



Vortex splitting on a planetary scale in the stratosphere by cyclogenesis on a subplanetary scale in the troposphere

Article

Accepted Version

O'Neill, A., Oatley, C. L., Charlton-Perez, A. J., Mitchell, D. M. and Jung, T. (2017) Vortex splitting on a planetary scale in the stratosphere by cyclogenesis on a subplanetary scale in the troposphere. *Quarterly Journal of the Royal Meteorological Society*, 143 (703). pp. 691-705. ISSN 1477-870X doi: <https://doi.org/10.1002/qj.2957> Available at <http://centaur.reading.ac.uk/69951/>

It is advisable to refer to the publisher's version if you intend to cite from the work.

Published version at: <http://dx.doi.org/10.1002/qj.2957>

To link to this article DOI: <http://dx.doi.org/10.1002/qj.2957>

Publisher: Royal Meteorological Society

All outputs in CentAUR are protected by Intellectual Property Rights law, including copyright law. Copyright and IPR is retained by the creators or other copyright holders. Terms and conditions for use of this material are defined in the [End User Agreement](#).

www.reading.ac.uk/centaur

CentAUR

Central Archive at the University of Reading

Reading's research outputs online



Vortex splitting on a planetary scale in the stratosphere by cyclogenesis on a sub planetary scale in the troposphere †

A. O'Neill^{1*}, C. L. Oatley¹, A. J. Charlton-Perez¹, D. Mitchell², T. Jung³

¹*Department of Meteorology, University of Reading, Reading, RG6 6BB, UK*

²*Environmental Change Institute, Oxford University Centre for the Environment, Oxford, OX1 3QY, UK*

³*Alfred Wegener Institute, Bussenstrasse 24, D-27570 Bremerhaven, Germany, and Department of Physics and Electrical Engineering, University of Bremen, D-28334 Bremen, Germany*

*Correspondence to: Department of Meteorology, University of Reading, Reading, RG6 6BB

It is hypothesized that a splitting of the stratospheric polar vortex and a sudden warming can result when the polar vortex is elongated and a closed cyclonic circulation develops on a sub planetary scale in the troposphere beneath one of its tips. The hypothesis is supported by studying the splitting event in the southern hemisphere during spring 2002. Potential vorticity inversion and an inverse modelling technique using the adjoint of a fully nonlinear dynamical model are used to confirm that splitting is sensitive to sub planetary-scale cyclogenesis when it is strong. Examples of stratospheric vortex splitting events in the northern hemisphere are consistent with the hypothesis. The proposed mechanism for splitting contrasts with the commonly accepted one that it is caused by the upward propagation of a planetary wave from the troposphere. It is suggested that the phenomenon is better understood as an example of a vortex interaction rather than as a wave, mean-flow interaction.

Key Words: polar vortex splitting; stratospheric sudden warmings; tropospheric cyclogenesis; PV inversion; sensitivity analysis; vortex interactions; finite-amplitude instability.

Received ...

1. Introduction

The central idea in this paper is that splitting of the polar vortex in the stratosphere during a major stratospheric sudden warming (SSW) can be caused by the development of a single, closed cyclonic circulation in the troposphere on a sub planetary scale. For the splitting to happen by this mechanism (a) the polar vortex needs to be elongated initially, and (b) the cyclogenesis must occur approximately under a tip of the elongated vortex, such that the curvatures of these systems match approximately in the vertical. The cyclogenesis occurs when a high-amplitude (non linear), baroclinically unstable Rossby wave with zonal wave number 4 or thereabouts - of the kind seen in winter in the mid troposphere - “breaks” to form locally a closed cyclonic circulation and associated anomaly in potential vorticity. Such a structure can persist for several days or longer, which should increase the predictability of the troposphere after the event, at least locally. According to our hypothesis, vortex splitting in the stratosphere is a dramatic manifestation of an interaction between planetary-scale structure in the stratosphere, ultimately associated with topography, and sub planetary-scale structure in the troposphere, associated with baroclinic instability.

The first recorded split of the stratospheric polar vortex occurred in the northern hemisphere during winter 1962, about 10 years after the phenomenon of the sudden warming was first discovered. The event inspired the seminal numerical experiments

of Matsuno (1971) into the cause of sudden warmings. His numerical model of the stratosphere had its lower boundary near the tropopause, and comprised a zonal-mean wind and a single longitudinal harmonic in geopotential height of wavenumber 2. By increasing in time the amplitude of the wave-2 harmonic at the lower boundary, Matsuno was able to simulate a split in the stratospheric polar vortex that had similarities to the event of 1962.

Since this work, the predominant dynamical paradigm for sudden warmings has been built on the theory of wave, mean-flow interactions (Andrews *et al.* 1987). It is proposed that a warming occurs when a planetary wave grows to large amplitude over a week or so in the troposphere at mid latitudes and propagates upward on the zonal-mean flow into the stratosphere, with the structure of the initial zonal-mean flow having a strong influence on whether a so-called major warming occurs. Smaller scale disturbances in the troposphere, such as develop during mid-latitude cyclogenesis, are not supposed to have a strong influence on such events in accord with the results of Charney and Drazin (1961). Their theoretical calculations using a linear, steady-state model indicated that such sub planetary-scale disturbances would be trapped in the troposphere.

Dynamical mechanisms for this signature of rapid planetary-wave growth in the upper troposphere remain a matter for investigation, however. For the typical zonal-mean state of the troposphere the dominant scale for disturbances to grow by baroclinic instability is sub planetary (synoptic) scale,

not the larger planetary scale. On the basis of a weakly nonlinear evolution equation, [Plumb \(1981\)](#) proposed instead that a resonance mechanism could be responsible for the growth of topographically forced planetary-scale waves in the troposphere-stratosphere system during sudden warmings. By using a simplified, one-layer model, [Esler and Scott \(2005\)](#) and [Esler et al. \(2006\)](#) proposed a mechanism for splitting that involves resonant excitation of a “barotropic mode” of the polar vortex. By using a similar model, [Mathewman and Esler \(2011\)](#) found that it was not necessary to invoke an anomalous tropospheric planetary wave source to explain vortex splitting events. They found that vortex splitting could arise as a spontaneous bifurcation of the forced vortex system when only a slight change was made in the forcing. [Kushner and Polvani \(2005\)](#) noted a spontaneous vortex splitting event after several decades of integration under perpetual solstice conditions of their simplified atmospheric general circulation model. Their model simulated baroclinic instability but had no forced planetary waves. Notwithstanding the dynamical insights afforded by such models, their simplifications and the flow regime they represent mean that their relevance to mechanisms for polar vortex splitting in the real atmosphere is an open question.

Another perspective on the dynamics of sudden warmings was offered by [O’Neill and Pope \(1988\)](#). They noted that if the timescale for vortex breakdown during sudden warmings is taken to be a week or so (the observed timescale of rapid changes in polar temperature and zonal wind in the stratosphere), then such events develop in mid winter not from stratospheric states close to being zonally symmetric but from highly asymmetric states comprising an elongated stratospheric polar vortex and an adjacent, well-developed anticyclone, i.e. a cyclone/anticyclone vortex pair or “planetary wave 1” precursor. In terms of the distinction often made between sudden warmings as being of the vortex displacement kind or of the vortex splitting kind, they see the former as involving vortex displacement and elongation in the nonlinear, growing perturbation, and the latter as involving this behaviour with the extra feature that the vortex goes on to split. On the basis of idealised experiments with a fully nonlinear model, they contended that aspects of the dynamics of sudden warmings were better described in terms of vortex-vortex interactions rather than in terms of planetary waves propagating upward on a zonal-mean flow.

Does such initial asymmetry in the stratosphere have a bearing on the mechanisms that can trigger a major warming as well as on the evolution of the event itself? [Charlton et al. \(2005\)](#) proposed that it does. They studied the only recorded split of the polar vortex in the stratosphere of the southern hemisphere, which took place during September 2002. They noted that, during the run up to the split, strong cyclogenesis occurred on a sub planetary scale through the depth of the troposphere under a tip of a stratospheric polar vortex that was already elongated in the presence of a quasi-stationary anticyclone. Such a pattern is a characteristic of the stratosphere in the southern hemisphere during spring (e.g. [Mechoso et al. 1988](#)), though there were some differences from normal conditions to be noted later.

The basic dynamical concept is illustrated schematically in [Figure 1](#). The polar vortex is elongated by planetary-scale structure in the troposphere and lower stratosphere, ultimately associated with topography. The tip of the vortex has strong local curvature on a sub planetary scale. Underneath it in the upper troposphere is developing a strong, sub planetary-scale cyclone (local zonal wave number about 4) owing to baroclinic instability in the extra-tropical jet stream. The potential vorticity associated with the cyclone is envisaged as inducing an anomaly wind field near the tip of the polar vortex, tending to form a closed circulation. Owing to the tendency of the elongated

distribution of potential vorticity in the rest of the polar vortex to circularize ([Dritschel 1990](#)), the effect is to split the polar vortex into distinct cyclones in the middle or upper stratosphere. The figure depicts schematically an example of dynamical up-scaling, whereby localized cyclogenesis in the troposphere induces a much larger scale response in the stratosphere.

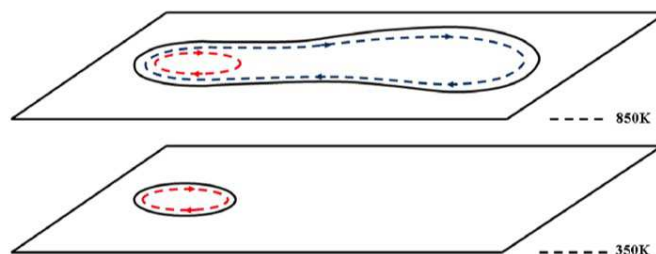


Figure 1. Schematic illustrating, for the southern hemisphere, the hypothesized mechanism for splitting the stratospheric polar vortex. The solid black lines denote the notional edges of the upper and lower vortices defined by their potential vorticity gradients. Dashed lines denote isopleths of potential vorticity, with the red circular isopleth in the upper panel signifying an induced effect of the cyclone vertically below.

The purpose of our paper is to test this hypothesis specifically with reference to the southern hemisphere event of 2002, and also to consider the relevance of the hypothesis to similar vortex splitting more commonly observed in the winter stratosphere of the northern hemisphere. Our analysis includes use of an inverse modelling or adjoint technique to explore the sensitivity of the stratosphere to perturbations in the troposphere. This approach, which is based on the use of tangent linear and adjoint versions of a fully nonlinear atmospheric general circulation model, was inspired by the work of [Jung and Barkmeier \(2006\)](#). They focused on the dynamical impact on the troposphere of the changes in the stratosphere to assess the need to represent the stratosphere in extended range weather forecasting. Our study, by contrast, focuses on the converse question of how the troposphere affects the stratosphere.

This paper is organized as follows. In section 2 we give an overview of changes in the large-scale circulation of the southern hemisphere during the vortex splitting (major warming) event of spring 2002, drawing attention to concurrent developments in the troposphere. Section 3 uses a potential vorticity (PV) inversion algorithm to compute the instantaneous (time-independent) response of the wind field in the stratosphere to a localized PV cyclonic anomaly in the troposphere when the polar vortex is elongated. We conclude that the splitting cannot be explained simply in these terms, but requires a model for (nonlinear) dynamical evolution. Section 4 uses such a model, as well as its tangent-linear and adjoint versions, to test the sensitivity of polar vortex splitting during 2002 to localized forcing in the troposphere located in the region of the observed cyclogenesis. Results support our main hypothesis. Vortex splitting events in the stratosphere of the northern hemisphere, considered briefly in section 5, show similar behaviour to their counterpart in the southern hemisphere, allowing us to suggest that the same dynamical mechanism is at work. The discussion Section 6 raises a number of issues arising from our work. These include how our hypothesis can be tested further, and how our proposed mechanism relates to polar vortex

displacements and to the possible role in SSWs of tropospheric blocking. Our conclusions are in Section 7.

2. Summary of the Southern Hemisphere (SH) 2002 event

The split in the polar vortex in the SH on 24 September 2002 was the only such event to have been recorded in the SH, though such events are witnessed sporadically in the northern hemisphere (NH), roughly two or three times a decade. [Charlton *et al.* \(2005\)](#) have described the synoptic evolution of the event in the stratosphere. We draw attention to some elements of particular dynamical interest here.

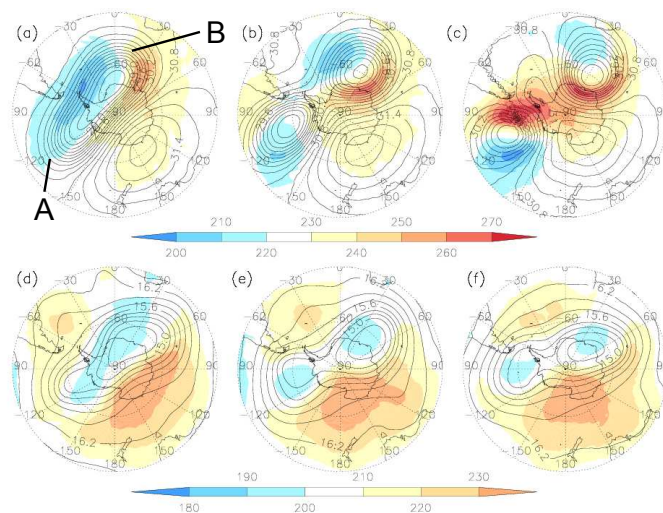


Figure 2. Fields of geopotential height (contours at 0.2 km intervals) and temperature in K (shaded) for the 10 hPa surface of the southern hemisphere on (a) 22, (b) 24, (c) 26 September 2002. Panels (d)–(f) show the corresponding plots for the 100 hPa surface. The labels A and B in (a) identify respectively the barotropic (upright) and baroclinic (westward leaning) tips of the polar vortex. The Greenwich Meridian is at the top on each map. Data from ECMWF Operational Analysis (available from [BADC \(2016\)](#))

Figure 2, panels (a), (b) and (c), show the evolution of the event in the middle stratosphere in terms of geopotential height and temperature. (Related fields of PV will be referred to in section 4, Figure 9.) The stratospheric precursor state shows (panel (a)) a highly elongated vortex in the presence of a strong quasi-stationary anticyclone – the classical wave-1 pattern during a SH final warming. The highest temperature anomaly is at the jet entrance owing to adiabatic descent ([Fairlie *et al.* 1990](#)). Because of the strong temperature gradients, we shall refer to the nearby tip of the vortex as the “baroclinic tip,” in contrast to the other tip of the vortex, where temperature gradients are weaker, which we shall refer to as the “barotropic tip” (marked respectively as B and A on the figure). As the vortex starts to split (panel (b)), the emerging cyclone pair are about the same size, but with one cyclone (near 30°E) westward leaning (baroclinic) and the other (near 120°W) vertically upright (barotropic). This barotropic/baroclinic asymmetry on the point of splitting is a noteworthy feature of such events in the NH (Section 5). It is destroyed when the vortex actually splits, and there is strong shear in the flow as the anticyclone penetrates between them over the pole (involving material advection as evidenced by the evolution of PV).

Figure 2, panels (d), (e) and (f), show the corresponding evolution in the lower stratosphere. The planetary-scale high, quasi-stationary temperature anomaly (panel (d)) is a characteristic of the springtime final warming, as noted by [Mechoso *et al.* \(1988\)](#), and is evidently associated with the presence of the anticyclone in the mid stratosphere (by simple considerations of atmospheric

thickness). The polar vortex is starting to elongate and to develop a sharper curvature at 120°W to form, on subsequent days, two cyclonic centres within a surrounding westerly flow. This structure contrasts with the complete split of the vortex in the middle and upper stratosphere, and with the complete cutting off of the pair of cyclones from each other.

Figure 3 shows fields of geopotential height at 500 hPa in the mid troposphere shortly before the split (including a three-day forecast field with a numerical weather prediction model described in section 4). The fields exhibit zonal asymmetry on both a planetary scale (note the area of generally low values of geopotential height in the top right quadrant of the panels) as well as structure on a sub planetary scale, in particular, the presence of a cyclone near 120°W (to the west of the southern tip of S. America) in the mid-latitude, meandering jet stream. This feature developed as a slowly eastward progressing, large-amplitude Rossby wave train “broke” locally to form a closed circulation under a collocated tip of the polar vortex (see Figure 2, panels (a) and (d)). The cyclone extended from the surface through the whole depth of the troposphere, and had local zonal wave number around 3 to 4. The splitting of the polar vortex led to a vertical extension of a closed cyclonic circulation to the stratopause. Other strong cyclonic systems, which lay under regions of low curvature in the polar vortex, were locally evanescent in the vertical.

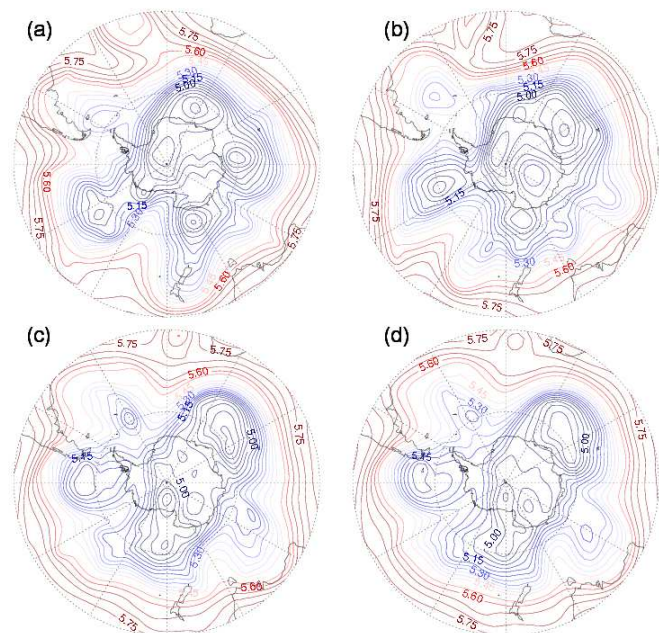


Figure 3. Geopotential height at 500 hPa in the mid troposphere for the southern hemisphere 2002 on days: (a) 20; (b) 22; (c) 24 September. Data from ECMWF Operational Analysis ([BADC \(2016\)](#)). (d) Numerical forecast for 24 September, initialized on 21 September using an ECMWF model, as discussed in section 4. The Greenwich Meridian is at the top of each map. Contour interval: 0.05 km.

Figure 4 gives the potential vorticity perspective of these developments. It shows a sequence of PV fields on an isentropic surface in the upper troposphere/lower stratosphere. There is a planetary-scale area of low (cyclonic) PV (roughly in the sector 0°E to 90°E) with some sub planetary-scale internal structure. Above this area, the polar vortex is displaced about 20° to the west. The dominant development in relation to the polar vortex appears to be the extrusion from the large polar area of low (cyclonic) PV and rolling up of the extruded PV near 120°W. This resulted in the formation of the closed cyclonic circulation directly below a tip of the vortex on a similar horizontal scale and with similar curvature. The essence of our hypothesis is that judiciously

located cyclogenesis was a *necessary* condition for splitting in this case, and in other cases considered later. Thus sub planetary-scale developments in the troposphere well removed from the polar vortex, and therefore under low PV gradients in the stratosphere, would not be expected to be significant for vortex splitting, for instance the PV extrusion and associated cyclogenesis in the quadrant 90°E to 180°E in Figures 3 and 4.

The tropospheric evolution witnessed in the PV fields seems to be a classical example of baroclinic instability of the LC-2 class, as described by [Thorncroft *et al.* \(1993\)](#). The scale of the 350 K PV anomaly is sub planetary scale, approximately local wave number 3 to 4 (the PV anomaly fits roughly into a 50 degree longitudinal sector), which is a common scale for such structure in the mid latitude jet stream in the troposphere in winter.

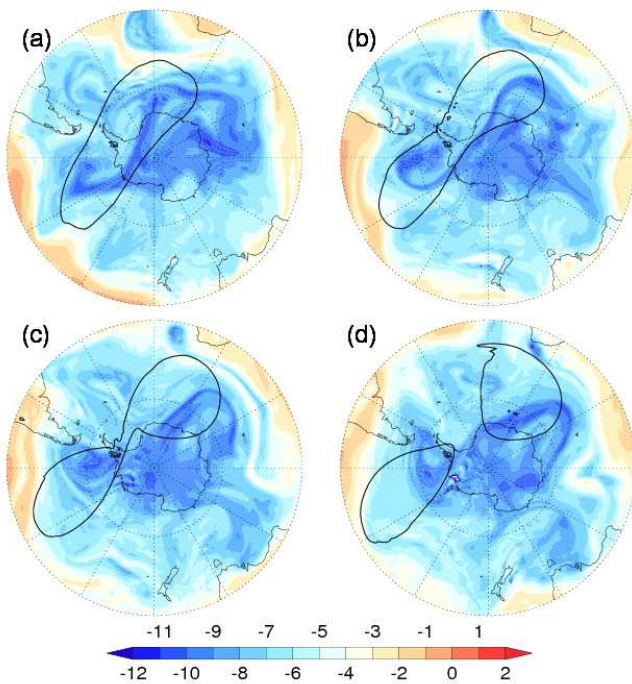


Figure 4. Fields of Ertel's potential vorticity on the 350 K isentropic surface for the southern hemisphere in 2002 on days: (a) 21, (b) 22, (c) 23 and (d) 24 September. Units: PV units, where 1 PV unit = $10^{-6} \text{ K m}^2 \text{ kg}^{-1} \text{ s}^{-1}$. The single isopleth of PV (solid black line) on the fields marks the edge of the polar vortex on the 850 K isentropic surface in the middle stratosphere. The Greenwich Meridian is at the top of each map. Data from ECMWF Operational Analysis (available from [BADC \(2016\)](#)).

The vertical structure of the splitting and the rapid transition from sub planetary scale to planetary scale are further illustrated in Figure 5. The barotropic/baroclinic structure referred to earlier is seen clearly here. The barotropic (upright) geopotential “low” lies above the cyclogenesis we have described, and corresponds to a closed cyclonic circulation extending as a vertically aligned column from the ground to the stratopause. The baroclinic (westward leaning) geopotential “low” was present in the anticyclone/elongated vortex precursor state. This structure is also a characteristic signature of vortex splitting events in the mid winter stratosphere of the northern hemisphere (section 5). Consistent with the structure shown in Figure 5, the splitting was accompanied by the growth of planetary wave number 2 in the lower stratosphere and by an increase in the associated vertical component of the Eliassen-Palm flux at 100 hPa, as found by [Newman and Nash \(2005\)](#). The structure shown in Figure 5 shows, however, that we cannot consider the splitting solely in terms of the upward propagation of planetary-wave 2 from the troposphere; the structure cannot be described simply as a “vertically twisted wave 2.” Moreover, we contend that the mechanism for this development was not one involving the

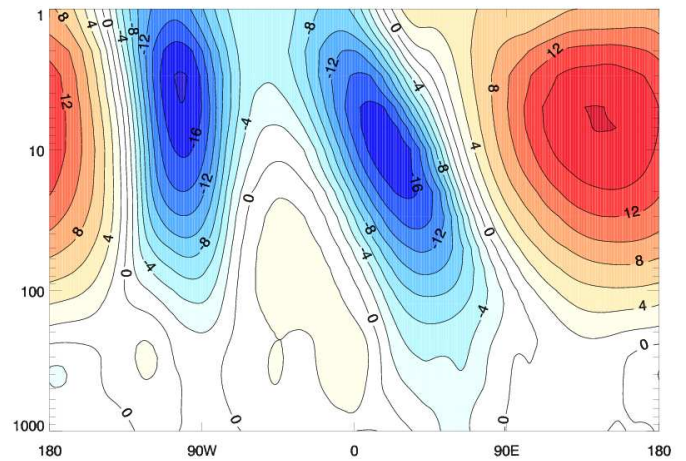


Figure 5. Longitude-height section at 60°S on 24 September 2002, just before the split of the polar vortex, showing the departure of geopotential height from the zonal average. Contours at 2 km intervals. The upright (barotropic) low anomaly to the left lies above a sub planetary-scale, closed cyclonic circulation in the troposphere. Data from ECMWF Operational Analysis (available from [BADC \(2016\)](#))

preferential selection for growth of a planetary-wave 2, but the one we have proposed, appearing as a dynamical up-scaling from the sub planetary scale in the troposphere to the planetary scale throughout the stratosphere (including 100 hPa). In the context of this mechanism, therefore, involving a barotropic vortex interaction embedded in a baroclinic flow, the Eliassen-Palm flux cannot unequivocally be interpreted as indicating wave propagation, since it is a non-local (zonally averaged) diagnostic.

The barotropic/baroclinic structure exhibited in Figure 5 is a property of the polar vortex, considered as a physical entity, as can be inferred from the PV fields shown in Figure 4, where one can infer a columnar PV structure at one tip of the polar vortex and a westward leaning structure at the other. This observation (confirmed by inspection of the other splitting events shown in Section 5) contrasts with the paper of [Mathewman and Esler \(2011\)](#), who concluded that splitting events are typically barotropic.

Further support for our proposed mechanism is given by a very similar event that occurred in the SH earlier in September 2002. Figure 6 shows that, on this occasion, the polar vortex became exceptionally elongated, but did not split. Cyclogenesis with wave breaking (seen as PV roll up) occurred under a tip of an already elongated polar vortex, and as with the event later in the month this happened near 120°W . Following the approach of [Vaugh \(1997\)](#), [Mitchell *et al.* \(2011\)](#) showed that the ellipticity of the polar vortex in the middle stratosphere (ratio of its semi-major to its semi-minor axis) reached the value of 3.5 for the early September event when the vortex did not split, but the value of 5.5 for the late September event when the vortex did split. One reason for the difference between the events seems to be that for the earlier of the two, the closed cyclonic circulation in the troposphere broke down quickly after the time of maximum ellipticity of the polar vortex.

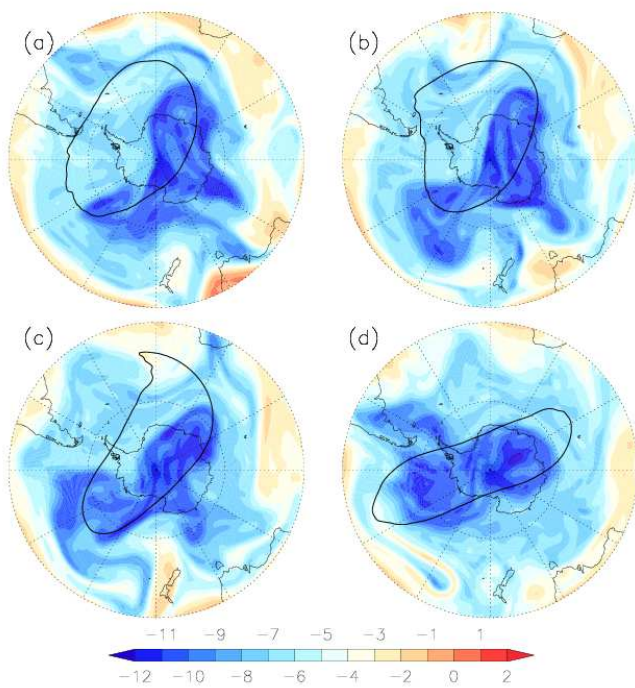


Figure 6. Fields of Ertel's potential vorticity on the 350 K isentropic surface for the southern hemisphere in 2002 on days: (a) 2, (b) 3, (c) 4 and (d) 6 September. Units: PV units, where 1 PV unit = $10^{-6} \text{ K m}^2 \text{ kg}^{-1} \text{ s}^{-1}$. The single isopleth of PV (solid black line) on the fields marks the edge of the polar vortex on the 850 K isentropic surface in the middle stratosphere. The Greenwich Meridian is at the top of each map. Data from ECMWF Operational Analysis Program (available from [BADG \(2016\)](#)).

3. PV anomalies and PV inversion

In this section, we use the technique of PV inversion to consider, with respect to a reference state comprising an elongated stratospheric polar vortex, the velocity field induced by an idealized PV anomaly extending through the troposphere. The size and magnitude of the anomaly are chosen to be similar to those of the closed cyclonic circulation discussed in section 2. The induced velocity field can be considered as an instantaneous response to the anomaly at a particular time under assumed conditions of dynamical balance. Since the induced velocity vectors are found to cut across the isopleths of PV in the reference state, the response must be dynamic rather than static, an aspect we consider in section 4. The PV inversion does, however, answer the preliminary question for the instantaneous case of whether position with respect to the polar vortex - in particular position under regions of high cyclonic curvature - makes a difference.

The results of the PV inversions shown here are derived by using the algorithm developed by P. Berrisford (pers. comm.) and described by [Charlton *et al.* \(2005\)](#) and by [Oatley \(2010\)](#). The PV inverter is nonlinear, and should therefore be more accurate than quasi-geostrophic PV inverters, which are based on linearizing about a reference state. It is three-dimensional, and includes terms that represent the horizontal curvature of the flow. The inversion is performed on isentropic surfaces on a sphere, assuming hydrostatic and geostrophic balance. The reference state was taken to be the PV distribution in the real atmosphere on 23 September 2002, just before the polar vortex split.

Figure 7 shows the induced anomaly velocity field in the stratosphere when the PV anomaly was placed in the troposphere under the tip of the polar vortex where cyclogenesis occurred. The curvature in the reference wind field is locally strong. There is a cyclonic wind anomaly in the stratosphere, coincident with a tip of the vortex, directly above the PV anomaly in the troposphere - i.e. the three-dimensional structure of the anomaly is barotropic,

as was found in the observed event. Anomaly wind vectors cut across the middle of the elongated vortex. If the anomaly winds grew strong enough, such a structure would be expected to cut off a portion of the polar vortex. The nonlinear, time evolving numerical experiments reported in next section support this contention.

Theoretical estimates of the Rossby scale height of a PV anomaly can be made both when the wind field in the reference state has no curvature at the position of the underlying PV anomaly and also when it does. Assuming no wind curvature, and making other simplifying assumptions as discussed by [Hoskins *et al.* \(1985\)](#), the Rossby scale height, H_{Rossby} , is given by

$$H_{\text{Rossby}} \sim fL/N,$$

where f is the planetary vorticity, L is the horizontal length scale of the PV anomaly and N is the Brunt-Väisälä frequency. Wind curvature modifies the Rossby scale height ([Hoskins *et al.* 1985](#)), which we denote H_{curv} , given by

$$H_{\text{curv}} \sim f_{\text{loc}}L/N,$$

where the planetary vorticity now has a local value $f_{\text{loc}} = f + 2v/r$ for a wind field with local radius of curvature r and speed v .

For various positions of the PV anomaly under the polar vortex and for the same reference state just before the split, we compared these theoretical estimates of the vertical scale height of an anomaly with those deduced from the full PV inversion. We found the following: (1) the anomaly scale height from direct PV inversion is greatest, about 12 km, above the actual position of the PV anomaly in the troposphere where the curvature of the wind curvature in the polar vortex is greatest; (2) the scale height at a “straight” edge of the elongated vortex is lower, about 9.5 km; (3) H_{Rossby} underestimates the actual scale height at a tip of the vortex by about 2.5 km; and (4) H_{curv} is in much better agreement with the actual scale height, with a difference of ± 0.2 km.

The key result of this section is that, while theory and full PV inversion agree that local curvature in the horizontal wind field of a static (time-independent) reference state does increase the scale height of response to the PV anomaly, the induced anomaly wind field is weak at the level of the middle stratosphere (a few ms^{-1} , [Figure 7](#)), too weak, it might seem, to disrupt a vortex surrounded by strong PV gradients.

For the actual time-evolving situation, however, we can reason heuristically as follows. The PV anomaly in the troposphere would induce a wind anomaly above it (at heights less than H_{curv} , say), which would advect the PV at a higher level changing the PV and wind distributions more than in the static case. In conditions where the polar vortex is susceptible to splitting, such “upward burrowing” might extend to a much greater height than estimated by H_{curv} , as we suggest occurred during the splitting event of 2002.

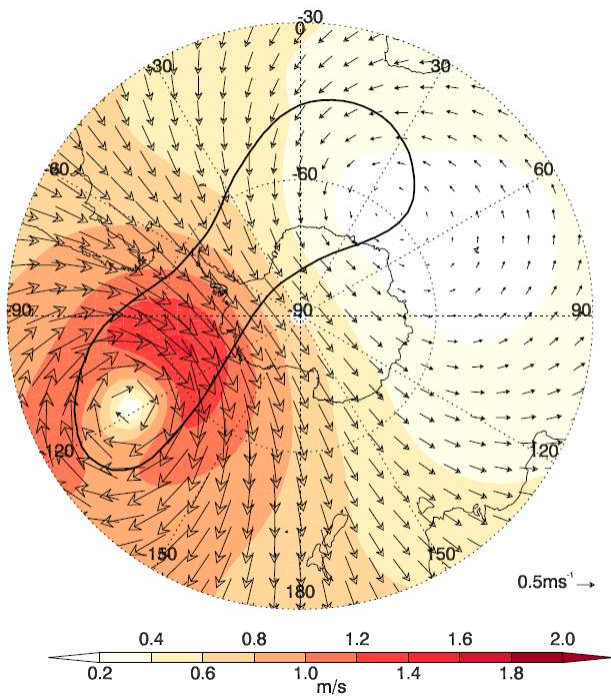


Figure 7. Instantaneous (static) response of the wind field at 10 hPa to an idealised (circular) PV anomaly of magnitude, scale and location similar to that of the cyclogenesis near 120°W. The anomaly extended vertically in the troposphere from about the 305 K isentropic surface to the 380 K isentropic surface. The anomaly wind field was calculated by using a fully nonlinear PV inversion algorithm for a sphere. The black line locates the edge of the polar vortex at 90°K on 23 September 2002. The Greenwich Meridian is at the top of the map.

4. Sensitivity of vortex splitting to tropospheric cyclogenesis

Moving beyond the results and limitations of static PV inversion, we now describe a time-dependent method to test our hypothesis that the split of the stratospheric polar vortex during September 2002 was triggered by localized cyclogenesis in the troposphere. The method has two steps. (1) By using the linearized adjoint of a fully nonlinear model, we determine a “sensitivity field,” which quantifies the sensitivity of polar vortex splitting to (multi-variate) changes in the troposphere. (2) By then taking sub-regions of this field, we construct a localized forcing in the troposphere to see whether, in a fully nonlinear simulation, reducing the intensity of tropospheric cyclogenesis on a sub planetary scale can significantly impede the splitting.

4.1. Models used

We used a fully nonlinear atmospheric general circulation model. It was a version of the ECMWF Integrated Forecasting System, cycle 32r2, used operationally between March and September 2004. The horizontal resolution corresponded to T_L159 , using a linear, Gaussian grid with a horizontal spacing of approximately 125 km. There were 60 levels in the vertical, extending up to 0.1 hPa, half of which were located above the tropopause. It successfully simulated, as an initialized forecast, the splitting event of 2002 (Simmons *et al.* 2005). In particular, it gave an excellent forecast of the cyclogenesis in the troposphere near 120°W of particular concern here (Figure 3: panel (c) observed field; panel (d) three-day forecast).

In order to study the sensitivity of the stratosphere to perturbations in the troposphere, we utilize tangent-linear and adjoint versions of this model, both of which involve a linearization of the governing equations (with linearized diabatic processes) about a trajectory produced by the fully nonlinear model. See Jung and Barkmeier (2006) for details.

4.2. Description of the methodology

With reference to the Appendix for mathematical details and to the schematic Figure 14, the basic idea is as follows. We constructed a localized forcing field in the troposphere, one designed specifically to inhibit the observed cyclogenesis. We then applied this forcing field in a full nonlinear simulation, which when unforced produced the split in the stratospheric vortex. We thereby sought to drive the stratosphere back towards a chosen prior target state about a week before the split, that of 17 September 2002 - not in fact the complete atmospheric state on that date but just the polar region in the middle stratosphere south of 60°S. Conceptually, this past state is a very rough surrogate for a future (unrealized) state that the atmosphere might have achieved in the absence of the cyclogenesis.

To derive the appropriate forcing in a systematic way, we used the adjoint model. We first defined a scalar measure of the “distance” between the target state and a forced model state at the end of a two-day period of integration. We refer to this period as the optimization period. By using the adjoint model, we then computed the gradient of this distance, or “miss-hit,” with respect to the applied forcing field. The resulting “sensitivity gradient” is three-dimensional and multi-variate, involving all the prognostic variables of the model.

We then took the forcing field to be constant in time and proportional to the negative of the sensitivity gradient. The forcing field, considered as a vector, was therefore chosen to point in a direction that would move the forced model state closer to the (unsplit) target state. A notable feature of the forcing field is that, like the sensitivity gradient it is derived from, it is multi-variate. It is also approximately dynamically balanced. This is because the adjoint model inherits this property directly from the forward model. Moreover, the forcing field extends through the whole depth of the troposphere; it is not just applied at a single level near the base of the stratosphere.

The algorithm used to calculate a finite-amplitude forcing field from the sensitivity gradient was iterative (details in Jung and Barkmeier (2006)). It required that, *for the two-day optimization period only*, the perturbation dynamics - the difference between model trajectories with and without forcing, not the trajectories themselves - could be described by the tangent-linear version of the fully nonlinear model. This requirement limited the magnitude of the forcing that could be applied. Oatley (2010) showed that our numerical experiments adhered to this constraint. Nevertheless, the forcing we applied was large enough to have substantial nonlinear effects when the simulations were run beyond the two-day optimization period.

Errico (1997) gives a wider perspective on the use of adjoint models in meteorology, and notes their ample scope to advance dynamical understanding.

4.3. The sensitivity gradient

Using the (nonlinear) ECMWF model, we ran an ensemble of 6 10-day forecasts, starting on successive days from 19 to 24 September 2002 during the week before the split. Each one was initialized with the observational data used by ECMWF in operational forecasting at that time. Following the procedure set out in sections 4.2 and in the Appendix, we calculated the sensitivity gradient, the left-hand side of Equation 6, at the start of each of these two-day optimization periods. There was little qualitative change in the sensitivity gradient during that period, so we confine attention to results for the forecast initialized on 21 September.

Figure 8 shows the sensitivity gradient for that day for the forcing fields of temperature and wind at 500 hPa in the mid troposphere. (The computational algorithm returns values of the

sensitivity gradient for all model variables and at all model grid-points; the figure is just a subset.) In line with Equation 6, if a forcing field is applied to the model forecasts with values proportional to those shown, but of *opposite* sign, then the model trajectory is pushed *down* the gradient of $\|\mathbf{m}\|$ to smaller values, i.e. toward the target un-split state of 17 September. It can be seen that the calculation of the sensitivity gradient with the adjoint model has retained the property that the large-scale fields are in approximate geostrophic balance.

Before interpreting the sensitivity gradient further, we note the following. (1) The sensitivity gradient indicates the *linear* response of the stratosphere to forcing as measured by $\|\mathbf{m}\|$; it does not necessarily indicate the response beyond a few days when linearity breaks down. Fully nonlinear simulations are required to do that, and to determine whether a particular feature in the sensitivity gradient corresponds to an important feature of the circulation for vortex splitting. (2) The sensitivity gradient indicates (in this linear sense) the *optimal* (down-gradient) “direction” in which to force the system to the target, unsplit state of the polar vortex. It does not necessarily correspond to any actual *developments* in the tropospheric circulation that were instrumental in the splitting (the route to splitting, or conversely to preventing splitting, is not unique). Such developments need to be identified in other fields, such as those of PV. (3) The target state is a very rough surrogate conceptually for a hypothetical (unrealized) stratospheric state that might have arisen (if our hypothesis is correct) in the absence of the tropospheric cyclogenesis. The sensitivity gradient and the forcing field derived from it are subject to these limitations. The main virtue of the sensitivity gradient, however, is that it gives a systematic way to derive a dynamically motivated, multi-level, multi-variate (internally consistent) forcing field to test hypotheses about the role of particular features in the circulation.

Two features of Figure 8 will be of particular interest in what follows: the one at position A, which is associated with the sub planetary-scale cyclogenesis and PV roll-up described earlier; and the one at position B, which is associated with a larger scale PV distribution below the polar vortex, where there was no obvious development such as there was at position A, at least in terms of PV. The sensitivity pattern shows, not unexpectedly, that the *optimal* way to suppress the split - i.e. to push the model trajectory directly down the gradient of $\|\mathbf{m}\|$ - would be in general to make the tropospheric circulation more zonally symmetric, in particular by simultaneously increasing the temperature and correspondingly decreasing the cyclonic circulations at A and B. We recall, however, point (2) above about optimality, which implies that such simultaneity is not necessary. The sensitivity pattern as computed does not distinguish between actual and hypothetical developments in the atmosphere. In the present context, the evolution of the PV fields is needed to do that.

We now go beyond the above linear analysis by using the sensitivity gradient to construct localized, finite-amplitude forcing fields using Equation 7 to study the nonlinear impact on vortex splitting of circulation features in the troposphere, specifically the cyclonic development corresponding to position A in Figure 8, and the larger scale, sub polar vortex PV distribution corresponding to position B. We selected only those components of the forcing \mathbf{f} given by Equation 7 within a mask of 60° longitude by 40° latitude centred first on position A and secondly on position B. All values of \mathbf{f} were set to zero outside the mask in the troposphere and everywhere above the tropopause.

4.4. Results of localized forcing experiments

We begin by focusing attention on the cyclogenesis at position A under the tip of the polar vortex. We used the reduced, multi-variate, multi-level forcing \mathbf{f} , constructed as describe above, to

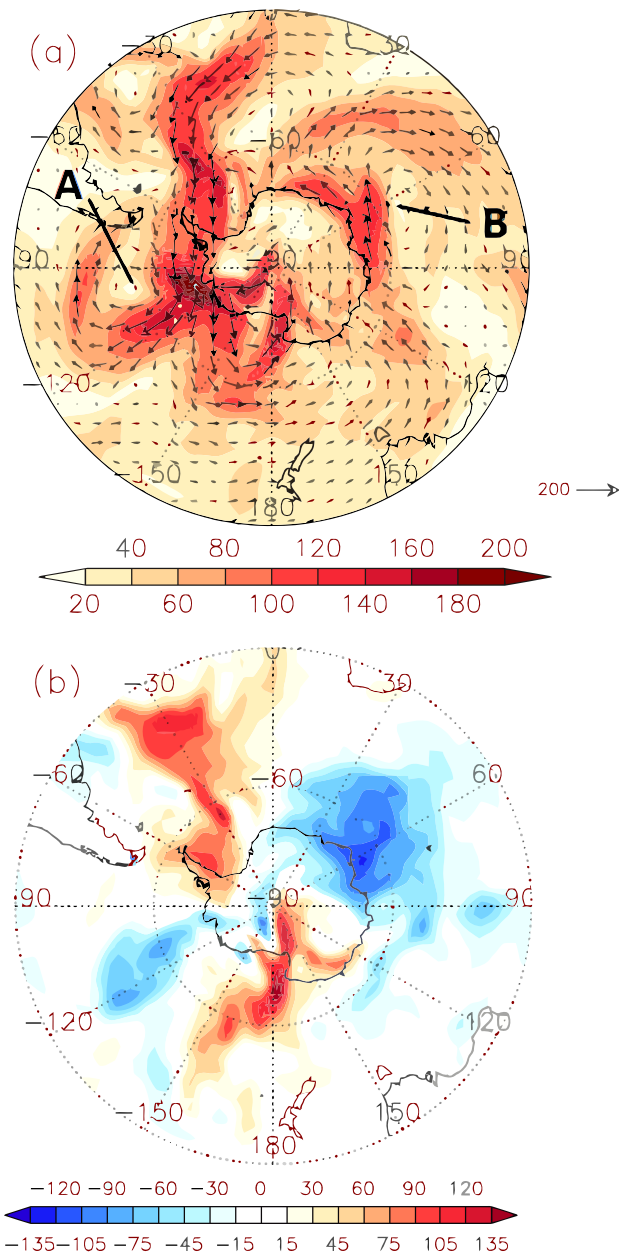


Figure 8. Sensitivity gradient at 500 hPa: (a) with respect to wind. Units: $10^4 \text{ J} (\text{ms}^{-1})^{-1} \text{ kg}^{-1} \text{ s}^{-1}$; (b) with respect to temperature. Units: $10^4 \text{ J K}^{-1} \text{ kg}^{-1} \text{ s}^{-1}$. For plotting purposes, the sensitivity gradient can be written as $\left. \frac{\partial \|\mathbf{m}\|}{\partial f_{ij}} \right|_{\mathbf{f}=0}$ to convert a vector quantity to an array of numbers. In this notation, f_{ij} is a forcing field for a particular model variable at a particular level at (longitude, latitude) grid-point (i, j) .

study the impact on the splitting of reducing the growth-rate of the cyclogenesis. We ran an ensemble of 6 fully nonlinear simulations with the ECMWF model initialized one day apart during the period 19 to 24 September 2002, each simulation lasting 10 days. One set comprised unforced simulations; the other set comprised forced simulations in which a constant localized forcing, constructed as previously described, was applied to partly weaken locally the sub planetary-scale cyclone in the troposphere. The forcing corresponded to a spin-down time of the cyclone of roughly 5 days.

The impact on the stratosphere was considerable. In the unforced simulations, the evolution of the zonal-mean wind very closely followed that of the ECMWF operational analyses, with a reversal from westerly to easterly characteristic of a major SSW. In the forced simulations, the reversal was prevented at 60°S out to 10 days, even for the simulation initialized just one day before the split (Oatley 2010). For model simulations initialized

on 21 September, 3 days before the split, Figure (9) compares the evolution of PV on the 850 K isentropic surface in the mid stratosphere without forcing (top set of plots) with the evolution with forcing (bottom set of plots). In the control simulation, which closely mimicked the actual event, the highly elongated vortex developed a distinct dumbbell shape and then split on 24 September (at 72 hours in the figure) into two distinct cyclones, roughly equal in size and intensity. In the forced simulation, on the other hand, the vortex elongated but did not split by 24 September, and the PV anomaly was significantly weaker and filamentary in the western hemisphere (near 120°W) above the tropospheric forcing (position A in Figure 9) than it was in the control experiment. It is evident that, by this time, the localized forcing had affected the whole of the vortex structure, including the formation of the cyclonic vortex in the eastern hemisphere, whose own formation into a cut-off feature in association with PV roll up was delayed. Subsequently, the two extremities of the polar vortex in the forced experiment did exhibit vortex roll up. In contrast to the unforced experiment, however, the cyclonic pair formation was highly asymmetrical, with a much smaller cyclone in the western hemisphere on 25 September, and with later evolution also significantly different.

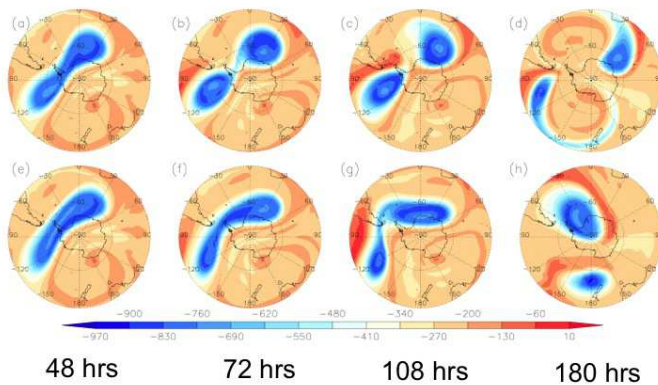


Figure 9. Fields of Ertel potential vorticity at 850 K for the unforced (control) forecast initialised on 21 September 2002, top four fields, and for the corresponding forecast with forcing confined in the troposphere to the sector 75°W to 135°W underneath a tip of the elongated polar vortex, bottom four fields. Units: PV units where 1 PV unit = $10^{-6} \text{ K m}^2 \text{ kg}^{-1} \text{ s}^{-1}$. The Greenwich Meridian is at the top of each map.

Although the localized, sub planetary-scale forcing in the troposphere had a significant impact on the vortex dynamics in the stratosphere, it did not completely suppress the development of a cyclone pair, albeit that a much more asymmetric pair materialized as a result of forcing than actually occurred. This is hardly surprising given the simplifications and empirical choices made in the algorithm to compute the forcing described in section 4.2. Besides the degree of arbitrariness in the choice of target state, perhaps the most significant simplification relates to the limited dynamical information contained in a sensitivity gradient (a linear concept) used to compute a steady forcing when the stratospheric circulation was evolving highly non-linearly, and was on the verge of irreversible breakdown and instability through vortex splitting.

The strong sensitivity exhibited by the stratosphere to the localized forcing occurred even though the forcing was on the sub planetary scale of the cyclogenesis, in apparent contravention of the Charney and Drazin (1961) criterion, which favours the vertical propagation of planetary-scale Rossby waves. The Charney-Drazin criterion is derived, however, for simplified conditions of small perturbations on zonally symmetric

atmospheric states. Its apparent restrictions can evidently be broken when the atmospheric state is highly zonally asymmetric and the forcing is large amplitude.

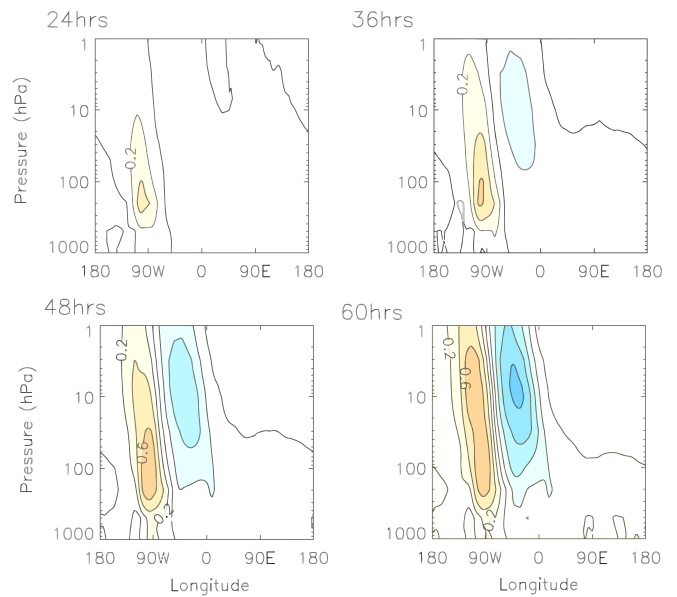


Figure 10. Difference fields (forced - unforced) of geopotential height at 60°S for forecasts initialised on 21 September 2002. The forcing is confined in the troposphere to the sector 75°W to 135°W underneath a tip of the elongated polar vortex. Contours at 0.2 km intervals.

The upward propagation of the effect of the localized tropospheric forcing is shown in Figure 10, which shows the difference in geopotential height between the forced and unforced simulations initialised on 21 September 2002. Over the first two days, the positive geopotential height perturbation extended vertically upwards rapidly through the depth of the troposphere into the stratosphere. The initial response was largely barotropic with little longitudinal tilt with height. This initial structure of the perturbation field was in line with that inferred from the PV inversion described in section 3. At this stage, two days into the simulation, the perturbation dynamics were approximately linear about the nonlinear trajectory of the unforced evolution, as noted in section 4.2. Thereafter, however, the perturbation developed very rapidly in the middle and upper stratosphere. The barotropic structure of the geopotential height perturbation near 90°W was retained to the upper stratosphere, acting to inhibit the formation of a closed, cut-off cyclone as part of the split. The apparent wave train developing from this point in Figure 10 corresponds to bodily shifts in the elongated vortex between the forced and unforced simulations, rather than to a simple Rossby wave propagating on a basic state.

It is evident therefore that the effect of the forcing in the troposphere was much bigger in the stratosphere in the time evolving case than would have been inferred from considerations of the static PV inversion alone (Figure 7). The very rapid growth of the effect in the stratosphere after an initial quasi-linear phase is indicative of a finite-amplitude instability in the system. Considering now not the “inverse” dynamics of cyclone suppression but the “forward” or actual dynamics of the event, our results support the notion that a finite-amplitude, sub planetary-scale development in the troposphere was instrumental in bringing the polar vortex to a point of no return - a bifurcation point - beyond which splitting and irreversible PV rearrangement occurred. The idea that vortex splitting in the stratosphere in 2002 is an example of a finite-amplitude instability was previously proposed by Esler and Scott (2005), but in the context of a different hypothesis for vortex splitting, that of barotropic resonance involving planetary wave number 2.

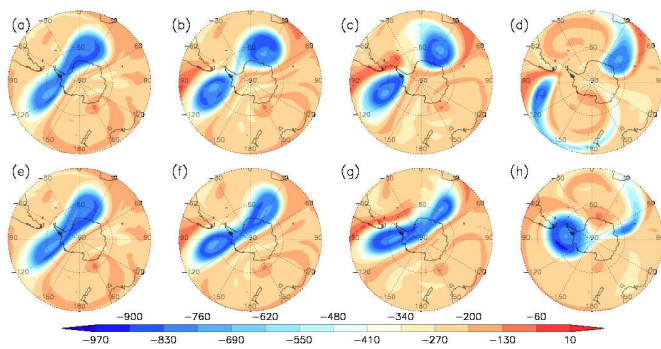


Figure 11. Fields of Ertel potential vorticity at 850 K for the unforced (control) forecast initialised on 21 September 2002, top four fields, and for the corresponding forecast with forcing confined in the troposphere to the sector 15°E to 75°E , bottom four fields. Units: PV units where $1\text{ PV unit} = 10^{-6}\text{ K m}^2\text{ kg}^{-1}\text{ s}^{-1}$. The Greenwich Meridian is at the top of each map. The times are the same as those in Figure 9

We turn now to the results of analogous forcing experiments, whereby localized forcing was applied in the troposphere at position B in Figure (8). The forcing was constructed as before, but this time with the effect of spinning down (over a time scale of about 5 days, as before) the cyclonic circulation associated with the large-scale region of low (cyclonic) PV underneath the polar vortex (Figure 4). For model simulations initialized on 21 September, Figure 11 compares the evolution of PV on the 850 K isentropic surface in the mid stratosphere without forcing (top set of plots) with the evolution with forcing (bottom set of plots). Again, even though the forcing was localized and sub planetary scale, the effect was considerable, and similar in outcome for the polar vortex (ignoring the approximately 180° longitudinal shift) to when forcing was applied at position A. In particular, vortex splitting was impeded, with asymmetric filamentation now in the eastern hemisphere. The larger remaining area of low (cyclonic) PV moved over the top of the tropospheric cyclone at 120°W , where it rolled to form an approximately axially symmetric vortex.

Noting this positional shift of the remainder of the polar vortex as a physical entity, our interpretation is that although there was not apparently any sub planetary-scale Rossby wave breaking in the troposphere at position B as there was at A, the low (cyclonic) planetary-scale PV distribution in the vicinity of B through the troposphere provided an “anchor” for the polar vortex through the PV-induced circulation above. We surmise that the presence of this PV anchor meant that the extrusion and roll-up of PV in the troposphere near A could induce the necessary stretching velocity field in the stratosphere leading to vortex elongation and eventual splitting. Weakening this anchor allowed the polar vortex to move bodily with less stretching.

Understanding the fluid dynamics of such vortex interactions, in particular of the details of vortex stretching and roll up, will require an extension of the kinds of idealised studies of vortex dynamics by, for example, Kida (1981), Reinaud *et al.* (2003) and Scott and Dritschel (2005), who analysed the behaviour of ellipsoidal vortices in shear flows. Fully three-dimensional studies of the two-vortex problem, treating interactions between the polar vortex and a tropospheric cyclone, should progress to studies of the three-vortex problem, treating in addition interactions with the planetary-scale anticyclone in the stratosphere that accompanies polar vortex elongation.

5. Comparison with vortex splitting events in the Northern Hemisphere (NH)

We now propose that sub planetary-scale cyclogenesis in the troposphere under the tip of an elongated polar vortex is also the mechanism behind polar vortex splitting in the stratosphere

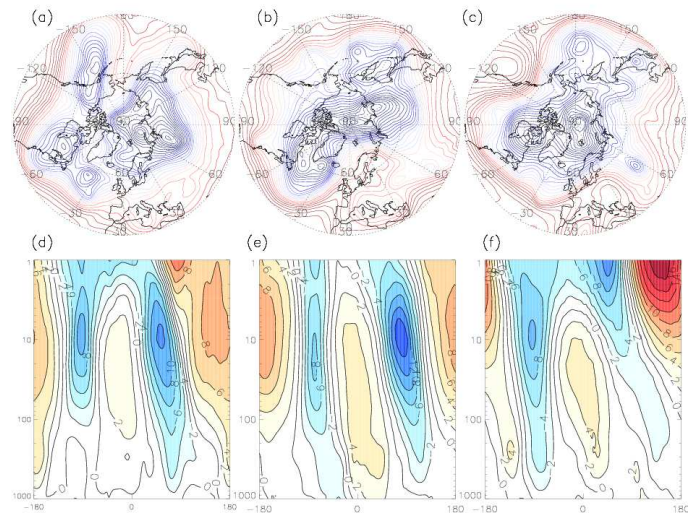


Figure 12. Fields of geopotential height at 500 hPa and corresponding longitude-height sections of the departure of geopotential height from the zonal average at 60°N : (a) and (d) 16 January 1971; (b) and (e) 19 February 1979; (c) and (f) 12 January 2009. The Greenwich Meridian is at the bottom of each map and at the centre of each vertical section. Contour interval for the maps: 0.05 km. Data from ERA-Interim Re-analysis (available from BADC (2016))

of the NH, and is not just a feature of the rare splitting event observed in the SH. We give 6 examples, which we consider only briefly, drawing attention to similarities and differences, and emphasizing elements in common with the SH event of 2002. They are considered in two sets: 1971, 1979 and 2009; and 1963, 1984 and 1989. For the first set, the relevant cyclogenesis occurred approximately over the north eastern side of N. America. On the basis of the typical mid-winter state of the polar vortex in the NH, described below, we expect this set to exemplify the more common, or “classical,” type of polar vortex splitting in the NH. For the second set, the relevant cyclogenesis occurred approximately over the north eastern side of Eurasia in 1984 and 1989, and over north central Eurasia in 1963.

We have examined all 14 vortex splitting events in the NH listed by Martius *et al.* (2009) for the period 1958 to 2001, a list which includes all the above examples with the exception of 2009. With the exception of the event in December 1987, a so-called Canadian warming - which we believe was wrongly identified as vortex splitting when what actually occurred was a recombination with the polar vortex of previously eroded high (cyclonic) PV air - we judge that our proposed mechanism is applicable to all of them, with common features among some examples but with others showing individuality beyond our scope to detail.

The typical mid-winter state of the stratosphere in the NH comprises an elongated polar vortex and an anticyclone over the Pacific Ocean - the so-called Aleutian High (see, for example, O’Neill (2003)). On the basis of numerical experiments, O’Neill and Pope (1988) argued that the Aleutian High is the first anticyclonic anomaly in an upward and downstream propagating (nonlinear) Rossby wave train emanating from a climatological feature in the troposphere, the so-called E. Asian Low. They noted that an intensification of this cyclone/anticyclone vortex pair precedes any subsequent splitting of the cyclonic vortex, referring to a wave-1 precursor for splitting and to a nonlinear transition from wave number 1 to wave number 2 as a zonal harmonic signature of such events.

For our first set, the precursor state of the stratosphere just before polar vortex splitting was as just described. Figure 12 shows fields of geopotential height at 500 hPa in the troposphere and the corresponding vertical sections of geopotential height variation around 60°N about 2 or 3 days before the polar vortex split. In each case, an intense, closed cyclonic circulation has

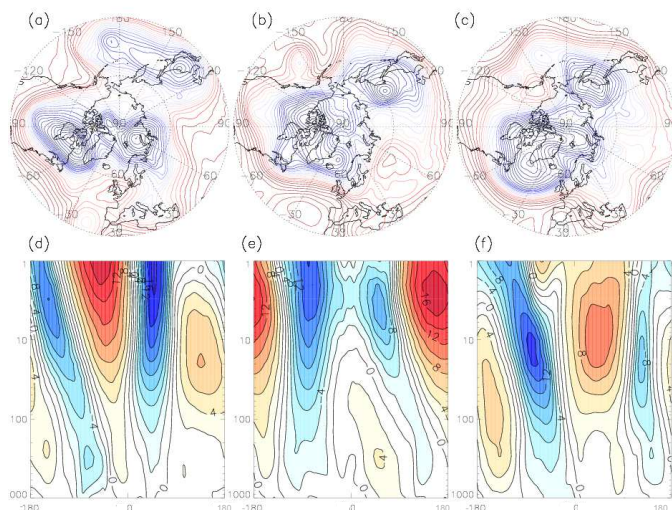


Figure 13. Fields of geopotential height at 500 hPa and corresponding longitude-height sections of the departure of geopotential height from the zonal average at 60°N: (a) and (d) 23 January 1963; (b) and (e) 21 December 1984; (c) and (f) 18 February 1989. The Greenwich Meridian is at the bottom of each map and at the centre of each vertical section. Contour interval for the maps: 0.05 km. Data from ERA-Interim Re-analysis (available from [BADC \(2016\)](#))

developed in the troposphere on a sub planetary scale over N.E. America under the tip of the polar vortex downstream from the Aleutian High. The closed cyclonic circulation extended as a locally barotropic (upright) structure through the troposphere into the middle or upper stratosphere. The barotropic/baroclinic structure is common to all events shown, and is very similar to that shown for the SH in Figure 5, where again the tropospheric cyclogenesis occurred downstream from the prevailing anticyclone in the stratosphere.

Figure 13, corresponding to our second set of NH splitting events, shows common features with the first set, notably the barotropic/baroclinic structure just before the split, but with clear geographical differences and with some noteworthy variety. The key difference with the previous set was that, before the split, the dominant cyclonic feature in the troposphere was not over N.E. Asia as before but over N.E. America, leading to a shift of the polar vortex and to a downstream anticyclone in the stratosphere over the Atlantic rather than over the Pacific. The split in the polar vortex resulted when sub planetary-scale cyclogenesis occurred over N.E. Eurasia in 1984 and 1989, and over central Eurasia in 1963. In 1963 and 1989, the barotropic structure above the cyclogenesis is evident, matching all our previous examples. In 1984, the corresponding structure over the cyclogenesis is more baroclinic; the barotropic structure is over the pre-existing cyclone over N. America. The event in 1984 seems to be a hybrid of the other events in the two sets. We suggest that an explanation for the difference between the sets can be found in the occasional switching of the dominant cyclonic structure affecting the precursor state of the polar vortex from N.E. Eurasia, the climatological position, to N.E. America, and that the 1984 event is an example of an in-between state.

6. Discussion

We now mention some related dynamical questions and issues that arise from our results.

- *What are the factors that determine the sensitivity of the polar vortex to splitting by cyclogenesis in the troposphere?*

The “inverse modelling” approach that we have adopted in this paper to test our ideas has the advantage of being directly relevant to an actual observed event. Trying to

undo, however, what the atmosphere has done in a highly nonlinear, rapidly evolving situation is difficult to achieve, and trying to understand the “inverse dynamics” is somewhat unnatural. To gain a better understanding of the dynamical role of sub planetary-scale cyclogenesis in polar vortex splitting, and to answer the question posed above, a more traditional “forward modelling” approach is needed. Numerical experiments using a fully nonlinear model - in which localized, sub planetary-scale potential vorticity anomalies, on various spatial scales, could be applied in the troposphere at various points beneath the stratospheric polar vortex in various states of elongation and zonal asymmetry - should be illuminating. A multi-level, nonlinear, mechanistic model in which a lower boundary forcing can be prescribed, of the kinds described by [Fisher \(1987\)](#) based on the primitive equations, or by [Dritschel and Viúdez \(2003\)](#) formulated around PV, could be used for this purpose.

- *Is the idea of a **localized** development of a cyclone in the troposphere as the essential ingredient for causing a split a valid dynamical simplification?*

The idea of localization is predicated on the idea that the relevant PV anomaly is associated with a closed, quasi-stationary circulation that can persist, rather than with a wavelike system that is mobile and therefore locally transient. Nevertheless, this kind of deconstruction of a nonlinear flow must be a simplification. In particular, the formation of closed cyclones in Rossby wave trains in the troposphere is commonly associated with some downstream or upstream ridging, leading to associated low (anticyclonic) PV anomalies that might need to be considered. In addition, the development over several days of an intense, closed cyclonic circulation through localized instability will have some dynamical consequences even farther afield, possibly triggering downstream cyclogenesis also relevant to splitting the polar vortex. In addition, during splitting the vertical alignment of a large PV anomaly in the stratosphere with that of the cyclone in the troposphere could increase the intensity, longevity and predictability of the cyclone itself. This last idea is supported by [Charlton *et al.* \(2004\)](#), who showed that a large change in stratospheric PV during an SSW can have a small but statistically significant effect on the intensity of tropospheric synoptic systems a week or so after the SSW. Numerical modelling will be needed to explore these issues.

- *Is it only through its projection onto planetary-wave amplitudes in the troposphere that cyclogenesis can lead to polar vortex splitting in the stratosphere?*

This question is prompted by the results of [Charney and Drazin \(1961\)](#) which show that, when the conditions of the theory apply, the winds in the stratosphere filter out the vertical propagation from the troposphere of all but the planetary waves. We have shown, however, by using an adjoint modelling technique, that an elongated polar vortex in the stratosphere is sensitive (in a linearized perturbation sense) to structures in the troposphere on a sub planetary scale, and more significantly that it is strongly sensitive to finite-amplitude, localized forcing on that scale. We have proposed a dynamical mechanism involving the vertical penetration of a localized, sub planetary-scale PV anomaly and the accompanying locally barotropic structure of the polar vortex as it splits. Delocalization by filtering out all but the planetary-wave components, and dynamical ideas based on the vertical propagation of planetary waves, do not seem to fit naturally with these ideas, and seem inconsistent with the spatial similarity of the tropospheric cyclone and the lobe of the polar vortex above it prior to splitting shown in

Figure 4 (left side of panels (a) and (b)). The conditions for splitting, involving a large-amplitude (localized) disturbance and an initially asymmetric polar vortex, are significantly different from those assumed by Charney and Drazin (1961). Numerical simulations with nonlinear mechanistic models allowing prescribed lower boundary forcing would be needed to investigate the question further.

- *How does our proposed mechanism for polar vortex splitting relate to how frequently they occur?*

Our proposed mechanism prompts new, hypothesis-driven statistical analyses in connection with vortex splitting. Regarding the exceptional behaviour exhibited by the SH polar vortex in September 2002, Oatley (2010) used a Lagrangian-based feature-tracking technique (Hodges 1994) to compare cyclogenesis in the troposphere of the SH in 2002 with that in the 1958–2008 climatology. She showed that storm track density and intensity were more intense in mid latitudes in the troposphere during winter and spring 2002 than in the climatology, and moreover that the PV anomaly associated with cyclogenesis and polar vortex splitting tracked well poleward of the climatological storm track. Regarding the polar vortex splitting in the NH, their much greater frequency than in the SH suggests that, besides the stronger and longer lasting elongation of the polar vortex due to stronger planetary-scale structure in the troposphere, there is some topographically induced alignment between that structure and the formation of the relevant intense, long-lived (cut-off) cyclones on the eastern side of the large continental land masses. The examples we have shown indeed tend to illustrate this.

7. Conclusions

The textbook mechanism for stratospheric sudden warmings (SSWs), comprising both vortex splitting and vortex displacement events, is typically stated along the following lines: that SSWs are caused by the amplification of quasi-stationary planetary waves in the troposphere, primarily zonal wave numbers 1 and 2, followed by propagation into the stratosphere and mean-flow deceleration (e.g. Holton and Hakim 2012, pp. 430–432). At least for SSWs involving a splitting of the stratospheric polar vortex, we hypothesize a different mechanism, and adduce evidence from observational data and a numerical model to support our hypothesis for the splitting event in the SH during spring 2002, and in less detail for similar events in the stratosphere of the NH.

We contend that the splitting of the polar vortex in the stratosphere of the SH during spring 2002 was caused by the development of a closed cyclonic circulation in the troposphere on a sub planetary scale below a tip of the vortex, which was elongated in the presence of a quasi-stationary anticyclone at the time of the (early) final warming. The initial zonal asymmetry and the associated strong curvature at the tip of the polar vortex, was central to its sensitivity to the sub planetary-scale in the troposphere, a scale that would be trapped in the troposphere were the circulation in the stratosphere westerly and zonally symmetric (Charney and Drazin 1961). That it was asymmetric depends in the first place on the presence of planetary-scale structure in the troposphere and lower stratosphere, which because of its geographical preferences must be ultimately associated with topography. The cyclogenesis, on the other hand, occurred through baroclinic instability of a high-amplitude (nonlinear) Rossby wavetrain in the middle and upper troposphere with local zonal wave number of about 4, reminiscent of “wave breaking” during idealised simulations of baroclinic instability of the LC-2 kind (Thorncroft *et al.* 1993). We suggest that splitting by this mechanism might be an example of a finite-amplitude instability,

whereby an elongated polar vortex has a tendency to split in two when one of its tips is forced to close off by a strong cyclonic perturbation beneath it.

Polar vortex splitting in the NH in winter is proposed to operate in a similar way, with the relevant cyclogenesis tending to occur approximately over the eastern seaboard of one or other of the two continental landmasses, with exceptions (e.g. during the splitting event of 1963). The variety between individual cases warrants further study.

The development and persistence of a closed cyclonic circulation in a nonlinear Rossby wavetrain lends credibility to the notion that it is worthwhile to consider, as we have done, its impact on the stratosphere in isolation from that of accompanying structure elsewhere (i.e. anticyclonic ridges or blocking patterns), at least to begin with. The persistence and quasi-stationarity of closed circulation system within an otherwise mobile Rossby wavetrain makes the troposphere inherently more predictable at the time of polar vortex splitting and SSWs.

We demonstrate the strong sensitivity of vortex splitting to tropospheric cyclogenesis by using an inverse modelling technique to construct a multivariate forcing field to suppress the cyclogenesis locally. We go on to suggest forward modelling experiments to test our ideas more fully. The dynamical interpretation of such experiments will need to go beyond ideas of wave, mean-flow interaction based on zonal averaging, which although self consistent mathematically are rather arbitrary physically, and may be inappropriate when wave amplitudes are large (Andrews *et al.* 1987, p.271). An alternative approach of general applicability (for balanced flows) is to use potential vorticity and “PV thinking,” which may be more suitable when the dynamics involve, as we have proposed here, localized vortex developments and vortex interaction, rather than wave propagation and wave, mean-flow interaction.

Our work proposing that, in the right circumstances, the development of sub planetary-scale systems in the troposphere can strongly impact the stratosphere invites a fresh look at some questions about the dynamics of SSWs of both the polar vortex spitting and vortex displacement kinds. (1) The role of tropospheric blocking in SSWs has a long history of enquiry (e.g. Quiroz 1986; Martius *et al.* 2009; Garfinkel *et al.* 2010), though no definitive mechanism has been advanced for how blocking affects SSWs. It does not seem intuitively that our ideas about the possible impact of a sub planetary-scale cyclonic PV anomaly would apply to that of an anticyclonic PV anomaly of similar scale in isolation. When the development of a blocking pattern accompanies cyclogenesis in a highly nonlinear Rossby wavetrain in the troposphere, however - for an excellent example see the period 14–21 January 1963 (maps readily plotted at ESRL (2016)) - then our ideas may prove relevant. (2) Observational and numerical modelling studies are needed to see whether major SSWs involving polar vortex displacement (but not splitting) can also be triggered by sub planetary scale cyclogenesis in the troposphere, given that the precursor state of the stratosphere again comprises an elongated polar vortex. We expect that an appropriate positional alignment between the cyclogenesis and the polar vortex will be needed in this case too.

Acknowledgements

Paul Berrisford kindly gave us his algorithm for potential vorticity inversion on a sphere, used in section 3. We thank the following for helpful discussions: Gavin Esler, Peter Hitchcock and John Methven. The paper has benefited from the reviews of two anonymous referees.

Appendix

Mathematical details of the methodology

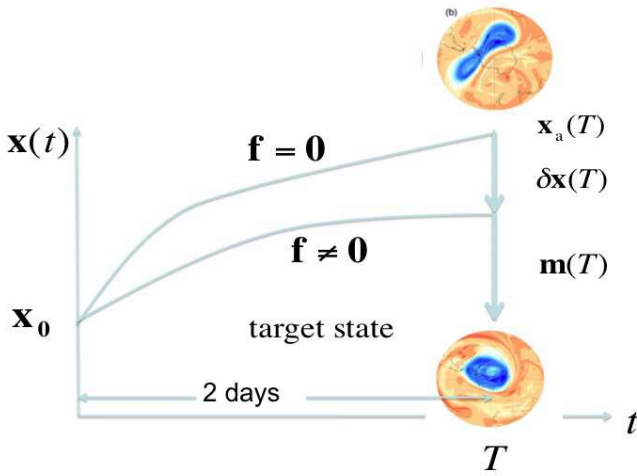


Figure 14. Schematic illustrating the calculation of the sensitivity gradient

Referring to Figure 14, let the model evolve according to the equation:

$$\frac{dx_a}{dt} = \mathcal{M}(x_a), \quad (1)$$

where x_a denotes the actual evolution of the model, its actual trajectory, as determined by the nonlinear model, \mathcal{M} .

Applying a small forcing \mathbf{f} , then the departure from the actual unperturbed trajectory, $\delta\mathbf{x}$, evolves according to:

$$\frac{d\delta\mathbf{x}}{dt} = \mathbf{M}\delta\mathbf{x} + \mathbf{f}, \quad (2)$$

where \mathbf{M} (a Jacobian matrix) is the linearisation of the nonlinear model \mathcal{M} along the actual, unperturbed, trajectory, i.e.:

$$\mathbf{M} = \left. \frac{\partial \mathcal{M}}{\partial \mathbf{x}} \right|_{\mathbf{x}=\mathbf{x}_a} \quad (3)$$

The difference, \mathbf{m} , between the target state \mathbf{x}_{tar} and the perturbed state at time T is given by:

$$\mathbf{m} = \mathbf{x}_{tar} - (\mathbf{x}_a(T) + \delta\mathbf{x}(T)) \quad (4)$$

To define a magnitude for this difference vector, \mathbf{m} , which we call the miss-hit vector, we need to define a suitable scalar product, or norm, in the vector space to which \mathbf{m} belongs. Following Buizza (1994), we use a total energy norm, which measures the difference in total energy between the target state before the vortex split and the perturbed state at time T .

The norm, $\|\mathbf{m}\|$, a measure of the miss-hit, is given by:

$$\|\mathbf{m}\| = \langle \mathbf{m}^T C_{TE} \mathbf{m} \rangle, \quad (5)$$

where $()^T$ denotes matrix transpose. The quantity C_{TE} is a matrix of constant coefficients, specified (Buizza 1994) using a climatological reference state, and the angle brackets $\langle \dots \rangle$ represent an integral over a chosen volume of the simulated atmosphere. To increase the sensitivity of the miss-hit $\|\mathbf{m}\|$ to vortex splitting, the volume of integration is restricted to a 20 km thick layer of the polar cap in the stratosphere, centred at a height of 30 km (near 10 hPa) and extending from the south pole to 60°S.

We wish to determine the sensitivity of this miss-hit $\|\mathbf{m}\|$ to the forcing \mathbf{f} - the "sensitivity gradient" - which we derive as follows:

$$\left. \frac{\partial \|\mathbf{m}\|}{\partial \mathbf{f}} \right|_{\mathbf{f}=\mathbf{0}} = \left(\left. \frac{\partial \mathbf{m}}{\partial \mathbf{f}} \right)^T \right|_{\mathbf{f}=\mathbf{0}} \left. \frac{\partial \|\mathbf{m}\|}{\partial \mathbf{m}} \right|_{\mathbf{f}=\mathbf{0}}. \quad (6)$$

The second factor on the right-hand side can be computed directly from the definition of $\|\mathbf{m}\|$ above. The first factor on the right is the transpose of a Jacobian matrix. Its computation involves integrating the adjoint of the linearised version of the model, \mathbf{M}^T , backwards in time along the actual, unforced, trajectory, $\mathbf{x}_a(t)$. The approach is standard practice in 4D-variational data assimilation (e.g. Kalnay 2002). We note again that a linearization in the adjoint model does not assume that the actual dynamics are linear. Also, the backwards integration of the model is a computational device, and does not assume that the dynamics are reversible, which they manifestly are not.

The forcing field is then set to be:

$$\mathbf{f} = -\alpha \left. \frac{\partial \|\mathbf{m}\|}{\partial \mathbf{f}} \right|_{\mathbf{f}=\mathbf{0}} \quad (7)$$

where α is a positive proportionality constant, which in principle could be chosen arbitrarily, but in practice was chosen to be small enough to satisfy the linear constraint during the optimization period. Interpreted in \mathbf{f} -space, \mathbf{f} is a multi-dimensional vector pointing locally down the gradient of the (scalar) miss-hit quantity $\|\mathbf{m}\|$ towards smaller values (closer states to the target state). Since the components of \mathbf{f} are simply numbers corresponding to different model variables (e.g. temperature) and different grid points, we can display sets of components of \mathbf{f} as fields on a map.

References

- Andrews DG, Holton JR, Leovy CB. 1987. *Middle atmosphere dynamics*. Academic Press.
- BADC. 2016. British Atmospheric Data Centre. URL <http://badc.nerc.ac.uk/home/index.html>. Accessed: 5 April 2016.
- Buizza R. 1994. Sensitivity of optimal unstable structures. *Quarterly Journal of the Royal Meteorological Society* **120**(516): 429–451.
- Charlton AJ, O'Neill A, Lahoz WA, Berrisford P. 2005. The splitting of the stratospheric polar vortex in the Southern Hemisphere September 2002: dynamical evolution. *Journal of the Atmospheric Sciences* **62**(3): 590–602.
- Charlton AJ, O'Neill A, Lahoz WA, Massacand AC. 2004. Sensitivity of tropospheric forecasts to stratospheric initial conditions. *Quarterly Journal of the Royal Meteorological Society* **130**: 171–1792.
- Charney JG, Drazin PG. 1961. Propagation of planetary-scale disturbances from the lower into the upper stratosphere. *Journal of Geophysical Research* **66**(1): 83–109.
- Dritschel DG. 1990. The stability of elliptical vortices in an external straining flow. *Journal of Fluid Mechanics* **210**: 223–261.
- Dritschel DG, Viúdez A. 2003. A balanced approach to modelling rotating stably stratified geophysical flow. *Journal of Fluid Mechanics* **488**: 123–150.
- Errico RM. 1997. What is an adjoint model? *Bulletin of the American Meteorological Society* **78**: 2578–2591.
- Esler JG, Polvani LM, Scott RK. 2006. The antarctic stratospheric sudden warming of 2002: A self-tuned resonance? *Geophysical Research Letters* **33**(12): L12 804.
- Esler JG, Scott RK. 2005. Excitation of transient rossby waves on the stratospheric polar vortex and the barotropic sudden warming. *Journal of the Atmospheric Sciences* **62**: 3661–3682.
- ESRL. 2016. Earth System Research Laboratory. URL http://www.esrl.noaa.gov/psd/data/gridded/data.ncep_reanalysis.html. Accessed: 5 April 2016.
- Fairlie TDA, Fisher M, O'Neill A. 1990. The development of narrow baroclinic zones and other small-scale structure in the stratosphere during simulated major warmings. *Quarterly Journal of the Royal Meteorological Society* **116**(492): 287–315.
- Fisher M. 1987. The Met.O.20 stratosphere-mesosphere model. dynamical climatology technical note no.52. Technical report, Meteorological Office, U.K.
- Garfinkel CI, Hartmann DL, Sassi F. 2010. Tropospheric precursors of anomalous Northern Hemisphere stratospheric polar vortices. *Journal of Climate* **23**: 3282–3299.
- Hodges KI. 1994. A general method for tracking analysis and its application to meteorological data. *Monthly Weather Review* **123**: 2573–2586.
- Holton JR, Hakim GJ. 2012. *An introduction to dynamic meteorology*. Academic Press, 5 edn.
- Hoskins BJ, McIntyre ME, Robertson AW. 1985. On the use and significance of isentropic potential vorticity maps. *Quarterly Journal of the Royal Meteorological Society* **111**: 877–946.
- Jung T, Barkmeier J. 2006. Sensitivity of the tropospheric circulation to changes in the strength of the stratospheric polar vortex. *Monthly Weather Review* **134**(8): 2191–2207.
- Kalnay E. 2002. *Atmospheric modelling, data assimilation and predictability*. Cambridge University Press.
- Kida S. 1981. Motion of an elliptic vortex in a uniform shear flow. *Journal of the Physical Society of Japan* **50**(10): 3517–3520.
- Kushner PJ, Polvani LM. 2005. A very large, spontaneous stratospheric sudden warming in a simple agcm: a prototype for the southern hemisphere warming of 2002? *Journal of the Atmospheric Sciences* **62**: 890–897.
- Martius O, Polvani LM, Davies HC. 2009. Blocking precursors to stratospheric sudden warmings. *Geophysical Research Letters* **36**(L14806).
- Mathewman NJ, Esler JG. 2011. Stratospheric sudden warmings as self-tuning resonances. part i: vortex splitting events. *Journal of the Atmospheric Sciences* **68**: 2481–2504.
- Matsuno T. 1971. A dynamical model of the stratospheric sudden warming. *Journal of Atmospheric Sciences* **28**: 1479–1494.
- Mechoso CR, O'Neill A, Pope VD, Farrara JD. 1988. A study of the stratospheric final warming of 1982 in the Southern Hemisphere. *Quarterly Journal of the Royal Meteorological Society* **114**(484): 1365–1384.
- Mitchell DM, Charlton-Perez AJ, Gray LJ. 2011. Characterizing the variability and extremes of stratospheric polar vortices using 2d moment analysis. *Journal of the Atmospheric Sciences* **68**: 1194–1213.
- Newman PA, Nash ER. 2005. The unusual southern hemisphere stratosphere winter of 2002. *Journal of the Atmospheric Sciences* **62**(3): 614–628.
- Oatley CL. 2010. The impact of synoptic-scale development in the troposphere on the planetary-scale polar vortex in the stratosphere: a case study of the unprecedented southern hemisphere major warming event of 2002. PhD thesis, University of Reading.
- O'Neill A. 2003. Stratospheric sudden warmings. *Encyclopedia of Atmospheric Sciences*, Eds. Holton, J. R. and Pyle, J. A. and Curry, J. A., Elsevier : 1342–1353.
- O'Neill A, Pope VD. 1988. Simulations of linear and nonlinear disturbances in the stratosphere. *Quarterly Journal of the Royal Meteorological Society* **114**(482): 1063–1110.
- Plumb RA. 1981. Instability of the distorted polar vortex: a theory of stratospheric warmings. *Journal of the Atmospheric Sciences* **38**: 2514–2531.
- Quiroz RS. 1986. The association of stratospheric warmings with tropospheric blocking. *Journal of Geophysical Research* **91**(D4): 5277–5285.
- Reinaud JN, Dritschel DG, Koudella CR. 2003. The shape of vortices in quasi-geostrophic turbulence. *Journal of Fluid Mechanics* **474**: 175–192.
- Scott RK, Dritschel DG. 2005. Quasi-geostrophic vortices in compressible atmospheres. *Journal of Fluid Mechanics* **530**: 305–325.
- Simmons A, Hortal M, Kelly G, McNally A, Untsch A, Uppala S. 2005. Ecmwf analyses and forecasts of the stratospheric polar vortex breakup: September 2002 in the Southern Hemisphere and related events. *Journal of the Atmospheric Sciences* **62**: 668–689.
- Thorncroft CD, Hoskins BJ, McIntyre ME. 1993. Two paradigms for baroclinic-wave life-cycle behaviour. *Quarterly Journal of the Royal Meteorological Society* **119**: 17–55.
- Waugh DN. 1997. Elliptical diagnostics of stratospheric polar vortices. *Quarterly Journal of the Royal Meteorological Society* **123**(542): 1725–1748.