

Comparing Sudden Stratospheric Warming Definitions in Reanalysis Data*

FROILA M. PALMEIRO

Departamento de Física de la Tierra II, Universidad Complutense de Madrid, Madrid, Spain

DAVID BARRIOPEDRO AND RICARDO GARCÍA-HERRERA

Departamento de Física de la Tierra II, and Instituto de Geociencias, Centro Mixto del Consejo Superior de Investigaciones Científicas, Universidad Complutense de Madrid, Madrid, Spain

NATALIA CALVO

Departamento de Física de la Tierra II, Universidad Complutense de Madrid, Madrid, Spain

(Manuscript received 23 December 2014, in final form 15 June 2015)

ABSTRACT

Sudden stratospheric warmings (SSWs) are characterized by a pronounced increase of the stratospheric polar temperature during the winter season. Different definitions have been used in the literature to diagnose the occurrence of SSWs, yielding discrepancies in the detected events. The aim of this paper is to compare the SSW climatologies obtained by different methods using reanalysis data. The occurrences of Northern Hemisphere SSWs during the extended-winter season and the 1958–2014 period have been identified for a suite of eight representative definitions and three different reanalyses. Overall, and despite the differences in the number and exact dates of occurrence of SSWs, the main climatological signatures of SSWs are not sensitive to the considered reanalysis.

The mean frequency of SSWs is $6.7 \text{ events decade}^{-1}$, but it ranges from 4 to 10 events, depending on the method. The seasonal cycle of events is statistically indistinguishable across definitions, with a common peak in January. However, the multidecadal variability is method dependent, with only two definitions displaying minimum frequencies in the 1990s. An analysis of the mean signatures of SSWs in the stratosphere revealed negligible differences among methods compared to the large case-to-case variability within a given definition.

The stronger and more coherent tropospheric signals before and after SSWs are associated with major events, which are detected by most methods. The tropospheric signals of minor SSWs are less robust, representing the largest source of discrepancy across definitions. Therefore, to obtain robust results, future studies on stratosphere–troposphere coupling should aim to minimize the detection of minor warmings.

1. Introduction

The winter stratospheric polar circulation is characterized by strong westerly winds referred to as the polar vortex. This circulation is disturbed by upward

propagating waves from the troposphere that dissipate in the stratosphere (e.g., Andrews et al. 1987). An extreme manifestation of this wave–mean flow interaction can lead to a dramatic weakening of the polar vortex and a rapid warming of the polar stratosphere (e.g., Matsuno 1971), referred as a sudden stratospheric warming (SSW). This phenomenon was detected for the first time during the 1952 winter (Scherhag 1952). SSWs are a clear manifestation of stratosphere–troposphere coupling, and the downward propagation of the anomalies from the stratosphere to the troposphere after SSW occurrence can be used to improve the Northern Hemisphere winter weather forecasts (e.g., Sigmund et al. 2013). This has launched international initiatives that aim to better understand the precursor

* Supplemental information related to this paper is available at the Journals Online website: <http://dx.doi.org/10.1175/JCLI-D-15-0004.s1>.

Corresponding author address: Froila M. Palmeiro, Departamento de Astrofísica y Ciencias de la Atmósfera, Facultad de Física, Universidad Complutense de Madrid, Plaza Ciencias, s/n, Ciudad Universitaria, 28040 Madrid, Spain.
E-mail: fm.palmeiro@fis.ucm.es

TABLE 1. Methodologies for SSW detection. For each method, the first six columns show the original reference, the acronym used in the text, the winter period of SSW detection, the basic criterion, and the region and pressure level employed in the definition. The last column indicates the minimum time interval demanded between two consecutive events.

Definition	Acronym	Period	Criterion	Latitude	Level	Time of recovery
WMO (McInturff 1978)	U&T ^a	Nov–Apr	$\bar{u}_{60N} < 0 \text{ m s}^{-1}$	60°–90°N	10 hPa	20 days
Baldwin and Dunkerton (2001)	EOFz ^b	Nov–Apr	NAM < -3.0	20°–90°N	10 hPa	60 days
Limpasuvan et al. (2004)	EOFu ^c	Oct–Apr	SZI < -1.0	20°–90°N	50 hPa	40 days
Taguchi and Hartmann (2005)	Tanom ^d	Oct–May	ZM Temperature anomalies $> 30 \text{ K}$	88°N	10 hPa	60 days
Kodera (2006)	Urate ^e	Dec–Feb	$\Delta[\bar{u}_{50-70N}]_{7\text{days}} < -2.0 \text{ m s}^{-1}$	50°–70°N	10 hPa	—
Nakagawa and Yamazaki (2006)	Trate ^f	Oct–May	$\Delta[\bar{T}_{80-90N}]_{6\text{days}} > 20 \text{ K}$	80°–90°N	10, 20, 30 hPa	30 days (the weakest event is discarded)
Charlton and Polvani (2007)	U60	Nov–Mar	$\bar{u}_{60N} < 0 \text{ m s}^{-1}$	60°N	10 hPa	20 days
Seviour et al. (2013)	MOM	Dec–Mar	Centroid latitude $< 66^\circ\text{N}$ (displacements) Aspect ratio > 2.4 (splits)	50°–90°N	10 hPa	30 days

^a We herein adopt the Limpasuvan et al. (2004) criterion, according to which the temperature difference has to be positive for at least 5 days within the period from 10 days before to four days after the first day of wind reversal. Note that U&T is basically the WMO definition with some added specifications.

^b See text for details.

^c The SZI index is calculated as the PC1 of the daily ZMZW area-weighted anomalies at 50 hPa, north of 20°N and from October to April. The resulting PC1 is then standardized for the same period and low-pass filtered with a cutoff value of 15 days. The midpoint since the 15-day low-passed SZI drops below -1 standard deviation and raises again above -1 standard deviation is the detection date in the original definition. We take instead the first day when the 15-day low-passed SZI drops below -1 standard deviation (see text for details).

^d The onset date is defined when the temperature anomaly becomes maximum.

^e For each day, the ZMZW tendency, U_t , is constructed from centered running means as the difference between two seven-day mean values spaced by eight days. The onset date is the day with the largest deceleration among those below $-2 \text{ m s}^{-1} \text{ day}^{-1}$.

^f The area-weighted zonal mean temperature north of 80°N is first computed at 10, 20, and 30 hPa. The warming rate is computed at each level. The onset date occurs when the warming rate at 10 hPa is maximum as long as it exceeds 20 K per 6-day period in one of the three levels. If the criterion is only satisfied for one day at 10 hPa the event is discarded.

forcings, the underlying dynamics, and the potential predictive skill of these extreme events, such as the Stratospheric Network for the Assessment of Predictability (SNAP; e.g., Tripathi et al. 2015).

The World Meteorological Organization (WMO) distinguishes between two types of events: 1) major midwinter warmings, characterized by a “complete circulation reversal,” and 2) minor warmings, with “limited circulation changes” (WMO/IQSY 1964). Based on this general form of the WMO definition, minor warmings have traditionally been detected as a reversal of the meridional temperature gradient over the polar cap at 10 hPa, whereas an additional reversal of the zonal-mean zonal wind (ZMZW) at 10 hPa is often required for major warmings (e.g., Labitzke 1981). On the other hand, the term “final warming” is often employed to refer to those SSWs that do not display a return to westerly winds and hence mark the transition to the easterly summer circulation (e.g., Labitzke and Naujokat 2000). In the last decade, many authors have identified SSWs modifying the former WMO definition or applying different diagnostic variables (Table 1). Here, we examine whether the climatological signatures of SSWs

depend on the definition used. Thus, we review all definitions of SSWs found in the literature, including those publications that do not deal specifically with SSWs, but with polar vortex extreme events in general. Figure 1 summarizes these definitions and classifies them according to the nature of the basic field used in the diagnosis and the specific methodology applied. Some definitions only refer to major SSWs, although most methods do not discriminate between major and minor events. Several methods include final warmings, while others filter them out by imposing conditions to the timing and characteristics of the events. All these differences highlight the different perceptions of SSWs, and contribute to the discrepancies in the detected events. Furthermore, some methods allow differentiation of events in types, according to 1) the morphology of the polar vortex, which leads to displacement SSWs (in which the vortex is displaced off the pole), and splitting SSWs (when the polar vortex is divided into two pieces) (e.g., Andrews et al. 1987), and 2) the dominant wavenumber signatures in the polar stratosphere preceding the SSW, which leads to events of wavenumbers 1 to 2 (e.g., Bancelá et al. 2012; Barriopedro and Calvo 2014).

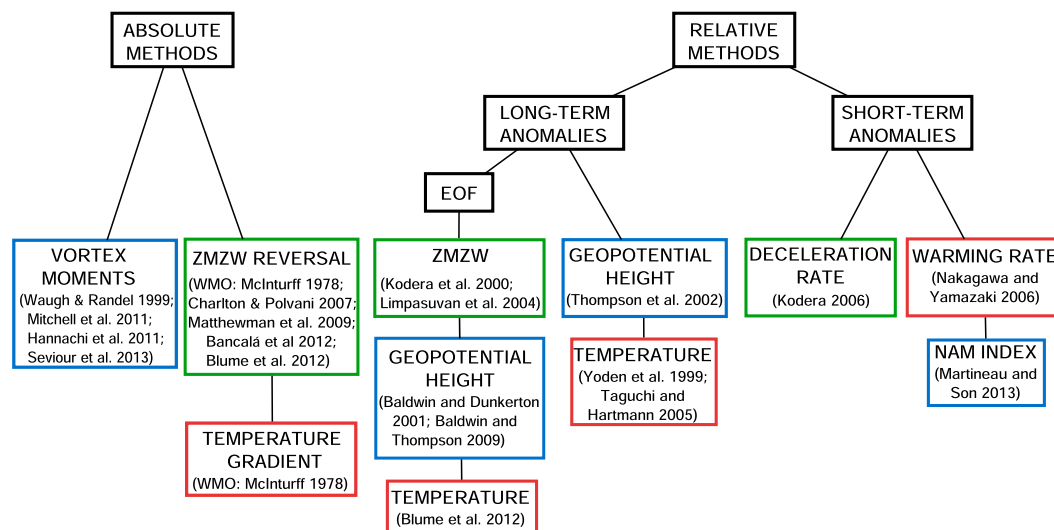


FIG. 1. Review of SSW detection methodologies. Methods are classified according to the nature (absolute and relative fields) and the specific diagnostic applied in the detection (colored boxes). The color indicates the field used to identify the events: geopotential height (blue), zonal-mean zonal wind (green), and temperature (red). (See text for details).

The first category of methods shown in Fig. 1 includes those based on imposed thresholds over absolute fields. Within this group, and similar to the WMO's definition of major SSWs (U&T; see acronyms for the eight methods listed in Table 1), many authors consider exclusively the 10-hPa ZMWZ reversal at 60°N to diagnose the occurrence of SSWs (e.g., Charlton and Polvani 2007; Matthewman et al. 2009; U60). There are also more sophisticated methods such as those based on vortex moments (e.g., Vaugh and Randel 1999; Hannachi et al. 2011). In particular, Mitchell et al. (2011) perform elliptical diagnoses of the polar vortex through potential vorticity (PV) fields to diagnose SSWs, and this methodology has been recently adapted to 10-hPa geopotential height input data, yielding similar results (Seviour et al. 2013; MOM).

Definitions based on relative fields appear in the right-hand side of Fig. 1. These methods do not distinguish between major and minor SSWs and can, in turn, be classified into two groups, depending on whether the departure fields are defined as 1) anomalies with respect to a climatological long-term mean or 2) rates of change (i.e., tendency), computed as the difference between two consecutive short-term periods, ranging from one day to one week. The first group of definitions considers methods that impose thresholds on the anomaly field (e.g., Yoden et al. 1999; Thompson et al. 2002; Taguchi and Hartmann 2005; Tanom) and those based on principal component analysis (PCA). Among those involving PCA, Baldwin and Dunkerton (2001; EOFz) use the northern annular mode (NAM) index, defined as the

projection of the geopotential height anomalies at 10 hPa onto the first empirical orthogonal function (EOF) pattern. Similar to Kodera et al. (2000), Limpasuvan et al. (2004; EOFu) employ the stratospheric zonal index (SZI), which is defined as the first principal component (PC1) of the ZMWZ latitudinal distribution at 50 hPa, while Blume et al. (2012) use the PC1 of the 10-hPa polar cap temperature. Methods based on short-term tendencies include the definition of Nakagawa and Yamazaki (2006; Trate), which sets a minimum warming rate at several pressure levels, and that of Martineau and Son (2013), employing the NAM index tendency at 10 hPa to identify SSWs. Finally, Kodera (2006; Urute) demands a minimum deceleration rate of the 10-hPa ZMWZ over the polar cap.

It is therefore clear that these methods differ not only in the basic field employed to detect SSWs, but also in the data treatment (zonal means, anomalies, etc.), the specific region of the polar stratosphere considered (i.e., a given latitude or the polar cap average, the vertical level chosen), and the different nature of the events (i.e., minor, major, and final warmings). Some of these issues have been noticed by Butler et al. (2015), who found differences in the total frequency of the events resulting from small changes in the demanded criteria. Most of the methodologies have been applied to reanalysis data, and some differences have also been obtained for different reanalysis products, revealing that the specific reanalysis can be an additional source of discrepancy. In fact, different reanalyses may involve time lags in the detection of the same event and different frequencies of occurrence

(e.g., [Charlton and Polvani 2007](#)) and hence potential differences in the SSW signatures.

The aim of our study is to perform a systematic comparison of the SSW definitions used in the literature in reanalysis datasets. We have applied the original methods (or slightly modified versions, for the sake of fair comparisons) to three different reanalyses over the same time period ([section 2](#)). To assess whether the SSW signatures are sensitive to the chosen definition, an intercomparison exercise is performed among all methods, focusing on the intraseasonal and decadal distributions of events ([section 3a](#)), the SSW characteristics in the middle stratosphere ([section 3b](#)), the downward propagation anomalies, and the surface signals before and after events ([section 3c](#)). Conclusions are presented in [section 4](#).

2. Data and methods

We have used daily mean data from 1958 to 2014 from the NCEP–NCAR ([Kalnay et al. 1996](#)), the JRA-55 ([Ebita et al. 2011](#)) and the ERA (ERA-40 for 1957–2002 plus ERA-Interim for 2002–14; [Uppala et al. 2005](#); [Dee et al. 2011](#)) reanalyses. All datasets were first interpolated to a common regular grid of $2.5^\circ \times 2.5^\circ$ spatial resolution. The basic fields computed in this study include zonal means of temperature, zonal wind, and geopotential height at various vertical levels, as required from the different definitions of SSWs. In addition, mean sea level pressure (MSLP) anomalies and several products were derived at daily time scales. They include the zonal mean meridional eddy heat flux $\overline{v'T'}$ at 100 hPa averaged over 45° – 75° N (where the overbar indicates the zonal mean and the primes deviations from it) and the NAM index. To calculate this index we use the daily anomalies of the zonal mean geopotential height north of 20° N for the entire year. Then, we perform a PCA for each pressure level separately, and the resulting PC1 (standardized for the whole year) is taken as the NAM index. In all the results presented here, latitudinal averages are always weighted by the cosine of latitude, and anomalies are computed with respect to a daily-based climatology over the 1958–2014 period. Different ways of merging ERA data products were tested, all leading to similar results. An additional comparison of the ERA-40 and ERA-Interim reanalyses for their common period (1979–2002) revealed negligible differences in the results of this study.

We have used eight definitions of SSWs (see [Table 1](#)), which are considered representative of all of those shown in [Fig. 1](#): U&T, EOFz, EOFu, Tanom, Urate, Trate, U60, and MOM. The detection of SSWs has been carried out by applying the original definitions given in the corresponding papers for the three reanalyses

(except MOM, for which the onset dates of SSWs in ERA were directly provided by the authors). Although the WMO distinguishes between major and minor SSWs, we only used its definition for major events since the inclusion of minor warmings led to a disproportionate number of SSWs as compared with the rest of the methods. We are aware that the U60 and U&T definitions can be considered redundant as they both are based on the reversal of the ZMWZ at 10 hPa. However, the U&T definition additionally requires a reversal of the temperature gradient, and, given the popular use of these definitions, we decided to include both in our analysis. Note also that all methods except Urate demand a minimum time interval between consecutive events that ranges from 20 to 60 days. The Urate definition instead picks for each winter the event with the largest wind deceleration among those satisfying its criteria, so only one event per winter can be detected.

The dates of detection of SSWs will be referred to hereafter as onset dates. In some methods, the onset corresponds to the day with the largest value of the diagnostic parameter, while in others it is defined as the first time the required conditions are satisfied. In this regard, some minor modifications were introduced in some original definitions to provide a fair comparison across methods. EOFu required a readjustment in the definition of the onset dates of SSWs since there was a systematic lag of about 20 days in the dates of the events in comparison with the other methods. This is not surprising, since in their original study, [Limpasuvan et al. \(2004\)](#) already denoted the beginning of the SSW as the $[-37, -23]$ -day period before the detection date. This is the midpoint between the day when the SZI exceeded -1 standard deviation and the day when the SZI returned to values below that threshold. However, this methodology depends on the persistence of the event and hence it can depart considerably from the timing used in the other definitions (the beginning or the peak of the anomalous period). Thus, in our study, and for the EOFu definition only, we settled the onset of the warming as the first day the SZI becomes lower than -1 standard deviation, which yields results more comparable with the other methods. In the case of EOFz, we followed the methodology described by [Baldwin and Dunkerton \(2001\)](#), although we have taken unfiltered data for the entire year. In addition, zonal mean geopotential height anomalies have been used to obtain the first EOF, instead of the full 2D field that was employed in the original study, as recommended later by [Baldwin and Thompson \(2009\)](#).

U60 is the only method that explicitly defines final warmings as those for which the ZMWZ does not return to westerlies for at least 10 consecutive days before

30 April. We have applied this criterion to all methods in order to identify and exclude these events from our analyses. Note that Tanom and Urate do not need this consideration because their period of detection ends in February. Table S1 in the supplementary material (available online at <http://dx.doi.org/10.1175/JCLI-D-15-0004.s1>) lists the SSWs identified by the different definitions, with events in bold indicating final warmings. Note that there are SSWs that are detected by several methods, albeit with different onset dates. These events will be hereafter referred to as common events, and appear in the same row of the table. For all definitions, events reaching the wind reversal (according to the U60 definition) are denoted as major SSWs. The remaining events will be classified as minor SSWs, even if they do not satisfy the WMO temperature gradient condition. However, similar results are obtained if the minor warming group only includes those events that are catalogued as such by the WMO. A separated analysis between major and minor SSWs will be performed when indicated. Otherwise, all events in Table S1 except final warmings will be considered.

Note that our study does not classify the events with respect to either the spatial structure of the stratospheric polar vortex (i.e., vortex splits and displacement SSWs) or the preconditioning of the polar vortex (i.e., events of wavenumber 1 and 2). This is because there is not a unique criterion to perform these classifications. For example, Charlton and Polvani (2007) and Mitchell et al. (2011) have their own criteria to classify SSWs into splitting/displacement events, and discrepancies in the classification of their common events were reported in the latter (Mitchell et al. 2011, their Table 1). Additionally, the split/displacement catalogue is sensitive to the reanalysis product (Charlton and Polvani 2007, their Table 1). Consequently, the arrangement of SSWs by their type is not consistent across reanalyses and methods, and would add unnecessary complexity to the intercomparison exercise.

In the following analyses, two types of composites will be used. The first is an SSW-based composite, which is specific for each definition according to its detected events. All SSWs are included in the composites of each method, regardless of its winter period, unless otherwise stated. Our results hold when the analysis is performed over the December–February period (common to all definitions). The second is a multimethod mean (MMM), constructed from the SSW-based composites of all methods derived from the first type. Similar results were obtained using other compositing approaches that minimize the influence of outliers (e.g., scaled composites weighted by the standard deviation). The standard deviation of a SSW-based composite (intramethod spread)

will be denoted as σ , while σ_M will indicate the intermethod spread associated to the MMM. To assess the statistical significance of the first type, we compute a Monte Carlo test of 1000 samples with the same number of cases as in the composite. In each sample, the days and months of the selected cases are fixed to those of the original SSW onset dates and only the years are chosen randomly. The signal is statistically significant when the corresponding value in the SSW-based composite is outside of the 5th to 95th percentiles of the Monte Carlo distribution. The robustness of the MMM signal is assessed by computing the percentage of methods that agree on the sign and significance. The SSW signal is considered robust across definitions when the agreement is higher than 75%.

3. Results

Our analyses have been performed on the eight selected definitions and applied to ERA, JRA-55, and NCEP–NCAR reanalysis data for the 1958–2014 period. We have chosen these three datasets because they are the only ones that include stratospheric data and extend back beyond 1979. While for several methods we have identified differences across datasets in the exact dates of SSW occurrence or even in the number of the detected SSWs, a pairwise *t* test comparison of the reanalysis results for the decadal frequency of SSWs revealed no significant differences at the 95% confidence level in any of the methods analyzed in this study. In addition, the results shown later are not sensitive to the reanalysis product, and hence the conclusions of this paper are not affected by the reanalysis used, which is in agreement with Martineau and Son (2010). Given that one of the methods is only available for the ERA datasets, we will only show results from this reanalysis, unless otherwise stated. Some of the corresponding results for the JRA-55 and NCEP–NCAR datasets can also be found in the supplementary material. Further comparison among reanalysis products will be included in the ongoing Stratosphere–Troposphere Processes and Their Role in Climate (SPARC) Reanalysis Intercomparison Project (S-RIP) report (<http://s-rip.ees.hokudai.ac.jp/>).

a. Time distribution

The MMM frequency of SSWs is 6.7 events per decade, although there is considerable variability among definitions. Trate and Urate show frequencies larger than 9 events per decade because they detect a large number of events that are catalogued as minor warmings (Table S1). On the contrary, MOM and EOFz show the lowest frequencies (~5 events per decade). This is

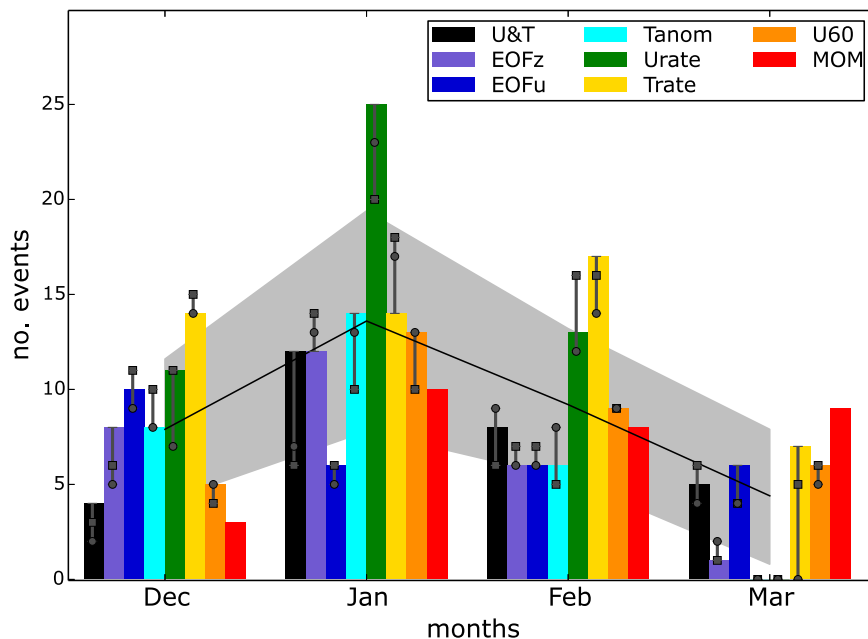


FIG. 2. Total monthly mean frequency distribution of SSWs for the 1960–2009 period. Color bars represent the number of events detected by each method (see legend). The black line represents the multimethod mean monthly frequency and the gray shadow denotes the $\pm 2\sigma_M$ interval across methods. Gray squares (circles) indicate the corresponding values for the NCEP–NCAR (JRA-55) reanalysis.

related to the highly demanding threshold imposed on the NAM index in EOFz, and to the MOM tendency to capture many events in March, some of which were catalogued as final warmings (Table S1) and excluded from our analysis, as explained in section 2. To test whether the SSW frequencies are significantly different across methods, we have performed a pairwise comparison of the mean decadal frequencies. A t test revealed that 11 out of a total of 28 possible combinations were significantly different at the 95% confidence level. A binomial test was applied to assess collectively the significance of these differences, indicating that the probability of obtaining this result by chance is lower than 1%. Thus, the SSW frequency depends on the chosen definition (at 99% confidence level).

Figure 2 shows the monthly frequency distribution of SSWs for the eight different methods. The black line is the MMM, and the gray shading denotes the corresponding $2\sigma_M$ interval. We restricted the analysis of Fig. 2 to the December–March period, which is covered by all methods except Urate and Tanom, whose analysis ends in February. Some methods also include October (EOFu and Trate), April (EOFu, U&T, and EOFz) or even May (Trate). However, no SSWs were detected later than March or earlier than November, and for the period of study only EOFu found a considerable number of SSWs in November, which partially results from our

redefinition of the onset dates (see section 2). In the remainder of the paper, all SSWs will be included, regardless of the winter period defined by each method.

Figure 2 shows similar distributions of SSWs, with the largest frequency in January in most definitions, except for Trate and EOFu. To evaluate the degree of dependence of the monthly distributions of events on the specific method, an analysis of variance (ANOVA; Wilks 2011) has been performed. This test is based on the comparison of the variance within two groups (e.g., the methods and the seasonal distribution) with the total variance. The ratio of these variances is given by the F factor, whose distribution follows a Fisher's F . Then, assuming the null hypothesis of similar population means within groups, the F factor is evaluated under an F test, thus determining whether the seasonal distribution depends or not on the method used. According to the ANOVA test, there is a significant seasonal variability in the occurrence of SSWs, which is statistically indistinguishable across methods at the 95% confidence level. This means that the seasonal cycle of SSWs is independent of the chosen method.

The decadal distribution of SSWs from 1960 to 2009 is shown in Fig. 3. In this case, the ANOVA test reveals that there is a significant amount of decadal variability associated with the occurrence of SSWs, but its decadal distribution does depend on the definition employed (at

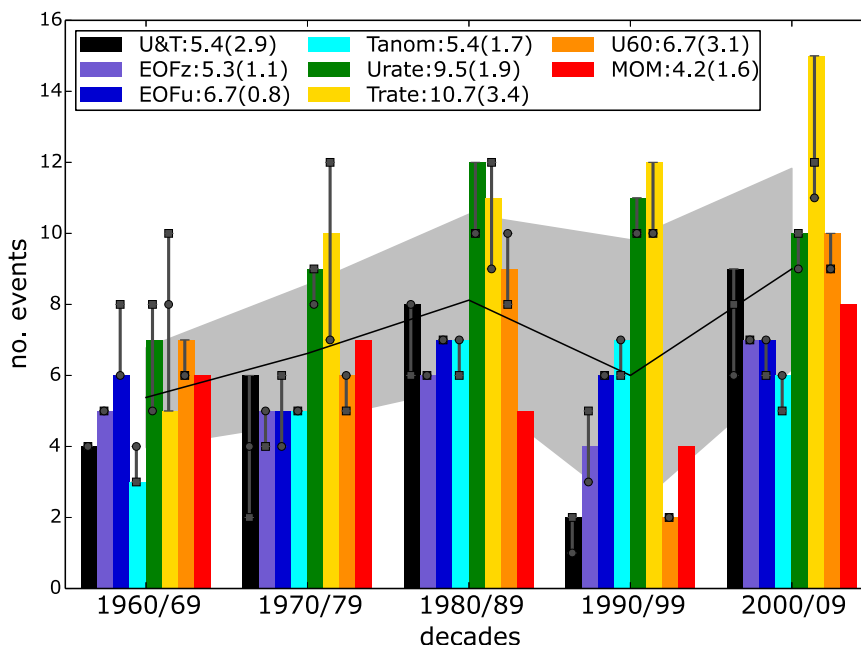


FIG. 3. As in Fig. 2, but for the total decadal distribution of events from the 1960s to the 2000s (only complete decades are shown). Numbers next to the legend show the average decadal frequency of SSWs for each method with its interdecadal standard deviation in parentheses. Gray squares (circles) indicate the corresponding values for the NCEP–NCAR (JRA-55) reanalysis.

the 95% confidence level). EOFz and EOFu have the lowest decadal variability in their distributions, while U&T and U60 show the largest variances, with a pronounced minimum in the 1990s. The latter is in agreement with relatively cold stratospheric conditions (Naujokat and Pawson 1996) and fewer occurrences of wind reversals at 60°N during the 1990s (Butler et al. 2015). Interestingly, methods based on other diagnostic variables (e.g., Tanom) or methodological approaches (e.g., Urate and Trate) do not display anomalously low frequencies of SSWs in the 1990s. Thus, the widely reported drop in the occurrence of SSWs during the 1990s is not significantly different from the behavior in other decades when definitions other than U&T or U60 (i.e., major warmings) are used, and hence it must be considered method dependent. As methods including minor SSWs do not show lower frequencies in the 1990s, this result also implies a near-normal occurrence of minor warmings in this decade. Similar results are obtained for the seasonal and decadal distribution of SSWs in NCEP–NCAR and JRA-55 reanalyses (Figs. 2 and 3).

b. Characteristics of SSWs

1) LIFE CYCLE

To assess the performance of the methods in capturing the main signatures of SSWs in the polar stratosphere

and their temporal evolution, we have computed composites of different diagnostic variables for each day of the $[-40, 40]$ -day period around the SSW onset (Fig. 4). Figure 4a shows the daily evolution of the 10-hPa ZMW at 60°N for each definition. While U&T, EOFz, and U60 cross the 0 m s^{-1} threshold near the onset, Tanom, Trate, Urate, EOFu, and MOM do not reach the wind inversion, although the latter two remain close to it. However, when the 10-hPa ZMW is analyzed at higher latitudes (e.g., 65°N), EOFu and MOM do cross the zero wind line, indicating certain latitudinal dependence of the ZMW reversal (not shown).

Several methods display the minimum ZMW some days later than the others (Fig. 4a). This time lag among definitions is also clearly seen in the composites of wind tendency (Fig. 4b) and the intensity of the warming (Fig. 4c), particularly for those methods based on short-term tendencies (Urate, Trate). The maximum wind deceleration (Fig. 4b) occurs some days before the SSW onset except in Urate and Trate, for which it peaks at the time of the SSW, as expected from their tendency-based approach to establish the onset dates. In addition to the different diagnostics used in the detection, the time lags are also influenced by the specific criterion adopted to set the onset day. While some definitions consider the onset date as the crossing-threshold day, others use the day when the polar vortex is more perturbed. All this

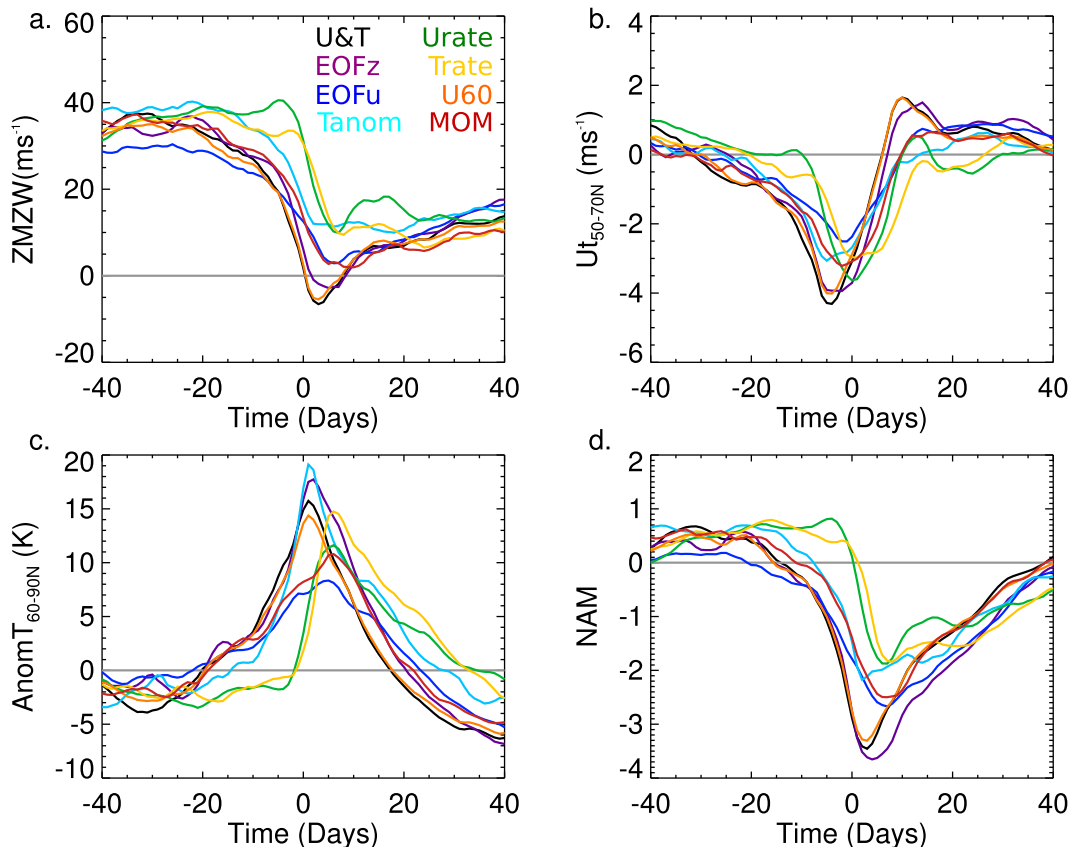


FIG. 4. Composites of different diagnostics at 10 hPa for the SSWs' life cycle: (a) zonal mean zonal wind (m s^{-1}), (b) time tendency of the zonal mean zonal wind area-weighted over 50° – 70°N (m s^{-2}), (c) zonal mean temperature anomalies area-weighted over 60° – 90°N (K), and (d) the NAM index (in standard deviation units). Each color line denotes a different method (see legend in the top left panel). The abscissa denotes time (in days) since the SSW onset.

explains why common events can be detected at different times of their life cycle in different methods (see Table S1). Note also that, for all definitions, the minimum in wind tendency occurs before the minimum ZMW (Fig. 4a) and the largest warming (Fig. 4c). This is in agreement with theoretical expectations, as the minimum in the wind tendency is related to the strong wave dissipation in the polar stratosphere preceding the breakdown of the polar vortex (e.g., Andrews et al. 1987; Kadera 2006). Similarly, for all methods, the amplitude of the maximum warming is in good agreement with the magnitude of the wind deceleration (cf. Figs. 4b and 4c), as expected from the thermal wind balance.

The NAM index presents a minimum around the onset date in all methods (Fig. 4d). Overall, the evolution of the NAM index is very similar to that of the ZMW. Thus, some methods place the minimum NAM value some days later than the detection of the event, and those that show the strongest easterly winds (Fig. 4a) also show the largest (negative) NAM values. In particular, the definitions that impose a wind reversal

(U&T and U60) reach NAM index values around -3 and show a similar behavior to EOFz, which identifies SSWs from a NAM index crossing threshold.

The overall comparison of all metrics shown in Fig. 4 reveals that the life cycle of the SSWs detected by EOFu displays weaker signatures than those reported by the other methods. As this is the only method based on data at 50 hPa (Table 1), we recomputed the life cycle composites by applying the EOFu definition at 10 hPa. In that case (not shown), the results displayed much better agreement with the other methods. This implies that the level chosen to detect SSWs can influence the life cycle of SSWs.

2) DYNAMICAL BENCHMARKS

Charlton and Polvani (2007) defined some benchmarks for SSWs based on time-averaged parameters around the SSW onset dates and they have been used in other studies (e.g., de la Torre et al. 2012) to validate the models' performance to reproduce SSW characteristics. However, the temporal windows for the calculations

were subjectively chosen according to the onset dates of the U60 definition and are not necessarily compatible with other methods included herein. To compare the different benchmarks across definitions, in our study, the time intervals have been modified to avoid biasing the results toward certain methods. Thus, we constructed the following benchmarks: 1) the amplitude of the SSWs in the midstratosphere, defined as the maximum 10-hPa warming rate over 50° – 90° N and for the $[-20, 20]$ -day period relative to the onset date of the SSW; 2) the maximum 10-hPa ZMW deceleration rate at 60° N for the $[-20, 20]$ -day period of the SSW; 3) the amplitude of the SSWs in the lower stratosphere, defined as in 1) but at 100 hPa; and 4) the troposphere–stratosphere coupling, as measured by the maximum anomaly of the zonal mean meridional eddy heat flux averaged over $(45\text{--}75)^{\circ}$ N at 100 hPa during the $[-30, 0]$ -day period (i.e., an indicator of the upward propagation of tropospheric Rossby waves preceding SSWs). To avoid assigning short-lasting (i.e., daily) values to the benchmarks, the rating changes defined in benchmarks 1–3 are calculated as centered differences of 7-day mean periods separated by 8 days. As the time lags among the onsets of common events are usually lower than 30 days, this procedure also ensures that the same value of the benchmark is taken for common events, regardless of the definition employed in the detection.

To investigate the downward propagation of the SSW signatures through the lower stratosphere, a new benchmark has also been constructed. It accounts for the relative number of SSWs (with respect to the total number of SSWs) that display a sizable NAM signal response through the middle to the lower stratosphere and will be referred to as the ratio of propagating SSWs. These events have been identified by tracking negative NAM values in a time–height cross section from 10 hPa to the lower stratosphere. Our criterion of propagation is that the NAM value stays equal to or lower than -0.5 standard deviations as we descend in the stratosphere. We start by searching for the latest day (after the SSW onset) when the NAM value criterion is satisfied at 10 hPa. Then, for the so-detected day we move down to the following pressure level. From this point, we step forward (or backward) in time searching for the latest day with NAM values reaching that threshold. This procedure is repeated until 200 hPa; if at this level the criterion is satisfied at least 10 days after the onset, the event is considered as a propagating SSW. The -0.5 standard deviation value was chosen as a threshold because it provides an upper limit to the significant signal of the NAM composites (shown later in Fig. 6). Qualitatively, the results do not vary substantially if similar thresholds are used instead.

Note that a propagating event is not required to reach the troposphere.

Figures 5a–d show the SSW-based composites of each benchmark computed for each of the eight methods (colored squares) with their $\pm 2\text{-}\sigma$ levels, together with the MMM (black circle) and the associated $\pm 2\text{-}\sigma_M$ interval. The most outstanding result is the large dispersion of values within methods, which highlights a strong case-to-case variability for all definitions. These within-method changes are much larger than the intermethod spread, making the differences in the dynamical benchmarks among methods not statistically significant. The overall good agreement of benchmarks across methods confirms that the method discrepancies observed in Fig. 4 can be largely alleviated by accounting for the lags in the times of detection (as done in Fig. 5).

Even though the differences in the benchmarks are not significant, EOFu shows the smallest SSWs amplitudes (Fig. 5a) and wind deceleration rates (Fig. 5b) in the midstratosphere and the largest warming in the lower stratosphere (Fig. 5c). Again, this is related to the choice of 50-hPa data for the detection of SSWs (not shown). Interestingly, the signal-to-noise ratio (MMM/σ_M) for the SSW amplitude at 10 hPa is around 5 times larger than at 100 hPa (Figs. 5a,c). This indicates an increasing intermethod spread of the benchmarks toward the lower stratosphere, and suggests that discrepancies among methods in the SSWs signatures increase as we move down from the level of detection.

Figure 5e shows the relative number of propagating SSWs detected by each method. On average, nearly 70% of the SSWs are propagating events. There is agreement between the two definitions that explicitly demand ZMW reversal (i.e., U&T and U60) but although they only include major warmings, they show lower ratios of propagating SSWs than EOFz and EOFu. This agrees with Baldwin and Thompson (2009), who showed that NAM-like indices can lead to stronger stratosphere–troposphere coupling than SSWs based on ZMW reversal at 60° N. On the other hand, there is large variability in the number of propagating events among the methods with the largest percentage of minor warmings (i.e., Tanom, Trate, and Urate). This may indicate discrepancies in the propagating behavior of minor SSWs. Note, however, that benchmarks are affected by a large dispersion in all methods, which makes it difficult to establish robust conclusions based solely on the mean values of these diagnostics. Thus, in the next section, we will analyze in more detail how major and minor SSWs contribute to the discrepancies in the resulting tropospheric signal and surface impacts of SSWs.

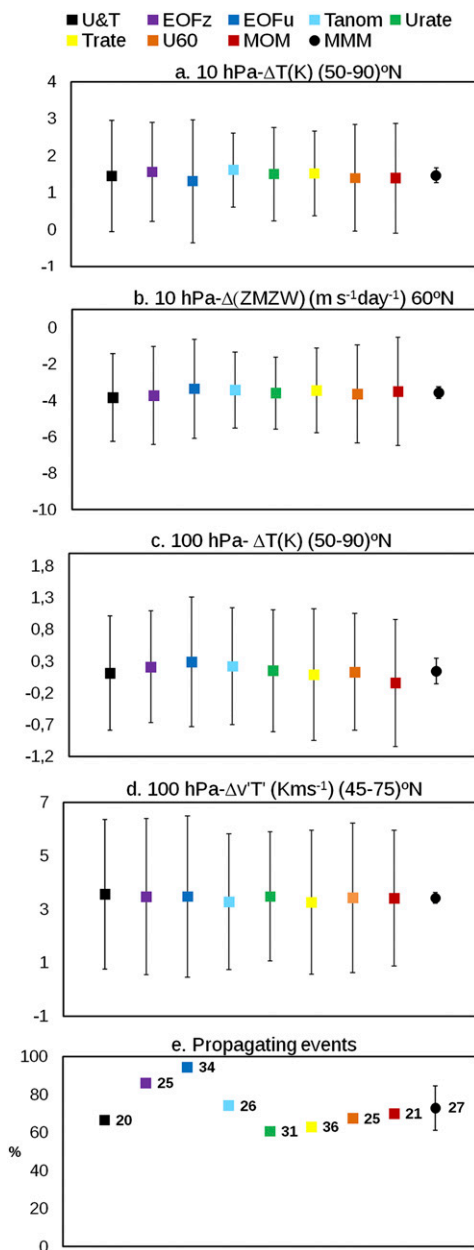


FIG. 5. SSW-based composites of dynamical benchmarks for the different methods (colored squares; see legend) and the corresponding $\pm 2\sigma$ interval (whiskers): (a) maximum 10-hPa warming rate area-weighted over 50° – 90° N (in K), (b) maximum 10-hPa zonal mean zonal wind deceleration rate at 60° N (in $\text{m s}^{-1} \text{ day}^{-1}$), and (c) maximum 100-hPa warming rate area-weighted over 50° – 90° N (in K). For (a)–(c), daily rates are computed as centered differences of 7-day mean periods and the maximum value is chosen over the $[-20, 20]$ -day interval around the onset date. Also shown are (d) maximum 100-hPa $\overline{v'T'}$ anomaly area-weighted over 45° – 75° N in the $[-30, 0]$ -day period before the onset date (in K ms^{-1}) and (e) percentage of propagating events into the lower stratosphere (relative to the total number of SSWs), with numbers showing the absolute number of propagating events (see text for details). Black circles and the associated whiskers are the multimethod mean (MMM) of each benchmark and the $\pm 2\sigma_M$ interval among methods, respectively.

c. Downward propagation signal and surface effects of SSWs

1) DOWNWARD PROPAGATION

The downward propagating signal of the SSWs can be better illustrated by computing the cross-section SSW-based composite of the NAM index for the $[-90, 90]$ -day period around the onset dates of the SSWs, as shown in Fig. 6 for each of the eight methods. All definitions show the typical “dripping paint” pattern of the NAM illustrated by Baldwin and Dunkerton (2001), with persistent negative NAM values propagating downward after the occurrence of SSWs. However, not all the methods show this stratosphere–troposphere coupling with equal intensity. The strongest tropospheric NAM response is found for EOFz and EOFu. As mentioned above, this is very likely related to the NAM-based definition of EOFz, and a similar reasoning could be sustained for EOFu, as both definitions account for the first mode of variability in the winter stratosphere. However, the level used in EOFu to detect SSWs (50 vs 10 hPa) also plays an important role in modulating the tropospheric response. In fact, when the EOFu procedure is applied at 10 hPa, the downward signal weakens (not shown). In addition, these two methods are those showing the largest ratios of propagating events into the lower stratosphere (Fig. 5e). This could indicate a relationship between the ratio of propagating events and the amplitude of the tropospheric response. However, this behavior is not observed in methods with a large fraction of minor warmings (see Fig. 6). For example, Tanom shows a high ratio of propagating events into the lower stratosphere (70%) but the tropospheric response is one of the weakest (together with Urate and Trate; cf. Figs. 6d–f). One possible explanation could be that the largest relative frequency of minor SSWs in these methods is weakening the NAM signal observed for the other definitions. This possible influence of minor SSWs will be analyzed later.

To evaluate the level of agreement among methods, Figs. 7a and 7b show the MMM composite of the NAM signal (computed from panels of Fig. 6) and the intermethod spread, respectively. Despite the considerable dispersion of NAM values around the onset date of SSWs, the MMM displays a robust downward propagating NAM pattern in the stratosphere across methods. On the contrary, there are substantial differences among methods in the significance and even the sign of the NAM response in the troposphere, as reflected by the reduced multimethod agreement therein (cross-hatched areas in Fig. 7a). The intermethod spread in the NAM

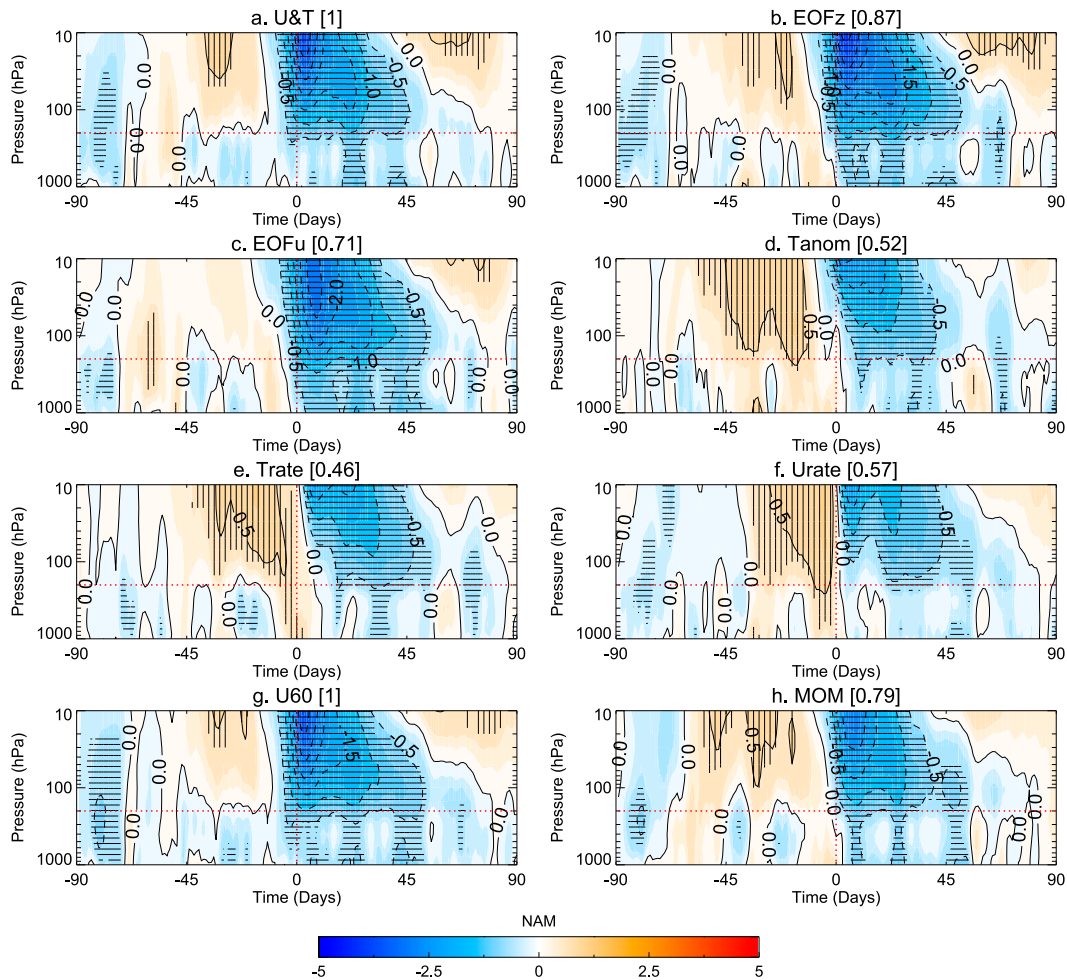


FIG. 6. Time–height composites of the northern annular mode (NAM) index (in standard deviation units) for SSWs events detected by (a) U&T, (b) EOFz, (c) EOFu, (d) Tanom, (e) Trate, (f) Urate, (g) U60, and (h) MOM. Numbers next to the titles indicate the relative frequency of major SSWs (with respect to the total number of events) detected by each method. Solid (dashed) lines denote positive (negative) NAM values. The abscissa denotes days relative to the SSW onset date. The red horizontal line highlights the 200-hPa pressure level (approximately the extratropical tropopause). Horizontal (vertical) hatched areas indicate negative (positive) anomalies that are statistically significant at the 95% confidence level according to a 1000-trial Monte Carlo test.

values around the day of detection in the stratosphere (Fig. 7b) is largely due to discrepancies in dating the SSWs. To illustrate this, we have readjusted all the SSW onsets to the date of minimum NAM index at 10 hPa. Thus, for every event, the onset is reassigned by searching the day with the minimum NAM value in a temporal window from 10 days before the earliest detection to 10 days after the latest detection among methods. This condition is applied to all SSWs, not only to common events. The MMM composite with the readjusted dates (Fig. 7c) shows stronger signals and better agreement across methods, as shown by the intermethod spread σ_M (Fig. 7d), which is now largely reduced. To test the robustness of these results to the

reanalysis product, we have repeated the MMM analysis using the NCEP–NCAR and JRA-55 data products (Fig. S1). Despite the weaker signal in the NCEP–NCAR MMM, the three reanalyses show a robust downward NAM propagation across methods. From this point, all the analyses will be performed using the readjusted onset dates. Next, we evaluate to what extent major and minor events contribute to the discrepancies among methods, as previously suggested. To do so, we have computed the MMM and intermethod spread for major and minor events separately. The downward propagation of major SSWs (Fig. 7e) shows a similar picture to the MMM of all events, but the NAM signals around the onset and the tropospheric response are

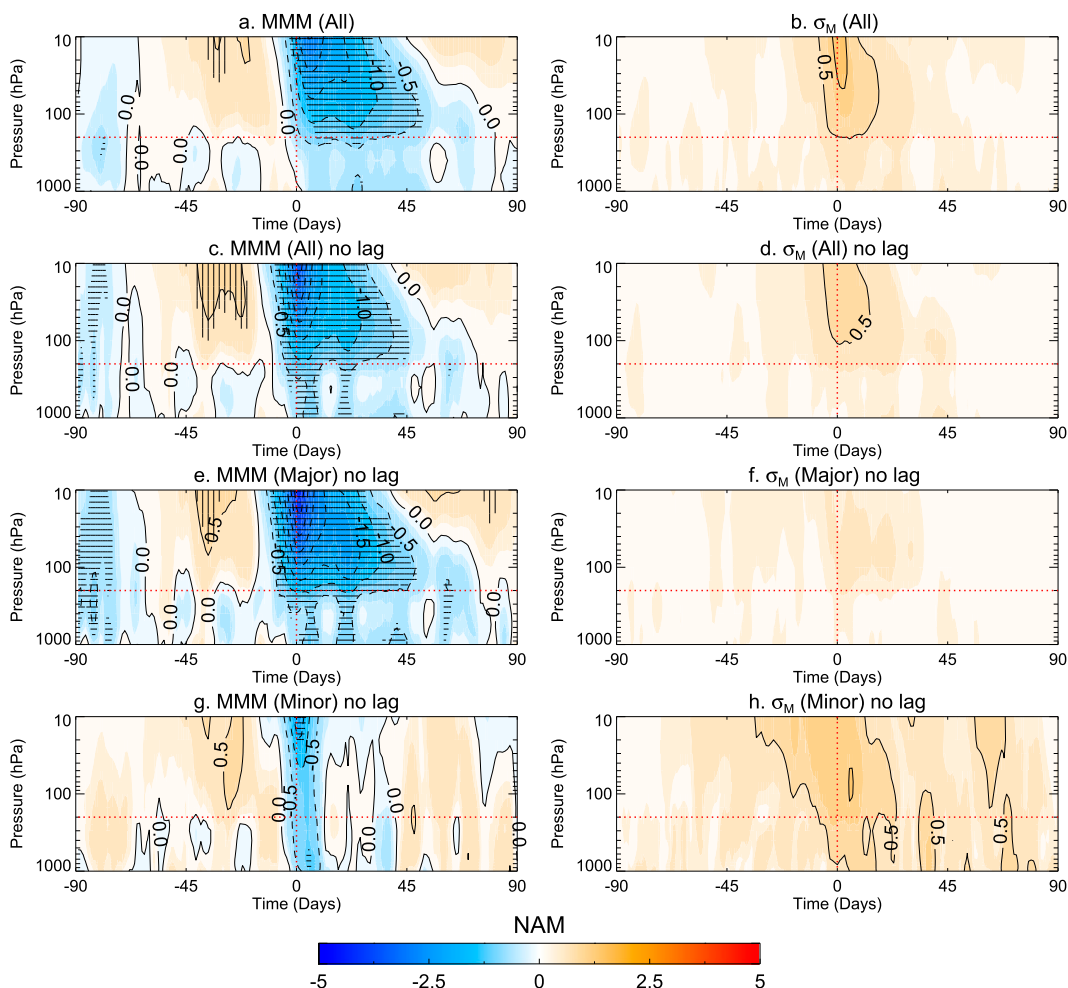


FIG. 7. (left) Multimethod means (MMMs) of the time–height NAM index composites (in standard deviation units) for (a) all SSWs, and (c),(e),(g) all, major, and minor SSWs with the readjusted onset dates (see text for details). Note that U&T and U60 are not included in the minor SSW MMM. (right) Intermethod standard deviation (σ_M) of the time–height NAM composites for (b) all SSWs, and (d),(f),(h) all, major, and minor SSWs with the readjusted onset dates. Solid (dashed) lines denote positive (negative) NAM values. The abscissa denotes days relative to the SSW onset date. The red horizontal line indicates the 200-hPa pressure level (i.e., approximately the extratropical tropopause). Hatched areas indicate regions where more of 75% of the methods agree in the sign and significance of the NAM signal.

stronger and more robust. Moreover, the intermethod spread σ_M (Fig. 7f) is noticeably reduced as compared to that of all SSWs (Fig. 7c). For minor warmings, the MMM (Fig. 7g) displays a weak and short-lasting downward propagation after the SSW onset, and large discrepancies among methods, as indicated by σ_M (Fig. 7h). In fact, the composites of minor SSWs for individual definitions show very different results (Fig. S2). In particular, EOFz and EOFu display significant propagation signals, albeit less persistent than that for major SSWs, while the others show weak negative NAM values around the onset, without clear downward propagation. Therefore, the large rates of minor SSWs

in Tanom, Urate, and Trate can explain the weaker NAM propagation signals found in their all-SSWs composites.

Finally, the very small discrepancies (σ_M) among the composites of major SSWs suggest that these events may be detected by several methods and thus may be common events. Table 2 corroborates this hypothesis. It reveals that the conditional probability of an event for being major SSW grows with the number of methods that capture it. Thus, if one event is detected by half or more of the methods (i.e., 4 out of 8), the probability of being a major SSW is $\sim 88\%$. On the contrary, minor SSWs are less prone to be common events. Therefore,

TABLE 2. Probability of an event being a major/minor SSW given the number of methods that detect it. Each cell shows the conditional probability (i.e., the number of major/minor events detected in n or more methods divided by the total number of events detected in n or more methods). Note that for each n , the conditional probability for minor warmings is complementary to that of major warmings.

	≥ 2 methods (9 events)	≥ 3 methods (15 events)	≥ 4 methods (7 events)	≥ 5 methods (3 events)	≥ 6 methods (24 events)
Major	0.66	0.69	0.88	0.96	1.0
Minor	0.34	0.31	0.12	0.04	0.0

given the inherent case-to-case variability of SSWs (see section 3b), the inclusion of minor SSWs contributes notably to the intermethod discrepancies in the zonal mean tropospheric signals of SSWs, as minor events are more likely to be exclusive of each method.

2) SURFACE IMPACT AND TROPOSPHERIC PRECURSORS

Figures 8 and 9 show the MMMs of the MSLP anomalies [5, 35] days after and $[-40, -10]$ days before the events, respectively. These time intervals were selected according to the NAM composites of Fig. 7 and the onset dates were readjusted as previously described. However, and similar to the previous section, the conclusions here remain if the original SSW onsets are used, although the signal is not so strong (not shown). The MMMs of major and minor SSWs are also shown in Figs. 8 and 9 (middle and right panels, respectively), together with their σ_M values (bottom panels). Individual composites for each method are shown in Figs. S3 and S4.

Overall, the MMM of MSLP after all SSWs (Fig. 8a) shows positive anomalies over the polar cap and negative anomalies over Europe, in agreement with Fig. 7 and previous studies (e.g., Limpasuvan et al. 2004; Charlton and Polvani 2007). However, high agreement among methods is mainly restricted to the polar cap only. This negative NAM pattern is more robust across methods when including only major warmings (Fig. 8b), and becomes weaker and not robust in the MMM of minor SSWs (Fig. 8c). Similar to the downward propagation of the SSW signals, the intermethod spread reveals better agreement across definitions in the major warming signatures (Fig. 8e), since most of them are common events, while the largest differences among methods are associated with minor SSWs (Fig. 8f). In fact, Fig. S3 indicates that there is not a unique response pattern across definitions after minor SSWs. EOFz and EOFu show similar NAM patterns after major and minor SSWs, albeit much weaker for the latter, consistent with the results shown for the downward propagation of the NAM signal. However, the other definitions display different patterns after minor warmings, which vary from method to method and show significant responses

over small regions only, thus revealing a strong method dependence on the surface impact of these events.

Finally, we compare the MSLP precursor signal of SSWs, computed for the $[-40, -10]$ -day period before the onset dates (Fig. 9). The MMM shows negative anomalies over northern North America and North Pacific and positive anomalies in Eurasia and is qualitatively similar to that obtained in previous studies for certain individual definitions (e.g., Limpasuvan et al. 2004; Cohen and Jones 2011). The MMM precursor pattern of SSWs shows higher agreement across methods than the MMM response to SSWs (Fig. 8a) and is also robust when only major warmings are considered (Fig. 9b). However, the MMM precursor signal of minor SSWs does not show a robust pattern (Fig. 9c), in agreement with the large σ_M values (Fig. 9f). In addition, the discrepancies between major and minor SSWs precursors are larger than those found for the SSWs responses (mainly in the Atlantic). Note that this result does not imply the absence of surface precursors for minor SSWs. Instead, Fig. S4 reveals significant surface signals prior to minor SSWs, but these are largely variable among methods, leading to a weak agreement in the MMM. Again, this corroborates that minor SSWs—those warmings that do not reverse the circulation—are the main source of discrepancies among the definitions.

4. Conclusions and discussion

In this study we have compared the occurrence of SSWs and their signatures among eight different definitions of SSWs, using three reanalysis datasets. Overall, the differences among reanalyses are much smaller than those across definitions. More specifically, no significant differences were found in the decadal frequencies of SSW among ERA, NCEP–NCAR, and JRA-55 reanalysis for any of the definitions, and the conclusions shown here are fairly robust to the reanalysis. Our main findings in the intermethod comparison are the following:

- 1) The mean frequency of SSWs is 6.7 events per decade, but it is method dependent, with some of

[5,35]-day period

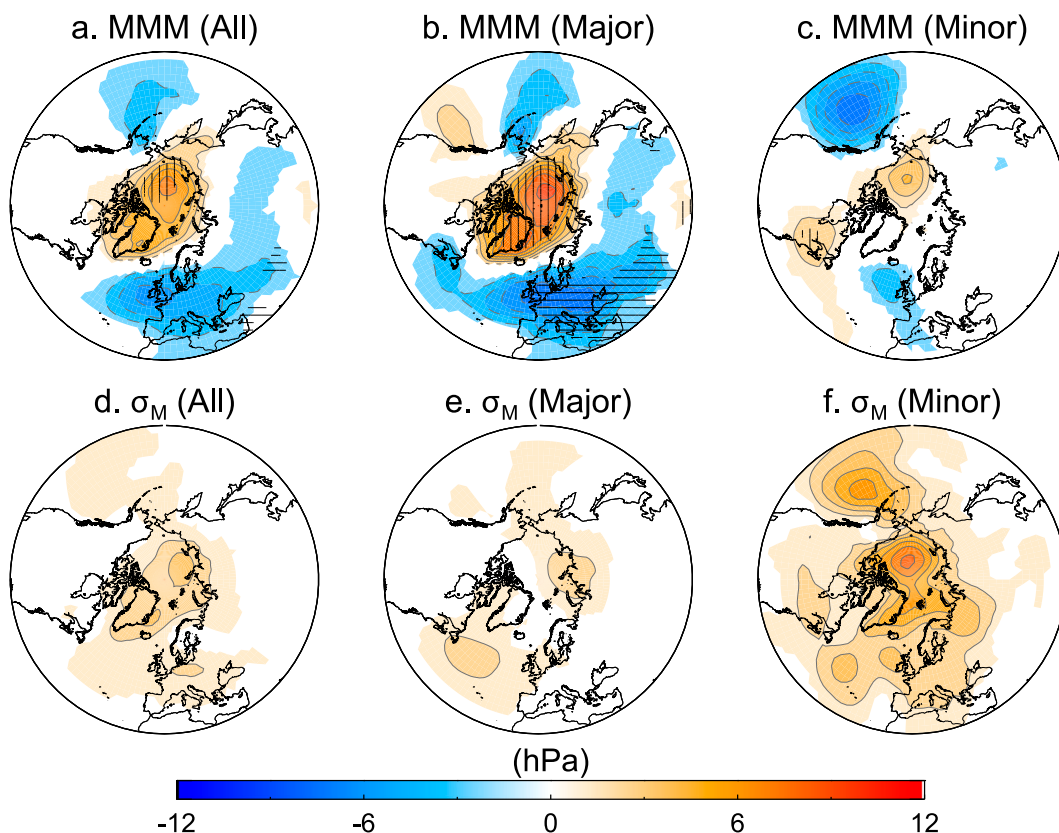


FIG. 8. (top) Multimethod means (MMM) of the mean sea level pressure (MSLP) anomalies (in hPa) for the [5, 35]-day period after (a) all, (b) major, and (c) minor SSWs. (bottom) Intermethod standard deviation (σ_M) of the MSLP composites for (d) all, (e) major, and (f) minor SSWs. Values between -0.5 and 0.5 are unshaded. Hatched areas indicate regions where more of 75% of the methods agree in the sign and significance of the MSLP signal. Results are shown for the readjusted onset dates.

the definitions that consider minor warmings reaching frequencies larger than 10 events per decade. All methods show indistinguishable intraseasonal distributions of SSWs at the 95% confidence level, with the largest occurrence in January. In contrast, the decadal variability of SSWs depends on the method. Only definitions based on wind reversal at 60°N show significant minimum frequencies in the 1990s.

- 2) The temporal evolution of different variables in the stratosphere through the SSW life cycle reveals lags among some definitions. These time lags are due to the use of different variables, approaches (instantaneous or rating changes values), and criteria adopted for dating the onset (e.g., peak values or crossing thresholds). These methodological issues involve different events and detection dates across definitions. In particular, methods based on wind and temperature rates tend to detect SSWs earlier than

the others. Nevertheless, these lags are not a major issue and can be easily corrected by readjusting the onset dates (e.g., by redefining the onset as the day of minimum NAM index in a given time interval around the detection).

- 3) The mean values of the SSW dynamical benchmarks are not statistically different across definitions due to large case-to-case variability within methods. Although the multimethod agreement decreases for lower stratospheric benchmarks, the intramethod variability is still larger than the intermethod spread, which highlights the strong differences among events for a given definition.
- 4) One of the methods included herein (i.e., EOFu) is based on data at 50 hPa, instead of the traditional 10-hPa level included in the other definitions. Using this lower level leads to discrepancies with other methods in several SSW features. This suggests that the

[-40,-10]-day period

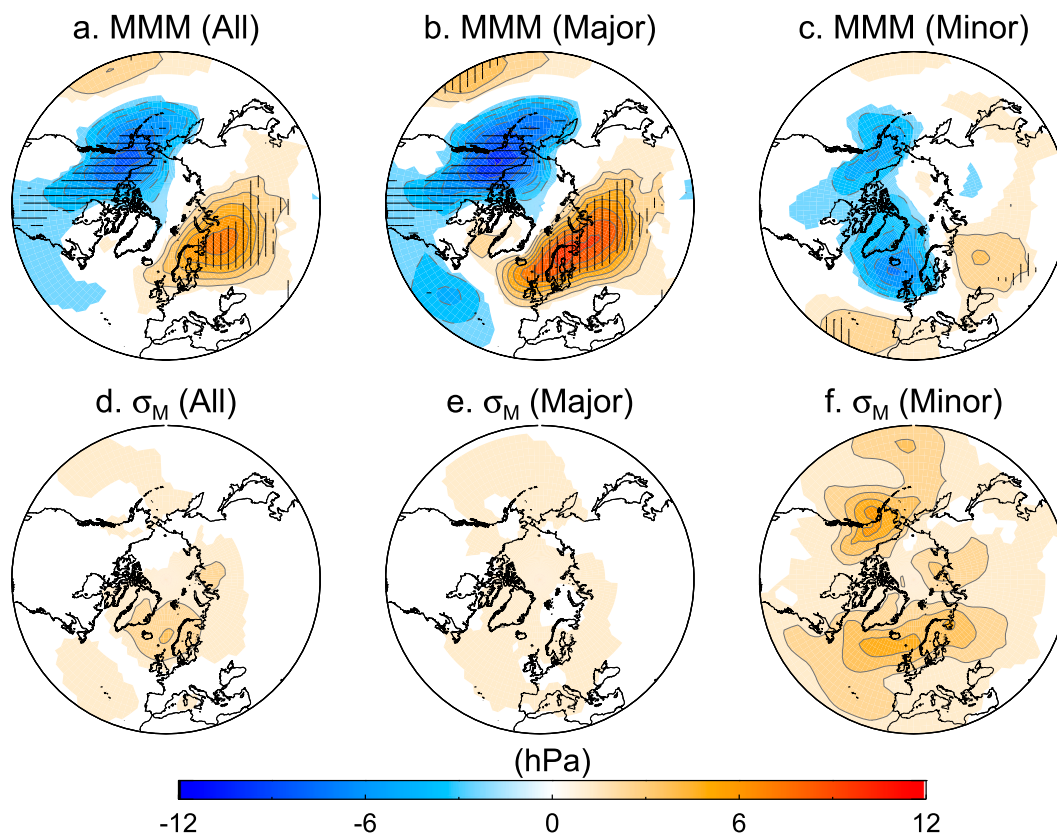


FIG. 9. As in Fig. 8, but for the [-40, -10]-day period prior to events.

chosen level for the detection plays a role in modulating the SSW signatures.

- 5) All methods show a significant downward propagation of the negative NAM signal from 10 hPa to the lower stratosphere, persisting therein for more than 45 days after the SSW onset. However, not all methods show the same level of stratosphere–troposphere coupling. The strength of the coupling, as measured by the NAM index, is affected by the relative frequency of minor SSWs (with respect to the total number of events) detected in each definition. Overall, methods with larger ratios of minor SSWs involve weaker NAM propagating signals.
- 6) Minor SSWs are also the main source of uncertainty in the precursor and response signals of SSWs at the surface. In contrast, major SSWs show significant NAM-like patterns at the surface that are robust across definitions, since they are more likely detected by most methods.

Therefore, any of the definitions analyzed here would be equally suitable for further research on the

seasonal cycle, dynamical benchmarks, and life cycle of SSWs. However, the decadal variability of SSWs is sensitive to the chosen definition, which calls for caution in studies of low-frequency variability and trends of SSWs. There are also substantial differences among methods in the tropospheric signal before and after SSWs, with the relative frequency of minor SSWs being an important source of discrepancy. This indicates that only major warmings in which wind reverses its sign should be considered to obtain robust results. This is particularly relevant when SSW occurrence is used to improve winter weather predictability or to explore tropospheric precursors of SSWs.

Since a discussion on a new SSW definition is undergoing (Butler et al. 2014), the results presented here lead us to suggest the following recommendations, which may contribute to the decision making:

- Revision of the vertical level of detection. We have found that the pressure level used to detect SSWs plays a role in modulating the downward propagation

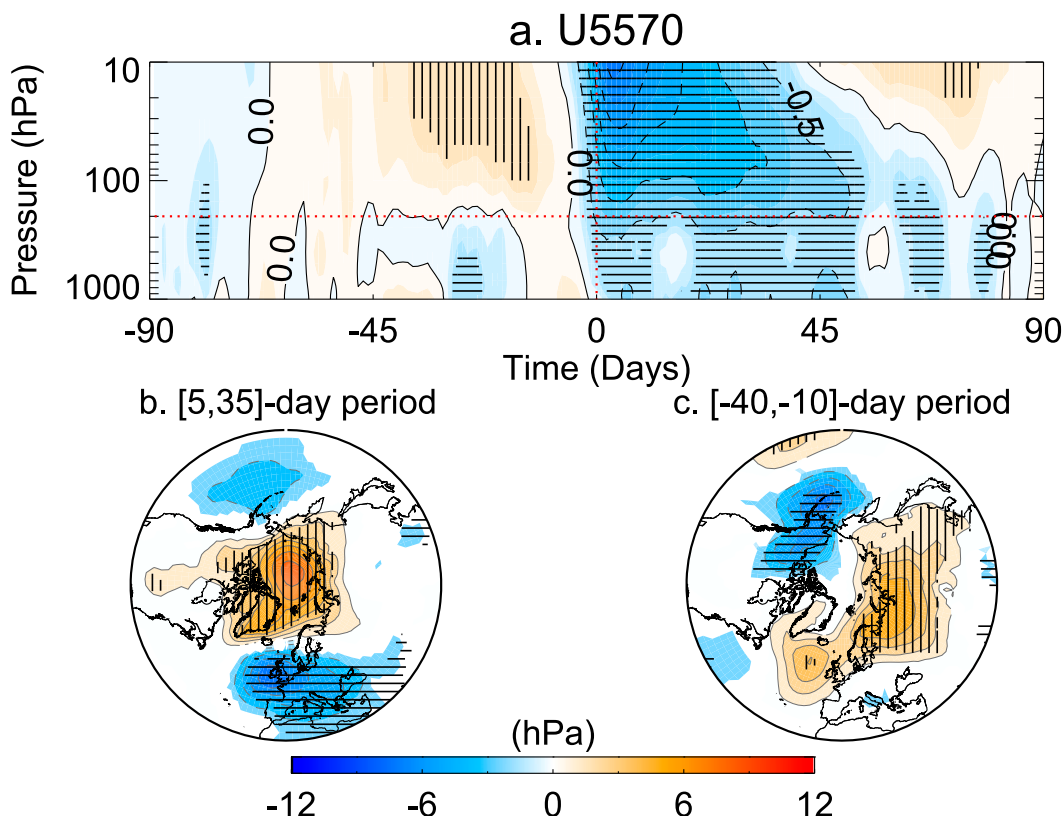


FIG. 10. (a) As in Fig. 6, but for the U5570 method. Also shown are MSLP anomalies composites for (b) the [5–35]-day period after and (c) the [–40, –10]-day period before SSW occurrence. Values of the MSLP anomalies between –0.5 and 0.5 are unshaded. Horizontal (vertical) hatched areas indicate negative (positive) anomalies that are statistically significant at the 95% confidence level according to a 1000-trial Monte Carlo test.

signal, with 50 hPa leading to stronger responses in the troposphere than the traditional 10-hPa level. While this may argue for choosing the lower level, our view is that the detection of the SSW should be independent of its impacts. On the other hand, previous studies have shown that 10 hPa may not be the most suitable level to define SSWs because of potential artifacts at this specific level associated with the incorporation of satellite data in reanalyses extending back beyond 1979 (Gómez-Escobar et al. 2012, and references therein) and hence it should be revised.

- Revision of the latitude of detection. Previous analyses (e.g., Butler et al. 2015) have shown that the SSW detection performed by the wind-reversal methods depends on the latitude chosen, and several alternatives (65°N or a latitudinal average) have been suggested. We rather propose evaluating the ZMWZ reversal within a latitudinal range. This has the advantage of assuring that the detection includes the polar vortex edge, even in climate change scenarios and models with vortex biases.

To test this methodology we have identified events for which the 10-hPa ZMWZ reversal occurs in at least one of the latitudes between 55° and 70°N (U5570 hereafter). The results obtained with this definition are consistent with the MMM values found in this paper and do not show outliers (see Fig. 10 and Fig. S5). Similar to most of the methods explored here, the minimum occurrence of major SSWs in the 1990s (as found in U&T and U60) diminishes, and although the frequency of occurrence in U5570 is comparable to methods including minor warmings (e.g., Trate), the new captured events show a major warming-like behavior in the downward propagation signal, which is similar to that shown by the definitions with the strongest stratosphere–troposphere coupling (cf. Figs. 6 and 10a). Additionally, the surface responses and precursors captured by U5570 show significant and coherent patterns, similar to those depicted by the major MMM composites (Figs. 10b,c).

- Minimizing minor SSWs detection. As shown in this study, the specific variables and criteria adopted in the

new definition might not be as relevant as long as it keeps the detection of minor SSWs to a minimum. Thus, efforts to define SSWs should aim to minimize minor warming events.

Acknowledgments. This study was supported by the Spanish Ministry of Science and Innovation (MCINN) through the MATRES (CGL2012-34221) project and the EU FP7 program through the StratoClim project (603557). We thank W. Seviour and D. Mitchell for providing the onset dates of the SSWs detected with their definition in the ERA reanalyses. We thank A. Butler and two anonymous reviewers for their useful comments and recommendations.

REFERENCES

- Andrews, D. G., J. R. Holton, and C. B. Leovy, 1987: *Middle Atmosphere Dynamics*. International Geophysics Series, Vol. 40, Academic Press, 489 pp.
- Baldwin, M. P., and T. J. Dunkerton, 2001: Stratospheric harbingers of anomalous weather regimes. *Science*, **294**, 581–584, doi:[10.1126/science.1063315](https://doi.org/10.1126/science.1063315).
- , and D. W. J. Thompson, 2009: A critical comparison of stratosphere–troposphere coupling indices. *Quart. J. Roy. Meteor. Soc.*, **135**, 1661–1672, doi:[10.1002/qj.479](https://doi.org/10.1002/qj.479).
- Bancalá, S., K. Krüger, and M. Giorgetta, 2012: The preconditioning of major sudden stratospheric warmings. *J. Geophys. Res.*, **117**, D04101, doi:[10.1029/2011JD016769](https://doi.org/10.1029/2011JD016769).
- Barriopedro, D., and N. Calvo, 2014: On the relationship between ENSO, stratospheric sudden warmings, and blocking. *J. Climate*, **27**, 4704–4720, doi:[10.1175/JCLI-D-13-00770.1](https://doi.org/10.1175/JCLI-D-13-00770.1).
- Blume, C., K. Matthes, and I. Horenko, 2012: Supervised learning approaches to classify sudden stratospheric warming events. *J. Atmos.*, **69**, 1824–1840, doi:[10.1175/JAS-D-11-0194.1](https://doi.org/10.1175/JAS-D-11-0194.1).
- Butler, A. H., E. P. Gerber, D. Mitchell, and W. Seviour, 2014: New efforts in developing a standard definition for sudden stratospheric warmings. *SPARC Newsletter*, No. 43, SPARC Office, Toronto, ON, Canada, 23–24.
- , D. J. Seidel, S. C. Hardiman, N. Butchart, T. Birner, and A. Match, 2015: Defining sudden stratospheric warmings. *Bull. Amer. Meteor. Soc.*, doi:[10.1175/BAMS-D-13-00173.1](https://doi.org/10.1175/BAMS-D-13-00173.1), in press.
- 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, doi:[10.1175/JCLI3996.1](https://doi.org/10.1175/JCLI3996.1).
- Cohen, J., and J. Jones, 2011: Tropospheric precursors and stratospheric warmings. *J. Climate*, **24**, 6562–6572, doi:[10.1175/2011JCLI4160.1](https://doi.org/10.1175/2011JCLI4160.1).
- 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, doi:[10.1002/qj.828](https://doi.org/10.1002/qj.828).
- de la Torre, L., R. R. Garcia, D. Barriopedro, and A. Chandran, 2012: Climatology and characteristics of stratospheric sudden warmings in the Whole Atmosphere Community Climate Model. *J. Geophys. Res.*, **117**, D04110, doi:[10.1029/2011JD016840](https://doi.org/10.1029/2011JD016840).
- Ebita, A., and Coauthors, 2011: The Japanese 55-year Reanalysis “JRA-55”: An interim report. *SOLA*, **7**, 149–152, doi:[10.2151/sola.2011-038](https://doi.org/10.2151/sola.2011-038).
- Gómez-Escolar, M., S. Fueglistaler, N. Calvo, and D. Barriopedro, 2012: Changes in polar stratospheric temperature climatology in relation to stratospheric sudden warming occurrence. *Geophys. Res. Lett.*, **39**, L22802, doi:[10.1029/2012GL053632](https://doi.org/10.1029/2012GL053632).
- Hannachi, A., D. Mitchell, L. Gray, and A. Charlton-Perez, 2011: On the use of geometric moments to examine the continuum of sudden stratospheric warmings. *J. Atmos. Sci.*, **68**, 657–674, doi:[10.1175/2010JAS3585.1](https://doi.org/10.1175/2010JAS3585.1).
- Kalnay, E., and Coauthors, 1996: The NCEP/NCAR 40-Year Reanalysis Project. *Bull. Amer. Meteor. Soc.*, **77**, 437–471, doi:[10.1175/1520-0477\(1996\)077<0437:TNYRP>2.0.CO;2](https://doi.org/10.1175/1520-0477(1996)077<0437:TNYRP>2.0.CO;2).
- Kodera, K., 2006: Influence of stratospheric sudden warming on the equatorial troposphere. *Geophys. Res. Lett.*, **33**, L06804, doi:[10.1029/2005GL024510](https://doi.org/10.1029/2005GL024510).
- , Y. Kuroda, and S. Pawson, 2000: Stratospheric sudden warmings and slowly propagating zonal-mean zonal wind anomalies. *J. Geophys. Res.*, **105**, 12 351–12 359, doi:[10.1029/2000JD900095](https://doi.org/10.1029/2000JD900095).
- Labitzke, K., 1981: Stratospheric-mesospheric midwinter disturbances: A summary of observed characteristics. *J. Geophys. Res.*, **86**, 9665–9678, doi:[10.1029/JC086iC10p09665](https://doi.org/10.1029/JC086iC10p09665).
- , and B. Naujokat, 2000: The lower Arctic stratosphere in winter since 1952. *SPARC Newsletter*, No. 15, SPARC Office, Toronto, ON, Canada, 11–14.
- 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, doi:[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).
- Martineau, P., and S.-W. Son, 2010: Quality of reanalysis data during stratospheric vortex weakening and intensification events. *Geophys. Res. Lett.*, **37**, L22801, doi:[10.1029/2010GL045237](https://doi.org/10.1029/2010GL045237).
- , and —, 2013: Planetary-scale wave activity as a source of varying tropospheric response to stratospheric sudden warming events: A case study. *J. Geophys. Res. Atmos.*, **118**, 10 994–11 006, doi:[10.1002/jgrd.50871](https://doi.org/10.1002/jgrd.50871).
- Matsuno, T., 1971: A dynamical model of the stratospheric sudden warming. *J. Atmos. Sci.*, **28**, 1479–1494, doi:[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., J. G. Esler, A. J. Charlton-Perez, and L. M. Polvani, 2009: A new look at stratospheric sudden warmings. Part III: Polar vortex evolution and vertical structure. *J. Climate*, **22**, 1566–1585, doi:[10.1175/2008JCLI2365.1](https://doi.org/10.1175/2008JCLI2365.1).
- McInturf, R. M., 1978: Stratospheric warmings: Synoptic, dynamic and general-circulation aspects. NASA Ref. Publ. 1017, 19 pp.
- Mitchell, D. M., A. J. Charlton-Perez, and L. J. Gray, 2011: Characterizing the variability and extremes of the stratospheric polar vortices using 2D moment analysis. *J. Atmos. Sci.*, **68**, 1194–1213, doi:[10.1175/2010JAS3555.1](https://doi.org/10.1175/2010JAS3555.1).
- Nakagawa, K. I., and K. Yamazaki, 2006: What kind of stratospheric sudden warming propagates to the troposphere? *Geophys. Res. Lett.*, **33**, L04801, doi:[10.1029/2005GL024784](https://doi.org/10.1029/2005GL024784).
- Naujokat, B., and S. Pawson, 1996: The cold stratospheric winters 1994/1995 and 1995/1996. *Geophys. Res. Lett.*, **23**, 3703–3706, doi:[10.1029/96GL03614](https://doi.org/10.1029/96GL03614).
- Scherhag, R., 1952: Die explosionsartige Stratosphärenwärmungen des Spätwinters, 1951–1952. *Ber. Dtsch. Wetterdienst*, **38**, 51–63.
- Seviour, W., D. M. Mitchell, and L. J. Gray, 2013: A practical method to identify displaced and split stratospheric polar vortex events. *Geophys. Res. Lett.*, **40**, 5268–5273, doi:[10.1002/grl.50927](https://doi.org/10.1002/grl.50927).

- Sigmond, M., J. F. Scinocca, V. V. Kharin, and T. G. Shepherd, 2013: Enhanced seasonal forecast skill following stratospheric sudden warmings. *Nat. Geosci.*, **6**, 98–102, doi:[10.1038/ngeo1698](https://doi.org/10.1038/ngeo1698).
- Taguchi, M., and D. L. Hartmann, 2005: Interference of extratropical surface climate anomalies induced by El Niño and stratospheric sudden warmings. *Geophys. Res. Lett.*, **32**, L04709, doi:[10.1029/2004GL022004](https://doi.org/10.1029/2004GL022004).
- Thompson, D. W. J., M. P. Baldwin, and J. M. Wallace, 2002: Stratospheric connection to Northern Hemisphere wintertime weather: Implications for prediction. *J. Climate*, **15**, 1421–1428, doi:[10.1175/1520-0442\(2002\)015<1421:SCTNHW>2.0.CO;2](https://doi.org/10.1175/1520-0442(2002)015<1421:SCTNHW>2.0.CO;2).
- Tripathi, O. P., and Coauthors, 2015: The predictability of the extratropical stratosphere on monthly time-scales and its impact on the skill of tropospheric forecasts. *Quart. J. Roy. Meteor. Soc.*, **141**, 987–1003, doi:[10.1002/qj.2432](https://doi.org/10.1002/qj.2432).
- Uppala, S. M., and Coauthors, 2005: The ERA-40 Re-Analysis. *Quart. J. Roy. Meteor. Soc.*, **131**, 2961–3012, doi:[10.1256/qj.04.176](https://doi.org/10.1256/qj.04.176).
- Waugh, D. W., and W. J. Randel, 1999: Climatology of Arctic and Antarctic polar vortices using elliptical diagnostics. *J. Atmos. Sci.*, **56**, 1594–1613, doi:[10.1175/1520-0469\(1999\)056<1594:COAAAP>2.0.CO;2](https://doi.org/10.1175/1520-0469(1999)056<1594:COAAAP>2.0.CO;2).
- Wilks, D. S., 2011: *Statistical Methods in the Atmospheric Sciences*. 3rd ed. International Geophysics Series, Vol. 100, Academic Press, 676 pp.
- WMO/IQSY, 1964: International Years of the Quiet Sun (IQSY) 1964–1965: Alert messages with special references to stratospheric warmings. WMO/IQSY Rep. 6, Secretariat of the World Meteorological Organization, 258 pp.
- Yoden, S., T. Yamaga, S. Pawson, and U. Langematz, 1999: A composite analysis of the stratospheric sudden warmings simulated in a perpetual January integration of the Berlin TSM GCM. *J. Meteor. Soc. Japan*, **77**, 431–445.