

# Why Are Stratospheric Sudden Warmings Sudden (and Intermittent)?

NOBORU NAKAMURA AND JONATHAN FALK

*Department of the Geophysical Sciences, University of Chicago, Chicago, Illinois*

SANDRO W. LUBIS

*Department of Mechanical Engineering, Rice University, Houston, Texas*

(Manuscript received 13 September 2019, in final form 25 December 2019)

## ABSTRACT


This paper examines the role of wave–mean flow interaction in the onset and suddenness of stratospheric sudden warmings (SSWs). Evidence is presented that SSWs are, on average, a threshold behavior of finite-amplitude Rossby waves arising from the competition between an increasing wave activity  $A$  and a decreasing zonal-mean zonal wind  $\bar{u}$ . The competition puts a limit to the wave activity flux that a stationary Rossby wave can transmit upward. A rapid, spontaneous vortex breakdown occurs once the upwelling wave activity flux reaches the limit, or equivalently, once  $\bar{u}$  drops below a certain fraction of  $u_{\text{REF}}$ , a wave-free, reference-state wind inverted from the zonalized quasigeostrophic potential vorticity. This fraction is 0.5 in theory and about 0.3 in reanalyses. We propose  $r \equiv \bar{u}/u_{\text{REF}}$  as a local, instantaneous measure of the proximity to vortex breakdown (i.e., preconditioning). The ratio  $r$  generally stays above the threshold during strong-vortex winters until a pronounced final warming, whereas during weak-vortex winters it approaches the threshold early in the season, culminating in a precipitous drop in midwinter as SSWs form. The essence of the threshold behavior is captured by a semiempirical 1D model of SSWs, similar to the “traffic jam” model of Nakamura and Huang for atmospheric blocking. This model predicts salient features of SSWs including rapid vortex breakdown and downward migration of the wave activity/zonal wind anomalies, with analytical expressions for the respective time scales. The model’s response to a variety of transient wave forcing and damping is discussed.

## 1. Introduction

Stratospheric polar vortex in the Arctic winter is significantly more volatile than its Antarctic counterpart, the culmination of which being stratospheric sudden warmings (SSWs). Recurring every 1.5–2 winters on average, SSWs involve a rapid transformation of the polar vortex, reversal of the circumpolar winds and a sudden increase in the polar temperatures (Scherhag 1952; Quiroz 1975; McIntyre and Palmer 1983, 1984; Limpasuvan et al. 2004; Charlton and Polvani 2007; Butler et al. 2015). SSWs not only disrupt stratospheric circulation but also affect tropospheric circulation of the extratropics in weeks to follow (Kodera and Chiba 1995; Baldwin and Dunkerton 1999, 2001; Kushner and

Polvani 2004; Polvani and Waugh 2004; Gerber et al. 2009; Hitchcock and Simpson 2014).

SSWs are arguably the most dramatic manifestation of wave–mean flow interaction in Earth’s atmosphere, for which transience in the planetary-scale Rossby waves plays a central role (McIntyre 1982; Limpasuvan et al. 2004). In the first successful modeling of SSW, Matsuno (1971) assumes that wave transience is a forced response of the stratosphere to tropospheric processes such as atmospheric blocking, a paradigm adopted by many subsequent studies (e.g., Holton 1976; Polvani and Saravanan 2000; Scott and Haynes 2000; Sjöberg and Birner 2012), with substantial observational support (Randel and Boville 1987; Polvani and Waugh 2004; Martius et al. 2009; Woollings et al. 2010). An impulsive wave forcing at the tropopause produces a vertically propagating packet of quasi-stationary Rossby waves, and the converging Eliassen–Palm (E–P) flux at the leading edge of the wave packet decelerates the zonal-mean zonal wind (Matsuno 1971).

 Denotes content that is immediately available upon publication as open access.

Corresponding author: Noboru Nakamura, nnn@uchicago.edu

DOI: 10.1175/JAS-D-19-0249.1

© 2020 American Meteorological Society. For information regarding reuse of this content and general copyright information, consult the AMS Copyright Policy ([www.ametsoc.org/PUBSReuseLicenses](http://www.ametsoc.org/PUBSReuseLicenses)).

An alternative paradigm of SSW asserts that wave transience arises from the internal dynamics of the stratosphere. [Holton and Mass \(1976\)](#) demonstrate with an idealized model that even a steady wave forcing at the tropopause generates stratospheric vacillation in wave activity and the zonal-mean zonal wind when a critical forcing amplitude is exceeded. [Plumb \(1981\)](#) proposes instability due to interaction between the steady wave forcing and near-resonant free waves in the stratosphere as a cause of wave transience (see also [Geisler 1974](#); [Tung and Lindzen 1979](#); [Smith 1989](#)). Unlike the transient forcing paradigm, the resonance theory predicts the amplification of the stationary wave as part of the solution. Resonance and wave-mean flow interaction also produce multiple equilibria (cf. [Charney and DeVore 1979](#); [Held 1983](#)), prompting some to characterize SSW as a transition from a “high-index” (nearly zonal) state to a “low-index” (perpetual resonance) state of the polar vortex ([Chao 1985](#); [Yoden 1987, 1990](#); [Ruzmaikin et al. 2003](#); [Birner and Williams 2008](#); [Yasuda et al. 2017](#)). The resonance idea has been applied to nearly barotropic, split-type SSWs ([Esler and Scott 2005](#); [Matthewman and Esler 2011](#); [Liu and Scott 2015](#)).

Whether one adopts the first paradigm or the second, modeling studies have long shown that the properties of wave transience are sensitive to the initial profile of the stratospheric winds and their vertical shear ([Holton and Dunkerton 1978](#); [Chen and Robinson 1992](#); [Scott and Polvani 2006](#)). This suggests that the condition of the stratosphere has a strong filtering effect on the forcing from below ([Hitchcock and Haynes 2016](#)). Generally, weak (but not too weak) zonal winds are conducive to SSWs, but such “preconditioning” of stratospheric winds in turn depends on prior wave events (or lack thereof) ([Polvani and Waugh 2004](#); [Jucker 2016](#); [de la Cámara et al. 2017](#)).

Although the short-term predictability of SSWs has improved over time ([Mukougawa and Hirooka 2004](#); [Stan and Straus 2009](#); [Tripathi et al. 2016](#); [Taguchi 2016](#); [Rao et al. 2018](#)), there remains considerable uncertainty in GCMs’ ability to predict the frequency of SSWs and its trend under a changing climate ([Labitzke and Naujokat 2000](#); [Charlton-Perez et al. 2008](#); [McLandress and Shepherd 2009](#); [Ayarzagüena et al. 2018](#)). This may be due partly to variations in the definition of SSW ([Butler et al. 2015](#)), but more fundamentally, our understanding of SSW’s onset is still incomplete, limiting our ability to interpret the disparate model results. Stationary waves in the stratosphere sometimes attain large amplitude and yet do not produce a full-blown SSW; other times they do ([Solomon 2014](#)). When an SSW does occur, deceleration and reversal of the

zonal-mean zonal wind proceed swiftly, exhibiting a sense of “suddenness.” What separates SSW and non-SSW conditions and why is vortex breakdown abrupt? The resonance paradigm partially addresses this question but supportive evidence from observation is still circumstantial at best (e.g., [Smith 1989](#); [Esler et al. 2006](#)).

In what follows, we propose a view that the timing and suddenness of SSWs are determined, on average, by a threshold behavior of Rossby waves after they attain finite-amplitude wave activity flux in the vertical, regardless of how they attain it. Based on the observed finite-amplitude wave activity [FAWA; [Nakamura and Zhu \(2010\)](#), hereafter [NZ10](#), [Nakamura and Solomon \(2010\)](#), hereafter [NS10](#)] and its vertical flux, we present evidence that there is an upper bound on the upward E–P flux (“transmission capacity”) of a stationary Rossby wave for a given altitude and flow condition. A rapid, spontaneous vortex breakdown occurs once the incident E–P flux from below reaches this capacity. As we will see, this happens when the zonal-mean zonal wind  $\bar{u}$  is decelerated to a certain fraction of the zonal wind associated with a hypothetical wave-free reference state  $u_{\text{REF}}$  (to be defined below). We thus propose the ratio  $r \equiv \bar{u}/u_{\text{REF}}$  as a local, instantaneous measure of the proximity to vortex breakdown. This ratio reveals the threshold behavior of SSW and proves useful for quantifying preconditioning of the stratosphere.

The dynamics underlying the threshold behavior is akin to that discussed by [Wang and Fyfe \(2000\)](#), who use a Boussinesq contour dynamics model of the polar vortex to show that the onset of wave breaking occurs once the zonal-mean zonal wind drops below about one-half of the initial value. However, unlike Wang and Fyfe, our diagnostic applies to instantaneous data without a need for solving an initial-value problem. We will also construct an idealized 1D model of SSWs that encapsulates the essence of the threshold behavior. Mathematically similar to the “traffic jam” model of [Nakamura and Huang \(2018\)](#) for atmospheric blocking, this model predicts salient features of SSWs including a rapid vortex breakdown and a downward migration of wave activity/zonal wind anomalies that follows.

In the next section we will review the diagnostic formalism and describe wave-mean flow interaction observed during the life cycles of SSWs in terms of FAWA,  $\bar{u}$ , and  $u_{\text{REF}}$ . [Section 3](#) explains how wave-mean flow interaction makes the vertical component of E–P flux a nonlinear function of FAWA and how this leads to a threshold behavior. We will demonstrate the threshold behavior in response to transient wave forcing with an idealized, semiempirical 1D model of SSW based on the WKB theory. In [section 4](#) we will examine the effects of

TABLE 1. List of symbols.

Symbol	Description
$t$ (s)	Time
$\lambda \in [0, 2\pi]$	Longitude
$\phi \in [-0.5\pi, 0.5\pi]$	Latitude
$\mu = \sin\phi \in [-1, 1]$	Sine of latitude
$\phi_e \in [-0.5\pi, 0.5\pi]$	Equivalent latitude [Eq. (2)]
$z \equiv -H \ln(p/p_0)$ (m)	Pressure pseudoheight
$H = 7 \times 10^3$ m	Density scale height
$p$ (hPa)	Pressure
$p_0 = 1000$ hPa	Reference pressure
$a = 6.378 \times 10^6$ m	Radius of Earth
$R = 287 \text{ J K}^{-1} \text{ kg}^{-1}$	Gas constant
$c_p = 1004 \text{ J K}^{-1} \text{ kg}^{-1}$	Specific heat at constant pressure
$\kappa = R/c_p$	
$u(\lambda, \phi, z, t)$ ( $\text{m s}^{-1}$ )	Zonal wind (positive eastward)
$v(\lambda, \phi, z, t)$ ( $\text{m s}^{-1}$ )	Meridional wind (positive northward)
$\theta(\lambda, \phi, z, t)$ (K)	Potential temperature
$\theta_0(z, t)$ (K)	Area- (hemispheric-) average $\theta$ (Time dependence is virtually negligible.)
$\Omega = 7.29 \times 10^{-5} \text{ s}^{-1}$	Rotation rate of Earth
$f = 2\Omega \sin\phi$ ( $\text{s}^{-1}$ )	Coriolis parameter
$N^2(z)$ ( $\text{s}^{-2}$ )	Square of Brunt–Väisälä frequency
$q(\lambda, \phi, z, t) = (1/a \cos\phi) \{ (\partial v / \partial \lambda) - [\partial(u \cos\phi) / \partial \phi] \}$ $+ f[1 + e^{z/H}(\partial/\partial z)[e^{-z/H}(\theta - \theta_0)/(\partial\theta_0/\partial z)]]$ ( $\text{s}^{-1}$ )	Potential vorticity
$\overline{(\dots)} = \frac{1}{2\pi} \int_0^{2\pi} (\dots) d\lambda$	Zonal mean
$(\dots)' = (\dots) - \overline{(\dots)}$	Eddy
$A(\phi_e, z, t)$ ( $\text{m s}^{-1}$ )	Finite-amplitude wave activity [Eq. (1)]
$u_{\text{REF}}(\phi, z, t)$ ( $\text{m s}^{-1}$ )	Zonal wind of a wave-free reference-state [Eq. (7)]
$q_{\text{REF}}(\phi, z, t)$ ( $\text{s}^{-1}$ )	Potential vorticity of a wave-free reference-state
$dS = a^2 \cos\phi d\lambda d\phi$ ( $\text{m}^2$ )	Spherical area element
$\varepsilon = \{(f^2 a^2 H) / [R(1 - \mu^2)]\} [e^{\kappa z/H} / (\partial\theta_0/\partial z)]$ ( $\text{m}^2$ )	$\approx a^2(f^2/N^2)/(1 - \mu^2)$
$S$ ( $\text{m s}^{-2}$ )	Frictional and nonadiabatic sources-sinks of $A$
$\bar{X}$ ( $\text{m s}^{-2}$ )	Frictional sources-sinks of $\bar{u}$
$\bar{\theta}$ ( $\text{K s}^{-1}$ )	Nonadiabatic sources-sinks of $\bar{\theta}$
$\alpha$	Ratio of $1 - \bar{u}/u_{\text{REF}}$ to $A/u_{\text{REF}}$
$n > 0$	Integer zonal wavenumber
$k = n/a$ ( $\text{m}^{-1}$ )	Equatorial zonal wavenumber
$l$ ( $\text{m}^{-1}$ )	Meridional wavenumber in the Mercator coordinate
$L$ (m)	Meridional scale of wave activity $A$ [Eq. (14)]
$m$ ( $\text{m}^{-1}$ )	Vertical wavenumber [Eq. (A9)]
$c_{gz}$ ( $\text{m s}^{-1}$ )	Vertical component of group velocity [Eq. (17)]
$F_z$ ( $\text{m}^2 \text{ s}^{-2}$ )	Vertical component of wave activity (E–P) flux

radiative damping and mixing on the threshold behavior. In section 5 we present an observational support for the theory and demonstrate the utility of  $r \equiv \bar{u}/u_{\text{REF}}$  as a diagnostic for preconditioning and its distinct behaviors during winters with and without SSWs. Summary and a discussion follow in section 6.

## 2. Wave-mean flow interaction during SSW life cycles

### a. FAWA and reference state

In this study, we quantify the amplitude of Rossby waves using FAWA, defined as the meridional displacement of

quasigeostrophic potential vorticity (PV) from equivalent latitude  $\phi_e$  [e.g., Eq. (4) of NS10]:

$$A(\phi_e, z, t) \equiv \frac{1}{2\pi a \cos\phi_e} \left( \int_{D_1} q dS - \int_{D_2} q dS \right), \quad (1)$$

where  $q(\lambda, \phi, z, t)$  is PV (see Table 1 for the list of symbols) and  $dS = a^2 \cos\phi d\lambda d\phi$ . Domains  $D_1$  and  $D_2$  and equivalent latitude  $\phi_e$  are defined by

$$D_1: q \geq Q(\phi_e, z, t); \quad D_2: \phi \geq \phi_e, \quad (2)$$

where  $Q(\phi_e, z, t)$  specifies the PV contour that encloses the same area as the polar cap bounded by  $\phi = \phi_e$ , that

is, the areas of  $D_1$  and  $D_2$  are the same [see Fig. 1 of NS10]. By definition FAWA  $A(\phi_e, z, t)$  is nonnegative and vanishes only when the PV contour is zonally symmetric.

The zonal-mean zonal wind  $\bar{u}(\phi, z, t)$  and FAWA  $A(\phi_e, z, t)$  are related through the following relation, assuming  $\phi_e = \phi$  [e.g., Eq. (10) of NS10 and Eq. (3) of Nakamura and Solomon (2011), hereafter NS11]:

$$\begin{aligned} & \left[ \frac{\partial^2}{\partial \mu^2} + e^{z/H} \frac{\partial}{\partial z} \left( \varepsilon e^{-z/H} \frac{\partial}{\partial z} \right) \right] \left[ \frac{\partial}{\partial t} \left( \frac{\bar{u} \cos \phi}{f} \right) \right] \\ &= - \frac{\partial^2}{\partial \mu^2} \left[ \frac{\partial}{\partial t} \left( \frac{A \cos \phi}{f} \right) - \frac{S \cos \phi}{f} - \frac{\bar{X} \cos \phi}{f} \right] \\ & \quad - e^{z/H} \frac{\partial^2}{\partial \mu \partial z} \left( \frac{a e^{-z/H} \bar{\theta}}{\partial \theta_0 / \partial z} \right), \end{aligned} \quad (3)$$

where  $\mu \equiv \sin \phi$  and

$$\varepsilon = \frac{f^2 a^2 H}{R(1 - \mu^2)} \frac{e^{\kappa z/H}}{\frac{\partial \theta_0}{\partial z}}, \quad (4)$$

whereas  $S$ ,  $\bar{X}$  and  $\bar{\theta}$  denote frictional and nonadiabatic sources-sinks of  $A$ ,  $\bar{u}$ , and  $\bar{\theta}$ , respectively. According to Eq. (3), if the dynamics is conservative ( $S = \bar{X} = \bar{\theta} = 0$ ) and wave activity is steady ( $\partial A / \partial t = 0$ ),  $\bar{u}$  does not change provided that the boundary conditions are also time independent. This is the non-acceleration theorem formulated by Charney and Drazin (1961) and Andrews and McIntyre (1976) for small-amplitude waves and later generalized to finite amplitude by NZ10.

NZ10 and NS11 also define a wave-free, reference-state wind  $u_{\text{REF}}(\phi, z, t)$  that satisfies the following:

$$\begin{aligned} & \left[ \frac{\partial^2}{\partial \mu^2} + e^{z/H} \frac{\partial}{\partial z} \left( \varepsilon e^{-z/H} \frac{\partial}{\partial z} \right) \right] \left\{ \frac{\partial}{\partial t} \left[ \frac{(\bar{u} - u_{\text{REF}}) \cos \phi}{f} \right] \right\} \\ &= - \frac{\partial^2}{\partial \mu^2} \left[ \frac{\partial}{\partial t} \left( \frac{A \cos \phi}{f} \right) \right], \end{aligned} \quad (5)$$

$$\begin{aligned} & \left[ \frac{\partial^2}{\partial \mu^2} + e^{z/H} \frac{\partial}{\partial z} \left( \varepsilon e^{-z/H} \frac{\partial}{\partial z} \right) \right] \left[ \frac{\partial}{\partial t} \left( \frac{u_{\text{REF}} \cos \phi}{f} \right) \right] \\ &= \frac{\partial^2}{\partial \mu^2} \left( \frac{S \cos \phi}{f} + \frac{\bar{X} \cos \phi}{f} \right) \\ & \quad - e^{z/H} \frac{\partial^2}{\partial \mu \partial z} \left( \frac{a e^{-z/H} \bar{\theta}}{\partial \theta_0 / \partial z} \right). \end{aligned} \quad (6)$$

Note that summing Eqs. (5) and (6) recovers Eq. (3). The reference-state wind  $u_{\text{REF}}$  is related to the reference-state

PV  $q_{\text{REF}}(\phi, z, t)$ , which is obtained by zonally averaging the instantaneous PV ( $q$ ) through area-preserving map on each  $z$  surface (NZ10, NS10):

$$\begin{aligned} & \frac{\partial}{\partial \mu} \left[ \frac{1}{f} \frac{\partial}{\partial \mu} (u_{\text{REF}} \cos \phi) \right] + e^{z/H} \frac{\partial}{\partial z} \left[ \frac{\varepsilon e^{-z/H}}{f} \frac{\partial}{\partial z} (u_{\text{REF}} \cos \phi) \right] \\ &= -a \frac{\partial}{\partial \mu} \left( \frac{q_{\text{REF}}}{f} \right). \end{aligned} \quad (7)$$

Notice that the time derivative of Eq. (7) is nearly equal to Eq. (6).<sup>1</sup> Since  $q_{\text{REF}}$  changes only in response to nonconservative processes,  $u_{\text{REF}}$  is invariant under conservative dynamics [Eq. (6), see Methven and Berrisford (2015) for a related discussion]. In what follows,  $u_{\text{REF}}$  is inverted hemispherically from  $q_{\text{REF}}$  with Eq. (7) with no-slip lower boundary condition [see supplementary material of Nakamura and Huang (2018)].

#### b. Reversible and irreversible components of wave-mean flow interaction

Equations (5) and (6) partition wave-mean flow interaction into reversible (integrable) and irreversible components. In particular, by integrating Eq. (5) in time (and assuming  $\bar{u} = u_{\text{REF}}$  when  $A = 0$ ) one obtains

$$\begin{aligned} & \left[ \frac{\partial^2}{\partial \mu^2} + e^{z/H} \frac{\partial}{\partial z} \left( \varepsilon e^{-z/H} \frac{\partial}{\partial z} \right) \right] \left[ \frac{(\bar{u} - u_{\text{REF}}) \cos \phi}{f} \right] \\ &= - \frac{\partial^2}{\partial \mu^2} \left( \frac{A \cos \phi}{f} \right). \end{aligned} \quad (8)$$

Thus, the departure of  $\bar{u}$  from  $u_{\text{REF}}$ ,  $\Delta u \equiv \bar{u} - u_{\text{REF}}$ , is directly related to  $A$ . Their relationship at 60°N and  $z = 32$  km (10.34 hPa) during SSW's life cycle is demonstrated in Fig. 1a for the 2009 event (Harada et al. 2010) [data source: ERA-Interim (Dee et al. 2011)]. Vortex conditions at four stages of this splitting event, labeled A–D, are illustrated in Fig. 2. As the vortex is distorted,  $A$  increases and  $\Delta u$  decreases, following a nearly linear relationship (Fig. 1a, red dots). Note that  $\Delta u$  is always negative, meaning that  $\bar{u}$  is slower than  $u_{\text{REF}}$  and it becomes even more so as the event matures. As the flow recovers after the event (blue dots),  $A$  decreases and  $\Delta u$  increases, tracing closely the linear path of the lead.

<sup>1</sup>The rhs of Eq. (6) is the nonconservative sources and sinks of the rhs of Eq. (7). The slight difference in the first term of Eqs. (6) and (7) arises from the approximation made in deriving Eq. (3)—latitudinal dependence of  $f$  does not cause divergence in the horizontal wind—a benign but necessary assumption to avoid mathematical complication. This discrepancy does not arise in the  $\beta$ -plane approximation.

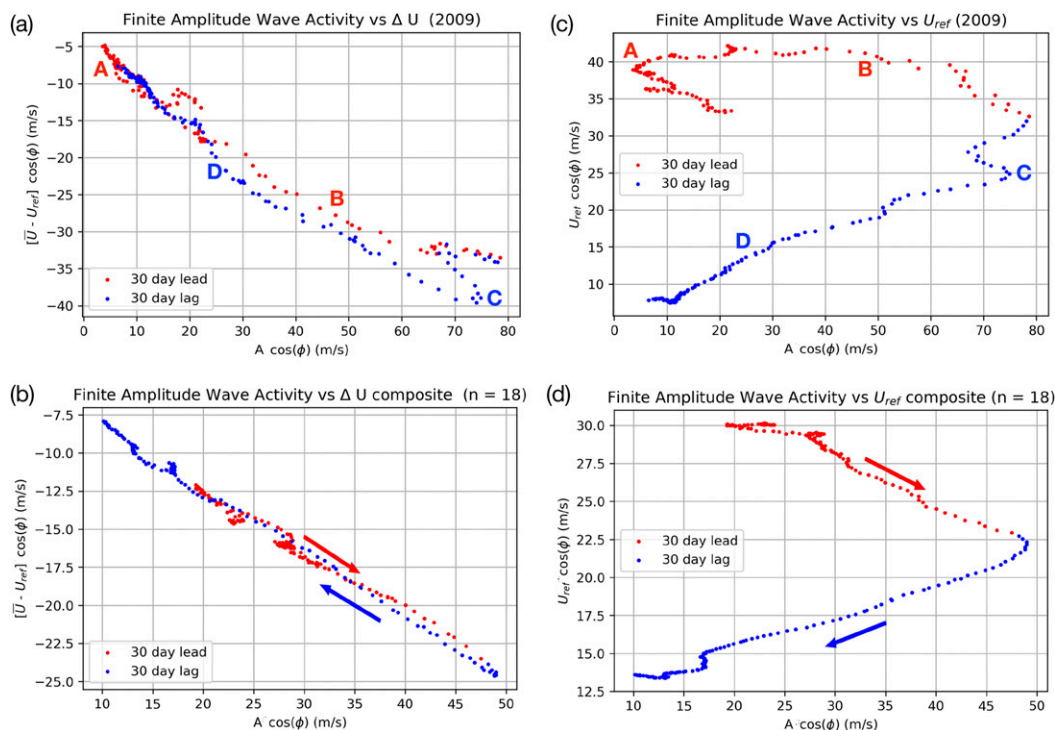


FIG. 1. (a)  $A \cos \phi$  (horizontal axis) vs  $\Delta u \cos \phi$  (vertical axis) at  $60^\circ\text{N}$  and  $z = 32 \text{ km}$  ( $10.34 \text{ hPa}$ ) during the life cycle of SSW event in 2009 (30-day lead plus 30-day lag). The color changes from red to blue when the zonal-mean zonal wind is first reversed. Labels A–D in the panels correspond to the four stages of the SSW illustrated in Fig. 2: A—8 Jan; B—19 Jan; C—28 Jan; D—5 Feb. Data source: ERA-Interim 6-hourly dataset (Dee et al. 2011). (b) As in (a), but composite of 18 SSW events between 1979 and 2016. (See Table 2 for the list of events.) (c), (d) As in (a) and (b), but the vertical axis is  $u_{\text{REF}}$ .

Both  $A$  and  $\Delta u$  almost return to their original values. Figure 1b shows the same diagram for a composite of 18 SSW events between 1979 and 2016. (To obtain a typical picture of SSWs, we have excluded a few late-season events and those that are preceded by another significant event during the 30-day lead time. See Table 2 for the list of the events.) Despite the broad type variations from one event to another (Table 2),

and despite the visually evident irreversibility in Fig. 2, the reversible, linear relationship between  $A$  and  $\Delta u$  is robust through the life cycles of SSWs.

Remarkably, this linear relationship also extends through the column of the stratosphere. Figure 3 shows a composite of  $\bar{u}/u_{\text{REF}}$  versus  $A/u_{\text{REF}}$  during 60-day leads of SSW events on different pressure levels, using both ERA-Interim and MERRA-2 (Gelaro et al. 2017).

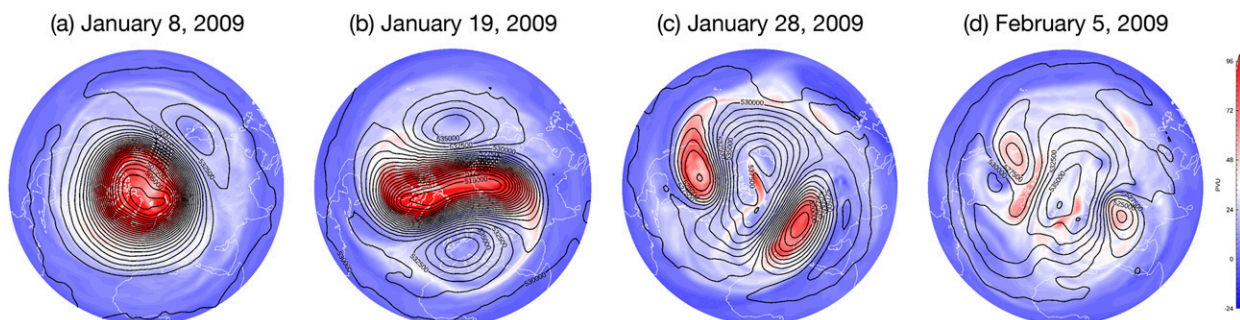


FIG. 2. Potential vorticity (color shading;  $\text{PVU}$ ;  $1 \text{ PVU} = 10^{-6} \text{ K kg}^{-1} \text{ m}^2 \text{ s}^{-1}$ ) and Montgomery streamfunction (contours;  $\text{m}^2 \text{ s}^{-2}$ ) on 850-K isentropic surface (approximately the same altitude as in Fig. 1) during the 2009 SSW: (a) 8 Jan; (b) 19 Jan; (c) 28 Jan; (d) 5 Feb. The four panels correspond to the four stages labeled A–D in Figs. 1a and 1c.



TABLE 2. List of SSWs included and excluded in Figs. 1a, 1b, and 11. SSW type categories are based on Charlton and Polvani (2007) and Kodera et al. (2016).

Included events		
Date	Split (S) or displacement (D)	Absorbing (A) or reflecting (R)
29 Feb 1980	D	R
4 Mar 1981	S	R
4 Dec 1981	D	R
1 Jan 1985	S	R
23 Jan 1987	D	A
8 Dec 1987	S	R
16 Dec 1998	D	R
20 Mar 2000	D	R
11 Feb 2001	D	A
31 Dec 2001	D	R
18 Jan 2003	S	R
6 Jan 2004	D	A
21 Jan 2006	S	A
24 Feb 2007	S	R
22 Feb 2008	D	R
24 Jan 2009	S	A
9 Feb 2010	D	A
6 Jan 2013	S	A
Excluded events		
Date	Reason for exclusion	
24 Feb 1984	Secondary warming	
14 Mar 1988		
21 Feb 1989		
26 Feb 1999	Secondary warming	
17 Feb 2002	Secondary warming	
13 Mar 2008	Secondary warming	
29 Mar 2008	Secondary warming	
24 Mar 2010	Secondary warming	

Data points in both panels cluster around a linear relationship expressed approximately<sup>2</sup>

$$\bar{u} \approx u_{\text{REF}} - \alpha A, \quad \alpha \approx 0.4, \quad \phi = 60^\circ\text{N}. \quad (9)$$

This relation suggests that the rhs of Eq. (8) is approximated between the first and second terms of the lhs at a constant ratio throughout the polar stratosphere.  $\alpha \approx 0.4$  means that about 40% of FAWA is used to drive  $\bar{u}$  away from  $u_{\text{REF}}$ , whereas the remainder is used to drive the residual circulation (NS10). Although the local relation [Eq. (9)] cannot be expected directly from Eq. (8) (except for barotropic case in which  $\alpha = 1$ ), given the robust observational support, we will use Eq. (9) as a key empirical constraint in the subsequent theoretical development.

<sup>2</sup> We ignore the slight offset in  $y$  and the variation of  $\alpha$  in data. These do not affect the subsequent results significantly.

In contrast to the reversible nature of  $\Delta u$ , the behavior of  $u_{\text{REF}}$  during SSW is distinctly irreversible. Figures 1c and 1d plot  $u_{\text{REF}} \cos \phi$  against  $A \cos \phi$  for the corresponding periods in Figs. 1a and 1b at  $60^\circ\text{N}$  and  $z = 32$  km. During the 30-day lead,  $u_{\text{REF}}$  remains more or less steady for small values of  $A$ . For larger values of  $A$ ,  $u_{\text{REF}}$  starts to decrease steadily and it does not recover even after  $A$  reverses its course. The steady decline of  $u_{\text{REF}}$  is driven by nonconservative processes [Eq. (6)], and although part of this driving is seasonal radiative forcing, the majority of it comes from mixing (Lubis et al. 2018a,b; Martineau and Son 2015). The irreversible loss of  $u_{\text{REF}}$  through mixing delays the vortex recovery in some SSW events (Fig. 9 of Lubis et al. 2018a).

### 3. Theory of threshold behavior

In the previous section we described how finite-amplitude Rossby waves modify the zonal-mean zonal wind during the life cycles of SSWs. In this section we discuss how this affects the propagation of Rossby waves and gives rise to a threshold behavior, using a semiempirical theory.<sup>3</sup>

#### a. Vertical group velocity of a stationary Rossby wave

Suppose a Rossby wave is propagating in the atmosphere undamped, and its propagation is roughly governed by the linear quasigeostrophic dynamics with WKB approximation.<sup>4</sup> The same theory also governs FAWA at small amplitude under conservative dynamics (NZ10). To maximize the affinity of quasigeostrophic theory for data, we formulate the dispersion relation for a near plane wave in the Mercator coordinate (e.g., section 5b of Hoskins and Karoly 1981). This leads to the following expression for the vertical group velocity for the stationary Rossby wave (see appendix for derivation):

$$c_{gz} = \frac{2f^2}{N^2} \bar{u} \left[ \frac{\frac{\bar{u}}{\cos \phi}}{\left(\frac{1}{a}\right) \left(\frac{\partial \bar{q}}{\partial \phi}\right)} \right] km. \quad (10)$$

Here  $k = n/a$  is equatorial wavenumber and  $n > 0$  is the integer zonal wavenumber, whereas  $m$  is vertical

<sup>3</sup> Semiempirical because observation is used to constrain theory [e.g., Eqs. (9) and (11)].

<sup>4</sup> Slow variation in the wave packet relative to the wave phase is admittedly not always a best approximation in the real atmosphere. Here we use the WKB theory as a guide and verify its outcome with observed data in section 5.

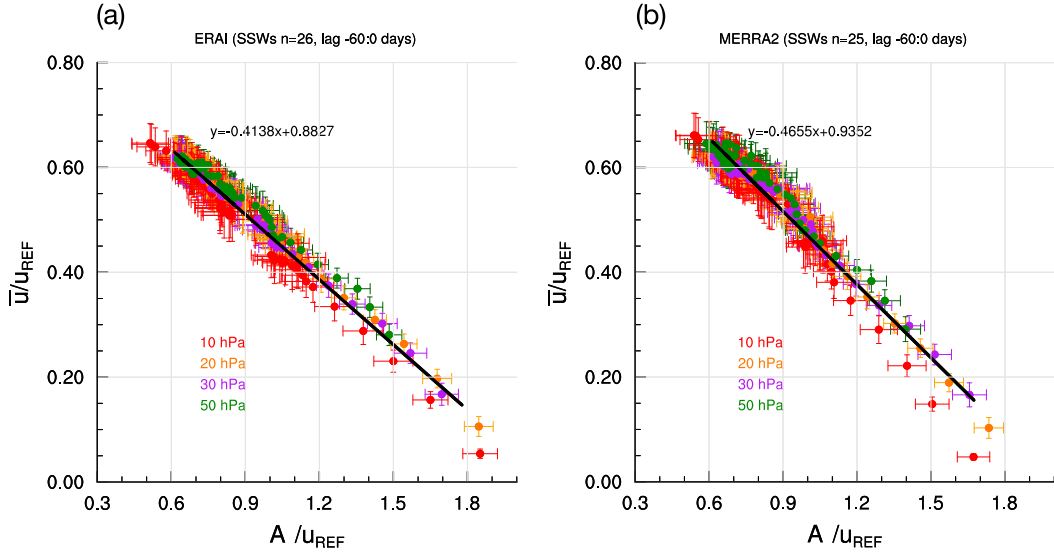


FIG. 3. Relationship between  $A/u_{\text{REF}}$  (abscissa) and  $\bar{u}/u_{\text{REF}}$  (ordinate) at  $\phi = 60^\circ\text{N}$  during 60 days leading to SSW events. Red: 10 hPa. Orange: 20 hPa. Purple: 30 hPa. Green: 50 hPa. To emphasize the average structure, error bars are limited to  $\pm 0.25$  standard deviations. (a) ERA-Interim (Dee et al. 2011) (daily composite of 26 cases). (b): MERRA-2 (Gelaro et al. 2017) (daily composite of 25 cases). The event generally progresses from upper left to lower right. See Table 1 of Lubis et al. (2018a) for the dates of SSWs.

wavenumber given by Eq. (A9). We assume a positive root of  $m$  and  $\bar{u} > 0$  for a wave packet propagating upward.

In evaluating the rhs of Eq. (10) with data, we find a few factors nearly constant in time and height. For example, using ERA-Interim we find  $N$  nearly constant with height in the stratosphere ( $\approx 2.1 \times 10^{-2} \text{ s}^{-1}$ ). The factor  $(\bar{u}/\cos\phi)/[(1/a)(\partial\bar{q}/\partial\phi)]$  involves two zonal-mean quantities that are both time-dependent, but their ratio proves fairly steady even during SSWs. Figure 4 plots  $\bar{u}/\cos\phi$  against  $(1/a)(\partial\bar{q}/\partial\phi)$  at  $60^\circ\text{N}$  in the lower stratosphere during 60 days leading up to SSW events using ERA-Interim (composite of 26 events) and MERRA-2 (25 events). Although both quantities vary significantly, they form an approximate linear relationship and the slope is remarkably similar at all levels, suggesting that

$$\frac{\bar{u}}{\cos\phi} \approx 8.4 \times 10^{11} \text{ m}^2, \quad \phi = 60^\circ\text{N} \quad (11)$$

is a reasonable first approximation. Apart from Eq. (9) and the constant  $N$ , this is the only other empirical constraint we use to simplify the theory.

The near constancy of this ratio can be used to infer the value of meridional wavenumber  $l$  [and through Eq. (A9),  $m$ ] as follows. Invariance of Eq. (11) through wave-mean flow interaction suggests

$$\frac{\bar{u}}{\cos\phi} \approx \frac{u_{\text{REF}}}{\cos\phi} \cdot \left( \frac{1/a}{\partial\bar{q}/\partial\phi} \right) \approx \left( \frac{1/a}{\partial q_{\text{REF}}/\partial\phi} \right). \quad (12)$$

With Eq. (9) obtained earlier and Eq. (19) of NZ10 (replacing the meridional coordinate with  $a\mu$ ),<sup>5</sup> the lhs of Eq. (12) may be rewritten as

$$\frac{u_{\text{REF}} - \alpha A}{\cos\phi} \approx \frac{u_{\text{REF}}}{\cos\phi} \cdot \left( \frac{1/a}{\partial q_{\text{REF}}/\partial\phi} \right). \quad (13)$$

If we assume a negative curvature in  $A$  with respect to latitude:

$$\frac{\cos^2\phi}{a^2} \frac{\partial^2 A}{\partial\mu^2} \approx -\frac{A}{L^2}, \quad (14)$$

then from Eqs. (13), (14), and (11):

$$L^2 = \frac{1}{\alpha} \frac{u_{\text{REF}}}{\cos\phi} \approx \frac{1}{\alpha} \frac{\bar{u}}{\cos\phi} \approx 2.1 \times 10^{12} \text{ m}^2, \quad (15)$$

<sup>5</sup> Equation (19) of NZ10 relates the PV gradient of the reference state with that of the instantaneous zonal-mean state through the curvature of FAWA.

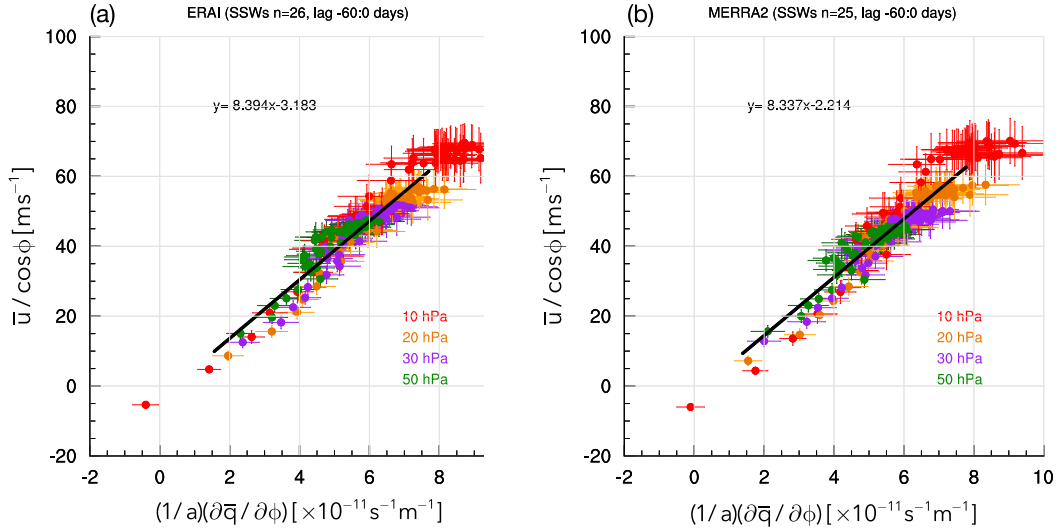


FIG. 4. Relationship between  $(1/a)(\partial\bar{q}/\partial\phi)$  (abscissa) and  $\bar{u}/\cos\phi$  (ordinate) at  $\phi = 60^\circ\text{N}$  during 60 days leading up to SSW events. Red: 10 hPa. Orange: 20 hPa. Purple: 30 hPa. Green: 50 hPa. Error bars indicate  $\pm 0.25$  standard deviation. (a) ERA-Interim (daily composite of 26 cases). (b) MERRA2 (daily composite of 25 cases). The event generally progresses from upper right to lower left.

where we assumed  $\alpha = 0.4$  in the last approximation. This gives the meridional scale of FAWA,  $L \approx 1.45 \times 10^6 \text{ m}$ , at  $60^\circ\text{N}$ . Then assuming that  $q'$  is modal in latitude and proportional to  $\sin y$  [ $y$  is Mercator latitude; Eq. (A1)],<sup>6</sup>

$$l = \frac{\cos 60^\circ}{2L} \approx 1.73 \times 10^{-7} \text{ m}^{-1}, \quad (16)$$

which gives, from Eq. (A9),  $m = 2.36 \times 10^{-4} \text{ m}^{-1}$  for  $n = 1$  and  $2.18 \times 10^{-4} \text{ m}^{-1}$  for  $n = 2$ . Substituting Eq. (15) into Eq. (10), we obtain the vertical group velocity of the stationary Rossby wave as

$$c_{gz} = C\bar{u}, \quad C = \frac{2f^2}{N^2} \alpha L^2 \text{ km}. \quad (17)$$

The nondimensional coefficient  $C$  is the ratio of  $c_{gz}$  to  $\bar{u}$ . Using the values obtained from the foregoing analysis, we find that  $C$  is nearly independent of height and  $C \approx 2.23 \times 10^{-3}$  for  $n = 1$  and  $\approx 4.11 \times 10^{-3}$  for  $n = 2$  at  $60^\circ\text{N}$ . Equation (17) highlights the proportionality between the group velocity and  $\bar{u}$ . Since the latter is modified by  $A$  according to Eq. (9),

$$c_{gz} \approx C(u_{\text{REF}} - \alpha A), \quad (18)$$

<sup>6</sup> In the small-amplitude limit,  $A$  is a quadratic function of  $q'$  so the meridional wavenumber of  $q'$  should be half of that of  $A$ . Furthermore, from Eq. (A2), wavenumber in the Mercator coordinate is  $\cos\phi$  times the wavenumber in the local Cartesian coordinate.

that is, the group velocity becomes a function of FAWA through wave–mean flow interaction.

#### b. Wave activity budget and threshold

Under the WKB approximation, the wave activity equation for a vertically propagating wave packet reads<sup>7</sup>

$$\frac{\partial}{\partial t}(Ae^{-z/H}) = -\frac{\partial}{\partial z}(c_{gz}Ae^{-z/H}), \quad (19)$$

where  $c_{gz}Ae^{-z/H} \equiv F_z$  is the vertical component of wave activity (E–P) flux density. Here the convergence of the meridional wave activity flux and nonconservative terms are ignored. These omissions would make Eq. (19) a poor approximation for describing the full life cycle of SSW (e.g., O'Neill and Taylor 1979; Lubis et al. 2018a,b). Yet there is mounting evidence that early development of SSW is dictated by the upward E–P flux and its convergence (Polvani and Waugh 2004; Jucker 2016), so we take Eq. (19) as a model for the SSW's onset. [Alternatively, Eq. (19) may be thought of as the budget of meridionally integrated wave activity (e.g., Plumb and Semeniuk 2003).] With Eq. (18) we have

$$F_z = c_{gz}Ae^{-z/H} = C(u_{\text{REF}} - \alpha A)Ae^{-z/H} \quad (20)$$

or equivalently

<sup>7</sup>  $\partial/\partial T$  and  $\partial/\partial Z$  have been rescaled to  $\partial/\partial t$  and  $\partial/\partial z$  using  $(Z, T) = (\gamma z, \gamma t)$ ,  $\gamma \ll 1$ , and the  $\cos\phi$  factor is dropped.



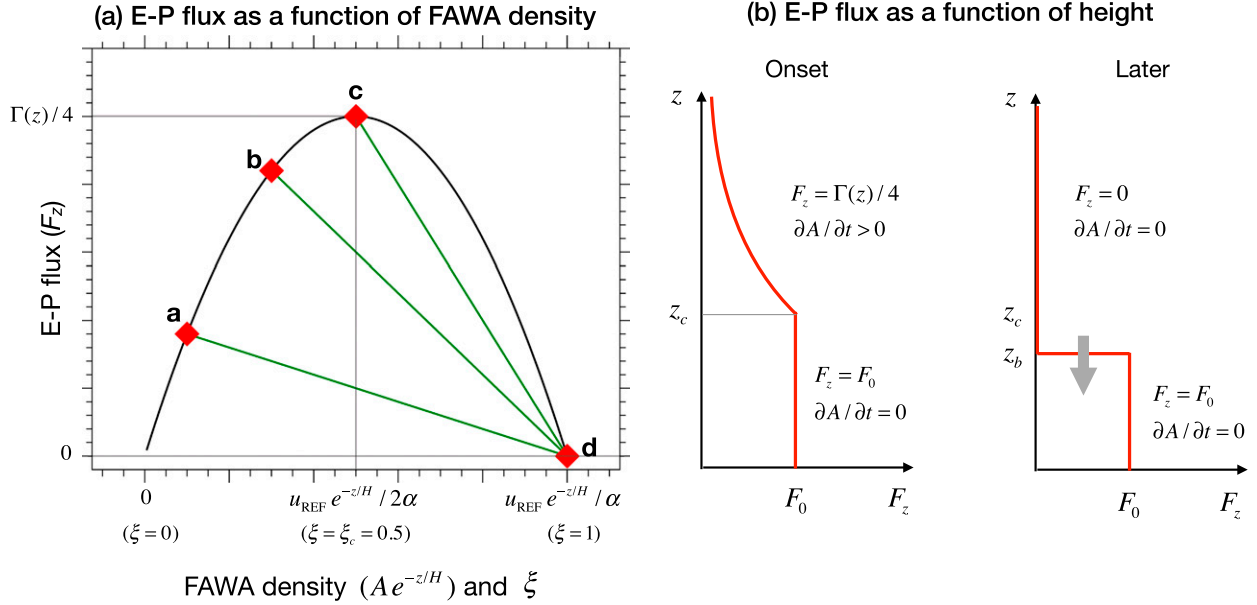


FIG. 5. (a) Schematic diagram for the relationship between FAWA density ( $Ae^{-z/H}$ ; horizontal axis) and E-P flux ( $F_z$ ; vertical axis) at a given altitude. The corresponding values of  $\xi$  are also labeled on the horizontal axis. For  $\xi < \xi_c = 0.5$  (“a” and “b”), a stable steady state is possible. Once  $\xi$  exceeds  $\xi_c$ , the state moves quickly from “c” to “d” (i.e., the state of vortex breakdown), through a positive feedback. The slopes of the line segments provide the migration speeds of the resulting shocks. (b) Schematic diagrams for the vertical profile of  $F_z$ . (left) At the onset of SSW,  $F_z$  above  $z_c$  is in state “c” in (a). (right) After SSW,  $F_z$  above  $z_b$  is in state “d” in (a). Because of the gap in the flux,  $z_b$  gradually descends. See text for details.

$$F_z = \Gamma(z)(1 - \xi)\xi, \quad \Gamma(z) \equiv \frac{Cu_{\text{REF}}^2(z)e^{-z/H}}{\alpha}, \quad \xi \equiv \frac{\alpha A}{u_{\text{REF}}}. \quad (21)$$

Because  $c_{gz}$  is a linear function of  $A$ ,  $F_z$  is a quadratic function of  $A$ . We will compare this theoretical prediction of  $F_z$  with observation in section 5. For now we demonstrate how the nonlinear flux of FAWA brings about a threshold behavior.

Suppose a wave packet is steady so the lhs of Eq. (19) vanishes. Then the wave activity flux must be constant with height ( $=F_0 > 0$  for upward propagation):

$$F_0 = \Gamma(z)(1 - \xi)\xi. \quad (22)$$

The flux  $F_0$  may be determined, for example, by a steady wave forcing at the lower boundary. Solving Eq. (22) for a steady-state  $\xi$ :

$$\xi(z) = \frac{1}{2} \left[ 1 - \sqrt{1 - \frac{4F_0}{\Gamma(z)}} \right] \quad \text{if } F_0 \leq \frac{\Gamma(z)}{4}. \quad (23)$$

The other root for  $\xi$  is unstable and hence unrealizable (we will see why shortly). The discriminant condition in Eq. (23) means that there is a height-dependent upper bound on the wave activity flux that a steady stationary

wave can transmit. This places a “transmission capacity” for finite-amplitude Rossby waves, whose vertical transmission is guaranteed at small amplitude (Charney and Drazin 1961). For a given  $F_0$  and  $\Gamma(z)$ , Eq. (23) predicts the altitude at which this upper bound is first reached. Because of the exponential factor,  $\Gamma(z)$  is a decreasing function of  $z$  at sufficiently high altitude [Eq. (21)], and therefore even if  $F_0 < \Gamma(z)/4$  at low altitudes,  $F_0$  eventually equals  $\Gamma/4$  at some  $z$ . This altitude  $z_c$  is lower for larger  $F_0$ . At this altitude, the following conditions are met:

$$F_0 = \frac{\Gamma(z_c)}{4}, \quad \xi = \xi_c = \frac{1}{2}, \quad \frac{\bar{u}}{u_{\text{REF}}} = 1 - \xi = \frac{1}{2}, \quad z = z_c. \quad (24)$$

Note that the zonal-mean zonal wind is decelerated to one-half of  $u_{\text{REF}}$  at  $z = z_c$ , and this criterion does not depend on the value of  $\alpha$  as long as it is positive. The above conditions correspond to the vertex of the quadratic curve in Fig. 5 (point “c”) touching the line  $F_z = F_0$ . For  $z > z_c$ , the vertex will be lower than  $F_0$ , so Eq. (22) has no real root. To ensure a real value for  $\xi$ ,  $F_z$  must be capped at the transmission capacity:

$$F_z(z) = \frac{\Gamma(z)}{4} < F_0, \quad \xi = \xi_c = \frac{1}{2}, \quad z > z_c. \quad (25)$$

However, this means  $\partial F_z / \partial z < 0$  above  $z_c$  (see the left panel of Fig. 5b), which is incompatible with the steady-state assumption:  $A$  (and hence  $\xi$ ) must increase with time according to Eq. (19). From Eqs. (19), (25), and (21), the growth rate of wave activity is estimated to be

$$-\frac{1}{Ae^{-z/H}} \frac{\partial F_z}{\partial z} = -\frac{1}{\frac{u_{\text{REF}}}{2\alpha} e^{-z/H}} \frac{d}{dz} \left[ \frac{\Gamma(z)}{4} \right] = \frac{C}{2} \left( \frac{u_{\text{REF}}}{H} - 2 \frac{\partial u_{\text{REF}}}{\partial z} \right), \quad z > z_c. \quad (26)$$

For typical midstratospheric values, the first term in the last expression dominates and it gives an  $e$ -folding time of 1.5–3 days, consistent with the suddenness of SSW. The growth of  $\xi$  beyond  $\xi_c$  causes  $F_z$  to decrease further, since  $F_z$  is a decreasing function of  $\xi$  for  $\xi > \xi_c$  [Eq. (21), Fig. 5a]. This sets up a positive feedback between wave activity and its flux: the decreasing flux above the threshold altitude  $z_c$  reinforces the flux convergence and further increase in  $A$ , which then promotes further reduction in  $F_z$ , provided that there is a continued supply of incident flux from below. Therefore, above  $z = z_c$  where the threshold (24) is first met, wave activity increases rapidly until it reaches

$$\xi = 1; \quad \bar{u} = F_z = 0, \quad (27)$$

a state marked “d” in Fig. 5a, which is the only stable steady solution of Eq. (19) for  $\xi > \xi_c$  and corresponds to “vortex breakdown.”<sup>8</sup> The threshold behavior and spontaneous breakdown of the vortex arises only because the E–P flux is a nonlinear function of  $\xi$  and unimodal.

Note that vortex breakdown starts when  $\bar{u}$  is still positive and finite [Eq. (24)]. Therefore, unlike linear theory (e.g., Dickinson 1969), presence of a critical line ( $\bar{u} = 0$ ) is not required for wave transience. Rather, the zero wind line emerges as a result of the threshold behavior.

### c. Numerical experiments

To demonstrate the transient behavior of  $A$  and  $\bar{u}$  in response to wave forcing, we integrate Eq. (19) numerically with Eq. (18). The computational domain is  $10 \leq z \leq 110$  km and discretized with grid spacing of 100 m. A small numerical diffusion with a diffusion coefficient of  $10 \text{ m}^2 \text{ s}^{-1}$  is added to the rhs to smooth out noise. We assume  $C = 3 \times 10^{-3}$ ,  $H = 7$  km, and  $\bar{u}(z, t)$  is diagnosed

at each time step from  $A$  as  $\bar{u} = u_{\text{REF}}(z) - \alpha A(z, t)$  with  $\alpha = 0.4$ . The initial condition is

$$A(z, 0) = 0, \quad \bar{u}(z, 0) = u_{\text{REF}}(z), \quad (28)$$

$$u_{\text{REF}}(z) = \begin{cases} 0.0015 \text{ s}^{-1} z + 5 \text{ m s}^{-1} & 10 \leq z \leq 50 \text{ km} \\ 80 \text{ m s}^{-1} & 50 \text{ km} < z. \end{cases} \quad (29)$$

Thus,  $u_{\text{REF}}$  increases linearly from  $20 \text{ m s}^{-1}$  at  $z = 10$  km to  $80 \text{ m s}^{-1}$  at  $z = 50$  km and remains  $80 \text{ m s}^{-1}$  above 50 km. This is broadly consistent with the climatology of  $u_{\text{REF}}$  at the beginning of the winter (see Fig. 12 below). At the lower boundary ( $z = 10$  km), wave forcing is prescribed as

$$A(10 \text{ km}, t) = \Delta A \left[ \tanh \left( \frac{t - t_{\text{on}}}{t_0} \right) + 1 \right]; \quad \Delta A = 3.13 \text{ m s}^{-1}, \quad (30)$$

whereas at the top boundary ( $z = 110$  km) a radiation condition is assumed.

In the first experiment, we turn on the wave forcing around day 10 ( $t_{\text{on}} = 10$  days) relatively quickly ( $t_0 = 6$  h) and leave it on. The result is shown in the top row of Fig. 6 in height–time cross sections for  $z \leq 50$  km. After forcing is turned on, a trace of wave activity moves up and quickly amplifies toward the top of the stratosphere (Fig. 6a). During days 12–15,  $A$  undergoes a rapid increase in the upper stratosphere, whereas  $\bar{u}$  shows a precipitous drop (Fig. 6b). A new quasi-steady state emerges, characterized by a large, height-dependent  $A$  and zero  $\bar{u}$  (“vortex breakdown”). Notice that the transition to the new state is nearly simultaneous between 30 and 50 km with little time lag across altitudes (Figs. 6a,b). Subsequently the transition progresses to lower altitudes, causing the vortex terminus to descend rapidly at first and then more gradually, until it reaches the base of the stratosphere around day 28. The descending edge of the vortex is very sharp and the intense negative vertical shear across the edge implies, by virtue of thermal wind balance, a sudden and intense reversal of the meridional temperature gradient. This simple model thus captures two common aspects of SSWs: rapid breakdown of vortex and a gradual descent of zonal-mean zonal wind anomaly (Fig. 4 of Matsuno 1971; Baldwin and Dunkerton 2001).

The progression of vortex breakdown is understood as follows. Soon after the wave forcing is turned on,  $F_z$  reaches a steady value of  $F_0 = 7.87 \times 10^{-2} \text{ m}^2 \text{ s}^{-2}$  in the lower stratosphere. For this value of  $F_0$ , the threshold (24) is reached at  $z_c = 27.5$  km. Therefore, a runaway

<sup>8</sup> This is in essence a critical line where the WKB approximation breaks down formally.

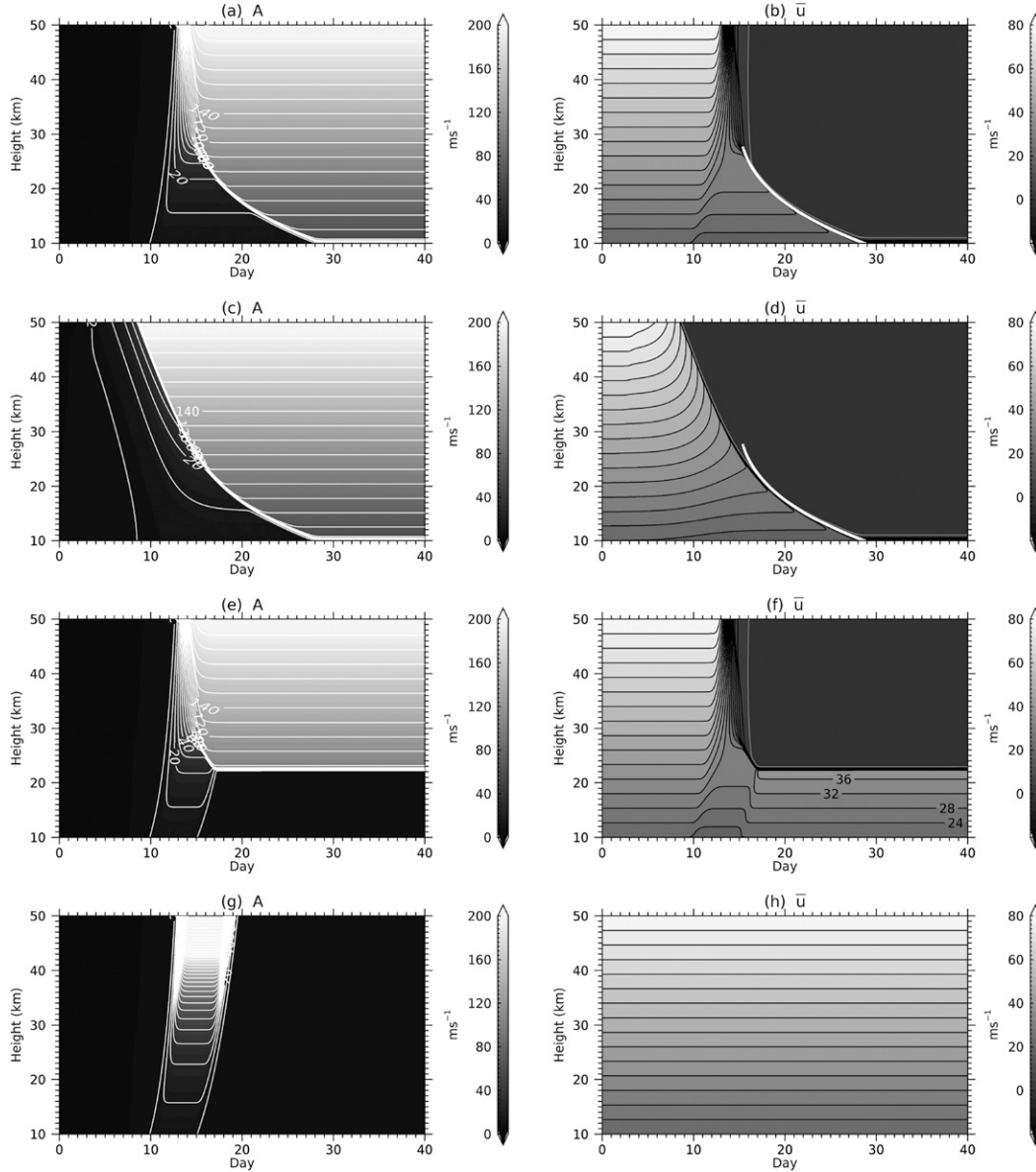


FIG. 6. Numerical solutions of Eq. (19) with transient wave forcing applied at the lower boundary. (a),(c),(e)  $A$ , with contours plotted for 2, 10, 20, ...  $\text{m s}^{-1}$ . (b),(d),(f)  $\bar{u}$ , with a contour interval of  $4 \text{ m s}^{-1}$ . (a),(b)  $t_0 = 6 \text{ h}$ ,  $t_{\text{on}} = 10$  days (“quick spinup”). (c),(d)  $t_0 = 96 \text{ h}$ ,  $t_{\text{on}} = 10$  days (“slow spinup”). (e),(f)  $t_0 = 6 \text{ h}$ ,  $t_{\text{on}} = 10$  days,  $t_{\text{off}} = 15$  days (“short pulse”). (g),(h) As in (e) and (f), but for  $\alpha = 0$ . Contour interval above  $200 \text{ m s}^{-1}$  in (g) is variable. The thin gray contour in (b), (d), and (f) marks  $\bar{u} = 0$ . The white curve in (b) and (d) is the theoretical prediction of the vortex terminus below  $z = z_c$ . See text for details.

flux convergence and vortex breakdown occur above this altitude first, shifting the vortex state from “c” to “d” in Fig. 5a. Subsequently the region of vortex breakdown (a stagnant reservoir of FAWA like critical layer) expands downward by absorbing the incident wave activity flux from below, which causes the vortex terminus to move down (right panel of Fig. 5b). This is akin to a traffic jam on a highway, where the jam expands backward by absorbing the traffic from behind

(Lighthill and Whitham 1955; Richards 1956). Nakamura and Huang (2017, 2018) and Paradise et al. (2019) demonstrate that a similar dynamics is at play in the formation of atmospheric blocking.

If we characterize the vortex terminus as a migrating shock of  $A$  and define its altitude as  $z_b(t)$ , we can estimate its migration speed  $dz_b(t)/dt$  using the Rankine–Hugoniot relation [LeVeque 2002, chapter 3]. Just above  $z_b$ , from Eq. (27):

$$F_z(z_b^+) = 0, \quad A(z_b^+) = \frac{u_{\text{REF}}(z_b)}{\alpha}, \quad (31)$$

whereas just below  $z_b$ , from Eq. (23):

$$F_z(z_b^-) = F_0, \quad A(z_b^-) = \frac{u_{\text{REF}}(z_b)}{2\alpha} \left[ 1 - \sqrt{1 - \frac{4F_0}{\Gamma(z)}} \right]. \quad (32)$$

Then from the Rankine–Hugoniot relation:

$$\begin{aligned} \frac{dz_b}{dt} &= \frac{F_z(z_b^+) - F_z(z_b^-)}{[A(z_b^+) - A(z_b^-)]e^{-z_b/H}} \\ &= -\frac{Cu_{\text{REF}}(z_b)}{2} \left[ 1 - \sqrt{1 - \frac{4F_0}{\Gamma(z)}} \right]. \end{aligned} \quad (33)$$

To derive the last expression, we used Eq. (22). At  $z_b = z_c$  where the shock first forms,  $dz_b/dt = -Cu_{\text{REF}}(z_c)/2$ . This corresponds to the slope of the line segment  $\overline{cd}$  in Fig. 5a. The less steep slopes of segments  $\overline{bd}$  and  $\overline{ad}$  in Fig. 5a correspond to the decreased migration speed below  $z_c$ . We have computed  $z_b(t)$  by numerically integrating Eq. (33) and show it in Fig. 6b. It captures the descent of the vortex boundary very well below  $z_c = 27.5$  km.

When the wave forcing is switched on more slowly ( $t_0 = 96$  h,  $t_{\text{on}} = 10$  days), the result is qualitatively similar but the vortex breakdown occurs earlier and the descent of the vortex terminus is noticeably slower in the upper stratosphere (Figs. 6c,d). In this case, a weak wave activity flux during the early stage of spinup manages to penetrate higher and initiate vortex breakdown from  $z_c > 50$  km. The vortex terminus is very sharp throughout the stratosphere.

Since the WKB theory does not permit wave reflection, the upwelled flux is entirely absorbed into  $z > z_b$ . In reality, we expect some wave reflection to occur at  $z = z_b$ . For example, Killworth and McIntyre (1985) show that a Rossby wave critical layer changes from an absorber to a perfect reflector of waves as it matures. We can crudely model the effects of wave reflection by making the wave forcing shorter, since the reflected (downward) wave packet partially cancels the upward wave activity flux and reduces the persistence of the net wave forcing (Harnik 2009; Koder et al. 2016). In Figs. 6e and 6f, we apply wave forcing only for 5 days. We accomplish this by adding

$$-\Delta A \left[ \tanh\left(\frac{t - t_{\text{off}}}{t_0}\right) + 1 \right], \quad t_{\text{off}} = 15 \text{ days}, \quad t_0 = 6 \text{ h} \quad (34)$$

to Eq. (30). The response of the stratosphere is initially similar to Figs. 6a and 6b, but the vortex breakdown is terminated halfway at around  $z = 22$  km, below which  $A$  and  $\bar{u}$  return to their original state after the forcing is turned off. Once the forcing is off, wave activity flux is zero both below and above  $z_b$ , therefore  $z_b$  remains steady. Greater and longer forcings cause the vortex terminus to reach lower altitudes (not shown).

To highlight the role of wave–mean flow interaction in the threshold behavior, in Figs. 6g and 6h we repeat the same experiment with  $\alpha = 0$ . Without wave–mean flow interaction, the pulse of wave activity propagates upward without modifying  $\bar{u}$ , and no regime transition occurs.

#### 4. Effects of damping

So far we have considered quasi-conservative dynamics, but the role of radiative damping and mixing can be significant during the life cycle of SSW, particularly in the recovery stage (Lubis et al. 2018a,b). To examine the effects of these nonconservative processes, we extend Eq. (19) to

$$\frac{\partial}{\partial t} A = -e^{z/H} \frac{\partial}{\partial z} (c_{gz} A e^{-z/H}) + S, \quad (35)$$

$$\frac{\partial}{\partial t} \bar{u} = \alpha e^{z/H} \frac{\partial}{\partial z} (c_{gz} A e^{-z/H}) + \bar{X}, \quad (36)$$

$$\frac{\partial}{\partial t} u_{\text{REF}} = \alpha S + \bar{X}. \quad (37)$$

Here  $S$  is nonconservative sources-sinks of  $A$ , whereas  $\bar{u}$  is driven by eddy forcing and the zonal-mean friction  $\bar{X}$  [Eq. (36)], and  $c_{gz}$  is given by Eq. (18).<sup>9</sup> Note that  $u_{\text{REF}}$  now varies with time due to nonconservative processes. Note also that Eqs. (35)–(37) ensure

$$\frac{\partial \bar{u}}{\partial t} = \frac{\partial}{\partial t} (u_{\text{REF}} - \alpha A), \quad (38)$$

consistent with  $\bar{u} = u_{\text{REF}} - \alpha A$  [Eq. (9), see also Figs. 9a and 9c below].

##### a. Radiative damping

Here we relax  $(A, \bar{u})$  toward the initial condition  $[0, \bar{u}(z, 0)]$  with radiative time scale  $\tau(z)$ :

$$S_{\text{rad}} = \frac{-A}{\tau(z)}; \quad \bar{X}_{\text{rad}} = -\frac{\bar{u} - \bar{u}(z, 0)}{\tau(z)}. \quad (39)$$

<sup>9</sup> We ignore the effect of damping on the group velocity in this study.

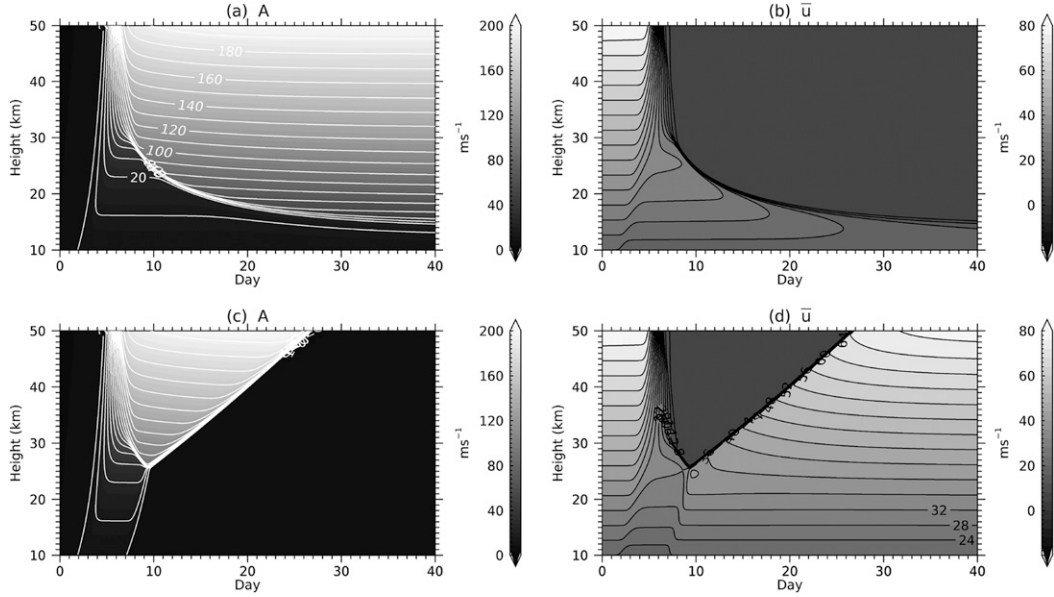


FIG. 7. Numerical solutions of Eqs. (35) and (36) with radiative damping. Transient wave forcing is applied at the lower boundary. (a), (c)  $A$ . (b), (d)  $\bar{u}$ . (a), (b)  $t_0 = 6$  h,  $t_{\text{on}} = 2$  days (quick switch-on). (c), (d)  $t_0 = 6$  h,  $t_{\text{on}} = 2$  days,  $t_{\text{off}} = 7$  days (short pulse). In (b) and (d), the minimum contour level is  $4 \text{ m s}^{-1}$  and the minimum value of  $\bar{u}$  is  $\sim 3.85 \text{ m s}^{-1}$ . See text for details.

In Eq. (39)  $\tau^{-1}$  increases linearly from  $(30 \text{ days})^{-1}$  at  $z = 10 \text{ km}$  to  $(5 \text{ days})^{-1}$  at  $z = 50 \text{ km}$  and remains  $(5 \text{ days})^{-1}$  for  $z > 50 \text{ km}$ . Similar values are used by Holton and Mass (1976) and Plumb and Semeniuk (2003). With (39) and the same diffusion coefficient as before, we solve Eqs. (35)–(37) numerically.

In Figs. 7a and 7b wave forcing is identical to Figs. 6a and 6b except it is turned on earlier ( $t_{\text{on}} = 2$  days,  $t_0 = 6$  h). The subsequent evolution is qualitatively similar to the undamped case, except the descent of the vortex terminus is much slower: 38 days after the forcing is switched on, it is still 4–5 km above the lower boundary, whereas it reached the boundary in less than 20 days in the adiabatic case (Figs. 6a,b). Also, after the vortex breakdown,  $\bar{u}$  does not vanish completely but maintains a small positive value. To understand these behaviors, it is useful to consider the steady state of Eq. (35):

$$0 = -\frac{\partial}{\partial z} [C(u_{\text{REF}} - \alpha A)Ae^{-z/H}] - \frac{Ae^{-z/H}}{\tau(z)}. \quad (40)$$

Since  $\tau^{-1}$  varies in  $z$  much more slowly than  $\exp(-z/H)$ , the approximate solution to Eq. (40) is

$$F_1 \approx C \left[ \left( u_{\text{REF}} - \frac{H}{C\tau} \right) - \alpha A \right] Ae^{-z/H}, \quad (41)$$

where  $F_1$  is independent of height and  $F_1 = 0$  for  $z > z_b$  since the rhs of Eq. (41) vanishes as  $z \rightarrow \infty$ . This leads

to  $\bar{u} = u_{\text{REF}} - \alpha A = H/(C\tau) > 0$ , so  $\bar{u}$  remains finite even after the vortex breaks down. Equation (41) is similar to Eq. (22), and the previous results all hold if we replace  $u_{\text{REF}}$  with  $u_{\text{REF}} - H/(C\tau)$ ,  $F_0$  with  $F_1$ , and  $F_z$  with  $F_z - (H/\tau)Ae^{-z/H}$ . Then the descent rate of the vortex terminus will be roughly proportional to  $u_{\text{REF}} - H/(C\tau)$  in Eq. (33), that is, radiative damping slows down the descent.

When the wave forcing is applied for only 5 days [ $t_{\text{on}} = 2$ ,  $t_{\text{off}} = 7$  days in Eqs. (30) and (34)], not only is the vortex breakdown interrupted halfway, but the vortex recovery proceeds in a discontinuous fashion (Figs. 7c,d). In a stark contrast with Figs. 6c and 6d,  $A$  and  $\bar{u}$  change abruptly across a boundary ( $z = z_b$ ) that moves up at a near constant rate. The upward movement of this “recovery line” is due to a gap in the wave activity flux across it. After the wave forcing is shut off,  $F_z = C\bar{u}Ae^{-z/H} \approx 0$  below  $z_b$  since  $A \approx 0$ . However,  $F_z$  is positive for  $z > z_b$ , since  $\bar{u} \approx H/(C\tau)$  and  $A \approx [u_{\text{REF}} - H/(C\tau)]/\alpha > 0$ . Due to this gap in the flux, the recovery line  $z = z_b$  must move up, and the rate of the movement is given by

$$\begin{aligned} \frac{dz_b}{dt} &= \frac{F_z(z_b^+) - F_z(z_b^-)}{[A(z_b^+) - A(z_b^-)]e^{-z_b/H}} \approx \frac{F_z(z_b^+)}{A(z_b^+)e^{-z_b/H}} \\ &= C\bar{u}(z_b^+) \approx \frac{H}{\tau}. \end{aligned} \quad (42)$$

Therefore the steepness of the recovery line in Figs. 7c and 7d is largely determined by the radiative



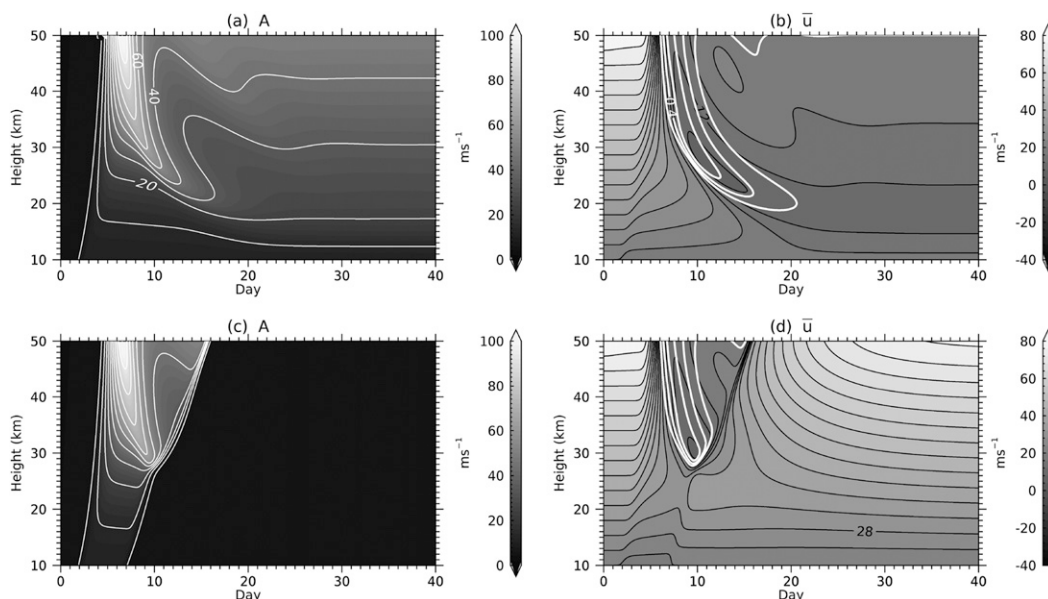


FIG. 8. As in Fig. 7, but including mixing parameterization for  $A$ . (a), (c)  $A$ . (b), (d)  $\bar{u}$ . (a), (b)  $t_0 = 6$  h,  $t_{\text{on}} = 2$  days (quick switch-on). (c), (d)  $t_0 = 6$  h,  $t_{\text{on}} = 2$  days,  $t_{\text{off}} = 7$  days (short pulse). White contours in (b) and (d) show  $\bar{u}/u_{\text{REF}}$  and are plotted for 0.3, 0.4, and 0.5. See text for details.

time scale. For  $H = 7$  km and  $\tau \approx 5$  days, the ascent rate is  $1.47 \text{ km day}^{-1}$ , roughly consistent with Fig. 7c ( $\approx 1.4 \text{ km day}^{-1}$ ).

### b. Mixing

Lubis et al. (2018a) demonstrate that mixing plays a significant role after the peak of SSW events, particularly those with slow recovery. The primary role of mixing is to damp FAWA (NZ10) and delay the vortex recovery (Martineau and Son 2015; Lubis et al. 2018a,b). To incorporate the effects of mixing to the foregoing model, we replace the radiative damping of  $A$  in Eq. (39) with a flow-dependent damping through  $S$ :

$$S = \frac{-A}{\tau_{\text{mix}}(z, t)}; \quad \tau_{\text{mix}}(z, t) = \left[ 0.2 + 0.8 \left( \frac{\bar{u}}{u_{\text{REF}}} \right)^2 \right] \tau(z), \quad (43)$$

where  $\tau(z)$  is the radiative damping time introduced earlier. When  $\bar{u} = u_{\text{REF}}$  ( $A = 0$ ),  $\tau_{\text{mix}}$  recovers the radiative damping time  $\tau$ . When  $\bar{u} = 0$ ,  $\tau_{\text{mix}}$  minimizes to  $0.2\tau$ . The above formula is ad hoc and meant as a crude parameterization of mixing inhibition by the mean flow (e.g., Ferrari and Nikurashin 2010).

When Eq. (43) is used instead of  $S_{\text{rad}}$  in Eq. (39), the numerical results change substantially. (In this set of experiments, we also triple the diffusion coefficient to  $30 \text{ m}^2 \text{ s}^{-1}$ ). Figures 8a and 8d correspond to Figs. 7a and

7d, the only difference being  $S$  used in Eq. (35). In both the switch-on and the short pulse forcings, FAWA is heavily damped by mixing after it reaches maximum in the upper stratosphere (Figs. 8a,d). In fact, in both cases  $A$  survives only for a short duration in a very similar way, suggesting an important role of mixing in determining the duration of a wave event. The main difference between the two forcing types is the response of  $\bar{u}$ . When the wave forcing is maintained after the spinup, a band of negative wind anomalies descends slowly after the vortex breakdown. The wind speed in the upper stratosphere also remains weak after the event, with little sign of recovery (Fig. 8b). Yet when the wave forcing is short lived,  $\bar{u}$  recovers quickly in about 10–15 days after the event (Fig. 8d). The discontinuous recovery line identified in Fig. 7d is also visible in Fig. 8d but its slope is significantly steeper because mixing shortens the damping time scale. This leads to a somewhat counterintuitive conclusion that mixing *promotes* vortex recovery when the wave forcing is short. White contours in Fig. 8b identify regions in which  $\bar{u}/u_{\text{REF}} < 0.5$ , that is, where Eq. (23) is violated. These supercritical regions are now localized in time and altitude, consistent with the relatively short duration of wave events. The supercritical region extends farther down when the forcing is persistent (Fig. 8b). Overall, Figs. 8b and 8d bear a strong resemblance to Fig. 8 of Lubis et al. (2018a) for slow- and rapid recovery composites based on reanalysis, which lends

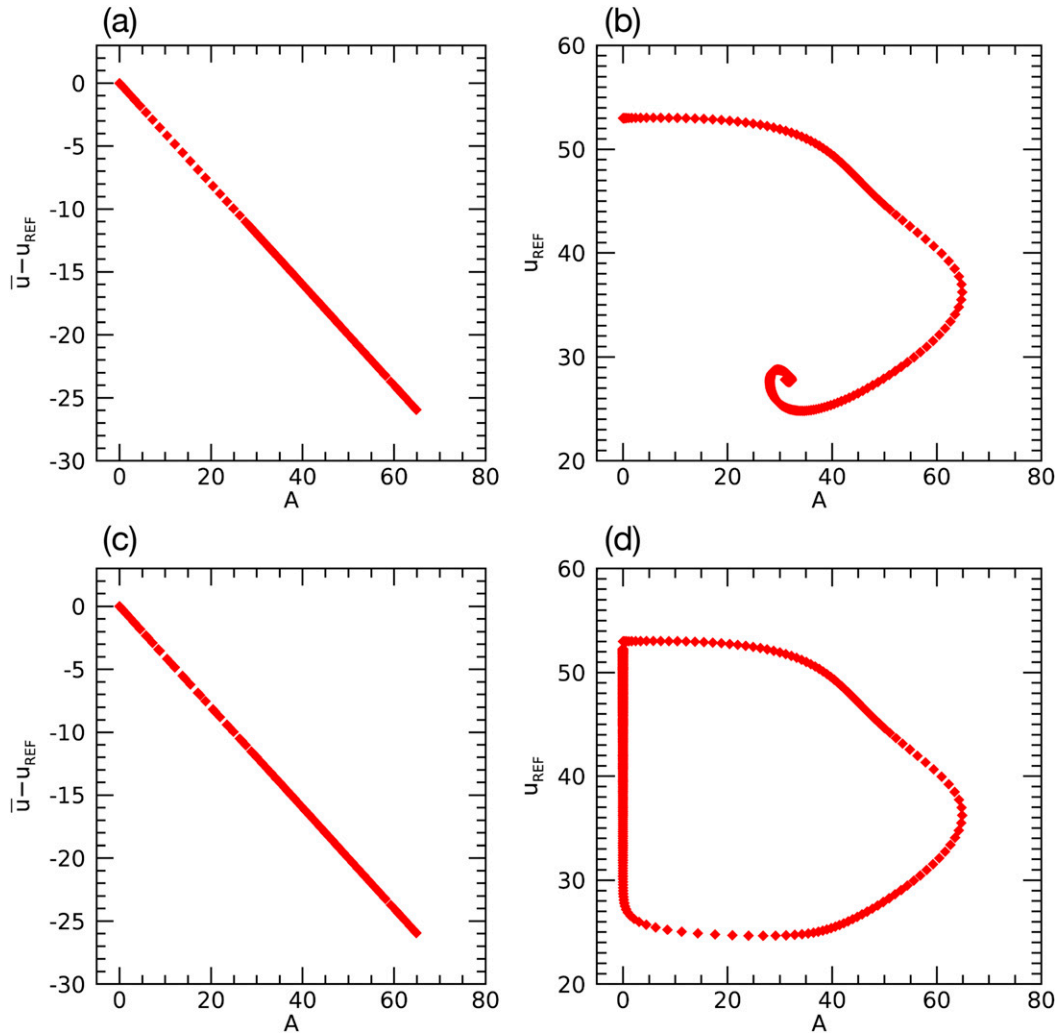


FIG. 9. (a)  $A$  (horizontal axis) vs  $\bar{u} - u_{\text{REF}}$  (vertical axis) at  $z = 32$  km during the experiment shown in Figs. 8a and 8b. (b) As in (a), but the vertical axis is  $u_{\text{REF}}$ . (c), (d) As in (a) and (b), but for the experiment shown in Figs. 8c and 8d.

support to the model results reported herein. On the other hand, the zonal-mean zonal wind in Fig. 8 does not drop below zero, which is one of the most unrealistic aspects of the model.

Figure 9 shows relationships between  $A$  and  $\Delta u = \bar{u} - u_{\text{REF}}$  and between  $A$  and  $u_{\text{REF}}$  at  $z = 32$  km during the two experiments shown in Fig. 8. In both cases  $A$  and  $\Delta u$  form a linear relation as expected (Figs. 9a,c), whereas  $u_{\text{REF}}$  monotonically decreases as  $A$  grows and decays. Only in the recovery stage of Fig. 9d,  $u_{\text{REF}}$  is radiatively restored and the hysteresis loop is complete.<sup>10</sup> The qualitative resemblance between Figs. 9 and

1 (except that Fig. 1d does not form a closed loop partly due to the seasonal demise of the polar vortex) is another evidence that the model is capturing the rudimentary dynamics of the observed SSWs.

## 5. Observational evidence for threshold behavior

The main conclusions from the foregoing semiempirical theory are: (i) wave activity flux  $F_z$  is a nonlinear (unimodal) function of FAWA due to wave–mean flow interaction and (ii) the mode of  $F_z$  defines the threshold for a spontaneous growth of FAWA. The WKB theory based on a vertically propagating stationary Rossby wave predicts  $F_z$  to be a quadratic function of  $A$  [Eq. (20)], but the foregoing conclusions hold for a more generalized form of Eqs. (19) and (21):

<sup>10</sup> The area enclosed by the loop is proportional to the pseudoenergy lost to mixing.

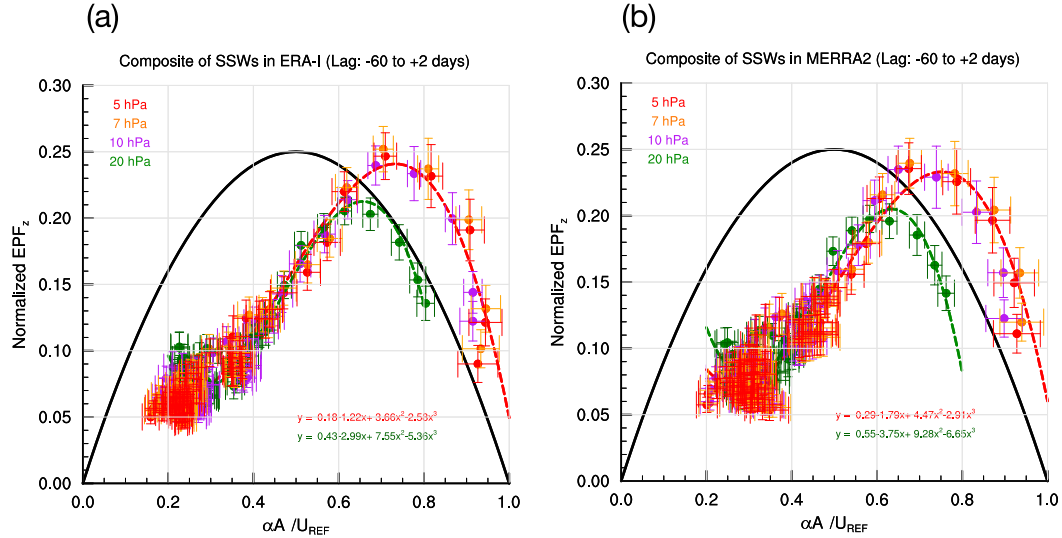


FIG. 10. (a)  $\xi = \alpha A / u_{\text{REF}}$  (horizontal axis) vs  $M(\xi)$  (vertical axis) at  $60^\circ\text{N}$  during 60 days before and 2 days after SSW events. Composite of 26 events in ERA-Interim between 1979 and 2016. Each dot shows a daily composite on a pressure level. Dots generally progress from low  $\xi$  to high  $\xi$  during this period. Error bars indicate  $\pm 0.25$  standard deviation. Red: 5 hPa. Orange: 7 hPa. Purple: 10 hPa. Green: 20 hPa. Third-order polynomial fits are shown for 5, 7, and 10 hPa (red dashed curve) and for 20 hPa (green dashed curve). Black curve is the theoretical prediction from section 4. (b) As in (a), but based on MERRA-2 (1979–2016, composite of 25 events). See text for details.

$$\frac{\partial}{\partial t}(Ae^{-z/H}) = -\frac{\partial}{\partial z}[\Gamma(z)M(\xi)], \quad \xi \equiv \frac{\alpha A(z, t)}{u_{\text{REF}}(z)}, \quad (44)$$

where  $M(\xi)$  is any smooth, nonnegative, unimodal function that does not depend on  $z$ , and  $\Gamma(z)$  is defined in Eq. (21). For the WKB theory  $M(\xi) = (1 - \xi)\xi$ . In the steady state, the lhs of Eq. (44) vanishes and Eq. (23) is replaced by

$$F_0 < \Gamma(z)M_{\text{max}}, \quad \xi < \xi_c, \quad (45)$$

where  $M(\xi_c) = M_{\text{max}}$ . The threshold behavior occurs at  $\xi = \xi_c$ . If the observed flux fits the above form and if we can demonstrate the modality of  $M$  with respect to  $\xi$  and exceedance of  $\xi_c$  during SSWs, it will be a strong evidence for a threshold behavior, irrespective of the WKB theory or how the wave activity flux is generated. Since [e.g., Andrews et al. (1987), their Eq. (3.5.6); NS10, their Eq. (7)]

$$F_z \approx \frac{f\overline{v'\theta'}}{\frac{\partial\theta_0}{\partial z}} e^{-z/H}, \quad (46)$$

we evaluate  $M$  at multiple altitudes using reanalysis data as

$$M = \frac{F_z}{\Gamma(z)} \approx \frac{f\overline{v'\theta'}}{\frac{\partial\theta_0}{\partial z}} \frac{\alpha}{Cu_{\text{REF}}^2} \quad (47)$$

and examine its dependence on  $\xi$ . If the data matches the WKB theory, we should observe  $M(\xi) = (1 - \xi)\xi$  at all altitudes.

Figure 10 shows  $M(\xi)$  during the growing stage of SSW (60-day lead plus 2-day lag) for 5, 7, 10, and 20 hPa at  $60^\circ\text{N}$  (daily composites of 26 events in ERA-Interim and 25 events in MERRA-2). In Eq. (47), we have assumed  $n = 2$  for all events to evaluate  $C$  [Eq. (17)] but used instantaneous  $u_{\text{REF}}$  since it varies significantly from one season to another. In doing so, we are effectively treating  $\Gamma(z)$  as a function of time also. This brings about some scatter in our analysis, but for both reanalysis products, data points for 5-, 7-, and 10-hPa collapse on average on a single curve (dashed red) with a clearly defined peak around  $\xi \approx 0.7 (= \xi_c)$ . The 20-hPa data trace a similar curve (dashed green) but its peak is lower and shifted to the left of the other curve. On both curves, the last few days around the event are on the right-hand side of the peak ( $\xi > \xi_c$ ). This is highly suggestive of a threshold behavior, and at least for 5–10 hPa, the height-independent  $M(\xi)$  is consistent with Eq. (44). On the other hand, although the diagnosed  $M_{\text{max}}$  is close to theory (0.25), its location  $\xi_c$  is shifted from the theoretical prediction (0.7 as opposed to 0.5). It may be that the weak dependence of  $u_{\text{REF}}$  on  $A$  due to mixing toward the event (Figs. 1c and 1d) modifies the form of  $M(\xi)$  from theory. Some discrepancies are undoubtedly related to the limitation of the WKB approximation employed in section 4. For example,  $M(\xi)$  remains positive

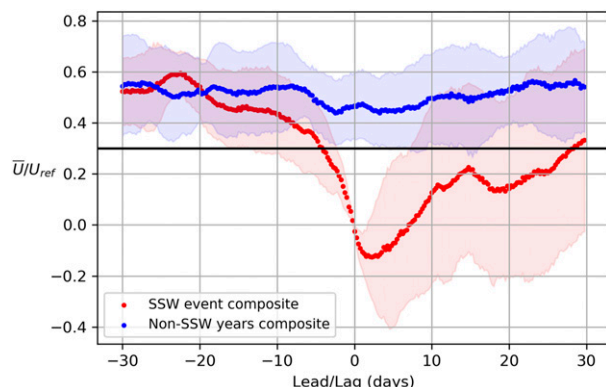


FIG. 11. The red line shows a composite time series of  $r = \bar{u}/u_{\text{REF}}$  during the life cycle of SSW (30-day lead plus 30-day lag) at  $60^\circ\text{N}$  and  $z = 32\text{ km}$  (10.34 hPa). The same 18 events as in Fig. 1 (listed in Table 1) are used. Lag 0 corresponds to the first reversal of  $\bar{u}$ . The shading indicates plus and minus one standard deviation. The blue line is similar to the red line but shows a composite for non-SSW winters. For this set, lag 0 corresponds to 31 Jan (date on which seasonal climatology of wave activity maximizes). Data source: ERA-Interim 6-hourly dataset over 1979–2016. The black horizontal line indicates  $r = 0.3$ . See text for details.

even after the zonal-mean zonal wind is reversed; this is likely due to a memory effect of rapidly evolving Rossby waves – an aspect that cannot be handled by the quasi-steady assumption in the WKB theory.

Since  $\xi_c \approx 0.7$ , a threshold behavior is expected to occur when  $r = \bar{u}/u_{\text{REF}} = 1 - \xi$  drops below about 0.3 in the reanalyses, rather than 0.5 as predicted by Eq. (24). This is indeed confirmed in Fig. 11, where composites of 60-day time series of  $r$  at  $60^\circ\text{N}$  and  $z = 32\text{ km}$  are compared for 18 winters with SSWs and non-SSW winters using ERA-Interim. While  $r$  stays above 0.3 for the majority of the non-SSW winters, the winters with SSW experience rapid decrease in  $r$  after it reaches 0.3, which marks the onset of the event. There is a wide spread after the event, as some events recover quickly (“reflecting” type in Table 2) while others recover more slowly (“absorbing” type), but the threshold behavior during the onset period is shared by all events analyzed. There were a few winters in which  $r$  dropped below 0.3 temporarily but a full SSW did not follow, presumably because the wave forcing from below was terminated prematurely.

Given the ability of  $r$  to identify an SSW threshold, we propose it as a useful instantaneous diagnostic for the proximity of the polar vortex to breakdown and a tool to understand the intermittency of SSWs. Figure 12 shows height–time cross sections of  $u_{\text{REF}}$  (color shading) and  $r$  (contours, plotted for 0.4 or less) at  $60^\circ\text{N}$  for strong-vortex winters and weak-vortex winters. To differentiate vortex strengths, we have computed the

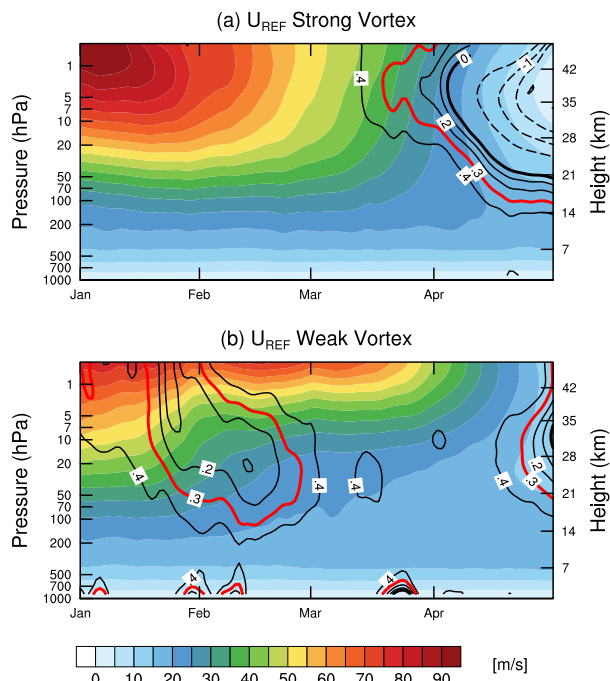


FIG. 12. (a) Composite of height–time cross sections of  $u_{\text{REF}}$  (color shading) and  $\bar{u}/u_{\text{REF}}$  (contours; plotted for 0.4 and below) for 15 strong-vortex winters (January–April 1982, 1986, 1989, 1990, 1993–97, 2000, 2005, 2007, 2011, 2014, 2015). (b) As in (a), but for 11 weak-vortex winters (1985, 1987, 1999, 2001, 2002, 2004, 2006, 2009, 2010, 2012, 2013). The 0.3 and 0 contours for  $\bar{u}/u_{\text{REF}}$  are highlighted in red and black, respectively. Data source: ERA-Interim 1979–2016. See text for details.

January–March NAM index based on the first EOF of the 50-hPa geopotential height anomaly north of  $30^\circ\text{N}$ . We have chosen 15 years in which the index exceeded 0.5 standard deviation as strong-vortex winters, and 11 years in which it exceeded  $-0.5$  standard deviation as weak-vortex winters (Perlwitz and Graf 2001; Baldwin and Dunkerton 1999). (See caption for the list of winters in each category.) Composite for the strong-vortex winters shows a strong and unperturbed  $u_{\text{REF}}$  in the stratosphere through January–February, and  $r$  drops below threshold only at the end of March (final warming, Fig. 12a). SSWs are characteristically absent, as exemplified by the streak of SSW-free winters during the mid-1990s. Whereas during the weak-vortex winters, both  $u_{\text{REF}}$  and  $r$  are significantly smaller already at the beginning of January (Fig. 12b). Then  $r$  drops below 0.3 from late January through February, starting at the top of the stratosphere and subsequently at lower altitudes, as SSWs form around this time of winter. Mixing associated with SSWs sharply decreases  $u_{\text{REF}}$  from the mid- to lower stratosphere, and it remains weak through the rest of the winter. However, persistence of  $u_{\text{REF}}$  in the upper stratosphere increases after SSW, and this

significantly delays the timing of final warming and easterly transition compared to the strong-vortex winters (Fig. 12b), consistent with the findings of previous studies (Hu et al. 2014; Lubis et al. 2017).

A weaker  $u_{\text{REF}}$  is more conducive to SSWs because it lowers  $\Gamma$  [Eq. (21)], allowing the threshold to be reached in the mid- to lower stratosphere with less wave forcing [Eq. (45)]. On the other hand, if  $u_{\text{REF}}$  is too weak, it reduces the growth rate, and hence suddenness, of SSWs [Eq. (26)]. This may be the reason why final warming is more pronounced during the strong-vortex winters (Fig. 12a). Since  $u_{\text{REF}}$  depends on mixing and radiative damping associated with preceding wave events, early season troposphere–stratosphere coupling likely plays a key role in the preconditioning of  $u_{\text{REF}}$  for SSWs. Interestingly, during the weak-vortex winters there are pockets of below-threshold  $r$ s in the lower troposphere (Fig. 12b). Their connection to SSWs is unclear from this figure alone, but it strongly suggests that wave coupling between the troposphere and the stratosphere is more active during the weak-vortex winters.

## 6. Summary and discussion

In the spirit of Jucker (2016), we have sought canonical properties of SSWs that transcend type variations. In particular, we have highlighted the role of wave–mean flow interaction [Eq. (9)] giving rise to suddenness and intermittency of SSWs through a threshold behavior. The threshold arises from a competition between an increasing wave activity  $A$  and a decreasing zonal-mean zonal wind  $\bar{u}$  in shaping the wave activity flux  $F_z$ . This renders  $F_z$  a quadratic function of  $A$  [Eq. (20)] and the mode of  $F_z$  defines the threshold [Eq. (24)]. Once  $A$  exceeds the threshold value, vortex breakdown proceeds spontaneously through a positive feedback between  $A$  and  $F_z$ , until  $\bar{u} = F_z = 0$  [Eq. (27)] is reached. A sharp boundary (“vortex terminus”) forms between the steady propagation regime and the vortex breakdown regime, and since there is a gap in  $F_z$  between the two, the boundary migrates downward (Fig. 6). Mathematically this process is analogous to a traffic jam on a highway, where the jam expands backward by absorbing the incoming traffic. The resultant 1D model is akin to the traffic jam model of atmospheric blocking considered by Nakamura and Huang (2017, 2018) and Paradise et al. (2019). Further simplified from previous theoretical models (e.g., Holton and Mass 1976; Plumb and Semeniuk 2003), this model permits certain analytical predictions besides the threshold, including the growth rate of SSW [Eq. (26)], migration speeds of the vortex terminus [Eq. (33)] and the

recovery line [Eq. (42)]. The model’s response has been tested for a variety of wave forcing and damping and shown to capture the salient features of observed SSWs (Fig. 8) despite the obvious limitations of the WKB theory it is built on.

Examination of reanalysis data lends strong support for the threshold behavior during the onset of SSWs. In the mid- to upper stratosphere, the observed wave activity follows a canonical unimodal function of FAWA and the onset of SSW lies on the right-hand side of its mode (Fig. 10). Consistent with this, the time series of  $r$  suggests  $r \approx 0.3$  as a “tipping point” for SSW’s onset (Fig. 11). The utility of  $r$  as an instantaneous diagnostic for the proximity to vortex breakdown has been illustrated (Fig. 12).

Given that the threshold is reached on average 4–5 days prior to an SSW event (Fig. 11),  $r$  should provide a forecast skill of SSW with a few days of lead time without a need for running a numerical weather prediction model. More importantly, the threshold may be used to better constrain the effects of climate change on the frequency of SSWs. Currently climate models’ confidence in the future projection of SSW frequency is low (e.g., Ayarzagüena et al. 2018). The lack of convergence among the model projections is likely due to model-to-model variations in (i) the threshold itself and (ii) the fashion in which the threshold is reached. On the first point, analyses like Figs. 10 and 11 should reveal model biases. Models with  $\xi_c > 0.7$  likely underestimate SSW frequencies for the same wave forcing, and vice versa. On the second point, one must ultimately understand how  $u_{\text{REF}}$  and  $A$  are determined and how they respond to climate perturbations. Particularly important is the early season behavior (preconditioning) of  $u_{\text{REF}}$  as it determines the subsequent wave forcing required to reach the threshold [Eq. (47)]. Sensitivity of the model prediction of  $u_{\text{REF}}$  and its impact on the SSW frequency to the early season troposphere–stratosphere wave coupling is a worthy topic for further investigation.

**Acknowledgments.** This work has been supported by NSF Grants AGS1563307 and AGS1909522. The authors thank two anonymous reviewers for constructive critiques on the early version of the manuscript.

## APPENDIX

### Dispersion Relation in the Mercator Coordinate

Following Hoskins and Karoly (1981), the Mercator coordinate is defined as

$$x \equiv a\lambda, \quad y \equiv a \ln[(1 + \sin\phi)/\cos\phi], \quad (\text{A1})$$



$$\frac{1}{a \cos \phi} \frac{\partial}{\partial \lambda} = \frac{1}{\cos \phi} \frac{\partial}{\partial x}, \quad \frac{1}{a} \frac{\partial}{\partial \phi} = \frac{1}{\cos \phi} \frac{\partial}{\partial y},$$

$$\nabla_H^2 = \frac{1}{\cos^2 \phi} \left( \frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2} \right). \quad (\text{A2})$$

The PV equation in this coordinate, linearized about the zonal-mean state and multiplied by  $\cos^2 \phi$ , reads<sup>A1</sup>

$$\left( \frac{\partial}{\partial t} + \frac{\bar{u}}{\cos \phi} \frac{\partial}{\partial x} \right) \left\{ \left( \frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2} \right) \psi' \right. \\ \left. + e^{z/H} \frac{\partial}{\partial z} \left[ \frac{(f \cos \phi)^2 e^{-z/H}}{N^2(z)} \frac{\partial \psi'}{\partial z} \right] \right\} \\ + \frac{\partial \psi'}{\partial x} \left( \frac{\cos \phi}{a} \frac{\partial \bar{q}}{\partial \phi} \right) = 0. \quad (\text{A3})$$

In the above,  $\psi'$  is the perturbation geostrophic streamfunction and the following coefficients are considered “slowly varying”:

$$\frac{\bar{u}}{\cos \phi}, \quad \frac{(f \cos \phi)^2}{N^2(z)}, \quad \frac{\cos \phi}{a} \frac{\partial \bar{q}}{\partial \phi}. \quad (\text{A4})$$

We assume that the wave structure is nearly planar and its amplitude and phase are slowly modulated according to

$$\psi' = \Gamma(Y, Z, T) e^{i\Theta(Y, Z, T)} e^{z/(2H)} e^{ikx},$$

$$\left( \frac{\partial \Theta}{\partial Y}, \frac{\partial \Theta}{\partial Z}, \frac{\partial \Theta}{\partial T} \right) \equiv (l, m, -\omega), \quad (\text{A5})$$

where  $k = n/a$  ( $n$  is the integer zonal wavenumber) is the equatorial wavenumber and  $(Y, Z, T) \equiv (\gamma y, \gamma z, \gamma t)$ , ( $\gamma \ll 1$ ) is the slow coordinate in the meridional plane and time. Substitution in Eq. (A3) yields the following dispersion relation:

$$\omega = \frac{\bar{u}k}{\cos \phi} - \frac{\frac{\cos \phi}{a} \frac{\partial \bar{q}}{\partial \phi} k}{k^2 + l^2 + \frac{f^2 \cos^2 \phi}{N^2} \left( m^2 + \frac{1}{4H^2} \right)}. \quad (\text{A6})$$

The vertical group velocity is

$$c_{gz} = \frac{\partial \omega}{\partial m} = \frac{\frac{2f^2}{N^2} \frac{1}{a \cos \phi} \frac{\partial \bar{q}}{\partial \phi} km}{\left[ \frac{k^2 + l^2}{\cos^2 \phi} + \frac{f^2}{N^2} \left( m^2 + \frac{1}{4H^2} \right) \right]^2}. \quad (\text{A7})$$

<sup>A1</sup> Equation (A3) takes on the familiar form because the  $x$  derivative commutes with the  $1/\cos^2 \phi$  factor in the Laplacian. This is not the case for the nonlinear PV equation, which involves a  $y$  derivative of PV and takes a considerably more complicated form.

For a stationary wave ( $\omega = 0$ ) Eq. (A6) gives

$$\frac{1}{\frac{k^2 + l^2}{\cos^2 \phi} + \frac{f^2}{N^2} \left( m^2 + \frac{1}{4H^2} \right)} = \frac{\bar{u}}{\frac{1}{a} \frac{\partial \bar{q}}{\partial \phi}} \quad (\text{A8})$$

and

$$m^2 = \frac{N^2}{f^2} \left[ \frac{\frac{1}{a} \frac{\partial \bar{q}}{\partial \phi}}{\bar{u}} - \frac{k^2 + l^2}{\cos^2 \phi} - \frac{f^2}{4N^2 H^2} \right]. \quad (\text{A9})$$

From Eqs. (A7) and (A8), one obtains Eq. (10) for the vertical group velocity of a stationary wave.

## REFERENCES

- Andrews, D., and M. McIntyre, 1976: Planetary waves in horizontal and vertical shear: The generalized Eliassen–Palm relation and the mean zonal acceleration. *J. Atmos. Sci.*, **33**, 2031–2048, [https://doi.org/10.1175/1520-0469\(1976\)033<2031:PWIHAV>2.0.CO;2](https://doi.org/10.1175/1520-0469(1976)033<2031:PWIHAV>2.0.CO;2).
- , J. Holton, and C. Leovy, 1987: *Middle Atmosphere Dynamics*. Academic Press, 489 pp.
- Ayarzagüena, B., and Coauthors, 2018: No robust evidence of future changes in major stratospheric sudden warmings: A multi-model assessment from CCMI. *Atmos. Chem. Phys.*, **18**, 11 277–11 287, <https://doi.org/10.5194/acp-18-11277-2018>.
- Baldwin, M. P., and T. J. Dunkerton, 1999: Propagation of the Arctic oscillation from the stratosphere to the troposphere. *J. Geophys. Res.*, **104**, 30 937–30 946, <https://doi.org/10.1029/1999JD900445>.
- , and —, 2001: Stratospheric harbingers of anomalous weather regimes. *Science*, **294**, 581–584, <https://doi.org/10.1126/science.1063315>.
- Birner, T., and P. D. Williams, 2008: Sudden stratospheric warmings as noise-induced transitions. *J. Atmos. Sci.*, **65**, 3337–3343, <https://doi.org/10.1175/2008JAS2770.1>.
- Butler, A. H., D. J. Seidel, S. C. Hardiman, N. Butchart, T. Birner, and A. Match, 2015: Defining sudden stratospheric warmings. *Bull. Amer. Meteor. Soc.*, **96**, 1913–1928, <https://doi.org/10.1175/BAMS-D-13-00173.1>.
- Chao, W. C., 1985: Sudden stratospheric warmings as catastrophes. *J. Atmos. Sci.*, **42**, 1631–1646, [https://doi.org/10.1175/1520-0469\(1985\)042<1631:SSWAC>2.0.CO;2](https://doi.org/10.1175/1520-0469(1985)042<1631:SSWAC>2.0.CO;2).
- Charlton, A. J., and L. M. Polvani, 2007: A new look at stratospheric sudden warmings. Part I: Climatology and modeling benchmarks. *J. Climate*, **20**, 449–469, <https://doi.org/10.1175/JCLI3996.1>.
- Charlton-Perez, A. J., L. M. Polvani, J. Austin, and F. Li, 2008: The frequency and dynamics of stratospheric sudden warmings in the 21st century. *J. Geophys. Res.*, **113**, D16116, <https://doi.org/10.1029/2007JD009571>.
- Charney, J., and P. Drazin, 1961: Propagation of planetary-scale disturbances from the lower into the upper atmosphere. *J. Geophys. Res.*, **66**, 83–109, <https://doi.org/10.1029/JZ066i001P00083>.
- , and J. DeVore, 1979: Multiple flow equilibria in the atmosphere and blocking. *J. Atmos. Sci.*, **36**, 1205–1216, [https://doi.org/10.1175/1520-0469\(1979\)036<1205:MFEITA>2.0.CO;2](https://doi.org/10.1175/1520-0469(1979)036<1205:MFEITA>2.0.CO;2).

- Chen, P., and W. A. Robinson, 1992: Propagation of planetary waves between the troposphere and stratosphere. *J. Atmos. Sci.*, **49**, 2533–2545, [https://doi.org/10.1175/1520-0469\(1992\)049<2533:POPWBT>2.0.CO;2](https://doi.org/10.1175/1520-0469(1992)049<2533:POPWBT>2.0.CO;2).
- de la Cámara, A., J. R. Albers, T. Birner, R. R. Garcia, P. Hitchcock, D. E. Kinnison, and A. K. Smith, 2017: Sensitivity of sudden stratospheric warmings to previous stratospheric conditions. *J. Atmos. Sci.*, **74**, 2857–2877, <https://doi.org/10.1175/JAS-D-17-0136.1>.
- Dee, D. P., and Coauthors, 2011: The Era-Interim reanalysis: Configuration and performance of the data assimilation system. *Quart. J. Roy. Meteor. Soc.*, **137**, 553–597, <https://doi.org/10.1002/qj.828>.
- Dickinson, R. E., 1969: Theory of planetary wave-zonal flow interaction. *J. Atmos. Sci.*, **26**, 73–81, [https://doi.org/10.1175/1520-0469\(1969\)026<0073:TOPWZF>2.0.CO;2](https://doi.org/10.1175/1520-0469(1969)026<0073:TOPWZF>2.0.CO;2).
- Esler, J. G., and R. K. Scott, 2005: Excitation of transient Rossby waves on the stratospheric polar vortex and the barotropic sudden warming. *J. Atmos. Sci.*, **62**, 3661–3682, <https://doi.org/10.1175/JAS3557.1>.
- , L. M. Polvani, and R. K. Scott, 2006: The Antarctic stratospheric sudden warming of 2002: A self-tuned resonance? *Geophys. Res. Lett.*, **33**, L12804, <https://doi.org/10.1029/2006GL026034>.
- Ferrari, R., and M. Nikurashin, 2010: Suppression of eddy diffusivity across jets in the Southern Ocean. *J. Phys. Oceanogr.*, **40**, 1501–1519, <https://doi.org/10.1175/2010JPO4278.1>.
- Geisler, J. E., 1974: A numerical model of sudden stratospheric warming mechanism. *J. Geophys. Res.*, **79**, 4989–4999, <https://doi.org/10.1029/JC079I033P04989>.
- Gelaro, R., and Coauthors, 2017: The Modern-Era Retrospective Analysis for Research and Applications, version 2 (MERRA-2). *J. Climate*, **30**, 5419–5454, <https://doi.org/10.1175/JCLI-D-16-0758.1>.
- Gerber, E. P., C. Orbe, and L. M. Polvani, 2009: Stratospheric influence on the tropospheric circulation revealed by idealized ensemble forecasts. *Geophys. Res. Lett.*, **36**, L24801, <https://doi.org/10.1029/2009GL040913>.
- Harada, Y., A. Goto, H. Hasegawa, N. Fujikawa, H. Naoe, and T. Hirooka, 2010: A major stratospheric sudden warming event in January 2009. *J. Atmos. Sci.*, **67**, 2052–2069, <https://doi.org/10.1175/2009JAS3320.1>.
- Harnik, N., 2009: Observed stratospheric downward reflection and its relation to upward pulses of wave activity. *J. Geophys. Res.*, **114**, D08120, <https://doi.org/10.1029/2008JD010493>.
- Held, I., 1983: Stationary and quasi-stationary eddies in the extratropical troposphere: Theory. *Large-Scale Dynamical Processes in the Atmosphere*, B. Hoskins and R. Pearce, Eds., Academic Press, 127–168.
- Hitchcock, P., and I. R. Simpson, 2014: The downward influence of stratospheric sudden warmings. *J. Atmos. Sci.*, **71**, 3856–3876, <https://doi.org/10.1175/JAS-D-14-0012.1>.
- , and P. H. Haynes, 2016: Stratospheric control of planetary waves. *Geophys. Res. Lett.*, **43**, 11 884–11 892, <https://doi.org/10.1002/2016GL071372>.
- Holton, J. R., 1976: A semi-spectral numerical model for wave-mean flow interactions in the stratosphere: Application to sudden stratospheric warmings. *J. Atmos. Sci.*, **33**, 1639–1649, [https://doi.org/10.1175/1520-0469\(1976\)033<1639:ASSNMF>2.0.CO;2](https://doi.org/10.1175/1520-0469(1976)033<1639:ASSNMF>2.0.CO;2).
- , and C. Mass, 1976: Stratospheric vacillation cycles. *J. Atmos. Sci.*, **33**, 2218–2225, [https://doi.org/10.1175/1520-0469\(1976\)033<2218:SVC>2.0.CO;2](https://doi.org/10.1175/1520-0469(1976)033<2218:SVC>2.0.CO;2).
- , and T. Dunkerton, 1978: On the role of wave transience and dissipation in stratospheric mean flow vacillations. *J. Atmos. Sci.*, **35**, 740–744, [https://doi.org/10.1175/1520-0469\(1978\)035<0740:OTROWT>2.0.CO;2](https://doi.org/10.1175/1520-0469(1978)035<0740:OTROWT>2.0.CO;2).
- Hoskins, B. J., and D. J. Karoly, 1981: The steady linear response of a spherical atmosphere to thermal and orographic forcing. *J. Atmos. Sci.*, **38**, 1179–1196, [https://doi.org/10.1175/1520-0469\(1981\)038<1179:TSLROA>2.0.CO;2](https://doi.org/10.1175/1520-0469(1981)038<1179:TSLROA>2.0.CO;2).
- Hu, J., R. Ren, and H. Xu, 2014: Occurrence of winter stratospheric sudden warming events and the seasonal timing of spring stratospheric final warming. *J. Atmos. Sci.*, **71**, 2319–2334, <https://doi.org/10.1175/JAS-D-13-0349.1>.
- Jucker, M., 2016: Are sudden stratospheric warmings generic? Insights from an idealized GCM. *J. Atmos. Sci.*, **73**, 5061–5080, <https://doi.org/10.1175/JAS-D-15-0353.1>.
- Killworth, P., and M. McIntyre, 1985: Do Rossby-wave critical layers absorb, reflect, or over-reflect? *J. Fluid Mech.*, **161**, 449–492, <https://doi.org/10.1017/S0022112085003019>.
- Kodera, K., and M. Chiba, 1995: Tropospheric circulation changes associated with stratospheric sudden warmings—A case study. *J. Geophys. Res.*, **100**, 11 055–11 068, <https://doi.org/10.1029/95JD00771>.
- , H. Mukougawa, P. Maury, M. Ueda, and C. Claud, 2016: Absorbing and reflecting sudden stratospheric warming events and their relationship with tropospheric circulation. *J. Geophys. Res. Atmos.*, **121**, 80–94, <https://doi.org/10.1002/2015JD023359>.
- Kushner, P. J., and L. M. Polvani, 2004: Stratosphere–troposphere coupling in a relatively simple AGCM: The role of eddies. *J. Climate*, **17**, 629–639, [https://doi.org/10.1175/1520-0442\(2004\)017<0629:SCIARS>2.0.CO;2](https://doi.org/10.1175/1520-0442(2004)017<0629:SCIARS>2.0.CO;2).
- Labitzke, K., and B. Naujokat, 2000: The lower Arctic stratosphere in winter since 1952. *SPARC Newsl.*, **15**, 11–14.
- LeVeque, R., 2002: *Finite Volume Methods for Hyperbolic Problems*. Cambridge University Press, 558 pp.
- Lighthill, M. J., and G. B. Whitham, 1955: On kinematic waves II. A theory of traffic flow on long crowded roads. *Proc. Roy. Soc. London*, **229**, 317–345, <https://doi.org/10.1098/rspa.1955.0089>.
- Limpasuvan, V., D. W. J. Thompson, and D. L. Hartmann, 2004: The life cycle of the Northern Hemisphere sudden stratospheric warmings. *J. Climate*, **17**, 2584–2596, [https://doi.org/10.1175/1520-0442\(2004\)017<2584:TLCOTN>2.0.CO;2](https://doi.org/10.1175/1520-0442(2004)017<2584:TLCOTN>2.0.CO;2).
- Liu, Y. S., and R. K. Scott, 2015: The onset of the barotropic sudden warming in a global model. *Quart. J. Roy. Meteor. Soc.*, **141**, 2944–2955, <https://doi.org/10.1002/QJ.2580>.
- Lubis, S. W., V. Silverman, K. Matthes, N. Harnik, N.-E. Omrani, and S. Wahl, 2017: How does downward planetary wave coupling affect polar stratospheric ozone in the Arctic winter stratosphere? *Atmos. Chem. Phys.*, **17**, 2437–2458, <https://doi.org/10.5194/acp-17-2437-2017>.
- , C. S. Y. Huang, and N. Nakamura, 2018a: Role of finite-amplitude eddies and mixing in the life cycle of stratospheric sudden warmings. *J. Atmos. Sci.*, **75**, 3987–4003, <https://doi.org/10.1175/JAS-D-18-0138.1>.
- , —, —, N. E. Omrani, and M. Jucker, 2018b: Role of finite-amplitude Rossby waves and nonconservative processes in downward migration of extratropical flow anomalies. *J. Atmos. Sci.*, **75**, 1385–1401, <https://doi.org/10.1175/JAS-D-17-0376.1>.
- Martineau, P., and S.-W. Son, 2015: Onset of circulation anomalies during stratospheric vortex weakening events: The role of planetary-scale waves. *J. Climate*, **28**, 7347–7370, <https://doi.org/10.1175/JCLI-D-14-00478.1>.

- Martius, O., L. M. Polvani, and H. C. Davies, 2009: Blocking precursors to stratospheric sudden warming events. *Geophys. Res. Lett.*, **36**, L14806, <https://doi.org/10.1029/2009GL038776>.
- Matsuno, T., 1971: A dynamical model of the stratospheric sudden warming. *J. Atmos. Sci.*, **28**, 1479–1494, [https://doi.org/10.1175/1520-0469\(1971\)028<1479:ADMOTS>2.0.CO;2](https://doi.org/10.1175/1520-0469(1971)028<1479:ADMOTS>2.0.CO;2).
- Matthewman, N. J., and J. G. Esler, 2011: Stratospheric sudden warmings as self-tuning resonances. Part I: Vortex splitting events. *J. Atmos. Sci.*, **68**, 2481–2504, <https://doi.org/10.1175/JAS-D-11-07.1>.
- McIntyre, M., 1982: How well do we understand the dynamics of stratospheric warmings? *J. Meteor. Soc. Japan*, **60**, 37–65, [https://doi.org/10.2151/jmsj1965.60.1\\_37](https://doi.org/10.2151/jmsj1965.60.1_37).
- , and T. Palmer, 1983: Breaking planetary waves in the stratosphere. *Nature*, **305**, 593–600, <https://doi.org/10.1038/305593a0>.
- , and —, 1984: The “surf zone” in the stratosphere. *J. Atmos. Terr. Phys.*, **46**, 825–849, [https://doi.org/10.1016/0021-9169\(84\)90063-1](https://doi.org/10.1016/0021-9169(84)90063-1).
- McLandress, C., and T. G. Shepherd, 2009: Impact of climate change on stratospheric sudden warmings as simulated by the Canadian middle atmosphere model. *J. Climate*, **22**, 5449–5463, <https://doi.org/10.1175/2009JCLI3069.1>.
- Methven, J., and P. Berrisford, 2015: The slowly evolving background state of the atmosphere. *Quart. J. Roy. Meteor. Soc.*, **141**, 2237–2258, <https://doi.org/10.1002/QJ.2518>.
- Mukougawa, H., and T. Hirooka, 2004: Predictability of stratospheric sudden warming: A case study for 1998/99 winter. *Mon. Wea. Rev.*, **132**, 1764–1776, [https://doi.org/10.1175/1520-0493\(2004\)132<1764:POSSWA>2.0.CO;2](https://doi.org/10.1175/1520-0493(2004)132<1764:POSSWA>2.0.CO;2).
- Nakamura, N., and A. Solomon, 2010: Finite-amplitude wave activity and mean flow adjustments in the atmospheric general circulation. Part I: Quasigeostrophic theory and analysis. *J. Atmos. Sci.*, **67**, 3967–3983, <https://doi.org/10.1175/2010JAS3503.1>.
- , and D. Zhu, 2010: Finite-amplitude wave activity and diffusive flux of potential vorticity in eddy–mean flow interaction. *J. Atmos. Sci.*, **67**, 2701–2716, <https://doi.org/10.1175/2010JAS3432.1>.
- , and A. Solomon, 2011: Finite-amplitude wave activity and mean flow adjustments in the atmospheric general circulation. Part II: Analysis in the isentropic coordinate. *J. Atmos. Sci.*, **68**, 2783–2799, <https://doi.org/10.1175/2011JAS3685.1>.
- , and C. S. Y. Huang, 2017: Local wave activity and the onset of blocking along a potential vorticity front. *J. Atmos. Sci.*, **74**, 2341–2362, <https://doi.org/10.1175/JAS-D-17-0029.1>.
- , and —, 2018: Atmospheric blocking as a traffic jam in the jet stream. *Science*, **361**, 42–47, <https://doi.org/10.1126/science.aat0721>.
- O'Neill, A., and B. F. Taylor, 1979: Study of the major stratospheric warming of 1976/77. *Quart. J. Roy. Meteor. Soc.*, **105**, 71–92, <https://doi.org/10.1002/QJ.49710544306>.
- Paradise, A., C. B. Rochas, P. Barpanda, and N. Nakamura, 2019: Blocking statistics in a varying climate: Lessons from a “traffic jam” model with pseudostochastic forcing. *J. Atmos. Sci.*, **76**, 3013–3027, <https://doi.org/10.1175/JAS-D-19-0095.1>.
- Perlwitz, J., and H. F. Graf, 2001: Troposphere-stratosphere dynamic coupling under strong and weak polar vortex conditions. *Geophys. Res. Lett.*, **28**, 271–274, <https://doi.org/10.1029/2000GL012405>.
- Plumb, R., 1981: Instability of the distorted polar night vortex: A theory of stratospheric warmings. *J. Atmos. Sci.*, **38**, 2514–2531, [https://doi.org/10.1175/1520-0469\(1981\)038<2514:IOTDPN>2.0.CO;2](https://doi.org/10.1175/1520-0469(1981)038<2514:IOTDPN>2.0.CO;2).
- , and K. Semeniuk, 2003: Downward migration of extratropical zonal wind anomalies. *J. Geophys. Res.*, **108**, 4223, <https://doi.org/10.1029/2002JD002773>.
- Polvani, L. M., and R. Saravanan, 2000: The three-dimensional structure of breaking Rossby waves in the polar winter-time stratosphere. *J. Atmos. Sci.*, **57**, 3663–3685, [https://doi.org/10.1175/1520-0469\(2000\)057<3663:TDSOB>2.0.CO;2](https://doi.org/10.1175/1520-0469(2000)057<3663:TDSOB>2.0.CO;2).
- , and D. W. Waugh, 2004: Upward wave activity flux as a precursor to extreme stratospheric events and subsequent anomalous surface weather regimes. *J. Climate*, **17**, 3548–3554, [https://doi.org/10.1175/1520-0442\(2004\)017<3548:UWAFAA>2.0.CO;2](https://doi.org/10.1175/1520-0442(2004)017<3548:UWAFAA>2.0.CO;2).
- Quiroz, R. S., 1975: The stratospheric evolution of sudden warmings in 1969–74 determined from measured infrared radiation fields. *J. Atmos. Sci.*, **32**, 211–224, [https://doi.org/10.1175/1520-0469\(1975\)032<0211:TSEOSW>2.0.CO;2](https://doi.org/10.1175/1520-0469(1975)032<0211:TSEOSW>2.0.CO;2).
- Randel, W. J., and B. A. Boville, 1987: Observations of a major stratospheric warming during December 1984. *J. Atmos. Sci.*, **44**, 2179–2186, [https://doi.org/10.1175/1520-0469\(1987\)044<2179:OOAMSW>2.0.CO;2](https://doi.org/10.1175/1520-0469(1987)044<2179:OOAMSW>2.0.CO;2).
- Rao, J., R. Ren, H. Chen, Y. Yu, and Y. Zhou, 2018: The stratospheric sudden warming event in February 2018 and its prediction by a climate system model. *J. Geophys. Res. Atmos.*, **123**, 13 332–13 345, <https://doi.org/10.1029/2018JD028908>.
- Richards, P. I., 1956: Shock waves on the highway. *Oper. Res.*, **4**, 42–51, <https://doi.org/10.1287/opre.4.1.42>.
- Ruzmaikin, A., J. Lawrence, and C. Cadavid, 2003: A simple model of stratospheric dynamics including solar variability. *J. Climate*, **16**, 1593–1600, <https://doi.org/10.1175/1520-0442-16.10.1593>.
- Scherhag, R., 1952: A review of uncertainties in global temperature projections over the twenty-first century. *Ber. Dtsch. Wetterdienstes*, **6**, 51–63.
- Scott, R. K., and P. H. Haynes, 2000: Internal vacillations in stratosphere-only models. *J. Atmos. Sci.*, **57**, 3233–3250, [https://doi.org/10.1175/1520-0469\(2000\)057<3233:IVISOM>2.0.CO;2](https://doi.org/10.1175/1520-0469(2000)057<3233:IVISOM>2.0.CO;2).
- , and L. M. Polvani, 2006: Internal variability of the winter stratosphere. Part I: Time-independent forcing. *J. Atmos. Sci.*, **63**, 2758–2776, <https://doi.org/10.1175/JAS3797.1>.
- Sjoberg, J. P., and T. Birner, 2012: Transient tropospheric forcing of sudden stratospheric warmings. *J. Atmos. Sci.*, **69**, 3420–3432, <https://doi.org/10.1175/JAS-D-11-0195.1>.
- Smith, A. K., 1989: An investigation of resonant waves in a numerical model of an observed sudden stratospheric warming. *J. Atmos. Sci.*, **46**, 3038–3054, [https://doi.org/10.1175/1520-0469\(1989\)046<3038:AIORWI>2.0.CO;2](https://doi.org/10.1175/1520-0469(1989)046<3038:AIORWI>2.0.CO;2).
- Solomon, A., 2014: Wave activity events and the variability of the stratospheric polar vortex. *J. Climate*, **27**, 7796–7806, <https://doi.org/10.1175/JCLI-D-13-00756.1>.
- Stan, C., and D. M. Straus, 2009: Stratospheric predictability and sudden stratospheric warming events. *J. Geophys. Res.*, **114**, D12103, <https://doi.org/10.1029/2008JD011277>.
- Taguchi, M., 2016: Connection of predictability of major stratospheric sudden warmings to polar vortex geometry. *Atmos. Sci. Lett.*, **17**, 33–38, <https://doi.org/10.1002/asl.595>.
- Tripathi, O. P., and Coauthors, 2016: Examining the predictability of the stratospheric sudden warming of January 2013 using multiple NWP systems. *Mon. Wea. Rev.*, **144**, 1935–1960, <https://doi.org/10.1175/MWR-D-15-0010.1>.

- Tung, K. K., and R. S. Lindzen, 1979: A theory of stationary long waves. Part I: Simple theory of blocking. *Mon. Wea. Rev.*, **107**, 714–734, [https://doi.org/10.1175/1520-0493\(1979\)107<0714:ATOSLW>2.0.CO;2](https://doi.org/10.1175/1520-0493(1979)107<0714:ATOSLW>2.0.CO;2).
- Wang, X. H., and J. Fyfe, 2000: Onset of edge wave breaking in an idealized model of the polar stratospheric vortex. *J. Atmos. Sci.*, **57**, 956–966, [https://doi.org/10.1175/1520-0469\(2000\)057<0956:OOEWBI>2.0.CO;2](https://doi.org/10.1175/1520-0469(2000)057<0956:OOEWBI>2.0.CO;2).
- Woollings, T., A. Charlton-Perez, S. Ineson, A. G. Marshall, and G. Masato, 2010: Associations between stratospheric variability and tropospheric blocking. *J. Geophys. Res.*, **115**, D06108, <https://doi.org/10.1029/2009JD012742>.
- Yasuda, Y., F. Bouchet, and A. Venaille, 2017: A new interpretation of vortex-split sudden stratospheric warmings in terms of equilibrium statistical mechanics. *J. Atmos. Sci.*, **74**, 3915–3936, <https://doi.org/10.1175/JAS-D-17-0045.1>.
- Yoden, S., 1987: Bifurcation properties of a stratospheric vacillation model. *J. Atmos. Sci.*, **44**, 1723–1733, [https://doi.org/10.1175/1520-0469\(1987\)044<1723:BPOASV>2.0.CO;2](https://doi.org/10.1175/1520-0469(1987)044<1723:BPOASV>2.0.CO;2).
- , 1990: An illustrative model of seasonal and interannual variations of the stratospheric circulation. *J. Atmos. Sci.*, **47**, 1845–1853, [https://doi.org/10.1175/1520-0469\(1990\)047<1845:AIMOSA>2.0.CO;2](https://doi.org/10.1175/1520-0469(1990)047<1845:AIMOSA>2.0.CO;2).