Received: 06 March 2018 | Revised: 21 December 2018 | Accepted: 15 January 2019 | Published on: 27 February 2019

DOI: 10.1002/qj.3474

RESEARCH ARTICLE

Quarterly Journal of the Royal Meteorological Society RMetS

Vortex dynamics of stratospheric sudden warmings: A reanalysis data study using PV contour integral diagnostics

Frank Kwasniok | Robin Beaumont | John Thuburn

College of Engineering, Mathematics and Physical Sciences, University of Exeter, UK

Correspondence
Frank Kwasniok, University of Exeter, Harrison Building, North Park Road, Exeter EX4 4QF, UK.
Email: F.Kwasniok@exeter.ac.uk

Funding information
EPSRC Doctoral Training Grant

The dynamics of the polar vortex underlying stratospheric sudden warming (SSW) events are investigated in a data-based diagnostic study. Potential vorticity (PV) contour integral quantities on isentropic surfaces are discussed in a unified framework; new expressions for their time evolution, particularly suitable for evaluation from data, are presented and related to previous work. Diagnostics of mass and circulation are calculated from ERA-40 reanalysis data for the stratosphere in case-studies of seven winters with different SSW behaviour. The edge of the polar vortex is easily identifiable in these diagnostics as an abrupt transition from high to low gradients of PV, and the warming events are clearly visible. The amount of air stripped from the vortex as part of a preconditioning leading up to the warming events is determined using the balance equation of the mass integral. Significant persistent removal of mass from the vortex is found, with several such stripping events identifiable throughout the winter, especially in those during which a major sudden warming event occurred. These stripping episodes are visible in corresponding PV maps, where tongues of high PV are being stripped from the vortex and mixed into the surrounding surf zone. An attempt is made to diagnose from the balance equation of the circulation the effect of frictional forces such as gravity wave dissipation on the polar vortex.

KEYWORDS
polar vortex, PV contour integral diagnostics, preconditioning, stratospheric sudden warming events, stripping of filaments of air

1 | INTRODUCTION

The polar vortex is a prominent feature of the winter stratosphere. It extends from the lower stratosphere up to the stratopause and is comprised of strong westerly winds increasing with height. In the Northern Hemisphere at a rate of about 0.6 major events per year, these persistent westerlies can become disturbed, replaced by easterly winds and accompanied by an increase of temperature in the polar region of the order of 10 K/day. These events are referred to as stratospheric sudden warmings (SSWs). They are of major interest as stratospheric conditions and in particular SSWs can affect surface weather (Baldwin and Dunkerton, 2001; Thompson *et al.*, 2002). Following events in which the polar vortex is anomalously strong or weak, zonal wind or geopotential height anomalies of the same sign are often evident in the troposphere.

At the beginning of winter, a large-scale cyclonic vortex is formed under radiative forcing. During the course of the winter a stripping process takes place; filaments of air are stripped from the edge of the vortex, reducing the area of the vortex and sharpening its edge, and are mixed into the surrounding well-mixed surf zone (McIntyre and Palmer, 1984). This vortex erosion process is related to Rossby waves which originate in the troposphere and propagate upward into the stratosphere where they eventually break, pulling

tongues of air from the vortex edge (McIntyre and Palmer, 1983; Waugh *et al.*, 1994; Polvani and Saravanan, 2000). The vortex stripping is thought to precondition the vortex, rendering it more susceptible to sudden warming events (Labitzke, 1981; Limpasuvan *et al.*, 2004).

Vortex erosion has been extensively studied in simulations with both single- and multi-layer models. Basically two main approaches have been taken. Many studies have imposed a wave-like bottom boundary forcing on the vortex. Mixing in a model vortex under such wave-like forcing has been investigated (Norton, 1994; Sobel and Plumb, 1999) as well as the influence of the forcing amplitude and wavenumber on the vortex erosion process (Matthewman and Esler, 2011; Esler and Matthewman, 2011; Beaumont *et al.*, 2017). Esler and Matthewman (2011) found that, for a given background flow, there is a threshold for the forcing amplitude at which a bifurcation occurs causing the system to transition into a state in which a vortex-splitting event can take place. Also Beaumont *et al.* (2017) found a threshold of the forcing amplitude for significant vortex stripping to occur; the relationship between the stripping process and the occurrence of an SSW appears to be quite complex. Other studies have embedded a vortex in an externally imposed barotropic flow (Mariotti *et al.*, 1994; Legras *et al.*, 2001). This simulates the effect of other vortices on the erosion of the polar vortex. If the externally imposed shear is weak, the vortex remains intact; if it is increased, the contours near the edge of the vortex become stripped until the vortex eventually breaks down.

Various diagnostics of the flow in the stratosphere have been used to investigate the structure and dynamics of the polar vortex, especially around the time of SSW events. The simplest of these diagnostics involve the Eulerian zonal means of various quantities of interest such as zonal-mean zonal wind, which is used in the WMO definition of SSWs, as well as the mean meridional circulation. A slightly more Lagrangian view of the stratospheric circulation is provided by the transformed Eulerian mean (TEM) circulation (Andrews *et al.*, 1987), which combines Eulerian means with the leading-order effects of eddies to provide an approximation to a Lagrangian mean circulation. However, Eulerian mean-based diagnostics can be unreliable, blurring features which have high zonal asymmetry. This is especially true in the case of a disturbed vortex which is no longer centred over the pole, such as following SSWs. The averaging process can also cause the edge of the vortex and other sharp gradients of PV as well as small-scale features such as filaments of high-PV air to become less well defined. Additionally the association between changes in stratospheric circulation and certain non-conservative processes is not well represented in Eulerian mean diagnostics. Lagrangian quantities on isentropic PV contours such as the vortex area and the related equivalent latitude (Butchart and Remsberg, 1986) as well as the more dynamically motivated integral diagnostics of mass and circulation (Nakamura, 1995; Thuburn and Lagneau, 1999; Nakamura and Solomon, 2011; Methven and Berrisford, 2015; Beaumont *et al.*, 2017) are more robust, in the sense that they are not strongly affected by reversible displacements of the vortex. Mass and circulation integrals only evolve through non-conservative processes such as diabatic heating, friction and irreversible PV mixing. PV contour integral diagnostics of mass and circulation and their balance equations have been used to study vortex stripping and erosion in model simulation data (Nakamura, 1995; Thuburn and Lagneau, 1999; Beaumont *et al.*, 2017).

The present paper extends this line of research to reanalysis data. It has three main purposes:

(a) Mass and circulation integrals on PV contours on isentropic surfaces are introduced as diagnostic tools to monitor the dynamics of the stratospheric polar vortex in reanalysis data.
(b) New expressions for the time evolution of mass and circulation integrals are presented which are suitable for evaluation from data and they are related to earlier work.
(c) The balance equation of the mass integral is used to diagnose stripping of filaments of air from the polar vortex as part of a preconditioning leading up to an SSW. The balance equation of the circulation is used to estimate the effect of frictional forces such as gravity wave drag on the polar vortex.

The research is motivated and guided by the following questions:

(a) How do SSWs show up in mass and circulation integral plots and are there differences between splitting and displacement events?
(b) Is there a relationship between the amount of mass in the vortex or the circulation of the vortex and the occurrence of an SSW?
(c) How much mass stripping does occur and what is the timing of this stripping?
(d) Is there a relationship between the amount of mass stripping and the occurrence of an SSW?
(e) Is there any difference in the stripping process between splitting and displacement SSWs?

Vortex-displacement events are dynamically distinct from vortex-splitting events (Matthewman *et al.*, 2009). In particular, splitting events are essentially barotropic whereas displacement events have a significant baroclinic component. This may manifest itself, for example, in the nature and timing of the mass-stripping process at different isentropic surfaces.

The paper is structured as follows: In Section 2 the dynamical setting is outlined. The dataset and pre-processing steps are described in Section 3. Section 4 introduces the integral diagnostics of mass and circulation and discusses their balance equations. The results are presented in Section 5, and conclusions are drawn in Section 6. Four Appendices are presented (A) to list and derive useful identities, (B) to give the detailed derivation of the balance equations for the PV contour integral diagnostics, (C) to give the balance equation for the vortex area and (D) to describe the numerical integration

scheme for the evaluation of surface integrals on PV contours from data.

2 | DYNAMICAL FRAMEWORK

Maps of PV on isentropic surfaces are an extremely useful tool to study atmospheric processes and in particular the stratospheric polar vortex, because of the conservation properties of PV as well as the PV inversion principle (Hoskins *et al.*, 1985; McIntyre, 2015). The motion within the stratosphere is predominantly horizontally layer-wise; isentropic surfaces are approximately equal to material surfaces as radiative heating is weak and slow. For adiabatic and frictionless flow, PV is materially conserved. In the winter stratosphere this condition is approximately true on time-scales of 6–10 days (Hoskins *et al.*, 1985).

The Rossby–Ertel PV is defined as

$$Q = \frac{\boldsymbol{\zeta}_{\mathrm{a}} \cdot \nabla\theta}{\rho}, \tag{1}$$

where ρ is the density, $\boldsymbol{\zeta}_{\mathrm{a}}$ is the vector of absolute vorticity, θ is the potential temperature and ∇ denotes the three-dimensional gradient operator. Under the small-slope and hydrostatic approximations, the PV is given as

$$Q = -g(\zeta_\theta + f)\frac{\partial\theta}{\partial p}. \tag{2}$$

Here, g is acceleration due to gravity, f is the Coriolis parameter, p is pressure, and

$$\zeta_\theta = \mathbf{k} \cdot (\nabla_\theta \times \mathbf{v}) = \left(\frac{\partial v}{\partial x} - \frac{\partial u}{\partial y}\right)_\theta \tag{3}$$

is the isentropic (relative) vorticity, for simplicity expressed in local Cartesian coordinates, where $\mathbf{v} = (u, v)$ is the horizontal velocity vector and ∇_θ is the horizontal gradient operator along isentropic surfaces. The subscript θ denotes that the horizontal derivatives are taken at constant θ.

PV increases exponentially with height; therefore there is no straightforward way of identifying equivalent PV contours, for example, on the vortex edge, between any two isentropic surfaces. More generally, PV refers to the class of quantities of the form

$$Q^* = \frac{\boldsymbol{\zeta}_{\mathrm{a}} \cdot \nabla\eta(\theta)}{\rho}, \tag{4}$$

with an arbitrary function $\eta(\theta)$, where the choice $\eta(\theta) = \theta$ recovers the usual Rossby–Ertel PV. In particular, Lait (1994) introduced a modified PV which was generalized by Müller and Günther (2003) to take the form

$$Q^* = \frac{\boldsymbol{\zeta}_{\mathrm{a}} \cdot \nabla\theta}{\rho}\left(\frac{\theta}{\theta_0}\right)^{-\alpha}, \tag{5}$$

where θ_0 is a reference potential temperature and α is a constant to be chosen based on the potential temperature profile for the conditions under consideration. The choice $\alpha = 4$ as used by Nakamura (1995) corresponds to a profile similar to that typically observed in the winter stratosphere. We use this here as an objective way of identifying equivalent contours on different isentropic surfaces, but when numerical values of PV are given these refer to the usual Rossby–Ertel PV. The method could be refined by fitting a slightly different exponent α for each SSW (Müller and Günther, 2003; Matthewman *et al.*, 2009), but we drop this here for simplicity.

The dynamical setting in this paper is given by the primitive equations. We use isentropic coordinates, that is, we have the potential temperature θ as vertical coordinate and the diabatic heating rate $\dot\theta = \mathrm{D}\theta/\mathrm{D}t$ as vertical velocity. For the horizontal representation we have spherical coordinates with longitude λ and latitude ϕ. The continuity equation, expressing mass conservation, reads

$$\frac{\partial\sigma}{\partial t} + \nabla_\theta \cdot (\sigma\mathbf{v}) + \frac{\partial(\sigma\dot\theta)}{\partial\theta} = 0, \tag{6}$$

where $\sigma = \rho/|\nabla\theta|$ is the isentropic mass density. The PV advection equation is

$$\frac{\mathrm{D}Q}{\mathrm{D}t} = \frac{\partial Q}{\partial t} + \mathbf{v} \cdot \nabla_\theta Q + \dot\theta\frac{\partial Q}{\partial\theta} = \dot Q \tag{7}$$

The source term $\dot Q$ contains non-conservative processes such as diabatic heating, friction and mixing. For adiabatic and frictionless flow ($\dot Q = 0$), there is material conservation of PV ($\mathrm{D}Q/\mathrm{D}t = 0$). Combining Equations 6 and 7 using the identity Equation A6, the PV evolution equation is obtained as

$$\frac{\partial(\sigma Q)}{\partial t} + \nabla_\theta \cdot (\sigma Q\mathbf{v}) + \frac{\partial(\sigma Q\dot\theta)}{\partial\theta} = \sigma\dot Q. \tag{8}$$

In the following, the contour PV $= Q$ on the isentropic surface with potential temperature θ is denoted by $\Gamma_{Q,\theta}$. The unit normal vector to $\Gamma_{Q,\theta}$ in the direction of increasing PV, that is, inwards on the Northern Hemisphere is given as

$$\hat{\mathbf{n}} = \frac{\nabla_\theta Q}{|\nabla_\theta Q|}. \tag{9}$$

The area on the isentropic surface enclosed by $\Gamma_{Q,\theta}$, that is, the area with PV $\geq Q$, is denoted by $S_{Q,\theta}$. It may be composed of several distinct pieces.

3 | DATA AND PRE-PROCESSING

The diagnostic study is based on the ERA-40 dataset (Uppala *et al.*, 2005) produced by the European Centre for Medium-Range Weather Forecasts (ECMWF) which is retrieved from the British Atmospheric Data Centre (BADC; http://badc.nerc.ac.uk/; accessed 31 January 2019). The ERA-40 dataset is a 45-year reanalysis dataset covering the period from September 1957 to August 2002. The meteorological fields used here are PV, temperature and the horizontal velocity components. These data are provided on 23 pressure levels between 1,000 hPa and 1 hPa at 6-hourly intervals for the period of the Northern Hemispheric winter which is here taken to be the five-month period from 1 November to 31 March. The data are provided at a spectral resolution of T159 and are transferred to a regular Gaussian grid of size 320×160 on the Northern Hemisphere.

In the calculations following below, it is required to compute the radiative heating rate $\dot{\theta}$. This was done using a radiation scheme based on that of Shine and Rickaby (1989) to compute $\dot{T}$, which was then transformed into $\dot{\theta}$ using temperature profiles from the ERA-40 data. The values of $\dot{T}$, and hence $\dot{\theta}$, produced by the radiation scheme were found to have a net cooling bias, due to the fact that the long-wave cooling was up to 50% too large. To correct this, we used the fact that, globally,

$$\int \sigma\dot{\theta}\,\mathrm{d}S = 0 \tag{10}$$

on a surface of constant pressure. Invoking the small-slope and hydrostatic approximations (cf. Equation 13) a correction $c(p)$ is determined for each pressure level such that

$$\int \frac{\partial p}{\partial\theta}\dot{\theta}_c\,\mathrm{d}S = 0, \tag{11}$$

where $\dot{\theta}_c = \dot{\theta} + c(p)$ is the corrected value of $\dot{\theta}$.

The data are linearly interpolated from the pressure levels to isentropic surfaces from 450 to 1,050 K at intervals of 50 K. Some fields in the ERA-40 dataset are also provided directly on isentropic surfaces but the highest isentropic surface provided is 850 K which is far below the highest isobaric surface provided which is at 1 hPa. PV fields on isentropic surfaces were used to check the validity of the interpolation procedure from pressure to isentropic surfaces.

4 | INTEGRAL DIAGNOSTICS

Lagrangian integral diagnostics of mass and circulation on isentropic PV contours are used to study the structure and dynamics of the polar vortex. These are invariants of adiabatic and frictionless flow and are changed only by non-conservative processes, that is, by diabatic heating and frictional forces (including real or numerical mixing). A unified treatment of a more general class of integral quantities and their evolution equations is given in Appendix B.

4.1 | Mass integral

The mass (more precisely, the mass per unit potential temperature) within the contour $\Gamma_{Q,\theta}$ is given as

$$\mathcal{M}_{Q,\theta} = \int_{S_{Q,\theta}} \sigma\,\mathrm{d}S. \tag{12}$$

Under the small-slope and hydrostatic approximations, the isentropic mass density is

$$\sigma = -\frac{1}{g}\frac{\partial p}{\partial\theta}. \tag{13}$$

The time evolution of the mass integral is described without any approximations by the balance equation

$$\frac{\partial\mathcal{M}_{Q,\theta}}{\partial t} = \mathcal{V}_{Q,\theta} + \mathrm{CrIs} \tag{14}$$

with

$$\mathcal{V}_{Q,\theta} = \mathrm{CrCo1} + \mathrm{CrCo2} \tag{15}$$

and

$$\mathrm{CrCo1} = \frac{\partial}{\partial Q}\int_{S_{Q,\theta}} \sigma\dot{\theta}\frac{\partial\widetilde{Q}}{\partial\theta}\,\mathrm{d}S, \tag{16}$$

$$\mathrm{CrCo2} = -\frac{\partial}{\partial Q}\int_{S_{Q,\theta}} \sigma\dot{Q}\,\mathrm{d}S, \tag{17}$$

$$\mathrm{CrIs} = -\int_{S_{Q,\theta}} \frac{\partial(\sigma\dot{\theta})}{\partial\theta}\,\mathrm{d}S, \tag{18}$$

as can be seen from Equation B3 in Appendix B with the choice $G(Q) = 1$. Here and in the following, $\widetilde{Q}$ is a variable of integration as opposed to Q which denotes the PV value of the contour under consideration. The surface term CrIs is the cross-isentropic mass flux. The term $\mathcal{V}_{Q,\theta}$ is the cross-contour mass flux. Cross-contour mass flux occurs only through non-conservative processes on the contour PV $= Q$. It consists of two contributions: the cross-contour mass flux due to horizontal motion along the isentropic surface and the cross-contour mass flux due to vertical motion through a vertically tilted contour. It has been argued that, under realistic conditions, the latter may often be the dominant one (Mo *et al.*, 1998; Bühler and Haynes, 1999). Note, however, that the terms CrCo1 and CrCo2 do not represent a decomposition into these two effects as each of these terms is not invariant under a change of definition of PV (cf. Equations 4 and 5) and the vertical slope of PV $\partial\widetilde{Q}/\partial\theta$ does not necessarily correspond to a vertical tilt of the vortex edge in real space. The physical interpretation thus depends on the choice of PV. The total cross-contour mass flux $\mathcal{V}_{Q,\theta}$ is invariant under redefinition of PV, as can be seen from the invariance of $\partial\mathcal{M}_{Q,\theta}/\partial t$ and CrIs.

To the best knowledge of the authors, the partition of terms presented in Equations 14–18 is novel and has not been given in the literature before. Put in our notation, eq. 2.3 in Nakamura (1995) reads

$$\frac{\partial\mathcal{M}_{Q,\theta}}{\partial t} = \mathrm{CrCo2} + \mathrm{Vert}, \tag{19}$$

with

$$\mathrm{Vert} = -\frac{\partial}{\partial\theta}\int_{S_{Q,\theta}} \sigma\dot{\theta}\,\mathrm{d}S, \tag{20}$$

which is consistent with our result as Vert $=$ CrCo1 $+$ CrIs by virtue of identity Equation A4. Note that Equations 19 and 20 do not describe a decomposition into cross-contour and cross-isentropic mass flux. Moreover, both terms on the right-hand side of Equation 19 are not invariant under redefinition of PV which renders their physical interpretation difficult.

Our formulation of the mass balance is particularly suitable for evaluation from data as it involves only surface integrals which are numerically stable to calculate (as opposed to line integrals). We here calculate the mass integral according to Equation 12 from the reanalysis data invoking Equation 13 for the isentropic mass density. Then the cross-contour mass

flux $\mathcal{V}_{Q,\theta}$ is diagnosed as the residual from Equation 14. The terms $\partial\mathcal{M}_{Q,\theta}/\partial t$ and CrIs are calculated from the timeseries of the mass integral and the reanalysis data. Finite-difference approximations are used for the time and vertical spatial derivatives. The details of the numerical calculation of surface integrals are given in Appendix D.

4.2 | Circulation

The (absolute) circulation around the contour $\Gamma_{Q,\theta}$ is defined as

$$C_{Q,\theta} = \oint_{\Gamma_{Q,\theta}} \mathbf{u}_\mathrm{a} \cdot \mathrm{d}\mathbf{r} = \int_{S_{Q,\theta}} \sigma\widetilde{Q}\,\mathrm{d}S, \tag{21}$$

where $\mathbf{u}_\mathrm{a}$ is the three-dimensional fluid velocity in an absolute reference frame. The second equality in Equation 21 follows from Stokes's theorem and is exact, not involving any fluid-dynamical approximations.

The time evolution of the circulation $C_{Q,\theta}$ is given as

$$\frac{\partial C_{Q,\theta}}{\partial t} = Q\mathcal{V}_{Q,\theta} - \int_{S_{Q,\theta}} \frac{\partial(\sigma\widetilde{Q}\dot{\theta})}{\partial\theta}\,\mathrm{d}S + \int_{S_{Q,\theta}} \sigma\dot{Q}\,\mathrm{d}S. \tag{22}$$

This follows from Equation B3 in Appendix B with the choice $G(Q) = Q$. The goal is now to isolate the effect of frictional forces on the circulation. The source term in the PV equation in Cartesian coordinates is (Haynes and McIntyre, 1987)

$$\rho\dot{Q} = \nabla\cdot(\dot{\theta}\boldsymbol{\zeta}_\mathrm{a} + \mathbf{F}\times\nabla\theta) = \boldsymbol{\zeta}_\mathrm{a}\cdot\nabla\dot{\theta} + (\nabla\times\mathbf{F})\cdot\nabla\theta. \tag{23}$$

Here $\mathbf{F} = (F_1, F_2, 0)$ is a body force (per unit mass) which summarizes all frictional and other forces. The main contribution in the context of the polar vortex may well be gravity wave dissipation. We now invoke the small-slope and hydrostatic approximations and, for the moment, use local Cartesian coordinates as horizontal coordinates to simplify the algebra. We then get

$$\begin{aligned}\sigma\dot{Q} &= \mathbf{k}\cdot(\nabla_\theta\times\mathbf{F}) + (\zeta_\theta + f)\frac{\partial\dot{\theta}}{\partial\theta} \\ &\quad + \frac{\partial v}{\partial\theta}\left(\frac{\partial\dot{\theta}}{\partial x}\right)_\theta - \frac{\partial u}{\partial\theta}\left(\frac{\partial\dot{\theta}}{\partial y}\right)_\theta\end{aligned} \tag{24}$$

All horizontal derivatives are taken at constant θ. Using the relationship

$$\sigma Q = \zeta_\theta + f \tag{25}$$

from Equations 2 and 13 we have

$$\frac{\partial(\sigma Q\dot{\theta})}{\partial\theta} = (\zeta_\theta + f)\frac{\partial\dot{\theta}}{\partial\theta} + \dot{\theta}\frac{\partial\zeta_\theta}{\partial\theta}, \tag{26}$$

and thus

$$\sigma\dot{Q} - \frac{\partial(\sigma Q\dot{\theta})}{\partial\theta} = \mathbf{k}\cdot\left[\nabla_\theta\times\left(\mathbf{F} + \dot{\theta}\frac{\partial\mathbf{v}}{\partial\theta}\right)\right]. \tag{27}$$

Recognising Stokes's theorem yields

$$\begin{aligned}\frac{\partial C_{Q,\theta}}{\partial t} &= Q\mathcal{V}_{Q,\theta} + \oint_{\Gamma_{Q,\theta}} \mathbf{F}\cdot\mathrm{d}\mathbf{r} + \oint_{\Gamma_{Q,\theta}} \dot{\theta}\frac{\partial\mathbf{v}}{\partial\theta}\cdot\mathrm{d}\mathbf{r} \\ &= Q\mathcal{V}_{Q,\theta} + \oint_{\Gamma_{Q,\theta}} \mathbf{F}\cdot\mathrm{d}\mathbf{r} + \int_{S_{Q,\theta}} \mathbf{k}\cdot\left[\nabla_\theta\times\left(\dot{\theta}\frac{\partial\mathbf{v}}{\partial\theta}\right)\right]\mathrm{d}S.\end{aligned} \tag{28}$$

The balance equation for the circulation depends only on fields on the contour $\Gamma_{Q,\theta}$; there is no cross-isentropic flux in accordance with the impermeability theorem for PV (Haynes and McIntyre, 1987). Equation 28 is consistent with the PV evolution equation in flux form (Haynes and McIntyre, 1987)

$$\frac{\partial(\sigma Q)}{\partial t} + \nabla_\theta\cdot(\sigma Q\mathbf{v}) + \nabla_\theta\cdot\mathbf{J} = 0, \tag{29}$$

with the non-advective PV flux $\mathbf{J} = \mathbf{J}_{\dot{\theta}} + \mathbf{J}_\mathbf{F}$ being given to within the hydrostatic approximation by

$$\mathbf{J}_{\dot{\theta}} = \left(\dot{\theta}\frac{\partial v}{\partial\theta}, -\dot{\theta}\frac{\partial u}{\partial\theta}, 0\right) \tag{30}$$

and

$$\mathbf{J}_\mathbf{F} = (-F_2, F_1, 0) \tag{31}$$

The circulation balance equation given here is consistent with expressions given previously by Bühler and Haynes (1999) and Thuburn and Lagneau (1999), yet the present form appears to be novel, suitable for evaluation from data.

Again, only surface integrals are evaluated. Here, timeseries of the circulation are calculated from reanalysis data using Equation 21 with the isentropic mass density given by Equation 13. Then the frictional term $\oint_{\Gamma_{Q,\theta}} \mathbf{F}\cdot\mathrm{d}\mathbf{r}$ is diagnosed as the residual from Equation 28. The cross-contour mass flux $\mathcal{V}_{Q,\theta}$ is already known from the mass balance; $\partial C_{Q,\theta}/\partial t$ is calculated from the timeseries of the circulation and the surface integral of the last term of Equation 28 is represented and evaluated in spherical coordinates. Time derivatives as well as horizontal and vertical spatial derivatives are approximated by finite differences.

4.3 | Vortex area and equivalent latitude

More traditional Lagrangian diagnostics to monitor the state of the polar vortex are the vortex area (Butchart and Remsberg, 1986)

$$A_{Q,\theta} = \int_{S_{Q,\theta}} \mathrm{d}S \tag{32}$$

and the related equivalent latitude

$$\phi_\mathrm{eq} = \arcsin\left(1 - \frac{A_{Q,\theta}}{2\pi a^2}\right), \tag{33}$$

where a denotes the radius of the earth. The balance equation for the vortex area is given in Appendix C (cf. Butchart and Remsberg, 1986). The vortex area is not an invariant of adiabatic and frictionless flow. There is still a surface integral over the horizontal divergence of the flow in the balance equation. For typical conditions in the winter stratosphere, the vortex area can be expected to be approximately invariant in the absence of non-conservative processes.

Another possibility is to use the mass-based (rather than area-based) equivalent latitude as the canonical meridional coordinate for both PV and circulation. Slightly different methods for converting mass to equivalent latitude have been described by Nakamura and Solomon (2011) and Methven and Berrisford (2015).

TABLE 1 Investigated winters with dates of SSWs and their classifications according to Charlton and Polvani (2007) and Mitchell *et al.* (2013)

Winter	Charlton and Polvani (2007)	Mitchell et al. (2013)
1978/1979	22 February – S	18 February – S
1979/1980	29 February – D	15 February – S; 16 March – D
1981/1982	4 December – D	21 January – S
1983/1984	24 February – D	26 February – D
1984/1985	1 January – S	25 December – S
1987/1988	7 December – S; 14 March – S	1 December – M; 10 March – D
1990/1991	—	—

S denotes a vortex-splitting SSW, D a vortex-displacement SSW and M a mixed event defined by Mitchell *et al.* (2013) as one which changed in state between split and displaced.

5 | RESULTS

Firstly composites of mass and circulation are presented to identify the features of a typical winter. Then we look in detail at the winter 1978/79 which has been well studied as it displays a particularly prominent and well-defined vortex-splitting SSW. This is followed by a series of case-studies of six other winters, four of which included a single warming event, one included two warmings and one with no warmings observed. Table 1 gives a list of the winters to be looked at here. The table also gives the dates of warming events as identified by Charlton and Polvani (2007) and Mitchell *et al.* (2013). These dates will be indicated in the following Figures (not including the composites), with those of Charlton and Polvani (2007) given by a red vertical line and those of Mitchell *et al.* (2013) by a magenta vertical line.

To identify when an SSW occurs Charlton and Polvani (2007) use the WMO definition (a reversal of the zonal-mean zonal wind at 60°N and 10 hPa). These were then classified as a splitting event if two vortices exist with circulations having a ratio of at least 2:1 and a displacement event if not. Mitchell *et al.* (2013) calculated the area, centroid latitude and aspect ratio of the vortex. They then used a hierarchical clustering algorithm to classify the vortex as split, displaced or stable. If within 5 days of the date of the warming the event changed from split to displaced, it was classified as a mixed event.

We mainly focus on the $\theta = 850$ K isentropic surface as representative of the middle of the stratosphere, typically located at a height of about 32 km close to the North Pole, but also look at the $\theta = 600$ K, $\theta = 950$ K and $\theta = 1,050$ K isentropic surfaces which are typically located at heights of about 24, 35 and 37 km, respectively.

5.1 | Composites

Figure 1 shows the composites of mass $\mathcal{M}_{Q,\theta}$ and circulation $\mathcal{C}_{Q,\theta}$ at $\theta = 850$ K from 1 November to 31 March for the winters in Table 1. The pole is at the bottom of the figures, where the higher valued PV contours also lie. Lower valued PV contours contain more mass and have higher circulation than the higher valued ones. The typical locations of the vortex and the surf zone at the beginning of the winter are indicated, keeping in mind that the year-to-year variability is considerable.

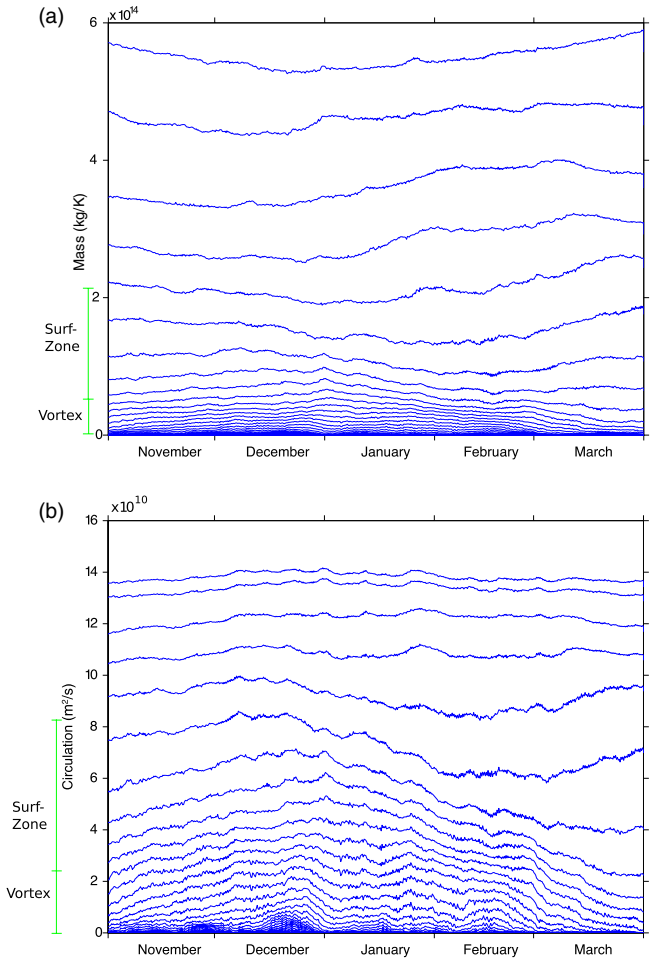

FIGURE 1 Composite of (a) mass integral $\mathcal{M}_{Q,\theta}$ and (b) circulation $\mathcal{C}_{Q,\theta}$ at $\theta = 850$ K, with 33 PV contours, equally spaced between −43.6 PVU and 1,838 PVU [Colour figure can be viewed at wileyonlinelibrary.com]

At the bottom of the mass plot, the gradient of PV can be seen to increase as the PV contours become more densely packed. This can be interpreted as the edge of the vortex which is characterized by a transition from high to low gradients of PV. From the start of November, the mass within the vortex increases steadily until the end of the month when it remains

fairly similar for the duration of the winter until it starts to decrease again at the end of February. The mass between the contours surrounding the vortex increases over the course of the winter, and can be identified as the surf zone.

In the circulation plot, the edge of the vortex is not as well defined as in the mass plot. Here the contours at the edge of the vortex have circulation values of roughly $2.5{\times}10^{10}$ m^2/s at the start of the winter. While the circulation does not show the difference between the vortex edge and the surf zone as clearly as the mass does, the circulation provides a much clearer view of the area inside the vortex, and changes in circulation over the winter are slightly clearer in some cases. Many of the features that are seen in one plot can be seen in the other, as would be expected from the mass-circulation relation of Equation B9. Due to the added detail of the circulation, the mid-winter period of relatively steady levels of mass has a more changeable circulation with an average of decreasing circulation over the period, though at a much slower rate than the decrease in circulation towards the end of the winter.

The mass and circulation composites at $\theta = 600$ K, 950 K and 1,050 K (not shown) display very similar features, with features generally more prominent at $\theta = 950$ K and 1050 K and less prominent at $\theta = 600$ K.

5.2 | Winter 1978/1979

Figure 2 shows the equivalent latitude at $\theta = 850$ K for the winter 1978/1979. A signature of the warming event can clearly be seen at the end of February. This is characterized by the area of the vortex decreasing immediately after the sudden warming for a period of approximately 2 weeks. Another prominent feature is the shrinking of the area of the vortex in the time leading up to the vortex-splitting event. This is seen as increasing equivalent latitude from about mid-January to the splitting event. It is also possible to observe a sharpening of the PV gradient at the vortex edge, seen as a greater concentration of the contours towards the top of the figure.

Figure 3 shows the mass integral for the three isentropic surfaces $\theta = 1,050$ K, 850 K and 600 K. In the mass plot for $\theta = 850$ K, it is possible to see the decreasing mass of air inside the vortex in the lead-up to the warming event as air is stripped from the main vortex and mixed into the surrounding surf zone. This is a similar picture to that seen in the equivalent latitude plot, although the stripping of air in the lead-up to the warming event is much clearer in the mass plot, where the mass inside some of the outer contours increases quite dramatically as the winter progresses. Also the sharpening of the PV gradient at the vortex edge before and during the SSW is more prominent in the mass plot than in the more traditional equivalent latitude plot. In fact, the vortex boundary is characterized by a sudden change from high to low PV gradients.

Complementarily, mass as a function of PV at $\theta = 850$ K is displayed in Figure 4 for various instants during the winter. In early November the "PV staircase" of alternating steep

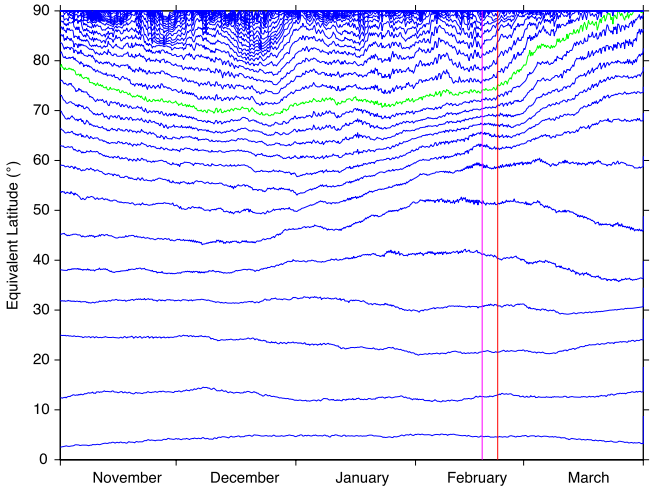

FIGURE 2 Equivalent latitude ϕ_{eq} for the winter 1978/1979 at $\theta = 850$ K, with 49 PV contours, equally spaced between −43.6 PVU and 1367.6 PVU. Marked in green is the PV contour $Q = 338.6$ PVU on the vortex edge for which the cross-contour mass flux $\mathcal{V}_{Q,\theta}$ is shown in Figure 5. Vertical lines mark dates of warming events as identified by Charlton and Polvani (2007) (red) and Mitchell *et al.* (2013) (magenta) [Colour figure can be viewed at wileyonlinelibrary.com]

and relatively shallow PV gradients (Dritschel and McIntyre, 2008) is clearly present. By mid-January the maximum value of PV has increased and the PV increases more sharply towards the centre of the vortex. The "step" that appears between around 50 and 300 PVU in early November is not as clearly visible in mid-January, though some features still remain at around 150 PVU. At the time of the SSW the staircase structure can be seen again as well as the fact that the maximum value of PV has dramatically reduced since mid-January. In mid-March the maximum value of PV has again increased, though not to the same value it attained in mid-January. The difference in gradient between the vortex edge and surrounding area of surf zone is much sharper than it is earlier in the winter.

In the mass plot for $\theta = 1,050$ K, similar features to those on the $\theta = 850$ K surface can be identified, though with some notable differences. Because the mass density is smaller on this higher isentropic level, the mass of air within the vortex is smaller than for 850 K. There also appears to be less contrast between the vortex and the surrounding air. As expected, the PV of the contours making up the edge of the vortex takes higher values than on the 850 K surface. The vortex seems to be more established at the start of November at 1,050 K than it is at 850 K where the mass of the vortex increases until late in November. After this, the evolution of the vortex appears similar on both levels until the signal of the warming starts to become visible in the diagnostic. Over the course of the winter, the gradient of PV at the vortex edge gradually sharpens, becoming steeper until mid-February when the breakdown of the vortex causes it to become shallower. The relative change in mass within PV contours on this level is much larger over the course of the winter than for those on the lower level. The decrease in mass within the vortex associated with the

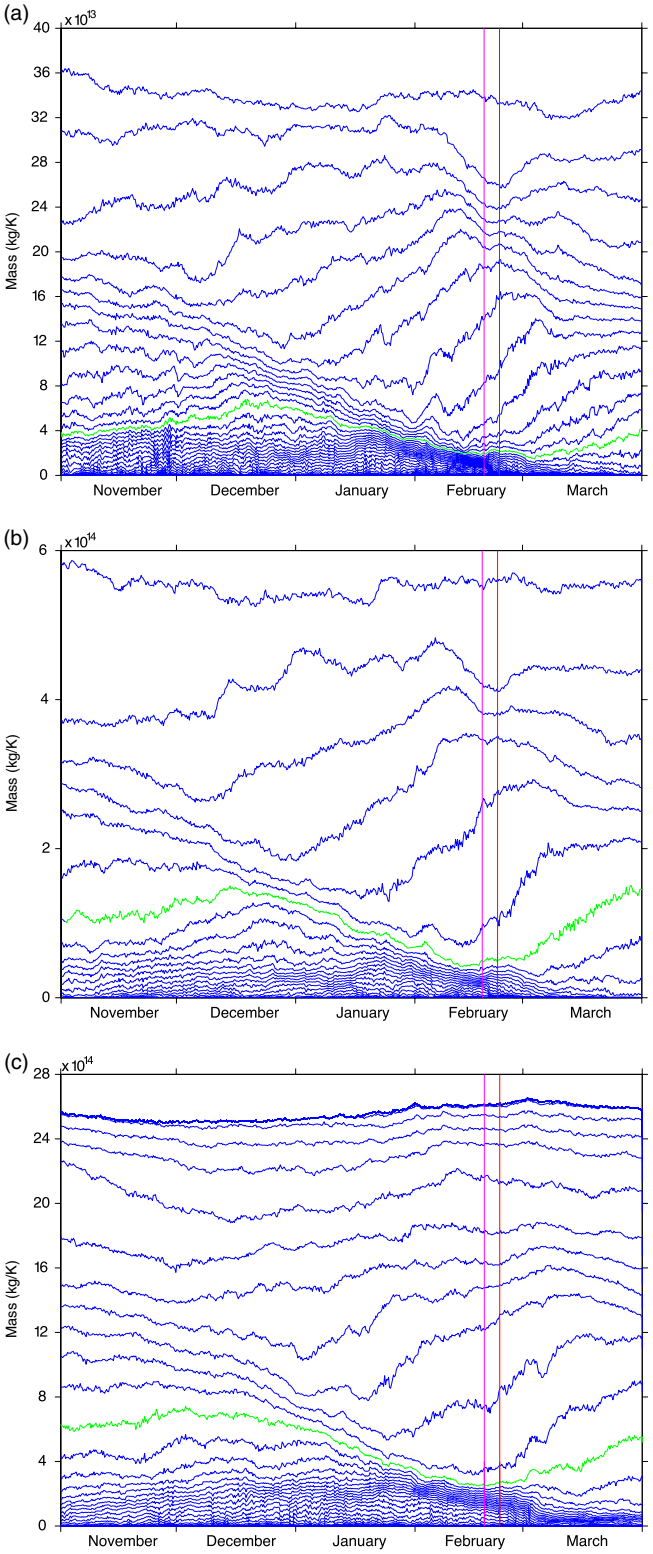

FIGURE 3 Mass integral $\mathcal{M}_{Q,\theta}$ for the winter 1978/1979 (a) at $\theta = 1,050$ K. with 46 PV contours, equally spaced between 15.2 PVU and 2661.2 PVU, (b) at $\theta = 850$ K, with 23 PV contours, equally spaced between −14.2 PVU and 1279.4 PVU, and (c) at $\theta = 600$ K, with 30 PV contours, equally spaced between 0.5 PVU and 213.65 PVU. Marked in green in all three panels are the PV contours on the vortex edge for which the cross-contour mass flux $\mathcal{V}_{Q,\theta}$ is shown in Figure 5 [Colour figure can be viewed at wileyonlinelibrary.com]

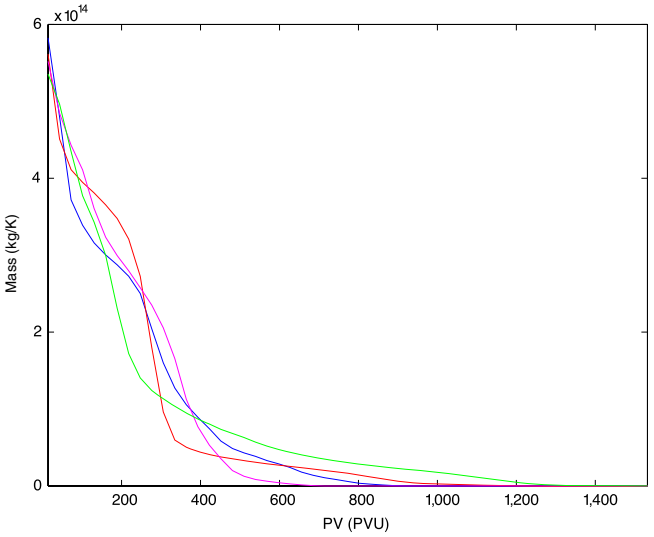

FIGURE 4 Mass integral $\mathcal{M}_{Q,\theta}$ against PV at $\theta = 850$ K for various times during the winter 1978/1979: early November (blue line), mid-January (green line), the time of the SSW in late February (red line), and mid-March (magenta line) [Colour figure can be viewed at wileyonlinelibrary.com]

warming seems to become visible earlier for 1,050 K than for 850 K, especially on the inner contours of the vortex, but the vortex remains relatively intact until the signal of the warming becomes visible on 850 K when a similar picture occurs on both levels.

In the mass plot for the 600 K isentropic level, the effect of the splitting event around 20 February can be seen less clearly, though it is still possible to identify its signature. Unlike the other two levels, however, in which the signal of the warming can be seen to occur at roughly the same time, on this level the signal cannot be seen to occur until slightly later, around the end of February. Comparing PV maps for the three different isentropic levels for the period in question (not shown) reveals that the vortex appears to split nearly simultaneously on all levels, as noted in Matthewman *et al.* (2009). The difference in the signal appearing in the mass plots is not caused by the split occurring later on the lower level, but the fact that the two "daughter" vortices appear to be longer-lived on the 600 K surface than they are on the 850 and 1,050 K surfaces. This could be related to the longer radiative time-scale in the lower stratosphere. The smaller vortex which breaks down allowing the other to recover to its pre-warming state seems to become eroded much later and at a faster rate on the lower level, while on the upper levels the higher PV air in the centre of the vortex appears to mix into the rest of the air in the interior giving the vortex a more uniform PV structure inside. This vortex will then break down slightly earlier, but at a slower rate than it does at 600 K. The other obvious feature of the plot is the fact that the values of PV are much smaller here than on the higher levels, as well as the fact that there is a much higher mass within the vortex than on the higher levels. There is also a much smaller relative change in mass due to the erosion of the vortex during the warming event than there is on the higher levels. The poleward trend of outer PV contours resulting in a

sharpening of the gradient on the vortex edge that can be seen on the other levels is also present here. After the occurrence of the sudden warming, the mass within the vortex decreases similarly to the other levels along with a decrease in the PV gradient on the vortex edge.

Figure 5 displays the cross-contour mass flux $\mathcal{V}_{Q,\theta}$ for a contour on the vortex edge on the three isentropic surfaces $\theta = 1,050\,\mathrm{K}$, 850 K and 600 K. These mass fluxes can be interpreted as mass fluxes into or out of the vortex. Positive values of $\mathcal{V}_{Q,\theta}$ correspond to the mass inside the vortex increasing, while negative values correspond to an outflux of mass of air being mixed into the surrounding surf zone.

Here and in the following, the PV contours on the vortex edge are identified as follows. For $\theta = 850\,\mathrm{K}$ a PV contour on the vortex edge is chosen by eye from visual inspection of PV maps before and at the time of the sudden warming. Equivalent PV contours on other isentropic surfaces are then identified as those with the same value of generalized PV Q^* as given by Equation 5. Small deviations occur as only a discrete set of PV contours is considered at each isentropic level. PV values given in the figures always refer to Rossby–Ertel PV.

We remark that the definition of the vortex boundary is often slightly ambiguous. A particular PV contour does not necessarily stay on the vortex edge for the whole of the season; in fact, the same PV contour may reside in the vortex before the SSW and may be expelled to the surf zone afterwards.

For the 850 K isentropic surface, a comparison with Figure 3 reveals that most of the changes to the mass within the vortex at this level are due to the cross-contour mass flux $\mathcal{V}_{Q,\theta}$. At the start of the winter there are small amounts of air constantly moving into and out of the PV contour. It is hard to tell how much of these fluctuations are a real signal, maybe due to the diurnal cycle, and how much is due to noise in the data. The net mass flux up to well into December appears to be around zero. In the run-up to the SSW, it is possible to identify in the plot several vortex-stripping events. These are much longer-lived than the regular fluctuations, on the scale of 5 to 10 days. One such event, for example, can be observed in mid-January. Figure 6 shows the PV map for an instant during the identified stripping event. It can clearly be seen that this event corresponds to a filament of air being pulled from the vortex, and examining a time series of PV maps (not shown) reveals it being mixed into the surf zone surrounding the vortex. Other such events can be observed throughout the winter, reducing the size of the vortex as mass is removed from it, up until mid- to late February when the vortex-splitting SSW occurred in this winter. The net cross-contour mass flux between mid-December and the SSW is clearly negative. After the warming event, the size of the vortex increases again with a strongly positive net mass influx up to the end of March, although the PV gradient at the edge of the vortex does not return to the same level as at the start of the winter.

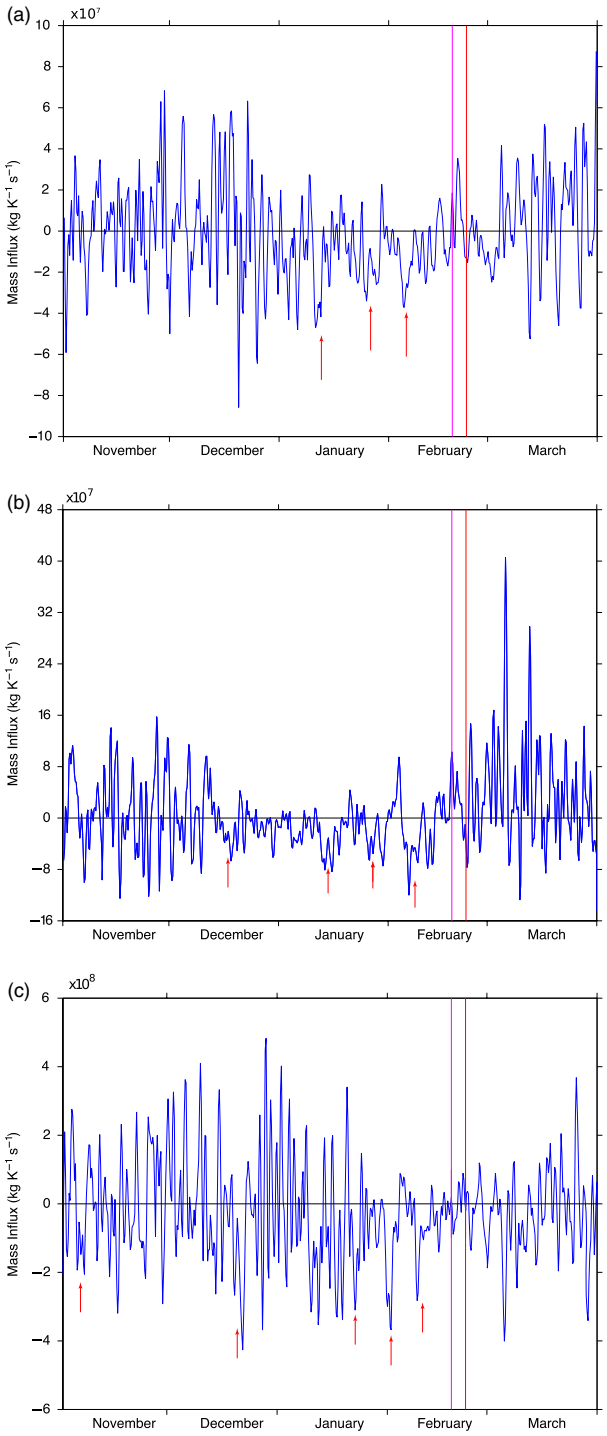

FIGURE 5 Cross-contour mass flux $\mathcal{V}_{Q,\theta}$ for the winter 1978/1979 for (a) $Q = 779.6\,\mathrm{PVU}$ and $\theta = 1,050\,\mathrm{K}$, (b) $Q = 338.6\,\mathrm{PVU}$ and $\theta = 850\,\mathrm{K}$, and (c) $Q = 88.7\,\mathrm{PVU}$ and $\theta = 600\,\mathrm{K}$. Red arrows in all three panels mark significant mass-stripping events [Colour figure can be viewed at wileyonlinelibrary.com]

At the 1,050 K level a very similar pattern of air being moved into and removed from the vortex can be seen. Between mid-December and the SSW there are a couple of persistent events of stripping of PV filaments. At the start of the winter and after the SSW there is a positive net mass influx into the vortex. A comparison with the corresponding mass plot in Figure 3 shows evidence of vertical movement of air through diabatic effects. The cross-contour term $\mathcal{V}_{Q,\theta}$ is much

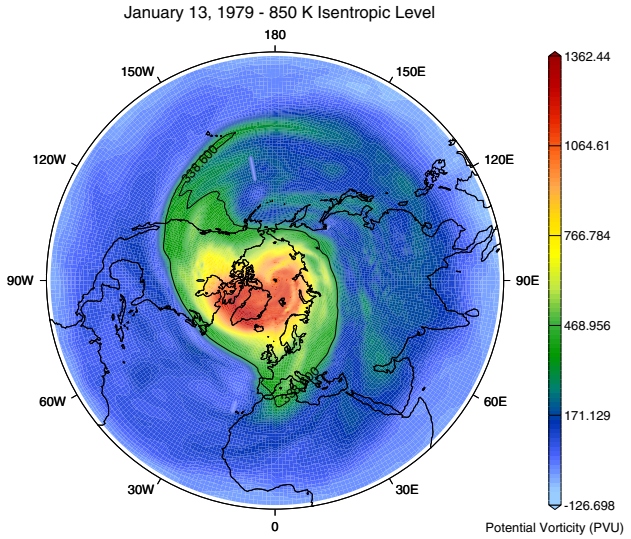

FIGURE 6 PV map for 13 January 1979 at $\theta = 850$ K. The PV contour $Q = 338.6$ PVU on the vortex edge is marked. A filament of air can be seen being stripped from the vortex and mixed into the surrounding surf zone [Colour figure can be viewed at wileyonlinelibrary.com]

more positive than the change in mass would indicate. This means that there is a big negative contribution to the mass balance from the cross-isentropic term CrIs which is indicative of vertical movement of air downward within the vortex, as noted by for instance by Waugh *et al.* (1994).

On the 600 K level, the stripping process is less localized in time than on the two higher levels with stripping events occurring throughout the winter. As with the higher levels, these can also be matched up to tongues of air being pulled from the vortex by looking at PV maps of this isentropic surface (not shown). The cross-contour mass flux $\mathcal{V}_{Q,\theta}$ is predominantly negative, even at the start of the winter and after the SSW. Integrating $\mathcal{V}_{Q,\theta}$ on the 88.7 PVU contour over the period of the winter shows a transport of an amount of air out of the vortex of the order of 10^{15} kg/K which is more than the amount of air inside the contour at the start of the winter. This indicates that air is constantly being moved vertically within the vortex onto this isentropic surface replacing that being removed via stripping. Indeed, some of the mass moving out of the vortex into the surf zone is probably due to vertical movement across the slanted edge of the vortex rather than necessarily from horizontal movement across the boundary.

Comparison of the plots of the cross-contour mass flux for the three isentropic levels shows that the more prominent stripping events can be identified as being common to each of the levels. They appear to occur at a similar time on all surfaces, although they do not seem to occur exactly simultaneously on all levels. There does not seem to be a preferred direction for the propagation of these events. Some events occur first on the upper level and propagate downwards, while others appear first on the lower level and propagate upward to the higher levels. Corresponding PV maps (not shown) confirm this picture, where it is possible to see filaments starting to be removed on one level slightly before they become visible on the other levels and the propagation of these either upward or downward.

Before and during the SSW on all of the three isentropic levels, the mass decreases in the polar vortex and increases on the poleward side of the surf zone, but the rapid increase in the surf zone more than compensates the stripping loss of the polar vortex. At the same time there is a vertical downward movement of air within the vortex. This gives some indication of a mass flux from the Tropics into the surf zone (poleward stripping). A concise and closed mass budget quantifying the contributions of cross-contour mass flux $\mathcal{V}_{Q,\theta}$ and cross-isentropic mass flux CrIs could be the topic of a further study.

The circulation for the three isentropic levels $\theta = 1,050$ K, 850 K and 600 K is shown in Figure 7. The plot for $\theta = 850$ K shows a similar picture to those of the equivalent latitude and mass plots with the warming event clearly visible around 20 February. While the vortex edge and formation of the surf zone are not as striking as in the mass plot, the circulation has more resolution around the area of the vortex. The circulation of the contours inside the vortex increases initially until mid-December to early January when it reaches its peak. The contours nearest to the outside of the vortex reach their maximum circulation earliest and begin falling, followed by the inner contours. The circulation then stabilizes briefly around mid-February in the lead-up to the vortex splitting before it decreases as the warming progresses. Following the warming, the circulation then starts to increase for the outer PV contours, though many of the inner PV contours disappear as the circulation goes towards zero.

The circulation on the 1,050 K surface takes larger values than on the 850 K level. The circulation on the contours making up the main part of the vortex takes values roughly twice as large at the start of the winter, though when the circulation starts to decrease around the time of the sudden warming the circulations on the two levels become very similar. The circulation can be seen to decrease slightly earlier on the 1,050 K level than on the 850 K level on some of the outer contours, similar to the way that the mass starts to decrease earlier on the higher level. Closer to the centre of the vortex, the reduction of the circulation becomes more synchronized between the two levels. Following the splitting event, the circulation increases at a similar rate on the higher level as it does on the 850 K level. The outer contours start to decrease in circulation from mid-December, slightly earlier than they do on the 850 K level, with a much sharper decrease occurring on this level. There are some contours on the outside of the vortex whose circulation starts to rise before the sudden warming, though there does not appear to be a significant amount more than on the 850 K level.

On the 600 K level the circulation at the vortex edge is similar to that on the upper two surfaces. The changes in circulation on the vortex's outer contours are again smaller on this lower level, increasing in size on each of the higher levels in turn. The decrease in circulation that characterizes the

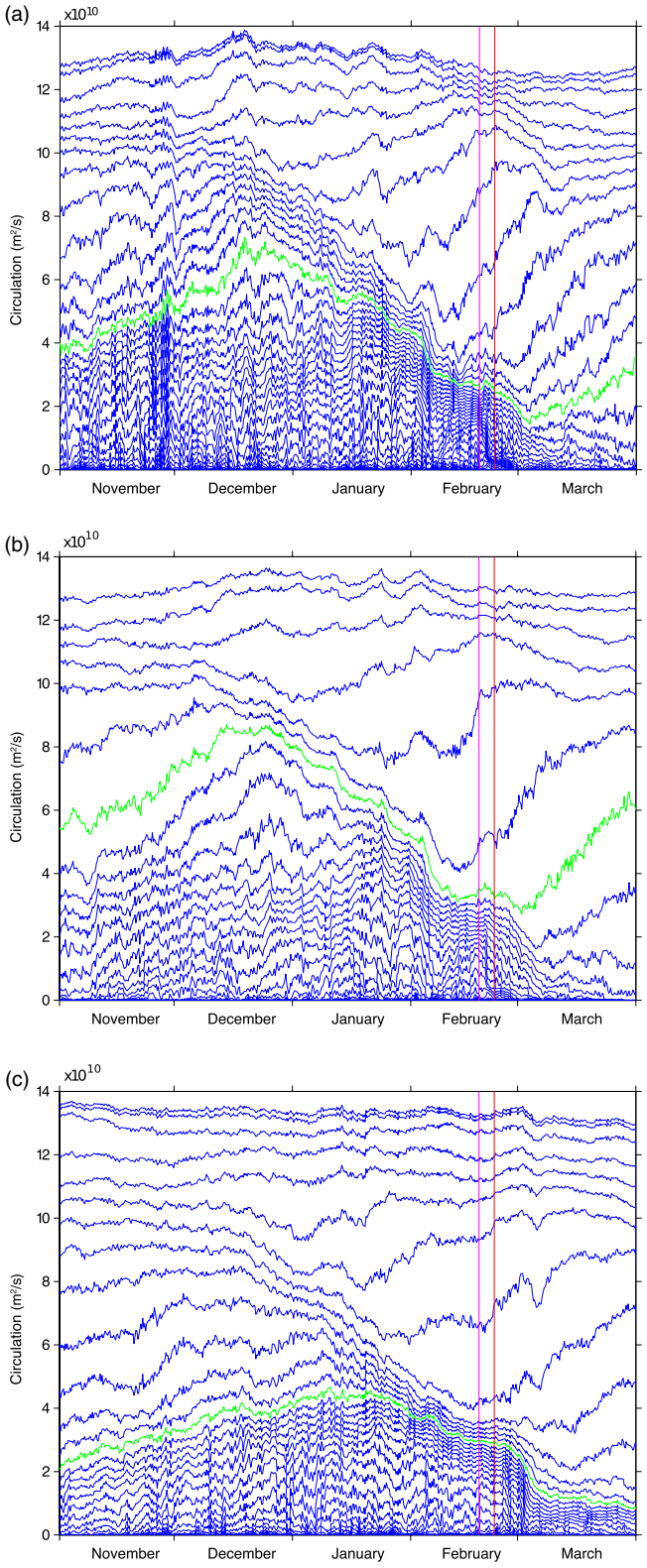

FIGURE 7 Circulation $C_{Q,\theta}$ for the winter 1978/1979 at (a) $\theta = 1,050$ K, with 46 PV contours, equally spaced between 15.2 PVU and 2661.2 PVU, (b) $\theta = 850$ K, with 23 PV contours, equally spaced between −14.2 PVU and 1279.4 PVU, and (c) $\theta = 600$ K, with 34 PV contours, equally spaced between −6.85 PVU and 235.7 PVU. Marked in green in all three panels are the PV contours on the vortex edge for which the cross-contour mass flux $\mathcal{V}_{Q,\theta}$ is shown in Figure 5 [Colour figure can be viewed at wileyonlinelibrary.com]

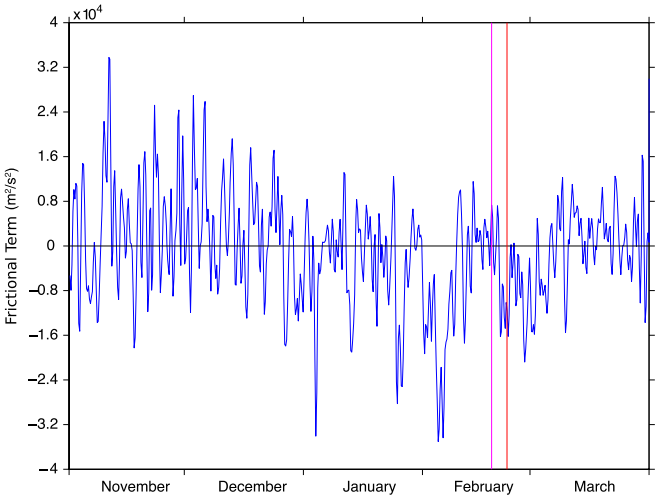

FIGURE 8 Frictional term $\oint_{\Gamma_{Q,\theta}} \mathbf{F} \cdot \mathrm{d}\mathbf{r}$ for the winter 1978/1979 for $Q = 338.6$ PVU and $\theta = 850$ K [Colour figure can be viewed at wileyonlinelibrary.com]

splitting event is seen much later on this level than on the upper levels in analogy to the mass plots; the splitting cannot be seen on this level until the very end of February and into the beginning of March while on the higher levels the signal is constrained to mid- to late February. Unlike the other levels, the PV contours on this surface which do decrease in circulation around the advent of the sudden warming do not increase again afterwards. There is a period immediately before the signal of the warming starts to become visible in this plot where the change in circulation slows down, so that for a short period of time the circulation stays nearly steady. Following this, the signal of the warming starts to become visible as the circulation decreases sharply. This plateau appears on both of the higher levels, but for a much shorter period of time.

Figure 8 shows a plot of the frictional term as diagnosed from the balance equation of the circulation. These plots are generally very noisy which makes it hard to identify any significant genuine signal. For the winter 1978/1979, between mid-January and mid-February in the run-up to the SSW there are two persistent events of removal of circulation from the contour due to frictional processes such as gravity wave drag which coincide with stripping events. However, generally when looking at other winters these events are much less well-defined than the vortex-stripping events. They occur scattered throughout the winter and there is often no correspondence to mass-stripping events or the occurrence of an SSW.

5.3 | Case-studies of other winters

5.3.1 | Winter 1984/1985

The features observed above for the winter 1978/1979 can be seen also in other winters which contain a single splitting SSW event. Figure 9 displays the mass integral on the 850 K surface for the winter 1984/1985 in which a vortex-splitting SSW event occurred at the end of the year. The area of high

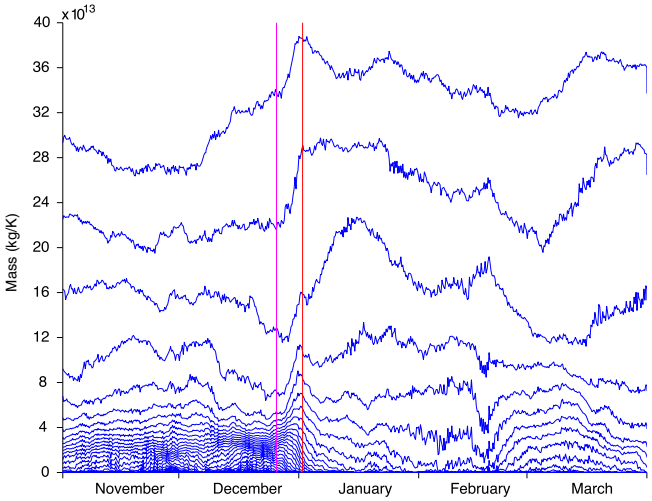

FIGURE 9 Mass integral $\mathcal{M}_{Q,\theta}$ for the winter 1984/1985 at $\theta = 850$ K, with 29 PV contours, equally spaced between 191.6 PVU and 1838 PVU [Colour figure can be viewed at wileyonlinelibrary.com]

PV gradient making up the polar vortex can clearly be seen up until the vortex-splitting event occurs. Following the warming event, the vortex reforms towards the end of February. After reforming the PV gradient is not as steep as it is in the pre-warming vortex, though a recognizable vortex is still visible in the mass plot.

Following the warming there is a period of time where the vortex is relatively undefined in the mass plot. Comparing this to maps of PV for the same period (not shown) reveals that, while there is still an area of high PV over the pole, the vortex itself is not well defined. The PV values are lower than those at the start of the winter and are spread more thinly over the pole. Over the period of January they start to become drawn together and by the start of February a more recognisable vortex is visible located over the pole.

The winter 1984/1985 also features several events where filaments of high-PV air are moved toward the edge of the vortex and pulled from it in the month leading up to the SSW event. Starting around 1 January, there is a fairly large stripping event lasting for roughly 15 days. This coincides with the splitting of the vortex and the breakdown of the smaller of the two daughter vortices.

5.3.2 | Winter 1983/1984

The winter 1983/1984 displayed a particularly prominent vortex displacement event toward the end of February. This is now investigated in an attempt to identify any differences in the pattern of stripping prior to the onset of the warming.

Figure 10 shows the mass integral and the circulation at $\theta = 850$ K as well as the cross-contour mass flux on the vortex edge at $\theta = 950$ K and 850 K. On the 850 K level, the signal of the vortex displacement event can be seen around the end of February when the mass of the vortex decreases following the displacement event. Unlike the splitting events above, the loss of mass of the PV contours does not extend to the contours around the vortex until near to the occurrence of the SSW. Most of the change in mass occurs in the early winter, while the vortex remains relatively stable throughout the mid-winter. Just prior to the date of the displacement event, there is again a poleward trend of PV contours at the edge of the vortex, before the mass inside the main body of the vortex starts to decrease as the vortex returns from being displaced to be located once more over the pole. Another feature is that the mass within the inner contours of the main vortex is fairly constant throughout the period of the winter up until the displacement event occurs.

In the circulation plot, the vortex can be seen strengthening between the start of November and mid-December when the circulation of the contours making up the vortex increases and the gradient of PV sharpens. This strengthening of the circulation of the contours making up the vortex is more gradual than in the winter 1978/1979. The circulation in the main part of the vortex remains fairly stable for most of the mid-winter period, from around mid-December until the decrease in circulation caused by the onset of the displacement SSW towards the end of February. Following the warming, a large number of the PV contours from inside the vortex decrease to zero circulation; only the outer contours start to increase in circulation again to reform the vortex before the final warming occurs. The outer contours do not show as much change for this winter as they do for 1978/1979, though this is common with the difference visible in the mass plot above.

The cross-contour mass flux at $\theta = 850$ K clearly shows several stripping events, similar to those seen for the vortex-splitting events above. These events account for significant amounts of mass being removed from the vortex as well as some events where mass is added back into the vortex. The mass-stripping events seem to occur throughout the winter, though the events become more frequent in the month or so leading up to the warming, as well as several large events in the period prior to this. At the start of the winter the PV gradient on the edge of the vortex is not as defined as later in the winter. The stripping of filaments in the early winter acts to remove air from the contours on the edge and strengthens the PV gradient there. Following the sudden warming event, and the subsequent breakdown of the vortex, there is a significant amount of air moved back into the vortex, as it reforms and starts to recover to its pre-warming state.

The mass and circulation plots for the 950 and 600 K isentropic surfaces (not shown) are both very similar to those on the 850 K surface, with the features more pronounced on the higher level and less pronounced on the lower level. The dropping in mass and circulation as the vortex breaks down following the sudden warming occurs slightly earlier on the higher surfaces than on the lower surfaces, similar to the finding above for the winter 1978/1979.

The cross-contour mass flux at $\theta = 950$ K shows similar features to that at $\theta = 850$ K, but is more pronounced and less noisy. There is a very persistent vortex-stripping event visible in the run-up to the warming event. Actually, between

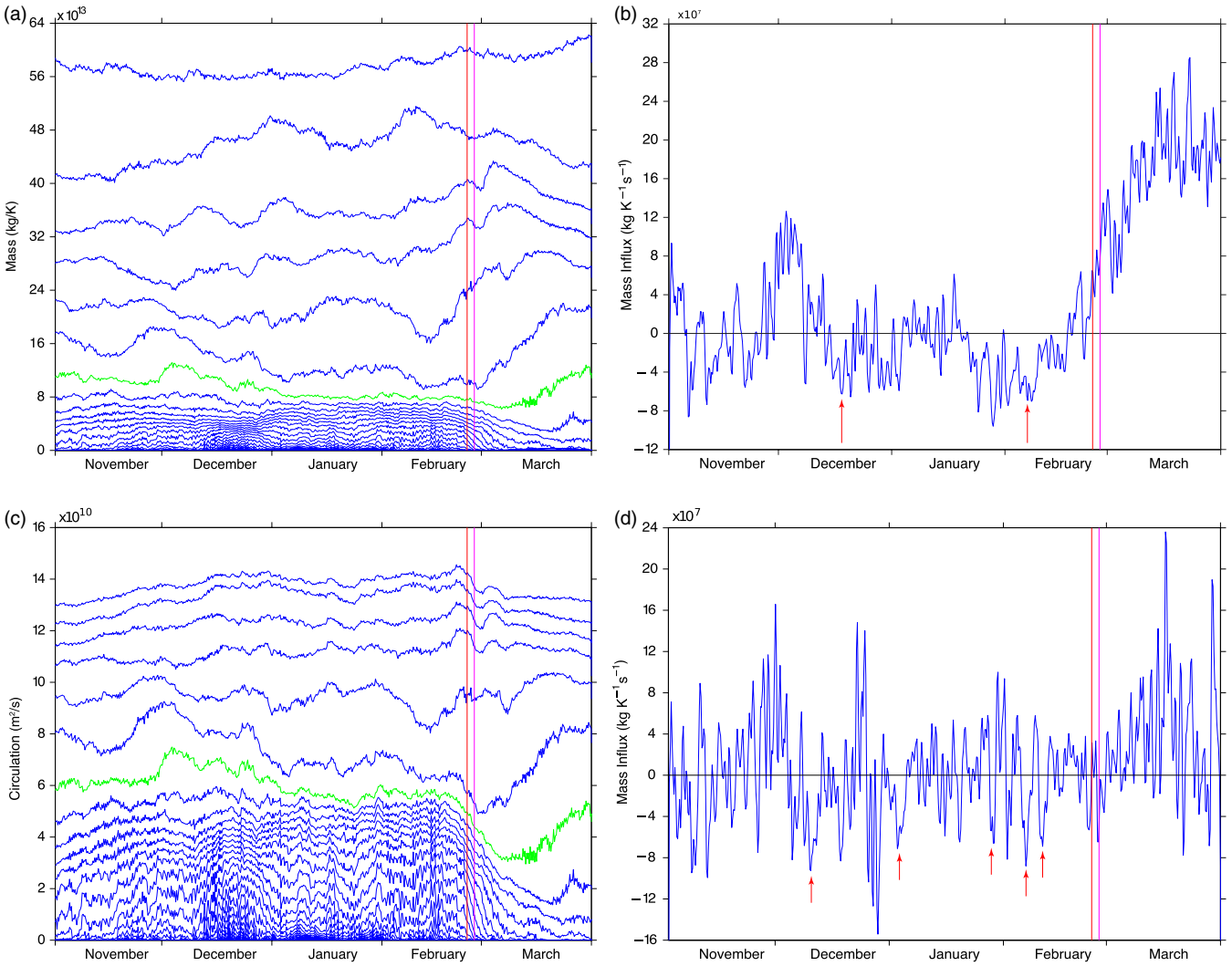

FIGURE 10 Winter 1983/84: (a) Mass integral $\mathcal{M}_{Q,\theta}$ at $\theta = 850$ K, with 27 PV contours, equally spaced between 15.2 PVU and 1544 PVU. (c) Circulation $C_{Q,\theta}$ at $\theta = 850$ K, with 30 PV contours, equally spaced between 15.2,PVU and 1720.4 PVU. Marked in green in (a) and (c) is the PV contour $Q = 368$ PVU on the vortex edge for which the cross-contour mass flux $\mathcal{V}_{Q,\theta}$ is shown in (d). (b) Cross-contour mass flux $\mathcal{V}_{Q,\theta}$ for $Q = 603.2$ PVU and $\theta = 950$ K. (d) Cross-contour mass flux $\mathcal{V}_{Q,\theta}$ for $Q = 368$ PVU and $\theta = 850$ K. Red arrows in (b) and (d) mark significant mass-stripping events [Colour figure can be viewed at wileyonlinelibrary.com]

around 20 January and 20 February the cross-contour mass flux is constantly negative for about one month. In the recovery phase of the vortex after the SSW, the cross-contour mass flux is strongly positive. The plot for $\theta = 600$ K (not shown) displays similar features but with a general bias to more negative values. This is again indicative of vertical downward movement of air within the vortex.

5.3.3 | Winter 1981/1982

For the winter 1981/1982, Charlton and Polvani (2007) identified a vortex displacement event to have occurred on 4 December. There was no event identified for December 1981 by Mitchell *et al.* (2013). However, there is a vortex-splitting event identified on 21 January by Mitchell *et al.* (2013) which was not classified as a major warming by Charlton and Polvani (2007). Around this date the vortex can be seen to split into two distinct parts, although one of the resultant vortices is relatively undefined and breaks down within a couple of days.

The mass integral and the cross-contour mass flux on the vortex edge at the 850 K isentropic level for the winter 1981/1982 are displayed in Figure 11. In the mass plot, there is not the clear signal of the vortex displacement event in December that there was for the more prominent splitting and displacement events above. During a displacement event, the vortex will be displaced from the pole before it moves back to be centred on the pole again. If the vortex as a whole remains more or less intact, there is not necessarily any movement of a significant mass of air from the interior to the exterior of the vortex for the signal to show up in the mass plot. Another possible reason is that the vortex is not yet well formed as the warming is very early in the winter. This may result in the recovering vortex becoming stronger than the pre-warming vortex as following the SSW event the vortex seems to undergo a strengthening during early December.

There is a reduction in mass seen in the mass plot around the time of the Mitchell *et al.* (2013) vortex-splitting event, which

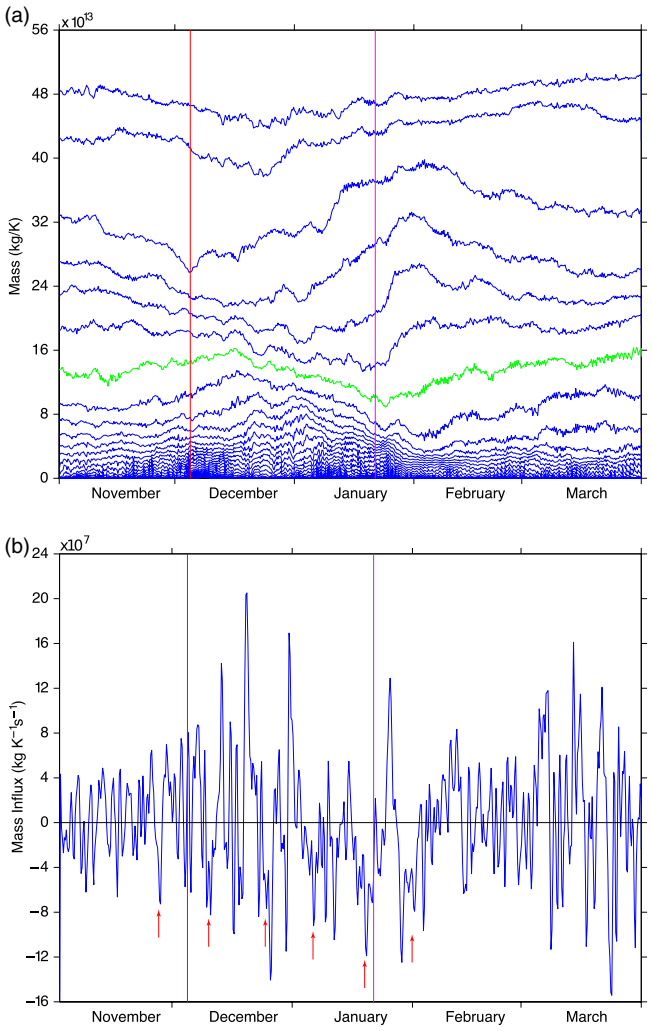

FIGURE 11 Winter 1981/1982: (a) Mass integral $\mathcal{M}_{Q,\theta}$ at $\theta = 850$ K, with 31 PV contours, equally spaced between 132.8 PVU and 1308.8 PVU. Marked in green is the PV contour $Q = 368$ PVU on the vortex edge for which the cross-contour mass flux $\mathcal{V}_{Q,\theta}$ is shown in (b), where red arrows mark significant mass-stripping events [Colour figure can be viewed at wileyonlinelibrary.com]

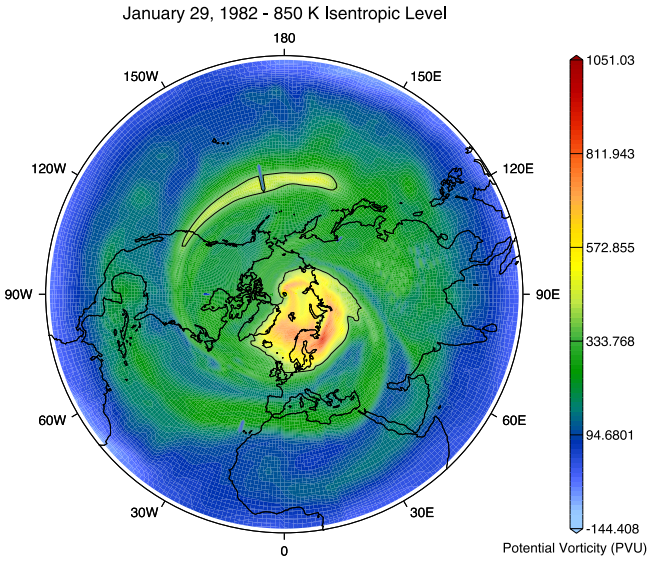

FIGURE 12 PV map for 29 January 1982 at $\theta = 850$ K. The PV contour $Q = 368$ PVU on the vortex edge is marked. A filament of air can be seen being stripped from the vortex and mixed into the surrounding surf zone [Colour figure can be viewed at wileyonlinelibrary.com]

can be considered the signal of this event, though it is not quite as pronounced as the signals of the warmings common to both identification criteria. The PV gradient on the edge of the vortex can be seen increasing prior to the splitting, as well as the decrease of the mass within it as the smaller weaker sub-vortex breaks down.

The plot of the cross-contour mass flux is similar to those for the winters examined above, indicating that most of the change in mass at $\theta = 850$ K is due to stripping of PV filaments rather than diabatic effects moving air vertically throughout the vortex. There are several stripping events in the lead-up to the vortex displacement on 4 December. These events are not as large as the corresponding events in the winters studied above, but here the vortex is not as strong as in larger stripping events. Much larger stripping events occur later in the winter once the vortex has recovered. Most of these larger events occur in the period immediately before the splitting event on 21 January. This is similar to the pattern of stripping in the lead-up to the 1978/1979 and 1984/1985 splitting events, with a large stripping event immediately before the breakdown and several other large events at roughly 15–30 days previously.

The stripping events visible in the cross-contour mass flux can again be matched up to PV maps, where filaments of PV are being removed and mixed into the surrounding air. There are in particular two large and persistent stripping events, one immediately prior to the vortex-splitting event and one shortly after it. Figure 12 displays the PV map at a time during the latter one.

The vertical structure of the mass integral (not shown) is similar to the winters examined above, with features throughout the winter occurring slightly earlier on the higher levels and propagating down through the stratosphere.

5.3.4 | Winter 1979/1980

The differing criteria of Charlton and Polvani (2007) and Mitchell *et al.* (2013) also disagree on the classifications of the SSW(s) which occurred in the winter of 1979/1980. The former classifies a single displacement to occur on 29 February, while the latter classifies a splitting on 15 February and a displacement on 16 March.

The timeseries of the mass integral for $\theta = 850$ K is shown in Figure 13. The signal of a warming event can be seen to have occurred at the end of February. This coincides in timing with the classification of a late-February displacement of Charlton and Polvani (2007). There is only a clear signal of one warming in this winter, favouring this classification over that of Mitchell *et al.* (2013). Comparing this winter, along with the prominent displacement event in 1983/1984, with the splitting events of 1978/1979 and 1984/1985, the gradient of PV that makes up the vortex appears to be greater during the

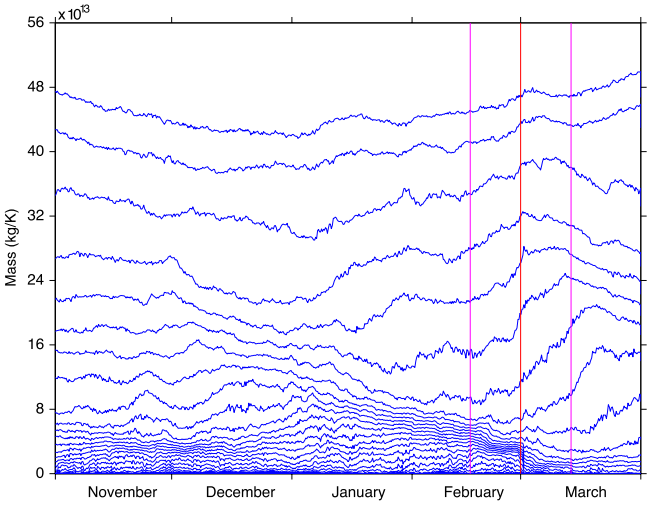

FIGURE 13 Mass integral $\mathcal{M}_{Q,\theta}$ for the winter 1979/1980 at $\theta = 850$ K, with 25 PV contours, equally spaced between 132.8 PVU and 1073.6 PVU [Colour figure can be viewed at wileyonlinelibrary.com]

winters in which splitting events occurred than in winters containing displacement events. The PV throughout the surf zone surrounding the vortex is roughly the same for each type of event, though the maximum value of the PV within the vortex on a given surface is much higher in winters containing splitting events.

5.3.5 | Winter 1987/1988

For the winter 1987/1988, the criteria of Charlton and Polvani (2007) identify two splitting events on 7 Decemberand on 14 March, while those of Mitchell *et al.* (2013) identify a mixed event, where the vortex changes in state between split and displaced, to occur on 1 December and a displacement event on 10 March.

Figure 14 shows the mass integral and the circulation at $\theta = 850$ K. The signals of the two SSW events can clearly be seen in the timeseries of the mass integral where in early December and in mid-March the mass within the vortex has depleted. Following the early splitting/mixed event, the vortex recovers more quickly than for the two splitting events of 1978/1979 and 1984/1985. For example, the splitting event of early January 1985 does not see the vortex fully recover until late February, with no clearly defined vortex visible in the intervening period. In contrast, the splitting/mixed event of December 1987 sees a drop in mass of the vortex, followed by a recovery by late December and by January the vortex is almost fully recovered to its pre-warming state. The other warming event which occurred in early winter (in 1981/1982) also shows a similar fast recovery to its pre-warming state. Following the splitting event, the vortex slowly grows in mass, with the contours immediately surrounding the main area of steep PV gradients losing mass to become part of this area of highly concentrated PV contours. By mid-February the vortex has become stronger than it was before the initial early winter event occurred. Once it reaches a peak in mass in mid- to late

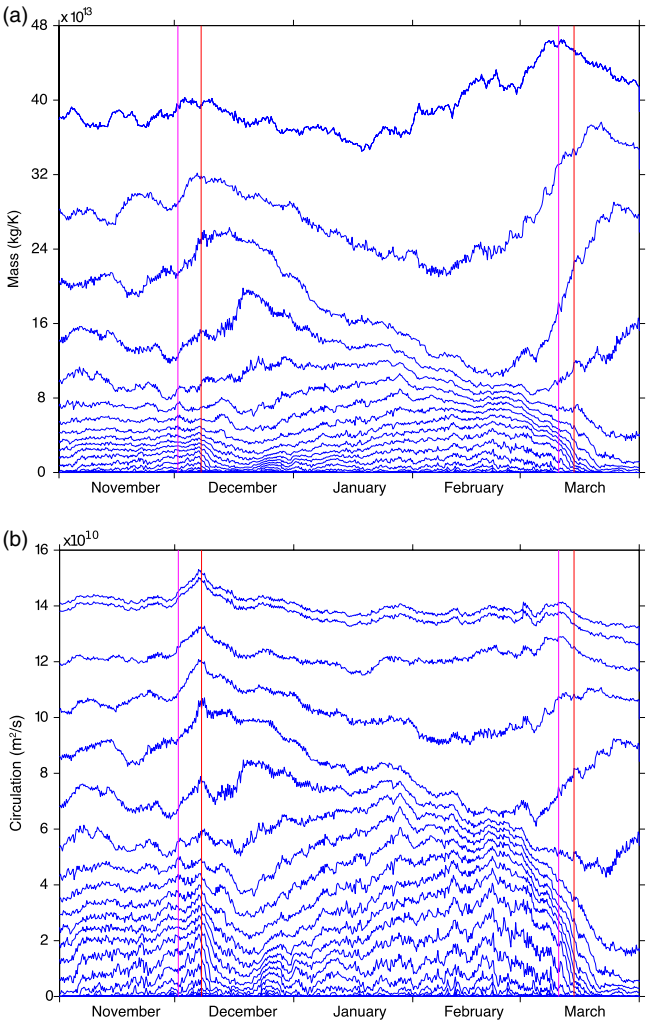

FIGURE 14 Winter 1987/1988: (a) Mass integral $\mathcal{M}_{Q,\theta}$ at $\theta = 850$ K, with 17 PV contours, equally spaced between 132.8 PVU and 1073.6 PVU. (b) Circulation $\mathcal{C}_{Q,\theta}$ at $\theta = 850$ K, with 21 PV contours, equally spaced between 15.2 PVU and 1191.2 PVU [Colour figure can be viewed at wileyonlinelibrary.com]

February, it starts to become eroded, undergoing a short phase of erosion in early March before the breaking in mid-March. The recovery following the early March warming is more like that seen in other winters with a longer period where the vortex remains relatively weak until the end of March.

The two phases of erosion in the March breakdown event are more pronounced in the circulation than they appear in the mass integral. The slowing down in erosion, and even a slight rise in circulation on some of the inner contours, can be seen before a rapid reduction in circulation for all contours making up the vortex. There is a fairly rapid rise in the circulation inside the vortex immediately prior to the fall as the December event occurs. The fall in circulation is much more rapid than the loss in mass for the same event. The initial gain in circulation which marks the start of the recovery is also much faster than the rest of the recovery period from late December to late February/early March. From around mid-December, just after the vortex splits, the contours closest to the vortex start to decrease in circulation bringing them closer to the vortex

and increasing the PV gradient on the edge. These contours continue losing circulation until the vortex starts to displace from the pole, when these outer contours are stripped from the vortex as it breaks down. These contours are only loosely attached to the vortex as they do not follow the same rapid breakdowns as the other inner contours.

The signal of the early winter vortex breakdown appears at roughly the same time on the 950 K level (not shown) as it does on the 850 K level. The 600 K level still shows the delayed signal of the warming seen in other winters (not shown). This means that the breakdown appears on the upper two levels nearly simultaneously but takes a period of a few days to propagate down to the lower surface. This is the case with the later March event as well, with the breakdown appearing much later and occurring at a much slower rate at 600 K than at 850 K and 950 K.

5.3.6 | Winter 1990/1991

So far, all the winters examined have been ones in which SSW events have occurred. These are of main interest here, but examining winters in which no major warming took place is a useful way to differentiate between features which can be seen as typical of an undisturbed polar vortex and those which are specific to either splitting or displacement events. In the winter 1990/1991, no major warming event occurred by the criteria of either Charlton and Polvani (2007) or Mitchell *et al.* (2013).

Figure 15 displays the mass integral, the circulation and the cross-contour mass flux on the vortex edge at $\theta = 850$ K. The mass plot shows a vortex which looks similar to the undisturbed parts of some of the other winters examined at this vertical level, for example the early part of the winter 1978/1979 up to the start of January, before the vortex begins to decrease in mass in the run-up to the late winter splitting event. However, there are notable differences, for example the gradient of PV making up the edge and body of the vortex for the winter 1978/1979 is much steeper than for the winter 1990/1991. For the portion of the winters surrounding major warming events, the picture described by the mass plots is very different however, with much more change in the mass of the vortex seen in the disturbed vortices exhibiting warming events.

While the mass plot shows relatively little variation with no clear events visible, the timeseries of circulation features much more relative change over the course of the winter. The change in the circulation is still less than in winters in which SSWs occurred but, unlike in the mass plot, it is possible to identify several small events over the period of the winter. For example, at the start of December there is a small increase in circulation followed by another about 10 days later. At the end of January and the beginning of February there are two small drops in circulation where the circulation drops back to just below the state at the start of November. This drop is not as large and is less sudden than is seen around the time

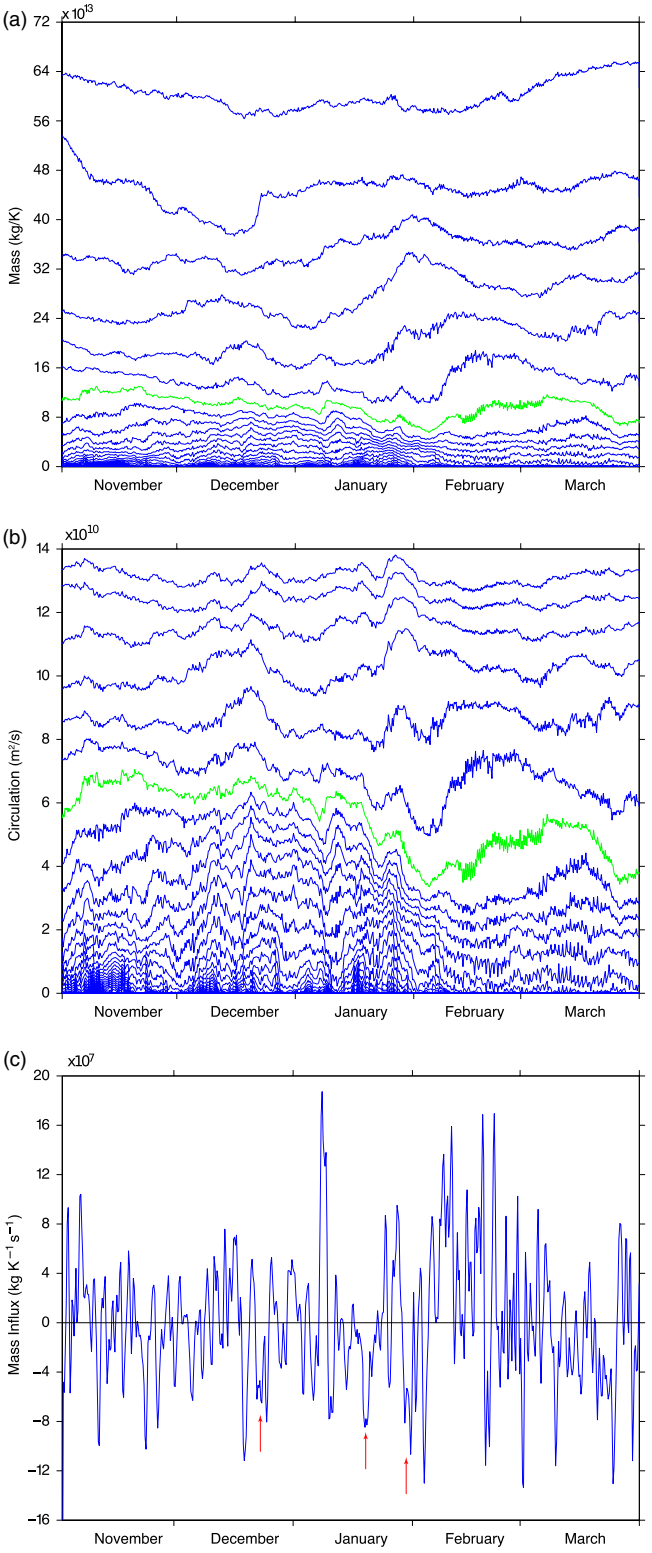

FIGURE 15 Winter 1990/1991: (a) Mass integral $\mathcal{M}_{Q,\theta}$ at $\theta = 850$ K, with 28 PV contours, equally spaced between 15.2 PVU and 1602.8 PVU. (b) Circulation $\mathcal{C}_{Q,\theta}$ at $\theta = 850$ K, with 31 PV contours, equally spaced between 15.2 PVU and 1779.2 PVU. Marked in green in (a) and (b) is the PV contour $Q = 368$ PVU on the vortex edge for which the cross-contour mass flux $\mathcal{V}_{Q,\theta}$ is shown in (c), where red arrows mark significant mass-stripping events [Colour figure can be viewed at wileyonlinelibrary.com]

of major warmings and is more of a gradual reduction in the circulation of the vortex. Although there are more features visible in the circulation plot than in the mass plot, the changes in circulation for the winter 1990/1991 occur at a much slower rate than in other more eventful winters.

The cross-contour mass flux for a PV contour on the vortex edge shows that there is much less movement of air across the contour for this uneventful winter than there is in the disturbed winters which result in SSWs. There are also vortex-stripping events visible, but they are smaller and less persistent and there is more of a balance with events adding mass to the vortex. There is in the region of three times as much mass moved across the vortex edge over the whole winter 1978/1979 than in the winter 1990/1991. There is also around twice as much mass transport across the contour in the winter 1984/1985 than in the winter 1990/1991.

6 | CONCLUSIONS

PV contour integral quantities of mass and circulation and their balance equations have been introduced as a tool to monitor and diagnose the state and the dynamics of the stratospheric polar vortex in reanalysis data.

Compared to more traditional diagnostics such as the vortex area and the equivalent latitude, the signal of the SSWs is much clearer and stronger in the new mass and circulation diagnostics. In particular, mass and circulation offer a more detailed and differentiated picture of the dynamical changes in the time period prior to the SSW. They are more informative, in that the relationships between mass/circulation and PV are very distinct between the polar vortex and the surf zone in the run-up to and during SSWs. The PV gradients in these coordinates change abruptly across the vortex edge, so this criterion (sudden change in PV gradients), rather than the maximum gradient in the equivalent latitude, may be a better definition of the vortex edge (like the definition of tropopause).

Mass and circulation plots usually display very similar features as they are not independent (cf. Equation B9). Nevertheless it is often advantageous to examine both side by side. The mass integral tends to show a clearer view of the vortex edge and the formation of the surf zone, whereas the circulation offers more resolution in the area inside of the vortex. Circulation plots tend to be noisier than mass plots as they involve a higher level of spatial differentiation.

Timeseries of the diagnostics have been calculated for several winters, both those in which major warming events occurred and those where no such events were present. Following most of these events, the mass of air within the vortex dramatically decreased and the vortex became fairly undefined in the mass plots. The events where this decrease is less visible, or not at all, are those events where the differing criteria of Charlton and Polvani (2007) and Mitchell *et al.* (2013) disagree on the occurrence of an SSW event. Also the early winter events show less dramatic decreases in mass than those later in the winter. The profiles of mass at the start of the winter, in early November, can vary significantly from year to year. The winters where warmings occurred early do have similar profiles of mass in early November. There does not seem to be any similarity between those that occurred later in the winter, or between those in which splittings or displacements occurred. The behaviour of the vortex around the time of the SSWs also differs significantly from year to year. This is similar to the findings of Matthewman *et al.* (2009), who noted that prior to the onset of SSWs there was a large amount of variability in the strength of the vortex, even when events were grouped into splitting and displacement events.

The mass within the vortex does not seem to attain any fixed value or threshold prior to the breakdown of the vortex, either for splitting or displacement events. There is also not a clear threshold for the amount of mass removed from the vortex via stripping of filaments before an SSW will occur. These findings are in line with a recent study of vortex erosion in a shallow-water model (Beaumont *et al.*, 2017). The circulation around the vortex edge does not seem to feature a threshold value prior to an SSW event either, although the circulation at the vortex edge for the events identified as displacements on all isentropic levels is in general higher (around 4 to $5 \times 10^{10}\mathrm{m}^2/\mathrm{s}$) than that immediately prior to those identified as splitting events (3 to $4 \times 10^{10}\mathrm{m}^2/\mathrm{s}$). This is in contrast to the finding by Matthewman *et al.* (2009) that immediately prior to splitting events the vortex appears stronger than it is in the lead-up to displacement events. However, they did find that there was a large amount of variability in the strength of the vortex prior to SSWs, which is in agreement with the results seen here.

In plots of the cross-contour mass flux there are several events throughout the course of the winter where mass is removed from the vortex. These events can be linked to PV maps where tongues of air can be seen to be pulled from the vortex and mixed into the surf zone. The amount of mass pulled from the vortex in one of these tongues can be a significant proportion of the mass within the vortex, typically of the order of 10% of the mass within the vortex. Using the classifications of Charlton and Polvani (2007), in the 10 days leading up to the dates of the warmings, stripping events lasting for periods of roughly 5 days occur in all the winters for which the diagnostics have been calculated. The stripping event seen close to the SSW is one of the largest events in each winter, with up to 25% of the mass of the vortex stripped for the largest of these events. These are not the only significant stripping events to occur in the period leading up to the SSWs. In those winters where sudden warmings occurred late in the winter, in late February and March, there are other large stripping events through the course of the winter lasting for several days of similar magnitudes to those immediately before the breakdown of the vortex. In contrast, for winters where the warming events occur early, in December and early January, there are several smaller stripping events lasting around 2 days

in length instead of the few larger events seen in the winters where warmings occur later.

It seems that large vortex-stripping events are a necessary, but not sufficient, condition for an SSW to occur; prior to all SSWs there is a large stripping event, the largest over the course of the winter, but there are several other stripping events of only slightly smaller magnitude which are not followed by an SSW. There are also large stripping events in winters in which no major warming occurs. These are similar in magnitude to the stripping events which occur throughout winters in which SSWs occur late in the winter.

There is not a significant difference in the pattern of stripping events in the lead-up to vortex splitting and vortex displacement events as defined by Charlton and Polvani (2007). This is similar to the findings of Miller *et al.* (2013) who found that the biggest differences between the different types of warming occurred following the date of the sudden warming, with only small differences visible before. The classifications of Mitchell *et al.* (2013) also show a similar pattern of stripping in the lead-up to SSWs with a much larger difference between early and late warming events than between splitting and displacement events. The events where the two criteria disagree as to whether a sudden warming occurs are those in which the signal of the warming is less visible in the stripping of filaments, though these can still be distinguished from the general background pattern. The reason for the lack of systematic difference in the stripping preceding splitting and displacement events seems to be the large amount of inter-event variability seen here and noted by Matthewman *et al.* (2009), who also noted that there were some splitting events which showed behaviour more characteristic of displacement events than splitting events.

An additional line of analysis may lie in considering a time-integrated version of the mass balance of Equation 14, that is, in examining cumulative mass fluxes instead of the (noisy) instantaneous mass fluxes, and comparing those with net changes in mass from a chosen initial condition (e.g. the start of the winter). That way one can quantify what fraction of mass change is attributable to cross-contour mass flux $\mathcal{V}_{Q,\theta}$ and what fraction to cross-isentropic mass flux CrIs.

An attempt was also made to diagnose from the balance equation of the circulation the impact of frictional forces such as gravity wave dissipation on the vortex. This is generally hampered by the high noise level in these timeseries due to the high level of differentiation involved. There are horizontal derivatives of the vertical shear of the horizontal velocities to be estimated numerically as well as a time derivative of the already noisy circulation timeseries. Only very prominent events can be expected to be detectable in this diagnostic. We remark that this diagnostic may be more useful when analyzing data from model simulations as these generally are less noisy. Again, the noisiness of the instantaneous flux term may be circumvented if the cumulative value is compared with the net change in circulation, that is, if a time-integrated form of the circulation balance of Equation 28 is considered. Also the other terms on the right-hand side of Equation 28 could be examined.

It might be an interesting question for further research whether the integral diagnostics of mass and circulation could be used to define dynamically based objective criteria for an SSW to have occurred and its classification as a splitting or displacement event.

ACKNOWLEDGEMENTS

We would like to acknowledge the British Atmospheric Data Centre (BADC) for providing the ERA-40 reanalysis dataset produced by the European Centre for Medium-Range Weather Forecasts (ECMWF). Robin Beaumont was supported during this research with a PhD studentship funded by an EPSRC Doctoral Training Grant.

ORCID

Frank Kwasniok https://orcid.org/0000-0003-1421-4010
John Thuburn https://orcid.org/0000-0002-4598-546X

REFERENCES

Andrews, D.G., Holton, J.R. and Leovy, C.B. (1987) *Middle Atmosphere Dynamics*. Cambridge, MA: Academic Press.

Baldwin, M.P. and Dunkerton, T.J. (2001) Stratospheric harbingers of anomalous weather regimes. *Science*, 294, 581–584.

Beaumont, R., Kwasniok, F. and Thuburn, J. (2017) Vortex erosion in a shallow-water model of the polar vortex. *Dynamics of Atmospheres and Oceans*, 78, 137–151.

Butchart, N. and Remsberg, E.E. (1986) The area of the stratospheric polar vortex as a diagnostic for tracer transport on an isentropic surface. *Journal of the Atmospheric Sciences*, 43, 1319–1339.

Bühler, O. and Haynes, P.H. (1999) Constraints on the mean mass transport across potential vorticity contours. *Journal of the Atmospheric Sciences*, 56, 942–947.

Charlton, A.J. and Polvani, L.M. (2007) A new look at stratospheric sudden warmings. Part I: climatology and modeling benchmarks. *Journal of Climate*, 20, 449–469.

Dritschel, D.G. and McIntyre, M.E. (2008) Multiple jets as PV staircases: the Phillips effect and the resilience of eddy-transport barriers. *Journal of the Atmospheric Sciences*, 65, 855–874.

Esler, J.G. and Matthewman, N.J. (2011) Stratospheric sudden warmings as self-tuning resonances. Part II: Vortex displacement events. *Journal of the Atmospheric Sciences*, 68, 2505–2523.

Haynes, P.H. and McIntyre, M.E. (1987) On the evolution of vorticity and potential vorticity in the presence of diabatic heating and frictional or other forces. *Journal of the Atmospheric Sciences*, 44, 828–841.

Hoskins, B.J., McIntyre, M.E. and Robertson, A.W. (1985) On the use and significance of isentropic potential vorticity maps. *Quarterly Journal of the Royal Meteorological Society*, 111, 877–946.

Labitzke, K. (1981) The amplification of height wave 1 in January 1979: A characteristic precondition for the major warming in February. *Monthly Weather Review*, 109, 983–989.

Lait, L.R. (1994) An alternative for potential vorticity. *Journal of the Atmospheric Sciences*, 51, 1754–1759.

Legras, B., Dritschel, D.G. and Caillol, P. (2001) The erosion of a distributed two-dimensional vortex in a background straining flow. *Journal of Fluid Mechanics*, 441, 369–398.

Limpasuvan, V., Thompson, D.W.J. and Hartman, D.L. (2004) The life cycle of the Northern Hemisphere stratospheric sudden warmings. *Journal of Climate*, 17, 2584–2596.
Mariotti, A., Legras, B. and Dritschel, D.G. (1994) Vortex stripping and the erosion of coherent structures in two-dimensional flows. *Physics of Fluids*, 6, 3954–3962.
Matthewman, N.J., Esler, J.G., Charlton-Perez, A.J. and Polvani, L.M. (2009) A new look at stratospheric sudden warmings. Part III: polar vortex evolution and vertical structure. *Journal of Climate*, 22, 1566–1585.
Matthewman, N.J. and Esler, J.G. (2011) Stratospheric sudden warmings as self-tuning resonances. Part I: Vortex splitting events. *Journal of the Atmospheric Sciences*, 68, 2481–2504.
McIntyre, M.E. (2015) Potential vorticity. In: North, G., Pyle, J. and Zhang, F. (Eds.). *Encyclopedia of Atmospheric Sciences,* Vol. 2, 2nd edition. Cambridge, MA: Academic Press, pp. 375–383.
McIntyre, M.E. and Palmer, T.N. (1983) Breaking planetary waves in the stratosphere. *Nature*, 305, 593–600.
McIntyre, M.E. and Palmer, T.N. (1984) The 'surf zone' in the stratosphere. *Journal of Atmospheric and Terrestial Physics*, 46, 825–849.
Methven, J. and Berrisford, P. (2015) The slowly evolving background state of the atmosphere. *Quarterly Journal of the Royal Meteorological Society*, 141, 2237–2258.
Miller, A., Schmidt, H. and Bunzel, F. (2013) Vertical coupling of the middle atmosphere during stratospheric warming events. *Journal of Atmospheric and Solar-Terrestrial Physics*, 97, 11–21.
Mitchell, D.M., Gray, L.J., Anstey, J., Baldwin, M.P. and Charlton-Perez, A.J. (2013) The influence of stratospheric vortex displacements and splits on surface climate. *Journal of Climate*, 26, 2668–2682.
Mo, R., Bühler, O. and McIntyre, M.E. (1998) Permeability of the stratospheric vortex edge: on the mean mass flux due to thermally dissipating, steady, non-breaking Rossby waves. *Quarterly Journal of the Royal Meteorological Society*, 124, 2129–2148.
Müller, R. and Günther, G. (2003) A generalized form of Lait's modified potential vorticity. *Journal of the Atmospheric Sciences*, 60, 2229–2237.
Nakamura, N. (1995) Modified Lagrangian-mean diagnostics of the stratospheric polar vortices. Part I: formulation and analysis of GFDL SKYHI GCM. *Journal of the Atmospheric Sciences*, 52, 2096–2108.
Nakamura, N. and Solomon, A. (2011) Finite-amplitude wave activity and mean flow adjustments in the atmospheric general circulation. Part II: analysis in isentropic coordinates. *Journal of the Atmospheric Sciences*, 68, 2783–2799.
Norton, W.A. (1994) Breaking Rossby waves in a model stratosphere diagnosed by a vortex-following coordinate system and a technique for advecting material contours. *Journal of the Atmospheric Sciences*, 51, 654–675.
Polvani, L.M. and Saravanan, R. (2000) The three-dimensional structure of breaking Rossby waves in the polar wintertime stratosphere. *Journal of the Atmospheric Sciences*, 57, 3663–3685.
Shine, K.P. and Rickaby, J.A. (1989) Solar radiative heating due to absorption by ozone. In: Bojkov, R.D. and Fabian, P. (Eds.) *Ozone in the Atmosphere* (Proceedings of the Quadrennial Ozone Symposium 1988 and Tropospheric Ozone Workshop. Hampton, VA: A. Deepak Publishing, pp. 597–600.
Sobel, A.H. and Plumb, R.A. (1999) Quantitative diagnostics of mixing in a shallow-water model of the stratosphere. *Journal of the Atmospheric Sciences*, 56, 2811–2829.
Thompson, D.W.J., Baldwin, M.P. and Wallace, J.M. (2002) Stratospheric connection to Northern Hemisphere wintertime weather: implications for prediction. *Journal of Climate*, 15, 1421–1428.
Thuburn, J. and Lagneau, V. (1999) Eulerian mean, contour integral, and finite-amplitude wave activity diagnostics applied to a single-layer model of the winter stratosphere. *Journal of the Atmospheric Sciences*, 56, 689–710.
Uppala, S.M., Kållberg, P.W., Simmons, A.J., Andrae, U., Bechtold, V., Fiorino, M., Gibson, J.K., Haseler, J., Hernandez, A., Kelly, G.A., Li X, Onogi K, Saarinen S, Sokka N, Allan RP, Andersson E, Arpe K, Balmaseda MA, Beljaars ACM, Van de Berg L, Bidlot J, Bormann N, Caires S, Chevallier F, Dethof A, Dragosazac M, Fisher M, Fuentes M, Hagemann S, Hólm EV, Hoskins BJ, Isaksen L, Janssen PAEM, Jenne R, McNally AP, Mahfouf J-F, Morcrette J-J, Rayner NA, Saunders RW, Simon P, Sterl A, Trenberth KA, Untch A, Vasiljevic D, Viterbo P and Woollen J (2005) The ERA-40 re-analysis. *Quarterly Journal of the Royal Meteorological Society*, 131, 2961–3012.
Waugh, D.W., Plumb RA, Atkinson RJ, Schoeberl MR, Lait LR, Newman P, Loewenstein M, Toohey D, Avallone L, Webster CR and May R (1994) Transport out of the lower stratospheric Arctic vortex by Rossby wave breaking. *Journal of Geophysical Research: Atmospheres*, 99, 1071–1088.

How to cite this article: Kwasniok F, Beaumont R, Thuburn J. Vortex dynamics of stratospheric sudden warmings: A reanalysis data study using PV contour integral diagnostics. *Q J R Meteorol Soc.* 2019;145:1013–1033. https://doi.org/10.1002/qj.3474

APPENDICES

A. USEFUL IDENTITIES

For any scalar function $h = h(\lambda, \phi, t)$ we have

$$\int_{S_{Q,\theta}} h \,\mathrm{d}S = \int_{\widetilde{Q}\geq Q} \oint_{\Gamma_{\widetilde{Q},\theta}} \frac{h}{|\nabla_\theta \widetilde{Q}|} \,\mathrm{d}s \,\mathrm{d}\widetilde{Q}, \tag{A1}$$

as $\mathrm{d}S = \mathrm{d}s\,\mathrm{d}l$ with $\mathrm{d}s$ and $\mathrm{d}l$ being scalar line elements tangent and normal, respectively, to the PV contour $\Gamma_{\widetilde{Q},\theta}$. We have $\mathrm{d}\widetilde{Q} = (\nabla_\theta \widetilde{Q})\cdot\mathrm{d}\mathbf{r}$ and thus $\mathrm{d}l = \mathrm{d}\widetilde{Q}/|\nabla_\theta \widetilde{Q}|$. Here and in the following $\widetilde{Q}$ is a variable of integration as opposed to Q which denotes the PV value of the contour under consideration. The identity

$$\frac{\partial}{\partial Q}\int_{S_{Q,\theta}} h \,\mathrm{d}S = -\oint_{\Gamma_{Q,\theta}} \frac{h}{|\nabla_\theta Q|} \,\mathrm{d}s \tag{A2}$$

immediately follows.

The temporal evolution of any surface integral is given by

$$\begin{aligned}\frac{\partial}{\partial t}\int_{S_{Q,\theta}} h \,\mathrm{d}S &= \int_{S_{Q,\theta}} \frac{\partial h}{\partial t} \,\mathrm{d}S + \oint_{\Gamma_{Q,\theta}} \frac{h\,\partial Q/\partial t}{|\nabla_\theta Q|} \,\mathrm{d}s \\ &\overset{(A2)}{=} \int_{S_{Q,\theta}} \frac{\partial h}{\partial t} \,\mathrm{d}S - \frac{\partial}{\partial Q}\int_{S_{Q,\theta}} h \frac{\partial \widetilde{Q}}{\partial t} \,\mathrm{d}S.\end{aligned} \tag{A3}$$

Analogously, the vertical derivative is given by

$$\begin{aligned}\frac{\partial}{\partial \theta}\int_{S_{Q,\theta}} h \,\mathrm{d}S &= \int_{S_{Q,\theta}} \frac{\partial h}{\partial \theta} \,\mathrm{d}S + \oint_{\Gamma_{Q,\theta}} \frac{h\,\partial Q/\partial \theta}{|\nabla_\theta Q|} \,\mathrm{d}s \\ &\overset{(A2)}{=} \int_{S_{Q,\theta}} \frac{\partial h}{\partial \theta} \,\mathrm{d}S - \frac{\partial}{\partial Q}\int_{S_{Q,\theta}} h \frac{\partial \widetilde{Q}}{\partial \theta} \,\mathrm{d}S.\end{aligned} \tag{A4}$$

We further have

$$\begin{aligned}\frac{\partial}{\partial Q}\int_{S_{Q,\theta}} hG(\widetilde{Q}) \,\mathrm{d}S &\overset{(A2)}{=} -\oint_{\Gamma_{Q,\theta}} \frac{hG(Q)}{|\nabla_\theta Q|} \,\mathrm{d}s \\ &= -G(Q)\oint_{\Gamma_{Q,\theta}} \frac{h}{|\nabla_\theta Q|} \,\mathrm{d}s \\ &\overset{(A2)}{=} G(Q)\frac{\partial}{\partial Q}\int_{S_{Q,\theta}} h \,\mathrm{d}S\end{aligned} \tag{A5}$$

for any function $G(Q)$ and in particular for $G(Q) = Q$.

Standard vector calculus yields the identity

$$\sigma\mathbf{v}\cdot\nabla_\theta Q = \nabla_\theta\cdot(\sigma Q\mathbf{v}) - Q\nabla_\theta\cdot(\sigma\mathbf{v}). \tag{A6}$$

With the two-dimensional divergence theorem, we get

$$\begin{aligned}\int_{S_{Q,\theta}}\nabla_\theta\cdot[\sigma G(\widetilde{Q})\mathbf{v}]\,\mathrm{d}S &= -\oint_{\Gamma_{Q,\theta}}\sigma G(Q)\mathbf{v}\cdot\hat{\mathbf{n}}\,\mathrm{d}s\\ &= -G(Q)\oint_{\Gamma_{Q,\theta}}\sigma\mathbf{v}\cdot\hat{\mathbf{n}}\,\mathrm{d}s\\ &= G(Q)\int_{S_{Q,\theta}}\nabla_\theta\cdot(\sigma\mathbf{v})\,\mathrm{d}S\end{aligned} \tag{A7}$$

for any function $G(Q)$ and in particular for $G(Q) = Q$.

We can now derive the theorem

$$\begin{aligned}&\frac{\partial}{\partial Q}\int_{S_{Q,\theta}}\sigma\mathbf{v}\cdot\nabla_\theta\widetilde{Q}\,\mathrm{d}S\\ &\overset{(A6)}{=}\frac{\partial}{\partial Q}\int_{S_{Q,\theta}}\left[\nabla_\theta\cdot(\sigma\widetilde{Q}\mathbf{v}) - \widetilde{Q}\nabla_\theta\cdot(\sigma\mathbf{v})\right]\mathrm{d}S\\ &\overset{(A7),(A5)}{=}\frac{\partial}{\partial Q}\left[Q\int_{S_{Q,\theta}}\nabla_\theta\cdot(\sigma\mathbf{v})\,\mathrm{d}S\right] - Q\frac{\partial}{\partial Q}\int_{S_{Q,\theta}}\nabla_\theta\cdot(\sigma\mathbf{v})\,\mathrm{d}S\\ &= \int_{S_{Q,\theta}}\nabla_\theta\cdot(\sigma\mathbf{v})\,\mathrm{d}S.\end{aligned} \tag{A8}$$

B. LAGRANGIAN INTEGRAL QUANTITIES ON PV CONTOURS AND THEIR BALANCE EQUATIONS

We consider the quantity $\sigma G(Q)$ for an arbitrary function $G(Q)$. It obeys the evolution equation

$$\frac{\partial[\sigma G(Q)]}{\partial t} + \nabla_\theta\cdot[\sigma G(Q)\mathbf{v}] + \frac{\partial[\sigma G(Q)\dot\theta]}{\partial\theta} = \sigma G'(Q)\dot{Q}, \tag{B1}$$

as can be seen by combining the continuity Equation 6 and the PV advection Equation 7 using identity Equation A6. We introduce the integral quantity

$$\mathcal{I}_{Q,\theta} = \int_{S_{Q,\theta}}\sigma G(\widetilde{Q})\,\mathrm{d}S. \tag{B2}$$

The time evolution of $\mathcal{I}_{Q,\theta}$ is given as

$$\begin{aligned}\frac{\partial\mathcal{I}_{Q,\theta}}{\partial t} &= \frac{\partial}{\partial t}\int_{S_{Q,\theta}}\sigma G(\widetilde{Q})\,\mathrm{d}S\\ &\overset{(A3)}{=}\int_{S_{Q,\theta}}\frac{\partial[\sigma G(\widetilde{Q})]}{\partial t}\,\mathrm{d}S - \frac{\partial}{\partial Q}\int_{S_{Q,\theta}}\sigma G(\widetilde{Q})\frac{\partial\widetilde{Q}}{\partial t}\,\mathrm{d}S\\ &\overset{(B1),(A5),(7)}{=}\int_{S_{Q,\theta}}\left\{-\nabla_\theta\cdot[\sigma G(\widetilde{Q})\mathbf{v}] - \frac{\partial[\sigma G(\widetilde{Q})\dot\theta]}{\partial\theta} + \sigma G'(\widetilde{Q})\dot{Q}\right\}\mathrm{d}S\\ &\quad - G(Q)\frac{\partial}{\partial Q}\int_{S_{Q,\theta}}\sigma\left[\dot{Q} - \mathbf{v}\cdot\nabla_\theta\widetilde{Q} - \dot\theta\frac{\partial\widetilde{Q}}{\partial\theta}\right]\mathrm{d}S\\ &\overset{(A7),(A8)}{=} G(Q)\underbrace{\left[\frac{\partial}{\partial Q}\int_{S_{Q,\theta}}\sigma\dot\theta\frac{\partial\widetilde{Q}}{\partial\theta}\,\mathrm{d}S - \frac{\partial}{\partial Q}\int_{S_{Q,\theta}}\sigma\dot{Q}\,\mathrm{d}S\right]}_{\mathcal{V}_{Q,\theta}}\\ &\quad - \int_{S_{Q,\theta}}\frac{\partial[\sigma G(\widetilde{Q})\dot\theta]}{\partial\theta}\,\mathrm{d}S + \int_{S_{Q,\theta}}\sigma G'(\widetilde{Q})\dot{Q}\,\mathrm{d}S.\end{aligned} \tag{B3}$$

The quantities $\mathcal{I}_{Q,\theta}$ are Lagrangian invariants of adiabatic and frictionless flow. For $G(Q) = 1$, $\mathcal{I}_{Q,\theta}$ is the mass integral $\mathcal{M}_{Q,\theta}$ and Equation B1 becomes the continuity Equation 6; for $G(Q) = Q$, $\mathcal{I}_{Q,\theta}$ is the circulation $\mathcal{C}_{Q,\theta}$ and Equation B1 becomes the PV evolution Equation 8. Other choices for $G(Q)$ are not discussed here as they have no immediate physical relevance. The term $G(Q)\mathcal{V}_{Q,\theta}$ is the cross-contour flux with $\mathcal{V}_{Q,\theta}$ being the cross-contour mass flux. The cross-contour fluxes depend only on fields on the contour PV $= Q$. The last two terms in Equation B3 (generally) are cross-isentropic fluxes. Further explanation is given in Section 4 for mass and circulation.

For completeness, an alternative way is mentioned for calculating the cross-contour flux (cf. Bühler and Haynes, 1999, Thuburn and Lagneau, 1999). The net influx associated with a scalar quantity $h = h(\lambda,\phi,t)$ through the PV contour $\Gamma_{Q,\theta}$ is given by

$$\oint_{\Gamma_{Q,\theta}} h(\mathbf{v}-\mathbf{v}_\mathrm{c})\cdot\hat{\mathbf{n}}\,\mathrm{d}s, \tag{B4}$$

where $\mathbf{v}_\mathrm{c}$ is the horizontal velocity of the PV contour $\Gamma_{Q,\theta}$. The component of $\mathbf{v}_\mathrm{c}$ normal to the contour is

$$\begin{aligned}\mathbf{v}_\mathrm{c}\cdot\hat{\mathbf{n}} &= -\frac{\partial Q/\partial t}{|\nabla_\theta Q|}\\ &\overset{(7)}{=}\frac{1}{|\nabla_\theta Q|}\left(\mathbf{v}\cdot\nabla_\theta Q + \dot\theta\frac{\partial Q}{\partial\theta} - \dot{Q}\right)\end{aligned} \tag{B5}$$

and we have

$$(\mathbf{v}-\mathbf{v}_\mathrm{c})\cdot\hat{\mathbf{n}} = \frac{1}{|\nabla_\theta Q|}\left(\dot{Q} - \dot\theta\frac{\partial Q}{\partial\theta}\right), \tag{B6}$$

in particular $(\mathbf{v}-\mathbf{v}_\mathrm{c})\cdot\hat{\mathbf{n}} = 0$ for adiabatic and frictionless flow. The component of $\mathbf{v}_\mathrm{c}$ tangent to the contour is undefined; it can be set to zero or, in the adiabatic and frictionless case, to the tangent component of $\mathbf{v}$, following material fluid elements. The cross-contour flux associated with $\sigma G(Q)$ is now

$$\begin{aligned}&\oint_{\Gamma_{Q,\theta}}\sigma G(Q)(\mathbf{v}-\mathbf{v}_\mathrm{c})\cdot\hat{\mathbf{n}}\,\mathrm{d}s\\ &\overset{(B6)}{=} G(Q)\oint_{\Gamma_{Q,\theta}}\frac{1}{|\nabla_\theta Q|}\left(\sigma\dot{Q} - \sigma\dot\theta\frac{\partial Q}{\partial\theta}\right)\mathrm{d}s\\ &\overset{(A2)}{=} G(Q)\mathcal{V}_{Q,\theta},\end{aligned} \tag{B7}$$

in accordance with the result above.

The integral quantities $\mathcal{I}_{Q,\theta}$ for different choices of $G(Q)$ are not independent. In PV space there is the relationship

$$\frac{\partial \mathcal{I}_{Q,\theta}}{\partial Q} \overset{(A5)}{=} G(Q)\frac{\partial \mathcal{M}_{Q,\theta}}{\partial Q} \tag{B8}$$

and in particular

$$\frac{\partial \mathcal{C}_{Q,\theta}}{\partial Q} = Q\frac{\partial \mathcal{M}_{Q,\theta}}{\partial Q}. \tag{B9}$$

This holds independent of any fluid-dynamical approximations or model and thus generalizes the result of Thuburn and Lagneau (1999) for a shallow-water model which was also used by Beaumont *et al.* (2017).

C. BALANCE EQUATION OF THE VORTEX AREA

The time evolution of the vortex area is given by

$$\begin{aligned}
\frac{\partial A_{Q,\theta}}{\partial t} &= \frac{\partial}{\partial t}\int_{S_{Q,\theta}} \mathrm{d}S \\
&\overset{(A3)}{=} -\frac{\partial}{\partial Q}\int_{S_{Q,\theta}} \frac{\partial \widetilde{Q}}{\partial t}\,\mathrm{d}S \\
&\overset{(7),(A2)}{=} \oint_{\Gamma_{Q,\theta}} \frac{1}{|\nabla_\theta Q|}\left(\dot{Q} - \mathbf{v}\cdot\nabla_\theta Q - \dot{\theta}\frac{\partial Q}{\partial \theta}\right)\mathrm{d}s \\
&= \int_{S_{Q,\theta}} \nabla_\theta \cdot \mathbf{v}\,\mathrm{d}S + \oint_{\Gamma_{Q,\theta}} \frac{1}{|\nabla_\theta Q|}\left(\dot{Q} - \dot{\theta}\frac{\partial Q}{\partial \theta}\right)\mathrm{d}s \\
&\overset{(A2)}{=} \int_{S_{Q,\theta}} \nabla_\theta \cdot \mathbf{v}\,\mathrm{d}S + \frac{\partial}{\partial Q}\int_{S_{Q,\theta}} \dot{\theta}\frac{\partial \widetilde{Q}}{\partial \theta}\,\mathrm{d}S - \frac{\partial}{\partial Q}\int_{S_{Q,\theta}} \dot{Q}\,\mathrm{d}S,
\end{aligned} \tag{C1}$$

where the last but one step uses the two-dimensional divergence theorem. Alternatively, we have

$$\begin{aligned}
\frac{\partial A_{Q,\theta}}{\partial t} &= -\oint_{\Gamma_{Q,\theta}} \mathbf{v}_\mathrm{c}\cdot\hat{\mathbf{n}}\,\mathrm{d}s \\
&\overset{(B5)}{=} \oint_{\Gamma_{Q,\theta}} \frac{1}{|\nabla_\theta Q|}\left(\dot{Q} - \mathbf{v}\cdot\nabla_\theta Q - \dot{\theta}\frac{\partial Q}{\partial \theta}\right)\mathrm{d}s,
\end{aligned} \tag{C2}$$

with the same result.

D. NUMERICAL EVALUATION OF SURFACE INTEGRALS

The various surface integrals occurring with the diagnostics of mass and circulation are approximated numerically as weighted sums over grid point values. The data are on a regular Gaussian grid. The grid points are (λ_i, ϕ_j); $i = 1, \ldots, N; j = 1, \ldots, M$, where the longitudes are given by $\lambda_i = 2\pi(i-1)/N$ and the latitudes are given by the Gaussian latitudes on the Northern Hemisphere with $\phi_j < \phi_{j+1}$. Each grid point represents a grid box extending half-way to the neighbouring grid points in longitude and latitude. The northernmost grid boxes extend to the North Pole and the southernmost gridboxes extend to the Equator. The weight of each grid point is given by the area of the corresponding grid box. For any function $h = h(\lambda, \phi)$, the surface integral over $S_{Q,\theta}$ is evaluated as

$$\int_{S_{Q,\theta}} h(\lambda,\phi)\,\mathrm{d}S = 2\pi a^2 \sum_{\substack{i,j \\ \widetilde{Q}(\lambda_i,\phi_j)\geq Q}} w_{ij} h(\lambda_i, \phi_j), \tag{D1}$$

where a is the radius of the Earth and the weights are

$$w_{i1} = \frac{1}{N}\sin\frac{\phi_1 + \phi_2}{2}, \tag{D2}$$

$$w_{ij} = \frac{1}{N}\left(\sin\frac{\phi_j + \phi_{j+1}}{2} - \sin\frac{\phi_{j-1} + \phi_j}{2}\right) \quad \text{with } j = 2, \ldots, M-1, \tag{D3}$$

$$w_{iM} = \frac{1}{N}\left(1 - \sin\frac{\phi_{M-1} + \phi_M}{2}\right). \tag{D4}$$

We have $\sum_{i,j} w_{ij} = 1$. A very similar integration scheme could be used in the case where the data are stored on a reduced Gaussian grid.