Transient Tropospheric Forcing of Sudden Stratospheric Warmings

JEREMIAH P. SJOBERG AND THOMAS BIRNER

Colorado State University, Fort Collins, Colorado

(Manuscript received 15 July 2011, in final form 17 April 2012)

ABSTRACT

The amplitude of upward-propagating tropospherically forced planetary waves is known to be of first-order importance in producing sudden stratospheric warmings (SSWs). This forcing amplitude is observed to undergo strong temporal fluctuations. Characteristics of the resulting transient forcing leading to SSWs are studied in reanalysis data and in highly truncated simple models of stratospheric wave–mean flow interaction. It is found in both the reanalysis data and the simple models that SSWs are preferentially generated by transient forcing of sufficiently long time scales (on the order of 1 week or longer). The time scale of the transient forcing is found to play a stronger role in producing SSWs than the strength of the forcing. In the simple models it is possible to fix the amplitude of the tropospheric forcing but to vary the time scale of the forcing. The resulting frequency of occurrence of SSWs shows dramatic reductions for decreasing forcing time scales.

1. Introduction

Sudden stratospheric warmings (SSWs) are large-scale events that are of first-order importance to understanding the wintertime dynamical variability in the polar stratosphere. Since the first identification of these events by Scherhag (1952), numerous attempts to model SSWs have been performed, perhaps most influentially by Matsuno (1971). The wave–mean flow interaction captured in Matsuno's nonlinear model was also shown to be resolved in simple, quasi-linear models (Geisler 1974; Holton 1976; Holton and Mass 1976). The justification for use of a simple model to investigate SSWs comes from the difficulty in isolating the response to very specific parameters from dynamical systems of large complexity. It has been suggested (Held 2005) that modeling hierarchies—where highly complex global climate models are supported by simple models of the relevant physics—are crucial for understanding atmospheric dynamical processes.

Perhaps the most well-known of these simple models—the Holton–Mass model—has been utilized in many studies because of its relative simplicity. This one-dimensional, β-plane channel model explores the interaction of waves specified by eddy potential vorticity and the zonal-mean zonal wind. It has been shown to capture zonal wind and eddy heat flux vacillations characteristic of SSWs from forcing by steady bottom boundary wave amplitudes (Holton and Mass 1976) and from forcing by sinusoidal bottom boundary wave amplitudes of varying period (Plumb and Semeniuk 2003; Harnik 2009). Analyses of stability states for this model may be performed either by quasi-analytical methods (Yoden 1987) or by numerical methods (Chao 1985; Christiansen 2000). Also, the Holton–Mass model may itself be simplified through vertical truncation down to a three-layer low-order model wherein all the modeled dynamical processes are contained in the middle layer and specified by the boundary conditions (Ruzmaikin et al. 2003). This strongly truncated low-order model has the great advantage that stationary solutions may be obtained analytically.

Most relevant of these simple modeling studies prior to this work are Holton and Mass (1976) and Harnik (2009), both of which investigate to some degree how the characteristics of the bottom boundary wave amplitudes affect the forcing of sudden warming solutions in the Holton–Mass model. The constant-amplitude tropospheric planetary waves considered in Holton and Mass (1976) do not exist in the real world. What are instead observed are pulses of wave amplitude more similar to those investigated in Harnik (2009). A similar investigation into both steady and time-dependent bottom

Corresponding author address: Jeremiah P. Sjoberg, Colorado State University, 1371 Campus Delivery, Fort Collins, CO 80521. E-mail: jeremiah@atmos.colostate.edu

DOI: 10.1175/JAS-D-11-0195.1

boundary forcing within a different stratospheric model may be found in Scott and Polvani (2006) and Scott et al. (2008).

Characteristics of observed wave activity preceding a sudden stratospheric warming have been discussed in Polvani and Waugh (2004). They show that extreme stratospheric events, such as SSWs, are associated with anomalously large values of the 40-day averaged upward wave activity flux at 100 hPa [confirming earlier results by Newman et al. (2001)]. They likewise show high anticorrelation values near -0.8 between this 40-day averaged upward wave activity flux and the strength of the stratospheric polar vortex, while these anticorrelation values become insignificant for time averaging of 10 days or less [see Fig. 3 of Polvani and Waugh (2004)]. These results suggest a linkage between long time scales of wave forcing and disturbances of the polar vortex. More recently, Harnik (2009) showed that SSWs are associated with upward wave activity of long duration, whereas upward wave activity of short duration tends to lead to wave reflection.

The relationship between tropopause-level upward planetary wave activity with time-varying amplitudes and forcing of SSWs is investigated in more detail in this study. Our analysis is based on both reanalysis data and simple models (the Holton–Mass model and a more severely truncated low-order version). We quantify the sensitivity of the forcing of SSWs to the time scale of the wave forcing. In the reanalysis data the duration of individual wave forcings preceding SSWs is analyzed and found to be significantly enhanced compared to the background winter climatology. Moreover, we find the strength of the wave forcing preceding SSWs to play a more secondary role: moderately strong forcings (just above the climatological mean value) are sufficient as long as they are long lasting (on the order of 1 week or longer). A similar sensitivity is found in the models, such that as the time scales of the wave forcing become shorter the number of forced SSW solutions decreases unless there is a concurrent, nonlinear increase in the magnitude of the wave amplitudes.

We will discuss the analysis procedure of and present results from the reanalysis data in section 2. In section 3 we describe the models utilized in this study and present results from two forcing experiments. Section 4 presents the conclusions of this study.

2. Observational analysis

a. Data and methods

An observational analysis meant to eke out the relationship between the time scales of wave forcing and SSWs is presented here. Stratospheric wave forcing is generally represented by the vertical component of the Eliassen–Palm (EP) flux vector, which may itself be approximated by the eddy meridional heat flux (Andrews et al. 1987; Newman and Nash 2000). Since these measures mainly differ by a static stability term that is nearly constant throughout the stratosphere, the present observational analysis is performed using the eddy meridional heat flux as a proxy for wave forcing. These heat flux data are derived from the interim European Centre for Medium-Range Weather Forecasts (ECMWF) Re-Analysis (ERA-Interim) dataset, are averaged from about 45° to 75°N as in previous studies (e.g., Newman and Nash 2000; Newman et al. 2001; Polvani and Waugh 2004) and cover a period of January 1979–February 2012 at a time resolution of 6 h. The observational analysis was also performed using 40-yr ECMWF Re-Analysis (ERA-40) data and was found to only show minor differences.

While 100 hPa is a typical choice for analyzing the relationship between eddy heat flux and SSWs (e.g., Polvani and Waugh 2004; Charlton and Polvani 2007), this level is also located within the bottom of the polar vortex. That is, at 100 hPa the wave forcing cannot be considered to be completely independent of the evolution of an SSW itself. A lower level, typically 300 hPa (e.g., Polvani and Waugh 2004; Charlton and Polvani 2007), is therefore often analyzed to more properly account for the tropospheric wave forcing entering the stratosphere. However, since the 300-hPa level is sometimes above the tropopause and sometimes below the tropopause within the range of 45°–75°N, the eddy heat flux data are no longer a proper proxy for wave activity (upward EP flux) because of differences in the static stability term between the upper troposphere and lower stratosphere. This study therefore utilizes the 200-hPa level—located below the polar vortex but above the tropopause.

The World Meteorological Organization (WMO) defines a major SSW event (hereafter simply SSW event, unless stated otherwise) as a wind reversal from westerlies to easterlies at 60°N and at 10 hPa. For the considered period of 34 winters [December–February (DJF)],[1] there are 18 winters with a total of 20 SSW events and 16 winters with no SSW events. This analysis will make use of three data periods based around the occurrence (or nonoccurrence) of an SSW event. The first period contains all winters in which no SSW

[1] Note: one of these winters is the period January–February 1979, which is included because a major warming occurred in late winter.

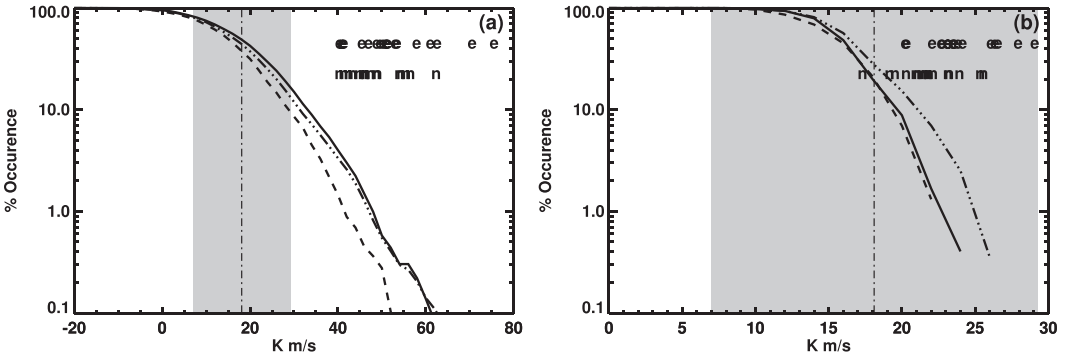

FIG. 1. Cumulative frequency distributions for (a) instantaneous heat flux and (b) 40-day averaged heat flux. In each of the plots here, the solid curve denotes the pre-event periods, the dash-triple dotted curve denotes the event winters, and the dashed curve denotes the nonevent winters (see text for definitions). Individual winter maximum heat flux values are also plotted for event winters (shown by ''e'') and for nonevent winters (shown by ''n''). The climatological winter (DJF) mean heat flux and one standard deviation are shown on each plot (dash-dotted with gray shading). Note that the *y* axes are plotted on a logarithmic scale.

occurred, hereafter termed nonevent winters. The second period contains all winters in which at least one SSW occurred, termed event winters. The third period contains the date of each SSW event and the preceding 45 days, termed the pre-event periods. The dates of the events contained in this dataset were identified by using the algorithm discussed in Charlton and Polvani (2007). For reference we note the climatological winter (DJF) mean heat flux at 200 hPa of approximately 18 K m s^{-1} (with standard deviation of 11 K m s^{-1}). For the following discussion, heat flux values larger than this background value are considered to be ''large.''

b. Results

Previous studies of SSWs show strong correlations between high values of meridional heat flux and major warming events (e.g., Baldwin and Dunkerton 1989; Polvani and Waugh 2004). We therefore first compare distributions of heat flux values for our three named periods (nonevent winters, event winters, and pre-event periods). Cumulative frequency distributions of the (6-hourly) heat flux (at 200 hPa, averaged over 45°–75°N) are shown in Fig. 1a. The pre-event periods (solid) and event winters (dash-triple dotted) display little difference throughout the range of heat flux values. In contrast, the frequency of occurrences of large heat flux values during nonevent winters (dashed) are systematically less than the frequency of occurrence during both the pre-event periods and event winters. For example, the probability to meet or exceed 20 K m s^{-1} is about 31% in nonevent winters compared to 43% in pre-event periods. The probability to meet or exceed 40 K m s^{-1} is about 1.3% in nonevent winters compared to 4.0% in pre-event periods.

However, while large heat flux values are less frequent in nonevent winters they still occur. For example, absolute counts (not shown) are similar in the range 30–40 K m s^{-1} between the pre-event periods and nonevent winters, and they drop below about 10 in each case for heat flux values greater than 50 K m s^{-1}. Furthermore, the maximum yearly heat flux values do not show a clear separation between event winters (indicated by ''e'' in the figure) and nonevent winters (indicated by ''n'' in the figure) but do show that these maxima are typically somewhat larger in event winters.

Polvani and Waugh (2004) argue that weak vortex events are highly correlated with large magnitudes of time-averaged heat fluxes, with highest correlations for time averaging of 30 days or longer. While not all weak vortex events are SSWs, all SSWs are weak vortex events. This suggests that large magnitudes of time-averaged heat flux are crucial to force SSW events. They propose that the 40-day averaged heat flux is perhaps best to capture this correlation, consistent with the earlier Newman et al. (2001) analysis. We may therefore expect a separation between occurrences of large 40-day averaged heat flux in nonevent winters and occurrences of large 40-day averaged heat flux in event winters and pre-event periods.

Figure 1b shows cumulative frequency distributions for the 40-day averaged heat flux. The 40-day averaged heat flux values are confined to within one standard deviation of the climatological mean instantaneous heat flux (indicated by gray shading). The frequency distributions show a difference between nonevent and event winters similar to that in Fig. 1a. For example, the probability of exceeding an averaged heat flux value of 20 K m s^{-1} is about 13% in nonevent winters, while it is

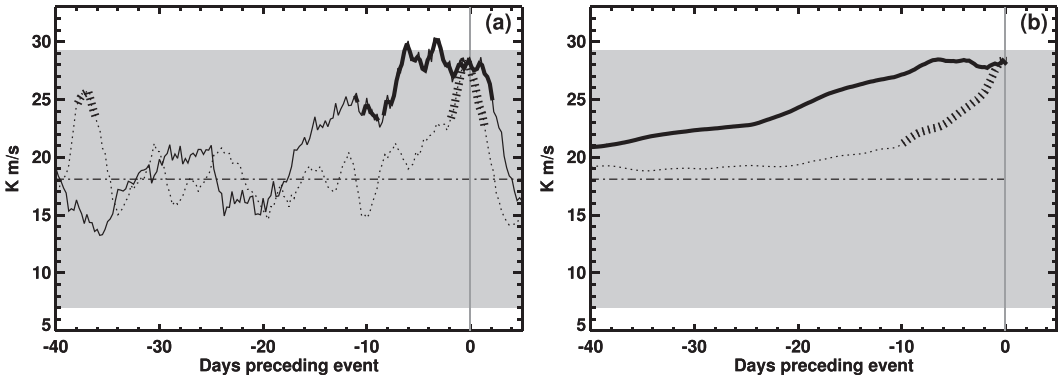

FIG. 2. (a) Composites of heat flux for the pre-event periods (solid), for the identified minor warming events (dashed), and for the climatological winter mean (dash-dotted) as a function of day preceding the event. (b) Composites of time-averaged heat flux for the same periods as in (a). The averaging is done over the number of days preceding the warming. Thick curves mark statistically significant differences—at the 95% level—relative to the full winter period. The gray shading denotes the one standard deviation range around the climatological winter mean heat flux.

about 21% in event winters. Further, the maximum yearly 40-day averaged heat flux values show a clear tendency toward larger values during event winters compared to nonevent winters. However, the pre-event period distribution follows much more closely that for nonevent winters and appears systematically different from that for event winters. A closer look into this at first surprising behavior reveals that for most SSW events the period of enhanced heat fluxes continues a few days past the event date (typically by 3–4 days), leading to largest averaged heat fluxes when these days past the event date are included. Postevent days are by definition only included in the event winters sample, but not in the pre-event periods. Modifying the pre-event periods to include 5 days past the SSW event puts the distributions for pre-event periods and event winters in much closer agreement. Nevertheless, the additional enhancement of the 40-day averaged heat flux past the SSW event date can hardly be said to precede and therefore cause the event.

The results from the 40-day averaged heat flux values reinforce the concept that SSW events are typically associated with large time-averaged heat fluxes near the tropopause. However, the 40-day averaged heat flux preceding SSWs does not appear to be strongly enhanced compared to nonevent winters. Furthermore, it is in principle not clear whether large time-averaged heat fluxes arise from an extremely large pulse of short duration within the averaging period, or from moderately large values extending over most of the averaging period. Results in Harnik (2009) suggest that short pulses preferentially lead to wave reflection with insufficient zonal wind deceleration to cause SSWs, while SSWs are associated with longer-lasting pulses. We thus more closely investigate the evolution of the heat flux preceding SSWs.

As a first step, a composite heat flux preceding SSWs is formed [similar to Fig. 3d in Harnik (2009)], averaging the heat flux relative to the SSW date and is plotted as the solid curve in Fig. 2a. Also shown in Fig. 2a are a composite minor warming heat flux (dashed; see below for minor warming definition) and the climatological winter-mean heat flux and corresponding standard deviation (dash-dotted with gray shading). The time-averaged heat flux for each of these is plotted in Fig. 2b as a function of the number of days preceding the SSW over which it is averaged.

A minor warming is identified here if the zonal-mean zonal wind at 10 hPa and 60°N drops below 20 m s^{-1} but stays above 0 m s^{-1}. That is, minor warming events are identified as weak vortex events that do not lead to a major SSW. We analyze these minor warming events so as to compare our (major) SSW heat flux composites with composites from cases of moderately strong wave forcing. The upper bound of this range is chosen because it determines periods of large departures from the climatological winter zonal-mean zonal wind speed (of approximately 30 m s^{-1} at 10 hPa and 60°N). We also impose the requirement that once the zonal-mean zonal wind becomes larger than 20 m s^{-1}, it must remain larger than 20 m s^{-1} for at least 10 days. As well, no minor warming event may fall within 20 days prior to or following a major SSW. Doing so forces these minor warming events to be well separated from other minor and major SSW events, helping separate analysis of the wave forcing affecting major SSW events and minor warming events. This minor warming identification finds 21 cases meeting these criteria in the ERA-Interim dataset.

The composite heat flux profile preceding SSWs only briefly exceeds 30 K m s^{-1} (Fig. 2a). Heat flux values of 30 K m s^{-1} roughly correspond to the climatological winter mean plus one standard deviation and the probability to exceed 30 K m s^{-1} is approximately 17% for the pre-event periods (cf. Fig. 1). While these constitute significantly elevated heat flux values, they could certainly not be classified as ''extreme'' values. Rather, it is more significant that the SSW composite profile exceeds the climatological-mean heat flux value for almost 20 days prior to the date of the event, with the statistically significant differences extending to about 10 days prior to the event date. In contrast, the composite minor warming profile shows a fairly short-lived pulse of heat flux that begins approximately 6 days prior to the minor warming event (statistical significance only extends to 2 days prior to the event). No such pulse exists in the SSW pre-event periods composite. The minor warming composite also shows a statistically significant similar pulse between 36 and 38 days preceding the event, suggesting a degree of periodicity in these events (reminiscent of vacillation cycles; e.g., Kuroda 2002) on a time scale of about 1 month.

Similar results may be seen in Fig. 2b, wherein the average heat flux for short averaging periods (which fall near to the event date) is dominated in the case of the minor warmings by the final pulse of heat flux. Statistically significant enhancements of the averaged heat flux are only found for averaging of up to 10 days preceding the minor warming event. The SSW pre-event periods composite, in contrast, shows statistically significant enhancements out to about 50 days (note that the figure only shows averaging periods up to 40 days). The averaged heat flux in this case remains relatively constant out to about 7 days of averaging and steadily decreases toward the climatological mean after that point.

Comparing the SSW pre-event composites with those from the winter mean and from our identified minor warming events gives a clearer picture of how meridional heat flux—and thus the wave forcing—evolves prior to an SSW event. In agreement with the results in Harnik (2009), the large time-averaged heat flux values that are well correlated with SSWs do not appear to arise from brief periods of extremely large heat fluxes but instead from long periods of moderately large heat fluxes. It is important to note that while the latter appears to be more common, the former in principle still remains a possibility (unless prevented through reflection events; Harnik 2009).

To test this result further, we compare the average durations over which the heat flux exceeds a given magnitude within each of the three periods considered in this study. From the analysis above, we expect that the average duration for large forcing values is significantly enhanced for the pre-event periods compared to the overall winter climatology. This comparison is accomplished by first averaging the durations for which the heat flux exceeds a specified value over the full winter period (including both event and nonevent years). To properly compare these with the pre-event periods, we first select the occurrence closest to the SSW date of an exceedance of the specified value. We constrain this final date of the heat flux exceedances to precede the SSW by 10 days or less. While we do not wish to analyze wave forcing well separated from the SSW event itself, there is typically a brief lead time for the maximum amplitude of wave forcing at 200 hPa relative to the date of the SSW. For the ERA-Interim dataset and our set of identified SSW events, the maximum heat flux leads the SSW date by about 6 days on average (with a standard deviation of about 4 days). This implies that most maximum heat flux values will fall within 10 days prior to the SSW event. Our method is not sensitive between 5 and 15 days for this chosen constraint. An additional constraint on the analyzed exceedances is that they must last longer than the resolution of the data (here 6 h) so as to disallow inclusion of possibly erroneous data points. The primary results of this comparison hold even if these point exceedances are included.

Figure 3 compares the average duration that the heat flux exceeds a given value for the pre-event periods (solid) and for all winter data (dashed).[2] Statistical significance at the 95% confidence level of the difference between these curves is assessed based on a Student's *t* test; thick lines mark regions of statistical significance. The degrees of freedom for the pre-event periods is simply the number of SSWs (20) found in the ERA-Interim dataset. The degrees of freedom per winter for the full winter period are taken to be the number of duration events if that number is below 40, and 40 otherwise [this number corresponds to the ratio of the approximate length of a winter (100 days) to the decorrelation time scale of the heat flux (about 2.5 days)]. The average duration is significantly larger (by a factor of 2–3) for the pre-event periods than for the winter climatology for heat flux thresholds between about 10–20 K m s^{-1} (recall the climatological winter value of 18 K m s^{-1}). At larger heat flux thresholds, in particular outside one standard deviation of the climatological mean, the average durations are similar between the two periods with statistically insignificant differences.

[2] The average duration for nonevent winters is only slightly reduced compared with all winters, by less than 1 day over the range of heat flux thresholds considered.

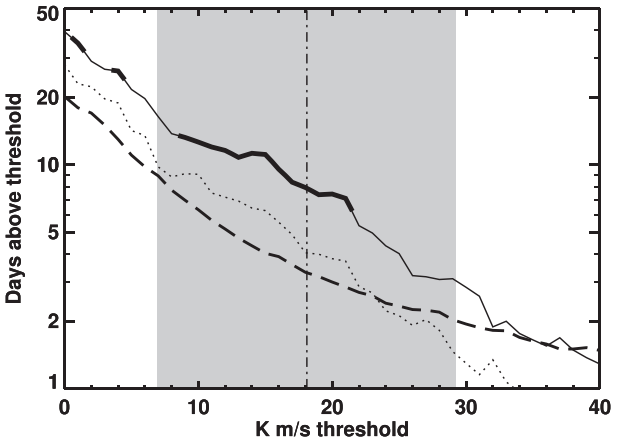

FIG. 3. Average number of days (ordinate) that the heat flux stays above a certain threshold (abscissa) for the pre-event periods (solid), for our identified minor warming events (dotted), and for the full winter period (long dashed). Thick curves mark statistically significant differences—at the 95% level—relative to the full winter period. The climatological winter mean heat flux and one standard deviation are shown with a dash-dotted line and gray shading, respectively. Note that the y axis is plotted on a logarithmic scale.

Isolated regions of statistically significant differences between the pre-event and the climatological curve are also obtained for strongly reduced heat flux thresholds (near 1 K m s^{-1} as well as ~4 K m s^{-1}), indicating that extended periods of strongly reduced heat fluxes are less likely to occur in the pre-event period.

For heat flux exceedances around the climatological mean value the average duration is on the order of 1 week in pre-event periods. This suggests that on average in order to force an SSW, the wave forcing entering the stratosphere needs to exceed climatological mean values for a period of at least 1 week. The strength of this wave forcing appears to be secondary; moderately large heat flux values, around or just above the climatological mean, appear to be sufficient. The duration for larger heat flux exceedances decays exponentially and is not much different between the overall climatology and the pre-event periods, indicating that these more extreme heat flux pulses are short-lived and less likely to force an SSW [according to Harnik (2009) short-lived wave forcings more likely lead to wave reflection events]. This analysis shows that the enhanced SSW composite heat flux values (above about 25 K m s^{-1}) during the 10 days preceding the event in Fig. 2a are a result of short-lived (2–3 days according to Fig. 3) wave pulses with slightly different timing relative to the individual SSWs. That is, forcing of the individual SSWs does not appear to be due so much to enhanced wave forcing as to the wave forcing not dropping below the climatological value for an extended period. Note that our analysis of durations for the pre-event periods by definition excludes the frequent heat flux exceedances a few days past the SSW date (according to Fig. 2a, the heat flux typically stays above the climatological value for 3–4 days past the event). Including these days past the event date would make the average duration for the pre-event periods even larger. However, heat flux exceedances past the SSW date can hardly be said to precede and therefore cause the event.

The duration analysis was also performed for the identified minor warming events (dotted line in Fig. 3). Even though these minor warmings also show enhanced durations compared to the climatology for heat flux exceedances in the range 10–20 K m s^{-1}, these are not statistically significant. Furthermore, the average durations for the minor warmings are about a factor of 2 smaller than for the (major) SSWs.[3] For large heat flux thresholds the minor warming curve drops below the climatology, largely because many minor warmings do not exhibit any exceedance (i.e., zero duration) at these large thresholds.

We have also tested the predictive capability of periods of long-lasting moderately enhanced wave forcing by identifying corresponding events in the 200-hPa heat flux time series and analyzing the zonal wind response at 10 hPa (not shown). We find, not surprisingly, that in general combinations of small heat flux thresholds (below the climatological value) with shorter minimum durations (below ~1 week) lead to a large number of such heat flux events, most of which are not followed by SSWs. Likewise, combinations of large heat flux thresholds (above the climatological value) with longer minimum durations (above ~1 week) lead to only a few heat flux events, almost all of which are followed by SSWs. However, only about half of the 20 identified SSWs in the ERA-Interim dataset could be ''predicted'' this way. A similar but slightly reduced predictive capability was achieved using time-averaged heat fluxes for time averaging of about 10 days. A more detailed analysis of the above relations is beyond the scope of the present study and represents a task for future work.

From our observational analysis we conclude that SSWs are preferentially forced by planetary wave forcing (meridional heat flux) of moderate strength that lasts 1 week or longer. The duration of the wave forcing

[3] Because of the relatively small number of degrees of freedom for both the minor and the major warming ensembles, the difference between the minor and major warming curves is only statistically significant at the 90% confidence level for a range of heat flux thresholds around the climatological mean value.

appears to be a stronger controlling factor in producing SSWs than the strength of the wave forcing, confirming and extending the results in Harnik (2009). The high correlation between time-averaged wave forcing and the occurrence of SSWs as suggested by Newman et al. (2001) and Polvani and Waugh (2004) appears to be a result of the minimum required forcing duration.

3. Simple modeling

a. The models

This study now turns to simple models of stratospheric wave–mean flow interaction to demonstrate the dependence on time scale for forcing SSW events. We consider the model of Holton and Mass (1976)—hereafter the Holton–Mass model—which describes the vertical propagation of a single planetary wavenumber and its interaction with the zonal-mean zonal wind in a quasigeostrophic β-plane channel. This model has been utilized in numerous previous studies (e.g., Chao 1985; Christiansen 2000; Hardiman and Haynes 2008; Harnik 2009) because it captures well the wave–mean flow interaction that is fundamental to forcing an SSW while only retaining the vertical dimension.

The primary model considered in this study is a low-order version of the Holton–Mass model, as introduced by Ruzmaikin et al. (2003). This low-order model derives from the Holton–Mass model by severe vertical truncation to only three vertical levels: the top and bottom boundaries (at 55 and 15 km, respectively) and one resolved middle layer. The form of and chosen parameters for this model utilized here are those of Birner and Williams (2008), except that the model radiative damping time scale (acting on the zonal wind) for our setup is about 20 days. At the cost of vertical resolution, this β-plane channel model contains the stratospheric wave–mean flow interactions of the Holton–Mass model but with only one resolved layer.

In the low-order model, the channel is centered about 60°N. This model assumes a constant bottom boundary zonal wind (here, 10 m s^{-1}) and includes simple Newtonian damping of the zonal-mean zonal wind to a radiative wind profile determined by constant vertical shear throughout the depth of the model (here, 1 m s^{-1} km^{-1}). The linearized wave equation contains only a single meridional and zonal wavenumber, and the quasigeostrophic potential vorticity equation reduces to an equation describing geopotential perturbations. The model is forced with a time-varying prescribed bottom boundary geopotential height perturbation $h(t)$ and assumes that all perturbation quantities vanish at the top boundary (while the zonal wind at the top boundary is assumed to increase at a rate equal to the vertical shear value). The zonal wind affects the amplitude of the waves, and the waves couple to the zonal-mean zonal wind through an EP flux divergence term. This coupling, and the relation with the bottom boundary condition, is such that if h is prescribed to be 0 m, then no wave–mean flow interaction can take place and the zonal-mean zonal wind relaxes to its radiative equilibrium profile. In this way, the wave–mean flow interactions modeled by this system are controlled primarily through the specified term h. Further details of our low-order model can be found in Ruzmaikin et al. (2003) and in Birner and Williams (2008).

Holton and Mass (1976) showed that for the Holton–Mass model, prescribing h either less than or greater than a critical value would result in the zonal-mean zonal wind either stabilizing to a steady state or oscillating about some value much less than the steady-state case, respectively. The existence of this bifurcation is why the Holton–Mass model is considered to qualitatively capture the dynamics of an SSW. The low-order model has also been shown to simultaneously contain two stable solution states (see Ruzmaikin et al. 2003; Birner and Williams 2008), both of which qualitatively match those from the Holton–Mass model. The SSW-like solution in the low-order model may be considered the simplest prototype SSW. Because of the large degree of similarity between the two models, we present only results using the low-order model and state here that the results from the Holton–Mass model show only small differences (Sjoberg 2010).

For the following analysis, we adopt terminology as in Birner and Williams (2008) to describe the state of the zonal-mean zonal wind. The stable state that contains zonal-mean zonal wind speeds very near the radiative equilibrium solution is called the radiative state (or radiative solution), while the stable state that contains zonal-mean zonal wind speeds considerably less than the radiative equilibrium solution is called the SSW state (or SSW solution). Because the simultaneous existence of both solutions only occurs with our choices of parameters for forcing of wavenumber 2, we present results only for wavenumber 2.

We comment here about the seeming inconsistency between the observational analysis of section 2 which considers meridional heat flux as the proxy for wave forcing and the modeling analysis which considers geopotential height perturbations as the proxy for wave forcing. At this point, to the authors' knowledge, no version of either the Holton–Mass or the low-order model exists that allows forcing by prescribed heat flux. However, we are able to diagnose the heat flux as computed by the model and compare its characteristics with

the prescribed geopotential height perturbation h. Since this study focuses on time scales of the forcing, we compared the time scales of the diagnosed heat flux with the time scales of the input h and found the differences to be small. We therefore assume that characteristics of the time scales of the forcing as represented by $h(t)$ carry over to characteristics of the forcing time scales of upward wave activity (represented by the heat flux), even though the two quantities describe different properties of upward-propagating planetary waves.

Birner and Williams (2008) show that inclusion of small-scale forcing, such as that due to quasi-random gravity wave activity, in the low-order model may produce qualitative effects. Using a white noise parameterization for this small-scale forcing, they show even small amounts of forcing to be capable of driving transitions to the SSW solution for bottom boundary wave amplitudes much less than the critical amplitude beyond which the radiative solution does not exist. Additionally, inclusion of such a randomly varying forcing allows one to produce large ensembles and therefore better model statistics. The strength of the quasi-random small-scale forcing is chosen such that it is much smaller than the strength of the planetary wave drag. The accelerations from planetary wave drag in the polar stratosphere are usually on the order of 5.0 m s^{-1} day^{-1} or larger (Randel et al. 1987). Thus we hold the parameterized small-scale (~gravity) wave drag at a maximum strength of 0.1 m s^{-1} day^{-1}. Additional details of modeling this white noise term are discussed in Birner and Williams (2008).

b. Idealized forcing

The model is forced with two forms of the incoming wave amplitude where temporal variations of the amplitudes are controlled. The first form of this amplitude forcing is an idealized wave amplitude pulse shown schematically in Fig. 4. This idealized pulse is initialized about some amplitude h_0 and is monotonically increased to a maximum wave amplitude h_{max}. The time over which the amplitude ramps to h_{max} is specified by τ_{ramp}. This system is subsequently forced with h_{max} for a specified time τ_{const}. After this, the forcing is ramped down to h_0 again over the period τ_{ramp}.

The model is initialized in the radiative solution with a bottom boundary wave amplitude of $h_0 = 80$ m and is allowed to stabilize for 50 days. Noise is introduced in the low-order model after this initialization period. Both τ_{ramp} and τ_{const} are specified at values between 2 and 50 days.

The idealized pulse mimics forcing by a quasi-stationary wave amplitude when τ_{ramp} and τ_{const} are large, whereas it mimics a highly transient wave amplitude when τ_{ramp} and τ_{const} are small. The aim of this idealized wave amplitude pulse experiment is to determine the smallest magnitude of h_{max} for which both quasi-stationary amplitudes and highly transient amplitudes force the system to the SSW solution by the end of the integration. As noise was included in these integrations, h_{max} is the smallest amplitude at which a large portion—chosen to be 75 out of 100 runs at each initial condition—of the ensemble of model runs stabilized to the SSW solution.

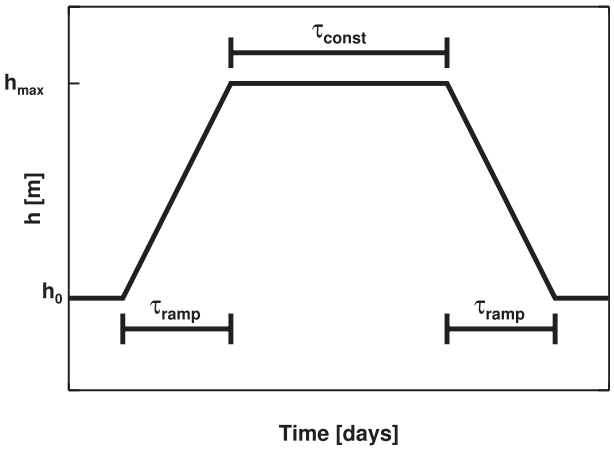

FIG. 4. Schematic of the idealized wave amplitude pulse used to force the model. The wave forcing—starting at an initial magnitude h_0—is ramped up over a specified period τ_{ramp} to a specified maximum magnitude h_{max}. The model is then forced with h_{max} over a specified period τ_{const}, followed by ramped-down forcing to h_0 over a period τ_{ramp}. For each chosen τ_{ramp} and τ_{const}, h_{max} is repetitiously increased until the model is observed to undergo a transition into the SSW solution.

Figure 5 plots contours of the smallest magnitude of h_{max} that forces an SSW solution as a function of both the ramping and the constant forcing time scales. At time scales of 10 days or longer for τ_{ramp} (τ_{const}), h_{max} does not vary much along the range of values of τ_{const} (τ_{ramp}). At time scales shorter than 10 days for τ_{ramp} (τ_{const}), h_{max} becomes large for τ_{const} (τ_{ramp}) less than 10 days. We note that this shows a similar sensitivity to approximately 10-day-long time scales as was shown in section 2.

To more closely tie these results with those from the observational analysis, Fig. 6 plots two curves of h_{max} as a function of one time scale while the other is fixed. In the reanalysis data, pulses of very large heat flux only last for short (on the order of 1 day) time scales. Thus for the solid curve, τ_{const} is fixed at 2 days such that the curve shows h_{max} as a function of τ_{ramp}. It is shown in section 2 that moderately large heat flux values with a time scale of 10 days or longer precede SSW events. Here we find

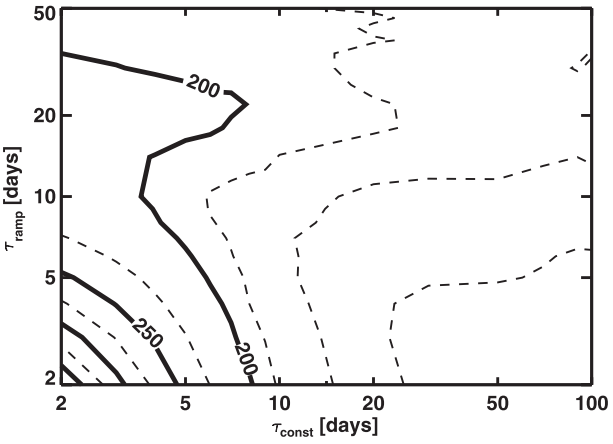

FIG. 5. Contour plot of h_{max} given as a function of the ramping period (abscissa) and of the constant forcing period (ordinate) for the low-order model. The solid contours are plotted at an interval of 50 m with the smallest solid contour being 200 m; the dashed contours to the left of 200 m are given at an interval of 25 m and the dashed contours to the right of 200 m are given at an interval of 10 m. As the time scales decrease, the h_{max} necessary to force an SSW solution increases. This effect is most notable at time scales shorter than 10 days.

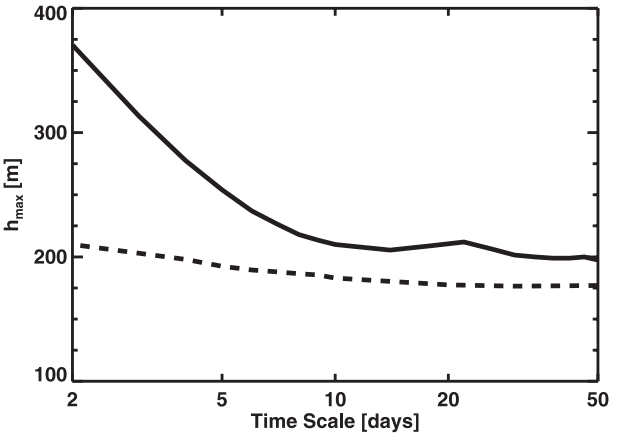

FIG. 6. The smallest h_{max} that forces an SSW in the idealized forcing experiment as a function of τ_{ramp} (solid) and as a function of τ_{const} (dashed). Here τ_{const} is set to be 2 days for the solid curve, while τ_{ramp} is set to be 10 days for the dashed curve. This shows the model sensitivity to ramping time scales of about 10 days, but also that the model is largely insensitive to the time scales of constant forcing once the ramping time scale is 10 days or longer.

from our simple, low-order model pulse experiments that h_{max} varies little beyond $\tau_{ramp} \sim 10$ days, confirming the observational result.

To further demonstrate this point, the dashed curve in Fig. 6 plots h_{max} as a function of τ_{const} where τ_{ramp} is fixed at 10 days. There is little change in the value of h_{max} across the range of τ_{const} values. This implies that the system is not sensitive to the choice of constant forcing time once the ramping time is on the order of or longer than the preferred time scale of approximately 10 days. This result suggests that—for sufficiently long ramping time—the time-varying forcing mimics constant amplitude forcing, which agrees with results from previous studies that consider slowly ramping forcing in the Holton–Mass model (Holton and Mass 1976; Yoden 1987; Christiansen 2000).

c. *Quasi-random forcing ensemble*

The second form of the controlled time-varying incoming wave amplitude is an ensemble of quasi-random wave amplitudes. The initial ensemble of amplitudes is generated by spectrally extracting zonal wavenumber 2 from the DJF 100-hPa, 60°N geopotential heights of the ERA-40 dataset. Using these extracted data produces a set of forcing profiles with highly variable time scales and physically reasonable amplitudes. While this ensemble already contains winter profiles with highly time-varying wave amplitudes, these profiles are altered such that the time scales of the amplitudes decrease. This is accomplished by first replicating and then "squeezing" the replicated occurrences of each initial profile of winter wave amplitudes onto the time period of the initial set. An example of this process is plotted in Fig. 7.

Figure 7a plots a simple sinusoid of maximum amplitude 150 m with period 180 days and an initial amplitude of 50 m. This example—but also the utilized DJF profiles—contains 50 days of ramp up to the initial amplitude (in the schematic, 50 m) and 50 days of ramp down from the final amplitude (50 m) for a total profile length of 190 days.[4] The 90 days that contain the sinusoid are then repeated after itself (we allow up to 20 replications); a single replication is shown in Fig. 7b. Since the 90-day forcing is repeated once, the profile length is now 280 days (180 days of forcing and 100 days of ramping). Since we wish to compare profiles of the same length, the 180 days of forcing in Fig. 7b are temporally interpolated down to the original 90 days. The forcing profile for a single replication of initial profile (Fig. 7a) is plotted in Fig. 7c. With each additional replication of the initial profile, the time scales of the wave amplitudes decrease while the magnitudes remain unaffected. To further illustrate this method, the profile for five replications of the example initial profile (Fig. 7a) is shown in Fig. 7d. The method outlined in this schematic example

[4] The ramping is a sinusoidal increase (decrease) from 0 m (the final amplitude) to the starting amplitude (0 m) over the 50 days of ramping.

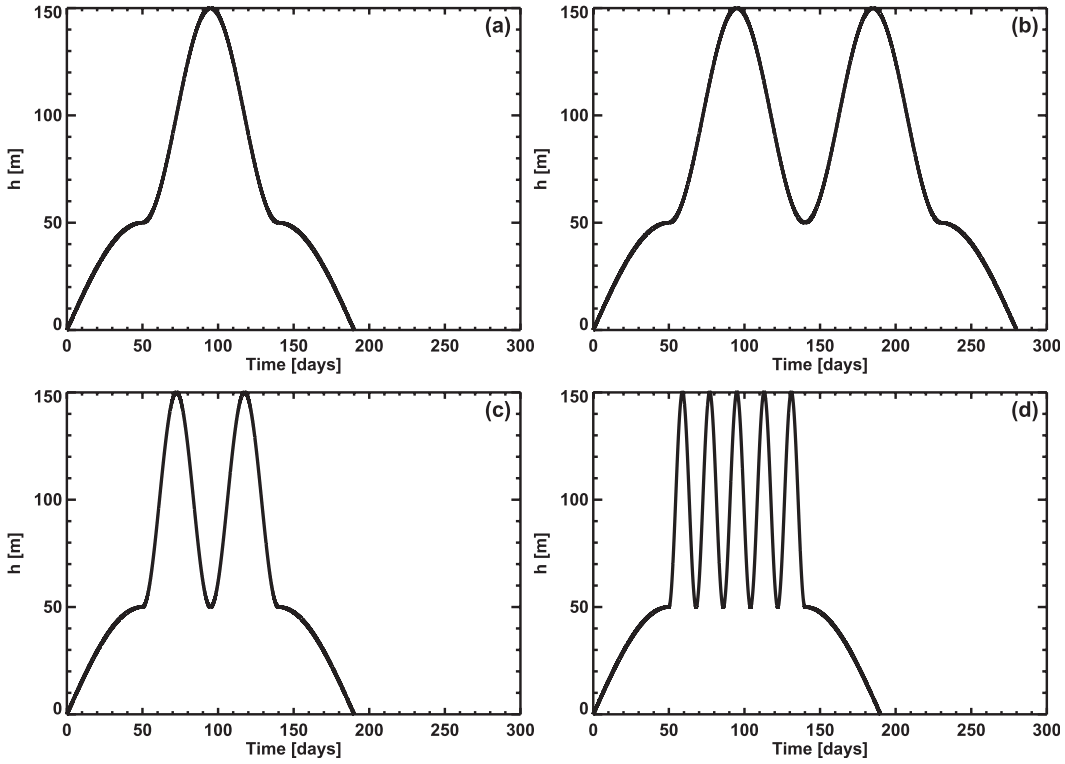

FIG. 7. Schematic of the replication experiment using a simple sinusoid for illustration (the actual replication experiment is performed using amplitudes from the reanalysis data). (a) An unaltered example profile. (b) A single replication of (a). (c) As in (b), but interpolated back to the initial period. (d) The final profile for five replications of (a). Note that the amplitudes—both of this example and of the reanalysis ensemble profiles—are not adjusted during this experiment.

is applied to the actual ensemble of forcing profiles obtained from ERA-40 (see above).

The aim of forcing the simple models with replications of the quasi-random wave amplitude ensemble is to determine how the number of forced SSW solutions varies as the time scales of the forcing amplitudes decrease. Since there is no associated alteration to the magnitudes of the amplitudes, one should expect from the idealized forcing experiment that as the time scales decrease—or as increasing replications are performed on the initial ensemble—the number of forced SSW solutions should decrease.

Figure 8 plots the number of forced SSW solutions as a function of the number of replications performed for the low-order model. It is noted here that zero replications correspond to the results from the unaltered wave amplitude time series. Since the low-order model is integrated 100 times at each initial condition because of the inclusion of noise and the ensemble contains 45 initial profiles, we are able to evaluate 4500 runs for each replication. As shown in the figure, as the time scales of the planetary wave amplitudes decrease while the magnitudes are unaltered, the number of forced SSW solutions decreases. This confirms our results from the idealized forcing experiments in the previous subsection.

One observes an interesting local enhancement of forced SSW solutions between five and six replications in Fig. 8. This behavior is interpreted to be a quasi-resonant response in the zonal-mean zonal wind from forcing the model near its internal damping time scales. We find this quasi-resonance to increase or decrease in amplitude as the parameters of the systems are tuned to longer or shorter time scales (not shown), which suggests that the quasi-resonance is an artifact of the simple system.

We note that through this replication method, the time-integrated geopotential perturbation entering at the bottom boundary is different for each number of replications performed. This experiment was also performed where the magnitudes were adjusted such that the integrated squared geopotential perturbation was the same between the initial profile and its associated replication profiles (not shown). Corresponding

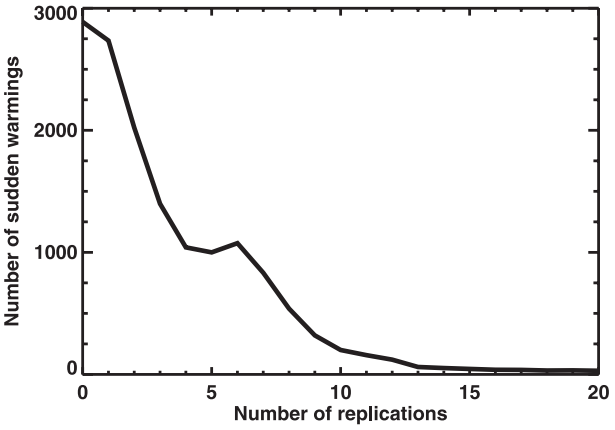

FIG. 8. The number of forced SSW solutions in the low-order model as a function of the number of replications performed on the reanalysis ensemble of quasi-random wave amplitudes. The zero-replications point denotes results from the unaltered ensemble. As the number of replications increases, the time scales of the wave amplitudes decrease while the number of SSW solutions forced in these integrations decreases.

results show only small differences from those presented here.

4. Conclusions

Characteristics of transient wave forcing leading to major stratospheric warming events are investigated in reanalysis data and simple models of stratospheric wave–mean flow interaction. The largest values of wave activity entering the stratosphere are not observed to vary dramatically between years with and without SSW events. Previous studies (Newman et al. 2001; Polvani and Waugh 2004) have shown the 40-day averaged heat flux values to correlate well with SSW events. We find a large part of the difference in 40-day averaged heat flux between nonevent winters and event winters to come from heat flux enhancements past the event date. These enhancements past the event date can hardly be said to cause the SSW event. Furthermore, the correlation between 40-day averaged heat flux and the occurrence of SSWs does not fully detail the evolution of the wave forcing prior to an SSW. This wave forcing could primarily take one of two forms: a large but rapid burst or a period of moderately large values. By compositing around SSW events, it is shown that the latter of the two forms—a sufficiently long period of anomalously large wave forcing—is most typical [in agreement with Harnik (2009)]. An additional composite for our identified minor warming events showed no such long-period anomalous wave forcing.

We verify these results by examining the duration of the heat flux exceeding a given threshold for the periods leading up to an SSW. For anomalously large heat flux thresholds (around or just above the climatological mean), this duration is found to be significantly larger during these pre-event periods (~1 week) compared with the all winter mean (~3 days). This difference in forcing duration becomes statistically insignificant for wave forcings stronger than about half a standard deviation above the climatological value, showing that strong and in particular extreme wave forcings are generally not very long lived preceding SSWs. This suggests that the strength of the forcing plays a secondary role in forcing SSWs; rather it appears crucial for the forcing to maintain a minimum strength (around or just above the climatological value) over a sufficiently long period. It is important to note that the above results are of statistical nature: large variations in forcing duration and strength preceding SSWs exist. Nevertheless, our results indicate certain characteristics of the wave forcing that appear favorable for producing SSWs. SSWs have often been linked to tropospheric blocking events (e.g., Martius et al. 2009). Blocking events are characterized by long time scales, consistent with our findings. However, the majority of blocking events do not produce an SSW, suggesting that internal stratospheric characteristics play an important role in allowing tropospheric wave activity of long time scale to propagate sufficiently deep into the stratosphere (cf. Scott and Polvani 2004; see also arguments in last paragraph below).

This sensitivity of the polar vortex to wave forcing with time scales of 10 days or longer is subsequently modeled using a low-order model of stratospheric wave–mean flow interaction. First, the model is forced with an idealized profile of wave amplitudes where both the time over which the profile ramps to its maximum amplitude and the time over which the model is forced with the maximum amplitude are controlled. This idealized forcing experiment shows that if the time over which the model is forced with the maximum amplitude is about 10 days or longer, the model behaves as if steady forcing is imposed. A similar result comes about when the ramping time scale of the idealized forcing is about 10 days or longer. For short-time-scale forcing the system requires significantly larger magnitudes of wave forcing to drive the system into the SSW solution. Second, the model is forced with an ensemble of wave amplitude profiles, the time scales of which are iteratively decreased with no corresponding alteration to the magnitudes of the waves. As the time scales of the forcing in the ensemble become shorter, the number of forced SSW solutions strongly decreases.

From both simple modeling experiments it is shown that, in order to force an SSW, the strength of the wave forcing must increase in a nonlinear fashion as the time scale of the wave forcing decreases. It is interesting to note that the critical forcing time scale to force SSWs as found by the simple, low-order model (~10 days) is similar to that found in our observational analysis (~7 days). Further, both the simple modeling and observational analyses suggest that the forcing time scales play a more important role in forcing SSWs than does the strength of the forcing. While the above listed agreement between the two approaches used here is encouraging, we note that there still remains uncertainty about whether this is a coincidental or physical agreement due to the simplicity of the model used.

The abovementioned nonlinearity in the simple model experiments is not as demonstrable in the observational results. From Harnik (2009), we expect that extremely large but short time-scale wave pulses will lead to wave reflection events. It is wave absorption—rather than reflection—that leads to the heating and jet deceleration associated with SSWs. Thus since these reflection events do not lead to SSWs, we are not able to identify SSWs that are forced by such extremely large but short-lived wave pulses. Instead, the ability of the simple model to still force SSWs with extreme but short-duration wave pulses may be suggestive of deficiencies in the model associated with realistically capturing wave reflection. On the other hand, the extreme forcings required to produce SSWs with short-duration wave pulses may simply not exist in the real atmosphere.

It is presently unclear why SSWs are preferentially produced by moderately strong forcing of long duration. Harnik (2009) finds that while short pulses of wave forcing lead to wave reflection events, these reflection events in turn are not causal in limiting the duration of the wave forcing. Results using simple models of the Holton–Mass type suggest that the simple interaction of a strictly vertically propagating monochromatic planetary wave with the mean flow captures the salient ingredients of the generation of SSWs by long-duration wave forcing. The only naturally occurring time scale in these simple models is the radiative relaxation time scale (e.g., in our model setup the zonal wind in the middle stratosphere is damped with ~20 days). Radiation acts to restore the vortex toward radiative equilibrium on this relaxation time scale. Hence, longer radiative relaxation time scales are expected to allow shorter wave pulses to force SSWs, whereas shorter radiative relaxation time scales are expected to require longer wave pulses to force SSWs. We have confirmed these expectations in our simple model by performing sensitivity calculations using longer and shorter radiative time scales. These simple model results suggest that the required forcing duration to produce SSWs is set by internal stratospheric characteristics.

Acknowledgments. The ERA-40 and ERA-Interim data used in this study have been obtained from the ECMWF data server. This research was supported by funding provided by the Climate and Large-Scale Dynamics Program of the U.S. National Science Foundation. We greatly appreciate the constructive criticism, as well as the many helpful comments and suggestions provided by the reviewers.

REFERENCES

Andrews, D. G., J. R. Holton, and C. B. Leovy, 1987: *Middle Atmosphere Dynamics*. Academic Press, 489 pp.

Baldwin, M. P., and T. J. Dunkerton, 1989: The stratospheric major warming of early December 1987. *J. Atmos. Sci.,* **46,** 2863–2884.

Birner, T., and P. D. Williams, 2008: Sudden stratospheric warmings as noise-induced transitions. *J. Atmos. Sci.,* **65,** 3337–3343.

Chao, W. C., 1985: Sudden stratospheric warmings as catastrophes. *J. Atmos. Sci.,* **42,** 1631–1646.

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; Corrigendum, **24,** 5951.

Christiansen, B., 2000: Chaos, quasiperiodicity, and interannual veriability: Studies of a stratospheric vacillation model. *J. Atmos. Sci.,* **57,** 3161–3173.

Geisler, J. E., 1974: A numerical model of the sudden stratospheric warming mechanism. *J. Geophys. Res.,* **79,** 4989–4999.

Hardiman, S. C., and P. H. Haynes, 2008: Dynamical sensitivity of the stratospheric circulation and downward influence of upper level perturbations. *J. Geophys. Res.,* **113,** D23103, doi:10.1029/2008JD010168.

Harnik, N., 2009: Observed stratospheric downward reflection and its relation to upward pulses of wave activity. *J. Geophys. Res.,* **114,** D08120, doi:10.1029/2008JD010493.

Held, I. M., 2005: The gap between simulation and understanding in climate modeling. *Bull. Amer. Meteor. Soc.,* **86,** 1609–1614.

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.

——, and C. Mass, 1976: Stratospheric vacillation cycles. *J. Atmos. Sci.,* **33,** 2218–2225.

Kuroda, Y., 2002: Relationship between the polar-night jet oscillation and the annular mode. *Geophys. Res. Lett.,* **29,** 1240, doi:10.1029/2001GL013933.

Martius, O., L. M. Polvani, and H. C. Davies, 2009: Blocking precursors to stratospheric sudden warming events. *Geophys. Res. Lett.,* **36,** L14806, doi:10.1029/2009GL038776.

Matsuno, T., 1971: A dynamical model of the sudden stratospheric warming. *J. Atmos. Sci.,* **28,** 1479–1494.

Newman, P. A., and E. R. Nash, 2000: Quantifying the wave driving of the stratosphere. *J. Geophys. Res.,* **105,** 12 485–12 497.

——, ——, and J. Rosenfield, 2001: What controls the temperature of the Arctic stratosphere during the spring? *J. Geophys. Res.,* **106,** 19 999–20 010.

Plumb, R. A., and K. Semeniuk, 2003: Downward migration of extratropical zonal wind anomalies. *J. Geophys. Res.,* **108,** 4223, doi:10.1029/2002JD002773.

Polvani, L. M., 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.

Randel, W. J., D. E. Stevens, and J. L. Stanford, 1987: A study of planetary waves in the southern winter troposphere and stratosphere. Part II: Life cycles. *J. Atmos. Sci.,* **44,** 936–949.

Ruzmaikin, A., J. Lawrence, and C. Cadavid, 2003: A simple model of stratospheric dynamics including solar variability. *J. Climate,* **16,** 1593–1600.

Scherhag, R., 1952: Die explosionsartigen Stratosphärenerwärmungen des Spätwinters 1951–52. *Ber. Dtsch. Wetterdienstes,* **6,** 51–63.

Scott, R. K., and L. M. Polvani, 2004: Stratospheric control of upward wave flux near the tropopause. *Geophys. Res. Lett.,* **31,** L02115, doi:10.1029/2003GL017965.

——, and ——, 2006: Internal variability of the winter stratosphere. Part I: Time-independent forcing. *J. Atmos. Sci.,* **63,** 2758–2776.

——, ——, and D. W. Waugh, 2008: Internal variability of the winter stratosphere. Part II: Time-dependent forcing. *J. Atmos. Sci.,* **65,** 2375–2388.

Sjoberg, J. P., 2010: Low-order models of sudden stratospheric warmings. M.S. thesis, Department of Atmospheric Science, Colorado State University, 89 pp.

Yoden, S., 1987: Bifurcation properties of a stratospheric vacillation model. *J. Atmos. Sci.,* **44,** 1723–1733.