Stratospheric Polar Vortices

Darryn W. Waugh

Department of Earth and Planetary Sciences, Johns Hopkins University, Baltimore, Maryland, USA

Lorenzo M. Polvani

Department of Applied Physics and Applied Mathematics and Department of Earth and Environmental Sciences, Columbia University, New York, New York, USA

The intense cyclonic vortices that form over the winter pole are one of the most prominent features of the stratospheric circulation. The structure and dynamics of these “polar vortices” play a dominant role in the winter and spring stratospheric circulation and are key to determining distribution of trace gases, in particular ozone, and the couplings between the stratosphere and troposphere. In this chapter, we review the observed structure, dynamical theories, and modeling of these polar vortices. We consider both the zonal mean and three-dimensional potential vorticity perspective and examine the occurrence of extreme events and long-term trends.

1. INTRODUCTION

The most prominent feature of the stratospheric circulation is the seasonal formation and decay of an intense cyclonic vortex over the winter pole. The strong circumpolar westerly winds at the edge of this “polar vortex” are in stark contrast to the weak easterlies that occur in the summer hemisphere. In both hemispheres, a polar vortex forms in the fall, reaches maximum strength in midwinter, and decays in later winter to spring. The structure and dynamics of these polar vortices play a dominant role in the winter and spring stratospheric circulation and are key to determining distributions of trace gases, in particular ozone, and the couplings between the stratosphere and troposphere.

There has been much interest in the structure and dynamics of stratospheric polar vortices ever since the discovery in the 1950s of this stratospheric “monsoon” circulation (westerlies in the winter and easterlies in the summer) and the recording of rapid warming events in the polar stratosphere (so-called stratospheric sudden warmings); see Hamilton [1999] and Labitzke and van Loon [1999] for historical reviews. However, there was a dramatic increase in interest in the stratospheric vortices in the 1980s with the discovery of the Antarctic ozone hole: Because polar vortices act as containment vessels and allow for the occurrence of extremely low temperatures, they play a critical role in polar ozone depletion and the annual formation of the Antarctic ozone hole [e.g., Newman, this volume; Solomon, 1999]. As a result, there has been a rapid growth in the last 2 decades of observational and modeling studies to better understand the structure and dynamics of polar vortices. Interest in the vortices has further intensified in recent years as numerous studies have shown that the polar vortices can influence tropospheric weather and climate. In particular, vortices are an important component of the dynamical stratosphere-troposphere couplings and so-called “annular modes” [e.g., Kushner, this volume].

In this chapter, we review the observed structure of polar vortices and briefly summarize recent advances in our understanding of their dynamics. We briefly touch upon aspects relevant to polar ozone depletion and stratosphere-troposphere coupling but leave detailed discussions of these issues to Newman [this volume] and Kushner [this volume], respectively. For earlier reviews of polar vortices, see Schoeberl and Hartmann [1991] and Newman and Schoeberl [2003].
The observed climatological structure and variability of the polar vortices is first summarized in section 2, focusing on zonal mean aspects. In section 3, polar vortices are examined from a potential vorticity (PV) perspective, followed by a discussion of dynamical theories and modeling based on PV, including Rossby wave propagation and “breaking” and formation of a “surf zone” surrounding the vortices. In section 4, we discuss the observations and theories of extreme vortex events, including so-called “stratospheric sudden warmings.” The coupling with the troposphere is discussed in section 5, including examination of the possible impacts of stratospheric polar vortices on tropospheric weather and climate. In section 6, we review observed trends over the past 4 decades and model projections of the possible impact of climate change on stratospheric polar vortex dynamics. Concluding remarks are given in the final section.

2. CLIMATOLOGICAL STRUCTURE

The general characteristics of stratospheric polar vortices can be seen in plots of zonal mean zonal winds. For example, Figure 1a shows the latitude-height variations of climatological zonal winds for July (left plot) in the Southern Hemisphere (SH) and January (right plot) in the Northern Hemisphere (NH). (See, for example, Andrews et al. [1987] and Randel and Newman [1998] for similar plots for other months and of zonal mean temperatures.) For both hemispheres, there is a strong westerly jet, the center of which corresponds roughly to the edge of the polar vortex. The westerly jets shown in Figure 1a form because of strong pole to equator temperature gradients, and there are very low temperatures over the winter polar regions (see below).

Stratospheric polar vortices form in fall when solar heating of polar regions is cut off, reach maximum strength in midwinter, and then decay in later winter to spring as sunlight returns to polar regions. This is illustrated in Figure 1b, which shows the latitude-seasonal variations of the zonal winds in the middle stratosphere (10 hPa). In both hemispheres, there are weak easterlies during summer months (June–August in the NH and December–February in the SH), which are replaced by westerlies in fall that grow in strength until there is a strong zonal flow in midwinter. These strong westerlies flow then decay through spring, and the flow returns to easterlies in the summer.

Although radiative processes (e.g., heating by absorption of solar radiation by ozone and cooling by thermal emission by carbon dioxide) play the forcing role in setting up the large-scale latitudinal temperature gradients and resulting zonal flow, the winter stratosphere is not in radiative equilibrium. Waves excited in the troposphere (e.g., by topography, landsea heating contrasts, or tropospheric eddies) propagate up into the stratosphere and perturb it away from radiative equilibrium, and the zonal winds shown in Figure 1 are weaker than predicted by radiative equilibrium [see Andrews et al., 1987]. Moreover, the propagation of such waves into the stratosphere varies with conditions in the stratosphere itself. Charney and Drazin [1961] showed that Rossby waves propagate upward only if their horizontal scale is large and if the flow is weakly eastward relative to their phase speed; that is, stationary waves only propagate through weak westerlies [see Andrews et al., 1987]. As a result, stationary Rossby waves propagate up into the stratosphere in the winter (when westerlies are prevalent) and not in the summer (when easterlies are prevalent), and the stratospheric flow is more disturbed in the winter than in the summer.

Large hemispheric differences in the polar vortices can be seen in Figure 1: The Antarctic vortex is larger, stronger (more rapid westerlies), and has a longer lifespan than its Arctic counterpart. These differences are caused by hemispheric differences in the wave generation and propagation. The
larger topography and land-sea contrasts in the NH excite more/larger planetary-scale Rossby waves that disturb the stratospheric vortex and push it farther from radiative equilibrium than in the SH. The hemispheric differences in the strength and, in particular, coldness of the polar vortices are extremely important for understanding ozone depletion, as explained below.

There are also significant hemispheric differences in the variability of the vortices, with the Antarctic vortex being less variable on both intraseasonal and interannual time scales. These differences can be seen in the evolution of minimum polar temperatures at 50 hPa, shown in Figure 2. Similar features are observed in other temperature diagnostics and in high-latitude zonal winds [e.g., Randel and Newman, 1998; Yoden et al., 2002]. The climatological minimum temperatures (thick curves) in the Antarctic are lower and stay colder longer than in the Arctic. Also, there is much larger variability in the Arctic temperatures than in the Antarctic: In the Arctic, a large range of temperatures can be observed from fall to spring (November to April), whereas in the Antarctic there is a fairly narrow range of values except during late spring (October–November). The range and quartiles in Figure 2 show that the distribution of Arctic temperatures is non-Gaussian and highly skewed; see Yoden et al. [2002] for more discussion.

The large variability in the Arctic occurs on interannual, intraseasonal, and weekly time scales. Within a single winter, there can be periods with extremely low temperatures as well as periods with extremely high temperatures, and the transition between these events can occur rapidly. These extreme events, and in particular weak events (so called “stratospheric sudden warming”), are discussed further in section 4.

The differences in polar temperatures shown in Figure 2 explain hemispheric differences in polar ozone depletion. In the Antarctic, midwinter minimum temperatures are lower than threshold temperatures for formation of polar stratospheric clouds (PSCs) every year (horizontal lines in Figure 2), and formation of PSC, chemical processing, and widespread ozone depletion occur every year. In contrast, Arctic temperatures fall below the threshold for PSC formation less frequently, and, as a consequence, ozone depletion in the Arctic is much less frequent and widespread. See Newman [this volume] for more details.

The interannual variability of the vortices is due to external forcing of the atmospheric circulation, e.g., solar variations, volcanic eruptions, and anthropogenic changes in composition (e.g., ozone and greenhouse gases (GHGs)), as well as internal variations within the climate system, e.g., the quasi-biennial oscillation (QBO), El Niño–Southern Oscillation (ENSO), and internal variability due to nonlinearities. See Gray [this volume] and Haigh [this volume] for more discussion of the influence of the QBO and solar variation, respectively, on the variability of the vortices.

3. POTENTIAL VORTICITY DYNAMICS

While examination of zonal mean quantities yields information on the general structure and variability of the vortices, examination of the three-dimensional structure is required for greater insight into the synoptic variability and dynamics of the vortices. A quantity that is particularly useful for understanding the structure and dynamics of the polar vortices is potential vorticity (PV), i.e.,

\[ PV = \rho^{-1} \zeta \cdot \nabla \theta, \]

where \( \rho \) is the fluid density, \( \zeta \) is the absolute vorticity, and \( \nabla \theta \) is the gradient of the potential temperature. Several properties of PV make it useful for studying the polar vortices. First, PV is materially conserved for adiabatic,
frictionless flows. Second, other dynamical fields can be determined from the PV distribution (assuming an appropriate balance and given boundary conditions) using “PV inversion” [e.g., Hoskins et al., 1985; McIntyre, 1992]. Finally, PV gradients provide the restoring mechanism for Rossby waves, so that the dynamics and propagation of Rossby waves are best understood by examining the distribution of PV [e.g., Hoskins et al., 1985].

Maps of PV on isentropic surfaces provide useful information on the structure and evolution of polar vortices and the associated transport of trace gases. On such maps, a polar vortex appears as a roughly circular, coherent region of high PV, with steep PV gradients at the edge of the high-PV region. (Note that PV is negative in the SH, and when we refer to high PV we are referring to high absolute value of PV in the SH.) The steep PV gradients are colocated with strong westerly winds, and steeper gradients correspond to stronger winds. As PV is materially conserved over the time scale of several days to a week, maps of PV can be used to trace the evolution of the polar vortices and also the transport of trace gases.

Maps of climatological mean PV show, as discussed above, that the Antarctic vortex is larger and stronger (steeper gradients tend to correspond to more rapid westerlies) than the Arctic vortex; for example, see Figure 3, which shows maps of the climatological mean PV on the 850 K isentropic surface (~10 hPa) for the NH in January (Figure 3a) and the SH in July (Figure 3b). These maps also show significant differences in the zonal variations, with the climatological Antarctic vortex being more symmetrical and centered nearer the pole than the Arctic vortex [e.g., Randel and Newman, 1998; Waugh and Randel, 1999; Karpechko et al., 2005]. The climatological Arctic vortex is shifted toward the Eurasian continent, and there is, climatologically, a stationary anticyclone over the Aleutian Islands [e.g., Harvey et al., 2002]. The vortices are smaller and, especially in the NH, more distorted in the lower stratosphere than in the upper stratosphere [e.g., Randel and Newman, 1998; Waugh and Randel, 1999]. Even though the Arctic vortex is more disturbed than the Antarctic vortex, the scale of the disturbances is much larger than those in the troposphere (e.g., disturbances in the stratosphere are typically between zonal wave numbers 1 and 3, whereas tropospheric disturbances are waves 5 to 7). This is typically understood in terms of the Charney-Drazin theory that describes the filtering of waves propagating into the stratosphere [e.g., Andrews et al., 1987].

As mentioned above, there can be large day-to-day variability in the structure of the Arctic vortex. Figure 4 illustrates this with PV maps for several days during January and February 1979. In mid-January, the Arctic vortex is centered near the pole, but by the end of the month the vortex has elongated and is displaced from the pole, and there is a “tongue” of PV extending from the vortex into the mid-latitudes. The vortex has returned to a circular shape by early February, but it is again disturbed in middle to late February. During this latter period, the vortex again becomes elongated, but in contrast to the late January event the vortex remains centered near the pole and splits into two smaller regions of high PV. The February event corresponds to a “stratospheric sudden warming” (SSW) and has been examined in numerous studies and by Andrews et al. [1987]. The occurrence and general characteristics of SSWs are discussed further in section 4.

PV is useful not only for visualizing the structure and evolution of the vortices, but it is also provides valuable insight into the dynamics of the vortices. McIntyre and Palmer [1983, 1984] examined maps of PV for the same period shown in Figure 4 and interpreted features in these

---

**Figure 3.** Maps of the climatological mean PV on the 850 K isentropic surface (~10 hPa) for (a) NH in January and (b) SH in July.
maps in terms of Rossby waves. They associated reversible distortions of the vortex with propagating Rossby waves and contrasted these with irreversible deformations where air with high PV is pulled off the vortex and stirred into midlatitudes (e.g., late January 1979, Figure 4). McIntyre and Palmer referred to the latter process as Rossby “wave breaking” and to the region surrounding the vortex as the stratospheric “surf zone.” Such Rossby wave breaking is very common at the vortex edge, occurs even when there are no SSW events, and has been documented in numerous observational studies [e.g., Baldwin and Holton, 1988; Abatzoglou and Magnusdottir, 2007].

Rossby wave breaking and accompanying formation of a surf zone has been extensively studied with numerical models. In a seminal study, Juckes and McIntyre [1987] performed high-resolution, single-layer simulations of an initially symmetric polar vortex disturbed by “wave 1” forcing (to mimic the impact of upward propagating Rossby waves). These simulations showed Rossby wave breaking at the edge of the vortex, resulting in filaments of vortex air being stripped from the vortex and stirred into middle latitudes (and filaments of tropical air also being entrained into midlatitudes), and formation of steep PV gradients. This is illustrated in Figure 5b, which shows the vortex evolution for similar calculations in a spherical, shallow water model [see also Juckes, 1989; Salby et al., 1990; Norton, 1994; Polvani et al., 1995]. This Rossby wave breaking, stirring of filaments into middle latitudes, and vortex erosion are ubiquitous features of polar vortex simulations, and they have been reported in a large number of subsequent studies using a hierarchy of models, ranging from simple planar models where the vortex is represented by a single region of uniform PV (Figure 5a) [Polvani and Plumb, 1992] to three-dimensional models (Figure 5c) [see also Haynes, 1990; Dritschel and Saravanam, 1994; Waugh and Dritschel, 1999; Polvani and Saravanam, 2000]. Furthermore, classical two-dimensional vortex flows also show the formation of steep gradients at the edge of vortices (so called “vortex stripping” [e.g., Legras et al., 2001]) and the stability of filaments in straining flows [Dritschel, 1989; Dritschel et al., 1991]. The latter helps explain the robustness of filaments in stratospheric simulations.

Three-dimensional (3-D) simulations of a forced vortex show two classes of wave breaking: “remote” wave breaking where Rossby waves propagate up the vortex edge and break in upper levels (e.g., Figure 5c) and “local” breaking where the wave breaking occurs at lower levels of the vortex and inhibits further wave propagation into the upper levels.

**Figure 4.** Maps of PV on the 750 K isentropic surfaces for several days during January and February 1979 from National Centers for Environmental Prediction/National Center for Atmospheric Research reanalyses.
The 3-D simulations also show the sensitivity of the wave propagation and wave breaking to the gradients at the edge of the vortex, with enhanced vertical propagation and breaking for steeper-edge gradients [Polvani and Saravanan, 2000; Scott et al., 2004]. This has potential implications for how well the dynamics of polar vortices is modeled in low- and moderate-resolution climate models.

The focus of the above simulations, and analysis of observations, has primarily been on the impact of upward propagating Rossby waves. However, the recent study of Esler and Scott [2005] revisited the idea of resonant excitation of free modes considered earlier by Tung and Lindzen [1979] and Plumb [1981]. Esler and Scott showed that in an idealized 3-D quasi-geostrophic model, there are not only upward and downward propagating Rossby waves, but there is also a barotropic mode. This latter mode can be forced by transient forcing and can dominate over the upward propagating waves. In fact, they showed that a “barotropic” sudden warming occurs if this mode is resonantly excited.

Furthermore, some observed NH vortex-splitting major sudden warmings, e.g., in 1979 and 2009, exhibit a very similar structure to the modeled barotropic sudden warming.

Although the dynamical evolution of polar vortices is controlled by relatively large zonal wave numbers, numerical simulations show the rapid formation of fine-scale features, e.g., PV filaments and very steep PV gradients (e.g., Figure 5). As these fine-scale features cannot be resolved in low-resolution stratospheric satellite observations or meteorological analyses (e.g., Figure 4), their realism might, at first glance, be doubted. However, high-resolution trajectory calculations performed using winds from meteorological analyses are able to produce fine-scale features, including filaments and steep gradients, that closely resemble those in the single and multilayer numerical simulations. For example, Figure 6 shows high-resolution simulations of the lower stratospheric vortex in January 1992. These maps show fine-scale features that are not seen in the analyzed PV fields, but they are consistent with features in the high-resolution dynamical simulations. Furthermore, the presence of these fine-scale features was confirmed by aircraft measurements of...
chemical tracers made during this period [e.g., Plumb et al., 1994; Waugh et al., 1994].

The breaking of Rossby waves at the vortex edge is important not only for the dynamics of vortices, but it also plays an important role in the transport of trace gases; see Plumb [2002] and Shepherd [2007] for comprehensive reviews of stratospheric tracer transport. As discussed above, Rossby wave breaking strips air from the edges of the vortices, forms steep edge gradients, and stirs vortex air into middle latitudes. This transport plays a major role in determining the distribution of trace gases in the stratosphere. Latitude-height contour plots of long-lived trace gases such as N$_2$O, CH$_4$, and CFCs (which have tropospheric sources and upper stratospheric/mesospheric sinks) show that tracer isopleths are depressed within the polar vortices compared to middle latitudes [e.g., Schoeberl and Douglass, this volume, Figure 1]. While larger polar descent contributes to the difference in tracers inside and outside the vortex, the dominant cause is the above difference in horizontal stirring. There is rapid stirring in the surf zone and very little mixing across the vortex edge, which leads to homogenization of tracers within the surf zone and steep latitudinal tracer gradients at the edge of the polar vortices (there are also steep gradients at the subtropical edge of the surf zone).

Although the vortex edge is a barrier to mixing, it is not a perfect barrier. There is evidence, primarily from tracer-tracer relationships, of mixing across the vortex edge [e.g., Plumb, 2007, and references therein]. The mechanisms for this mixing and when this mixing occurs is still a matter of debate. “Inward” Rossby wave-breaking events in which surf zone air is mixed into the vortex have been observed (e.g., Figure 6). However, these events are infrequent, and Rossby wave breaking that transports air from the vortex into middle latitudes is much more common.

4. EXTREME EVENTS

As discussed in section 2, there is large variability in the Arctic vortex during fall to spring, and there are periods when the vortex is anomalously strong and periods when it is weak (or even nonexistent), with rapid transitions between these states. The variability of the Arctic vortex and existence of extreme events is illustrated in Figure 7, which shows the time series of the Northern Annular Mode (NAM) index for 24 years [Polvani and Waugh, 2004]. The NAM is the dominant pattern of variability in the northern extratropical troposphere and stratosphere [Thompson and Wallace, 2000]. Since the polar vortex dominates the variability of the stratosphere, the NAM index at 10 hPa is a rough measure of the strength of the stratospheric vortex, with a positive NAM index corresponding to a strong vortex and a negative index corresponding to a weak vortex (e.g., the breakup of the Arctic vortex shown in Figure 4 corresponds to a period when the NAM is around $-3$).

There is virtually no variability in the NAM index during summer but large variability during winter and spring, with rapid transitions between strong and weak vortices. Extremely strong and weak events occur on average around once every other winter, although the occurrence of extreme events is not evenly spread: Two weak vortex events can occur in one winter (e.g., 1998/1999 winter), and there can be extended periods when there are few weak events but frequent strong events (e.g., early to mid-1990s [Manney et al., 2005]). The probability distribution function of the NAM index in winter [Baldwin and Dunkerton, 2001; Polvani and Waugh, 2004] is close to Gaussian, and the frequency of strong/weak vortex events is consistent with expectations for a random process.

The occurrence of strong and weak vortex events has been linked to the upward wave activity entering the stratosphere. Case studies and composite analyses have shown that anomalously strong wave activity nearly always precedes weak vortex events (SSWs), and, conversely, anomalously weak wave activity precedes strong vortex events [e.g., Christiansen, 2001; Polvani and Waugh, 2004]. This is illustrated in Figure 7 where the time series of eddy heat flux at 100 hPa integrated over the prior 40 days (a measure of the time-integrated wave activity entering the stratosphere) is
shown to be anticorrelated with the NAM index at 10 hPa (correlation coefficient of $-0.8$). Theoretical support for this observed relationship is provided by Newman et al. [2001], who showed that stratospheric polar temperatures (and, via geostrophic balance, stratospheric winds) on a given day are related not to the instantaneous upward wave activity but to its weighted integral over several weeks prior to that day [see also Esler and Scott, 2005]. There is also a clear relationship on interannual time scales between the state of the stratosphere and the time-integrated wave activity upwelling from the troposphere, with strong upward wave activity during winters with a weaker, warmer vortex that breaks up earlier [e.g., Waugh et al., 1999; Newman et al., 2001; Hu and Tung, 2002].

Periods when the NAM index is less than $-2.7$ are generally associated with major SSW events, which are traditionally, and more simply, defined by the reversal of the zonal mean zonal winds at 60°N and 10 hPa. There has been considerable research into SSWs, and there is an extensive literature on the vortex evolution during individual SSWs. However, until recently, there were few studies of the climatological nature of SSWs. Using the traditional definition of SSWs based on zonal winds, Charlton and Polvani [2007] reported an average of around six SSWs every decade (29 SSWs in the 44 winters between 1957/1958 and 2001/2002). This study and those of Limpasuvan et al. [2004] and Matthewman et al. [2009] also used composite-based analysis to examine the climatological nature of SSWs, including the evolution of temperature, zonal flow, and eddy fluxes during warming events as well as the three-dimensional structure of the vortices. (The climatological nature of strong vortex events has also been examined by Limpasuvan et al. [2005].)

Traditionally, SSWs were classified into wave 1 or wave 2 events depending on the amplitude of the longitudinal wave numbers. However, given the nonlinearity of the flow, such a classification can be misleading [see Waugh, 1997], and it is more appropriate to use a vortex-oriented classification. Charlton and Polvani [2007] and Matthewman et al. [2009] used such a classification and divided the SSWs into either “vortex displacement” or “vortex split” events. (The minor warming in January and major warming in February 1979, shown in Figure 4, are examples of vortex displacement and vortex split events, respectively.) They showed that the two types of events are dynamically different, with differences in the stratospheric and tropospheric flows before the events and in vortex evolution during the event. Although the life cycles of the displacement and split events are very different, there is much less variation between individual events of each class, both in terms of vertical structure and longitudinal orientation, and characteristics are well captured by the composite events [Matthewman et al., 2009]. In particular, during splitting events the vortex
deformations are highly barotropic, while for displacement events the vortex tilts westward with height.

The basic understanding of the dynamics of SSWs, based on the seminal Matsuno [1970, 1971] studies, involves the anomalous growth of upward propagating planetary-scale Rossby waves originating in the troposphere (see Andrews et al. [1987] for review). However, the cause of the rapid amplification of the Rossby wave and the role of the initial state of the stratospheric vortex are not fully understood. There is evidence that the vortex needs to be preconditioned for SSWs to occur [e.g., McIntyre, 1982; Limpasuvan et al., 2004]. While most analyses of SSWs have focused on upward propagating Rossby waves, alternate theories include resonant excitation of free modes [Tung and Lindzen, 1979; Plumb, 1981; Esler and Scott, 2005; Esler et al., 2006] and vortex interactions, in particular interactions between the polar vortex and Aleutian anticyclone [O’Neill and Pope, 1988; Scott and Dritschel, 2006].

Until recently, it was thought that sudden warmings were exclusively a NH phenomenon. However, a dramatic event occurred in the SH in September 2002, when the Antarctic vortex elongated and split into two. This is the only known SSW in the SH, and there has been considerable research into this event, much of which is summarized in the March 2005 special issue of the Journal of Atmospheric Sciences [see also Baldwin et al., 2003]. Although there have been many studies into the dynamics of this event (see the above mentioned special issue), the exact cause remains unknown. Most of the focus has been on upward propagating Rossby waves, but Esler et al. [2006] provide evidence that the event may have been the result of a self-tuned resonance. Kushner and Polvani [2005] documented the spontaneous occurrence of a sudden warming, resembling the observed SH event, in a long numerical simulation of a simple troposphere-stratosphere general circulation model with no stationary forcing, suggesting that the 2002 event may have been just a rare, random (“natural”) event.

There is large interannual variability not only in the midwinter vortices but also in the timing and characteristics of the final breakdown of the vortices (so called “final warming”) [see, e.g., Waugh and Rong, 2002]. The stratospheric final warmings influence not only the stratospheric circulation (e. g., a transition from winter stratospheric westerlies to summer easterlies) but also strongly organize the tropospheric circulation, with a rapid weakening of high-latitude tropospheric westerlies occurring for both NH and SH final warmings [e.g., Black and McDaniel, 2007a, 2007b].

5. STRATOSPHERE-TROPOSPHERE COUPLING

Stratospheric vortices have, until recently, been considered an interesting middle-atmosphere phenomenon, with little attention paid to their possible impact on the troposphere. However, increasing observational and modeling evidence in the last decade suggests that polar stratospheric vortices can have a significant influence on the tropospheric flow for a range of time scales [e.g., Baldwin and Dunkerton, 2001; Thompson and Solomon, 2002; Polvani and Kushner, 2002; Gillett and Thompson, 2003; Norton, 2003; Charlton et al., 2004].

Much of the evidence for a stratospheric impact on the troposphere focuses on the so-called annular modes: the Northern Annular Mode (NAM) and Southern Annular Mode (SAM) [e.g., Thompson and Wallace, 2000]. As discussed above, these modes are the dominant patterns of variability in the extratropical troposphere and stratosphere, and the NAM/SAM index in the stratosphere is a measure of the vortex strength (see Figure 7). Baldwin and Dunkerton [1999] showed that anomalous values in the NAM index are found to appear in the stratosphere first and subsequently progress downward over periods of several weeks. Moreover, subsequent studies showed that extreme stratospheric events can be followed by anomalous weather regimes at the surface that persist for up to 2 months [Baldwin and Dunkerton, 2001; Thompson et al., 2002]. The exact dynamical mechanism by which the stratosphere influences the troposphere is unknown, but there are several proposed theories, including direct PV inversion [Hartley et al., 1998; Ambaum and Hoskins, 2002; Black, 2002], changes in refractive properties and Rossby wave propagation [Hartmann et al., 2000] or wave reflection [Perlwitz and Harnick, 2004], and eddy-mediated feedbacks [Kushner and Polvani, 2004; Song and Robinson, 2004; Chen and Held, 2007]. See Kushner [this volume] for more detailed discussion of these mechanisms.

Although anomalous values of the annular mode index appear first in the upper stratosphere, it is important to note, as discussed in section 4, that these extreme events are preceded by anomalous wave activity entering the stratosphere (see Figure 7). While the fact that stratospheric extreme events are preceded by anomalous wave activity might indicate that the stratosphere is slave to the troposphere, this is not necessarily the case. Numerous studies, using a hierarchy of models, have shown that internal variability can be generated within the stratosphere, with vacillation cycles of strong (westerly) and weak (easterly) polar winds [e.g., Holton and Mass, 1976; Yoden, 1987; Scott and Haynes, 2000; Rong and Waugh, 2003; Scott and Polvani, 2004]. Furthermore, the Scott and Polvani [2004] simulations show cycles in wave activity entering the stratosphere that resemble those in observations (e.g., Figure 7) even though all forcings in their simple model are completely time-independent. This therefore suggests that the stratosphere plays a role in determining the wave activity entering from the troposphere.
Couplings between the stratospheric vortices and tropospheric circulation have also been found in the SH. Observations show a strengthening of westerlies (and corresponding increase in the SAM) in both the stratosphere and troposphere over the past 2–3 decades. The largest stratospheric trends occur in spring months, whereas the largest tropospheric trends occur in the summer. This is consistent with Antarctic ozone depletion strengthening the stratospheric vortex (see section 6) and a time lag for stratospheric anomalies to descend to the surface [Thompson and Solomon, 2002; Gillett and Thompson, 2003]. By modifying the SAM, a strengthening (or weakening) Antarctic vortex also has the potential to impact other aspects of the tropospheric circulation, including subtropical jets, storm tracks, the Hadley cell width, and subtropical hydrology [Perlwitz et al., 2008; Son et al., 2008, 2009].

6. TRENDS

Given the key role the polar vortices play in ozone depletion and stratosphere-troposphere coupling, it is important to understand how polar vortices have changed over the past few decades and how they might change in the future.

Several studies have examined decadal variability and trends in polar vortices over the past 4 decades using meteorological reanalyses [e.g., Waugh et al., 1999; Zhou et al., 2000; Langematz and Kunze, 2006; Karpetchko et al., 2005]. Because of the lack of stratospheric measurements in the SH in the presatellite era, reliable trends can only be determined from 1979 on. However, there are sufficient stratospheric observations in the 1960s and 1970s in the NH to perform trend analyses from 1960 to present day for the Arctic vortices. These studies have shown that there are significant trends in the springtime Antarctic vortex and that the vortex has become stronger, colder, and more persistent (i.e., breaks up later) since 1979; see Figure 8. The colder vortex and delay in breakup are attributed to the decreases in ozone within the vortex over this period (i.e., growth in the Antarctic ozone hole). As discussed above, these changes in the spring Antarctic vortex have caused changes in SH tropospheric circulation.

The large interannual variability of the Arctic vortex (e.g., Figure 2) makes detecting long-term trends very difficult. Although trends have sometimes been reported for shorter data records, there are no significant trends in the size or persistence in the Arctic vortex between 1958 and 2002 (Figure 8) [see also Karpetchko et al., 2005]. However, trends in the area or volume of temperatures below the threshold for PSC formation have been noted [e.g., Knudsen et al., 2004; Karpetchko et al., 2005; Rex et al., 2006]. The most intriguing result is the analysis of Rex et al. [2006] that indicates that there is a significant increase in “PSC volume” if coldest winters within 5 year periods are considered. As there is a high correlation between PSC volume and ozone depletion, this cooling of cold winters implies an increase in Arctic ozone depletion. The cause of the increased PSC volume for coldest winters is not known, and similarly, it is not known whether this trend will continue into the future.

There is a great deal of interest in possible future trends in the polar vortices. As significant changes in concentrations of key radiative gases in the stratosphere are expected over the 21st century (ozone is expected to increase as the concentrations of ozone-depleting substances decrease back to 1960 levels, and GHGs are expected to continue to increase), some changes in the polar vortices may be expected. It is also possible that increases in GHGs could lead to changes in wave activity and propagation from the troposphere, which could then lead to changes in the vortices. There have been numerous modeling studies examining possible changes in the stratosphere over the 21st century, but the majority of these studies have focused on changes in stratospheric ozone [e.g., Austin et al., 2003; Eyring et al., 2007; Shepherd, 2008] or circulation [e.g., Garcia and Randel, 2008; Oman et al., 2009; Butchart et al., 2009], and there have been limited detailed studies of changes in the vortices. However, there has been analysis of monthly mean temperatures in these simulations that provide insight into possible changes in the vortices.

An early study by Shindell et al. [1998] indicated a significant cooling (and strengthening) of the Arctic vortex as GHGs increased, leading to significant ozone depletion and formation of an Arctic ozone hole. However, more recent simulations with more sophisticated chemistry-climate models (CCMs) that have a better representation of the
dynamics and chemistry (and couplings between them) do not show a significant strengthening or formation of Arctic ozone holes during the 21st century [e.g., Austin et al., 2003; Eyring et al., 2007]. The long-term trends in Arctic temperatures in these chemistry-climate models are small, with no consistency as to whether the polar stratosphere will be warmer or colder [Butchart et al., 2009]. The CCMs also predict a limited impact of increased GHGs on the Antarctic vortex during the 21st century. However, the same simulations all predict an increase in the tropical upwelling as GHGs increase, which has been attributed to changes in subtropical wave driving [e.g., Garcia and Randel, 2008; Oman et al., 2009]. This change in tropical circulation appears not to be strongly connected with changes in polar regions [e.g., McLandress and Shepherd, 2009].

In terms of sudden warmings, it is important to note first that most state of the art chemistry-climate models severely underestimate their frequency [Charlton et al., 2007]. As for what might be expected in the 21st century, only one study is available [Charlton-Perez et al., 2008]: On the basis of several long integrations with a single model, the Charlton-Perez et al. suggest that the frequency of sudden warmings (currently six events per decade) might increase by one event per decade by the end of the 21st century. Needless to say, owing to the large interannual variability, such trends are very difficult to estimate, and the question remains largely open.

Finally, and perhaps most importantly, the recovery of Antarctic ozone is predicted to cause a positive trend in lower stratospheric temperatures and vortex strength in late spring to summer. As discussed in section 5, changes in the Antarctic lower stratospheric temperatures over the last 2 decades have been linked to changes in Southern Hemisphere climate. The ozone recovery over the next 4 to 5 decades is predicted to reverse these changes [e.g., Son et al., 2008; Perlwitz et al., 2008]. It is important to note that in the latter part of the 20th century, the impact of ozone depletion on the tropospheric circulation has been in the same sense as the impact of increasing GHGs. However, as ozone recovers, the stratospheric impact will oppose, and even reverse, some of the expected changes to increases in GHGs.

7. OUTSTANDING ISSUES

Over the past few decades, the combined use of theory, observations, and modeling has greatly advanced our understanding of, and ability to model, the dynamics of stratospheric polar vortices, their impact on the transport of chemical tracers, and stratosphere-troposphere couplings. Many of these advances have come from consideration of the vortex as a material entity. This has lead to an improved understanding of the three-dimensional propagation of Rossby waves, the impact of Rossby wave breaking on vortex dynamics and tracer transport, and the dynamical coupling between the stratosphere and troposphere. Two recent examples are the paper of Waugh et al. [2009], which showed the importance of zonal asymmetries in the Antarctic vortex and ozone hole in tropospheric climate change, and that of Martius et al. [2009], which showed a close connection between the type of atmospheric block and whether the stratospheric warming was a vortex displacement or splitting event. Nevertheless, some important questions remain unanswered.

The exact causes of the variability in the vortices, including the occurrence of SSWs, are not known. It has been shown that this variability is linked to wave activity entering the stratosphere, but the relative contributions from internal stratospheric, tropospheric, and coupled processes to the variability in wave activity are not known. Furthermore, the recent Esler and Scott [2005] and Esler et al. [2006] studies raise the possibility that the resonant excitation of free modes, as opposed to upward propagating Rossby waves, plays a larger role than previously thought. Most attention on extreme vortex events has been focused on the Northern Hemisphere, but the dramatic Southern Hemisphere SSW in 2002 has changed this. There is now a lot of focus on the cause and occurrence of SSWs in the SH.

Numerous recent observational and modeling studies have shown that changes in the stratospheric polar vortices can influence the tropospheric circulation, on both weather and climate time scales. However, there remains uncertainty in the precise dynamical processes involved. Several mechanisms have been proposed, including direct nonlocal dynamical effects, downward reflection of Rossby waves, and alteration of synoptic eddies in the upper troposphere, but more research is required to determine the relative importance of these processes.

A key issue for both the recovery of stratospheric ozone and the influence of the stratosphere on tropospheric climate is how the polar vortices will change, if at all, as greenhouse gases continue to increase. The stratosphere will cool because of the direct radiative effect of increased CO₂, but whether the polar vortices will be come stronger or weaker will likely depend on changes in wave activity entering the stratosphere. There is currently no agreement between climate models as to trends in either the wave activity entering the stratosphere or the strength of the polar vortex, although the trends are generally small in all models. It is unclear how much confidence can be put into the model projections of the vortices given that the models typically only have moderate resolution and that the climatological structure of the vortices in the models depends on the tuning of gravity wave parameterizations.


Given the above outstanding issues, there is need for continued research in the dynamics of the vortices and their representation in global models.

REFERENCES


Haynes, P. H. (1990), High-resolution three-dimensional modelling of stratospheric flows: Quasi-2D turbulence dominated by a single


McLandress, C., and T. G. Shepherd (2009), Simulated anthropogenic changes in the Brewer–Dobson circulation, including its extension to high latitudes, J. Clim., 22, 1516–1540.


L. M. Polvani, Department of Applied Physics and Applied Mathematics, Columbia University, New York, NY 10027, USA.
D. W. Waugh, Department of Earth and Planetary Sciences, Johns Hopkins University, Baltimore, MD 21218, USA. (waugh@jhu.edu)