Vortex Preconditioning due to Planetary and Gravity Waves prior to Sudden Stratospheric Warmings

JOHN R. ALBERS* AND THOMAS BIRNER

Colorado State University, Fort Collins, Colorado

(Manuscript received 7 February 2014, in final form 4 June 2014)

ABSTRACT

Reanalysis data are used to evaluate the evolution of polar vortex geometry, planetary wave drag, and gravity wave drag prior to split versus displacement sudden stratospheric warmings (SSWs). A composite analysis that extends upward to the lower mesosphere reveals that split SSWs are characterized by a transition from a wide, funnel-shaped vortex that is anomalously strong to a vortex that is constrained about the pole and has little vertical tilt. In contrast, displacement SSWs are characterized by a wide, funnel-shaped vortex that is anomalously weak throughout the prewarming period. Moreover, during split SSWs, gravity wave drag is enhanced in the polar night jet, while planetary wave drag is enhanced within the extratropical surf zone. During displacement SSWs, gravity wave drag is anomalously weak throughout the extratropical stratosphere.

Using the composite analysis as a guide, a case study of the 2009 SSW is conducted in order to evaluate the roles of planetary and gravity waves for preconditioning the polar vortex in terms of two SSW-triggering scenarios: anomalous planetary wave forcing from the troposphere and resonance due to either internal or external Rossby waves. The results support the view that split SSWs are caused by resonance rather than anomalously large wave forcing. Given these findings, it is suggested that vortex preconditioning, which is traditionally defined in terms of vortex geometries that increase poleward wave focusing, may be better described by wave events (planetary and/or gravity) that “tune” the geometry of the vortex toward its resonant excitation points.

1. Introduction

Sudden stratospheric warmings (SSWs) are midwinter events in which the polar cap temperature and circulation undergo large and abrupt changes (McInturff 1978) that regularly extend downward from the stratosphere to affect tropospheric weather and climate. It is because of this potential to perturb weather and climate over a deep layer of the atmosphere that makes the ability to predict how and under what circumstances SSWs occur a question of paramount importance. Unfortunately, answers to two central questions regarding the dynamics of SSWs remain elusive: namely, what types of wave phenomena are responsible for the explosive wave-amplitude growth that triggers an SSW and what does the evolution of the basic state look like prior to an SSW?

The predominant mechanism cited to answer the first question is that SSWs are triggered by the anomalous injection of wave activity from the troposphere. In this view, the stratosphere simply needs to be supplied with an adequate amount of wave forcing in order to trigger an SSW via wave–mean flow interaction (Matsuno 1971; Limpasuvan et al. 2004; Manney et al. 2009; Nishii et al. 2009; Ayarzagüena et al. 2011; Kuttippurath and Nikulin 2012). Indeed, many studies are able to find a high correlation between SSWs and various tropospheric phenomena that are favorable for generating wave activity, including tropospheric blocking, ENSO, and storm-track shifts (e.g., Martius et al. 2009; Garfinkel et al. 2010; Cohen and Jones 2011). Yet when the reverse question is posed, very little if any predictive power is realized. For example, while SSWs are seemingly well correlated with tropospheric blocking events, of the 782
blocks that occurred during the winter seasons between 1957 and 2001, only 52 occurred prior to an SSW (Martius et al. 2009). Results such as this call into question whether the various anomalous tropospheric phenomena are hallmarks of SSW triggers themselves or, conversely, whether these phenomena simply represent basic-state configurations that are broadly favorable for SSW development.

In contrast to the concept of anomalous wave-triggered SSWs, recent theoretical SSW studies have suggested that anomalous tropospheric upward wave fluxes are not necessary for triggering an SSW (Esler and Scott 2005; Scott and Polvani 2006; Matthewman and Esler 2011; Esler and Matthewman 2011). These studies suggest that the explosive growth of wave amplitude that occurs during an SSW is instead representative of nonlinear resonance that is triggered by the modulation of key geometric vortex parameters that are defined based on whether a split or displacement-type warming (Charlton and Polvani 2007) is under consideration. Viewed from this perspective, the anomalous stratospheric wave fluxes are a symptom of vortex breakdown rather than a trigger of the event itself. If this scenario is accurate, then the troposphere need only supply a nominal (potentially climatological in the Northern Hemisphere) amount of wave activity and it is the state of the stratosphere that determines whether an SSW occurs. This suggestion is supported by the recent observational analysis of T. Birner and J. Albers (2014, unpublished manuscript), who found that SSW-like events are not preceded by anomalous tropospheric upward wave-activity fluxes but rather that the anomalous stratospheric wave fluxes appear nearly instantaneously at all heights within the interior of the stratosphere at the onset of an SSW.

Determining which scenario—anomalous tropospheric wave forcing versus resonance—is responsible for triggering SSWs is even more important because it affects how we interpret the necessary basic-state conditions prior to an SSW (i.e., vortex preconditioning). If anomalous tropospheric wave forcing triggers SSWs, then vortex preconditioning is simply a vehicle for focusing a sufficient amount of wave activity into the polar upper stratosphere to begin the critical-layer cascade (Matsuno 1971). In contrast, if SSWs are triggered by resonance, then preconditioning needs to be interpreted in terms of the basic-state constraints needed to “tune” the vortex into a geometric structure that supports resonance (Clark 1974).

Unfortunately, while the basic dynamical principles associated with the contrasting theories of SSW triggers (anomalous wave forcing versus resonance) and vortex preconditioning (wave focusing versus resonant cavity formation) were insightfully summarized by McIntyre (1982), there is still no consensus regarding which of the theories most faithfully represents the actual phenomena. For example, several recent studies cite distinctly different basic-state configurations as being representative of vortex preconditioning. Limpasuvan et al. (2004) state that preconditioning is characterized by anomalously weak winds equatorward of 60° and anomalously strong winds poleward of 70° for all warming types. Likewise, Nishii et al. (2009) and Kutippurath and Nikulin (2012) suggest that both split- and displacement-type SSWs are characterized by precursor wave pulses that weaken and constrain the vortex about the pole, while Ayarzagüena et al. (2011) define a preconditioned vortex to be anomalously weak but state that two recent split and displacement SSWs (2009 and 2010, respectively) did not exhibit this feature at all. Although there are some minor differences among each of the studies just listed, they all have one underlying commonality: preconditioning is presumed to work in the same way for all SSW types because it is assumed that all SSWs are triggered via anomalous wave forcing.

In contrast, Charlton and Polvani (2007) and Bancala et al. (2012) provide evidence that split and displacement SSWs exhibit distinct prewarming evolution morphologies, a viewpoint that closely conforms to the “type 1–type 2” and “type A–type B” SSW distinctions first suggested by Quir oz et al. (1975) and Schoeberl (1978), respectively. In agreement with these studies, our analysis will show that split and displacement SSWs indeed have very distinct prewarming evolutions. However, in contrast to Charlton and Polvani (2007) and Bancala et al. (2012), who focused on planetary waves in the region below 30 km (10 hPa), we extend our analysis upward to 55 km (0.5 hPa) and analyze the combined effects of both planetary waves and gravity waves on vortex preconditioning and the resonant excitation theory of SSWs. In doing so, we focus particular attention on split-type SSWs for two reasons. First, the notions of precursor wave events and vortex preconditioning are observed to occur largely for the case of split SSWs (Labitzke 1977; Schoeberl 1978; Labitzke 1981; McIntyre 1982; Charlton and Polvani 2007); thus, the analysis of split SSWs provides a natural starting point for comparing the recent nonlinear resonance ideas of Matthewman and Esler (2011) and Esler and Matthewman (2011) to observational/reanalysis data. Second, from a practical standpoint, the dynamics of resonance due to the first baroclinic mode that is thought to be responsible for displacement SSWs is significantly more complicated (Esler and Matthewman 2011); thus, we defer detailed analysis of displacement SSWs in reanalysis data to a future study.
Our results provide the following insights into the dynamics of SSWs. First, we show that in a composite sense, the prewarming evolution of displacement and split SSWs are characterized by unique vertical structures that extend upward into at least the lower mesosphere. Second, we show that in the preconditioning phase of SSWs, both planetary waves and gravity waves play an important role in the geometric evolution of the vortex and that the majority of the combined wave forcing occurs in the region above approximately 30 km (~10 hPa). These two results highlight the need to analyze the wave and vortical structures of SSWs throughout the entire depth of the stratosphere instead of focusing on the region below 30 km, as has been done in many previous SSW climatologies (Limpasuvan et al. 2004; Charlton and Polvani 2007; Bancala et al. 2012).

While our composite analysis provides a broad view of the geometric evolution of SSWs, the averaging procedures inherently obscure a detailed view of the wave–vortex evolution necessary to evaluate the signatures of resonance. To address this, we conduct a case study of the well-studied 2009 split SSW where we use our composite analysis as a guide in order to help verify that the 2009 warming evolution is representative of split SSWs in general.

To determine whether the 2009 SSW was triggered via anomalous wave activity versus resonance, we take advantage of the fact that the two types of wave phenomena have unique vertical Eliassen–Palm (EP) flux signatures. Namely, if the SSW is triggered via the anomalous injection of wave activity from the troposphere, then large EP flux pulses should be traceable from the troposphere into the upper stratosphere at the standard group velocity for vertically propagating planetary waves. In contrast, if the SSW is triggered via resonance, then the vertical EP fluxes should appear nearly instantaneously at all heights. Our results reveal that while the 2009 prewarming period was characterized by vertically propagating waves, the split SSW itself was characterized by the signature of resonance. This result is in qualitative agreement with previous observational studies that showed that split-SSW prewarming periods are characterized by planetary waves with significant westward tilt, while the warmings themselves are characterized by wave events with little to no vertical phase tilt (Kanzawa 1980, 1982; Naujokat et al. 2002).

Given our analysis of resonantly triggered split SSWs, we discuss how the basic dynamical precepts of wave-guide formation, vortex erosion, and wave focusing originally outlined by McIntyre (1982)—and henceforth used by many researchers to support the notion of SSWs triggered via anomalous wave activity—can be equally and perhaps more appropriately applied to the formation of vortex geometries that favor resonance. In doing so, we show that, in addition to breaking planetary waves in the surf zone, gravity waves also play an important role in vortex preconditioning. Finally, we suggest how the spatial and temporal pattern of orographic gravity wave drag (GWD), which is robust among two reanalysis datasets, may reveal a potential pathway for gravity waves to provide additional wave forcing that helps to trigger resonance itself.

2. Reanalysis data and vortex classification

In the analysis that follows, we primarily use data from the Japanese Meteorological Agency and Central Research Institute of Electrical Power Industry 25-year Reanalysis (JRA-25) project (Onogi et al. 2007; Japan Meteorological Agency and Central Research Institute of Electric Power Industry 2012). The JRA-25 data used in this study span the years 1980–2011 and were calculated using T106L40 grid resolution with a model top at 0.4 hPa. The dominant contribution to the JRA-25 GWD in the stratosphere comes from orographic waves with wavelengths longer than 100 km and is based on a modified version of the Palmer et al. (1986) scheme. The JRA-25 data do not account for nonorographic gravity waves. These waves may indeed represent an important source of polar vortex variability as they make an important contribution to the total wave forcing above about 40 km (Scinocca 2003; McLandress and Scinocca 2005; Orr et al. 2010) and thus represent an important avenue for future research.

To verify the robustness of the space–time patterns of GWD in JRA-25, we also examined the Interim European Centre for Medium-Range Weather Forecasts (ECMWF) Re-Analysis (ERA-Interim) GWD (Dee et al. 2011). However, only the full zonal momentum physics tendency is available from ERA-Interim. Fortunately, at the altitude range of interest here, this physics tendency can be considered equal to the GWD. The ERA-Interim GWD consists of an orographic parameterization that is based on the Baines and Palmer (1990) scheme [see also Lott and Miller (1997) and IFS (2006)]. The differences in GWD between the two reanalyses are largely confined to differences in the overall magnitude of drag. For example, while GWD constitutes roughly 30% of the total wave forcing (resolved plus parameterized) in large areas between 30 and 50 km and poleward of 40° in the JRA-25 data, GWD consisted of about 10% of the total wave forcing in the same regions in ERA-Interim. Nevertheless, because the GWD patterns are extremely localized in their longitudinal extent, it is likely that zonal-mean gravity wave forcing anywhere in the 10%–30% range
represents an important wave forcing in specific regions along the vortex edge.

For the years 1980–2002, we differentiate between split and displacement SSWs based on the vortex classification of Charlton and Polvani (2007). For the years 2003–10, we differentiate split versus displacements based on Manney et al. (2009), Thurairajah et al. (2010), and Kuttippurath and Nikulin (2012) (Table 1). However, while we separate split and displacement events in accordance with the findings of Charlton and Polvani (2007), our definition of climatology, and therefore anomalies from climatology, is slightly different.

In JRA-25, an SSW occurred in 18 of the total 32 years of available data. Thus, if all of the SSW years were included in the climatology calculation, then more than 50% of the years included in that climatology would include an SSW. Moreover, because stratospheric radiative relaxation time scales are relatively long and vortex preconditioning may begin as far as 20–50 days in advance of the central warming date (Baldwin and Holton 1988), then with the possible exception of SSWs that occur in late February or March, virtually any winter that undergoes an SSW will introduce the vortex variability that we wish to isolate into the climatology to which we are comparing the events. We therefore construct our climatology and anomaly time series as described below.

First, we only include years in our climatology if no SSW occurred during that winter. The climatology therefore consists of 14 years of data averaged into one time series that spans October–March in 6-hourly increments. To create the anomalies from climatology, we subtract the time period before each SSW from the climatology for those same, seasonally specific days in the time series. For example, for the SSW that occurred on 24 February 1984, we subtract the days before the central warming date in the 1984 time series from the same days before 24 February in the climatology time series. This step helps account for seasonal variations in the vortex. We perform this procedure for each of the SSW events and then bin each anomaly time series according to whether the SSW was a split or displacement event. Finally, each set of split and displacement anomalies is averaged to create three separate time series: one for all SSWs, one for vortex splits, and one for vortex displacements.

We note that several of our plots show the uppermost model layer generated by each reanalysis; this was done so that readers can get a complete view of the entire reanalysis domain. However, we caution that the uppermost layer may potentially contain unphysical values due to model upper-boundary issues; this may be particularly true for the wave fluxes. To address this, we do not include values from the uppermost model layer in any of our calculations that involve the column-integrated GWD values (i.e., Figs. 6 and 7) or comparison of the relative importance of GWD versus EPFD.

3. Composite analysis

a. Polar vortex geometry of split and displacement SSWs

Displacement-type SSWs occur with roughly equal frequency throughout the winter and early spring months (December–March); split SSWs, on the other hand, occur almost exclusively during January and February (Charlton and Polvani 2007, their Fig. 2). The strong peak in split events during midwinter suggests that we look closer at the geometry of the vortex during January and February versus the surrounding months. Beginning in the fall season, the vortex forms as a funnel with a wide top and a narrow bottom. Then, during late December, the area of the upper-stratospheric and lower-mesospheric portion of the vortex decreases in size due in large part to breaking planetary waves on the periphery of the vortex (McIntyre and Palmer 1983, 1984). Because the lower-stratospheric portion of the vortex remains largely unchanged, the vortex becomes more vertically aligned in height (Harvey et al. 2002, their Fig. 11). The obvious question then is: does a vortex that is smaller in area and more invariant with height

<table>
<thead>
<tr>
<th>Date</th>
<th>Type</th>
</tr>
</thead>
<tbody>
<tr>
<td>29 Feb 1980</td>
<td>D</td>
</tr>
<tr>
<td>4 Dec 1981</td>
<td>D</td>
</tr>
<tr>
<td>4 Feb 1984</td>
<td>D</td>
</tr>
<tr>
<td>1 Jan 1985</td>
<td>S</td>
</tr>
<tr>
<td>23 Jan 1987</td>
<td>D</td>
</tr>
<tr>
<td>8 Dec 1987</td>
<td>S</td>
</tr>
<tr>
<td>14 Mar 1988</td>
<td>S</td>
</tr>
<tr>
<td>22 Feb 1989</td>
<td>S</td>
</tr>
<tr>
<td>15 Dec 1998</td>
<td>D</td>
</tr>
<tr>
<td>25 Feb 1999</td>
<td>D</td>
</tr>
<tr>
<td>20 Mar 2000</td>
<td>D</td>
</tr>
<tr>
<td>11 Feb 2001</td>
<td>D</td>
</tr>
<tr>
<td>30 Dec 2001</td>
<td>D</td>
</tr>
<tr>
<td>17 Feb 2002</td>
<td>D</td>
</tr>
<tr>
<td>18 Jan 2003</td>
<td>D</td>
</tr>
<tr>
<td>4 Jan 2004</td>
<td>D</td>
</tr>
<tr>
<td>21 Jan 2006</td>
<td>D</td>
</tr>
<tr>
<td>24 Feb 2007</td>
<td>D</td>
</tr>
<tr>
<td>22 Feb 2008</td>
<td>D</td>
</tr>
<tr>
<td>24 Jan 2009</td>
<td>S</td>
</tr>
<tr>
<td>9 Feb 2010</td>
<td>D</td>
</tr>
</tbody>
</table>
represent a geometric basic state that is more conducive to triggering a split SSW? With this question in mind, we consider the differences in vortex geometry for split versus displacement SSWs.

Figure 1 shows the composite zonal-mean zonal wind anomalies for displacement and split SSWs 10–20 and 5–10 days prior to the central warming date. Winds prior to displacement SSWs are anomalously weak between approximately 50° and 90° below 30 km and between approximately 30° and 90° above 30 km. This anomaly pattern is characteristic of a weak and broad vortex that is funnel shaped in height. While the strength of the vortex gets progressively weaker as the central warming date is approached, the funnel-shaped vertical structure
stays relatively constant throughout the prewarming period.

In contrast, the time evolution of the split SSW vortex geometry varies in both strength and horizontal and vertical structure. In the early prewarming period (Fig. 1c), the split-SSW wind profile is anomalously strong poleward of about 70° below 20 km and poleward of about 45° above 20 km. Equatorward of these same latitudes, the winds are anomalously weak. Thus, at the early stages of vortex development, the split-type vortex anomaly has a wide funnel-shaped geometry similar to the displacement anomaly but opposite in sign. As the central warming date approaches, however, the split vortex anomaly pattern becomes more vertically aligned as the upper-stratospheric and lower-mesospheric portion of the vortex decreases in area. This pattern represents a vortex that transitions from a strong, wide, and funnel-shaped vortex to a more narrow and vertically aligned vortex. In addition, as the central warming date is approached in the split SSW wind composite, the poleward-side positive wind anomalies get weaker (shown in red in Figs. 1c and 1d).

b. EP flux divergence and GWD for split and displacement SSWs

Before comparing the EPFD and GWD anomalies for split and displacement SSWs, it is instructive to understand the relative contributions of each wave type to the total wave forcing budget. Figure 2 shows the percentage of the total drag (EPFD plus GWD) that orographic gravity waves contribute in the 5–30 days prior to the central warming date for all SSWs in the JRA-25 data. Poleward of 45° and above 30 km, GWD constitutes between 15% and 80% of the total accumulated wave drag. In particular, in the bulk of the region between about 45° and 75° and 30 and 55 km (the region of particular interest for SSW dynamics), GWD contributes greater than 30% of the total forcing in the JRA-25 data and greater than 10% of the total forcing in the ERA-Interim data (see section 2). This is perhaps a rather surprising result given the relative lack of attention that gravity waves have received in the SSW literature, although there are a few notable exceptions (Dunkerton and Butchart 1984; Rind et al. 1988; Lawrence 1997; Pawson 1997; Limpasuvan et al. 2007; Birner and Williams 2008; Wang and Alexander 2009; Flury et al. 2010; Thurairajah et al. 2010; Yamashita et al. 2010). It should be noted that a very similar pattern also emerges when the percentage of GWD is computed for climatology.

Figures 3a, 3b, 4a, and 4b show the composite EPFD and GWD anomalies, respectively, for displacement SSWs 10–20 and 5–10 days prior to the central warming date. The EPFD anomalies in Figs. 3a and 3b reflect the patterns that one would expect given planetary wave propagation constraints (Charney and Drazin 1961). In particular, planetary wave propagation is directly
proportional to the meridional gradient of potential vorticity (PV) and inversely proportional to the magnitude of the zonal-mean zonal wind. While the planetary waveguide in the upper stratosphere is largely governed by the strong meridional PV gradient along the equatorward edge of the vertical axis of the polar vortex (Matsuno 1970; Albers et al. 2013), weakening of the zonal-mean wind will itself increase vertical wave propagation. Indeed, as the zonal wind anomaly becomes more strongly negative and extends farther downward, the EPFD anomaly also widens and extends downward (cf. Figs. 1a,b and 3a,b).

Likewise, the displacement SSW GWD anomalies appear to mirror what one would expect given wave
propagation constraints. Though in contrast to planetary waves, gravity wave propagation is enhanced where the zonal wind is strongest within the core of the polar night jet and suppressed in the regions of weak wind on either side of the jet core (Duck et al. 1998, 2001; Wang and Alexander 2009). This phenomenon is reflected in the suppressed GWD in the region of weakened zonal mean wind (cf. Figs. 1a,b and 4a,b). The overall relationship between the EPFD, GWD, and zonal-mean wind appear to evolve in a rather straightforward manner. As the central warming date is approached, the region of EPFD expands poleward and downward in a pattern that is a known hallmark of downward-propagating zonal-mean wind anomalies due to local wave–mean flow
interaction (Kodera et al. 2000; Plumb and Semeniuk 2003; Perlwitz and Harnik 2004). As a result of the weak wind anomalies, GWD is suppressed throughout the prewarming period.

In contrast to the displacement anomalies, the wave fields during split SSWs show a very different pattern. In the period 10–20 days before the central warming date of the split composite, the EPFD is anomalously weak poleward of about 55° and anomalously strong equatorward of about 55° (Fig. 3c). The GWD anomaly has an analogous spatial pattern to the EPFD but opposite in sign (Fig. 4c). In a similar fashion to the displacement wave anomalies, each of the split-SSW wave drag anomalies can be related to wave propagation conditions given the zonal-mean wind anomaly (Fig. 1c). Namely, the EPFD is anomalously large in regions of anomalously weak winds and is anomalously weak in the region where the wind is anomalously strong (and vice versa for the GWD). The ensuing wave–mean flow evolution, however, unfolds very differently for the split SSW composite. As pointed out in section 3a, the split SSW composite zonal-mean wind anomaly becomes weaker and becomes more constrained about the pole as the central warming date approaches (Figs. 1c,d). This wind evolution is associated with both types of wave drag.

By 5–10 days before the central warming date, the zonal-mean wind anomaly has very little vertical tilt throughout the stratosphere and lower mesosphere and is anomalously strong poleward of 60° (Fig. 1d). The strengthened wind in the polar region is reflected in both the EPFD and GWD anomalies that are anomalously weak and strong (Figs. 3d and 4d, respectively) throughout the polar region above 30 km. The wave drag anomalies shown in Figs. 3 and 4 clearly show that both planetary- and gravity-scale waves are acting on the polar vortex in the days and weeks prior to split-type SSWs. This prompts us to consider how each of the two wave types contributes to decreasing the area of the vortex in the upper stratosphere and aligning the vortex from a funnel shape to a more vertically aligned vortex. In doing so, we relate the prewarming evolution of the vortex geometry to two different dynamical scenarios that have been proposed as triggers for SSWs (see the introduction).

4. Vortex preconditioning and SSW-triggering scenarios

The key to interpreting our results hinges on understanding how planetary and gravity wave breaking may contribute to producing the changes in vortex geometry shown in Fig. 1 and how those changes are related to the formation of strong meridional PV gradients and critical lines involved in the two SSW-triggering scenarios outlined in the introduction. Because it is difficult to envision exactly how the two wave types act in concert to produce the prewarming vortex evolution, we first consider how idealized planetary wave and gravity wave breaking events individually and collectively affect the zonal-mean wind and meridional vorticity distributions. We follow with a brief overview of resonance and an outline of wave and vortex characteristics that allow us to differentiate between an SSW triggered by resonance versus the anomalous injection of wave activity from the troposphere. Readers who are sufficiently familiar with the tenets of preconditioning and external and internal mode resonance can skip to section 5.

a. Vortex preconditioning

Although planetary wave breaking is fundamentally governed by PV dynamics, the barotropic vorticity equation retains enough of the necessary physics required in order to provide insight into the effects of wave breaking on the polar vortex (Juckes and McIntyre 1987). Thus, in a similar manner to McIntyre (1982), we use barotropic vorticity inversion as a qualitative guide for interpreting and relating the wind and wave drag evolutions from our composite analysis to the 2009 SSW.

We begin by assuming that the stratosphere’s “climatological” basic state prior to an SSW consists of a reasonably undisturbed polar vortex with a broad jet maximum located near 55° (e.g., Labitzke 1981, their Fig. 4c). Because such a vortex does not have a sharp “edge,” it will have a PV gradient with a reasonably uniform slope between the subtropics and polar regions. This situation is depicted in Figs. 5a and 5b, which show initial idealized climatological absolute vorticity and zonal-mean wind profiles (dashed lines), where the vorticity profile was calculated from the wind profile. The lack of a sharp peak in the meridional PV gradient means that there will not be a strong planetary wave guide into the polar regions, and owing to the curvature of Earth, planetary waves will tend to bend equatorward as they propagate upward from the troposphere into the mid-to lower stratosphere (e.g., Dunkerton et al. 1981; Hoskins and Karoly 1981). The traditional notion of vortex preconditioning is then initiated via the breaking of a planetary wave in the extratropical surf zone located on the outer periphery of the vortex.

When a planetary wave breaks in the surf zone, considerable quasi-horizontal mixing occurs whereby regions of high- and low-PV air are mixed southward and northward, respectively. Figures 5a and 5b show wind and absolute vorticity profiles before and after an idealized planetary wave breaking event (the procedure
used to calculate the wind and vorticity perturbations is described in the appendix). While the meridional PV gradient on the outer vortex periphery becomes “smeared out” by the mixing, the high-PV air in the vortex core remains largely undisturbed. The net result is a sharpening of the vortex edge—characterized by an increase in the meridional PV gradient—and a poleward shift in the location of the jet core and PV gradient maximum. In combination, the change in the strength and location of the jet and PV gradient maximum has three important effects on the guiding of subsequent planetary waves.

First, planetary waves are guided along the vertical axis of the polar vortex where the meridional PV gradient is maximized (Palmer 1981, 1982; Karoly and Hoskins 1982). Thus, a precursor wave pulse that sharpens the vortex edge should help to preferentially guide waves upward into the interior of the stratosphere rather than allowing them to immediately bend equatorward as they enter the lower stratosphere. Second, as the pole is approached, the PV gradient term tends to be dominated by the wavenumber geometry term in the refractive index [i.e., the first versus second terms on the rhs of Eq. (3), respectively]. Under climatological conditions, this defocusing effect (McIntyre 1982) generally prevents waves from propagating poleward and breaking at high latitudes. However, a precursor wave event that shifts the polar night jet, and the sharpened PV gradient, toward the pole will generally counteract this balance and help to focus wave activity poleward. Third, as the vortex is constrained about the pole, its size and moment of inertia also decrease; this in turn increases the ability of a wave event of a given magnitude to cause large disruptions of the vortex (i.e., an SSW). Thus, in sum total, the precursor wave event will modify the basic-state circulation in such a way as to sharply focus subsequent wave pulses poleward into a smaller and less massive polar vortex.

Several additional important features regarding the wind response become apparent when gravity wave breaking perturbations are taken into account. Figures 5c and 5d show wind and absolute vorticity profiles before and after an idealized gravity wave breaking event, while Fig. 5f shows the initial wind profile (dashed line) and the wind profile that results from the combined effects of the planetary wave and gravity wave perturbations (solid line). The wind profile for the planetary wave breaking perturbation alone is also plotted for reference (dashed–dotted line). The vorticity mixing perturbation due to planetary wave breaking both
increases the speed of the jet maximum and shifts the maximum poleward. When the gravity wave perturbation is taken into account, the maximum wind speed of the jet is decreased and the jet is shifted farther poleward. What this means is that as the combined wave drag due to the two wave types accumulates as the central warming date approaches, we expect (i) the wave drag from both wave types to systematically constrain the vortex about the pole and sharpen the meridional PV gradient and (ii) the vortex wind speed poleward of the surf zone to be determined by the relative mixture of the wave drag from each wave type, with planetary wave breaking accelerating the jet and GWD decelerating the jet.

In a broad sense, the vortex evolution just outlined is apparent in the composite split-SSW zonal wind anomalies shown earlier. Specifically, Figs. 1c and 1d show that as the central warming date is approached for split SSWs, both the vortex becomes more constrained about the pole and the wind speed anomaly poleward of the surf zone becomes weaker. As described above, the decrease in vortex area could be due to wave breaking from either wave type. The decrease in wind speed poleward of the surf zone, however, is consistent with GWD rather than planetary wave breaking (at least within the context of simplified barotropic dynamics).

The vortex preconditioning concepts just described provide one particularly effective way to achieve the large planetary wave driving that is necessary to trigger an SSW via the anomalous tropospheric wave activity scenario outlined in the introduction. However, the formation of a sharp vertical waveguide also happens to be a vortex geometry that is favorable for resonance; we outline the basics of this principle next.

b. External- and internal-mode Rossby wave resonance

Triggering an SSW via resonance was suggested at least as far back as Clark (1974), and has been discussed by numerous authors since that time (Tung and Lindzen 1979a,b; Plumb 1981; Haynes 1985; Smith and Avery 1987; Smith 1989; Esler and Scott 2005; Matthewman and Esler 2011; Esler and Matthewman 2011). In a broad sense, there are two types of resonance that are relevant for SSWs. The first type involves the resonant interaction between an external (barotropic) Rossby wave normal mode and a quasi-stationary Rossby wave; the second type involves the interaction between a traveling internal Rossby wave normal mode (Charney–Drazin spectrum) and a quasi-stationary wave. While both types of resonance hinge on nonlinear wave amplitude growth to trigger an SSW, they each require slightly different vortex configurations owing to the fact that the first type involves an edge wave that does not propagate vertically (the external mode), while the second type involves a wave that does propagate vertically (the internal mode).

External Rossby waves can arise under two circumstances that can be differentiated via what supplies their horizontal nodal structure. The first circumstance is a local external mode where a “perfect” vertical waveguide supplies the horizontal nodal structure. This is the idea envisioned by Matthewman and Esler (2011), who modeled such a waveguide by employing a sharp stepwise increase in PV between the surf zone and the interior of the polar vortex. The second type of external Rossby wave arises as a global normal mode where Earth’s poles provide the horizontal nodal structure. In the text below, we briefly describe the properties of global external normal modes but note that a local external mode has similar characteristics.

For an isothermal atmosphere in the absence of wave damping and in uniform rotation, Laplace’s tidal equation yields global Rossby wave solutions that are external, barotropic modes (Hough modes). While these modes have exponential amplitude growth with height due to density effects, they have zero vertical phase tilt and thus their energy density decays exponentially away from Earth’s surface. These modes therefore have the vertical structure of an edge wave and are analogous to Lamb modes (Lamb 1911, 1932, 440–442; Gill 1982). Interestingly, the potential importance of these waves in the dynamics of SSWs was suggested as far back as Deland (1970).

Introducing wave damping and nonuniformities in wind and temperature modifies important aspects of the structure of the external Rossby waves including depressing, shifting, and broadening the wave response (Salby 1981a,b, 1984). While the energy of external Rossby waves is largely contained within the first three to four scale heights of the atmosphere, the exponential growth of the wave amplitude with height means that damping and nonuniformities in the mean fields above the troposphere can have important consequences for how these waves manifest themselves in the stratosphere. In particular, locally increasing the meridional PV gradient, or more generally the refractive index, can lead to a local amplification of the amplitude of a global normal mode (Geisler and Dickinson 1976; Salby 1981a).

In contrast to an SSW involving the resonant excitation of external Rossby waves, internal Rossby waves readily radiate their energy into the upper stratosphere and mesosphere under climatological conditions. Because damping effects are strong in the upper stratosphere, the radiation of energy makes it difficult to
achieve wave amplitudes in the interior of the stratosphere that are sufficient to trigger an SSW. Thus, an SSW associated with internal-mode resonance has the additional requirement that a three-sided wave cavity must be present so that the wave energy is trapped in the interior of the stratosphere. Such a cavity can be identified via the refractive index squared (Matsumo 1970) and would ideally have two vertically oriented critical lines—one in the midlatitudes and one in the polar regions—and a third critical line extending horizontally across the upper stratosphere. In general, the vertical midlatitude critical line and the horizontal critical line must be produced via wave–mean flow interactions, while the polar critical line is always present due to the \( k^2 = s^2/a \cos^2 \phi \) term in the refractive index [Eq. (3)]. While a perfect wave cavity is unlikely to be realized in the real atmosphere, one can nonetheless expect stronger resonant responses as the cavity becomes less “leaky.”

While internal- and external-mode resonances have slightly different vortex geometry requirements, they nevertheless exhibit vertical EP flux signatures that are nearly identical. That is, under both circumstances, the vertical EP fluxes should have virtually no vertical phase tilt. The lack of vertical tilt for an external wave is because it is an edge wave (described above). For internal-mode resonance, there should be minimal vertical tilt owing to the fact that reflection of the wave inside the wave cavity leads to the formation of a standing wave pattern (Harnik and Lindzen 2001; Harnik 2002). While the lack of vertical phase tilt makes it difficult to differentiate between the two types of resonance, it does allow us to differentiate between an SSW triggered by resonance versus the injection of anomalous wave activity from the troposphere. It is this difference in vertical phase tilt that we exploit in the next section as we analyze the 2009 split SSW.

5. January 2009 split-type SSW

The split-type SSW that occurred on 24 January 2009 was the strongest and most-long-lived warming on record (Manney et al. 2009) and was characterized by strong gravity wave activity prior to the central warming date (Thurairajah et al. 2010; Yamashita et al. 2010; Limpasuvan et al. 2011). Given this fact, it is perhaps not surprising that Kim and Flatau (2010) found that accurately hindcasting the 2009 SSW was extremely sensitive to the way in which gravity waves were parameterized in their model. In particular, they found that some orographic GWD schemes led to the enhancement of planetary wave activity prior to the SSW, while the implementation of other schemes eliminated the SSW altogether. In light of these results, and the fact that the 2009 warming has been extensively analyzed in the literature (e.g., Manney et al. 2009; Harada et al. 2010; Coy et al. 2011; Wang et al. 2011; Ayarzaguena et al. 2011), the 2009 SSW makes an ideal case study for examining the relationship between gravity waves, planetary waves, and the two SSW-triggering scenarios outlined in the introduction.

We analyze the prewarming period that extends back about 1 month prior to the 24 January 2009 SSW central warming date. The evolution of the wave and vortex features during this period are split into 5–10-day periods, where the time windows are chosen to coincide with pulses of planetary and gravity wave activity. The effect of each of the wave drag pulses is then evaluated by considering how the overall shape of the vortex evolves, as gauged by changes in geopotential height contours in the upper stratosphere, and by how the meridional gradient of PV and the spherical form of the quasigeostrophic refractive index evolve. For the case of GWD, we have included figures generated from both JRA-25 and ERA-Interim (Figs. 6 and 7). The fact that the reanalysis datasets depict an extremely robust spatial–temporal pattern of GWD provides support for the contention that GWD plays a systematic role in the evolution of the vortex during both the prewarming and triggering phases of SSW development.

Using the quasigeostrophic approximation, the dispersion relation for slowly varying [in a Wentzel–Kramers–Brillouin (WKB) sense] planetary waves in spherical geometry is (e.g., O’Neill and Youngblut 1982)

\[
\omega = \pi k - \frac{k q_{\phi}}{a} \left[ l^2 + k^2 + \frac{f^2}{N^2} m^2 + \left( \frac{f}{2N} \right)^2 \right]^{-1}, \tag{1}
\]

where

\[
q_{\phi} = 2\Omega \cos \phi - \frac{1}{a} \left( \frac{\pi \cos \phi}{\cos \phi} \right) \frac{- f^2}{\rho} \left( \frac{\pi \rho_0}{N^2} \right) \tag{2}
\]

is the zonal-mean quasigeostrophic PV gradient; \( \omega (=-kc) \) is the wave frequency; \( c \) is the wave phase speed; \( \pi \) is the zonal-mean zonal wind; \( \rho = \rho_0 \exp(-z/H) \) is the standard density in log-pressure coordinates; \( z \) is height; \( \rho_0 \) is the sea level reference density; \( \phi \) is latitude; \( a \) is the radius of Earth; \( f \) is the Coriolis parameter; \( N \) is the buoyancy frequency; \( H \) is the scale height; \( l, k, \) and \( m \) are the meridional, zonal, and vertical wavenumbers, respectively; and subscripts refer to derivatives with respect to the given variable. Equation (1) yields the refractive index squared:
\[ n_z^2(\phi, z) = f^2 + \frac{f_0^2}{N^2} \mu_0^2 = \frac{\nabla z}{\alpha (\mu - c)} - \left( \frac{f}{2NH} \right)^2. \]  

All of the refractive index plots are for quantized zonal planetary wavenumber 2 \((s = k \alpha \cos \phi = 2)\) with a frequency of 0.04 cpd, which corresponds to a 20-day traveling wave with a phase speed of about 1.6 m s\(^{-1}\) at 60°N. The 20-day wave period was chosen to be consistent with the internal mode resonance study of Smith (1989), but we note that traveling waves with periods ranging from 4 to 25 days are known to exist during northern winter (Hirota and Hirooka 1984; Hirooka and Hirota 1985, 1989; Sassi et al. 2012). Other phase speeds were explored when creating the refractive index plots, but it did not change the qualitative nature of our results.

We begin by considering the 5-day time period between 22 and 26 December. This period was chosen because the vortex was circular and largely undisturbed (Fig. 6a), had a very weak meridional PV gradient (Fig. 8a), and thus had a rather weak waveguide from the troposphere into the interior of the extratropical stratosphere (dark gray coloring between about 55° and 65°N). The PV gradient and refractive index are averaged over the whole 5-day period to help eliminate day-to-day variance, while the geopotential height contours in Fig. 6 are plotted toward the end of the time period. A similar averaging and plotting procedure is used for the remaining time periods.

During the next 10 days (27 December–5 January), there was a weak pulse of planetary wave 1 (see Fig. 10a), which was followed by a somewhat stronger pulse.
of planetary wave 2 (see Fig. 10b). The EPFD from these two pulses (averaged between 45° and 75°) can be seen in the upper stratosphere between about 35 and 45 km. During this same time period, there was also a strong burst of GWD that constituted approximately 40% of the total wave drag averaged between 15 and 48 km and 45° and 75°. Figure 6b shows the time evolution of geopotential height contours in the upper stratosphere overlaid with height-integrated GWD; the larger regions of GWD in Fig. 6b are 2σ events, while nearly all of the GWD values in Figs. 6c and 6d exceed the 2σ threshold. The combined effects of the EPFD and GWD pulses during the first half of the time period can be clearly seen by as early as 30 December and include several features that are consistent with our discussion of vortex erosion and jet sharpening (section 4a).

For example, a comparison of the geopotential height contours for 25 (Fig. 6a) and 30 December (Fig. 6b) reveals a tightening of the contours over most of Asia and Europe. The tightening of the contours on the vortex edge is associated with the combined effects of (i) vortex erosion due to PV filamentation from the amplifying planetary wave 1 over central Asia and (ii) GWD in the core of the vortex jet (Fig. 6b). The effect of the sharpened vortex edge on planetary wave propagation can be confirmed by comparing the meridional PV gradient and refractive index averaged over the first and second time periods (Fig. 8). The change in the strength and orientation of the PV gradient reflects two structural changes to the vortex. First, there is a large increase in the strength of the PV gradient, which is related to the jet-sharpening effect of the two types of wave drag. Second, the PV
Gradients reflect the orientation of the zonal-mean-wind jet axis (not shown), which becomes much more vertically oriented by 5 January. The alignment of the jet axis in the vertical is consistent with the split-type SSW composite that revealed a transition from a wide, funnel-shaped vortex to a narrow, vertically aligned vortex (Figs. 1c,d).

The change in the structure of the wind and the PV gradient results in two changes to the refractive index. First, the increase in the strength of the PV gradient near 65° leads to an increase in the strength of the refractive index within the core of the planetary waveguide between 50° and 70°N. Second, a three-sided wave cavity has begun to form, which is bounded by a vertically oriented region of negative refractive index between 12 and 35 km and centered near 45°N and a horizontal critical line between about 50° and 70°N at 45 km.

During the next 5-day period (6–10 January) there is virtually no planetary wave activity in the stratosphere.
(see Fig. 10) and thus the rather large GWD pulse (Fig. 6c) constitutes nearly 100% of the wave drag for the period. The net result of the GWD is visible in both the geopotential height contours at the end of the period and the meridional PV gradient and refractive index. In particular, the geopotential height contours (Fig. 6c) show that the area of the polar vortex has decreased and has become tightly constrained over the pole. This fact is also reflected in that the PV gradient (Fig. 9a) has not only grown much stronger in magnitude but is nearly perfectly aligned in the vertical. Again, the progression of the PV gradient and the vertical axis of the polar night jet (not shown) into vertical alignment is representative of the vortex geometry observed in the split-SSW composite (Figs. 1c,d).

The changes in the PV gradient due to the GWD pulse are also apparent in the refractive index, which shows that the conditions for wave propagation have further
increased within the interior of the stratospheric waveguide centered around 60°N (Fig. 9b). In addition, the horizontally aligned critical line has descended to about 40 km, which is the altitude that several authors have identified as the necessary critical-layer height for efficient resonant excitation of internal wave modes (Plumb 1981; Smith 1989). While the vertically oriented region of negative refractive index between 40° and 50°N no longer extends quite as high as it did 5 days earlier, the portion between 15 and 25 km has moved poleward by about 5°. In terms of the formation of a wave cavity, it is likely that the poleward shift of the negative region of the refractive index is more important than the decrease in vertical extent because of the strong tendency for planetary waves to bend equatorward within the lower stratosphere in the absence of a vertical critical line (Dunkerton et al. 1981; McIntyre 1982). In sum total, it appears that by 10 January there is strong evidence to support the notion that a robust three-sided wave cavity has formed for vertically propagating (internal) planetary waves.

By 15 January both the meridional PV gradient (Fig. 9c) and the three-sided wave cavity (Fig. 10h) have weakened considerably. However, the apparent weakening of the cavity may be misleading because by this point in the prewarming period the vortex is highly nonzonal and thus zonal-mean diagnostics such as the refractive index become increasingly poor at describing wave propagation (Palmer and Hsu 1983; Mahlman 1969). Nevertheless, the period between 10 and 15 January is characterized by sizable GWD (Fig. 6d) and planetary wave 2 activity (Fig. 10b) and the vortex itself appears to be in the early stages of splitting (Fig. 6d).

Discussion

Our analysis of the 2009 SSW shows that the stratospheric basic state had features that would promote the triggering of an SSW via both of the scenarios outlined in the introduction. While it is beyond the scope of this paper to rigorously explore the relative merits of each of the triggering scenarios, we offer several lines of reasoning that suggest that resonance is the most likely cause of split SSWs.

For example, it is well documented that strong- and weak-polar-vortex events are well correlated with the amount of wave activity entering the stratosphere (Coy et al. 1997; Pawson and Naujokat 1999; Polvani and Waugh 2004), Sjoberg and Birner (2012) found that much of the anomalous 40-day heat fluxes associated with SSWs actually occur after the central warming date. This calls into question the assertion that anomalous integrated heat flux events trigger SSWs because any heat flux that occurs after the onset of the warming can hardly be considered to be a trigger of the event itself. Indeed, a careful examination of the time–height evolution of the 2009 SSW reveals that the polar vortex began to split in the upper stratosphere by as early as 17 January, was clearly split by 20 January at 1 hPa (not shown), and was split by 21 January at 10 hPa [see Fig. 3 of Harada et al. (2010)]. This means that the bulk of the wave forcing shown in Fig. 10c occurred after the vortex split had already begun.

The fact that there is a multiday time lag between when the vortex begins to split and when the largest wave fluxes appear seems to support the contention put forth by Matthewman and Esler (2011) that split-type SSWs are characterized by nonlinear wave growth (i.e., a wave-amplitude bifurcation) rather than anomalous tropospheric wave forcing. The key but subtle implication to this idea is that, while an SSW is indeed characterized by a large-amplitude wave-2 vortex event, the anomalously large wave fluxes themselves are the result, not the cause, of the SSW. What this perspective means in practice is that if the basic-state vortex is sufficiently tuned to its resonant excitation point, then any small increase in wave forcing or PV sharpening will cause a bifurcation in wave amplitude that then leads to the sudden and catastrophic breakdown of the vortex and consequent large poleward heat fluxes.

To address the possibility of resonance, we examine the suggestion by Esler and Scott (2005) that the vertical propagation of (internal) vortex Rossby waves should lead to wave breaking and PV filamentation, while during a split SSW, the total wave field will be dominated by wave activity due to external waves. If this assertion is true, then the vortex-sharpening events that occurred in late December and early January (Figs. 6 and 8–10) should exhibit very different vertical EP flux signatures. Namely, the vortex filamentation wave pulses should be characterized by vertical propagation time scales given by the group velocity for internal Rossby waves, while the external wave pulses should exhibit nearly barotropic vertical structures owing to their minimal westward phase tilt (Salby 1984).

Esler and Scott (2005) calculate that vertically propagating vortex Rossby waves of zonal wavenumber 2 should radiate upward with a group velocity of about 5.5 km day$^{-1}$. To check this, we overlaid four identical vectors with 5.5 km day$^{-1}$ slopes emanating from the local maximums in the wavenumber-2 vertical EP flux in the lower stratosphere (∼12 km) for the time series shown in Fig. 10b. For the two wave events in the first half of the time series, which includes the vortex filamentation event described at the beginning of section 5 (Fig. 6b), the theoretical predictions of Esler and Scott (2005) match the vertical propagation of the EP flux.
FIG. 10. Time series of (a) vertical EP flux for quantized zonal planetary wave 1 ($s = 1$), (b) vertical EP flux for quantized zonal planetary wave 2 ($s = 2$), and (c) the EPFD for all planetary waves ($s = 1, 2, \ldots$). All quantities are area-weight summed between 45° and 75°N. The units of the vertical EP fluxes are $1 \times 10^{-7}$ kg m s$^{-2}$; the units of the EP-flux convergence and GWD are $1 \times 10^2$ m s$^{-1}$ day$^{-1}$. The GWD contours range from $-0.25$ to $-3 \times 10^2$ m s$^{-1}$ day$^{-1}$ from the outermost to innermost contours.
remarkably well. For the three wave events during the SSW, however, the local maxima in the EP fluxes appear almost instantaneously at all altitudes; this fact is especially apparent for the second and third pulses on 17 and 26 January. The near verticality of the peaks in the EP flux contours is suggestive of resonance rather than what would be expected given the time scales for vertically propagating internal waves. This notion is reinforced further when consideration is given to how changes in group velocity should evolve as the PV gradient becomes sharper upon approaching the central warming date.

While strictly valid only for slowly varying plane waves, the concept of group velocity nevertheless yields a qualitative guide for determining whether vertical-propagation time scales should increase or decrease as the PV gradient sharpens. For stationary, vertically propagating waves ($\omega = 0$ and $n > 0$), the vertical component of the Rossby wave group velocity determined from Eq. (1) is

$$
\epsilon_{\text{z}} = \frac{\partial \omega}{\partial m^2} \frac{4km^2\alpha T^2}{\bar{T}_\phi N^2}.
$$

Equation (4) makes clear that the progressive sharpening of the vortex edge ($\bar{T}_\phi$ increasing) between 30 December and 15 January (Figs. 8 and 9) should decrease the vertical group velocity and thus increase the time it takes for a given wave packet to reach the upper stratosphere. That is, the group velocity vectors for the 15 and 20 January wave-2 events should actually be more horizontally oriented than the vectors for the 15 and 20 December events (Fig. 10b). This means that the vertical EP flux pulses for 15 January onward are even more poorly matched with the group velocity vectors than is shown in Fig. 10b. An interpretation of the SSW solely in terms of vertically propagating wave fluxes is therefore even more unlikely. This provides further support for the view that the December wave events have the signature of vertically propagating Rossby waves, while the SSW wave events are indicative of resonance.

While the results above indicate that the 2009 SSW has the signature of resonance, it is difficult to determine which type of resonance—internal or external mode—is more likely. Nevertheless, as pointed out by several authors (e.g., Smith 1989; Matthewman and Esler 2011), the presence of wave damping makes internal-mode resonance a relatively inefficient mechanism for triggering an SSW. This leads us to consider how gravity and planetary waves might combine to trigger an SSW via external (barotropic)-mode resonance. In particular, we propose two ways in which GWD may help trigger a barotropic SSW, both of which involve gravity wave perturbations to vortex parameters explored in Matthewman and Esler (2011).

In their study of split-type SSWs, Matthewman and Esler (2011) considered a single-layer vortex with uniform PV rotating within an infinite background of uniform, but smaller PV. In essence, this model is physically analogous to the high-PV stratospheric polar vortex, which is surrounded by the lower-PV stratospheric surf zone. The model is forced by stationary topography, which generates an azimuthal velocity field. An SSW is triggered when the free-barotropic-wave mode comes into resonance with the azimuthal velocity field. The overall conceptual picture can be understood by considering the vortex’s geometry from barotropic (Figs. 11a–c) and zonal-mean (Figs. 11d–f) perspectives.

Under “typical” wintertime conditions with moderate planetary wave-2 forcing, the polar vortex will continually fluctuate between a circular shape and an elongated, elliptical shape (cf. Figs. 6a,c and Figs. 11a,b, respectively). During times when vertically propagating modes are excited, the vortex may be strongly deformed due to wave breaking and vortex filamentation. Such a situation can clearly be identified by comparing the classical “comma”-shaped vortex patch in Fig. 6b with the results of Polvani and Plumb (1992). If however the right collection of circumstances coalesces and the vortex is excited near its barotropic resonant frequency, the vortex will transition from an elongated geometry to a slightly pinched or peanut-shaped geometry (cf. Fig. 6d and Fig. 11c). It is at this point that the vortex has become tuned to its resonant excitation point and small changes to either of two key vortex–wave parameters will cause the nonlinear wave-amplitude growth that triggers a split SSW (i.e., a wave-amplitude bifurcation).

Matthewman and Esler’s (2011) first vortex parameter, termed the surf-zone parameter, is proportional to the ratio between the background and vortex PV distributions, while the second parameter governs the amplitude of the stationary topographic forcing (i.e., the azimuthal velocity field). As Matthewman and Esler (2011) point out, the surf-zone parameter controls variations in the “climate” of the stratosphere due to the seasonal cycle, changes in longwave radiation, or, in our case, changes in the geometry of the vortex due to the combined effects of GWD and planetary wave breaking. What is particularly important about the surf-zone parameter is the fact that it can be directly related to the azimuthal (angular) velocity field of the polar vortex [see Eqs. (1)–(4) of Polvani and Plumb (1992)]. Angular
velocity, in turn, can be related to the evolution of the polar vortex shown in Figs. 1c, 1d, 8, and 9 via the external-mode experiments of Salby (1981a).

In a suite of cleanly designed numerical experiments, Salby (1981a) showed that latitudinal steps in angular velocity greatly affect the meridional and vertical structure of external Rossby wave normal modes in the stratosphere [see also Geisler and Dickinson (1976) and Salby (1984)]. The results of these experiments for typical northern winter conditions are schematically depicted in Figs. 11d–f. For an atmosphere in uniform rotation and in the absence of damping, Earth's external global normal modes would have the basic amplitude structure (e.g., temperature) shown in Fig. 11d. In contrast, under climatological winter conditions the stratosphere has a strong westerly jet centered in the midlatitudes, where the jet core tilts equatorward with increasing height. Because strong westerlies suppress the vertical penetration of external Rossby modes, these waves will be strongly suppressed everywhere poleward of the jet core (Fig. 11e). Within the context of our results, such a suppression pattern is directly related to the transition from the wide funnel-shaped vortex to the vertically aligned vortex that is constrained about the pole. That is, during the early prewarming period, external Rossby waves will be suppressed everywhere inside the region of red shading in Fig. 1c.

However, as GWD and planetary wave breaking combine to erode the upper half of the vortex and align it in the vertical, the region in which external Rossby waves are suppressed becomes confined to the extreme high latitudes throughout the depth of the stratosphere (cf. the red shading patterns in Figs. 1d and 1f). The essence of this behavior is also reflected in the evolution of the PV gradient and refractive index (Figs. 8 and 9), which as explained in Salby (1981a) provides a simple means of understanding how changes in vortex angular velocity modulate external Rossby waves. Specifically, the sharpened PV gradient and vertical waveguide should lead to local amplification of either type of external normal mode: global or local. Within the context of the local external mode, the sharpened vortex edge is equivalent to increasing the strength of the PV jump [i.e., the surf-zone parameter of Matthewman and Esler (2011)]. Indeed it is the connection between the PV gradient, refractive index, and external modes just described that provides an alternate viewpoint from which to interpret the concepts of vortex erosion and jet sharpening that define vortex preconditioning.
Given that other wave types, including both internal planetary waves and gravity waves, help determine the strength and shape of the polar vortex, this provides one avenue in which other wave types can influence how the barotropic wave modes are manifest in the stratosphere prior to an SSW. This is potentially an important point given that, within the context of resonant barotropic SSWs, Esler et al. (2006) found that the vertical tilt of the polar vortex is an important factor in determining whether the vortex undergoes a partial or complete split throughout the depth of the stratosphere. Indeed, the next logical step will be to understand how barotropic resonance is modulated when the surf-zone parameter is applied to a model setting that allows for vertical variations in vortex angular velocity. Moreover, further work is also needed to clarify whether the PV gradient on the vortex edge can be sharpened sufficiently to support local external modes in addition to the global normal modes just described.

In addition to helping shape the prewarming geometry of the vortex, it is also conceivable that GWD may contribute to triggering a barotropic SSW via changes to the external Rossby wave amplitude itself [i.e., the wave amplitude parameter in Matthewman and Esler (2011)]. A similar hypothesis was put forward by Birner and Williams (2008), who suggested that even if planetary wave amplitudes are not sufficient for causing an SSW, bursts of GWD may be capable of triggering an SSW. While the link between GWD and SSWs has received little attention in the literature, it is well known that gravity waves and planetary waves interact.

For example, Dunkerton and Butchart (1984) explored the selective transmission of internal gravity waves due to planetary waves during SSWs, while others (Holton 1984; Schoeberl and Strobel 1984; McFarlane 1987; McLandress and McFarlane 1993; Smith 2003) describe the in situ generation of planetary waves due to GWD. Of particular note are two results from Holton (1984). First, if the generation of GWD is nonzonal in nature, then there is no a priori need for underlying planetary wave structures in order to produce an in situ planetary wave [see also McLandress and McFarlane (1993)]. Second, Holton found that zonally asymmetric GWD induces planetary waves with barotropic structure; this raises the question of how GWD events such as those in our results may project onto any barotropic external modes already present. In combination, these results indicate that GWD could indeed act to push the wave amplitude past its bifurcation point.

Given the assertion by Matthewman and Esler (2011) that once the polar vortex is near its resonant excitation point, any small (theoretically infinitesimal) increase in planetary wave forcing can trigger the wave-amplitude bifurcation responsible for a barotropic SSW, we consider the potential for GWD to provide such a perturbation. To do so, we compare the time–longitude evolution of GWD (Fig. 6) to the three-dimensional planetary wave fluxes (Plumb 1985) during the 2009 SSW described in Harada et al. (2010).

Harada et al. (2010) plot the evolution of the three-dimensional planetary wave fluxes for six time periods between 9 and 29 January (see their Fig. 6). Consistent with our Fig. 10, Harada et al. (2010) show that the bulk of the planetary wave fluxes appear in the stratosphere between 17 and 29 January. Of particular interest to our results, however, is the fact that the initial, and largest, wave event occurs between 15 and 25 January and appears strictly in the region between 60° and 160°W [see Figs. 6e–m in Harada et al. (2010)]. This is precisely the region where in the previous 5-day period, GWD was maximized and appears to be “pinching” the vortex in a way that is consistent with the evolution from an elongated-vortex shape to a pinched or peanut-shaped vortex, as described earlier in this section and depicted in our Figs. 6d and 11c. If GWD did in fact play a role in amplifying the already-present planetary wave 2, this result would provide evidence of the Holton (1984) mechanism operating in the middle to upper stratosphere.

6. Conclusions

Recent composite analysis of SSWs and vortex preconditioning have focused on the lower stratosphere (Charlton and Polvani 2007; Bancala et al. 2012). We have expanded on this work by using reanalysis data to show that split and displacement SSWs exhibit unique prewarming vortex evolutions that are characterized by coherent vertical structures that extend over a deep layer that includes the entire region between the tropopause and the lower mesosphere. In particular, our analysis reveals that 30 days prior to split SSWs, the polar vortex is characterized by a wide, funnel-shaped vortex geometry that is narrow in the lower stratosphere and wide in the upper stratosphere and is anomalously strong. As the central warming date approaches, vortex erosion in the upper stratosphere due to the combined effects of planetary wave and gravity wave breaking systematically constrain the vortex about the pole such that the vortex exhibits very little vertical tilt. In contrast, we find that displacement SSWs are characterized by a wide, funnel-shaped vortex that is anomalously weak throughout the prewarming period.

The fact that the two types of SSWs exhibit fundamentally different geometric evolutions prior to the central warming suggests that very different types of
wave events are responsible for producing the basic states that are conducive to triggering each of the respective warming types. For the case of split SSWs, the wave–vortex evolution reflects the concepts of vortex edge sharpening and waveguide formation that constrain the vortex to the extreme high latitudes, as outlined by McIntyre (1982). Displacements, on the other hand, result from wave events that simply weaken the strength of the vortex but do not change its generally climatological funnel shape. Thus, despite the fact that precursor wave pulses are regularly cited as evidence of vortex preconditioning for both types of SSWs (Limpasuvan et al. 2004; Nishii et al. 2009; Kuttippurath and Nikulin 2012; Ayarzagüena et al. 2011), our results suggest that precursor wave pulses alone are not sufficient to provide a meaningful definition of preconditioning. Instead, we suggest that preconditioning needs to have two separate definitions—one for displacement SSWs and another for split SSWs—where the definitions reflect the specific vortex geometries needed to support the types of dynamical phenomena that are responsible for triggering each type of warming.

Using the composite analysis as a guide, the 2009 split SSW was examined for evidence that split-type warmings are triggered by anomalous planetary wave forcing from the troposphere below or, alternatively, resonance due to either internal or external Rossby wave normal modes. Differentiating between the triggering mechanisms was facilitated by noting that a stratospheric vertical EP flux event due to either type of resonance will have a unique vertical structure that is distinct from an EP flux event that is due to an anomalously large planetary wave that propagates vertically, breaks, and exerts a drag on the mean flow.

In particular, resonant wave events of either type (internal or external) will exhibit vertical EP flux signatures with little to no vertical phase tilt. For the case of an internal Rossby wave, the lack of vertical tilt is due to the formation of a standing wave associated with reflection inside a high-latitude wave cavity. For external Rossby waves, the lack of vertical tilt is simply due to the fact that they are edge waves that do not propagate vertically away from the surface of Earth. In contrast, if a given vortex disruption is due to anomalously large wave forcing from the troposphere, then large wave fluxes should be traceable at standard group velocity speeds as they propagate up from the tropopause to their breaking region in the upper stratosphere.

Our analysis reveals that both types of wave forcing were active during the 2009 SSW period. For example, several wave pulses could be seen propagating from the tropopause to the upper stratosphere at the theoretically predicted group velocity for planetary waves in the month prior to the central warming date (Fig. 10). The wave events that actually triggered the SSW, however, have the clear signature of resonance in that the wave pulses associated with the warming itself have nearly no vertical phase tilt. Thus, at least for the case of the 2009 SSW, our results lend support to the suggestion by Matthewman and Esler (2011) that split SSWs are triggered by resonance rather than the anomalous forcing from the troposphere. That is not to say, however, that anomalous wave events from the troposphere are unimportant in the life cycle of split SSWs. Rather, we make the following suggestion regarding the concept of vortex preconditioning.

The most common notion of vortex preconditioning is predicated on the notion that SSWs are triggered via focusing anomalously large amounts of wave activity into the polar vortex. In this view, the anomalously large wave fluxes are themselves the cause of the SSW. If, on the other hand, SSWs are triggered via resonance as our results suggest, then the large wave fluxes that accompany an SSW are a characteristic of the vortex breakdown itself and not the cause. In this case, it may be more fruitful to view vortex preconditioning as the process of “tuning” the geometry of the vortex toward its resonant excitation point. From this perspective, once the precursor wave events have tuned the vortex sufficiently toward resonance, then a theoretically infinitesimal change in wave forcing or PV sharpening can trigger resonant self-tuning and the explosion in wave-amplitude growth that causes the vortex to split [i.e., a wave-amplitude bifurcation in the parlance of Matthewman and Esler (2011)]. In combination with the theoretical modeling results of Matthewman and Esler (2011) and Plumb (1981), our results suggest that such a tuning process will include the vortex undergoing the following geometric evolution.

Several weeks prior to a split SSW, the vortex will be characterized by a wide, funnel shape that is anomalously strong. As the central warming date approaches, vortex erosion due to a combination of planetary wave and gravity wave breaking events constrains the vortex about the pole and aligns the vortex in the vertical. As this alignment occurs, an increasing strong meridional PV gradient will form with its axis aligned in the vertical. Several additional features should also be evident dependent on whether internal- or external-mode resonance is responsible for triggering the warming. For internal-mode resonance, a three-sided wave cavity will need to form. In contrast, if the meridional PV gradient becomes strong enough, then the PV gradient itself forms the vertical cavity for the external (barotropic) mode. Indeed, the model that Matthewman and Esler (2011) employed in their study of split SSWs included...
a “perfect” vertical waveguide characterized by an intense and narrow step-wise jump in PV between the surf zone and the vortex core. While our results have provided evidence that either scenario was possible during the 2009 SSW, additional research will be needed to determine whether three-sided wave cavities regularly form to facilitate internal-mode resonance or, conversely, whether the meridional PV gradient can be sufficiently sharpened to provide the needed vertical waveguide for the resonant excitation of local or global external normal modes.

Our analysis also provides the first detailed observational model (reanalysis)–based support for the notion that gravity waves may play an important role in SSWs. For example, we find that gravity waves contribute roughly 10%–30% (ERA-Interim versus JRA-25 data, respectively) of the total wave drag (EPFD + GWD) in the extratropical and polar upper stratosphere in the 30 days prior to both types of SSWs. In particular, our results show that significant gravity wave pulses positively contributed to several features of vortex preconditioning as outlined above. Indeed, it is perhaps not surprising that gravity wave activity is enhanced prior to split SSWs in light of the fact that Charlton and Polvani (2007) found that tropospheric winds are anomalously strong prior to split warmings, which is consistent with the anomalous venting of gravity waves into the stratosphere.

Finally, our results provide support for the suggestion by Birner and Williams (2008) that GWD may potentially play a role in triggering the warming itself. In the vortex evolution suggested by Matthewman and Esler (2011), the vortex is near its resonant excitation point when it achieves a “peanut”-shaped geometry (viewed from a stereographic perspective). At this point, any nominal increase in wave forcing can potentially trigger the cascade of resonant self-tuning that leads to a split SSW. Interestingly, both the JRA-25 and ERA-Interim data suggest that significant gravity wave pulses occurred in a manner that appears to be pinching the vortex near the nadirs of the peanut shape (Fig. 6d). While this provides evidence that is far from conclusive, it may be of particular importance to the concept of external-mode resonance in light of the results of Holton (1984) who showed that zonally asymmetric GWD results in a barotropic planetary wave response in the region below the height of gravity wave breaking; this raises the interesting possibility that GWD events such as those shown in Fig. 6d may project constructively onto any external (barotropic) normal modes already present.

Acknowledgments. The authors appreciate the constructive comments of two anonymous reviewers. This work was supported by the NSF Climate Dynamics Program (TB and JRA), the NSF Atmospheric and Geospace Science Postdoctoral Fellowship program (JRA), and NASA Grant NNX13AM24G (JRA).

APPENDIX

Qualitative Wave Breaking Perturbations

In the absence of wave breaking, Earth’s meridional PV gradient would smoothly increase from the equator to the North Pole. When a planetary wave breaks in the stratospheric “surf zone,” it mixes high- and low-PV air southward and northward, respectively (McIntyre and Palmer 1984). After the mixing event has subsided, the meridional PV gradient will have decreased within the surf zone, while the PV gradient will have increased poleward of the surf zone. This process leads to the jet sharpening originally envisioned and schematically shown in Fig. 5 of McIntyre (1982) [see also Dritschel and McIntyre (2008)]. Within the context of the barotropic dynamics, we can qualitatively reproduce the effect of PV mixing by introducing a relative vorticity perturbation that represents the horizontal mixing of high- and low-vorticity air southward and northward, respectively, subject to the constraint that relative vorticity is conserved in the model domain. The perturbed zonal-mean wind field can then be recovered via vorticity inversion (Vallis 2006).

The vorticity and wind profiles shown in Figs. 5a and 5b very closely match the PV and zonal-wind distributions before and after wave breaking events calculated by Polvani et al. (1995) using a shallow-water model (see their Fig. 1). Interestingly, there is only one major difference between the qualitative PV and wind profiles first sketched in McIntyre (1982) and the profiles shown in our Fig. 5 and the shallow-water results of Polvani et al. (1995), namely, not only does the vorticity perturbation push the jet maximum poleward, but the maximum wind speed of the jet actually increases as well.

In contrast to planetary wave breaking that directly alters the meridional PV gradient by mixing PV in the horizontal plane, orographic gravity waves mix static stability in the vertical. In doing so, gravity waves transfer momentum from Earth’s surface to their breaking region higher in the atmosphere. However, while a single breaking planetary wave is large enough in scale to significantly alter the wind and PV distribution, a single breaking gravity wave has too small an impact to significantly alter the meridional PV gradient. Nevertheless, if enough gravity waves break in a localized region, then the aggregate effect on the zonal-mean
wind, and consequently the PV distribution, can be quite large (e.g., Chen et al. 2007; Limpasuvan et al. 2007). Because orographic gravity wave propagation is enhanced by strong wind speeds (Duck et al. 1998, 2001; Wang and Alexander 2009), we expect that the largest amount of GWD should occur in the core of the polar night jet. Indeed, this filtering pattern is reflected in the GWD patterns prior to SSWs shown in Fig. 4. Thus, to account for the effect of gravity wave breaking on the zonal-mean wind, we introduce a wind perturbation that maximizes at the jet core and decreases as the background wind speed decreases on either flank of the jet. Although Fig. 2 shows that GWD constitutes roughly 15%–80% of the total wave forcing, we chose a more conservative gravity wave wind perturbation that is 15%–80% of the total wave forcing, we chose a more

REFERENCES


McFarlane, N. A., 1987: The effect of orographically excited gravity wave drag on the general circulation of the lower stratosphere


