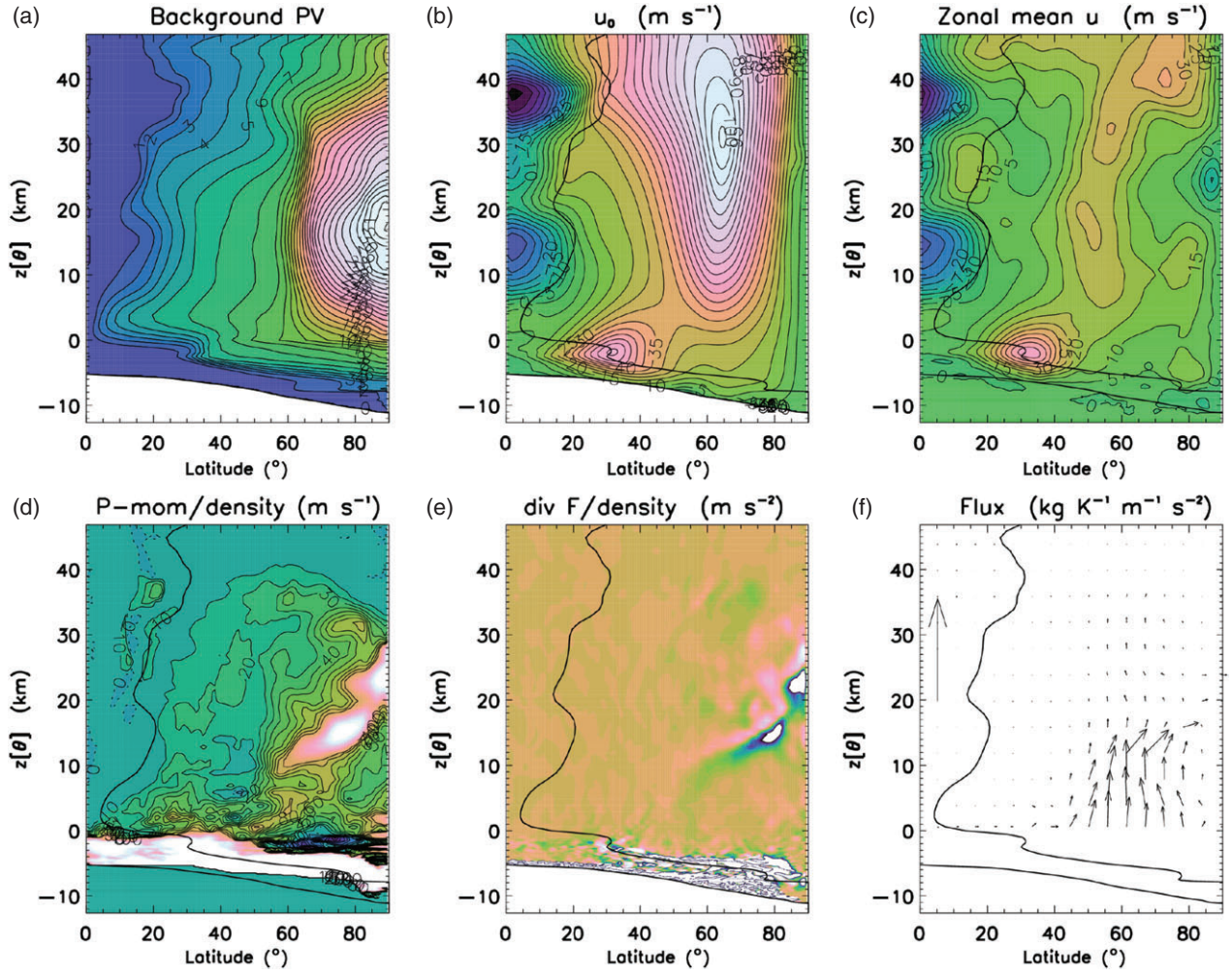

Figure 4. The atmospheric state at 1800 UTC on 22 January 2010, during a major stratospheric warming, shown over the full domain used for the calculation. The pseudoheight, $\hat{z}(\theta)$, is calculated from θ using (33). $\hat{z} = 0$ corresponds to $\theta = 380$ K. (a) Background-state Ertel PV (PVU) using Lait (1994) scaling for $\hat{z} > 0$ (34). (b) MLM zonal wind. (c) Eulerian zonal mean zonal wind. (d) Pseudomomentum divided by the density scaling $\sqrt{r_{00} r_o}$ (CI = 10 m s^{-1}). (e) Pseudomomentum flux divergence scaled by $\sqrt{r_{00} r_o}$ (CI = 5×10^{-4} m s^{-2}; blue to green is negative). (f) Pseudomomentum flux vector (see text for scaling of its components).

E and F are all to some extent associated with amplification of the subtropical jet. The relation between the *polar jet* and tropopause-level pseudomomentum is not immediately clear.

5. Stratospheric winter

5.1. Structure of the background state from the surface to the mesosphere

Figure 4(a)–(c) presents latitude–height cross-sections of the MLM PV and zonal flow alongside the Eulerian zonal mean zonal wind. The vertical coordinate used is a pseudoheight, which is a function of potential temperature alone, given by the formula

$$\hat{z}(\theta) = \frac{H}{\kappa} \ln\left(\frac{\theta}{\theta_{00}}\right), \tag{33}$$

where H is the pressure scale height (taken to be 6.5 km) and $\theta_{00} = 380$ K. The function equals height only for an isothermal profile. This is a good approximation in the lower stratosphere, but poor in the troposphere. The coordinate is 0 at the reference level θ_{00}, which has been chosen to be just above the tropical tropopause.

The Ertel PV has been rescaled above 380 K ($\hat{z} > 0$) using the method of Lait (1994), designed to factor out the increase in PV associated with the density decrease in an isothermal atmosphere:

$$\hat{q} = q\left(\frac{\theta}{\theta_{00}}\right)^{-9/2}, \tag{34}$$

where q stands for Ertel PV. Since the scaling factor is a function of θ only, the shape of the PV distribution on every isentropic surface is unaffected by the transformation. The transformation enables a regular contour interval (1 PVU) to be used throughout the stratosphere. The lower part of the plot ($\hat{z} < 0$) covers the domain already shown in Figure 1(a). Note that the abrupt change in $\partial q_o/\partial \hat{z}$ at 380 K or $\hat{z} = 0$ (Figures 4(a) and 1(a)) corresponds to the change from Ertel PV to Lait PV used in determining the PV levels for the calculations. The inverter results are not sensitive to the change, since the inversion grid itself is regular in latitude and the gradient in Ertel PV on the inverter grid does not have a step change across $\hat{z} = 0$. The stratosphere extends up to $\hat{z} \approx 30$ km (recall that this is a height above the 380 K surface). Throughout the stratosphere, PV is dominated by the winter polar vortex, which has a maximum meridional PV gradient at about 65°N at all levels in the background state. In the mesosphere above the polar vortex there is a rapid fall in Lait PV with altitude.

Figure 4(b) shows the zonal wind, u_o, obtained by the ELIPVI calculation. It is strongest (> 105 m s^{-1}) at high latitudes near the stratopause ($\hat{z}(\theta) \approx 30$ km). The stratospheric polar night jet is vertically aligned following the latitude of maximum PV gradient.

In contrast, the Eulerian zonal mean zonal wind (Figure 4(c)) is easterly in the upper stratosphere near the pole. This is the signature of a major stratospheric sudden warming (SSW), which had just begun at this time. The contrast with the MLM state in the extratropics is striking and has been shown previously by Nakamura (1995). It arises because the stratospheric polar vortex has a very strong circulation but it is distorted and displaced a long way off the pole. Recall that it has already been argued in section 4.3 that it is usually the case that $u_o \geq [u]$ and the greater the disturbance on a vortex, the greater the difference.

In the Tropics, above the tropopause, there is a clear alternation between easterlies and westerlies with height associated with the quasi-biennial oscillation (QBO) pattern. The MLM state is similar to the Eulerian zonal mean in this region, because PV contours in the analyses are nearer zonal symmetry than in the extratropics. However, PV has been modified in the equatorial regions of the MLM state (section A7). The inversion of background-state PV requires that it is everywhere positive in the Northern Hemisphere (so that the inversion equation is elliptic) and so any negative PV in the background state is set to be near 0. The modified region occupies most of $q_o < 1$ PVU in Figure 4(a) and easterlies are strongest where it protrudes further away from the Equator, because $Q_y < \mathrm{d}f/\mathrm{d}y$ and therefore the relative vorticity $\xi_o = r_o q_o - f$ is negative, which implies a positive meridional gradient in U.

5.2. Structure of wave activity in the stratosphere

Figure 4(d) shows the pseudomomentum calculated relative to the MLM state using the large-amplitude expression (29) and then zonally averaged.[‡] It is necessary to use an appropriate density weighting. If it is assumed that the density scale height is a constant, it can be shown from linear Rossby-wave theory (in height coordinates) that the pseudomomentum ($A = \frac{1}{2}\rho Q_y \eta^2$) is expected to be constant with height for the vertically propagating Rossby waves of the Charney–Drazin spectrum (e.g. James, 1994). Therefore, this would predict that the Rossby-wave component of pseudomomentum P_w (equivalent to A in isentropic coordinates) is approximately constant. However, Esler and Scott (2005) have shown that a barotropic mode exists in addition to the vertically propagating Charney–Drazin spectrum. Their approach was analytic, using simple contour dynamics theory for a polar vortex represented by a uniform patch of PV on every level. The barotropic mode possesses a pseudomomentum structure that falls with height as mass density does ($P/r_o \sim$ constant). A compromise between these two regimes for scaling density dependence is to divide the pseudomomentum P by the factor $\sqrt{r_{00} r_o}$, where r_{00} is a constant density, here taken to equal 10 kg K^{-1} m^{-2}, approximately the value at 380 K (see Figure 1(b)). The density-weighted wave activity, $P/\sqrt{r_{00} r_o}$, has units of m s^{-1}. In this measure, a packet of Charney–Drazin waves might be expected to increase with height as $r_o^{-1/2}$, while the barotropic mode would decrease with height as $r_o^{1/2}$. There is a distinct axis of maximum pseudomomentum that slopes upwards through the stratosphere from 55°N to the North Pole.

Figure 4(e) shows that the divergence of pseudomomentum flux (divided by the same density weight) is dominated by a dipole structure with a zero node aligned along the wave-activity maximum. The ramifications of this structure are discussed in section 5.4. Figure 4(f) shows the meridional and cross-isentropic components of the pseudomomentum flux depicted as (pseudo-)vectors in the meridional plane. The magnitude and direction of each vector is determined by the rescaled values $(F_\phi/Y, F_\theta/\Delta\theta)$ evaluated at the point coinciding with the vector base. They are shown at every eighth latitude point and every fourth isentropic level on the inverter grid ($\hat{z} > 0$ only). Here, $Y = \pi a$ and $\Delta\theta = (\theta_{\mathrm{top}} - \theta_{\mathrm{bot}})/2$ are related to the outer dimensions of the global domain and also serve to give the components the same units as flux divergence. The magnitude scaling is indicated by the vector at 5°N, the length of which represents an upward flux of 10^{-3} kg K^{-1} m^{-1} s^{-2}. There is clearly a strong upward pseudomomentum flux at 60N above the 380 K surface and the vectors are directed polewards just below the level of strong flux convergence. The weaker flux divergence immediately above is associated with the poleward tilt of the fluxes and the fact that they are untilted above this level. The zonally averaged flux vectors shown are calculated using almost the same expression as for Eliassen–Palm fluxes in isentropic coordinates. Indeed, the expression for the vertical component is the same and the meridional component differs only from the addition of nonlinear advection of pseudomomentum by the meridional flow (32). However, the key distinction is that the vectors shown here are calculated from perturbations relative to the MLM state, rather

[‡]The values in the troposphere and spanning the tropopause saturate in the plot, since the focus is on wave activity in the stratosphere ($\hat{z} > 0$).

than the Eulerian zonal mean. Since the MLM state is much more slowly varying during the SSW, the flux vectors are also much more consistent in time.

5.3. *Evolution on isentropic surfaces in the stratosphere*

The contrast between the MLM background state, u_o, and the traditional Eulerian zonal mean, $[u]$, is greatest in the upper stratosphere during SSWs. Figure 5 compares the MLM zonal flow and the Eulerian zonal mean on the 1411 K surface ($\hat{z} = 29.8$ km; $p \approx 1.5$–2.5 hPa) throughout the six-month period. SSWs are traditionally identified with periods when there are rapid decreases in the zonal mean zonal wind. In the case of major warmings, $[u]$ becomes easterly (negative). The brief dip in $[u]$ in early December (event A) is identified as a minor SSW. Prior to the minor SSW, there was a continuous acceleration of the MLM flow, representing the polar vortex circulation, and its strength did not dip during the SSW.

The abrupt appearance of zonal mean easterlies ($[u] < 0$) on 21 January 2010 (2 days before D) is the hallmark of a major SSW. This is the period of the most prominent difference between the MLM and Eulerian zonal mean states. Initially, the circulation did not decrease at all during the SSW. It took 8 days before the MLM maximum flow dropped below 90 m s^{-1} (Figure 5(a)). The MLM continued to decrease at 70°N for 30 days after this SSW onset; meanwhile, the flow spun up at 40°N as the polar vortex re-established itself.

The evolution of pseudomomentum at 1411 K (Figure 5(c)) is overlaid with background-state PV contours on the same surface to illustrate the nature of wave–mean flow interaction during the two SSW events. Pseudomomentum, before the minor SSW, first increased at 50°N and then amplified and migrated polewards, reaching the North Pole at the time (A) of the strong dip in $[u]$. This behaviour appears to mirror the amplification and poleward migration of pseudomomentum at the tropopause level (Figure 3(a)), which will be discussed further in section 5.5. As the pseudomomentum increased, the background-state PV contours on the equatorward flank of the pseudomomentum migrated polewards. The signature of PV filaments advected into the subtropics can be seen in the zonally averaged pseudomomentum at 25°N at time A (Figure 5(c)). The poleward migration of background-state PV stopped abruptly with the cessation of wave activity. This is a sign of the vortex erosion process, resulting in a more compact polar vortex with a tighter PV gradient and stronger circulation surrounding it.

The precursor build-up to the major SSW in January (D) was rather similar, although with the wave activity focused at higher latitudes, reflecting the position of the polar vortex PV gradients. As the wave activity amplified and reached the polar regions, mass was peeled away from the vortex edge into the surf zone by Rossby-wave breaking and mixing in the surf zone, resulting in the systematic poleward progression of the PV contours and tightening of the PV gradient on the polar vortex edge. The MLM flow, obtained through PV inversion, also exhibits a maximum that progresses polewards with the PV gradient (Figure 5(a)). The deceleration of the MLM flow only occurs when some of the highest PV values are mixed away completely at this level (the maximum PV at the pole reducing from 26 to 12 PVU during the SSW). PV poleward of 35°N builds up again after event E as a result of non-conservative processes, presumably long-wave radiative cooling. This accounts for the spin-up of the MLM flow and the zonal mean flow in the midlatitudes.

In the mid-stratosphere at 730 K ($\hat{z} = 14.8$ km; $p \approx 9$–16 hPa), the influence of the major SSW is dramatic (Figure 5(d)). Following the onset of the warming (D), the background-state PV contours progress polewards, with the maximum gradient moving from 65 to 80°N. Two major PV filamentation events are seen before markers E and F. The PV does not build up again at this level and the vortex circulation is not re-established. The minor SSW in December 2009 (A) is not as pronounced at this level as it was in the upper stratosphere.

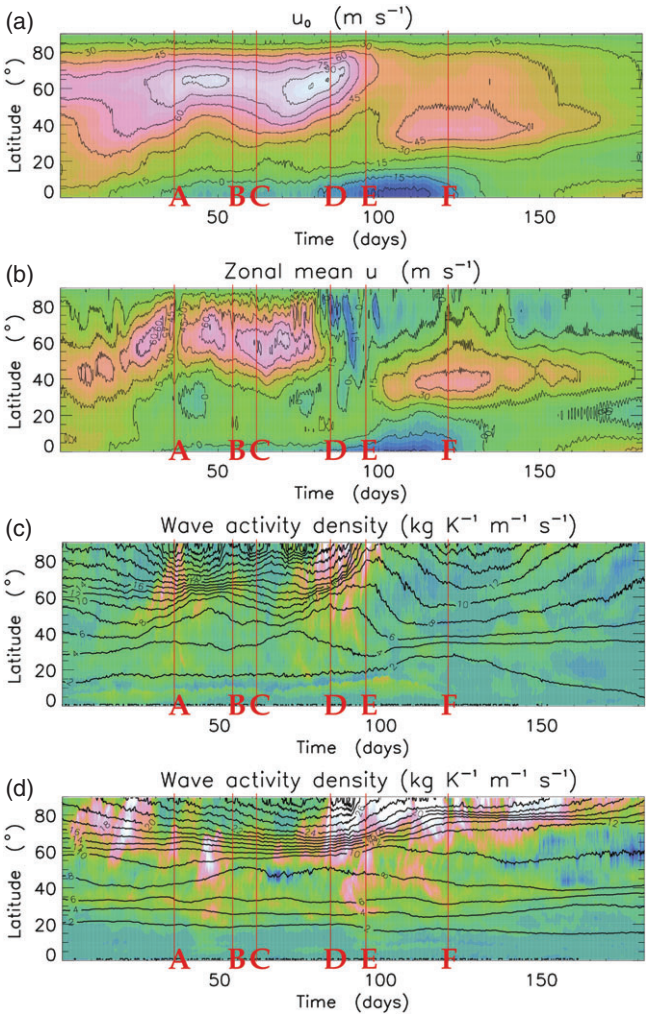

Figure 5. Evolution of the background state in the stratosphere. Time runs from 0000 UTC on 1 November 2009 to 0000 UTC on 1 May 2010. (a) MLM zonal wind (CI = 5 m s^{-1}) on the 1411 K surface, which lies ~ 3 km below the stratopause. (b) Eulerian zonal mean wind. (c) Pseudomomentum at 1411 K (CI = 2 kg K^{-1} m^{-1} s^{-1}) with background-state PV contours (PVU) overlaid using Lait scaling. (d) Pseudomomentum (CI = 10 kg K^{-1} m^{-1} s^{-1}) and background-state PV at 730 K. A–F refer to events in pseudomomentum at tropopause level (see Figure 3(b)).

5.4. *Downward migration of wave activity associated with SSWs*

One advantage of the partition of the full flow using the MLM as the background state is that the diagnosed pseudomomentum and its flux are also coherent in time. Figure 6(b) shows the vertical (cross-isentropic) component of the pseudomomentum flux in a time–height diagram (zonal and meridional average poleward of 30°N). The flux is almost always upwards (positive) and typically decays with height above 380 K ($\hat{z} > 0$). The upward flux is greatest and extends to highest altitudes in the period before and during SSWs (A and D). Correspondingly, the flux is convergent throughout the lower stratosphere during most of the time shown (Figure 6(c)) and is strongly convergent just above the stratopause before and during SSWs. During the major SSW, there is a strong signal of repeated downward migration of the locations of greatest flux convergence with local maxima in pseudomomentum density (Figure 6(d)). In contrast, the MLM zonal flow (Figure 6(a)) changes more slowly without the rapid downward migration signal during SSWs. The spin-down of the MLM flow through non-conservative effects is much greater during the major SSW (D to E) than the minor one (following event A). There is evidence

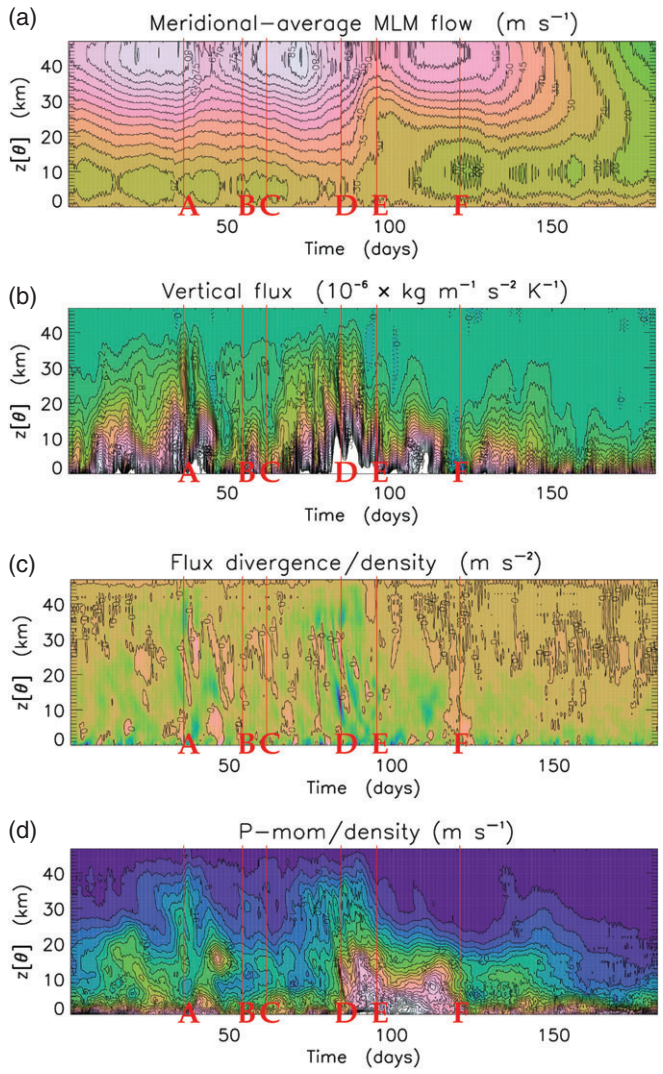

Figure 6. Evolution of background state and perturbation quantities versus pseudoheight (shown above 380 K; $\hat{z} > 0$) averaged poleward of 30°N on each isentropic surface. Time runs from 0000 UTC on 1 November 2009 to 0000 UTC on 1 May 2010. (a) MLM zonal wind (CI = 5 m s^{-1}). (b) Vertical component of pseudomomentum flux (CI = 2×10^{-6} kg K^{-1} m^{-1} s^{-2}). (c) Pseudomomentum flux divergence rescaled by density $\sqrt{r_{00} r_o}$ (CI = 2×10^{-5} m s^{-2}; blue and green are negative). (d) Pseudomomentum rescaled by density $\sqrt{r_{00} r_o}$ (CI = 4 m s^{-1}).

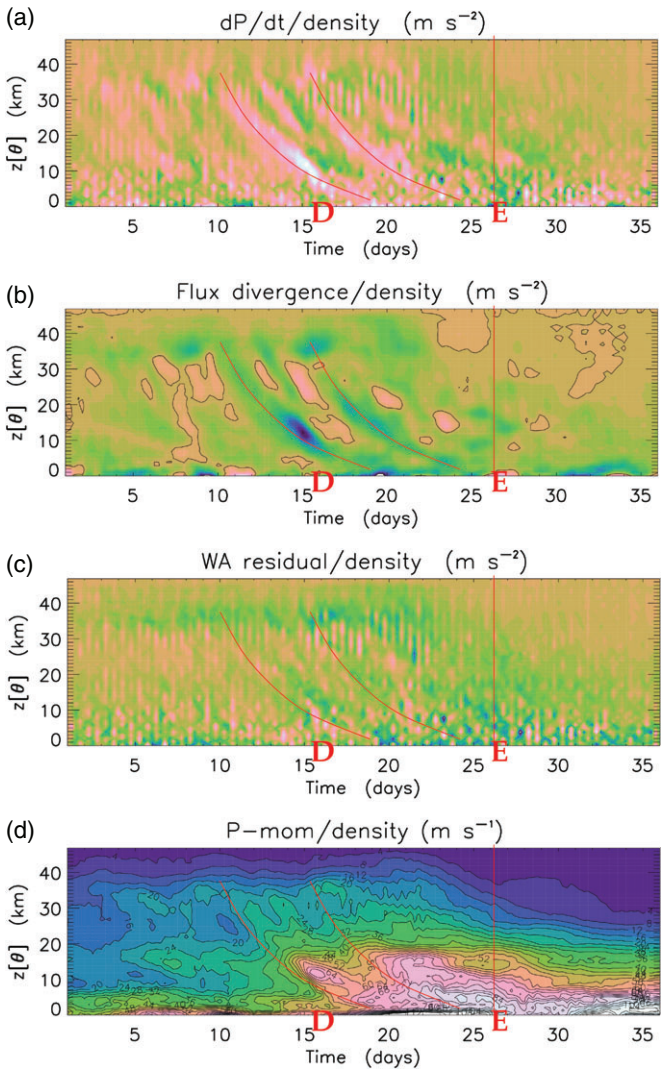

Figure 7. Evolution of terms in the pseudomomentum budget averaged poleward of 30°N on each isentropic surface and rescaled using the density $\sqrt{r_{00} r_o}$ (CI = 2×10^{-5} m s^{-2}; blue to green is negative). Time runs from 0000 UTC on 9 January 2010 to 0000 UTC on 13 February 2010. (a) Local time derivative of pseudomomentum. (b) Pseudomomentum flux divergence. (c) Residual from the pseudomomentum budget (see text). (d) Pseudomomentum rescaled by density $\sqrt{r_{00} r_o}$ (CI = 4 m s^{-1}).

for a much slower downward migration of the MLM flow during the final warming in spring.

The terms in the budget for conservation of pseudomomentum (28) are examined more closely during the major SSW in Figure 7. The tendency of pseudomomentum is often close to balancing flux convergence –as would be expected for a conservative fluid. However, the non-conservative term, S, calculated as a residual of the budget (Figure 7(c)) exhibits systematic behaviour. It is negative just above the stratopause ($\hat{z} \approx 35$ km) in the build-up to the SSW and strongly negative during the SSW event, indicative of the dissipation of pseudomomentum. In this region, the flux convergence and dissipation almost balance, such that the wave activity tendencies fluctuate about zero and the pseudomomentum is maintained at a high amplitude. Towards the end of the SSW, the dissipation wins and the pseudomomentum decays at stratopause level. Following 30 January 2010 ($t = 22$ days in Figure 7), the residual is negative throughout the stratosphere, showing wave dissipation at all levels until the pseudomomentum (Figure 7(d)) decays almost to 0 (E) in the upper stratosphere (when there are no longer any disturbances to dissipate).

During the SSW, the repeated downward migration of flux convergence is associated with downward migration of pseudomomentum. The curves in Figure 7 (red in the online article) have been drawn through the minima in $\nabla.\mathbf{F}$ and copied on to the other panels to highlight the relationship between variables. Flux convergence and positive pseudomomentum tendency occur immediately below local maxima in pseudomomentum, as required for a downward migration of the maxima. Although the residual signal is fairly noisy, in part due to the centred difference in time used to calculate the tendency from six-hourly data, wave dissipation is strongest where the wave activity is greatest (below $\hat{z} = 30$ km). Therefore, the downward migration is associated with non-conservative wave dissipation at the level of maximum wave activity, where the Rossby waves are breaking and strong filamentation occurs.

Vertical wave tilt is important to the upward flux. Equation (32) shows that $F_\theta = (1/ga)p'v'_g$, where v_g is the geostrophic meridional wind. In an undular Rossby wave without vertical tilt, pressure perturbations are in quadrature with v_g and therefore $F_\theta = 0$. The strong vertical flux in the build-up to the major SSW (Figure 6(b)) is associated with the prominent westward tilt of the distorted polar vortex, dominated by zonal wavenumber 1. Near the onset (event D) the vortex exhibits a westward tilt with its trough axis, turning from 90°E near 400 K to 90°W near the stratopause (a height separation of 35 km), very similar to the composite of vortex-displacement SSWs shown in figure 8

of Matthewman *et al.* (2009). This implies a vertical wavelength of 70 km or depth scale $H \approx 70/(2\pi) \approx 11$ km. Wave breaking and the associated PV stripping disrupt this coherent vertical tilt and weakens the upward pseudomomentum flux. Consequently, there must be flux convergence below the wave-breaking level, which results in local amplification of wave activity there, and the signal propagates downwards.

Although the situation is complex, it is instructive to consider a simplified model where the pseudomomentum is assumed to peak at some level, z_m, and across the region below, $z_m - H < z < z_m$, it is assumed that the flux falls linearly from value F_{up} at height $z_m - H$ to 0 at z_m, giving rise to the vertical flux convergence F_{up}/H. F_{up} is associated with westward-tilting, upward-propagating Rossby waves and the tilt is disrupted across the region of wave breaking approaching z_m. Wave dissipation is assumed to result in a decay at rate $P\tau^{-1}$ everywhere, but below $z_m - H$ the pseudomomentum, P_{up}, associated with upward propagating waves is maintained by a weak flux convergence, $(F(0) - F_{up})/(z_m - H) = P_{up}/\tau$, so that it is steady and uniform. Non-dimensionalizing P by $F_{up}\tau/H$, the pseudomomentum conservation law (28) becomes

$$\frac{\partial \hat{P}}{\partial t} \approx \frac{G(z_m - H, z_m)}{\tau} - \frac{\hat{P}}{\tau} \tag{35}$$

in the region $z > z_m - H$, where the top-hat function $G(z_m - H, z_m)$ is unity across $z_m - H < z < z_m$ and zero elsewhere. The maximum in $P(z)$ must move downwards, since P only increases below the maximum and dissipation occurs everywhere. The crucial assumption of the model is that the flux convergence layer moves downwards too, staying situated below the level of maximum pseudomomentum, $z_m(t)$. The justification is that wave breaking occurs where the pseudomomentum is greatest and the depth of the breaking region, H, is related to the vertical scale of the upward-propagating Rossby wave, which is assumed to be approximately invariant. Solving (35) numerically reveals that any initial pseudomomentum distribution $\hat{P}(z)$ with a single peak evolves to a sharply peaked structure, described approximately by

$$\begin{cases} P_m e^{-(z-z_m)/H} & \text{for } z > z_m, \\ P_{up} + (P_m - P_{up})\dfrac{z - (z_m - H)}{H} & \text{for } z_m - H < z < z_m, \\ P_{up} & \text{for } z < z_m - H, \end{cases} \tag{36}$$

which propagates downwards without change in form. If this structure function is assumed, it can be shown that its rate of downward migration is $(H/\tau)(1 - P_m)/(P_m - P_{up})$. Typically, the maximum, P_m, is found to be just less than 1/2 when $P_{up} \ll P_m$, yielding a downward migration rate slightly exceeding H/τ.

In the period immediately following the SSW, when the pseudomomentum flux from below shuts off (4 days before E), the pseudomomentum decay rate $\partial P/\partial t$ is given approximately by the non-conservative term S. Averaging over a five-day period from 0000 UTC on 30 January 2010 when the pseudomomentum experiences dissipation at all levels, it is found that the average dissipation time-scale ($\tau = -\overline{P}/\overline{S}$) is 3 days in the mesosphere ($\hat{z} > 32$ km) and increases approximately monotonically to 15 days at $\hat{z} = 0$ (just above the tropical tropopause level). Fels (1982) estimated scale-dependent radiative damping rates for temperature waves in the middle atmosphere associated with ozone and carbon dioxide, using a radiative transfer model. For a vertical wavelength of 12.6 km, he calculated a damping time-scale of 2 days at 50 km above ground, increasing to 6–7 days at 30 km. For an infinitely deep temperature anomaly, the time-scales are 6 and 25 days at the same two levels. Therefore the wave dissipation rate calculated here is faster than radiative damping alone. However, the dissipation rate for waves in potential vorticity will differ from the damping of the temperature field, depending on the PV anomaly aspect ratio (Haynes and Ward, 1993). The dissipation rate, S, also includes the effects of PV mixing and mechanical dissipation.

The downward migration speeds obtained from the slope of the curves in Figure 7(b) (red in the online article) are 6.7 km day^{-1} in the upper stratosphere, slowing monotonically to 1.3 km day^{-1} approaching the tropopause. This is consistent with descent at rate $\alpha H/\tau$, where $\tau(z)$ is obtained from the pseudomomentum budget residual and $\alpha H \approx 20$ km is constant. If H is taken as the height scale from the structure of the zonal wavenumber 1 disturbance (11 km), this implies $\alpha \approx 1.8$. This is consistent with the mechanism described above, given that solutions to the simple model (35) give $\alpha \geq 1$ and there are major approximations regarding the structure of wave breaking.

Andrews and McIntyre (1976) used a similar pseudomomentum argument to discuss the importance of wave dissipation for the downward propagation of zonal mean easterlies forming part of the QBO. They also highlighted the difference in the action of thermal and mechanical dissipation on the zonal mean flow. Here, the downward propagation signal is identified with a local maximum in wave-activity density, although this is coincident with a reduction in the Eulerian zonal mean zonal wind.

There are some important points relating to this downward migration mechanism.

(1) It does not involve downward wave-activity flux. The downward migration cannot be interpreted as the group propagation of a Rossby-wave packet.
(2) The mechanism does not invoke the zonal flow of the mean state, as is used in critical line arguments. Indeed, the relevant background zonal wind in this theoretical framework is the MLM flow, which is everywhere positive and therefore likely to be greater than the phase speeds of planetary Rossby waves. There are no critical lines with respect to this flow. Anomalies in the Eulerian zonal mean flow are a direct reflection of the wave-activity density, due simply to the disturbance of the vortex from zonal symmetry about the pole.

The mechanism deduced here from the ERA-Interim data is consistent with the argument of Plumb and Semeniuk (2003) based on a version of the Holton and Mass (1976) model and also idealized simulations using a 3D primitive equation model. In their experiments, by construction, the pseudomomentum flux is always upwards from the mechanically forced lower boundary and downward migration could not result from reflection of the upwelling Rossby-wave activity. It also does not result from 'downward control' via induced meridional circulations. They argue that downward migration results from 'the purely local interaction between the upward-propagating wave and the zonal mean flow' in a manner analogous to the QBO. They go further to state that 'these results imply that the similar downward migration observed in the Arctic Oscillation should not be taken to indicate any controlling influence of the stratosphere on the troposphere'. The fact that the pseudomomentum flux is upwards almost all the time in the stratosphere (Figure 6(b)) and the downward migration rate is well described by the dissipation rate calculated from analyses indicates that their argument applies here.

This does raise the question as to why Rossby waves break first near the stratopause, resulting in a local P maximum and the flux convergence below necessary for the mechanism. Dritschel and Saravanan (1994) constructed a contour dynamics quasi-geostrophic model of the stratospheric polar vortex, represented by a distorted PV patch (with a sharp edge) on every level and forcing by orography at the lower boundary. They showed two types of wave-breaking behaviour: wave breaking close to the lower boundary for strong orographic forcing and remote wave breaking occurring near the top boundary in regimes of weaker forcing. Compressibility (finite density scale height) was important to the remote wave breaking.

An argument for wave breaking near the vortex top is that pseudomomentum P is invariant following a Rossby-wave packet upwards, but density decreases exponentially, implying that meridional displacements of air parcels grow as $\eta \sim \exp(z/2H)$. Eventually, wave amplitude must saturate nonlinearly through the wave-breaking process, involving horizontal overturning of PV contours and irreversible deformation (McIntyre and Palmer, 1983). Dritschel and Saravanan (1994) imposed a rigid lid, which is clearly unrealistic. However, above the stratopause the (Lait) PV and circulation of the vortex fall with height (see Figure 4). In their theoretical analysis of displacement warmings, Esler and Matthewman (2011) used a model where the PV of the vortex stopped abruptly at the stratopause, but they did not impose a rigid lid using an unbounded domain instead. They also found wave breaking at the top of the vortex.

Note that inference of the wave-activity dissipation rate directly from data is only possible using the large-amplitude formulation of wave activity, which does not have any residual terms associated with nonlinear aspects of the adiabatic motions. This is especially important during a SSW, where the polar vortex is extremely distorted. The pseudomomentum conservation law requires the background state to be a zonally symmetric solution of the governing equations (without an eddy flux term forcing it) and the MLM state meets these criteria.

5.5. *Troposphere–stratosphere interaction*

5.5.1. *Upward pseudomomentum fluxes*

The last section has shown that the pseudomomentum flux is almost always upwards from the tropopause into the mesosphere. Typically, there is convergence of pseudomomentum throughout the stratosphere, but before and during SSWs the flux is deeper and converges strongly at the stratopause.

Figure 3(b) shows the meridional average of pseudomomentum poleward of 30°N at six isentropic levels in the stratosphere: 586, 730, 909, 1133, 1411 and 1758 K, corresponding to a regular spacing of 5 km in pseudoheight ($\hat{z} = 9.8, 14.8, 19.8, 24.8, 29.8, 34.8$ km). The minor SSW (event A) shows as a clear peak in pseudomomentum at the top three stratospheric levels shown. At 1411 K (near the stratopause), there was evidence of a build-up from 15 days beforehand, which is associated with the poleward extension of pseudomomentum at this level (Figure 5(c)), but also simultaneously at tropopause level (Figure 3(a)). The bold curve in Figure 3(b) shows that the arrival of pseudomomentum poleward of 70°N at the tropopause (A) coincides with the minor SSW onset at the stratopause. The poleward migration lags by several days in the mid-stratosphere (Figure 5(d)).

During the 20 day period before the major SSW, pseudomomentum is again seen to build up (Figure 3(b)) and extend polewards at stratospheric levels (Figure 5(c) and (d)), while at the tropopause there is a sequence of four events where pseudomomentum migrates into the Arctic (the last of which is event D). During this period, the density-scaled wave activity $P/\sqrt{r_{00}r_o}$ is approximately uniform with height in the stratosphere (Figure 3(b)), indicative of a vertically propagating Charney–Drazin spectrum. After the onset of the major SSW, a decreasing profile of $P/\sqrt{r_{00}r_o}$ with height emerges, which is consistent with the influence of a barotropic mode (see section 5.2). This vertical dependence continues throughout the remainder of the time series, including the final warming. In their idealized study, Dritschel and Saravanan (1994) also found that the vortex becomes more barotropic when wave forcing from the lower boundary is switched off, which they described as a vortex alignment effect.

Although not shown here, the peaks in pseudomomentum during the onset of the minor SSW (A) and in event E at tropopause level were dominated by $m = 2$. In contrast, the onset of the major SSW (D) was dominated by $m = 1$. Esler and Scott (2005) have shown that the $m = 1$ barotropic mode is unlikely to be excited and the vertical dependence of pseudomomentum during the early stages of the major SSW is consistent with a westward-tilted $m = 1$ structure. The strong fall in pseudomomentum with height emerges as the $m = 2$ pseudomomentum progresses polewards at the tropopause (reaching the Arctic at event E), indicative of the excitation of the $m = 2$ barotropic mode, as described in the theory of Matthewman and Esler (2011).

5.5.2. *Influence of the stratosphere on the troposphere*

Baldwin and Dunkerton (1999) diagnosed downward migration of Eulerian zonal mean zonal wind anomalies following SSWs. Zonal mean wind anomalies are primarily a marker for wave activity and therefore can migrate downwards with wave activity, as described in section 5.4. However, on longer time-scales, signals in zonal mean wind also reflect changes in the MLM background state (e.g. Figure 5). The evolution of the MLM state implies that mass crosses isentropic surfaces and/or PV contours on isentropic surfaces, so that the mass contained within PV contours in isentropic layers can change. The mass transport can be regarded as the meridional overturning circulation associated with the evolving background state (Nakamura, 1995). A persistent zonal torque in the upper stratosphere associated with Rossby-wave breaking will enable a meridional mass flux and then the 'downward control' theory of Haynes *et al.* (1991) indicates that the meridional circulation would burrow downwards in response, enabled by cross-isentropic mass flux associated with radiative processes.

Therefore, it is possible that the irreversible spin-down of the polar vortex circulation by the erosion mechanism, as seen cleanly in the MLM state (Figure 6(a)), is the factor that influences the troposphere below on longer time-scales. There have been many studies examining links to tropospheric circulation by imposing stratospheric perturbations. For example, Polvani and Kushner (2002) established that a zonally symmetric cooling perturbation to the polar lower stratosphere results in a robust poleward shift in the tropospheric jet. Haigh *et al.* (2005) conducted a suite of experiments to demonstrate that anomalous heating in the equatorial lower stratosphere also results in a poleward shift of the tropospheric jet. Reichler *et al.* (2005) used experiments with a GCM in idealized configurations to show that the rate of downward migration and appearance of the tropospheric signal were dependent on the radiative damping rate for thermal anomalies.

Song and Robinson (2004) argued that baroclinic eddies play a central role in transferring the signal to the troposphere via changes in Rossby-wave breaking behaviour and termed the mechanism 'downward control with eddy feedback' (DCWEF). Since the wave-breaking direction has a major influence on jet latitude via the vortex erosion mechanism, this could be the way in which small changes in the stratospheric state could have a major influence on the tropospheric circulation. Several mechanisms have been proposed for the influence of the lower stratosphere on baroclinic wave breaking. Wittman *et al.* (2007) demonstrated the sensitivity of the baroclinic wave breaking direction to stratospheric shear. Simpson *et al.* (2009) argued that the change in refractive index for the propagation of small-amplitude Rossby waves can explain the response arising from eddy feedback. They also found that changes in the meridional PV gradient, rather than the zonal mean flow, dominate the changes in refractive index. Rivière (2011) examined the link between increases in upper tropospheric baroclinicity and poleward shifts in the jet stream. He argued that the effect occurs via preferentially stronger growth of baroclinic waves with longer wavelengths, which tend to break anticyclonically and erode the jet polewards.

Following the end of the major SSW and the final downward migration of wave activity (a few days after event E), a steady

increase in PV is seen at high latitudes, both in the upper stratosphere and on the stratospheric side of the tropopause. The most notable tropospheric feature is that strong gradients in lower boundary θ at high latitudes progress over 10 days to the mid-latitudes (Figure 2(a)) and the polar jet in thermal wind balance with it dissipates (Figure 2(c)). Mass increases in isentropic layers at lower potential temperature, associated with an equatorward shift of the lower boundary in isentropic coordinates (e.g. below 265 K in Figure 1(b)). In pressure coordinates, this corresponds to an equatorward expansion of the dome of cold Arctic air. This is plausibly the result of stronger radiative cooling associated with less transient wave activity in the Arctic (Figure 3(a)) and large, persistent clear-sky regions. Methven (2013) also shows the existence of persistent $m = 2$ wave activity at this time and Figure 3(b) indicates the prevalence of barotropic modes.

After the lower boundary θ gradient becomes strongest at 60°N, there appears to be a change in regime for baroclinic wave breaking from cyclonic to anticyclonic at event F. The PV gradient at the tropopause is sharpened and migrates polewards as a result of the anticyclonic breaking (Figure 2(b)). The only time when the meridional average of the vertical pseudomomentum flux is substantially negative is immediately prior to event F (Figure 6(b)). This is associated with the structure of a baroclinic wave with an eastward zonal tilt of meridional wind structure with height above the tropopause. This is a signature of baroclinic wave trapping and is consistent with the argument of Song and Robinson (2004) that trapping of long baroclinic waves is important in the tropospheric response to stratospheric perturbations.

6. Conclusions

An atmospheric background state has been defined in such a way that it can be calculated as a diagnostic from a single global atmospheric analysis and yet is inherently slowly varying. This is achieved by defining it using the integral quantities –mass and circulation –that would be conserved under adiabatic and frictionless conditions. The state is classified as a modified Lagrangian mean (MLM), following McIntyre (1980), because it is defined relative to potential vorticity (PV) contours on isentropic surfaces. Although this would be a Lagrangian mean for adiabatic, frictionless flow, in a real flow PV contours are not exactly material contours (for example as a result of small-scale mixing). The MLM state is defined to be zonally symmetric and can be regarded as an adiabatic rearrangement from the disturbed atmospheric state to one, where all PV contours are aligned along latitude circles. The mass and circulation enclosed within PV contours on every isentropic surface describe the functionals $\mathcal{M}(Q,\Theta)$ and $\mathcal{C}(Q,\Theta)$, which by construction are the same for the background state as in the full 3D data. Since PV and potential temperature are materially conserved for adiabatic, frictionless flow, the functionals would also be invariant under these conditions and the background state steady.

A new method to obtain such a state has been developed, paying particular attention to the intersection of isentropic surfaces with the lower boundary. The latitudes of points along the lower boundary in isentropic coordinates are obtained as part of the solution procedure. This enables, for the first time, extraction of an MLM state from global data, which is suitable for examining the variation in background tropospheric baroclinicity associated with non-conservative processes. The state is obtained using an exact zonally symmetric PV inversion relation (13) and its numerical solution, including an iteration of equivalent latitudes describing the position of PV contours (on isentropic surfaces) and the location of the lower boundary in the background state. The method is named ELIPVI, standing for 'equivalent latitude iteration with PV inversion'.

The background state obtained is similar to the 'reference state' obtained by Nakamura and Solomon (2011), since both are derived from the mass and circulation functionals of atmospheric datasets. However, the key difference relates to the solution for the lower boundary of the background state. Nakamura and Solomon (2011) specify the lower boundary location using the Eulerian zonal mean θ on a pressure surface.

It is necessary to exclude some features of the atmospheric state from contributing to the background state. The lower boundary of the data is taken to be a particular terrain-following model level from the ERA-Interim reanalysis data. However, orography introduces local anomalies in potential temperature and geopotential at the lower boundary. These cannot be part of a zonally symmetric background state, so the approach taken is to redefine the geopotential height of the surface to be mean sea level and extrapolate variables below ground before calculating the mass and circulation integrals (1). The height above mean sea level of the lower boundary for the inversion domain is set equal to the height of the original lower boundary of the 3D data above the orography (approximately 30 m). Special consideration is also given to boundary-layer isentropic density and narrow PV anomalies such as those at fronts, which are regarded as perturbations. The method has been implemented so far only on the Northern Hemisphere domain and it is necessary to apply a boundary condition at the Equator and to eliminate negative PV values, which could prevent solution of the PV inversion equation by introducing hyperbolicity.

The MLM state at tropopause level is slowly varying and the most prominent feature is the seasonal cycle, as found by Nakamura and Solomon (2011). Non-conservative processes are central to the observed cycle. It was shown that Rossby-wave breaking, filamentation and subsequent dissipation are important in sharpening the PV gradient on isentropic surfaces crossing the tropopause. Moving into winter, the PV in the lowermost stratosphere increases through long-wave radiative cooling. Consequently, the polar vortex expands and the tropopause moves slowly equatorwards on every isentropic surface, continually being sharpened by wave breaking. After January, the tropopause begins to move polewards as the maintenance of the stratospheric PV reservoir by radiative cooling lessens and wave breaking results in inexorable poleward erosion of the vortex edge. Methven (2003) explored the parameter space in a single-layer model with orographic forcing and radiative damping and showed that regimes exist where wave activity erodes the polar vortex faster than it can be maintained until the vortex is eradicated. This erosion behaviour continues into late summer (e.g. Methven, 2009), when the Northern Hemisphere polar vortex at tropopause level has its minimum extent before the maintenance by radiative cooling builds again, moving into autumn.

All variability associated with adiabatic, frictionless motion is reflected in the perturbations relative to the MLM state. Since the MLM state is zonally symmetric, the evolution of perturbations can be described by the conservation law for pseudomomentum obtained by the zonal angular momentum–Casimir method of Haynes (1988). The conservation law expresses a relation between the local rate of change of pseudomomentum, pseudomomentum flux divergence and non-conservative effects. It is exact even for large-amplitude perturbations, when using the MLM as the background state. Both the pseudomomentum and 3D flux are defined at every point in physical space (λ,ϕ,θ). The zonal average of the flux has the same form as the Eliassen–Palm flux in isentropic coordinates, except that there is an additional nonlinear term expressing the meridional advection of pseudomomentum. However, a key difference is that perturbations are defined relative to the MLM state rather than the Eulerian zonal mean. This results in flux vectors that are much more smoothly varying in time. Indeed, it was found here that in the stratosphere the vertical component of the flux is almost always positive (upwards), even during the periodic events during a sudden stratospheric warming (SSW), when there is the characteristic downward migration of zonal mean zonal wind anomalies. This implies that the signal is not related to a downward flux of wave activity. Indeed, the

new diagnostic framework enables one to discern the physical mechanisms behind the phenomenon.

During the build-up to the SSW events, pseudomomentum migrates polewards on each isentropic surface, starting first at levels just above the stratopause, followed by similar behaviour at successive levels below. It is evident in the new diagnostics that this behaviour is associated with the descent of a pronounced maximum in pseudomomentum that slopes upwards in the meridional plane from approximately 55°N to the North Pole (Figure 4(d)). Rossby-wave breaking begins first above the stratopause. Filamentation in PV resulting from wave breaking is subject to strong dissipation, through both radiative damping of temperature anomalies and small-scale mixing.

The new ELIPVI background-state calculation, combined with the pseudomomentum conservation law of Haynes (1988), enables estimation of the non-conservative wave dissipation rate as a residual. The observations motivate a simple model (35) that describes the downward migration of wave activity and zonal mean zonal wind anomalies. It is sufficiently accurate to predict the slow-down of the migrating signal as it descends from the stratopause towards the lower stratosphere as a result of the increase in the time-scale of wave damping $\tau(z)$. The downward migration rate is $\alpha H/\tau$, where H is the depth scale associated with the tilted distortion of the polar vortex (zonal wavenumber 1 Rossby wave) and α is a constant slightly bigger than unity. Rossby-wave breaking disrupts the coherent westward-tilting structure of upward-propagating Rossby waves associated with the upward pseudomomentum flux. Therefore, there must always be strong convergence of the flux immediately below the wave-breaking region. The flux convergence results in a local rate of increase in pseudomomentum. Since this is below the maximum, the pseudomomentum signal must migrate downwards.

Note that it is not necessary to invoke a change in the background state to describe the migration. Indeed, the MLM zonal flow does not exhibit these almost periodic events. The MLM state changes on a slightly slower time-scale associated with the vortex erosion mechanism (see Figure 5(c) and (d)). In the minor SSW, although Rossby-wave breaking and a downward migration event occur, the wave breaking stops before the circulation of the polar vortex has been substantially depleted by the vortex erosion mechanism. In contrast, during the major SSW, continued upward pseudomomentum flux and repeated Rossby-wave breaking eventually halve the MLM zonal flow around the vortex and the strong PV gradient region is eroded right to the pole at the stratopause (Figure 5(c)). This change is irreversible and the PV reservoir cannot be replenished fast enough by radiative cooling before the end of winter. Note that the downward migration of the weakening in MLM flow can be linked to the meridional overturning circulation, which is dependent on non-conservative processes in the MLM framework (Nakamura, 1995). The rate of downward burrowing of the circulation was found to depend on the radiative damping time-scale in the GCM study of Reichler *et al.* (2005).

The interaction between the troposphere and stratosphere was explored using the new MLM framework. In the build-up to the major SSW in January 2010, the pseudomomentum density falls more slowly with height than during other periods (Figure 3(b)). It is argued that this is in part a signature of the upward-propagating Charney–Drazin spectrum of Rossby waves, dominated by a zonal wavenumber one structure showing a westward tilt with height throughout the stratosphere, closely resembling the composite of Matthewman *et al.* (2009).

During the build-up to the minor SSW in December 2009, there was a clear progression of pseudomomentum polewards from the midlatitudes almost simultaneously near the tropopause (Figure 3(a)) and stratopause (Figure 5(c)). The minor SSW occurred as the pseudomomentum maxima reached poleward of 70°N (event A). These are both related to the meridional transport of air over vast distances from the subtropics into the Arctic and then the diabatic modification of those air masses. In this event, the wave activity was dominated by zonal wavenumber 2 at both tropopause and stratopause levels, which is consistent with the description of wavenumber 2 warmings in terms of a barotropic mode (Esler and Scott, 2005).

An aspect warranting further study is the change to the troposphere following the major SSW. Following event E, when the upper stratospheric wave activity has dissipated and the MLM zonal flow reaches a minimum as a result of vortex erosion, there is a marked equatorward progression of θ contours at the lower boundary of the background state (Figure 2(a)). This must be associated with an increase in mass within lower tropospheric isentropic layers doming over the polar cap, but intersecting the lower boundary poleward of 60°N. One plausible explanation is greater diabatic descent into these layers associated with radiative cooling and the mean meridional circulation. As well as the rapid build-up of MLM PV in the polar upper stratosphere associated with downwards diabatic mass transport from the layer (and the fast radiative time-scale there), the lower stratospheric PV also increases at this time. This is consistent with net diabatic mass transport out of that layer (and concentration of PV substance there) and mass convergence into the lower troposphere.

The new calculation of a slowly evolving background state (ELIPVI) and perturbations to it presented in this article opens the way for many investigations of atmospheric dynamics and aspects such as the interaction between storm tracks and climate. The advantage of the partition is that it makes a clear distinction between the roles of adiabatic, frictionless motions and non-conservative processes. It avoids the use of spatial or temporal filtering, which cannot be supported in the absence of a spectral gap in the wave number or frequency domain. Furthermore, the MLM background state enables the use of the large-amplitude pseudomomentum conservation law and the partition of non-local influence via wave-activity fluxes from the local effects of non-conservative processes.

Acknowledgements

JM is grateful for support during the early stages of this work from NERC grant NE/D011507/1. The Cubehelix colour scale used in the plots was developed to couple a colour palette with monotonic increase in saturation and follows Green (2011), using code written by James R. A. Davenport.

Appendix A: Numerical details for PV inversion

A1. Calculation of isentropic variables

The data used here are obtained from the ERA-Interim analyses in their most fundamental form: spherical harmonic coefficients for surface pressure plus vorticity, horizontal divergence and temperature on terrain-following hybrid-pressure η levels (Simmons and Burridge, 1981). The spectral truncation is $T255$ on 60 levels from the Earth's surface to a pressure of 0.1 hPa. Inverse associated Legendre transforms and fast Fourier transforms are used to obtain surface pressure, temperature, $U = u\cos\phi/a$ and $V = v\cos\phi/a$ on a full Gaussian grid at η levels with $4 \times$ JG longitudes by $2 \times$ JG latitudes where the integer $JG = 128$ for the linear transform grid corresponding to T255. These fields are interpolated on to 131 isentropic surfaces. The level spacing is $\Delta\theta = 1.5$ K from 218–320 K and then is regular in pseudoheight above 450 K with $\Delta\hat{z} = 1$ km. A linearly increasing $\Delta\theta$ is used across the intermediate range 320–450 K.

The PV consistent with the finite-difference discretization used by the PV inverter to solve (13) (for Montgomery potential) is calculated as follows.

(1) U, V, T and p are found from the spectral transform at the same grid points, (λ_i, μ_j), on a Gaussian grid on η levels. Then $U_{j+\frac{1}{2}} = (U_j + U_{j+1})/2$ and $\theta_{j+\frac{1}{2}}$ are found and

assigned to the half-latitudes $\mu_{j+\frac{1}{2}} = (\mu_j + \mu_{j+1})/2$ for $1 \le j \le 2 \times JG - 1$, where $\mu_{\frac{1}{2}} = 1$ is the North Pole and $\mu_{2\times JG+\frac{1}{2}} = -1$ is the South Pole. The boundary condition $U = 0$ at the poles is used. Similarly, $V_{i+\frac{1}{2}}$ and $\theta_{i+\frac{1}{2}}$ are assigned to the half-longitudes $\lambda_{i+\frac{1}{2}}$ and periodic boundary conditions are used.

(2) $U_{i,j+\frac{1}{2}}$, $V_{i+\frac{1}{2},j}$ and $p_{i,j}$ are linearly interpolated in θ coordinates to the set of isentropic levels $\theta_{\rm m}$. If, at some location, $\theta_{\rm m}$ is less than $\theta_{\rm NL}$ (value at the lowest η level), then $U = U_{\rm NL}$ and $V = V_{\rm NL}$ are assigned to that latitude and longitude for level $\theta_{\rm m}$.

(3) $\zeta_{i,j}$ is calculated from (4) by the centred finite-difference of winds $U_{i,j+\frac{1}{2}}$ and $V_{i+\frac{1}{2},j}$ along isentropic levels.

(4) The top of the domain for the PV inversion algorithm is defined by an isentropic surface at pseudoheight $\hat{z}(\theta_{\rm top}) = \hat{z}(\theta_M) + \Delta\hat{z}/2$ where θ_M is the top data level and $\Delta\hat{z}$ is the spacing above 450 K. The pressure at $\theta_{\rm top}$ is found by interpolation from η levels (linearly in θ).

(5) Geopotential at the highest level, θ_M, is found by integrating the hydrostatic relation in η coordinates from the Earth's surface ($\eta = 1$) and then interpolating linearly in θ to location θ_M. The Montgomery potential there is found from (7) using pressure and geopotential.

(6) Hydrostatic balance (8) is integrated downward, using pressure, from the top boundary to the Earth's surface, yielding M at isentropic levels $\theta_{\rm m}$.

Note that the methods used here are consistent with the boundary conditions derived by Andrews (1983) for variables where θ surfaces go underground. Let $\Phi(\lambda, \mu, t)$ and $\Theta(\lambda, \mu, t)$ denote the location of the lower boundary. At each point (λ, μ, t), where $\theta < \Theta$, the variables p, U, V are constant (i.e. $p(\lambda, \mu, \theta, t) = p(\lambda, \mu, \Theta, t)$) but the isentropic density $r = 0$.

In the interior, a second-order finite difference in M is used to find Γ from (10) and isentropic density from (9). At the θ level closest to the lower boundary, a quadratic interpolation is used to find the derivatives of M (as in section 2.2.3). At the 'model top', $\theta_{\rm top}$, the pressure there is used to define $\partial M/\partial\theta$ using (8) and then (9) and (10) can be evaluated at the highest level, θ_M. Finally, (3) is used to calculate the Ertel PV.

A2. Treatment of orography

Mountain ranges can outcrop through θ surfaces to altitudes where potential temperature is higher, forming isolated hot spots of θ on the Earth's surface and on the terrain-following lower boundary. It does not make sense to rearrange the orographic height or potential temperature of the lower boundary. Therefore, such 'hot spots' cannot be regarded as part of the background state.

Data are adjusted by extrapolating below the geopotential height of the lower boundary in the 3D data, $z_{\rm LB}$, down to $z = 0$, which denotes mean sea level. Since the data defining the background state are in isentropic coordinates, the approach taken is to assume constant static stability (N) for the extrapolation below the orography across the height range $0 < z < z_{\rm LB}$:

$$\theta(z) = \theta_{\rm LB} - \frac{\partial\theta}{\partial z}(z_{\rm LB} - z); \qquad \frac{\partial\theta}{\partial z} = \frac{\theta_{00}}{g}N^2, \tag{A1}$$

where $N = 1.25 \times 10^{-2}\,{\rm s}^{-1}$ and $\theta_{00} = 300\,{\rm K}$. This equation is then used to find the heights, $z(\theta_{\rm m})$, corresponding to all the θ levels in the range $\theta(z = 0) < \theta_{\rm m} < \theta_{\rm LB}$. Integration of hydrostatic balance (8) downwards from the lower boundary yields pressure on these levels and therefore temperature. Horizontal wind and PV are assumed to be constant across the range $0 < z < z_{\rm LB}$. The adjusted lower boundary height is $\tilde{z}_{\rm LB} = z_{\rm LB} - z_{\rm s}$, where $z_{\rm s}(\lambda, \phi)$ denotes surface orographic height. The adjusted lower boundary potential temperature is $\tilde{\theta}_{\rm LB} = \theta(\tilde{z}_{\rm LB})$ from (A1). As mentioned in section 2.2.3, the lower boundary used here corresponds to an average height above the surface of only 30 m. The zonal average of $\tilde{z}_{\rm LB}$ is used in the lower boundary condition (21).

A3. Atmospheric boundary-layer PV anomalies

On the lowest few model levels, there are often very large values of PV (positive and negative) associated with shallow stable boundary layers, especially in Arctic winter, where the surface can become extremely cold and $r_{\rm o}$ small. The estimation of isentropic density is also poor where it varies rapidly in the vertical. The approach taken is to modify the density to be uniform across these layers with a value that ensures that the column mass is consistent with the pressure drop from bottom to top.

Specifically, the density is modified on the lowest $m_{\rm band}$ θ levels lying above the lower boundary at that location, defining the range $\theta_{\rm LB} < \theta < \theta_{\rm upper}$. The parameter $m_{\rm band} = 2$ was used for the results shown. Numerical integration with respect to θ is used to calculate the mass integral, $I_{\rm upper}$, in the column above these levels ($\theta_{\rm m} > \theta_{\rm upper}$) to the top of the inversion domain. The mass (per unit area) of total column is $(p_{\rm LB} - p_{\rm top})/g$. The difference gives the mass that should reside across the lowest levels and this is used to redefine isentropic density there, assuming that it is uniform:

$$r_{\rm o} = \frac{(p_{\rm LB} - p_{\rm top})/g - I_{\rm upper}}{\theta_{\rm upper} - \theta_{\rm LB}}. \tag{A2}$$

The influence of this modification on the background state can be seen in the thin wedge of high $r_{\rm o}$ nearest the lower boundary in Figure 1(b).

In addition, within the polar lower troposphere some extra constraints are placed on the 3D PV field before calculating the integrals for the background state. The polar lower troposphere is defined as the region polewards of a fixed position $\mu_{\rm mod}$, but only on isentropic surfaces where the background state intersects the lower boundary within this polar zone: $\mu_{\rm mod} < \mu_{\rm es}(\theta_{\rm m}) < 1$. At latitudes poleward of $\mu_{\rm mod}$, the first PV minimum searching upwards from the lower boundary in the polar lower troposphere is identified (if there is a minimum). The isentropic density at the level of the minimum and below is modified as described for the boundary layer (A2) and then the PV is recalculated. This is required to eliminate very variable density and PV in close proximity to high-latitude orography. Also, for $\mu > \mu_{\rm mod}$, the relative vorticity on isentropic surfaces in the 3D data is constrained to $\zeta - f < 2\Omega$, limiting the effects of strong vorticity at fronts.

Finally, once the first-guess background state has been calculated, the background PV is capped at the value $PV_{\rm polar}$, which is taken to be 0.75 PVU, close to the climatological value of $f/r_{\rm o}$ for the region. In the results shown, the parameter $\mu_{\rm mod} = 0.76$ and the region subject to modification is readily seen in Figure 1(a) as the region of almost uniform PV on the lowest isentropic surfaces (next to the pole).

A4. Calculation of mass and circulation integrals

The integrals (1) and (2) are evaluated by assuming that isentropic layers are shallow relative to structure in the PV field, so that the horizontal coordinates defining the PV contour do not vary across the layer. In this limit, it is not necessary to perform the vertical integral and the quantities within the integrands are interpreted as depth averages for the layer at locations (λ, μ).

The integrals over the areas enclosed by PV contours on each isentropic surface are evaluated by searching for all grid points where $q \ge Q$ and then summing these points weighted by the associated grid-box area multiplied by r or ζ, respectively.

A5. Interpolating from equivalent latitudes to the inverter grid and vice versa

The area, mass and circulation functionals are calculated for discrete values of PV, Q_k, on isentropic surfaces θ_m. The zonally symmetric background state is obtained by calculating the equivalent latitudes of every PV contour, $\mu_e(Q_k, \theta_m)$, together with the equivalent latitudes of the points where each isentropic surface intersects the lower boundary, $\mu_{es}(\theta_m)$. The first-guess equivalent latitudes are found from the area integrals using (6), but these latitudes are refined as part of the 'outer iteration' described in section 2.3. Every step in the outer iteration requires the following.

(1) A forward mapping of the PV levels, Q_k, and the functionals from the equivalent latitudes to the regular latitude grid, μ_j, used for PV inversion.
(2) Inversion of PV to obtain the Montgomery potential and therefore pressure, temperature, isentropic density and zonal flow. Note that the first guess for the zonal flow is obtained from (5) and that for the isentropic density is obtained from $\partial\mathcal{M}/\partial\mathcal{A}$ evaluated along each isentropic surface θ_m. The inversion algorithm uses a first guess for M calculated by inverting (9).
(3) The density and zonal flow from the inverter are used to calculate the new mass and circulation integrals $\mathcal{M}_{2D}(Q_k, \theta_m)$ and $\mathcal{C}_{2D}(Q_k, \theta_m)$.
(4) The new integrals are used to calculate the equivalent latitude shifts $\delta\mu_e^M$ (24) and $\delta\mu_e^C$ (26).
(5) The shifts are used to calculate the new equivalent latitudes, as described in section A10.
(6) The outer step repeats from point 1 above, unless the convergence criterion described in section A10 is met.

The forward mapping to find the PV distribution on the regular latitude grid μ_j used for inversion is found by cubic spline interpolation of PV values Q_k between the equivalent latitudes $\mu_e(Q_k, \theta_m)$. As discussed in section A7, the PV is assumed to be 0 on the Equator and $q \geq 0$ where isentropic surfaces intersect the lower boundary $\mu_{es}(\theta_m)$. The maximum value of PV from the 3D data is used as the boundary condition for the spline at the North Pole, unless the PV has been modified there by removal of a sharp PV spike (see section A6).

The mass functional is interpolated using $\mathcal{M}(Q_{max}, \theta_m) = 0$ as the spline boundary condition at the North Pole. The lower latitude boundary condition on the cubic spline is given by $\mathcal{M}(0, \theta_m) = \mathcal{M}_s(\theta_m)$. This corresponds to the mass of the entire isentropic layer above the lower boundary.

Zonal wind $U(Q_k, \theta_m)$ is found from the circulation integral using (5) and then interpolated on to the inversion grid to avoid errors associated with interpolating the planetary vorticity contribution (which is known exactly). $U = 0$ is used as the boundary condition on cubic spline interpolation at the North Pole. The lower latitude boundary condition on the spline is taken to be $U_s(\theta_m)$, obtained from the 'surface circulation integral' $\mathcal{C}_s(\theta_m)$.

Note that cubic splines generally gave better results than linear interpolation, because sometimes large distances between equivalent latitudes points are covered (see Figure 1(a)). However, overshoots or undershoots are possible. Since the background-state PV, mass and circulation must vary monotonically with latitude, the overshoots and undershoots are cut off. An inverse mapping is applied to the modified profile $q_o(\mu_j)$ to obtain a revised set of equivalent latitudes $\tilde{\mu}_e(Q_k)$ which, by construction, vary monotonically with Q_k and do not include a PV spike near the pole. Linear interpolation between grid points μ_j is sufficient for the inverse mapping, due to their close spacing. On the first step of the outer iteration only, the revised profile is used to modify the mass and circulation functionals appropriately so that the PV overshoots do not return in subsequent steps of the outer iteration.

A6. Removing spikes in background-state PV and θ_s

Diabatically generated high and low PV anomalies are often small-scale and particularly active in narrow regions along fronts. Often, PV dipoles are formed and due to their close proximity largely cancel in inversion to obtain flow and density anomalies. However, if they are included in the PV rearrangement, the highest values on each isentropic surface will be placed next to the North Pole and the lowest values next to the lower boundary or Equator. The artificial separation of the anomalies can introduce a large cyclonic circulation about the pole in the background state in the polar lower troposphere.

However, from a single snapshot of the atmosphere, it is not possible to partition PV anomalies that have arisen rapidly through non-conservative processes from PV anomalies associated with adiabatic advection from elsewhere. Such a partition would be dependent upon the recent history of the air. This introduces a problem for the definition of the background state. However, isolated spots of very high PV occupy a small area and contribute to a narrow spike in high PV at the pole following the calculation of first-guess equivalent latitudes. The spike is surgically removed from the background-state PV distribution on θ surfaces, by setting any equivalent latitudes where $\mu_{cut} < \mu_e < 1$ to 1 (i.e. removing the PV contour from the domain on that isentropic surface). The parameter is $\mu_{cut} = 0.985$ in the results shown.

In addition, the numerical stability of the inverter method is very sensitive to the precise shape of the θ_s curve approaching the Equator, because it becomes increasingly difficult to balance a horizontal temperature gradient as the Coriolis parameter $f \to 0$. The approach taken is that, once the first-guess background state has been found via the use of area integrals, the θ_s curve is modified equatorward of a chosen position μ_{tropic}. A smooth curve approaching the Equator with a functional form that can satisfy the balance relation is obtained by fitting the quadratic function

$$\theta_s(\mu) = \theta_s(\mu_{tropic}) + 0.99 \times \Delta\theta \left(1 - \frac{\mu^2}{\mu_{tropic}^2}\right). \qquad \text{(A3)}$$

Note that the parameter $\mu_{tropic} = 0.1$ is initially specified, but the μ position of the first θ surface intersecting the lower boundary equatorward of this point, $\mu_{es}(\theta_m) < 0.1$, is used to replace μ_{tropic} in (A3). The extrapolation to the Equator proceeds from there. Note that by design, the extrapolated profile hits the Equator at a θ coordinate just below the next θ surface, so that all the θ surfaces above θ_m do not intersect the lower boundary but run to the Equator. Since the modification results in eliminating intersections above θ_m (cutting off a θ_s spike near the Equator), it is necessary to modify the interior mass and circulation integrals at the points that have moved from 'below ground' to within the background-state domain. This is achieved by assuming that $r_o = 300\ \mathrm{kg\,K^{-1}\,m^{-2}}$ and $U_o = U_{os}(\theta_m)$ throughout the modification region and recalculating the integrals.

A7. Removing negative PV in the background state

The PV inverter can fail to solve (13) for M if the background PV is negative because the equation is no longer elliptic. Therefore, the negative PV region between the Equator and the equivalent latitude of the zero PV contour must be modified. This region covers a substantial area in the Tropics in the winter stratosphere and mesosphere (discussed in section 5). Since the $q = 0$ contour is moved to the Equator, it is necessary to set the integrals for area, mass and circulation for $q = 0$ equal to integrals across the entire isentropic layer in the Northern Hemisphere: $\mathcal{A}_s$, $\mathcal{M}_s$ and $\mathcal{C}_s$. Furthermore, since negative PV in the 3D data contributes to the circulation integral $\mathcal{C}_s(\theta_m)$, but does not contribute to $\mathcal{C}(Q > 0, \theta_m)$, it is possible to have a local maximum in $\mathcal{C}(Q)$

near $Q = 0$. The region where $\mathcal{C}(Q_k) > \mathcal{C}_s$ is modified by setting area, mass and circulation integrals equal to the values for the first PV contour, Q_{small}, with smaller circulation than $\mathcal{C}_s$. When mapped on to the inverter grid (section A5), this results in a modified region with PV increasing from 0 to value Q_{small} at $\mu_e(Q_{small}, \theta_m)$. Mass and circulation then decrease monotonically with PV from the Equator.

On isentropic surfaces that intersect the lower boundary, the lowest PV contour is sought where $\mu_e(Q_K, \theta_m) - \mu_{es}(\theta_m) > 10^{-3}$ and $Q_K > 0$. The PV at the lower boundary is then set to Q_{K-1} ($< Q_K$). Therefore, when interpolated on to the inverter grid, PV is positive and increases monotonically with latitude.

A8. Linear PV inversion algorithm

Using the previous estimate for M, the nonlinear Γ term in (13) is evaluated to yield a linear partial differential equation. The second-order finite-difference representation of this and the application of the boundary conditions results in a set of simultaneous equations which are solved with the D03EBF routine from the Numerical Algorithms Group (NAG) library, using the Strongly Implicit Procedure (Stone, 1968). This algorithm was chosen because it can be implemented on a grid with irregular boundaries.

The maximum number of allowed linear iterations is increased from 5 to 25 as the count of the nonlinear iterations increases. The maximum equation residual is set to decrease from no more than 1 to 0.8×10^{-8}, as the residual from the previous nonlinear iteration decreases (see below). These controls are designed to prevent too many iterations being performed when the nonlinear iteration residuals are large. Typically, convergence to this tolerance occurs in 5–16 linear iterations. The maximum fractional change in M over the last iteration is $< 10^{-11}$.

A9. Nonlinear iteration for PV inversion

The linear iteration treats the pressure function Γ (10) as given, as well as the values of M below the lower boundary obtained by extrapolation downwards (section 2.2.3) that are used to evaluate the meridional second derivatives in (13). A nonlinear iteration is introduced to update these terms. Nonlinear convergence is deemed to have been obtained when the maximum equation (13) residual is smaller than 0.6×10^{-7}. Typically, convergence to this tolerance occurs in 400–500 iterations. The maximum change in M during the last iteration is several orders of magnitude smaller than the residual. This iteration is very slow to converge, due to oscillatory behaviour near the lower boundary, but a faster method has not been found.

A10. Outer iteration to determine equivalent latitudes

The equivalent latitudes are revised by mixing the shifts calculated from the mass (24) and circulation (26) methods (section 2.3) with the shift from the previous outer iteration. This avoids introducing an alternating pull on equivalent latitudes in the outer iteration. The iteration is also made smoother by introducing a relaxation scheme,

$$\mu_k^n = \left(\mu_k^{n-1} + \frac{1}{2}\delta\mu_k^{n-1} + \frac{1}{4}\delta\mu_k^M + \frac{1}{4}\delta\mu_k^C + \frac{\overline{\mu}_k^{n-1}}{N_r} \right) \frac{N_r}{N_r+1},$$

$$\overline{\mu}_k^n = \left(\overline{\mu}_k^{n-1}(N_r - 1) + \mu_k^n \right) \frac{1}{N_r},$$

which applies to every PV contour $\mu_k = \mu_e(Q_k, \theta_m)$. It is analogous to a implementing a mixed explicit–implicit time step with relaxation towards the running mean, $\overline{\mu}_k$, on 'time-scale' N_r. Larger N_r implies weaker relaxation and the value $N_r = 5$ was used.

A similar procedure is used to obtain a new estimate for the equivalent latitudes of the surface potential temperature distribution, $\mu_{es}(\theta_m)$, using the latest MLM $\theta_s(\mu)$ estimate.

The convergence criterion for the outer iteration requires that the mean magnitudes of the shift in equivalent latitudes, $|\delta\mu_e(Q_k, \theta_m)|$, and the similar quantity for the surface equivalent latitudes, $|\delta\mu_{es}(\theta_m)|$, are less that $\alpha\Delta\mu_{min}$, where $\Delta\mu_{min}$ is the smallest grid spacing on the inverter grid (next to the pole) and the tolerance parameter α was set to 0.5. Note that the inverter grid is regular in latitude, with spacing 0.703°, so, in terms of sine of latitude, $\Delta\mu_{min} = 1.377 \times 10^{-4}$ while $\Delta\mu_{max} = 1.225 \times 10^{-2}$. For the dates examined, the outer iteration stopped on this criterion after 12–15 iterations.

References

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

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

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

Baldwin M, Dunkerton T. 1999. Downward propagation of the Arctic oscillation from the stratosphere to the troposphere. *J. Geophys. Res.* **104**: 30 937–30 946, doi: 10.1029/1999JD900445.

Berrisford P, Kallberg P, Kobayashi S, Dee D, Uppala S, Simmons A, Poli P, Sato H. 2011. Atmospheric conservation properties in ERA-Interim. *Q. J. R. Meteorol. Soc.* **137**: 1381–1399.

Bleck R. 1973. Numerical forecasting experiments based on the conservation of potential vorticity on isentropic surfaces. *J. Appl. Meteorol.* **12**: 737–752.

Bleck R. 1974. Short-range prediction in isentropic coordinates with filtered and unfiltered numerical models. *Mon. Weather Rev.* **102**: 813–829.

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

Chagnon J, Gray S, Methven J. 2013. Diabatic processes modifying PV in a North Atlantic cyclone. *Q. J. R. Meteorol. Soc.* **139**: 1270–1282.

Dee DP, Uppala SM, Simmons AJ, Berrisford P, Poli P, Kobayashi S, Andrae U, Balmaseda MA, Balsamo G, Bauer P, Bechtold P, Beljaars ACM, van de Berg L, Bidlot J, Bormann N, Delsol C, Dragani R, Fuentes M, Geer AJ, Haimberger L, Healy SB, Hersbach H, Hólm EV, Isaksen L, Kållberg P, Köhler M, Matricardi M, McNally AP, Monge-Sanz BM, Morcrette J-J, Park B-K, Peubey C, de Rosnay P, Tavolato C, Thépaut J-N, Vitart F. 2011. The ERA-Interim reanalysis: Configuration and performance of the data assimilation system. *Q. J. R. Meteorol. Soc.* **137**: 553–597.

Dritschel D, Saravanan R. 1994. Three-dimensional quasi-geostrophic contour dynamics, with an application to stratospheric vortex dynamics. *Q. J. R. Meteorol. Soc.* **120**: 1267–1297.

Esler J, Matthewman N. 2011. Stratospheric sudden warmings as self-tuning resonances. Part II: Vortex displacement events. *J. Atmos. Sci.* **68**: 2505–2523.

Esler J, Scott R. 2005. Excitation of transient Rossby waves on the stratospheric polar vortex and the barotropic sudden warming. *J. Atmos. Sci.* **62**: 3661–3682.

Fels S. 1982. A parameterization of scale dependent radiative damping rates in the middle atmosphere. *J. Atmos. Sci.* **39**: 1141–1152.

Green DA. 2011. A colour scheme for the display of astronomical intensity images. *Bull. Astron. Soc. India* **39**: 289.

Greenslade M, Haynes P. 2008. Vertical transition in transport and mixing in baroclinic flows. *J. Atmos. Sci.* **65**: 1137–1157.

Haigh J, Blackburn M, Day R. 2005. The response of tropospheric circulation to perturbations in lower-stratospheric temperature. *J. Clim.* **18**: 3672–3685.

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.

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

Haynes P, Marks C, McIntyre M, Shepherd T, Shine K. 1991. On the 'downward control' of extratropical diabatic circulations by eddy-induced mean zonal forces. *J. Atmos. Sci.* **48**: 651–678.

Held I. 1985. Pseudomomentum and the orthogonality of modes in shear flows. *J. Atmos. Sci.* **42**: 2280–2288.

Holton J, Mass C. 1976. Stratospheric vacillation cycles. *J. Atmos. Sci.* **33**: 2218–2225.

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

James I. 1994. *Introduction to Circulating Atmospheres*. Cambridge University Press: Cambridge, UK.

Lait L. 1994. An alternative form for PV. *J. Atmos. Sci.* **51**: 1754–1759.

Lee S, Kim H. 2003. The dynamical relationship between subtropical and eddy-driven jets. *J. Atmos. Sci.* **60**: 1490–1503.

Legras B, Dritschel D. 1993. Vortex stripping and the generation of high vorticity gradients in 2-D flows. *Appl. Sci. Res.* **51**: 445–455.

Magnusdottir G, Haynes P. 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. *Philos. Trans. R. Soc. London A* **296**: 129–148.

McIntyre M, Palmer T. 1983. Breaking planetary waves in the stratosphere. *Nature* **305**: 593–600.

McIntyre M, Shepherd T. 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.

Matthewman N, Esler J. 2011. Stratospheric sudden warmings as self-tuning resonances. Part I: Vortex splitting events. *J. Atmos. Sci.* **68**: 2481–2504.

Matthewman N, Esler J, Charlton-Perez A, Polvani L. 2009. A new look at stratospheric sudden warmings. Part III: Polar vortex evolution and vertical structure. *J. Clim.* **22**: 1566–1585.

Melander M, McWilliams J, Zabusky N. 1987. Axisymmetrization and vorticity-gradient intensification of an isolated 2-D vortex through filamentation. *J. Fluid Mech.* **178**: 137–159.

Methven J. 2003. The influence of PV inversion on polar vortex dynamics and passive tracer simulations in atmosphere-like regimes. *Q. J. R. Meteorol. Soc.* **129**: 1191–1215.

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

Methven J. 2013. Wave activity for large amplitude disturbances described by the primitive equations on the sphere. *J. Atmos. Sci.* **70**: 1616–1630.

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

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

Nakamura N, Solomon A. 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.

Nakamura N, Zhu D. 2010. Finite-amplitude wave activity and diffusive flux of potential vorticity in eddy-mean flow interaction. *J. Atmos. Sci.* **67**: 2701–2716.

Norton W. 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.

Plumb R, Semeniuk K. 2003. Downward migration of extratropical zonal wind anomalies. *J. Geophys. Res.* **108**: 4223, doi: 10.1029/2002JD002773.

Polvani L, Kushner P. 2002. Tropospheric response to stratospheric perturbations in a relatively simple general circulation model. *Geophys. Res. Lett.* **29**: 1114, doi: 10.1029/2001GL014284.

Reichler T, Kushner P, Polvani L. 2005. The coupled stratosphere–troposphere response to impulsive forcing from the troposphere. *J. Atmos. Sci.* **62**: 3337–3352.

Rivière G. 2011. A dynamical interpretation of the poleward shift of the jet streams in global warming scenarios. *J. Atmos. Sci.* **68**: 1253–1272.

Simmons A, Burridge D. 1981. An energy and angular-momentum conserving vertical finite-difference scheme in hybrid vertical coordinates. *Mon. Weather Rev.* **109**: 758–766.

Simmons A, Poli P, Dee D, Berrisford P, Hersbach H, Kobayashi S, Peubey C. 2014. Estimating low-frequency variability and trends in atmospheric temperature using ERA-Interim. *Q. J. R. Meteorol. Soc.* **140**: 329–353.

Simpson I, Blackburn M, Haigh J. 2009. The role of eddies in driving the tropospheric response to stratospheric heating perturbations. *J. Atmos. Sci.* **66**: 1347–1365.

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

Song Y, Robinson W. 2004. Dynamical mechanisms for stratospheric influences on the troposphere. *J. Atmos. Sci.* **61**: 1711–1725.

Stone H. 1968. Iterative solution of implicit approximations of multi-dimensional partial differential equations. *SIAM J. Numer. Anal.* **5**: 530–558.

Thuburn J, Lagneau V. 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.

Wittman M, Charlton A, Polvani L. 2007. The effect of lower stratospheric shear on baroclinic instability. *J. Atmos. Sci.* **64**: 479–496.