Chapter 6: Extratropical Stratosphere–Troposphere Coupling

Chapter lead authors

Edwin P. Gerber	Courant Institute of Mathematical Sciences, New York University	USA
Patrick Martineau	Japan Agency for Marine-Earth Science and Technology *previously at: Research Center for Advanced Science and Technology, the University of Tokyo*	Japan

Co-authors

Blanca Ayarzagüena	(1)Departamento de Física de la Tierra y Astrofísica, Universidad Complutense de Madrid (2)Instituto de Geociencias, CSIC-UCM	Spain
David Barriopedro	Instituto de Geociencias, CSIC-UCM	Spain
Thomas J. Bracegirdle	British Antarctic Survey	UK
Amy H. Butler	National Oceanic and Atmospheric Administration/Chemical Sciences Laboratory	USA
Natalia Calvo	Departamento de Física de la Tierra y Astrofísica, Universidad Complutense de Madrid	Spain
Steven C. Hardiman	Met Office Hadley Centre	UK
Peter Hitchcock	Department of Earth and Atmospheric Sciences, Cornell University	USA
Maddalen Iza	Departamento de Física de la Tierra y Astrofísica, Universidad Complutense de Madrid	Spain
Ulrike Langematz	Institut für Meteorologie, Freie Universität Berlin	Germany
Hua Lu	British Antarctic Survey	UK
Gareth Marshall	British Antarctic Survey	UK
Andrew Orr	British Antarctic Survey	UK
Froila M. Palmeiro	(1)Departamento de Física de la Tierra y Astrofísica, Universidad Complutense de Madrid (2)Barcelona Supercomputing Center	Spain
Seok-Woo Son	School of Earth and Environmental Sciences, Seoul National University	South Korea
Masakazu Taguchi	Department of Earth Science, Aichi University of Education	Japan

Abstract. This chapter assesses the representation of the two-way coupling between the troposphere and the stratospheric polar vortices in the reanalysis products. This coupling is evaluated over a broad range of time scales, from sub-seasonal to decadal, with a particular emphasis on Sudden Stratospheric Warming (SSW) events, which are among the clearest manifestations of coupling between the tropospheric and stratospheric circulations. Coupled variability on synoptic to seasonal time scales is evaluated by comparing the timing, evolution, and dynamical consistency of SSW events and Final Warming events, and the representation of the Annular Mode indices. Variability on interannual time scales is evaluated by comparing the modulation of sub-seasonal stratosphere-troposphere coupling by El Niño-Southern Oscillation (ENSO) and the Quasi-Biennial Oscillation (QBO). Finally, variability on decadal time scales is evaluated by comparing atmospheric circulation trends driven by the depletion of stratospheric ozone over Antarctica.

As the large-scale circulation cannot easily be characterized from direct observations, this chapter has largely focused on the consistency between the reanalyses, asking the question: would the characterization of stratosphere-troposphere coupling provided by a given reanalysis differ from that provided by another? The internal self consistency of reanalyses has also been evaluated, allowing for more objective grading of the reanalyses. In the satellite era, there is generally good agreement among full-input reanalyses (which assimilate all available observations, including satellite measurements) on stratosphere-troposphere coupling on synoptic to interannual time scales. In addition, conventional-input reanalyses (which exclude satellite observations, and hence full-input reanalyses before the introduction of satellites) are fairly consistent as far back as 1958 in the Northern Hemisphere. There is, however, demonstrable evidence of improvement in the more recent reanalyses. While results in prior studies based on older reanalyses will generally not be significantly different from comparable results based on the modern reanalyses, due to large sampling uncertainty, we strongly recommend that users discontinue use of older reanalyses such as NCEP R1, NCEP R2 and ERA-40 since they provide limited data (*i.e.*, lower model top) and are biased with respect to modern products.

The dominance of sampling uncertainty implies that our assessment of stratosphere-troposphere coupling is limited by the length of the reanalysis records. Consequently, the availability of high quality pre-satellite era reanalysis in the Northern Hemisphere reduces our uncertainty in the tropospheric response to SSWs by approximately 20 %.

Among the more modern reanalyses, a consistent trend in the coupled stratosphere-troposphere circulation is found, associated with ozone loss in the Southern Hemisphere. Caution should always be employed in the assessment of decadal variations and trends in stratosphere-troposphere coupling, however, due to changes in the observational network. It is also shown that uncertainties in older and conventional-input reanalyses increase with height, particularly above 10 hPa, and that satellite observations appear to be critical for an assessment of stratosphere-tropospheric coupling in the austral hemisphere. Finally, surface-input reanalyses have also been evaluated. While they should not be used in place of a full-input reanalysis, there is evidence that ERA-20C captures a substantial fraction of the variability between the troposphere and stratosphere, and so may be valuable for research into low frequency variations in stratospheric-troposphere coupling.

Contents

6.1 Introduction

In this chapter, we assess the representation of coupling between the troposphere and stratosphere across all reanalyses, with a focus on interaction through the stratospheric polar vortices. While this coupling is primarily manifested on daily to seasonal time scales, low frequency modulation by other modes of internal variability (*e.g.*, the Quasi-Biennial Oscillation) and external forcings (*e.g.*, stratospheric ozone loss) require an analysis across a wide range of time scales. The global nature of these low frequency changes also requires consideration of links between variability in the tropics and extratropics.

Our focus on the influence of the stratosphere on tropospheric weather and variability presented two challenges to this chapter. First, this report has sought to evaluate reanalyses against direct measurements, ideally measurements that are not assimilated into the reanalyses themselves. The large-scale weather and variability of the troposphere, however, is not easily characterized or verified with single measurement records. We have attempted to compare with observation-constrained measures where available, but generally, this chapter evaluates the consistency of the reanalyses, or lack thereof, as opposed to verifying them against some objective standard.

A second challenge that we face in this chapter are limitations to our understanding imposed by the natural variability of the atmosphere. A common theme is the relative importance of sampling errors, associated with the finite length of the reanalysis records, compared to the differences between the reanalyses themselves. We term the latter a "reanalysis uncertainty", to differentiate it from the sampling uncertainty. While we find evidence for an improvement in more recent reanalysis products, overall we find that our characterization of stratosphere-troposphere coupling is dominated by sampling uncertainty. As such, the choice of one reanalysis over another would not affect the scientific conclusions of a particular study, with certain exceptions, *e.g.*, the use of restricted input reanalyses, as documented below.

Sampling uncertainty can appear in subtle ways. Stratosphere-troposphere coupling is often evaluated through the analysis of events that are identified by threshold criteria, *e.g.*, a Sudden Stratospheric Warming (SSW) is identified by a reversal of the winds at 10 hPa and 60°. As a result, subtle differences between reanalyses can lead to the identification and examination of different events. This effectively aliases sampling error into a comparison of reanalyses, giving a false impression of disagreement between different reanalysis products. To address this concern, we suggest the use of a uniform set of events when evaluating different reanalysis products.

After a brief introduction to stratosphere-troposphere coupling (*Section 6.2*), we describe the reanalysis datasets in *Section 6.3*. We then present our methodology for identifying, characterizing, and evaluating SSW events in *Section 6.4*. Stratosphere-troposphere coupling on daily to seasonal time scales is further evaluated in *Sections 6.5* and *6.6*, where we evaluate the representation of the annular modes and final warming events, respectively. *Section 6.7* then examines the modulation of stratosphere-troposphere coupling on interannual time scales by El Niño-Southern Oscillation (ENSO) and the Quasi-Biennial Oscillation (QBO). *Section 6.8* compares the representation of the vertical coupling forced by ozone depletion over interdecadal time scales. Finally, *Section 6.9*, provides a summary of our results and conclusions and a compact list of key findings and recommendations.

6.2 Context and background

The troposphere and the stratosphere, the two lowermost layers of the Earth's atmosphere, contain together about 99 % of the atmospheric mass. The troposphere is the portion of the atmosphere in close contact with Earth's surface. It is the region where day-to-day weather systems evolve and impact human life; in this sense it could be viewed as a boundary layer, albeit one that occupies roughly 80 - 90 % of the atmospheric mass. The stratosphere is found from about 10 - 16 km, depending on the latitude, to about 50 km above the surface (*Andrews et al.*, 1987). What sets these two layers apart is mainly the stability of the layers: whereas temperature decreases with height in the troposphere at a rate of about 7 K per kilometer - making it nearly neutral to moist convection - stratospheric temperatures increase with height owing to the absorption of ultraviolet radiation by ozone. This stratification gave the "sphere of layers" its name.

The stratosphere's large stability sets it dynamically apart from the troposphere as it prevents the penetration of atmospheric convection from the surface, and inhibits the propagation and growth of baroclinic disturbances that make up a great fraction of tropospheric weather. Yet, depending on the season, it can be a dynamically active region subject to large variability. Large equator-to-pole temperature gradients favor the formation of strong westerly vortices in the winter stratosphere (*Waugh et al.*, 2017). These strong westerlies act as a window for the propagation of tropospheric disturbances, allowing planetary-scale waves to go through while preventing the propagation of synoptic-scale systems (*Charney and Drazin*, 1961).

When planetary-scale waves propagate vertically from the troposphere to the stratosphere, they interact with the mean flow and sometimes break (*McIntyre and Palmer*, 1983, 1984) causing an irreversible mixing of potential vorticity leading to a long-lasting weakening of the westerly winds. One of the most extreme examples of stratospheric variability, Sudden Stratospheric Warming (SSW) events, which are characterized by an abrupt deceleration and reversal of the zonal-mean zonal wind, are the result of such interactions between planetary-scale waves and the stratospheric vortex (*Matsuno*, 1971; *Limpasuvan et al.*, 2004; *Polvani and Waugh*, 2004).

The vortices in the Northern Hemisphere and Southern Hemisphere are known to behave quite differently. While the Northern Hemisphere vortex is often disturbed by SSW events in December - January - February, the Southern Hemisphere vortex is more quiescent. These differences are attributable mainly to differences in topography and land-sea temperature contrasts which are known to generate stronger planetary-scale waves in the Northern Hemisphere (*Plumb*, 1989, 2010; *Randel*, 1988). Because of the comparatively weaker wave drag in the Southern Hemisphere, zonal winds are too strong to allow vertical propagation of waves which limit wave-mean flow interactions and the variability of the vortex (*Plumb*, 1989).

As mentioned earlier, a large fraction of stratospheric variability is the result of temporal fluctuations in planetary-scale wave propagation from the troposphere to the stratosphere. It is therefore of great importance to understand how these waves are amplified or reduced in the troposphere. *Garfinkel and Hartmann* (2010) has shown that the intensification of wavenumber-1 and wavenumber-2 waves in the Northern Hemisphere are important precursors of stratospheric polar vortex weakening. One specific tropospheric circulation pattern, atmospheric blocking, has garnered particular attention due to its ability to modulate planetary-scale wave fluxes. *Nishii et al.* (2011), for instance, have shown that there are preferred regions where upward-propagating wave packets from blocking events can interfere constructively with stationary waves to produce large bursts of upward-propagating wave activity, ultimately causing SSW events. The role of such interference in modulating stratospheric variability was also discussed in *Smith and Kushner* (2012). The coupling between the stratosphere and the troposphere is not limited to an upward coupling where the evolution of the stratosphere is influenced by upward-propagating waves. The coupling is actually two-way. Events of weak stratospheric vortex anomalies, such as SSW events, were shown to affect weather at the surface by, notably, favoring the negative phase of the related North Atlantic Oscillation (NAO) and Northern Annular Mode (NAM) patterns, and shifting the storm track southward (*Baldwin and Dunkerton*, 2001). This coupling, which is often attributed to balance arguments (*e.g.*, *Black*, 2002; *Haynes et al.*, 1991) and eddy feedback mechanisms (*e.g.*, *Kushner and Polvani*, 2004; *Song and Robinson*, 2004), has implications for the predictability of tropospheric weather. Mounting evidence suggests that the state of the stratosphere influences the skill of numerical weather forecasts (*e.g.*, *Domeisen et al.*, 2019b; *Tripathi et al.*, 2015; *Sigmond et al.*, 2013; *Baldwin et al.*, 2003).

Although the extratropical stratosphere itself has no interannual memory, essentially due to the opacity to wave propagation of the summertime easterly circulation which resets the state of the vortex every year, it does vary on interannual time scales because of dynamical linkages with other modes of atmospheric variability. A clear example of such influence is the connection between ENSO and the fluxes of planetary-scale waves from the troposphere to the stratosphere which can modulate the frequency of SSW events (*Domeisen et al.*, 2019a; *Weinberger et al.*, 2019; *Song and Son*, 2018; *Calvo et al.*, 2017; *Cagnazzo and Manzini*, 2009). In fact, the stratosphere can play a significant role in setting the extratropical response to ENSO events (*Polvani et al.*, 2017; *Butler et al.*, 2015; *Iza and Calvo*, 2015). Unlike the extratropics, the equatorial stratosphere does possess an intrinsic interannual memory which manifests itself as the Quasi-Biennial Oscillation (QBO). The QBO is characterized by an oscillation between westerly and easterly winds which occurs approximately every 28 months (*Baldwin et al.*, 2001). The QBO can induce interannual variability in the extratropical stratosphere through the modulation of upward fluxes of planetary-scale wave activity in the extratropics (*Holton and Tan*, 1980) and influence atmospheric circulation at the surface (*Gray et al.*, 2018).

On interdecadal to longer time scales, the stratospheric state is influenced by modes of sea surface temperature variability such as the Atlantic Multidecadal Oscillation (*Omrani et al.*, 2014) and the Pacific Decadal Oscillation (*Woo et al.*, 2015) and anthropogenic forcing. Perhaps the clearest example of human influence to date is the destruction of ozone which cools the polar stratosphere. This perturbation of the stratosphere has in turn affected the tropospheric circulation by inducing a poleward shift of the storm track and mid-latitude westerly jet through changes in wave forcing and wave mean-flow interactions (*Son et al.*, 2018; *Orr et al.*, 2012). Finally, increasing greenhouse gas concentrations continue to cool the stratosphere (*e.g.*, *Steiner et al.;* 2020, *Ramaswamy et al.*, 2001), and may ultimately have the largest impact as the ozone hole recovers over the next decades.

A substantial fraction of the progress made in understanding these features of the stratospheric circulation and its coupling to the troposphere is owed to the development of reanalysis data sets which have greatly facilitated the study of the dynamical phenomena that regulate the coupling. Reanalysis systems integrate both forecasts from numerical models and observations through data assimilation to produce a best guess of the true state of the atmosphere. However, as discussed in *Chapter 2* (see also *Fujiwara et al.*, 2017), reanalysis data sets differ by the models, observations and assimilation techniques they utilize. As such, they produce different versions of the thermodynamic and kinematic properties of the atmosphere.

As a notable example of the differences in the representation of the stratosphere among reanalyses, *Charlton and Polvani* (2007), and more recently *Butler et al.* (2017), have highlighted discrepancies in the onset dates of SSW events between NCEP R1 and ERA-40 data sets. However, subsequent studies have revealed that the depiction of the evolution of SSW events is fairly similar among data sets (*Martineau et al.*, 2018b; *Butler et al.*, 2015; *Palmeiro et al.*, 2015; *Martineau and Son*, 2010).

The biases among reanalyses are limited enough as to not significantly alter our understanding of the physical processes regulating the evolution of SSWs. More generally, vortex variability was also shown to be similar among reanalyses during both strong and weak stratospheric vortex states (*Martineau et al.*, 2016). On interannual time scales, *Mitchell et al.* (2015) recently compared reanalysis datasets and found a remarkable consistency between them in the context of the variability of the circulation associated with volcanic eruptions, ENSO, QBO and the solar cycle. Despite these recent findings, there is a growing need to better quantify and understand the differences in the representation of atmospheric processes among reanalyses as the number of available data sets grows with the development of more sophisticated reanalyses incorporating advanced modeling and assimilation components.

6.3 Reanalysis datasets

The reanalyses assessed in this chapter are listed in **Table 6.1**. The reader is referred to *Fujiwara et al.* (2017) and *Chapter 2* for an exhaustive description of reanalyses. Variables analyzed include geopotential height, temperature and three-dimensional wind components, all of which are analyzed on pressure levels, as well as the mean sea level pressure.

In order to facilitate the comparison of zonal-mean quantities, a standardized data set of zonal mean dynamical and thermodynamical variables, the S-RIP: Zonal-mean dynamical variables of global atmospheric reanalyses on pressure levels (*Martineau et al.*, 2018c; *Martineau*, 2017), was prepared for this chapter and made public at **http://dx.doi.org/10.5285/b241a7f536a244749662360bd7839312**. Details about the variables archived, the grids and numerical methods are provided in *Martineau et al.* (2018c). Analyses of the zonal mean circulation in this chapter made use of this data, with the exception of *Section 6.6*. *Sections 6.4.2, 6.7.1* and *6.7.2*, involved additional analysis of the full three-dimensional circulation.

Table 6.1: *List of reanalysis data sets compared.*

Name	type	Reference
ERA-40	full-input	*Uppala et al.* (2005)
ERA-Interim	full-input	*Dee et al.* (2011)
ERA-20C	surface-input	*Poli et al.* (2016)
JRA-25	full-input	*Onogi et al.* (2007)
JRA-55	full-input	*Kobayashi et al.* (2015)
JRA-55C	conventional-input	*Kobayashi et al.* (2014)
JRA-55AMIP	SSTs only	*Kobayashi et al.* (2014)
MERRA [b]	full-input	*Rienecker et al.* (2011)
MERRA-2 [b]	full-input	*Gelaro et al.* (2017)
NCEP-R1	full-input	*Kalnay et al.* (1996)
NCEP-R2	full-input	*Kanamitsu et al.* (2002)
CFSR	full-input	*Saha et al.* (2010)
CFSv2	full-input	*Saha et al.* (2014)
20CR v2	surface-input	*Compo et al.* (2011)
20CR v2c	surface-input	*Compo et al.* (2011)

[b] For MERRA and MERRA-2, only the assimilated state (ASM) products are used (see discussion in *Chapter 2* and *Fujiwara et al.*, 2017).

6.4 Sudden stratospheric warming events

A Sudden Stratospheric Warming is a dramatic breakdown of the climatological stratospheric polar vortex in the winter hemisphere, first observed in post-war Berlin by *Scherhag* (1952). The name itself encapsulates the essential features of these events. They are sudden, or, in the original language of Scherhag, explosive: the entire vortex breaks down in a few days, being associated with a remarkable warming of the winter pole, typically on the order of 10s of degrees Celsius at 10 hPa, sometimes exceeding 50 °C or 60 °C. They are primarily a Northern Hemisphere phenomenon, and only one major SSW (in 2002) has been observed in the Southern Hemisphere. We therefore focus exclusively on SSWs in the Northern Hemisphere.

SSWs tend to come in two flavors, splits and displacements. In the former, the climatological vortex splits into two vortices of similar size at the time of the warming, while in the latter, the vortex shifts off the pole. In both cases, the vortex(ices) are ultimately sheared apart, leading to an irreversible mixing of potential vorticity and the deceleration of the polar vortex. Equivalently, splits are associated with comparatively more wavenumber-2 activity, while displacements are primarily associated with wavenumber-1. Recent work has suggested that the type of warming may have significant implications to the mechanism of the warming and its impact on the surface (*Mitchell et al.*, 2013; *Esler and Matthewman*, 2011; *Matthewman and Esler*, 2011) although this sensitivity is not observed in all studies (*White et al.*, 2019; *Maycock and Hitchcock*, 2015). While many SSWs can be characterized unambiguously (*e.g.*, 22 February 1979 is a classic split), a nontrivial number (roughly 1/3, as we will see) are not so easy to classify. There are also substantial sampling uncertainty issues, particularly when assessing the influence of anthropogenic forcing (*e.g.*, *Maycock and Hitchcock*, 2015).

A key result of this section is shown in **Table 6.2** and **Figure 6.2**, where a standardized list of SSW event dates and classifications for the period 1957 to 2011 are provided. We refer the reader to the SSW compendium (*Butler et al.*, 2017) for an up-to-date list of SSW events (**https://www.esrl.noaa.gov/csd/groups/csd8/sswcompendium/majorevents.html**). Results of SSW classification performed independently for each reanalysis are listed in the *Appendix*, **Tables A6.4 - A6.7.**

6.4.1 Identifying SSW events

A number of definitions have been proposed to characterize SSWs in reanalyses and models, all ultimately establishing a key threshold to define the onset of an event. This threshold nature of SSWs makes them sensitive to subtle differences between the reanalyses (*Butler et al.*, 2015). For example, the most commonly used criteria, as adopted by the World Meteorological Organization (*WMO; McInturff*, 1978), requires that the zonal mean zonal wind reverses at 60° and 10 hPa. If the zonal mean winds drop just below zero in one reanalysis, but to only +0.1 m s^{-1} in another, only one would count as an event (*Kim et al.*, 2017). Given the large variation between SSW events, this can alias sampling error into a comparison of events across reanalyses. In our hypothetical case above [1], a trivial difference in the reanalysis winds (0.1 m s^{-1} compared to a climatological variability on the order of 10 m s^{-1}) could mistakenly imply a large difference between two products that are actually very similar.

To account for this issue, we identify a standardized set of SSW dates for use across all reanalyses. This was obtained by first identifying events for each reanalysis individually, similarly to *Butler et al.* (2017), based on a reversal of the daily mean, zonal mean zonal wind at 60°N and 10 hPa from November to April, as listed in **Table A6.4**. The central date is defined by the day the daily mean wind first reverses, not necessarily the date on which the instantaneous zonal mean wind first reverses. Two criteria to ensure events are independent, and not the final reversal of the polar vortex to its summertime state, are also imposed. Following *Charlton and Polvani* (2007), the winds must return to a westerly direction for at least 20 consecutive days between independent events, and for at least 10 consecutive days prior to April 30.

The standard WMO definition also requires a reversal of the temperature gradient at 10 hPa. This gradient reversal is not well defined. Commonly it is interpreted that the zonal-mean temperature at the pole (here, 87.5°N is used to avoid the singular nature of the zonal mean at 90°N) must exceed the zonal-mean temperature at 60°N, but this puts a great deal of weight on the temperature near the pole. In practice, this criterion rarely matters; the stratosphere remains in geostrophic balance during an SSW, such that a reversal of the 10 hPa winds implies a reversal in the temperature gradients below 10 hPa, which are highly correlated with the 10 hPa temperatures. Only a few events would be excluded (two from NCEP R1, and just one from JRA-55, as delineated by the green boxes in **Table A6.4**). We therefore omit the temperature gradient criterion for classifying SSW events in this work.

To establish the standard set of dates listed in **Table 6.2**, events were defined when a majority of the reanalyses identify a SSW around the same time, *i.e.*, prior to 1979, 2 out of 3 reanalyses must detect the event and post 1979, at least 4 reanalyses must detect the event. The onset date was then set by taking the median across the dates given by each reanalysis. In recent decades, the dates rarely vary by more than a day or two across reanalyses, but there are a few events at the beginning of the reanalysis record, as in December 1965, where the spread was more than a week. In this case the date was set by the average of the two more modern reanalyses.

The frequency and seasonality of SSWs determined from each reanalysis separately were examined, as detailed in *Ayarzagüena et al.* (2019). In both periods, historical (1958 - 1978) and satellite (1979 - 2012), there is a good agreement in the mean frequency of SSWs between all reanalyses. This frequency is very similar in both eras, with 5.9 events per decade for the historical period and 6.5 events per decade for the satellite period.

In contrast, larger differences are found for the seasonality of SSWs. **Figure 6.1** shows the SSW decadal frequency distribution within ±10-day periods. The historical period shows the largest spread. ERA-40 and JRA-55 display an increasing SSW occurrence from early winter that maximizes in January and decreases by late winter (**Figure 6.1a**). On the contrary, the intraseasonal distribution of SSWs for NCEP R1 shows three sharp maxima in early, mid and late winter, in agreement with the evolution of the standard deviation of the polar night jet (PNJ) for this reanalysis.

In the satellite period, the results are similar across reanalyses (**Figure 6.1b**). For this time period, the maximum occurrence is shifted to late winter in all datasets, unlike the distributions of ERA-40 and JRA-55 in the historical period. Similar differences in the intra-seasonal distribution of events were already documented by *Gómez-Escolar et al.* (2012) for the pre/post 1979 periods. The distribution of events in the two periods were compared with a two-sample Kolmogorov-Smirnov test: the null hypothesis that both samples came from the same probability distribution can be rejected. This may indicate low frequency variations in the seasonality of SSWs, although we have less confidence in the pre-satellite distribution given differences between reanalyses.

6.4.2 Characterizing SSW events

Recent work has suggested that there may be fundamental differences between the two types of sudden warmings. For instance, *Matthewman et al.* (2009) have shown that while split events typically have deep equivalent-barotropic structures, vortex displacement events have clear baroclinic structures. In addition, the impact of the event on the troposphere may differ between the two types of events (*e.g.*, *Mitchell et al.*, 2013).

[1] This case is actually not hypothetical; a similar situation, for example, occurred in February 2002, when MERRA missed an event detected by MERRA-2 by only 0.07 m s^{-1}.

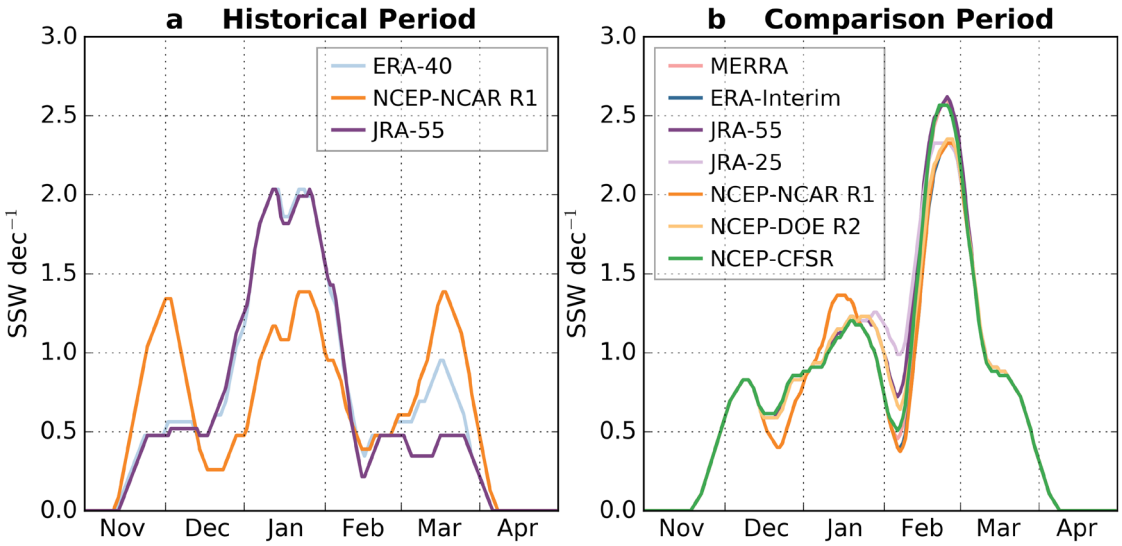

Figure 6.1: *Decadal frequency distribution of SSW events within ± 10 day-periods from the date displayed in the axis for: (a) the historical period (1958 - 1978) and (b) the satellite period (1979 - 2012). Data was smoothed with a 10-day running mean. Reproduced from Ayarzagüena et al. (2019).*

The limited sample size, however, leads to large uncertainty, such that alternative studies come to differing conclusions (*e.g.*, *Maycock and Hitchcock*, 2015). The topic is further muddied by the fact that different studies have utilized alternative definitions of SSWs, leading to a proverbial apples vs. oranges situation.

To provide greater clarity, while acknowledging that the topic is still an area of active research, we have taken the following approach. First, we consider only wind reversal events: classification schemes were applied to the 36 SSWs identified in the previous subsection. Second, we have applied four alternative classification schemes, described in more detail below, chosen to capture the range of ideas in the current literature. We provide a standardized classification of each event, listed in **Table 6.2**, based on the agreement of the classification applied separately to each reanalysis.

We compare three schemes designed to characterize whether the polar vortex is split (S) or displaced (D) during the warming event with another classification scheme that focuses primarily on the wave activity that precedes the vortex breakdown. The three schemes have been tuned to produce approximately the same rates of S and D events, all three reporting slightly more displacements than splits. The wave based diagnostic is different in that it focuses on the period leading up to the warming, as opposed to the evolution of the warming itself. It reflects the climatological dominance of wavenumber 1, classifying a clear majority of the events as wave 1-type. All the schemes are detailed below.

(1) The Seviour et al. (2013) classification scheme is based on geometric moment diagnostics of the geopotential height field at 10 hPa. The use of 10 hPa geopotential heights, which is output from all reanalyses, makes the scheme more practical than previous moment diagnostic techniques which rely on isentropic tracers, such as N_2O or potential vorticity (*Mitchell et al.*, 2011, 2013; *Waugh and Randel*, 1999; *Waugh*, 1997). The *Seviour et al.* (2013) approach was originally designed to characterize event dates as well; *e.g.*, a split event was triggered when the aspect ratio of the vortex remained higher than 2.4 for 7 days or more. However, only half of the major splits/displacements using this method are in common with those detected using the zonal-mean zonal wind reversal.

We therefore adapted the method to classify reversal events. We apply the same methodology as in *Seviour et al.* (2013), but only to days - 10 to + 10 surrounding the wind reversal. The diagnostic is based on both the aspect ratio of the vortex (the number of days the aspect ratio is above 2.4) and the displacement of the centroid (the number of days the centroid of the vortex stays below 66 degrees latitude). If the latter is greater than the former, then the event is classified as a displacement. Conversely, if the former is greater than the latter, the event is classified as a split. If the numbers are equal (or both are zero) we consider the event "unclassifiable". Note that if this adapted technique is applied to the events of *Seviour et al.* (2013), it yields identical classifications (W. Seviour, personal communication). **Table A6.5** shows results based on analysis of each individual reanalysis.

(2) The "Shibata" scheme was originally developed by Kiyotaka Shibata, and first described in *Ayarzagüena et al.* (2019). It focuses on non-zonal anomalies in the absolute vorticity at 10 hPa over a 16 day period starting 5 days before the central date of the SSW and ending 10 days later. Application of this scheme to each reanalysis is listed in **Table A6.6**.

The method is based on the algorithm suggested by *Charlton and Polvani* (2007), but with a few important modifications, as detailed in *Ayarzagüena et al.* (2019). Briefly, the algorithm identifies a local maximum in the vorticity.

Table 6.2: *Sudden Stratospheric Warming dates and classifications, according to the four schemes: D refers to a displacement, S to a split, and U, for an event that was unclassifiable, while W1 and W2 refer to events preceded by wave forcing at that number. The symbol * indicates that there was disagreement between the reanalyses; a † indicates that only a single reanalysis disagreed (after 1979 only).*

date	Seviour	Shibata	Lehtonen	Barriopedro and Calvo
30-Jan-58	D	S	S	W1
17-Jan-60	S*	D*	D*	W1
29-Jan-63	S	S	D*	W2
17-Dec-65	D	D*	D	W1
23-Feb-66	D*	D	S	W1
7-Jan-68	S	S	S	W2
28-Nov-68	D	D	D	W1
13-Mar-69	S*	D	D	W1
2-Jan-70	S	D	D*	W1
18-Jan-71	S	S	S	W2
20-Mar-71	D	D	D	W1
31-Jan-73	S	S	S	W1
9-Jan-77	S	D	S	W1
22-Feb-79	S	S	S	W2
29-Feb-80	D	D	D	W1
4-Mar-81	D	D	D	W1
4-Dec-81	U	D†	D	W1
24-Feb-84	D	D	D	W1
1-Jan-85	S	S*	S	W2
23-Jan-87	D	D	D	W1
8-Dec-87	S	S*	S	W1
14-Mar-88	S	S	S	W1
21-Feb-89	S†	S	S	W2
15-Dec-98	D*	S	D	W1
26-Feb-99	S	S	S	W1
20-Mar-00	U*	D†	D	W2
11-Feb-01	S	D†	S*	W1
31-Dec-01	S	D	D	W1
18-Jan-03	S	S†	S	W1
5-Jan-04	D	D	D	W1
21-Jan-06	D	D	D	W1
24-Feb-07	D	D	D	W1
22-Feb-08	D	D	D	W2
24-Jan-09	S	S	S	W2
9-Feb-10	U*	S	S	W1
24-Mar-10	D	D	D	W1

If two vorticity maxima are detected in diametrically opposing sectors, and the secondary maximum is at least half as strong as the first, the event is classified as a split. Otherwise it is a displacement. The main differences with the strategy of *Charlton and Polvani* (2007) consist in the definition of the sector around the strongest vorticity maximum, and the fact that the second sector must be located diametrically opposed to the first one.

(3) The Lehtonen and Karpechko (2016) classification, applied to all reanalyses in **Table A6.7**, shares features with both of the previous methods. It is based on a analysis of geopotential height at 10 hPa (as with the Seviour method), but with a goal similar to that of the Shibata approach: to separate cases where there are two independent vortices (as in a split event) from cases where there is essentially one vortex at any given time (as in a displacement).

The algorithm seeks out the two minima in the 10 hPa geopotential height, spaced apart by at least 1500 km in the horizontal and separated by a ridge of at least 375 m. If this condition is met on at least three consecutive days over the period from 5 days prior to the event onset to 10 days after, then the SSW is classified as a split. Otherwise, it is classified as a displacement. These parameters in the classification were selected to give the best agreement with the classification of major SSWs during 1958 - 2002 presented by *Charlton and Polvani* (2007).

Finally, the **(4) Barriopedro and Calvo (2014) method** classifies SSWs into wave 1 (W1) and wave 2 (W2) types by focusing on wave activity over just the period leading up to the SSW. The method is based on earlier work by *Bancalá et al.* (2012). It was applied to all reanalyses (with the exception of MERRA-2), and there was universal agreement on the classification of all 36 warmings across all the datasets. Briefly, this approach considers a Fourier decomposition of geopotential height anomalies at 50 hPa and 60 °N over an 11-day period; days -10 to 0 relative to onset. An SSW is classified as a W2 event if the amplitude over the 11-day period associated with wave 2 is equal or larger than that of wave 1, or if the wave 2 amplitude mean exceeds that of wave 1 by 200 m or more for at least one day of the period. Otherwise, the SSW is classified as a W1 event. In most cases, the former condition determines the type of SSW. The latter was included because the build up of W2 events is generally more abrupt than W1 events. The 50 hPa pressure level was chosen because wave 2 reaches its climatological maximum at this level.

As shown in **Figure 6.2**, 11 of the 36 SSWs observed between 1958 and 2011 are unanimously classified as splits by all three schemes, and 12 unanimously as displacements. The remaining 13 events differ depending on the classification scheme. These events, however, are more likely to be classified as a displacement: 8 events were displacements according to 2 of the 3 schemes, while only 5 were splits according to 2 of the 3 schemes.

We find that more than half of split events are preceded by enhanced wavenumber-2 activity (see the *Barriopedro and Calvo* (2014) method described above), as one might expect but the rest do not have prominent wavenumber-2 precursors. These may correspond to events that are preconditioned by wavenumber-1 forcing (*Bancalá et al.*, 2012; *Labitzke*, 1977) which reduces the necessity for large wavenumber-2 forcing prior to the onset in comparison to "pure" wavenumber-2 events. Perhaps more surprisingly, 2 displacement events (20 March 2000 and 22 February 2008) – one that was unambiguous across all classification schemes (the latter) – were also preceded by enhanced wavenumber-2 forcing.

Compared to the timing of event dates, there is more spread in the classification analysis between different reanalysis products. In a few instances, a tie had to be broken, in which case we gave greater weight to more modern reanalyses.

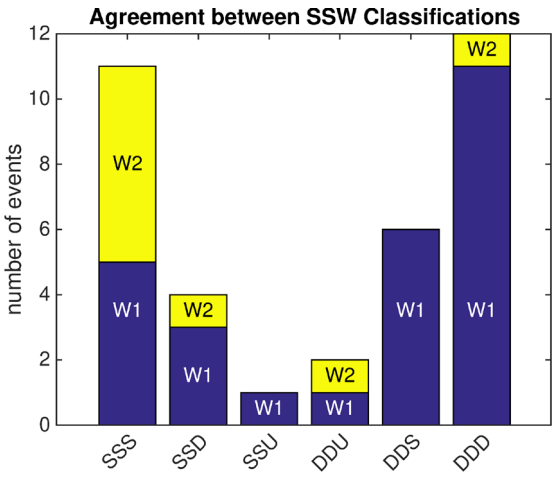

Figure 6.2: *Agreement between the SSW classification approaches. SSS and DDD refer to cases where all three schemes identified a split, displacement, respectively. SSD refers to cases where two schemes indicate the event is a split while one characterizes it as a displacement, SSU refers to a similar case, but where the third scheme was unable to classify the event, and so forth. Each bar is then divided into cases where the wave amplitude at 50 hPa over the 11 days preceding the event (see the Barriopedro and Calvo (2014) method for more detail) was primarily wave 1 (blue) or 2 (yellow).*

In addition, the Seviour scheme considers a few (3 of 36) events to be "unclassifiable", as they reflect too much of a mixture of properties of splits and displacements. In some cases, an event was unclassifiable for the individual reanalyses; in others, there was so much spread between products that we felt "unclassifiable" was the most reasonable designation. The classification schemes were applied to the wind reversal SSW events as above. There are a number of small differences in the dates and classifications based on individual reanalyses, as detailed in **Tables A6.4**, A**6.5**, A**6.6** and A**6.7**. Hence two studies based exclusively on two different reanalyses will not find the same SSW frequency, or produce the same composite fields. We find, however, that these differences are generally not significant if one accounts for sampling error. That is to say, the differences in the SSW frequency, or event composites, based on the two different reanalyses, would not be statistically significant.

As an example, consider a comparison of the dynamical evolution of W1 and W2 SSWs, classified with the *Barriopedro and Calvo* (2014) method, across different reanalyses. Key characteristics of SSWs, such as the warming of the lower and middle polar stratosphere, the deceleration of the polar vortex, and the injection of tropospheric wave activity, were compared across reanalyses by *Ayarzagüena et al.* (2019) based on the diagnostic benchmarks by *Charlton and Polvani* (2007).

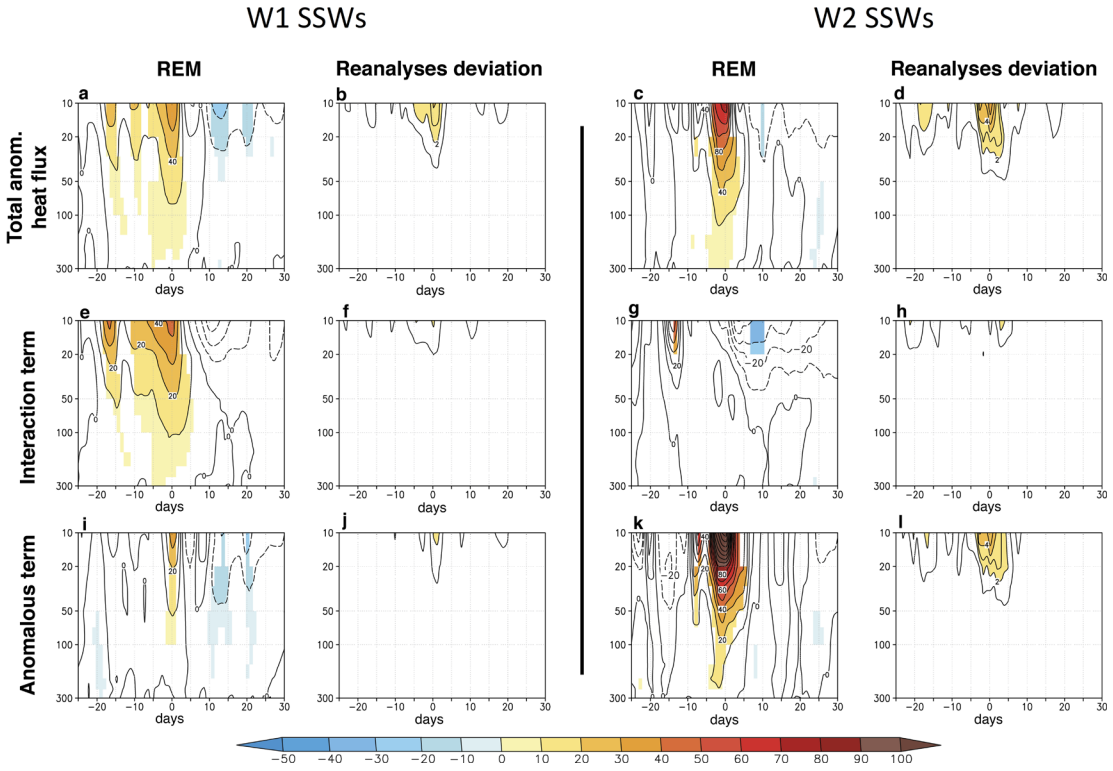

Figure 6.3: *(a) REM composited time evolution of the total anomalous eddy heat flux averaged over 45° N-75° N (K m s^{-1}) at different levels from 29 days before to 30 days after the occurrence of W1 SSWs in the comparison period. Contour interval: 20 K m s^{-1}. (b) Same as (a) but for the standard deviation of the reanalyses with respect to the REM. Contour interval: 2 K m s^{-1}. (e) and (f) Same as (a) and (b) but for the interaction between climatological and anomalous waves. Contour interval: 10 K m s^{-1}. (i) and (j) Same as (a) and (b) but for the contribution of the intrinsic wave activity associated with wave anomalies to the total anomalous heat flux. (c), (d), (g), (h), (k) and (l) Same as (a), (b), (e), (f), (i) and (j) respectively but for W2 SSWs. Shading in (a), (e), (i), (c), (g) and (k) denotes statistically significantly anomalies at a 95 % confidence level (Monte-Carlo test). Adapted from Ayarzagüena et al. (2019).*

Common events were considered to avoid possible discrepancies between reanalyses due to a different sampling. In both the pre- and post satellite periods of comparison, the agreement between datasets is very high. Only small discrepancies are found for the deceleration of the polar vortex at 10 hPa in the case of NCEP R1, particularly in the historical period. These discrepancies are probably related to the lowest model top and vertical resolution of the NCEP R1 model, since other SSW properties computed at lower levels do not present discrepancies between reanalyses.

As shown in **Figure 6.3**, anomalous meridional eddy heat flux (HF), averaged between 45 °N and 75 °N, and its different contributing terms (*Nishii et al.*, 2009) have been computed as a function of height about the onset date of SSWs. Since some previous studies have shown differences in mechanisms triggering different types of SSWs (*e.g.*, *Barriopedro and Calvo*, 2014; *Smith and Kushner*, 2012), the heat flux analysis is shown separately for W1 and W2 SSWs in the comparison period. The results of the Reanalysis Ensemble Mean (REM) resemble very much those by *Smith and Kushner* (2012) for D and S events, respectively, despite the lack of a one-to-one correspondence between W1 (W2) and D (S) SSWs.

W1 events are mainly triggered by the interaction between climatological and anomalous waves (**Figures 6.3a**, **e** and **i**) during persistent and moderately intense peaks of HF anomalies. Conversely, W2 events are related to intense but short pulses of HF arising from anomalous wave packets (**Figures 6.3c**, **g** and **k**). The comparison among reanalyses results reveals that all datasets can reproduce the different mechanisms involved in W2 and W1 SSWs. The spread is higher for W2 SSWs than for W1 SSWs particularly during the days immediately before the occurrence of SSWs (**Figures 6.3b**, **d**, **f**, **h**, **j**, and **l**). This is probably due to the smaller sample of W2 SSWs.

The tropospheric circulation associated with the occurrence of W1 and W2 SSWs in the satellite period has also been explored (**Figure 6.4**). The tropospheric patterns preceding the SSWs have been computed by analyzing the averaged geopotential height anomalies at 500 hPa in the [-10,0]-days prior to the central date of each type of SSW, while the surface signal after the occurrence of W1 and W2 SSWs has been analyzed by compositing the mean sea-level pressure (MSLP) anomalies in the [5,35]-days after these dates. The precursor signals for W1 SSWs and W2 SSWs show predominant W1-like and W2-like structures, respectively, that are similar to the precursors of the most intense events of stratospheric vortex deceleration (*Martineau and Son*, 2015).

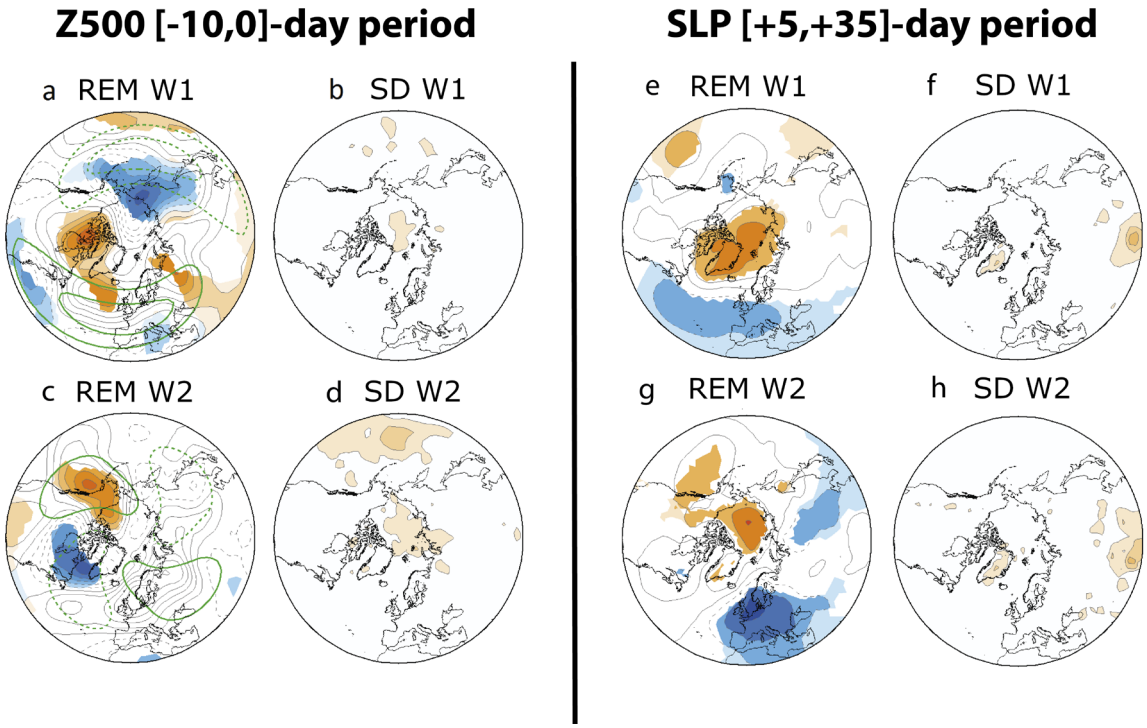

Figure 6.4: *(a) Reanalysis ensemble mean (REM) of W1 SSW-based composites of 500 hPa geopotential height anomalies (contour interval 20 m) over the [-10, 0]-day period before events for the comparison period (1979 - 2012). Only statistically significant anomalies at the 95 % confidence level of the same sign (Monte Carlo test) in at least 66.7 % of all reanalyses are shaded. (b) Standard deviation of the reanalyses with respect to the REM divided by the square root of the number of reanalyses for W1 SSWs (contour interval is 1 gpm). (c, d) Same as (a) and (b) but for WN2 SSWs, respectively. Green contours in (a) and (c) show the REM climatological W1 and W2 of 500 hPa geopotential height from November to March, respectively (contours: 40 and 80 gpm). To the right, the MSLP is composited over the [5, 35] day period after SSWs. The panels follow the same order as the Z500 precursors. Contour interval is 2 hPa for REM composites and differences and 0.1 hPa for the standard deviation of the reanalyses. Adapted from Ayarzagüena et al. (2019).*

We refer the reader to *Cohen and Jones* (2011) for earlier precursors. The SSW impact shows a negative Northern Annular Mode (NAM) pattern with positive MSLP anomalies over the polar cap in both cases, but some differences are found in lower latitudes of the Northeastern Pacific and Atlantic basins. The Pacific responses resemble the tropospheric precursor patterns therein, suggesting a possible remainder signal. In both cases (precursors and responses), the agreement among reanalyses is very good and almost no differences have been detected.

This analysis shows overall very good agreement among reanalyses in the representation of the main features, triggering mechanisms and surface fingerprint of SSWs. Despite this, some differences are found among reanalyses, particularly in the historical period and concerning the NCEP R1 reanalysis. Before 1979, SSWs in NCEP R1 show a lower mean frequency and a different seasonal distribution with respect to JRA-55 and ERA-40 (**Figure 6.1**). This disagreement also extends to climatological fields and their variability in upper levels. A plausible cause of this discrepancy is the strong artificial temperature trend affecting the early record of NCEP R1 (*Badin and Domeisen*, 2014). Arguably, the characteristics of the reanalysis models play an important role in this period, since the number of available data to be assimilated at upper levels is limited. Thus, we do not recommend the use of this reanalysis in the historical period for model evaluation initiatives.

6.4.3 Sampling uncertainty vs. reanalysis uncertainty

Studies of stratosphere-troposphere coupling are limited by the considerable dynamical variability present in both the stratosphere and the troposphere below. This variability introduces considerable sampling uncertainty into composite analyses, for example, and it is thus of interest to use all the data that is available. The amount of observational data increased considerably after 1979 when global satellite observations became broadly available. However, the basic theory underlying the occurrence of SSWs was formulated by *Matsuno* (1971), and several well-known reviews of the dynamics of these events were published well before a significant time series of satellite observations was available (*McIntyre*, 1982; *Labitzke*, 1977), indicating that the observational record largely based on radiosondes is of considerable value. This can be expected to be even more the case within the troposphere which is more easily observed with radiosondes.

Indeed, the uncertainty arising from dynamical variability that is intrinsic to the global circulation is far larger than the uncertainty arising from observational uncertainty and the process of assimilating this data into reanalysis products (*Hitchcock*, 2019). This is demonstrated in **Figure 6.5**. **Figure 6.5a** shows the time-series of zonal mean zonal wind at 10 hPa, 60 °N, around 36 major sudden stratospheric warmings from a single reanalysis, JRA-55. Events post 1979 are in solid lines, while those prior to 1979 are in dashed lines. The broad spread across events at all lags from the central date is evident, and the character of the variability in the two periods is not obviously different. This inter-event variability can be compared with the differences for individual events across reanalysis products.

Figure 6.5b shows the corresponding time series for one event (21 February 1989) during the post-1979 period, for each of the 12 reanalyses. With the exception of the two reanalyses that ingest only surface observations (ERA-20C, 20CR v2), the time-series are nearly indistinguishable relative to the inter-event variability highlighted in **Figure 6.5a**. This is even more the case if one omits NCEP-NCAR R1 and NCEP R2 whose forecast model top lies at 10 hPa. Although relatively few reanalyses extend prior to 1979 (and only one of the more modern products), this close agreement holds nearly as well for the pre-1979 period (**Figure 6.5c**).

Including the 21 years from 1958 to 1979, in addition to the 32 years from 1979 through 2010, can be expected to shrink confidence intervals by a factor of $\sqrt{32/53} = 0.78$; about a 20 % reduction. For instance, **Figure 6.6** shows the impact of including this period on the estimated frequency of SSWs. Here the *Lehtonen and Karpechko* (2016) classification method is used to define SSWs. Although not shown, similar results are found for the

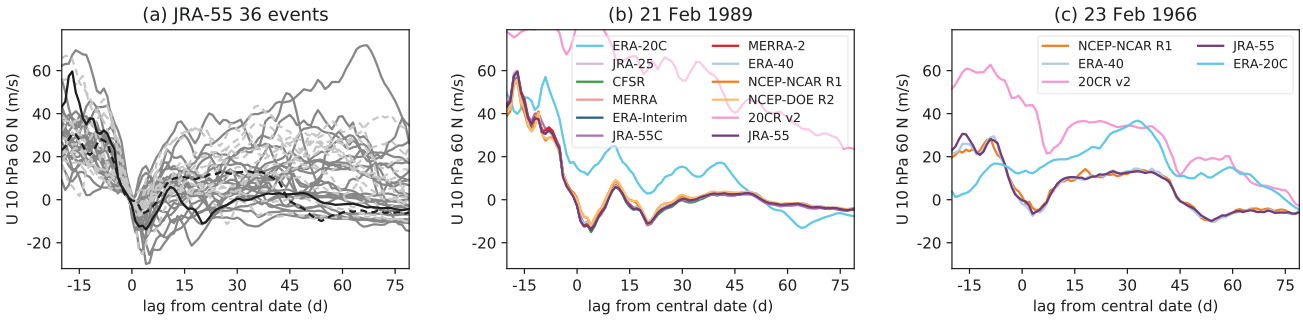

Figure 6.5: *(a) Zonal winds at 10 hPa and 60° N from JRA-55 for 36 sudden warmings. Events from the satellite period are in dark grey, those from the radiosonde period are in light grey and are dashed. (b) Winds for a single satellite-period event for all reanalyses; this event is shown by the black line in (a). (c) Zonal winds at 10 hPa and 60° N for a single radiosonde-period event for all reanalyses covering this period; this event is shown by the dashed black line in (a). Reproduced from Hitchcock (2019).*

other classification methods. The confidence intervals are generated by a bootstrapping procedure. For instance, for the post- 1979 period, sets of 32 years chosen at random (with replacement) from the period from 1958 through 2010; events that happen during these years are then used to generate an overall frequency. If a year is chosen multiple times, the events that occurred during these years are also included multiple times. This is carried out 10000 times; the 2.5th and 97.5th percentiles then define the confidence interval. A similar procedure is used for the confidence intervals on the whole 1958 - 2010 period but using sets of 53 years.

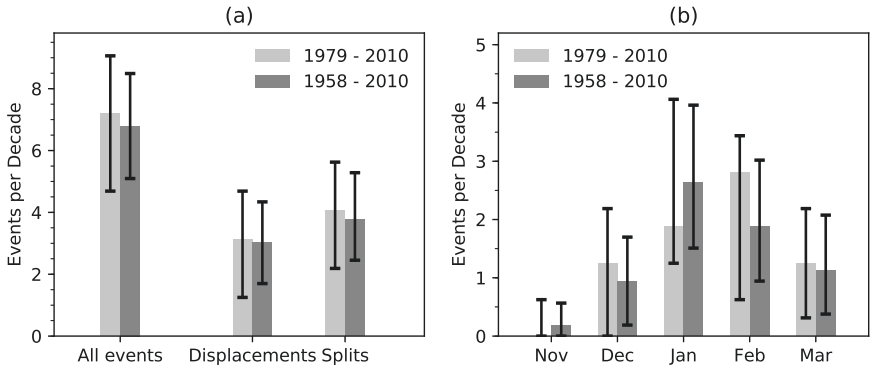

Figure 6.6: *(a) Frequency of all SSW events, and of events classified as splits or displacements for the satellite period versus the entire period where quality reanalyses are available. (b) Same as (a) but for each month of extended winter. Error bars indicate 95% confidence intervals, see text for details. Reproduced from Hitchcock (2019).*

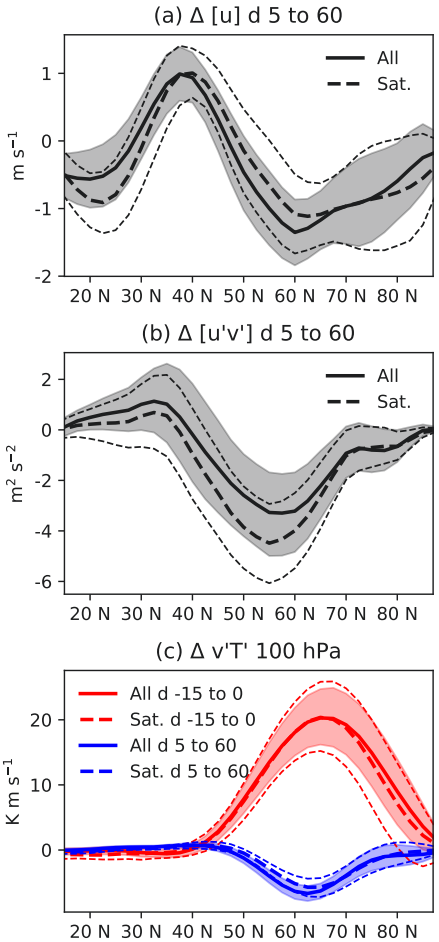

Figure 6.7: *(a) Composite mean of vertically averaged (100 to 1000 hPa) zonal wind anomalies, averaged over lags 5 to 60 days following major warmings. The solid line shows the composite for all events while the dashed line shows the composite for the satellite era alone. Confidence intervals for the whole period are shaded while those for the satellite era are indicated by thin dashed lines. (b) Similar but for vertically integrated momentum fluxes. (c) Similar but for meridional heat fluxes at 100 hPa, averaged over lags -15 to 0 (in red), and over lags 5 to 60 (in blue). Reproduced from Hitchcock (2019).*

The resulting confidence intervals are indeed reduced by a factor close to the 20 % estimate given above. The overall event frequency and the frequency of splits and displacements are somewhat reduced. The seasonal distribution of events is more substantially affected; within the broader record more events occur in January that in any other month; the period from 1979 - 2010 had relatively few January events and relatively many February events resulting in a rather different seasonal distribution as shown in **Figure 6.1** (though one well within sampling uncertainty).

Similar reductions in confidence intervals can be found for more dynamical quantities. **Figure 6.7** shows three such examples. **Figure 6.7a** shows the anomalous zonal wind, integrated from 1000 hPa to 100 hPa, from days 5 to 60 following the central date. **Figure 6.7b** shows the anomalous meridional momentum flux, also integrated from 1000 hPa to 100 hPa and averaged from days 5 to 60 following the central date. Finally, **Figure 6.7c** shows the meridional heat flux at 100 hPa, averaged from days -15 to 0, prior to the central date (in red), and averaged from days 5 to 60 following the central date (in blue). In all cases confidence intervals are generated by a similar bootstrapping procedure; however in this case the events themselves are sub-selected, rather than the years.

Again, in all cases, including the whole period results in a slightly different meridional structure. The low-latitude easterly response is somewhat weaker in **Figure 6.7a**, the momentum flux response is somewhat more positive at all latitudes (**Figure 6.7b**), and the heat fluxes in the recovery period are somewhat more reduced (**Figure 6.7c**). More importantly, the reduction in confidence intervals provides a stronger constraint for dynamical understanding and for model evaluation.

6.4.4 Assessing the internal consistency of SSW events in Reanalyses

Given that the sampling error tends to overwhelm differences in the representation of SSWs in different reanalysis products, we consider an alternative approach to evaluating their fitness: an assessment of their internal consistency. Many studies have investigated the evolution of zonal mean zonal wind using zonal-mean momentum budgets applied to reanalysis data (*e.g.*, *Martineau and Son*, 2015; *Limpasuvan et al.*, 2004). Reanalysis data sets, however, are known to present biases with respect to observations and with respect to each other. For instance, recent studies by *Lu et al.* (2015) and *Martineau et al.* (2016) have highlighted discrepancies among data sets concerning the momentum budget. Here we summarize and show key figures from the analysis of *Martineau et al.* (2018b), which quantified uncertainties in the zonal momentum budget among the reanalysis data sets.

The comparison is performed among all conventional reanalysis data sets except for ERA-40 whose deficiencies are well documented in the literature (*e.g.*, *Martineau et al.*, 2016) and which terminates in 2002, limiting the sample of SSW events. The common dates identified in **Table 6.2**, beginning with the 29-Feb-80 event and ending with the 24-Mar-10 event, are used to perform composites of the momentum budget for SSW events. The zonal-mean momentum budget can be written as follows:

$$\frac{\partial \bar{u}}{\partial t} = \underbrace{f\bar{v}}_{fv} \underbrace{-\frac{1}{a\cos^2\phi}\frac{\partial(\cos^2\phi\overline{u'v'})}{\partial\phi}}_{du'v'/dy} \underbrace{-\bar{v}\frac{1}{a\cos\phi}\frac{\partial(\bar{u}\cos\phi)}{\partial\phi}}_{Adv_\phi} \underbrace{-\bar{\omega}\frac{\partial\bar{u}}{\partial p}}_{Adv_p} \underbrace{-\frac{\partial(\overline{u'\omega'})}{\partial p}}_{du'\omega'/dp} + R \quad (6.1),$$

where f is the Coriolis parameter, u, v, ω are the zonal, meridional, and vertical components of wind, ϕ is the latitude, p is the pressure, and a is the mean radius of the Earth (6371 km). Overbars and primes denote zonal mean and anomalies with respect to the zonal mean, respectively. While the left-hand side term expresses the zonal-mean zonal wind tendency, terms of the right-hand side represent forcing terms. They are, in order, the acceleration due to the Coriolis torque, the meridional convergence of momentum fluxes, the advection of zonal momentum by the meridional wind, the vertical advection of zonal momentum by the vertical wind, and the vertical convergence of vertical momentum fluxes. The last term, R, is referred to as the residual and represents sub-grid scale processes such as gravity wave drag and numerical diffusion. It also includes imbalances in the momentum equation introduced by the data assimilation process (analysis increment), errors due to the interpolation from model levels to pressure levels, and errors related to the numerical methods employed to evaluate each term of the equation. All calculations are based on the zonal-mean data set of global atmospheric reanalyses on pressure levels (*Martineau et al.*, 2018c; *Martineau*, 2017) which provides dynamical variables on a common 2.5° by 2.5° latitude-longitude grid for all reanalysis datasets at six-hour intervals. The diagnostics presented here are markedly more sensitive to the choice of data set than horizontal resolution (*Martineau et al.*, 2018c).

Figure 6.8 shows the composite evolution of all terms of the zonal-mean momentum equation during SSW events. In addition to the terms evaluated and shown for each individual data set, the standard deviation among an ensemble of the latest reanalysis data (CFSR, ERA-Interim, JRA-55 and MERRA-2) is displayed. SSW events are characterized by an intense deceleration (up to -7 m s^{-1} day^{-1} at 3 hPa) of the zonal-mean zonal wind in the mid-stratosphere. Uncertainties in the zonal wind tendency are typically small in comparison to other terms of the momentum equation and are largest several days before the onset date (day 0). The dominant forcing terms are those that are typically included in the quasigeostrophic version of the momentum equation – *i.e.*, the acceleration due to the Coriolis torque and the convergence of meridional fluxes of momentum. These two forcings are strongly opposed, but not completely. Their sum results in a net deceleration before the onset of SSW events. Uncertainties in these forcing terms due to inter-reanalysis discrepancies typically peak several days before the onset of SSW events. Other forcing terms that are left out of the QG approximation have smaller magnitudes and show better agreement among the reanalyses. Finally, the residual is typically negative before the onset of SSW events, in part due to the exclusion of gravity wave drag from our analysis (*Martineau et al.*, 2016). It becomes more neutral after the onset, suggesting a more dynamically quiet period.

It is worth noting that preceding lag 0, JRA-25 shows a markedly larger residual in comparison to other reanalyses both in the mid and upper stratosphere. This large negative residual may be attributed to an underestimation of deceleration by the Coriolis torque in the mid stratosphere and an overly strong momentum flux convergence in the upper stratosphere in comparison to other reanalyses (not shown, see *Martineau et al.* (2016) for more details). Note that NCEP R1 and NCEP R2 are also clear outliers for these two forcings in the mid-stratosphere. Their residual is however not shown here since vertical motion is not provided in the stratosphere.

The vertical profiles of the forcing terms and their uncertainties are shown in **Figure 6.9**. Here, the inter-reanalysis standard deviation is shown separately for the ensemble of latest reanalyses and an ensemble of all reanalyses (listed in legend). Overall, all forcing terms display an exponential increase of uncertainties with height in the stratosphere. Again, the Coriolis torque and the convergence of meridional momentum fluxes dominate in terms of uncertainty. It is also noteworthy that uncertainties of the latest reanalysis ensemble are always smaller than the all reanalysis ensemble in the stratosphere which suggests an enhanced consistency in the representation of the atmospheric circulation in the modern reanalysis products.

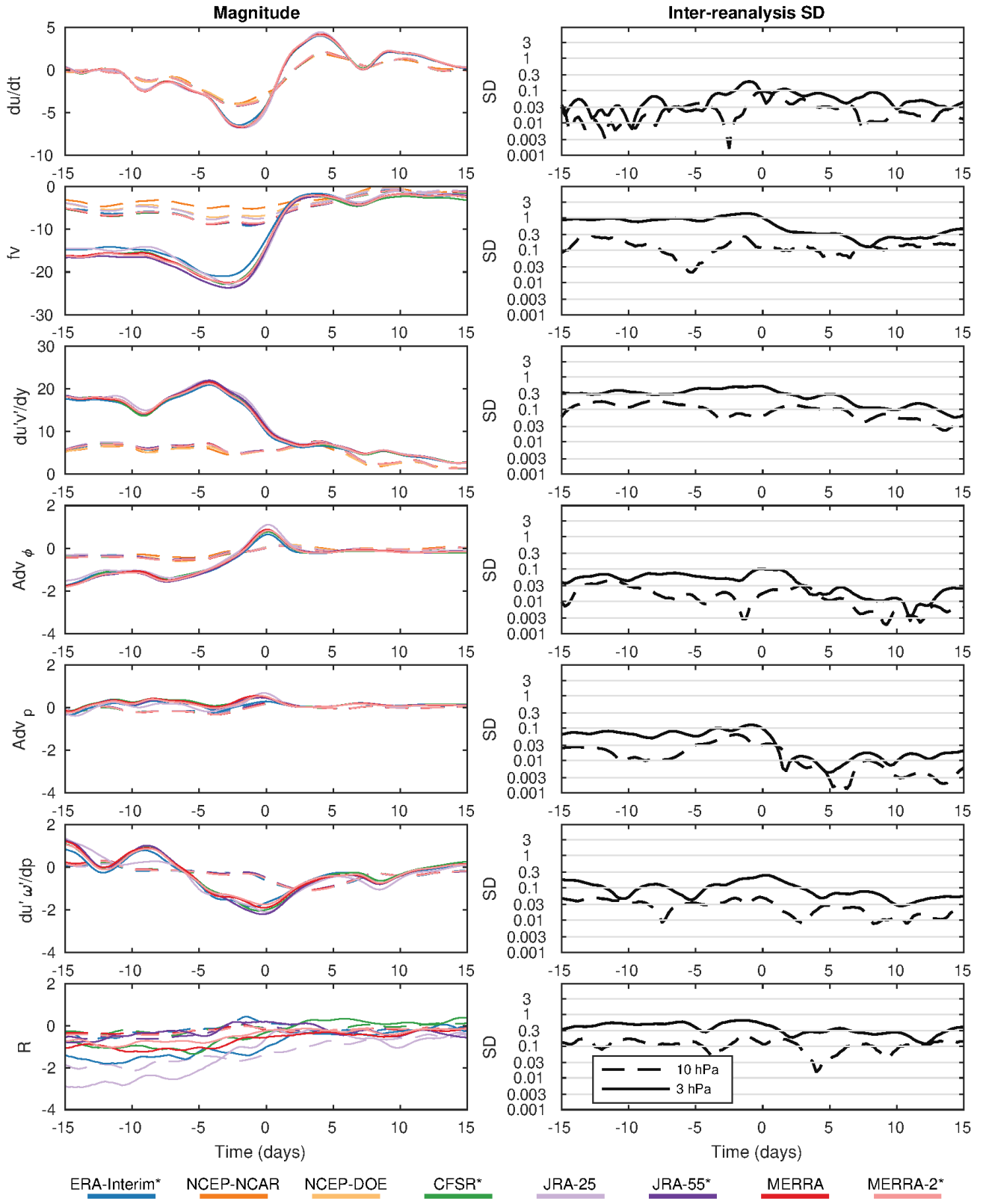

Figure 6.8: *Evolution of forcing terms of the zonal-mean momentum equation at 10 hPa (dashed lines) and 3 hPa (solid lines) in the course of SSW events. All variables are averaged from 45°N to 85°N. Note that the range of the y axis in each panel is different. (b) The inter-reanalysis spread (standard deviation) of the corresponding terms are shown for the latest reanalysis ensemble members (indicated with a * in the legend). The standard deviation is shown on a logarithmic scale: the spacing between tick marks represents a decrease or increase of the standard deviation by a factor of about 3. All quantities are expressed in m s^{-1}day^{-1}. Reproduced from Martineau et al. (2018b).*

Martineau et al. (2018a) have noted that not only the mean forcings between the Coriolis torque and momentum flux convergence are strongly opposed, but also the inter-reanalysis discrepancies in the Coriolis torque are often compensated for by inter-reanalysis discrepancies in the momentum fluxes. This results in a seemingly better self-consistency of the momentum equation (small residual) although the disagreement between data sets about the dominant momentum forcing terms can be large. This compensation could be the result of an induced meridional overturning circulation in response to biases in wave drag from planetary waves or gravity waves among the data sets. The meridional overturning circulation is an ageostrophic circulation and is thus not constrained by the thermal structure of the atmosphere like the zonal mean zonal winds which largely obey geostrophic and hydrostatic balance in the extratropics.

The aforementioned results characterized uncertainties of the momentum budget in reanalysis data sets by considering all SSW events but the study of *Martineau et al.* (2018a) provides a more thorough analysis by investigating differences between SSW events characterized by a split or displacement of the stratospheric polar vortex. The classification is done by both using vortex moment diagnostics (see *Section 6.4.2*) and by identifying the dominant fluxes of wave activity from the troposphere to the stratosphere (whether dominated by wavenumber 1 or 2) prior to the events. Overall, there is no striking difference in the uncertainties of the momentum budget between these different types of events. It is rather found that the intensity of the event, evaluated by the magnitude of the deceleration of zonal-mean zonal wind prior to the reversal, is more relevant for the agreement between reanalysis data sets. As is somewhat intuitive, the events that showed the strongest deceleration and largest forcing terms were shown to suffer from larger inter-reanalysis uncertainties.

In summary, there is generally a good agreement between the various terms of the zonal-mean momentum budget among reanalysis data sets. The discrepancies are small enough as to not introduce important uncertainties in our understanding of the dynamical evolution of SSW events. Inter-reanalysis uncertainty typically increases exponentially with height as the forcing terms also grow in magnitude. The dominant forcing terms, *i.e.*, momentum flux convergence and the Coriolis force, dominate the budget and have the largest uncertainties. The residual also increases with height, indicative of the greater role played by gravity waves in the momentum budget in the mid- to-upper stratosphere. Differences in the contribution of gravity waves to the momentum budget among reanalyses are hard to evaluate since gravity wave drag is not commonly provided for the reanalysis data sets; we therefore recommend that future data sets provide daily parameterized gravity wave drag on the standard pressure levels.

6.5 Annular modes

The annular modes have been used to quantify the coupling between the stratosphere and troposphere, particularly that associated with SSW events (*e.g.*, *Kushner*, 2010; *Baldwin and Dunkerton*, 2001; *Thompson and Wallace*, 2000). In the troposphere, the annular modes characterize meridional shifts in the extratropical jet streams; a positive index indicates the jet is located poleward of its climatological position. The jet streams are associated with the extratropical storm tracks, so that the annular modes are linked with shifts in storm activity, particularly in Northern Europe and eastern North America (*e.g.*, *Thompson and Wallace*, 1998). In the stratosphere, the annular modes chiefly characterize variations in the strength of the polar vortex. A positive index indicates a stronger than average vortex, so that the breakdown of the vortex in an SSW is associated with an abrupt shift to a very negative annular mode index in the stratosphere.

The negative shift in the stratospheric annular mode index associated with an SSW typically precedes a similar (albeit weaker) shift towards a negative annular mode index in the troposphere by a few days (*Baldwin and Dunkerton*, 2001; *Karpechko et al.*, 2017). The equatorward shift in the tropospheric jet stream persists on the order of 30 to 60 days, associated with the slow recovery time scale of the lower stratospheric vortex (*e.g.*, *Gerber et al.*, 2010) and potential feedback with baroclinic eddies in the troposphere (*e.g.*, *Song and Robinson*, 2004). SSWs are therefore important for seasonal to subseasonal forecasts (*e.g.*, *Butler et al.*, 2019; *Domeisen et al.*, 2019b; *Sigmond et al.*, 2013).

In addition, the annular modes have been used to investigate cases where the polar vortex is stronger than average (*Baldwin and Dunkerton*, 2001; *McDaniel and Black*, 2005). These "Polar Vortex Intensification" events (hereafter strong vortex events) are somewhat of an opposite analogue to a SSW, but lack a clear, abrupt onset. A stronger than average polar vortex (*i.e.*, a positive annular mode state in the stratosphere) is typically associated with a poleward shift in the tropospheric jet (*i.e.*, a positive annular mode in the troposphere).

6.5.1 Consistency of the annular mode index across reanalyses in the post and pre-satellite periods

As detailed by *Gerber and Martineau* (2018), we use a simplified procedure to compute the daily annular mode indices from the reanalyses. As proposed by *Baldwin and Thompson* (2009), the annular mode index is defined by the polar cap averaged geopotential height (all latitudes poleward of 65°), normalized to have zero mean and unit variance. To ensure that the annular mode indices characterize meridional shifts in geopotential height at all levels, the global mean geopotential height on each pressure level is first removed at each time step before computing the polar cap averages (*Gerber et al.*, 2010).

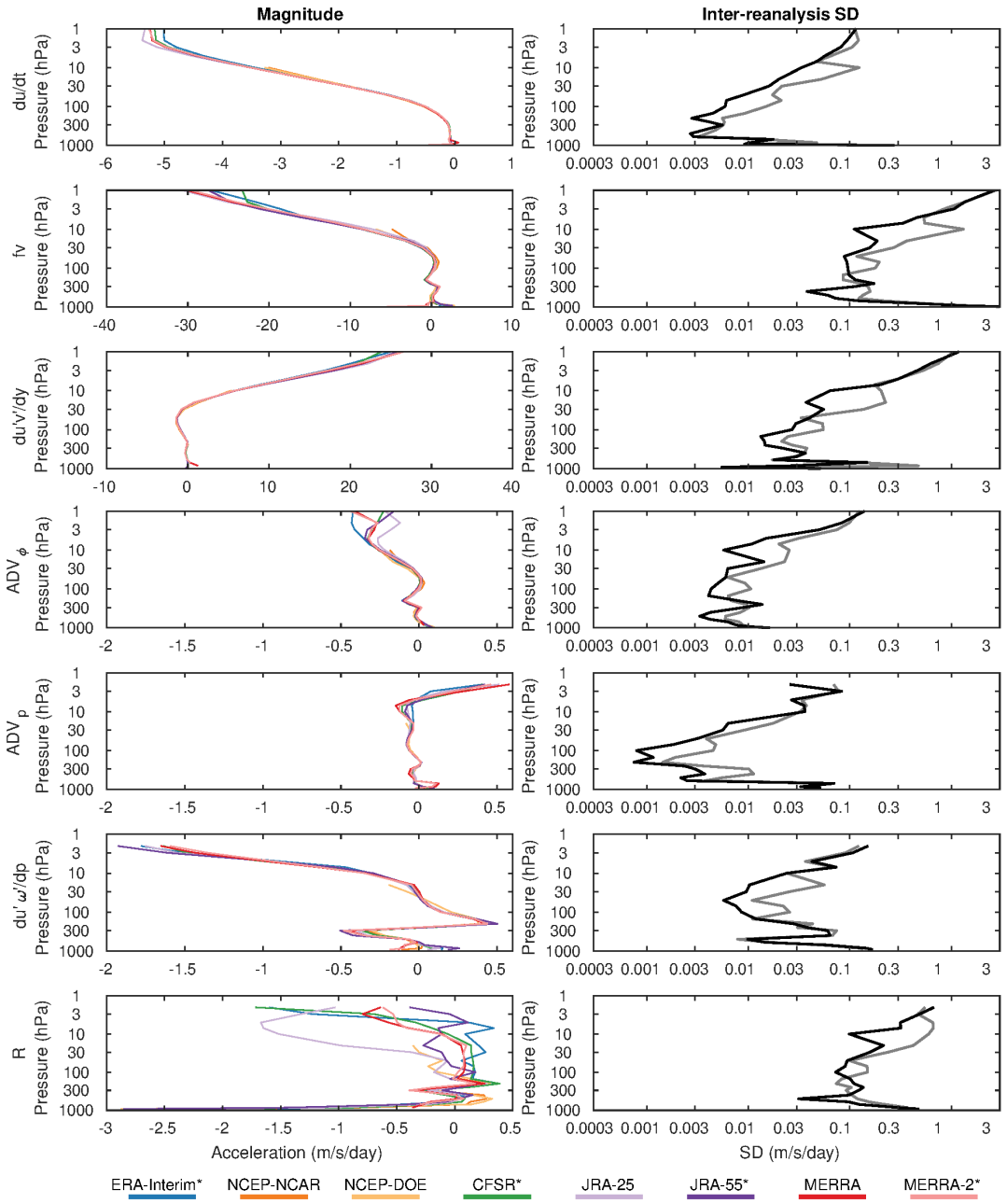

Figure 6.9: *Vertical profiles of each term in the momentum equation averaged from lags 5 to 0 days before SSW events. All variables are averaged between 45° N and 85° N. Individual reanalyses are shown to the left and the inter-reanalysis standard deviation is shown to the right on a logarithmic scale. The latter is shown for all reanalyses (grey) and for just the modern reanalyses (black; indicated with a * in the legend). All quantities are expressed in units of m s^{-1} day^{-1}. Reproduced from Martineau et al. (2018b).*

In keeping with the sign convention of *Thompson and Wallace* (2000), we also reverse the sign, so that a high index state is associated with a lower than average polar cap geopotential height.

This definition of the annular mode requires extrapolation of data to pressure levels below the surface in regions of high topography, which was done by the reanalysis centers with the exception of the MERRA products. To avoid introducing extrapolation errors, we omit MERRA and MERRA-2 from comparisons below 700 hPa. We focus on a subset of the pressure levels between 1000 hPa and 1 hPa that were shared by all reanalyses. Levels above 10 hPa, however, are unavailable for NCEP R1/NCEP R2 and 20CR v2/v2c reanalyses.

For the satellite era, 1979 onward, *Gerber and Martineau* (2018) found that a reanalysis ensemble mean (REM) constructed from the most recent reanalyses (ERA-Interim, JRA-55, and CFSR) provided a reliable benchmark for comparison. MERRA-2 was not included in the REM due to missing data below 700 hPa, but the results are nearly identical if it is included. The annular mode indices in the modern reanalyses are

correlated $R^2 > 0.96$ with each other at all levels in the Northern Hemisphere and up to 3 hPa in the Southern Hemisphere; CFSR's correlation with the others drops to $R^2 = 0.9$ at 1 hPa in the Southern Hemisphere. For the pre- satellite period, it was unclear if a REM was meaningful, particularly in the austral hemisphere. In the analysis shown in **Figure 6.10**, JRA-55 is chosen among modern full-input reanalyses as an arbitrary point of comparison.

Figure 6.10 contrasts consistency between the reanalyses in the post- and pre-satellite periods. To assess performance during the satellite era, **Figures 6.10a** and **c** correlate the annular modes computed from each individual reanalysis with the REM index over the standard WMO climatological period, 1981 - 2010. Essentially the same results would be found for any period after 1979, with some evidence of greater agreement in the last decades at upper levels (not shown). In the Northern Hemisphere, the annular mode indices computed from all of the full-input reanalyses are almost indistinguishable (the squared correlations are near one). In the Southern Hemisphere, there is reasonable agreement between all the full-input reanalyses ($R^2 > 0.95$ up to 10 hPa), but with evidence of tighter agreement amongst the more recent reanalyses ($R^2 \approx 0.99$ up to 3 hPa). While not shown here, an early output of ERA5 (2008 - 2016) was compared with the other modern reanalyses by *Gerber and Martineau* (2018) and shown to be as good as the other modern reanalyses.

In the Northern Hemisphere, the conventional-input JRA-55C reanalysis provides a very good estimate of the state of the annular mode up to 10 hPa. JRA-55C's annular mode index, however, is noticeably less correlated with the REM in the Southern Hemisphere, suggesting the satellite observations are critical for quantifying the large-scale circulation of the austral hemisphere. At the surface, and throughout most of the troposphere, the surface-input reanalyses 20CRv2/v2c and ERA-20C are also well correlated with the REM. The annular mode indices in the 20CR reanalyses, however, quickly decorrelate with the REM above the tropopause, suggesting that these reanalyses cannot effectively capture stratospheric variability. ERA-20C also loses skill in the stratosphere, but much more slowly, particularly in the Northern Hemisphere. The R^2 of approximately 0.6 at 10 hPa indicates that ERA-20C captures 60 % of the variance in the annular mode at this height in the stratosphere. As discussed in greater detail by *Gerber and Martineau* (2018), ERA-20C appears to capture approximately half of the observed SSWs, and simulates the same frequency of events overall.

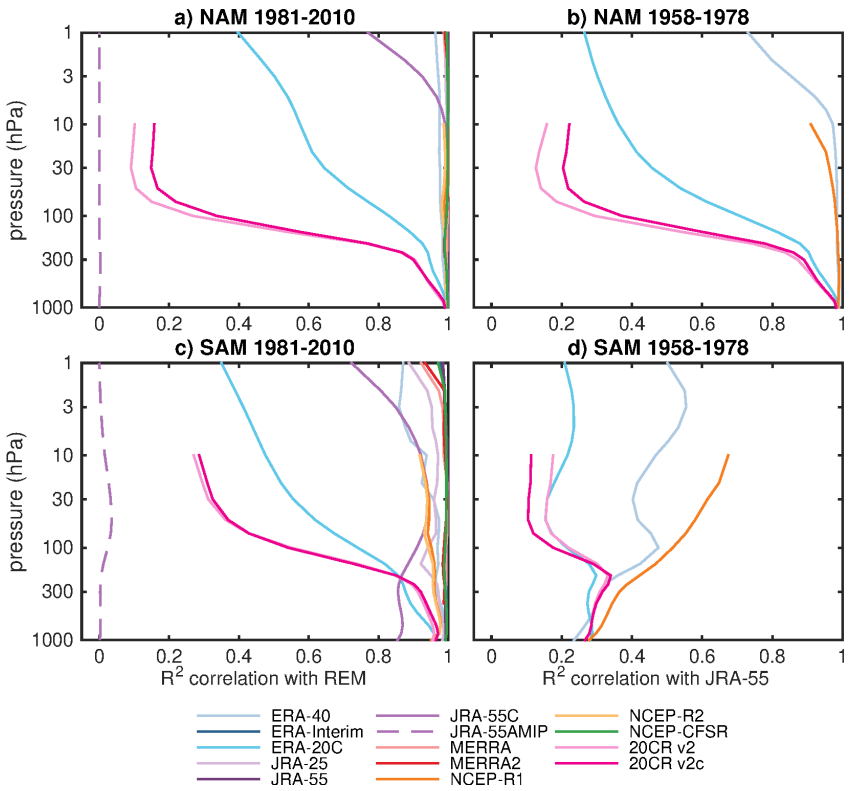

Figure 6.10: *The squared correlation between the (a, b) Northern and (c, d) Southern Annular Mode indices computed from each individual reanalysis with (a, c) a Reanalysis Ensemble Mean (REM) for the period 1981 - 2010, and (b, d) with the Annular Mode index of JRA-55 for the pre-satellite period, 1958 - 1978. As detailed in the text, the REM for the more recent period is constructed from three of the most recent reanalyses (ERA-Interim, JRA-55, and CFSR). In the pre-satellite period, a REM proved less meaningful. Comparable plots are obtained if NCEP R1 or ERA-40 are used instead of JRA-55. Adapted from Gerber and Martineau (2018).*

We note that the JRA-55AMIP integration does not meaningfully capture any of the annular mode variability. This was an expected result; this integration is not a reanalysis, but rather the JRA-55 model forced with observed SSTs, as in a standard Atmospheric Model Intercomparison Project (AMIP) simulation. Knowledge of the sea surface temperature is not sufficient to constrain the large-scale circulation of either hemisphere.

Only six reanalyses provide coverage in the pre-satellite era. Here we restrict ourselves to the period 1958 - 1979, as only NCEP R1 and the surface-input reanalyses extend further back in time, but *Gerber and Martineau* (2018) consider earlier periods. We have arbitrarily chosen JRA-55 as the reference time series among the modern full-input reanalyses for **Figures 6.10b** and **d**, but a qualitatively similar structure is found if ERA-40 or NCEP R1 is used instead. In the Northern Hemisphere, we find that the annular mode is consistently represented in the full-input reanalyses, with growing uncertainty above 10 hPa (where NCEP R1 is not available). This result is consistent with the ability of the conventional input reanalysis

JRA-55C to capture Northern Annular Mode variability in the satellite period.

While ERA-20C still captures more of the variability in the stratosphere in comparison to the 20CR reanalyses, the R^2 correlation is weaker in the pre-satellite period. At 10 hPa, ERA-20C captures only 40 % of the variability in the full-input reanalysis JRA-55 (or equivalently, ERA-40 and NCEP R1), compared to 60 % in in the satellite era. This could be due to fewer surface observations during this earlier period.

In the Southern Hemisphere, the situation is different. There is little agreement between JRA- 55 and the other reanalyses. Similarly poor agreement is found if NCEP R1 or ERA-40 is chosen as the reference time series (not shown), though we do find the NCEP R1 is somewhat better correlated with the surface-input reanalyses in the troposphere than either JRA-55 or ERA-40. The poor consistency between the reanalyses in the pre-satellite period was somewhat expected, given the inability of JRA-55C to capture the Southern Annular Mode in recent years. But the fact that JRA-55C still captures 85 % or more of the variance in the REM at nearly all levels suggests that a scarcity of conventional observations before 1979 is a larger part of the problem.

As discussed in *Gerber and Martineau* (2018), it is difficult to assess the synoptic variability of the Southern Annular Mode from direct measurements. On monthly time scales, *Marshall* (2003) has constructed a station based index that is correlated at approximately R = 0.85 with the 850 hPa Southern Annular Mode index in all reanalyses over the period 1979 - 2001. (This period was chosen to allow comparison with ERA-40.) For JRA-55 and ERA-40, this correlation drops markedly (to approximately 0.5) in the pre-satellite period 1958 - 1978. NCEP R1's correlation also weakens, but only drops to approximately 0.7. In contrast, the surface based reanalyses ERA-20C and 20CR maintain their correlation with the *Marshall* (2003) index.

The 20CR products, however, have been shown to miss most of stratospheric variability in earlier periods. Thus, for probing the large-scale circulation of the stratosphere-troposphere in the pre-satellite Southern Hemisphere atmosphere, ERA-20C might actually provide a more reliable estimate, even though NCEP R1, ERA-40, and JRA-55 assimilate radiosonde data and other free atmosphere observations.

6.5.2 Sampling uncertainty vs. reanalysis uncertainty

As found with the evolution of the stratosphere during an SSW event in *Section 6.4.3*, our ability to quantify the large-scale tropospheric response to SSWs and strong vortex events is primarily limited by the finite length of the reanalysis records, not differences between the reanalyses. **Figure 6.11** compares the sampling uncertainty in the "dripping paint" plots of *Baldwin and Dunkerton* (2001) to uncertainty associated with differences in the reanalyses. Panels (a) and (b) provide an update on the evolution of the annular mode index about weak and strong vortex events, now based on almost 6 decades of JRA-55 reanalysis. Following *Baldwin and Dunkerton* (2001), composites are centered about the date the 10 hPa index drops below -3 (rises above 1.5) standard deviations. The asymmetry in event criteria was based in part on the fact that the annular mode index at this level is skewed negative on account of SSWs, but 1.5 standard deviations is a much weaker threshold, such that more strong events are identified.

Gerber and Martineau (2018) show that using a consistent set of event dates is important for this comparison. The threshold nature of the event detection implies that very small differences between reanalyses can lead to the detection of different events (or more frequently, a shift in the timing of a given event). This effectively aliases sampling uncertainty in a comparison of reanalyses: the key is that the annular mode indices vary very little between reanalyses (differences are on the order of 1 %), but the inter-event variance is of order unity.

The weak vortex composite (**Figure 6.11a**) shows a rapid breakdown of the stratospheric polar vortex in the week preceding an event, evident first at upper levels, but become nearly synchronous in height by the time of onset. The stratospheric vortex then slowly recovers, from top to bottom, taking nearly three months in the lower stratosphere. During this long period of recovery, the tropospheric annular mode tends to be weakly negative, indicating an equatorward shift in the jet stream.

The strong vortex events (**Figure 6.11b**) exhibit a similar structure, but shifted earlier in time relative to date of event onset. The stratospheric vortex exhibits a positive annular mode (*i.e.*, is stronger than average) for over a month in advance, associated with a positive tropospheric annular mode (poleward shift in the tropospheric jet) that is already fully developed by the onset. This shift is partly due to the fact that strong vortex events tend to build slowly, on the time scale of radiative forcing, and so are harder to align in time. With respect to the amplitude of these events, pay close attention to the color scale. A weak vortex event is associated with a 3 standard deviation drop in the annular mode index in approximately 1 week, corresponding to a 1.4 km rise in the 10 hPa surface height at the pole. In contrast, the strong vortex event is associated with a more gradual 0.7 km drop in the 10 hPa surface over a month.

Figures 6.11c and **d** show the 1 standard deviation error bound on the weak and strong vortex composites, respectively. As shown by *Gerber and Martineau* (2018), inter-event variance of the annular mode indices is on the order of unity at all times except in the stratosphere at event onset (which occurs by construction: the 10 hPa index annular mode is always approximately -3 or 1.5 at lag 0). The sampling uncertainty of the composite is thus approximately 1 over the square root of the number of events. For weak vortex cases, where we have only 32 events, this is approximately 0.2, of the same order as the signal at any

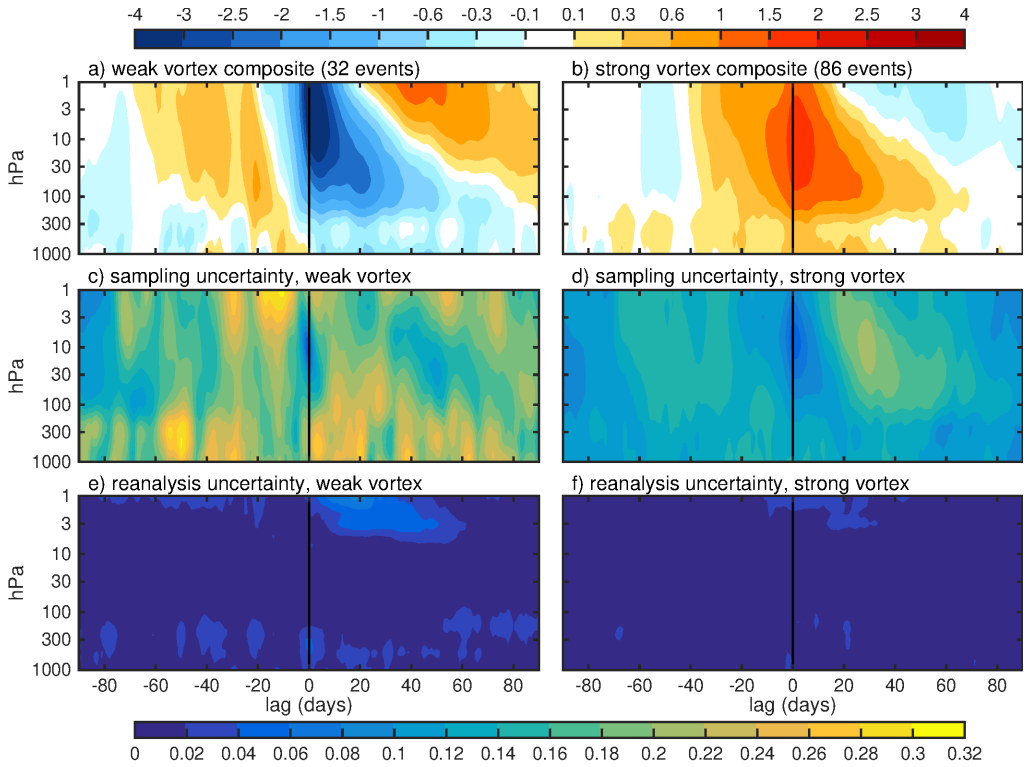

Figure 6.11: *Composites of the Northern Annular Mode indices as a function of lag and pressure for (a) weak and (b) strong vortex events, based on JRA-55 reanalyses over the period 1958-2016. Following Baldwin and Dunkerton (2001), weak (strong) events are identified when the NAM index at 10 hPa drops below -3 (rises above 1.5), and must be separated by a minimum of 30 days. The remaining panels quantify the uncertainty in the NAM index evolution as a function of lag and pressure. (c) and (d) show the sampling uncertainty in the mean weak/strong composites shown in* ***Figs. 6.11 (a) and (b)****, expressed as a one standard deviation error bound. Panels (e) and (f) show the reanalysis uncertainty: the standard deviation between composites of weak/strong vortex events based on the 4 most recent reanalysis products (ERA-Interim, JRA-55, MERRA-2, and CFSR/CFSv2) separately, for the period 1980 - 2016. As discussed in the text, a standardized set of event dates are used to prevent the aliasing of sampling error. Adapted from Gerber and Martineau (2018).*

given time! As argued by *Baldwin and Dunkerton* (2001), the tropospheric response is only significant if one averages over an extended period. This takes advantage of the fact that the tropospheric annular mode tends to exhibit memory on the order of 10 days (*e.g.*, *Gerber et al.*, 2010). If we ask for a 95 % confidence interval at any given time, we need the signal to be about equal to two standard deviations, requiring on the order of 100 events, a point we just approach in the case for strong vortex events.

Differences between the reanalyses are an order of magnitude smaller than the sampling error, as shown in **Figures 6.11e** and **f**. This measure of the "reanalysis uncertainty" was constructed by comparing weak and strong vortex composites based on the most recent reanalyses (ERA- Interim, JRA-55, MERRA-2, and CFSR/CFSv2) separately. We find that composites based on one reanalysis versus another are almost indistinguishable, provided one uses a standardized set of event dates. As the uncertainty is more than 10 times smaller than the sampling uncertainty, we'd need a record 100 times as long (*i.e.*, 6000 years!) for the choice of reanalysis to become as important as sampling uncertainty.

A similar conclusion applies to other measures of the coupling between the stratosphere and troposphere through the polar vortex, such as the variance and persistence of the annular mode indices as a function of season explored by *Baldwin et al.* (2003) and *Gerber et al.* (2010): results based on one reanalysis are not significantly different from those based on another with respect to the sampling uncertainty. This suggests that lengthening the reanalysis record has a substantial effect on our ability to quantify the coupling between the troposphere and stratosphere.

The sampling uncertainty shown in **Figures 6.11c** and **d** was based on JRA-55, which provides two additional decades (30 % more years) than the other most recent reanalyses which are restricted to the satellite era. As the sampling error decays with the square root of the number of events, these error bounds are 20 % smaller than could be obtained from the other modern reanalyses. This reduction depends on the assumption that JRA-55's reanalysis from 1958 - 1978 is of sufficiently high quality, supported by our comparison of the pre-satellite era reanalyses in **Figure 6.10**, and the fact that JRA-55C does a good job of capturing annular variability since 1979 without the aid of satellite observations. We look forward to assessing the ERA5 reanalysis, which is planned to extend back to 1950.

6.6 Stratospheric final warming events

The extratropical stratosphere exhibits a pronounced seasonal cycle with westerly winds in the winter hemisphere (with the exception of SSW events) and easterly winds in the summer hemisphere. The final transition from the westerlies to the easterlies, which occurs every year, is referred to as a Stratospheric Final Warming (SFW) event. Similar to SSW events, SFW events show a signature of zonal-mean zonal wind deceleration in the troposphere, indicative of a downward coupling, and a signature of enhanced upward Eliassen-Palm (EP) flux propagation to the stratosphere prior to the events (*Sun and Robinson*, 2009; *Black and McDaniel*, 2007). As such, they allow us to evaluate the representation of stratosphere-troposphere dynamical coupling in both hemispheres.

There is greater variability of final warmings in the Northern Hemisphere compared to the Southern Hemisphere, but stratospheric ozone loss has influenced their statistics in the Southern Hemisphere. Given their influence on the troposphere, the timing of the final warming has implications for seasonal forecasting (*e.g.*, *Butler et al.*, 2019; *Byrne and Shepherd*, 2018; *Lim et al.*, 2018; *Hardiman et al.*, 2011; *Ayarzagüena and Serrano*, 2009).

The final warming of the polar vortex is of key importance in chemistry-climate models. Once the polar vortex has broken down, ozone rich air can be transported to polar latitudes again. In the Southern Hemisphere, a late final warming in models will mean that the simulated Antarctic ozone hole persists longer through the year than is observed. A bias in the final warming time is also an indication of polar temperature biases, which will adversely affect the modelling of heterogeneous ozone destruction there (*Eyring et al.*, 2006). Adequate representation of the timing of the final warming in reanalysis data sets therefore has important implications for the evaluation of chemistry-climate models.

The final warming date is defined here as the day on which the zonal mean zonal wind at 60° becomes easterly for the final time during winter/spring. This can be sufficiently diagnosed using monthly mean data (calculating the day of the final warming using linear interpolation and assuming the monthly mean value represents the value on day 15 of the month) and occurs first in the mesosphere in the Southern Hemisphere (**Fig. 6.12**; shown only up to 1 hPa) but first in the mid-stratosphere in the Northern Hemisphere (Fig 6.12). With the exception of 20CR, all reanalysis products agree on the mean final warming date to within 6 days.

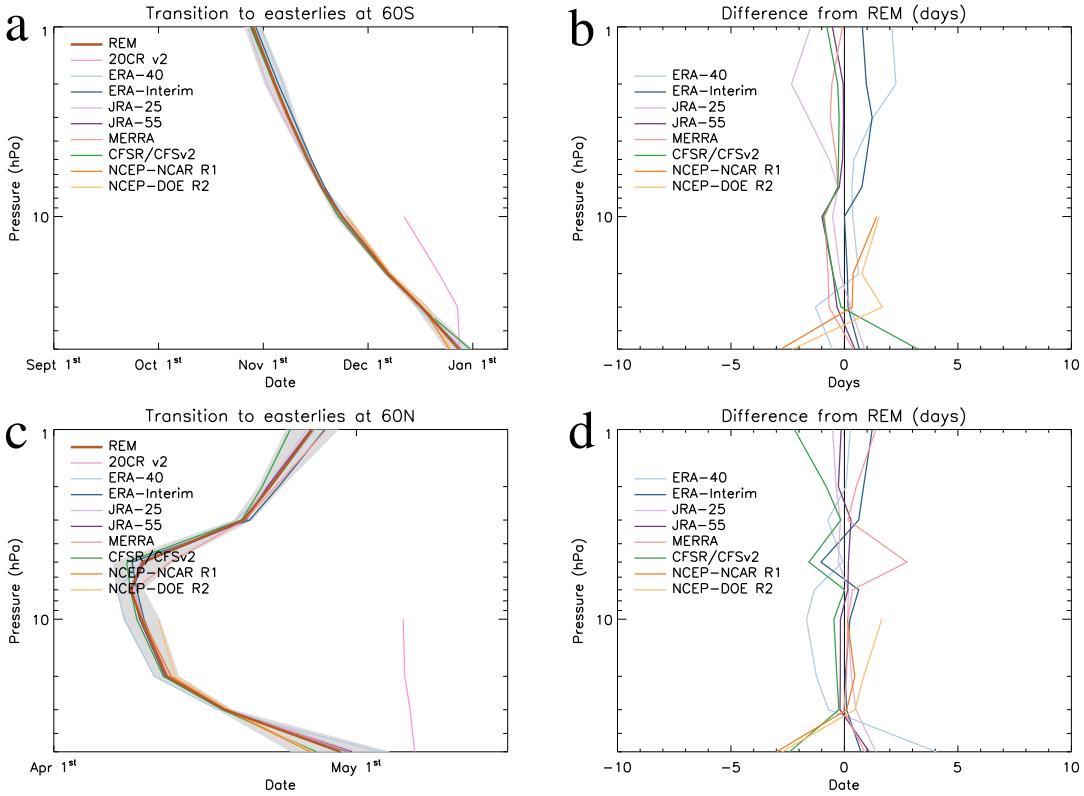

Figure 6.12: *The final transition of zonal mean zonal wind from westerly to easterly at (a) 60° S and (c) 60° N is shown for the period 1979 - 2010 for all reanalysis data products except 20CR v2 (which uses 1979 - 2009) and ERA-40 (which uses 1979 - 2002). The reanalysis ensemble mean (REM) is shown as a thick brown line, and uses data from all products except 20CR v2. The dark gray shading indicates the inter-reanalysis standard error (again excluding 20CR v2), scaled to represent a 95% confidence interval. The difference, in the final warming times shown in panels a and c, of each reanalysis from the multi-reanalysis mean is shown in (b,d). 20CR v2 is excluded from the REM since final warming times, especially in the Northern Hemisphere, are significantly later in this reanalysis and, given the remarkable agreement in final warming times across all other reanalysis datasets, the final warmings in 20CR v2 are very likely to be biased late.*

A closer study of the final warming in the Northern Hemisphere reveals that in some years the final warming occurs first in the mid-stratosphere ("10hPa-first years"), but in some years occurs first in the mesosphere ("1hPa-first years") (**Figure 6.13**). In 27 of the 32 years used, the reanalysis products all agree on the final warming type. Although there is generally a good agreement among full-input reanalyses, ERA-40 shows larger discrepancies in the mid- to lower- stratosphere transition date with respect to other data sets.

Correctly simulating the proportion of 10 hPa-first years and 1 hPa-first years is an area in which climate models do not currently perform well. In the reanalyses 68 - 79 % of years are 10 hPa-first years, whereas only 36 % of all modeled years, using the chemistry-climate models participating in phase 2 of the Chemistry-Climate model Validation activity (CCMVal-2) are 10 hPa-first years (*Hardiman et al.*, 2011). *Thiéblemont et al.* (2019) note a similar underestimation of 10 hPa-first years in the CESM and EMAC climate models.

6.7 Modulation of stratosphere-troposphere coupling by ENSO and QBO

The Northern-Hemisphere winter stratospheric polar vortex varies in strength from year to year with several external factors (*Yoden et al.*, 2002). One prominent source for this interannual variability is ENSO, the main mode of interannual variability in the tropical troposphere. During its warm phase (El Niño), Rossby wave trains propagate towards mid-latitudes in the Northern Hemisphere (NH) in boreal winter, strengthening the Aleutian low (*e.g.*, *Horel and Wallace*, 1981). As a consequence, upward propagation of planetary waves into the stratosphere is enhanced, which results in a weaker and a warmer polar stratosphere (*e.g.*, *Cagnazzo and Manzini*, 2009; *Brönnimann*, 2007). Although its teleconnectivity to the stratosphere is weaker than El Niño, the cold ENSO phase, La Niña, weakens the Aleutian low leading to reduced upward-propagating wave activity into the stratosphere and a strengthening of the polar vortex (*Iza et al.*, 2016; *Butler and Polvani*, 2011). For a comprehensive review of ENSO-stratosphere teleconnections, see *Domeisen et al.* (2019a).

ENSO's influence on the extratropical circulation is not limited to the time-mean flow. *Barriopedro and Calvo* (2014) found an ENSO modulation of the blocking precursors of SSWs, leading to distinctive wave signatures of SSWs during opposite ENSO phases: during El Niño, SSWs are predominantly associated with wavenumber-1 amplification in the lower stratosphere, whereas La Niña SSWs tend to occur after wavenumber-2 amplification (see also *Song and Son* (2018)). The way blocking events interfere with stationary waves and either amplify or damp the total injection of wave activity into the stratosphere depends critically on their location (*e.g.*, *Nishii et al.*, 2011; *Castanheira and Barriopedro*, 2010; *Woollings et al.*, 2010; *Martius et al.*, 2009).

Another source of interannual variability of the strength of the stratospheric polar vortex is the QBO which can modulate the nature and propagation of extratropical planetary-scale waves (*Garfinkel et al.*, 2012b; *Holton and Tan*, 1980). Several studies (*Taguchi*, 2015; *Richter et al.*, 2011; *Calvo et al.*, 2009; *Garfinkel and Hartmann*, 2007) further suggested some nonlinear influence of QBO and ENSO onto the stratospheric polar vortex such that when the QBO is in a westerly phase in the lower stratosphere, the polar night jet weakens and SSW probability increases for the warm ENSO phase (El Niño), whereas the changes are opposite for the QBO easterly winters.

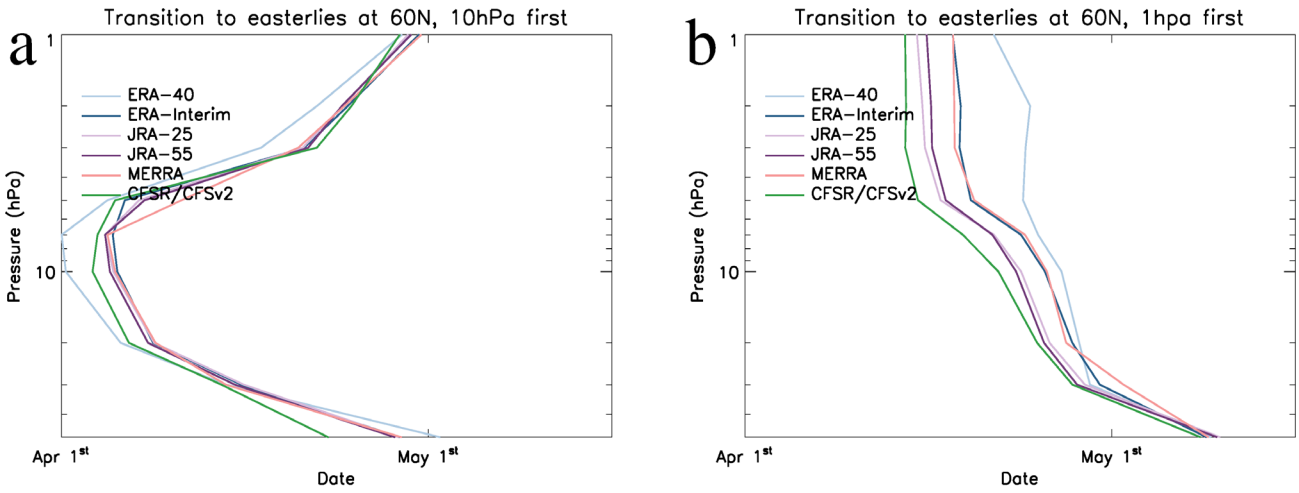

Figure 6.13: *Mean final warming date at 60 °N (as in* ***Figure 6.12****) composited over (a) 10 hPa-first years and (b) 1 hPa-first years (defined in text). The percentage of 10 hPa-first years is: 73.9 in ERA-40, 78.1 in ERA-Interim, 75.0 in JRA-25, 75.0 in JRA-55, 78.1 in MERRA, and 68.8 in CFSR. Data from the reanalyses 20CR, NCEP R1, and NCEP R2 does not extend above 10 hPa, so these products cannot be used for this diagnostic.*

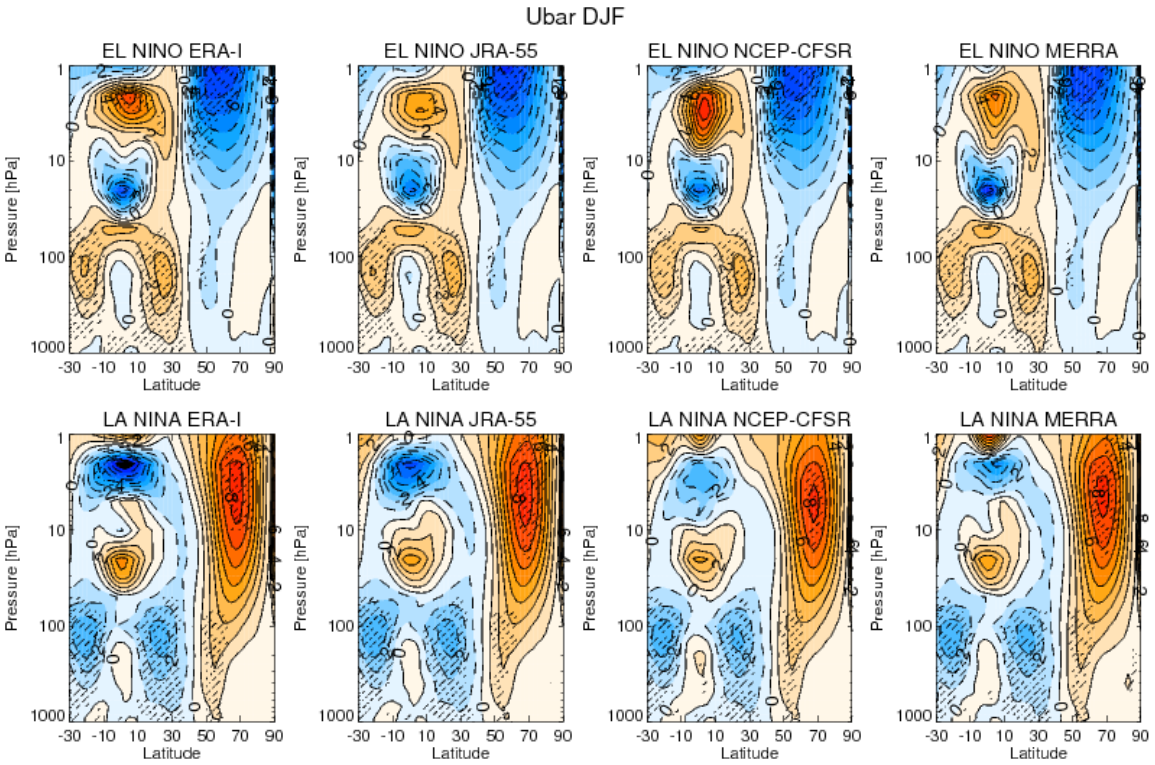

Figure 6.14: *Latitude-pressure cross sections of the composited DJF average of monthly zonal mean zonal wind anomalies for (top) El Niño and (bottom) La Niña events, from left to right for ERA-Interim, JRA-55, CFSR, and MERRA reanalyses. Contour intervals are ± 1 m s^{-1}. Solid (dashed) contours denote positive (negative) anomalies. Stippling indicates significance at the 95 % level.*

Alternative criteria have been used in the literature to define cold and warm ENSO phases. In the following analyses, we have focused on the most commonly used Niño 3.4 index based on monthly mean SST anomalies in the region from 5°S - 5°N and 170°E - 20°W with reference to 1981 - 2010 climatology. Standard El Niño and La Niña phases are defined by plus or minus 0.5 K anomalies in this region, as done in *Section 6.7.2*. In *Sections 6.7.1* and *6.7.3*, more restrictive criteria (1 standard deviation anomalies) were applied to focus on more extreme events. The period of averaging, DJF vs. a more extended winter season, was also varied depending on the scientific focus. The criteria for selecting warm and cold phases and the resulting years are therefore listed in each section; winters are identified by the year in January, *e.g.*, 1983 refers to the 1982 - 1983 winter.

6.7.1 Troposphere-stratosphere coupling through ENSO

The wintertime-mean stratospheric response to El Niño and La Niña conditions is first compared among reanalysis datasets. Monthly mean data from ERA-Interim, JRA-55, CFSR, and MERRA reanalyses are used. First, for each field, time series from 1979 to 2013 are detrended and anomalies are computed with respect to the 1981 - 2010 climatology. El Niño and La Niña events are defined using the standardized NDJF sea surface temperature anomaly of the Niño 3.4 index from the NCEP-CPC. El Niño (La Niña) winters are selected above (below) 1 SD (-1 SD). The composites include 7 El Niño winters (1983, 1987, 1992, 1995, 1998, 2003, 2010) and 5 La Niña winters (1989, 1999, 2000, 2008, 2011). The statistical significance of the composites is assessed with a Monte Carlo test at the 95 % confidence level.

Figure 6.14 shows the latitude-pressure cross-section of December-January-February (DJF) average of the zonal mean zonal wind anomalies composited for El Niño (up) and La Niña (bottom) events. In the polar stratosphere the El Niño (La Niña) signal is characterized by a robust weakening (strengthening) of the zonal mean zonal wind in all reanalyses. All reanalyses agree on the significant area and the sign of the anomalies, with the largest polar stratospheric signal peaking at -7 m s^{-1} for El Niño and 8 m s^{-1} for La Niña. Therefore, a good agreement across reanalyses is found for El Niño and La Niña polar stratospheric responses.

To quantify the relationship between the strength of the Aleutian low, modulated by ENSO, and the response of the stratospheric polar vortex, **Figure 6.15** shows the scatter plot of the Z index at 500 hPa (average of geopotential height anomalies between 40°N-60°N and 180°E - 210°E) versus the U index (zonal mean zonal wind averaged at 60°N between 10 hPa and 30 hPa), similar to *Cagnazzo et al.* (2009). It is important to note that these Z and U index values for each event are very similar among reanalyses. El Niño winters (squares) are associated with negative values of the Z and U indices. This corroborates that the deepened Aleutian low related

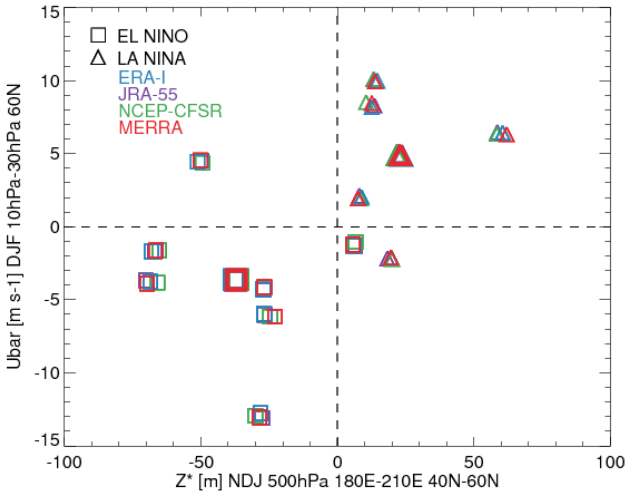

Figure 6.15: *Scatter plot of the NDJ mean Z index, versus DJF U index. Squares (triangles) represent each El Niño (La Niña) event and the corresponding larger symbols represent El Niño and La Niña events composite.*

to the negative Z index increases the upward wave propagation into the stratosphere leading to a weaker polar vortex. In contrast, La Niña winters (triangles) are mainly related to positive Z and U indices, due to an anomalously weak Aleutian low and in agreement with the observed positive wind anomalies respectively, since a weakened Aleutian low inhibits the upward wave propagation leading to a stronger polar vortex. Results show an excellent agreement among reanalyses. Therefore, we conclude that for the purpose of studying the coupling between the stratosphere and the troposphere during El Niño and La Niña events, any of the compared reanalyses is equally suitable, as *Iza et al.* (2016) noted for La Niña events.

6.7.2 Blocking patterns associated to SSWs and the modulation of ENSO

The intercomparison of ENSO's influence on the stratosphere among reanalyses is then extended to ENSO's influence on SSW events and their blocking precursors. The analysis contrasts inter- dataset uncertainties with the uncertainties associated with the definition of blocking events by using three different blocking definitions.

Daily mean geopotential height at 500 hPa (Z500) and 100 hPa (Z100) is used for this analysis which is performed for the full 1958 - 2012 period and the 1979 - 2012 satellite period. For the latter period, the REM is computed from the CFSR, ERA-Interim, JRA-55 and MERRA reanalyses. The REM of the full period is based on the NCEP R1, ERA-40 (completed with ERA-Interim from 2002 to 2012 since the two agree well over their overlapping period from 1979 to 2002) and JRA-55 reanalyses. Fields were interpolated (if required) to the same common 2.5° × 2.5° grid before any further analysis is carried out. Anomalies are defined with respect to the daily climatology of 1981 - 2010.

SSW central dates are chosen from the common dates identified in **Table 6.2**. ENSO winters were characterized by the NDJFM average of the monthly Niño 3.4 index (**http://www.cpc.ncep.noaa.gov/**). EN and LN winters were identified when Niño 3.4 ≥ 0.50 °C and Niño 3.4 ≤ 0.50 °C, respectively. The resulting warm phase years are 1958, 1966, 1969, 1973, 1983, 1987, 1988, 1992, 1995, 1998, 2003, 2005, 2007, and 2010. Cold phases were identified in 1962, 1963, 1965, 1967, 1968, 1971, 1972, 1974, 1975, 1976, 1984, 1985, 1989, 1996, 1999, 2000, 2001, 2006, 2008, 2009, 2011, and 2012.

We employed three blocking detection methods, which cover most approaches to blocking definition: 1) the absolute method (ABS), based on the detection of reversals in the meridional Z500 gradient; 2) the anomaly method (ANO), using Z500 anomalies above a given threshold; 3) the mixed method (MIX), a hybrid definition of the two previous approaches. These definitions are described in more details in *Woollings et al.* (2018). All methods give two preferred regions for blocking occurrence: one over the Atlantic and one over the Pacific basins, with maximum blocking frequencies of about 15 % of days in NDJFM. However, there are substantial differences among definitions in the blocking location within each basin as well as in the relative frequencies of Atlantic vs Pacific blocking (*Woollings et al.*, 2018).

Blocking precursors of SSWs were identified for each reanalysis by performing 2-D composites of blocking frequency for the [-10,0]-day period before the central dates of SSWs. This was carried out separately for SSWs occurring during El Niño and La Niña winters. The REM for the full period is shown in **Figure 6.16**. There is a spatial preference for different blocking precursors of SSWs depending on the ENSO phase, with enhanced (reduced) blocking frequencies over eastern North America and the North Alantic (eastern Pacific) during El Niño, and nearly opposite patterns for La Niña winters. Thus, SSWs are often preceded by North Atlantic sector blocking during El Niño, while eastern Pacific blocks are the preferred precursors of SSWs in La Niña winters. The comparison across reanalyses reveals a good agreement, with differences that are much smaller than among blocking definitions (everywhere except the blue dots in **Figure 6.16**). The intensity, significance and spatial extension of the signal weaken for the satellite period (1979 - 2012, not shown).

The composites of blocking precursors of SSWs for El Niño and La Niña winters are similar to those obtained for W1 and W2 SSWs, respectively (*Ayarzagüena et al.*, 2019; *Song and Son*, 2018), which hints at a modulation of the characteristics of SSW events. To further illustrate the association between ENSO and the dominant wave signatures of SSWs, the temporal evolution of Z100 wavenumber components are evaluated for the [-30, 30]-day period surrounding the central date of SSWs (**Figure 6.17**). The results confirm that SSWs are significantly preceded by wavenumber-1 amplification during El Niño, whereas SSWs preferably occur after wavenumber-2 amplification in La Niña winters (*Taguchi and Hartmann*, 2006).

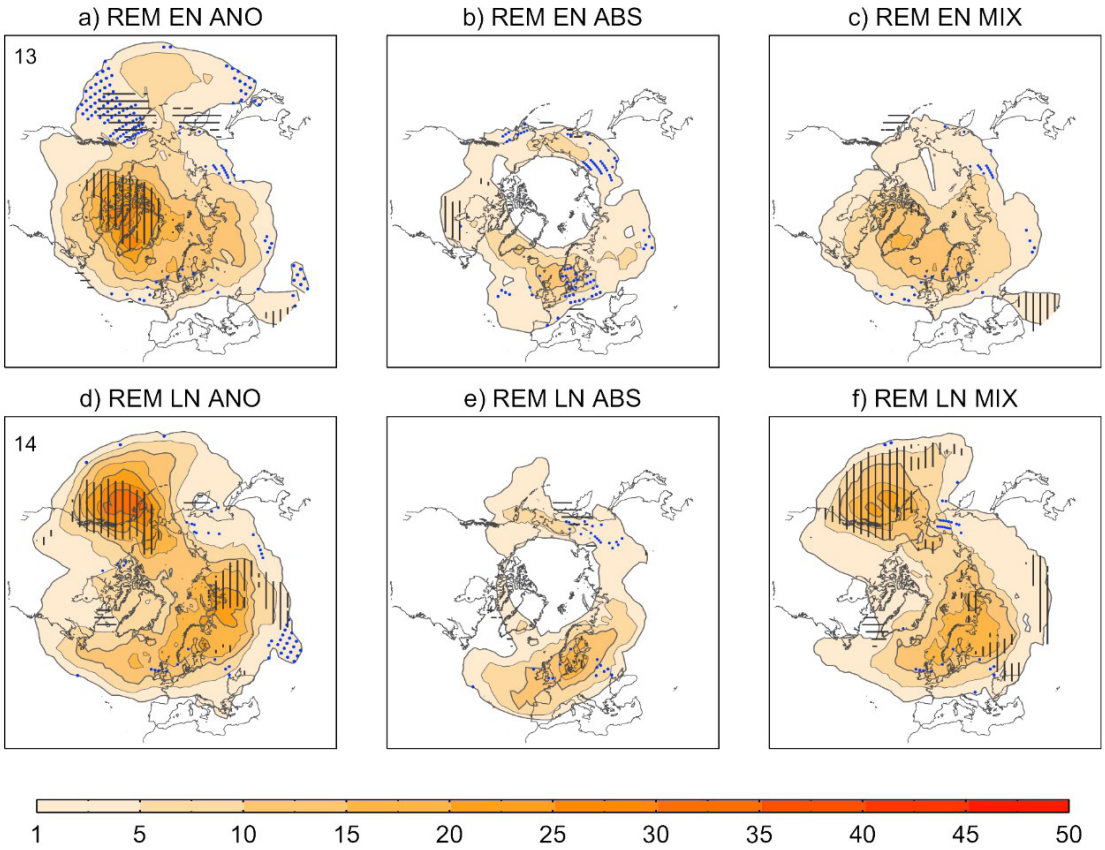

Figure 6.16: *Reanalysis ensemble mean composites of blocking frequency for the [-10,0]-day period before the central dates of SSWs occurring during El Niño (top) and La Niña (bottom) winters of the 1958 - 2012 period for three different blocking definitions (columns). The blocking frequency is expressed as the percentage of time (over the 11-day period) during which a blocking was detected at each grid point. Vertical (horizontal) black lines indicate regions with blocking activity significantly higher (lower) than the climatology at the 90 % confidence level in at least 66 % of the reanalyses. The significance is derived from a bootstrap of 1000 members, each one containing the same number of cases and dates as the SSWs of each composite but with random years of occurrence. Blue dots highlight grid points where the inter-reanalysis spread for a given blocking definition is larger than the spread across the reanalysis ensemble mean of blocking definitions. The numbers in the upper left corner of panels a), d) indicate the sample size of SSWs during El Niño and La Niña winters, respectively. Adapted from Ayarzagüena et al. (2019).*

During La Niña, the wavenumber-2 signal is accompanied by significant anomalies in wavenumber-1, albeit they are smaller and/or shorter-lasting. This difference in wave driving does not, however, necessarily affect the ratio of vortex splits to displacement (*Garfinkel et al.*, 2012a). This modulation of SSW characteristics by ENSO is achieved through a change in the preferred blocking location, which injects different scales of wave activity into the stratosphere, and thus forces different types of SSWs (*e.g.*, *Barriopedro and Calvo*, 2014). We note that the sensitivity of these results to the choice of reanalysis is very weak.

The modulation of SSW properties by ENSO is robustly observed across reanalyses when the 1958 - 2012 period is used, but less evident in the 1979 - 2012 period. This suggests decadal variability in the ENSO-blocking-SSW relationship (*e.g.*, *Rao et al.*, 2019), biases in the pre-satellite period or sampling issues affecting the shorter satellite period. The differences among blocking definitions are much larger than differences among reanalyses, likely contributing to the discrepancies in the blocking-SSW relationship reported in the literature.

6.7.3 Nonlinear modulation of the extratropical stratosphere by ENSO and QBO

Finally, we evaluate the representation of nonlinearities in the modulation of DJF-mean polar vortex strength and SSW occurrence with ENSO and QBO among reanalyses. SSW onset dates are defined by the common dates established in *Section 6.4.1* and the DJF zonal mean zonal wind at 60 °N and 10 hPa is used as a proxy for strength of the polar night jet. The analysis period ranges from 1979 to 2011, except for 20CR v2 (1979 - 2010). The DJF climatology for each reanalysis is based on the 1981 - 2010 period.

In order to define ENSO phases, the monthly Niño 3.4 index (provided by NOAA/CPC) is averaged over DJF. The DJF mean of the zonal mean zonal wind at the equator and 50 hPa in the respective reanalyses is used to define QBO phases. All DJF seasons are classified into six groups defined by three ENSO and two QBO conditions. Two of the three ENSO conditions are El Niño and La Niña,

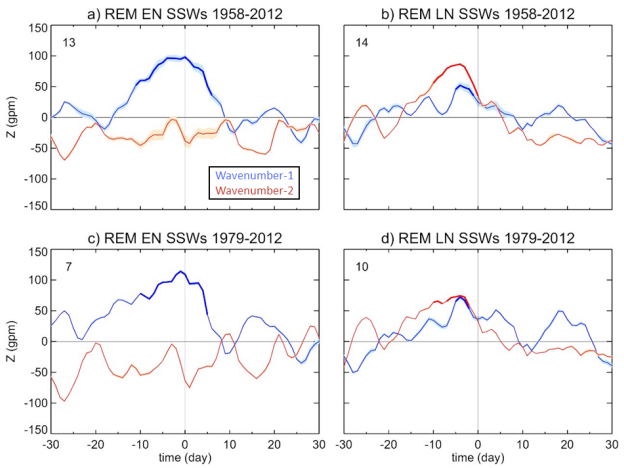

Figure 6.17: *REM composites of the temporal evolution of 100 hPa geopotential height wavenumber-1 (blue) and wavenumber-2 (red) amplitude anomalies at 60° N (gpm) for the [- 30,30]-day period around the central dates of SSWs occurring during El Niño (left) and La Niña (right) winters of the full (1958 - 2012, top) and satellite (1978 - 2012, bottom) period. Shading denotes the ±2 sigma level across reanalyses. The time intervals highlighted with thick lines indicate significant differences with respect to climatology at the 95% confidence level in at least 66% of the reanalyses. The significance is assessed with a bootstrap test of 1000 samples with the same number of cases and calendar days as the SSWs of each composite but with random years of occurrence. The numbers in the upper left corner of each panel indicate the sample size of the composite. This figure differs from* ***Fig. 6*** *of Barriopedro and Calvo (2014) due to the addition of an extra SSW and a different ENSO classification with updated ENSO indices.*

when the DJF mean Niño 3.4 index exceeds ±1 standard deviation (both inclusive); warm phase years (EN) were identified in 1983, 1987, 1992, 1998, and 2010, and cold phases (LN) in 1989, 1999, 2000, 2008, and 2011. The third ENSO condition is neutral (NT) for remaining years. The mean and standard deviation of the ENSO index are calculated for the 1981 - 2010 period.

The two QBO conditions are easterly and westerly, when the DJF mean zonal wind in the equatorial lower stratosphere is negative or positive (the latter includes zero). The resultant grouping is slightly different among the reanalyses, as the equatorial zonal wind is different. Alternatively, we also use 50 hPa equatorial zonal wind data archived at Free University Berlin to standardize the QBO classification for all reanalyses, which yields similar results (not shown). It is noted that the classification of about 30 years into the six groups implies that in some cases the sample size is small and therefore it is difficult to obtain statistically significant results. Before focusing on changes with ENSO and QBO, it will be useful to mention that except for 20CR v2, the interannual variability of the DJF mean vortex strength is highly correlated between the reanalyses, with correlation coefficients over +0.99.

Figure 6.18 shows heat maps based on composite zonal wind anomalies of each reanalysis for the six groups. As expected from the high correlations of the interannual variability of the zonal wind, the plots show that the changes in the vortex strength with ENSO and QBO are more or less similar among the reanalyses, except for 20CR v2 whose climatological vortex strength is notably more than twice that of other reanalyses. Specifically, the zonal wind anomalies tend to be slightly negative for easterly QBO winters regardless of the ENSO conditions. For westerly QBO conditions, the zonal wind anomalies exhibit a clear decreasing tendency with the ENSO SST conditions, from La Niña (positive wind anomalies) through neutral to El Niño (negative anomalies).

It is noted that the sample sizes of the six groups are not the same among all reanalyses, implying that the equatorial zonal wind and hence QBO classification are different in some cases. This may matter when one extracts changes with QBO conditions that are defined using an equatorial zonal wind index in the respective reanalyses, but this effect seems limited here since most reanalyses show similar results.

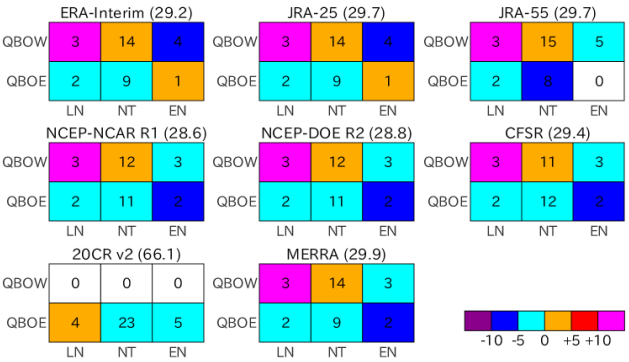

Figure 6.19: *Heat maps showing composite anomalies of the DJF-mean zonal wind at 60° N, 10 hPa for the six groups in the eight reanalyses as indicated above each panel. The number indicated in each cell denotes the sample size. Each panel also includes the DJF climatological wind value in ms^{-1} in parentheses.*

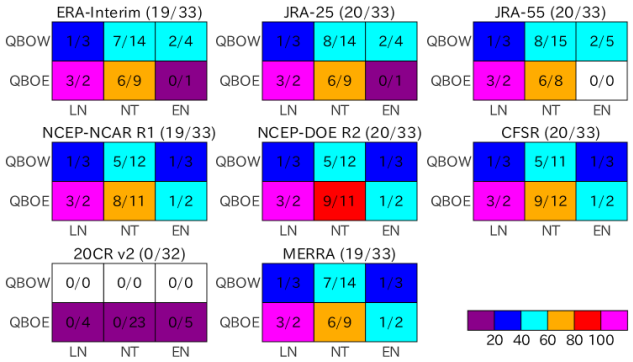

Figure 6.18: *Same as* ***Figure 6.18****, but for SSW probabilities (in %) computed as the ratio of the number of SSWs to the number of years for each group (indicated in each cell). The numbers in the title indicate total number of SSW / total number of years.*

Figure 6.19 similarly shows heat maps of SSW probabilities for the six groups. Here, for each group, the SSW probability is a ratio (times 100) of the number of SSWs to the number of years. The charts show that although the classification of QBO years is slightly different among the reanalyses, the changes in the SSW probabilities are similar among the data sets, except for 20CR v2 which has no SSWs. For easterly QBO, the probabilities tend to decrease from La Niña, through neutral, to El Niño. A characteristic feature is the highest SSW probability for La Niña and easterly QBO (group 1). On the other hand, for westerly QBO, the probabilities slightly increase in the opposite way, from La Niña, through neutral, to El Niño, consistent with the changes in the zonal wind anomalies (**Figure 6.18**). These changes in the SSW probabilities do not necessarily match changes in DJF zonal wind anomalies (**Figure 6.18**), since the occurrence or absence of a SSW during each winter depends not only on the DJF mean vortex strength but also on its variance.

6.8 Stratosphere-troposphere coupling through the antarctic hzone hole

In recent decades, severe stratospheric ozone depletion has led to the Antarctic ozone hole in austral spring (*Thompson and Solomon*, 2002). This has resulted in substantial cooling in the lower stratosphere, leading to an increase in the latitudinal temperature gradient and a consequent strengthening of the stratospheric polar vortex. Through mid-to-late spring, this mid-to-high latitude circulation anomaly descends from the lower stratosphere to reach the troposphere during austral summer (*e.g.*, *Son et al.*, 2018). The anomalous tropospheric circulation is associated with a noticeable increase in zonal mean sea level pressure difference between the mid and high latitudes, commonly referred to as an increase in the positive phase of the Southern Annular Mode (SAM). The positive SAM is generally marked by a poleward displacement and intensification of the tropospheric mid-latitude jet.

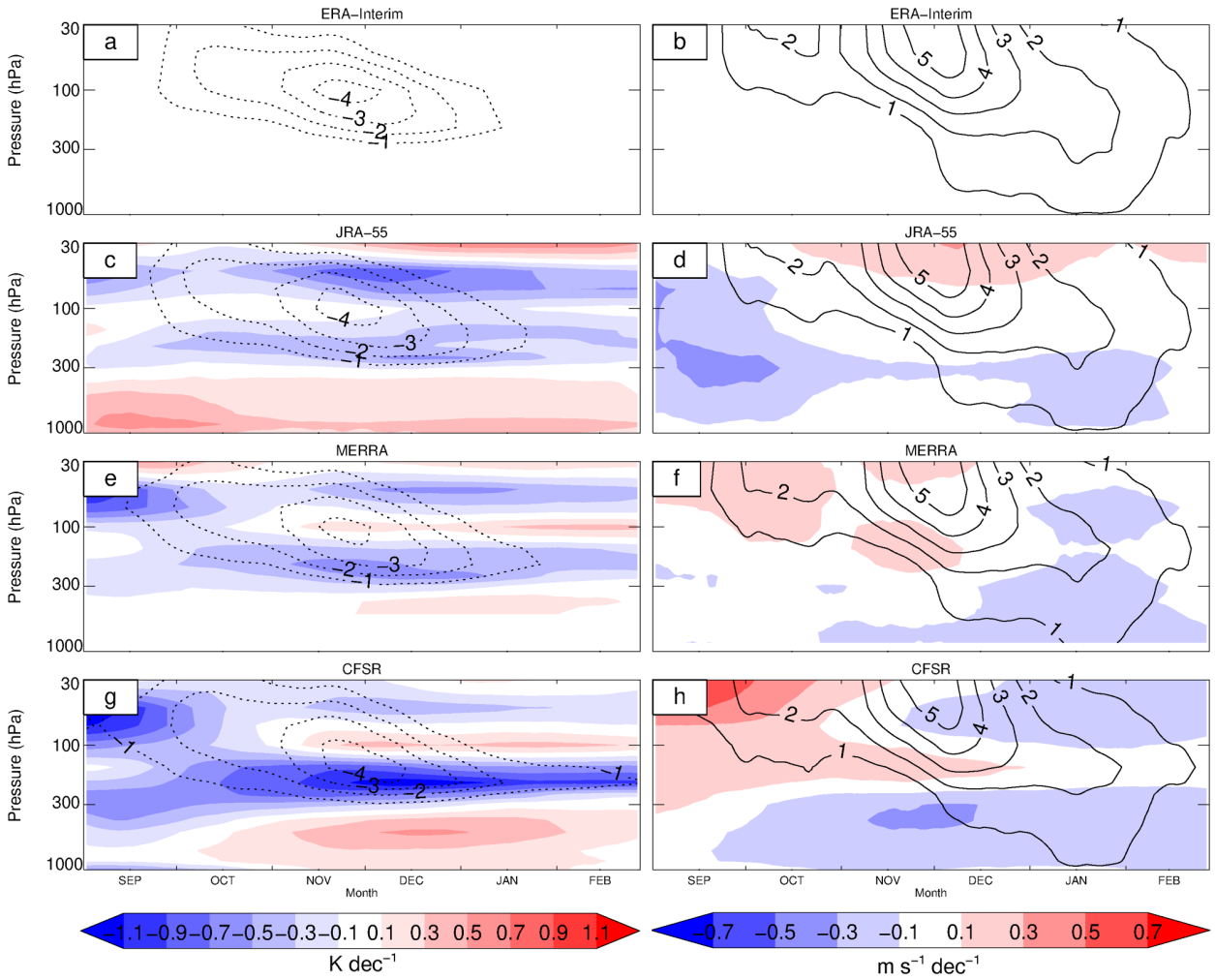

Figure 6.20: *(left) The dashed contours show time-height cross sections of zonal-mean temperature trend (with contour intervals of 1 K dec^{-1}) averaged over latitudes 60-90° S during 1979-2001 for a) ERA-Interim, c) JRA-55, e) MERRA, and g) CFSR. The shadings show the differences between the various reanalyses and ERA-Interim at intervals of ±0.1, ±0.3, ±0.5, ±0.7, ±0.9 and ±1.1 K dec^{-1}. (right) The contours show time-height cross sections of zonal-mean zonal wind trend (with contour intervals of 1 m s^{-1} dec^{-1}) averaged over latitudes 50-70° S during 1979 - 2001 for b) ERA-Interim, d) JRA-55, f) MERRA, and h) CFSR. The shadings show the differences between the various reanalyses and ERA-Interim at intervals of ±0.1, ±0.3, ±0.5, and ±0.7 m s^{-1} dec^{-1}.*

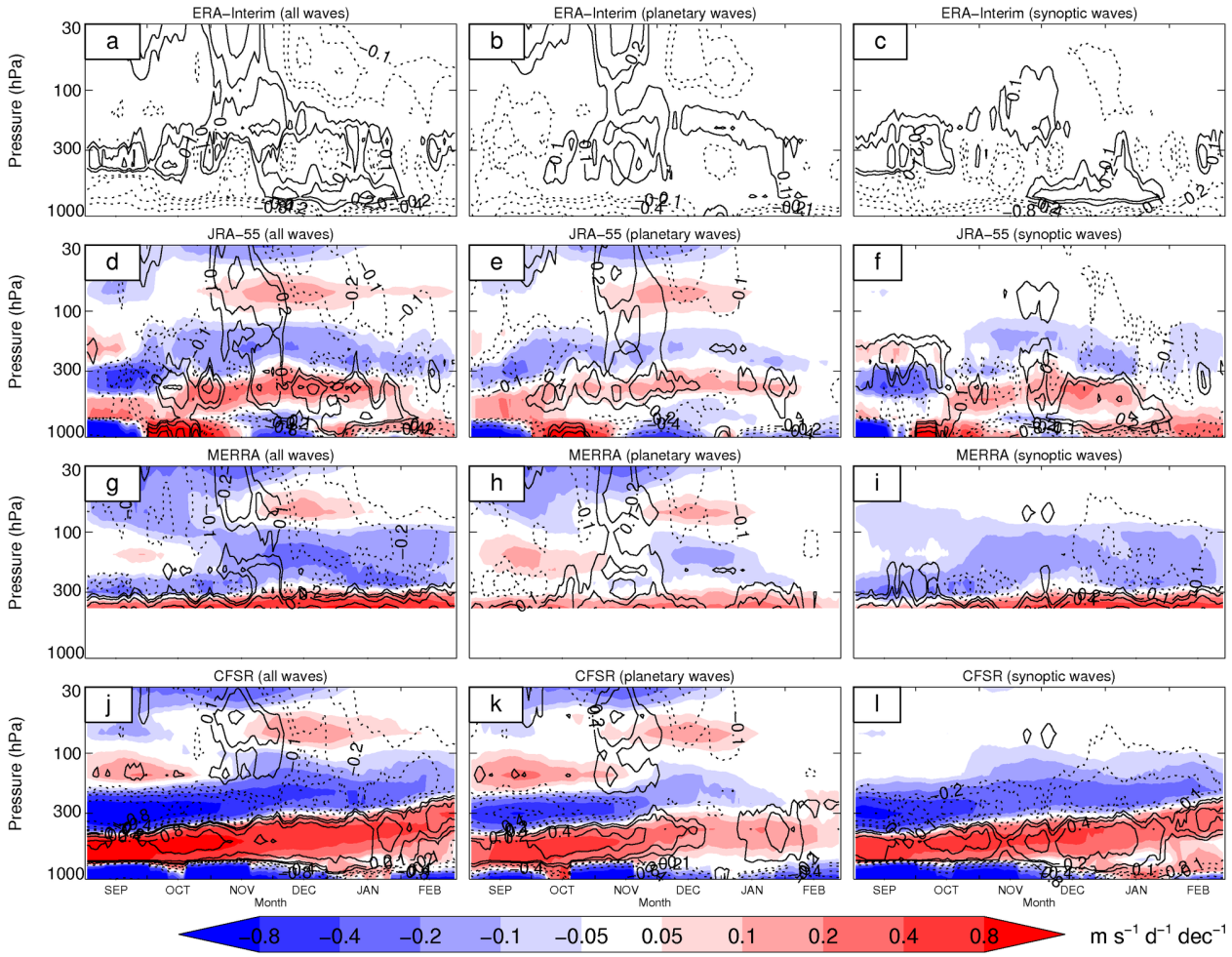

Figure 6.21: *The contours show time-height cross sections of the trend in the EP flux divergence due to all waves (left), planetary-scale waves (middle) and synoptic-scale waves (right) averaged over the latitude band of 40-80° S during 1979 - 2001 for (a-c) ERA-Interim, (d-f) JRA-55, (g-i) MERRA, and (j-l) CFSR at intervals of ±0.1, ±0.2, ±0.4 and ±0.8 m s^{-1} d^{-1} dec^{-1}. Solid and dashed contours indicate positive and negative values, respectively. The shadings show the differences between the various reanalyses and ERA-Interim at intervals of ±0.05, ±0.1, ±0.2, ±0.4 and ±0.8 m s^{-1} d^{-1} dec^{-1}. Note that MERRA data below 400 hPa is excluded.*

Although ozone loss has a direct impact on stratospheric temperatures by reducing the absorption of incoming solar radiation, a number of studies show that the anomalous circulation is strongly influenced by changes to wave forcing and wave mean-flow interaction (*Orr et al.*, 2012). Here, the impacts of the ozone hole on the dynamical coupling between the stratosphere and the troposphere in the spring and summer Southern Hemisphere are examined in the ERA-Interim, JRA-55, MERRA, and CFSR reanalyses datasets. A more detailed analysis is provided in Orr et al. (2021).

Figure 6.20 shows the trends in zonal mean temperature over the SH polar region between 1979 and 2001 for the four datasets. This period is chosen for two reasons. First, the size of the ozone hole increased steadily during this period (*Huck et al.*, 2007). Second, the trends in the four reanalyses were largest for this period (not shown), which allows to identify important differences between the datasets. In ERA-Interim, the cooling starts at 30 hPa in October and peaks at around 100 hPa between mid-November and early December (with trends reaching -4 K per decade), which is in good agreement with radiosonde data from Antarctica (*Thompson and Solomon*, 2002). The other three reanalyses all show broadly similar results with downward descent pattern from 30 hPa to 300 hPa. However, compared to ERA-Interim, CFSR shows considerably stronger and longer-lasting cooling (by up to -1 K dec^{-1}) between 100 hPa and 300 hPa, and enhanced warming below 300 hPa (by around 0.5 K dec^{-1}). This would lead to a comparative weakening of the atmospheric stability near the tropopause. In both CFSR and MERRA, the cooling also starts noticeably earlier than ERA-Interim.

Figure 6.20 also shows the corresponding trends in zonal wind over the SH polar regions, with all four reanalyses showing the expected strengthening of the SH circumpolar winds from the lower stratosphere down to the surface. In ERA-Interim, the strengthening starts in mid-to-late September at 30 hPa, peaks at around 5 m s^{-1} dec^{-1} between late November and early December, and reaches the lower troposphere in January. The results from ERA-Interim, JRA-55 and MERRA are in relatively good agreement, with differences not exceeding ± 0.3 m s^{-1} dec^{-1}.

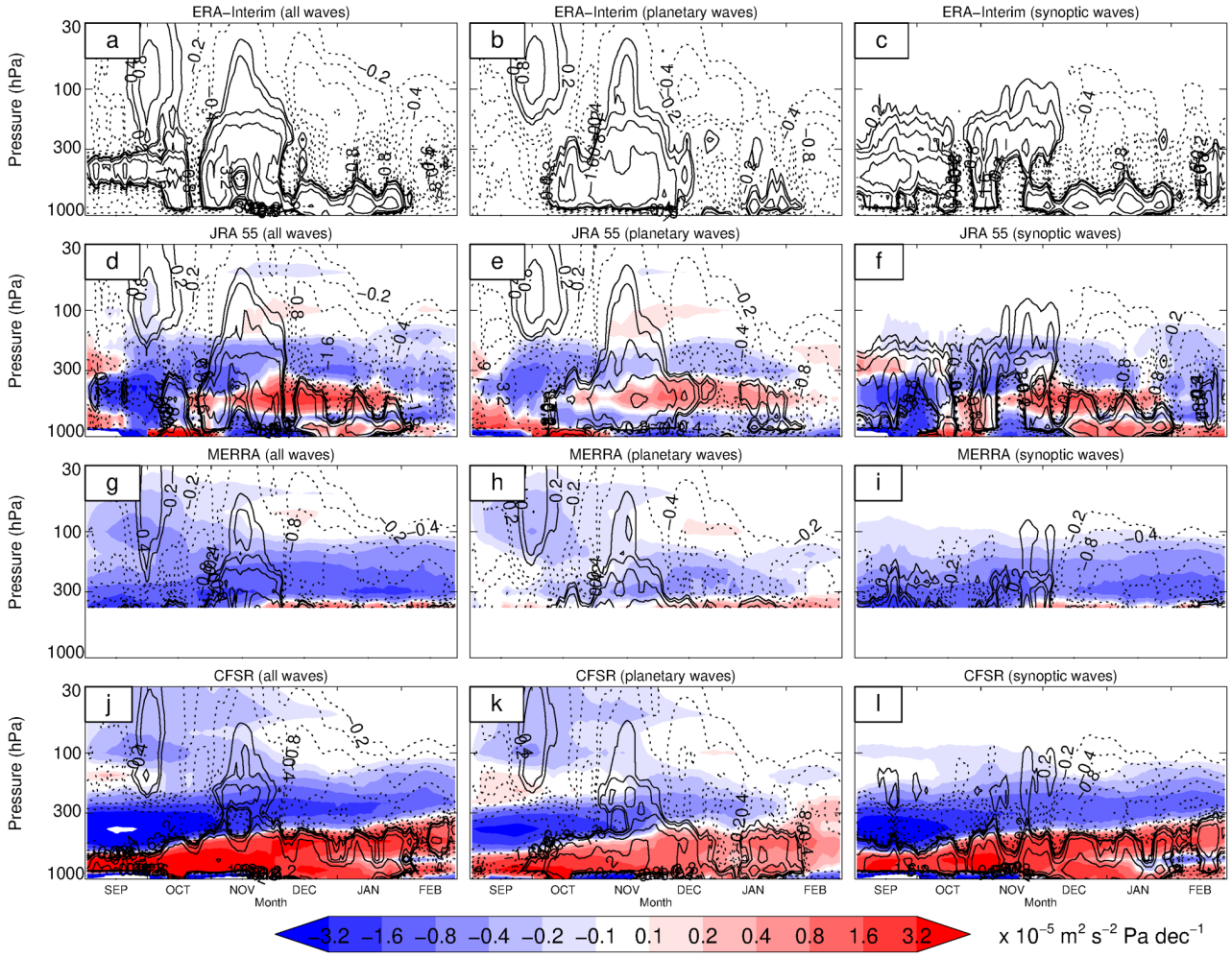

Figure 6.22: *The contours show time-height cross sections of the trend in the vertical component of EP flux due to all waves (left), planetary-scale waves (middle) and synoptic-scale waves (right) averaged over the latitude band of 40-80° S during 1979 - 2001 for (a-c) ERA-Interim, (d-f) JRA-55, (g-i) MERRA, and (j-l) CFSR at intervals of 0.2, 0.4, 0.8, 1.6, 3.2 and 10.0 x 10^{-5} m^2 s^{-2} Pa dec^{-1}. Solid and dashed contours indicate positive and negative values, respectively. The shadings show the differences between the various reanalyses and ERA-Interim at intervals of ±0.1, 0.2, 0.4, 0.8, 1.6 and 3.2 x 10^{-5} m^2 s^{-2} Pa dec^{-1}. Note that MERRA data below 400 hPa is excluded.*

However, in CFSR the initial strengthening of the winds in the lower stratosphere occurs earlier than in ERA-Interim, while in the lower troposphere they are delayed, indicating a comparatively slower downward descent rate in CFSR.

Figure 6.21 shows the trends of the total Eliassen-Palm (EP) flux divergence from 40 - 80 °S derived from the four datasets. We note that MERRA data is excluded from this analysis below 400 hPa because, unlike other reanalyses, data is not extrapolated below the surface and thus zonal-mean diagnostics are not comparable. The EP flux is assessed from a common grid for all data sets (*Martineau et al.*, 2018c; *Martineau*, 2017). In ERA-Interim, there are positive EP flux divergence anomalies from September to November and negative EP flux divergence anomalies from December to February in the lower stratosphere, which imply a strengthening of the polar vortex in spring followed by a delayed breakup of the vortex in summer. This is consistent with the circulation changes shown in **Figure 6.20**. In the stratosphere, the anomalies of EP flux divergence are dominated by planetary waves. In the troposphere, both planetary and synoptic waves are affected. In late austral spring, a region of positive EP flux descends from the upper troposphere down to the surface, which is dominated by planetary waves in the upper troposphere and synoptic waves in the lower troposphere. These wave forcing anomalies are consistent with the downward descent of strengthened circumpolar winds, shown in **Figure 6.20**. The other three reanalyses show a broadly similar pattern in the stratosphere, particularly JRA-55, although the negative EP flux divergence trend in summer is typically strongest in ERA-Interim (by around -0.2 m s^{-1} d^{-1} dec^{-1}). Considerable differences are detected when compared to MERRA and particularly CFSR, which take the form of alternating positive and negative horizontally-orientated bands in total (planetary and synoptic) wave contributions. The disagreement is most profound in the troposphere, with differences reaching ± 0.8 m s^{-1} d^{-1} dec^{-1}. In all four reanalyses the region of negative EP flux divergence descends into the upper troposphere during summer, but is less pronounced in ERA-Interim largely due to differences in the synoptic wave component.

We also found that the corresponding trend in the vertical component of EP flux (**Fig. 6.22**) is characterized by reduced planetary wave propagation from the troposphere into the stratosphere in austral spring and enhanced planetary wave propagation in austral summer. All four reanalyses demonstrate similar broad features. Nevertheless, it is apparent that ERA-Interim and JRA-55 show stronger and longer lasting upward wave propagation in austral spring compared to MERRA and CFSR. In the troposphere, the intensification of winds during summer is associated with anomalies of both vertical and horizontal (not shown) synoptic EP flux divergence anomalies. The results for CFSR, in particular, show considerable differences when compared with the other three reanalyses. The disagreement again takes the form of alternating positive and negative horizontally-orientated bands.

These banded features most likely originate from the stability parameter in the vertical component of the EP flux, which is affected by the banded structure of zonally-averaged temperature trend anomalies (**Fig. 6.20**). This may be due in part to model drift induced by radiative heating imbalance during data assimilation, rather than observational errors (*e.g.*, *Lu et al.*, 2015). Similar banded structures are observed in temperature anomalies (**Fig. 1** of *Long et al.*, 2017, see also *Chapter 3*) and may result from discontinuities in the assimilation of temperatures retrieved from satellite sensors, which are known to show vertical oscillations when compared among sensors.

The four modern reanalyses support the notion that ozone depletion leads to a strengthening of the stratospheric polar vortex and consequent downward movement of zonal mean anomalies. They broadly agree on characterising the dynamical evolution of circulation anomalies and associated wave forcing in high southern latitudes during the period of formation of the ozone hole (*Thompson et al.*, 2011; *Son et al.*, 2010, 2018). The wave driving characteristics associated with the circulation changes are in general agreement with the hypothesis examined by *Orr et al.* (2012). Noticeably large differences in EP fluxes and divergence are found in CFSR compared to the other three reanalyses datasets, which appear to be related to the aforementioned vertically alternating positive and negative anomalies in temperature.

6.9 Outlook, key findings, and recommendations

We have assessed the reanalyses' representation of large-scale coupling between the troposphere and the stratospheric polar vortices, which are present during the extended winter season (or, polar night) of each hemisphere. This coupling is chiefly effected through major Sudden Stratospheric Warming (SSW) events, which are found almost exclusively in the Northern Hemisphere. Much of our focus has thus been on the boreal extratropical atmosphere on synoptic to intraseasonal time scales (*Section 6.4*). The influence of the tropics on the Northern Hemisphere polar vortex, however, is felt through modulation of SSWs by the tropical ocean (ENSO) and stratosphere (QBO) on lower frequencies (*Section 6.7*). Large-scale coupling on synoptic to seasonal timescales in both hemispheres was assessed by comparing the annular mode indices and final warming events in *Sections 6.5* and *6.6*, respectively. Finally, anthropogenic induced ozone loss caused significant trends in the polar vortex over Antarctica, as assessed in *Section 6.8*. After summarizing the results of this chapter in this section, we conclude with a list of key findings and recommendations.

Our assessment has largely focused on the self-consistency of a given reanalysis, and the consistency between the different reanalyses, as opposed to a direct validation against measurements. The large-scale circulation cannot be easily assessed from measurements directly. Surface based observations (*e.g.*, radiosondes) generally provide a very localized (point) measurement, while satellite irradiance measurements provide indirect information about composition and temperature [2]. These measurements can of course be directly linked to the large-scale circulation, but the best way of doing so is through a reanalysis, which allows one to interpolate between localized measurements and incorporate retrieval information to infer temperature, and hence the balanced circulation.

Figures 6.23 and **6.24** provide an overview of reanalyses performance for the satellite (1979-) and pre-satellite (1958 - 78) periods respectively, based on metrics discussed in *Sections 6.4* to *6.8*. We have used the 4 point scale used by all chapters in this report. In some cases, we struggled to find entirely objective measures to provide these scores, and therefore urge the reader to consult the relevant sections of the report for a more careful analysis. Demonstrated suitable indicates that a reanalysis provides a self-consistent representation of the large-scale circulation that is very similar to other reanalyses at the same level. For the very large-scale structures (*e.g.*, planetary wave structure preceding an SSW), nearly all full-input reanalyses provide a comparable representation. As detailed in previous sections, on finer scales, and particularly at higher elevations, the more recent reanalyses become more clearly superior.

[2] The new European Space Agency Aeolus mission, launched in 2018 is an exception, designed to provide direct wind measurements.

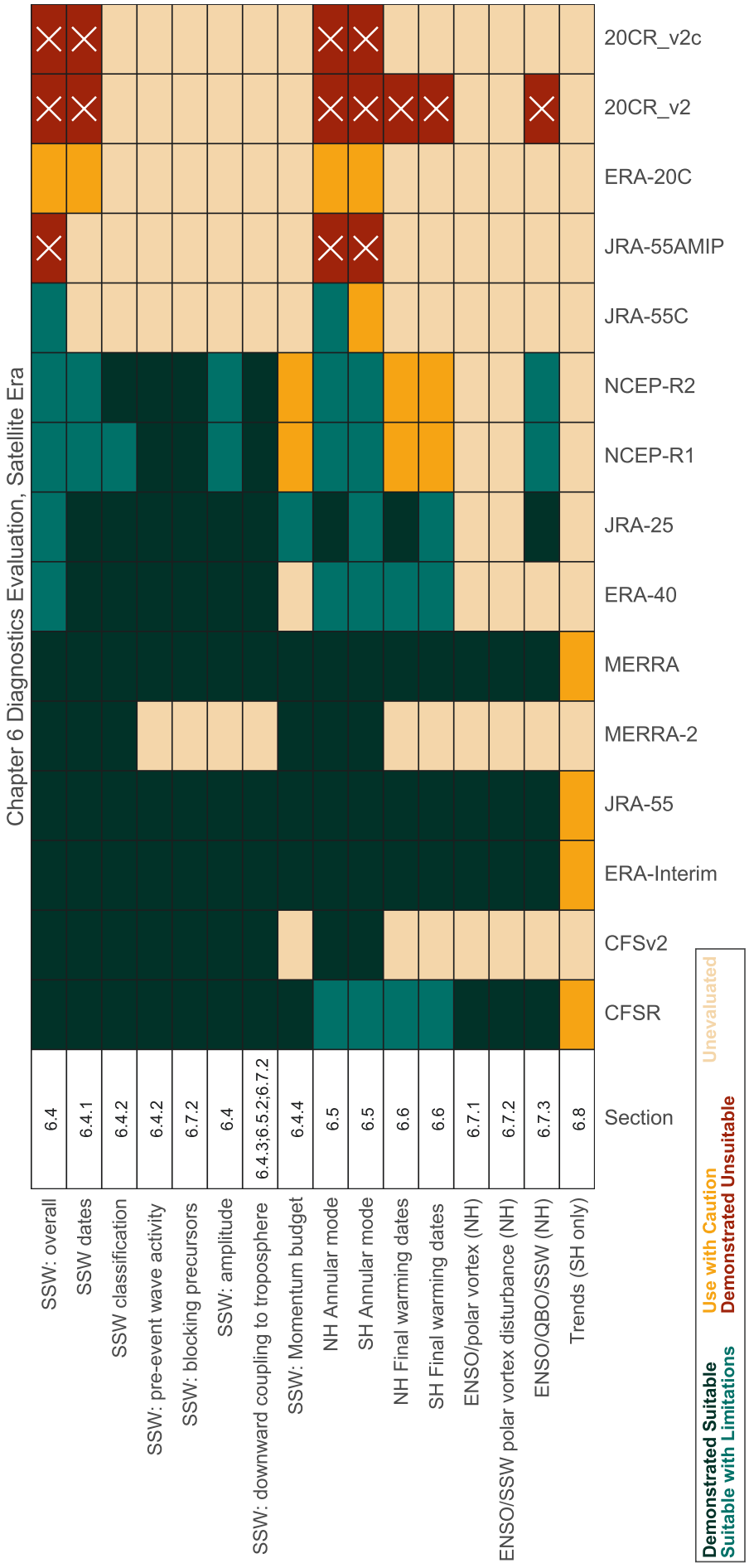

Figure 6.23: *Metric based evaluation of the reanalyses during the satellite era, 1979 onward. Please see text for further details.*

Suitable with Limitations indicates that a reanalysis provides a fairly consistent representation of circulation; conclusions from previous studies that used these reanalyses would not differ significantly if redone with a Demonstrated Suitable reanalysis. However, absent a compelling reason, these reanalyses should not be used for further research. Use with Caution has generally been applied only to the surface-input reanalyses, and the older NCEP products, which exhibit clear inconsistencies, particularly near their upper boundary at 10 hPa. Surface-input reanalyses are severely handicapped when it comes to the representation of the stratosphere, but in some cases could be used to explore variability on longer time scales. We also generally recommend to use caution when evaluating trends since reanalysis data is affected by artificial jumps caused by discontinuities in assimilated observations (*Long et al.*, 2017; *Lu et al.*, 2015; see also *Chapter 3*). Finally, as the name would imply, Demonstrated Unsuitable indicates the presence of clear problems in a reanalysis product. In particular, all of the full-input reanalyses show clear sign of divergence from basic measurements in the Southern Hemisphere before 1979. This is not to say, however, that there is no useful information in them. We also found significant biases in the mean state and variability of the polar vortex in the 20CR surface-input reanalysis, such that we do not recommend it for the purpose of investigating stratosphere-troposphere coupling.

We find that nearly all measures of large-scale coupling between the extratropical stratosphere and the troposphere are dominated by sampling uncertainty, as opposed to uncertainty in the reanalyses. As a result, conclusions based on any full (or conventional-input) reanalysis during the satellite era are generally valid. To put this more precisely, differences between the reanalyses are always smaller than the sampling uncertainty. One would not obtain results that are significantly different if you picked one reanalysis over another. The dominance of sampling uncertainty implies that our characterization of stratosphere-troposphere coupling is limited by the length of record; in a sense, we have a "small data" problem.

In the Northern Hemisphere, there is evidence that conventional observations are sufficient to constrain reanalyses from at least 1958 onward, as indicated in **Figure 6.24**. Given the dominance of sampling uncertainty, the longer record available in the boreal hemisphere is important. An additional two decades of high-quality reanalysis, as provided by JRA-55, reduces uncertainty in stratosphere-coupling processes by about 20 %. This reduction in uncertainty dwarfs the differences between the modern reanalysis over the satellite period, and makes a case for using JRA-55. We are excited that ERA5 will provide a reanalysis of the atmosphere from 1950, and it is a high priority for future work to more fully assess and compare this reanalysis.

The dominance of sampling uncertainty has implications for event based diagnostics, notably SSWs. Results based on different reanalyses may appear to diverge from one another more substantively if one does not compare the same events, i.e., use the same dates. This divergence, however, is really sampling uncertainty, aliasing into the signal.

All this said, we find that the modern reanalyses, ERA-Interim, JRA-55, MERRA 1 and 2, and to a slightly lesser extent, CFSR/CFSv2, are demonstrably superior to earlier reanalyses, providing a more dynamically consistent representation of the circulation. Over the limited period for which it is available, ERA5 also appears to be equally high quality as well. As a matter of best practice, we would urge all users to avoid earlier reanalyses unless there is specific need for them. As a practical note, modern reanalyses are available at reduced resolution. Based largely on anecdotal evidence, this appears to be a common reason why NCEP R1 is still used widely: it 's volume of data is smaller, and thus simply easier to download. Reduced resolution is appropriate for many analyses of the large-scale circulation, but it is recommended to use a modern reanalysis with a reduced resolution instead of NCEP R1. An exception is when real-time data availability is required but we note that by mid 2020, ERA5 will be provided five days behind real time.

The surface-input reanalyses are generally inferior in their representation of stratospheric variability, but may still provide research value. We do not find evidence that NOAA-20CR reanalyses accurately capture stratospheric variability; they are therefore not recommended for use. There is evidence that ERA-20C has accurate climatological variability in the stratosphere, and substantial skill in recent decades of capturing the actual variability. It is not recommended for use if restricted to periods where other reanalyses are available, but could be valuable for analysis of stratosphere- troposphere coupling on longer time scales. It should, however, be viewed as a mixture of a high quality free running model and a reanalysis, as stratospheric variability is only partially constrained by observations.

To conclude, we provide an overall, albeit more subjective, assessment of the reanalyses in **Table 6.3**. Full-input reanalyses, which make use of all available observations at a given time, have been marked recommended, consistent, or inconsistent. Recommended does not necessarily mean error-free, but indicates a self-consistent representation of the coupled variability, and consistency with other recommended reanalyses and observational constraints where available. We have marked other reanalyses consistent when differences between them and the recommended reanalyses are small relative to sampling uncertainty. Hence published results based on these reanalyses would not be significantly different if they were redone with a recommended reanalysis. A mark of inconsistent indicates that the reanalysis differs substantially with respect to other reanalysis data sets and/or available observational constraints. While "inconsistent" is meant to convey a

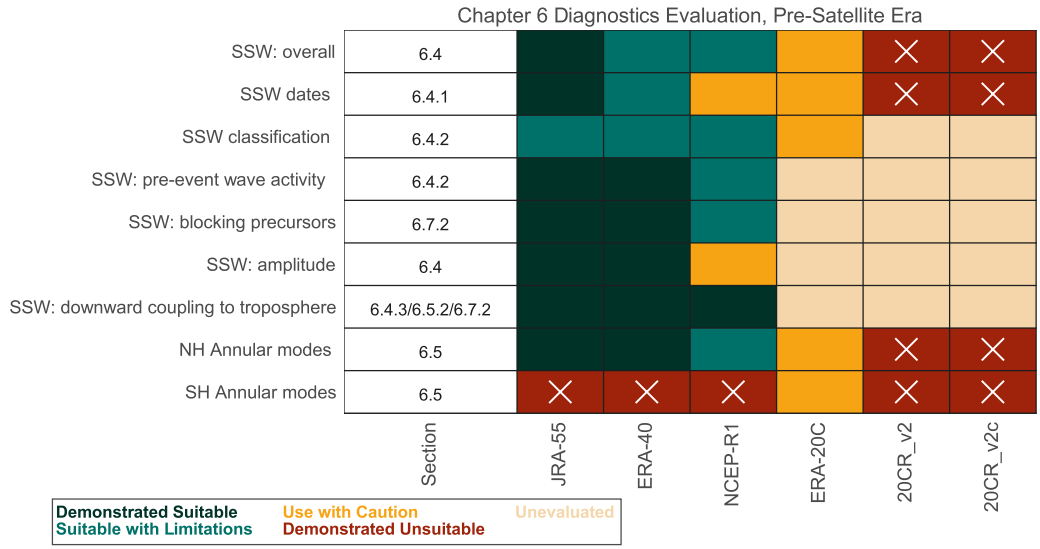

Figure 6.24: *Metric based evaluation of the reanalyses during the pre-satellite era from 1958-1978.*

clear warning, it does not imply that there is no useful information in these reanalysis products.

Given the dominance of sampling uncertainty, we may be able to glean additional confidence in stratosphere-troposphere coupling by careful use of earlier records and limited input reanalyses (*Hitchcock*, 2019). Use w/ caution has been applied to alternative reanalyses (JRA-55C and ERA- 20C), the latter of which can be used to explore variability on longer time scales. 20CR may be suitable for analysis of the troposphere, but exhibits clear biases in the variability of the stratosphere. ERA-20C, while clearly not as accurate as modern, full-input reanalysis, does appear capable of capturing information

Table 6.3: *Recommendations on the use of atmospheric reanalyses to evaluate the large-scale coupling between the stratospheric polar vortex and the tropospheric circulation on synoptic to interannual time scales. This endorsement does not include the analysis of trends, where greater caution must be employed, as discussed in Section 6.8.*

Name	post-satellite era, 1979-present		pe-satellite era, 1958-979	
	NH	SH	NH	SH
ERA-40	consistent	consistent	consistent *	inconsistent
ERA-Interim†	recommended	recommended	n.a.	n.a.
ERA-20C	use w/ caution	use w/ caution	use w/ caution	use w/ caution
JRA-25	consistent	consistent	n.a.	n.a.
JRA-55	recommended	recommended	recommended *	inconsistent
JRA-55C	consistent *	use w/ caution	n.a.	n.a.
JRA-55AMIP	inconsistent	inconsistent	inconsistent	inconsistent
MERRA	consistent	consistent	n.a.	n.a.
MERRA-2	recommended	recommended	n.a.	n.a.
NCEP-R1	consistent *	consistent *	consistent *	inconsistent
NCEP-R2	consistent *	consistent *	n.a.	n.a.
CFSR	recommended	recommended	n.a.	n.a.
CFSv2	recommended	recommended	n.a.	n.a.
20CR v2	inconsistent	inconsistent	inconsistent	inconsistent
20CR v2c	inconsistent	inconsistent	inconsistent	inconsistent

* There are few conventional observations above 10 hPa, and caution must be employed above this level (or the reanalysis itself does not extend past 10 hPa).

† ERA-Interim is being supplanted by the ERA5 reanalysis. Tentative analysis suggests thatERA5 is as good as ERA-Interim, if not better, but we do not have sufficient evidence to make a full recommendation. It will be particularly important to evaluate its performance in the Northern Hemisphere during the pre-satellite era.

about the variability of the stratosphere given only surface data. This feat alone establishes the remarkably tight coupling between the troposphere and stratosphere in our atmosphere.

Key findings

- In the satellite era (1979 - onward), the representation of large-scale stratosphere-troposphere circulation is very consistent across all full-input (including satellite observations) reanalyses. On synoptic scales, the more recent reanalyses (ERA-Interim, JRA-55, MERRA and MERRA 2, and to a slightly lesser extent, CFSR/CFSv2) become more clearly superior.

- Our ability to assess and understand stratosphere-troposphere coupling is primarily limited by sampling uncertainty, that is, by the comparatively large natural variability of the circulation relative to the length of the satellite record. As an example, various efforts have sought to characterize the break-down of the polar vortex during a Sudden Stratospheric Warmings (SSW) as a split or displacement event. Methodological differences among the classifications proposed in the literature, however, result in a partial agreement (for two-thirds of SSW events). In contrast, applying the same definition to different reanalyses yields nearly identical results.

- Although measures of stratosphere-troposphere coupling determined from earlier reanalyses are generally not statistically distinct from results obtained with a more recent reanalysis, the more recent products show demonstrable improvement, particularly with respect to internal consistency (*e.g.*, the momentum budget) and at higher levels (10 hPa and above).

- Reanalysis datasets broadly agree on temperture, wind, and wave forcing trends in the austral polar vortex related to ozone depletion from 1979 to 2001. In contrast, there are no discernible trends in Northern Hemisphere polar vortex variability over the same period.

- Pre-satellite era reanalyses (1958 - 1978) appear to be of good quality in the Northern Hemisphere, and therefore can be used to reduce sampling uncertainty in measures of stratosphere-troposphere coupling by approximately 20 %. We emphasize that this represents a more significant reduction in uncertainty than achieved by shifting from an earlier generation reanalysis to a more recent reanalysis.

- Pre-satellite era reanalyses of the Southern Hemisphere are generally of poor quality, and can only be used to reduce sampling uncertainty with great caution.

- A conventional-input (excluding satellite observations) reanalysis of the Northern Hemisphere (JRA-55C) matches full-input reanalyses well up to 10 hPa, supporting the validity of pre-satellite reanalysis products in this hemisphere. JRA-55C's representation of the Southern Hemisphere is not as accurate, suggesting that satellite measurements are more critical in this hemisphere due to the reduced density of conventional observations.

- Surface-input reanalyses have also been evaluated. ERA-20C captures not only the correct statistical climatology of the Northern Hemisphere stratospheric polar vortex, but also much of its actual variability (correctly representing the timing of about half of observed SSWs). This suggests it may be suitable for exploring low-frequency variability of the stratosphere-troposphere coupled system. The representation of the stratospheric vortex in NOAA 20CR v2/v2c, however, is demonstrably poor.

Recommendations

- We recommend the use of more recent reanalysis products. As a matter of best practice, we urge all users to avoid the use of earlier reanalyses unless the project requires the use of an older product, and special care is taken to justify that the older product is otherwise consistent with more recent reanalyses. In particular, we note for users that modern reanalyses can be obtained, in addition to their native high-resolution grids, at a coarser resolution that is comparable to that of earlier reanalyses and thus more manageable in size, but which still captures the best representation of the large-scale circulation.

- The consistency of trends associated with the Antarctic ozone hole (for the period 1979 forward) suggest that reanalyses may be reliably capturing the influence of stratospheric ozone loss. One must exercise great caution in the interpretation of trends in the reanalyses, however, as they can be spuriously caused by changes in the

observations assimilated over time, an issue that could systematically affect all products. Additional support from direct observations and/or understanding of the mechanism(s) help build confidence in trends found in the reanalyses.

- When an extended record is needed to reduce sampling uncertainty, we recommend the use of pre-satellite era reanalyses (1958 - 1978) in the Northern Hemisphere, but caution against their use in the Southern Hemisphere.

- Due to significant biases in the mean state and variability of the polar vortex in the 20CR surface-input reanalysis, we do not recommend it for the purpose of investigating stratosphere-troposphere coupling.

- ERA-20C may be suitable, with caution, for exploring the low-frequency variability of the stratosphere-troposphere coupled system.

- As our ability to quantify the large-scale coupling between the stratosphere and troposphere is primarily limited by sampling uncertainty, we recommend that future reanalysis products extend their analysis prior to the satellite era.

Code availability

Code can be provided by the authors upon request.

Data availability

The S-RIP: Zonal-mean dynamical variables of global atmospheric reanalyses on pressure levels (Martineau et al., 2018c; Martineau, 2017) is publicly available. More refined data can be provided by the authors upon request.

Acknowledgements

We thank Ilari Lehtonen for providing the classification of the split/displacement events shown in **Table A6.7**. EPG acknowledges support from the US NSF through grants AGS-1546585 and AGS-1852727. PM was a Japan Society for the Promotion of Science (JSPS) international research fellow. BA was funded by "Ayudas para la contratación de personal postdoctoral de formación en docencia e investigación en los departamentos de la UCM". BA, DB, NC, FMP and MI were supported by the European Project 603557-STRATOCLIM under program FP7-ENV.2013.6.1-2 and the Spanish Government through the PALEOSTRAT (CGL2015-69699-R) and JEDiS (RTI2018- 096402-B-I00) projects. The work of SCH was supported by the Met Office Hadley Centre Climate Programme funded by BEIS and Defra. The work by MT was supported by JSPS KAKENHI Grant Number JP24224011.

References

Andrews, D.G., J.R. Holton, and C.B. Leovy, 1987: Middle Atmosphere Dynamics. *Academic Press*, San Diego, xi + 489 pp., doi: 10.1002/qj.49711548612.

Ayarzagüena, B. and E. Serrano, 2009: Monthly Characterization of the Tropospheric Circulation over the Euro-Atlantic Area in Relation with the Timing of Stratospheric Final Warmings. *J. Climate*, **22**, 6313 - 6324, doi: 10.1175/2009JCLI2913.1.

Ayarzagüena, B., *et al.*, 2019: On the representation of major stratospheric warmings in reanalyses. *Atmos. Chem. Phys.*, **19**, 9469 - 9484, doi: 0.5194/acp-19-9469-2019.

Badin, G. and D.I. Domeisen, 2014: A Search for chaotic behavior in Northern Hemisphere stratospheric variability. *J. Atmos. Sci.*, **71**, 1494 - 1507, doi: 10.1175/JAS-D-13-0225.1.

Baldwin, M.P. and T.J. Dunkerton, 2001: Stratospheric Harbingers of Anomalous Weather Regimes. *Science*, **294**, 581 - 584, 2001, doi: 10.1126/science.1063315.

Baldwin, M.P. 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.

Baldwin, M.P., *et al.*, 2001: The Quasi-Biennial Oscillation. *Rev. Geophys.*, **39**, 179 - 229, doi: 10.1029/1999RG000073.

Baldwin, M.P., *et al.*, 2003: Stratospheric Memory and Skill of Extended-Range Weather Forecasts. *Science*, **301**, 636 - 640, doi 10.1126/science.1087143.

Bancalá, S., K. Krüger, and M. Giorgetta, 2012: The preconditioning of major sudden stratospheric warmings. *J. Geophys. Res. Atm.*, **117**, D04101, doi: 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.

Black, R.X., 2002: Stratopsheric Forcing of Surface Climate in the Arctic Oscillation. *J. Clim.*, **15**, 268 - 277, doi: 10.1175/1520-0442(2002)015<0268:SFOSCI>2.0.CO;2.

Black, R.X. and B.A. McDaniel, 2007: The Dynamics of Northern Hemisphere Stratospheric Final Warming Events. *J. Atmos. Sci.*, **64**, 2932 - 2946, doi: 10.1175/JAS3981.1.

Brönnimann, S., 2007: Impact of El Niño-Southern Oscillation on European climate. *Rev. Geophys.*, **45**, RG3003, doi: 10.1029/2006RG000199.

Butler, A.H. and L.M. Polvani, 2011: El Niño, La Niña, and stratospheric sudden warmings: A reevaluation in light of the observational record. *Geophys. Res. Lett.*, **38**, 1 - 5, doi: 10.1029/2011GL048084.

Butler, A.H., *et al.*, 2015: Defining Sudden Stratospheric Warmings. *Bull. Am. Meteorol. Soc.*, **96**, 1913 - 1928, doi: 10.1175/BAMS- D-13-00173.1.

Butler, A.H., J.P. Sjoberg, D.J. Seidel, and K.H. Rosenlof, 2017: A sudden stratospheric warming compendium. *Earth Syst. Sci. Data*, **9**, 63 - 76, doi: 10.5194/essd-9-63-2017.

Butler, A.H., *et al.*, 2019: Predictability of Northern Hemisphere Final Stratospheric Warmings and Their Surface Impacts. *Geophys. Res. Lett.*, **46**, 10 578 - 10 588, doi: 10.1029/2019GL083346.

Byrne, N.J. and T.G. Shepherd, 2018: Seasonal persistence of circulation anomalies in the Southern Hemisphere stratosphere, and its implications for the troposphere. *J. Clim.*, 3467 - 3483, doi: 10.1175/JCLI-D-17-0557.1.

Cagnazzo, C. and E. Manzini, 2009: Impact of the stratosphere on the winter tropospheric teleconnections between ENSO and the North Atlantic and European Region. *J. Clim.*, **22**, 1223 - 1238, doi: 10.1175/2008JCLI2549.1.

Cagnazzo, C., *et al.*, 2009: Northern winter stratospheric temperature and ozone responses to ENSO inferred from an ensemble of Chemistry Climate Models. *Atmos. Chem. Phys.*, **9**, 8935 - 8948, doi: 10.5194/acp-9-8935-2009.

Calvo, N., M.A. Giorgetta, R. Garcia-Herrera, and E. Manzini, 2009: Nonlinearity of the combined warm ENSO and QBO effects on the Northern Hemisphere polar vortex in MAECHAM5 simulations. *J. Geophys. Res.*, **114**, D13109, doi: 10.1029/2008JD011445.

Calvo, N., *et al.*, 2017: Northern Hemisphere Stratospheric Pathway of different El Niño flavors in stratosphere-resolving CMIP5 models. *J. Clim.*, **30**, 4351 - 4371, doi: 10.1175/JCLI-D-16-0132.1.

Castanheira, J.M. and D. Barriopedro, 2010: Dynamical connection between tropospheric blockings and stratospheric polar vortex. *Geophys. Res. Lett.*, **37**, 1 - 5, doi: 10.1029/2010GL043819.

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.

Charney, J.G. and P.G. Drazin, 1961: Propagation of Planetary-Scale Disturbances from the Lower into the Upper Atmosphere. *J. Geophys. Res.*, **66**, 83 - 109, doi: 10.1029/JZ066i001p00083.

Cohen, J. and J. Jones, 2011: Tropospheric Precursors and Stratospheric Warmings. *J. Clim.*, **24**, 6562 - 6572, doi: 10.1175/2011JCLI4160.1.

Compo, G.P., *et al.*, 2011: The Twentieth Century Reanalysis Project. *Q. J. R. Meteorol. Soc.*, **137**, 1 - 28, doi: 10.1002/qj.776.

Dee, D.P., *et al.*, 2011: The ERA-Interim reanalysis: configuration and performance of the data assimilation system, *Q. J. R. Meteorol. Soc.*, **137**, 553 - 597, doi: 10.1002/qj.828.

Domeisen, D.I., C.I. Garfinkel, and A.H. Butler, 2019a: The Teleconnection of El Niño Southern Oscillation to the Stratosphere. *Rev. Geophys.*, **57**, 5 - 47, doi: 10.1029/2018RG000596.

Domeisen, D.I.V., *et al.*, 2019b: The Role of the Stratosphere in Subseasonal to Seasonal Prediction: 2. Predictability Arising From Stratosphere-Troposphere Coupling. *J. Geophys. Res. Atmos.*, **125**, doi: 10.1029/2019JD030923.

Esler, J.G. and N.J. Matthewman, 2011: Stratospheric Sudden Warmings as Self-Tuning Resonances. Part II: Vortex Displacement Events. *J. Atmos. Sci.*, **68**, 2505 - 2523, doi: 10.1175/JAS-D-11- 08.1.

Eyring, V., *et al.*, 2006: Assessment of temperature, trace species, and ozone in chemistry-climate model simulations of the recent past. *J. Geophys. Res.*, *111*, D22 308, doi: 10.1029/2006JD007327.

Fujiwara, M., *et al.*, 2017: Introduction to the SPARC Reanalysis Intercomparison Project (S-RIP) and overview of the reanalysis systems. *Atmos. Chem. Phys.*, **17**, 1417 - 1452, doi: 10.5194/acp- 17-1417-2017.

Garfinkel, C.I. and D.L. Hartmann, 2007: Effects of the El Niño–Southern Oscillation and the Quasi- Biennial Oscillation on polar temperatures in the stratosphere. *J. Geophys. Res.*, **112**, D19112, doi: 10.1029/2007JD008481.

Garfinkel, C.I. and D.L. Hartmann, 2010: Tropospheric Precursors of Anomalous Northern Hemisphere Stratospheric Polar Vortices. *J. Clim.*, **23**, 3282 - 3299, doi: 10.1175/2010JCLI3010.1.

Garfinkel, C.I., *et al.*, 2012a: Why might stratospheric sudden warmings occur with similar frequency in El Niño and la Niña winters? *J. Geophys. Res. Atmos.*, **117**, 1 - 11, doi: 10.1029/2012JD017777.

Garfinkel, C.I., T.A. Shaw, D.L. Hartmann, and D.W. Waugh, 2012b: Does the Holton-Tan Mechanism Explain How the Quasi-Biennial Oscillation Modulates the Arctic Polar Vortex? *J. Atmos. Sci.*, **69**, 1713 - 1733, doi: 10.1175/JAS-D-11-0209.1.

Gelaro, R., *et al.*, 2017: The Modern-Era Retrospective Analysis for Research and Applications, Version 2 (MERRA-2). *J. Clim.*, **30**, 5419 - 5454, doi: 10.1175/JCLI-D-16-0758.1.

Gerber, E.P. and P. Martineau, 2018: Quantifying the variability of the annular modes: Reanalysis uncertainty vs. sampling uncertainty. *Atmos. Chem. Phys.*, **18**, 17 099 - 17 117, doi: 10.5194/acp-18-17099-2018.

Gerber, E.P., *et al.*, 2010: Stratosphere-Troposphere Coupling and Annular Mode Variability in Chemistry-Climate Models. *J. Geophys. Res.*, **115**, D00M06, doi: 10.1029/2009JD013770.

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, 2012.

Gray, L.J., *et al.*, 2018: Surface impacts of the Quasi Biennial Oscillation. *Atmos. Chem. Phys.*, **18**, 8227 - 8247, doi: 10.5194/acp-18- 8227-2018.

Hardiman, S.C., *et al.*, 2011: Improved predictability of the troposphere using stratospheric final warmings. J. *Geophys. Res. Atmos.*, **116**, 1 - 11, doi: 10.1029/2011JD015914.

Haynes, P.H., *et al.*, 1991: On the "Downward Control" of Extratropical Diabatic Circulations by Eddy-Induced Mean Zonal Forces. *J. Atmos. Sci.*, **48**, 651 - 678, doi: 10.1175/1520-0469(1991)048<0651:OTCOED>2.0.CO;2.

Hitchcock, P., 2019: On the value of reanalyses prior to 1979 for dynamical studies of stratosphere–troposphere coupling. *Atmos. Chem. Phys.*, **19**, 2749 - 2764, doi: 10.5194/acp-19-2749-2019.

Holton, J.R. and H.-C. Tan, 1980: The Influence of the Equatorial Quasi-Biennial Oscillation on the Global Circulation at 50 mb. *J. Atmos. Sci.*, **37**, 2200 - 2208, doi: 10.1175/1520-0469(1980)037<2200:TIOTEQ>2.0.CO;2.

Horel, J.D. and J.M. Wallace, 1981: Planetary-Scale Atmospheric Phenomena Associated with the Southern Oscillation, *Mon. Wea. Rev.*, **109**, 813 - 829, doi: 10.1175/1520-0493(1981)109<0813:PSAPAW>2.0.CO;2.

Huck, P.E., *et al.*, 2007: An improved measure of ozone depletion in the Antarctic stratosphere. *J. Geophys. Res.*, **112**, D11104, doi: 10.1029/2006JD007860.

Iza, M. and N. Calvo, 2015: Role of Stratospheric Sudden Warmings on the response to Central Pacific El Niño. *Geophys. Res. Lett.*, **42**, 2482 - 2489, doi: 10.1002/2014GL062935.

Iza, M., N. Calvo, and E. Manzini, 2016: The Stratospheric Pathway of La Niña. *J. Clim.*, **29**, 8899 - 8914, doi: 10.1175/JCLI-D-16-0230.1.

Kalnay, E., *et al.*, 1996: The NCEP/NCAR 40-Year Reanalysis Project. *Bull. Am. Meteorol. Soc.*, **77**, 437 - 472, doi: 10.1175/1520-0477(1996)077<0437:TNYRP>2.0.CO;2.

Kanamitsu, M., *et al.*, 2002: NCEP–DOE AMIP-II Reanalysis (R-2). *Bull. Am. Meteorol. Soc.*, **83**, 1631 - 1644, doi: 10.1175/BAMS-83-11-1631.

Karpechko, A.Y., P. Hitchcock, D.H. Peters, and A. Schneidereit, 2017: Predictability of Downward Propagation of Major Sudden Stratospheric Warmings. *Q. J. R. Meteorol. Soc.*, **60**, doi: 10.1002/QJ.3017.

Kim, J., S.-W. Son, E.P. Gerber, and H.-S. Park, 2017: Defining Sudden Stratospheric Warming in Climate Models: Accounting for Biases in Model Climatologies. *J. Clim.*, **30**, 5529 - 5546, doi: 10.1175/JCLI-D-16-0465.1.

Kobayashi, C., *et al.*, 2014: Preliminary Results of the JRA-55C, an Atmospheric Reanalysis Assimilating Conventional Observations Only. *SOLA*, **10**, 78 - 82, doi: 10.2151/sola.2014-016.

Kobayashi, S., *et al.*, 2015: The JRA-55 Reanalysis: General Specifications and Basic Characteristics. *J. Meteorol. Soc. Japan. Ser. II*, **93**, 5 - 48, doi: 10.2151/jmsj.2015-001, 2015.

Kushner, P.J., 2010: Annular modes of the troposphere and stratosphere. *Geophys. Monogr. Ser.*, **190**, 59 - 91, doi: 10.1029/2009GM000924.

Kushner, P.J. and L.M. Polvani, 2004: Stratosphere-Troposphere Coupling in a Relatively Simple AGCM: The Role of Eddies. *J. Clim.*, **17**, 629 - 639, doi: 10.1175/1520-0442(2004)017<0629:SCIARS>2.0.CO;2.

Labitzke, K., 1977: Interannual Variability of the Winter Stratosphere in the Northern Hemisphere. *Mon. Weather Rev.*, **105**, 762–770, doi: 10.1175/1520-0493(1977)105<0762:IVOTWS>2.0.CO;2.

Lehtonen, I. and A.Y. Karpechko, 2016: Observed and modeled tropospheric cold anomalies associated with sudden stratospheric warmings. *J. Geophys. Res. Atmos.*, **121**, 1591 - 1610, doi: 10.1002/2015JD023860.

Lim, E., H.H. Hendon, and D.W.J. Thompson, 2018: Seasonal Evolution of Stratosphere-Troposphere Coupling in the Southern Hemisphere and Implications for the Predictability of Surface Climate. *J. Geophys. Res. Atmos.*, **123**, 12,002 - 12,016, doi: 10.1029/2018JD029321.

Limpasuvan, V., D.W.J. Thompson, and D.L. Hartmann, 2004: The Life Cycle of the Northern Hemisphere Sudden Stratospheric Warmings. *J. Clim.*, **17**, 2584 - 2596, doi: 10.1175/1520-0442(2004)017<2584:TLCOTN>2.0.CO;2.

Long, C.S., *et al.*, 2017: Climatology and interannual variability of dynamic variables in multiple reanalyses evaluated by the SPARC Reanalysis Intercomparison Project (S-RIP). *Atmos. Chem. Phys.*, **17**, 14593 - 14629, doi: 10.5194/acp-17-14593-2017.

Lu, H., T.J. Bracegirdle, T. Phillips, and J. Turner, 2015: A Comparative Study of Wave Forcing Derived from the ERA-40 and ERA-Interim Reanalysis Datasets. *J. Clim.*, **28**, 2291 - 2311.

Marshall, G.J., 2003: Trends in the Southern Annular Mode from Observations and Reanalyses. *J. Climate*, **16**, 4134 - 4143, doi: 10.1175/1520-0442(2003)016<4134:TITSAM>2.0.CO;2.

Martineau, P., 2017: S-RIP: Zonal-mean dynamical variables of global atmospheric reanalyses on pressure levels. *Centre for Environmental Data Analysis*, doi: 10.5285/b241a7f536a244749662360bd7839312.

Martineau, P. and S.-W. Son, 2010: Quality of reanalysis data during stratospheric vortex weakening and intensification events. *Geophys. Res. Lett.*, **37**, doi: 10.1029/2010GL045237.

Martineau, P. and S.-W. Son, 2015: Onset of Circulation Anomalies during Stratospheric Vortex Weakening Events: The Role of Planetary-Scale Waves. *J. Clim.*, **28**, 7347 - 7370, doi: 10.1175/JCLI- D-14-00478.1.

Martineau, P., S.-W. Son, and M. Taguchi, 2016: Dynamical Consistency of Reanalysis Datasets in the Extratropical Stratosphere. *J. Clim.*, **29**, 3057 - 3074, doi: 10.1175/JCLI-D-15-0469.1.

Martineau, P., G. Chen, S.-W. Son, and J. Kim, 2018a: Lower-Stratospheric Control of the Frequency of Sudden Stratospheric Warming Events. *J. Geophys. Res. Atmos.*, **123**, 3051 - 3070, doi: 10.1002/2017JD027648.

Martineau, P., S.-W. Son, M. Taguchi, and A.H. Butler, 2018b: A comparison of the momentum budget in reanalysis datasets during sudden stratospheric warming events. *Atmos. Chem. Phys.*, **18**, 7169 - 7187, doi: 10.5194/acp-18-7169-2018.

Martineau, P., J.S. Wright, N. Zhu, and M. Fujiwara, 2018c: Zonal-mean data set of global atmospheric reanalyses on pressure levels. *Earth Syst. Sci. Data*, **10**, 1925 - 1941, doi: 10.5194/essd-10-1925- 2018.

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

Matsuno, T., 1971: A Dynamical Model of the Stratospheric Sudden Warming. *J. Atmos. Sci.*, **28**, 1479 - 1491, doi: 10.1175/1520-0469(1971)028<1479:ADMOTS>2.0.CO;2.

Matthewman, N.J. and J.G. Esler, 2011: Stratospheric Sudden Warmings as Self-Tuning Resonances. Part I: Vortex Splitting Events. *J. Atmos. Sci.*, **68**, 2481 - 2504, doi: 10.1175/JAS-D-11-07.1.

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. Clim.*, **22**, 1566 - 1585, doi: 10.1175/2008JCLI2365.1, 2009.

Maycock, A.C. and P. Hitchcock, 2015: Do split and displacement sudden stratospheric warmings have different annular mode signatures? *Geophys. Res. Lett.*, **42**, 10,943 - 10,951, doi: 10.1002/2015GL066754.

McDaniel, B.A. and R.X. Black, 2005: Intraseasonal Dynamical Evolution of the Northern Annular Mode, *J. Clim.*, **18**, 3820 - 3839, doi: 10.1175/JCLI3467.1.

McInturff, R.M., 1978: Stratospheric warmings: Synoptic, dynamic and general-circulation aspects. *NASA Reference Publ.* **NASA-RP-1017**, p. 174, available at **https://ntrs.nasa.gov/archive/nasa/casi.nt rs.nasa.gov/19780010687.pdf**.

McIntyre, E., 1982: How well do we Understand the Dynamics of Stratospheric Warmings? *J. Meteorol. Soc. Japan*, **60**, 37–65, doi: 10.2151/jmsj1965.60.1_37.

McIntyre, M.E. and T.N. Palmer, 1983: Breaking planetary waves in the stratosphere. *Nature*, **305**, 593 - 600, doi: 10.1038/305593a0, 1983.

McIntyre, M. and T. Palmer, 1984: The 'surf zone' in the stratosphere. *J. Atmos. Terr. Phys.*, **46**, 825 - 849, doi: 10.1016/0021-9169(84)90063-1.

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, 2011, doi: 10.1175/2010JAS3555.1.

Mitchell, D.M., *et al.*, 2013: The Influence of Stratospheric Vortex Displacements and Splits on Surface Climate. *J. Climate*, **26**, 2668 - 2682, doi: 10.1175/JCLI-D-12-00030.1.

Mitchell, D.M., *et al.*, 2015: Signatures of naturally induced variability in the atmosphere using multiple reanalysis datasets. *Q. J. R. Meteorol. Soc.*, **141**, 2011 - 2031, doi: 10.1002/qj.2492.

Nishii, K., H. Nakamura, and T. Miyasaka, 2009: Modulations in the planetary wave field induced by upward-propagating Rossby wave packets prior to stratospheric sudden warming events: A case-study. *Quart. J. Roy. Meteor. Soc.*, **135**, 39 - 52, doi: 10.1002/qj.359.

Nishii, K., H. Nakamura, and Y.J. Orsolini, 2011: Geographical Dependence Observed in Blocking High Influence on the Stratospheric Variability through Enhancement and Suppression of Upward Planetary-Wave Propagation. *J. Clim.*, **24**, 6408 - 6423, doi: 10.1175/JCLI-D-10-05021.1.

Omrani, N.E., N.S. Keenlyside, J. Bader, and E. Manzini, 2014: Stratosphere key for wintertime atmospheric response to warm Atlantic decadal conditions. *Clim. Dyn.*, **42**, 649 - 663, doi: 10.1007/s00382-013-1860-3.

Onogi, K., *et al.*, 2007: The JRA-25 Reanalysis. *J. Meteor. Soc. Japan*, **85**, 369 - 432, doi: 10.2151/jmsj.85.369.

Orr, A., *et al.*, 2012: Possible Dynamical Mechanisms for Southern Hemisphere Climate Change due to the Ozone Hole. *J. Atmos. Sci.*, **69**, 2917 - 2932, doi: 10.1175/JAS-D-11-0210.1.

Orr, A., *et al.*, 2021: Is our dynamical understanding of the circulation changes associated with the Antarctic ozone hole sensitive to the choice of reanalysis dataset? *Atmos. Chem. Phys.*, **21**, 7451 - 7472, doi: 10.5194/acp-21-7451-2021.

Palmeiro, F.M., D. Barriopedro, D., R. García-Herrera, and N. Calvo, 2015: Comparing Sudden Stratospheric Warming Definitions in Reanalysis Data. *J. Clim.*, **28**, 6823 - 6840, doi: 10.1175/JCLI- D-15-0004.1.

Plumb, R.A., 1989: On the seasonal cycle of stratospheric planetary waves. *Pure Appl. Geophys.* **130**, 233 - 242, doi: 10.1007/BF00874457.

Plumb, R.A., 2010: Planetary Waves and the Extratropical Winter Stratosphere. In *The Stratosphere: Dynamics, Transport, and Chemistry* (eds. L.M. Polvani, A.H. Sobel and D.W. Waugh). doi: 10.1002/9781118666630.ch2.

Poli, P., *et al.*, 2016: ERA-20C: An atmospheric reanalysis of the twentieth century. *J. Clim.*, **29**, 4083 - 4097, doi: 10.1175/JCLI-D-15-0556.1.

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. Clim.*, **17**, 3548 - 3554, doi: 10.1175/1520-0442(2004)017<3548:UWAFAA>2.0.CO;2.

Polvani, L.M., *et al.*, 2017: Distinguishing stratospheric sudden warmings from ENSO as key drivers of wintertime climate variability over the North Atlantic and Eurasia. *J. Clim.*, **30**, 1959 - 1969 doi: 10.1175/JCLI-D-16-0277.1.

Ramaswamy, V., *et al.*, 2001: Stratospheric temperature trends: Observations and model simulations. *Rev. Geophys.*, **39**, 71 - 122, doi: 10.1029/1999RG000065.

Randel, W.J., 1988: The seasonal evolution of planetary waves in the Southern Hemisphere stratosphere and troposphere. *Q. J. R. Meteorol. Soc.*, **587**, 1385 - 1409, doi: 10.1002/qj.49711448403.

Rao, J., C.I. Garfinkel, and R. Ren, 2019: Modulation of the Northern Winter Stratospheric El Niño–Southern Oscillation Teleconnection by the PDO. *J. Clim.*, **32**, 5761 - 5783, doi: 10.1175/JCLI-D-19-0087.1.

Richter, J.H., K. Matthes, N. Calvo, and L.J. Gray, 2011: Influence of the quasi-biennial oscillation and El Niño–Southern Oscillation on the frequency of sudden stratospheric warmings. *J. Geophys. Res.*, **116**, D20111, doi: 10.1029/2011JD015757.

Rienecker, M.M., *et al.*, 2011: MERRA: NASA's modern-era retrospective analysis for research and applications. *J. Clim.*, **24**, 3624 - 3648, doi: 10.1175/JCLI-D-11-00015.1.

Saha, S., *et al.*, 2010: The NCEP Climate Forecast System Reanalysis. *Bull. Am. Meteorol. Soc.*, **91**, 1015 - 1057, doi: 10.1175/2010BAMS3001.1.

Saha, S., *et al.*, 2014: The NCEP Climate Forecast System Version 2. *J. Climate*, **27**, 2185 - 2208, doi: 10.1175/JCLI-D-12-00823.1.

Scherhag, R., 1952: Die Explosionsartige Stratosphärenerwärmung des Spätwinters 1951/52. *Ber. Deut. Wetterdienst* **38**, 51 - 63.

Seviour, W.J.M., 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.

Sigmond, M., J.F. Scinocca, V.V. Kharin, and T.G. Shepherd, 2013: Enhanced seasonal forecast skill following stratospheric sudden warmings. *Nat. Geosci.*, **6**, 1 - 5, doi: 10.1038/ngeo1698.

Smith, K.L. and P.J. Kushner, 2012: Linear interference and the initiation of extratropical stratosphere- troposphere interactions. *J. Geophys. Res.*, **117**, doi: 10.1029/2012JD017587.

Son, S.-W., *et al.*, 2010: The Impact of Stratospheric Ozone on Southern Hemisphere Circulation Change: A Multimodel Assessment. *J. Geophys. Res.*, **115**, D00M07, doi: 10.1029/2010JD014271.

Son, S.-W., *et al.*, 2018: Tropospheric jet response to Antarctic ozone depletion: An update with Chemistry-Climate Model Initiative (CCMI) models. *Environ. Res. Lett.*, **13**, 054024, doi: 10.1088/1748-9326/aabf21.

Song, K. and S.-W. Son, 2018: Revisiting the ENSO-SSW relationship. *J. Clim.*, **31**, 2133 - 2143, doi: 10.1175/JCLI-D-17-0078.1.

Song, Y. and W.A. Robinson, 2004: Dynamical Mechanisms for Stratospheric Influences on the Troposphere. *J. Atmos. Sci.*, **61**, 1711 - 1725, doi: 10.1175/1520-0469(2004)061<1711:DMFSIO>2.0.CO;2.

Steiner, A.K. *et al.*, 2020: Consistency and structural uncertainty of multi-mission GPS radio occultation records. *Atmos. Meas. Tech.* **13**, 2547 - 2575, doi: 10.5194/amt-13-2547-2020.

Sun, L. and W.A. Robinson, 2009: Downward influence of stratospheric final warming events in an idealized model. *Geophys. Res. Lett.*, **36**, L03819, doi: 10.1029/2008GL036624.

Taguchi, M., 2015: Changes in Frequency of Major Stratospheric Sudden Warmings with El Niño/Southern Oscillation and Quasi-Biennial Oscillation. *J. Meteorol. Soc. Japan. Ser. II*, **93**, 99 - 115, doi: 10.2151/jmsj.2015-007.

Taguchi, M. and D.L. Hartmann, 2006: Increased Occurrence of Stratospheric Sudden Warmings during El Niño as Simulated by WACCM, *J. Clim.*, **19**, 324 - 332, doi: 10.1175/JCLI3655.1.

Thiéblemont, R., *et al.*, 2019: Drivers and Surface Signal of Interannual Variability of Boreal Stratospheric Final Warmings. *J. Geophys. Res. Atmos.*, **124**, 5400 - 5417, doi: 10.1029/2018JD029852.

Thompson, D.W.J. and J.M. Wallace, 1998: The Arctic Oscillation signature in the wintertime geopotential height and temperature fields. *Geophys. Res. Lett.*, **25**, 1297 - 1300, doi: 10.1029/98GL00950.

Thompson, D.W.J. and J.M. Wallace, 2000: Annular modes in the extratropical circulation. Part I: Month-to-month variability. *J. Climate*, **13**, 1000 - 1016, doi: 10.1175/1520-0442(2000)013<1000:AMITEC>2.0.CO;2.

Thompson, D.W.J. and S. Solomon, 2002: Interpretation of recent Southern Hemisphere Climate Change. *Science*, **296**, 895 - 899, doi: 10.1126/science.1069270.

Thompson, D.W.J., et al., 2011: Signatures of the Antarctic ozone hole in Southern Hemisphere surface climate change. *Nat. Geosci.*, **4**, 741 - 749, doi: 10.1038/ngeo1296.

Tripathi, O.P., A. Charlton-Perez, M. Sigmond, and F. Vitart, 2015: Enhanced long-range forecast skill in boreal winter following stratospheric strong vortex conditions. *Environ. Res. Lett.*, **10**, 104007, doi: 10.1088/1748-9326/10/10/104007.

Uppala, S.M., *et al.*, 2005: The ERA-40 re-analysis. *Q. J. R. Meteorol. Soc.*, **131**, 2961 - 3012, doi: 10.1256/qj.04.176.

Waugh, D.W., 1997: Elliptical diagnostics of stratospheric polar vortices. *Q. J. R. Meteorol. Soc.*, **123**, 1725 - 1748, doi: 10.1002/qj.49712354213.

Waugh, D.W. and Randel, W.J., 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.

Waugh, D.W., A.H. Sobel, and L.M. Polvani, 2017: What Is the Polar Vortex and How Does It Influence Weather? *Bull. Am. Meteorol. Soc.*, **98**, 37 - 44, doi: 10.1175/BAMS-D-15-00212.1.

Weinberger, I., C.I. Garfinkel, I.P. White, and L.D. Oman, 2019: The salience of nonlinearities in the boreal winter response to ENSO: Arctic stratosphere and Europe. *Clim. Dyn.*, **53**, 4591 - 4610, doi: 10.1007/s00382-019-04805-1.

White, I., *et al.*, 2019: The downward influence of sudden stratospheric warmings: Association with tropospheric precursors. *J. Clim.*, **32**, 85 - 108, doi: 10.1175/JCLI-D-18-0053.1.

Woo, S.-H., M.-K. Sung, S.-W. Son, and J.-S. Kug, 2015: Connection between weak stratospheric vortex events and the Pacific Decadal Oscillation. *Clim. Dyn.*, **45**, 3481 - 3492, doi: 10.1007/s00382- 015-2551-z.

Woollings, T., A. Hannachi, B. Hoskins, and A. Turner, 2010: A regime view of the North Atlantic oscillation and its response to anthropogenic forcing. *J. Clim.*, **23**, 1291 - 1307, doi: 10.1175/2009JCLI3087.1.

Woollings, T., *et al.*, 2018: Blocking and its Response to Climate Change. *Curr. Clim. Chang. Reports*, **4**, 287 - 300, doi: 10.1007/s40641-018-0108-z.

Yoden, S., M. Taguchi, and Y. Naito, 2002: Numerical Studies on Time Variations of the Troposphere- Stratosphere Coupled System. *J. Meteorol. Soc. Japan*, **80**, 811 - 830, doi: 10.2151/jmsj.80.811.

Appendix A: Detection and classification of major SSW events

The onset dates of SSW events identified independently for each reanalysis data sets are listed in **Table A6.4**. Then, for the common dates whose identification is described in *Section 6.2*, events are classified as to whether they are splits or displacements according a method adapted from *Seviour et al.* (2013) (**Table A6.5**), the Shibata method (**Table A6.6**) and the method of *Lehtonen and Karpechko* (2016) (**Table A6.7**). These methods are described in more detail in *Section 6.4.2*.

Table A6.4: *Identification of major SSW events in reanalyses.The criterion for the detection is a reversal of zonal-mean zonal wind at 60° N and 10 hPa (see Section 6.4.1 for more details). Cases where the reanalysis deviates from the "common" events are highlighted in bold. Events that do not show a positive meridional temperature gradient at the same level within 5 days of the zonal wind reversal are highlighted in green.*

common	NCEP R1	CFSR	ERA-40	ERA-Interim	JRA-25	JRA-55	MERRA	MERRA-2	NCEP-R2
30-Jan-58	30-Jan-58		**31-Jan-58**			30-Jan-58			
—	**30-Nov-58**		—			—			
17-Jan-60	**16-Jan-60**		17-Jan-60			17-Jan-60			
29-Jan-63	****		**28-Jan-63**			**30-Jan-63**			
—	**23-Mar-65**		—			—			
17-Dec-65	**8-Dec-65**		**16-Dec-65**			**18-Dec-65**			
23-Feb-66	**24-Feb-66**		23-Feb-66			23-Feb-66			
7-Jan-68	****		7-Jan-68			7-Jan-68			
28-Nov-68	**27-Nov-68**		28-Nov-68			**29-Nov-68**			
13-Mar-69	13-Mar-69		13-Mar-69			****			
2-Jan-70	2-Jan-70		2-Jan-70			2-Jan-70			
18-Jan-71	17-Jan-71		18-Jan-71			18-Jan-71			
20-Mar-71	20-Mar-71		20-Mar-71			20-Mar-71			
31-Jan-73	**2-Feb-73**		31-Jan-73			31-Jan-73			
9-Jan-77	****		9-Jan-77			9-Jan-77			
22-Feb-79	22-Feb-79	22-Feb-79	22-Feb-79	22-Feb-79	22-Feb-79	22-Feb-79	22-Feb-79		22-Feb-79
29-Feb-80	29-Feb-80	29-Feb-80	29-Feb-80	29-Feb-80	29-Feb-80	29-Feb-80	29-Feb-80	29-Feb-80	29-Feb-80
—	—	—	—	—	**6-Feb-81**	**6-Feb-81**	—	—	—
4-Mar-81	****	**3-Mar-81**	4-Mar-81	4-Mar-81	4-Mar-81	4-Mar-81	4-Mar-81	****	****
4-Dec-81	4-Dec-81	4-Dec-81	4-Dec-81	4-Dec-81	4-Dec-81	4-Dec-81	4-Dec-81	4-Dec-81	4-Dec-81
24-Feb-84	24-Feb-84	24-Feb-84	24-Feb-84	24-Feb-84	24-Feb-84	24-Feb-84	24-Feb-84	24-Feb-84	24-Feb-84
1-Jan-85	**2-Jan-85**	1-Jan-85	1-Jan-85	1-Jan-85	1-Jan-85	1-Jan-85	1-Jan-85	1-Jan-85	**31-Dec-84**
23-Jan-87	23-Jan-87	23-Jan-87	23-Jan-87	23-Jan-87	23-Jan-87	23-Jan-87	23-Jan-87	23-Jan-87	23-Jan-87
8-Dec-87	8-Dec-87	8-Dec-87	8-Dec-87	8-Dec-87	8-Dec-87	8-Dec-87	8-Dec-87	8-Dec-87	8-Dec-87
14-Mar-88	14-Mar-88	14-Mar-88	14-Mar-88	14-Mar-88	14-Mar-88	14-Mar-88	14-Mar-88	14-Mar-88	14-Mar-88
21-Feb-89	**22-Feb-89**	21-Feb-89	21-Feb-89	21-Feb-89	21-Feb-89	21-Feb-89	21-Feb-89	21-Feb-89	**22-Feb-89**
—	—	—	—	—	—	—	—	—	**5-Feb-95**
15-Dec-98	15-Dec-98	15-Dec-98	15-Dec-98	15-Dec-98	15-Dec-98	15-Dec-98	15-Dec-98	15-Dec-98	15-Dec-98
26-Feb-99	**25-Feb-99**	26-Feb-99	26-Feb-99	26-Feb-99	26-Feb-99	26-Feb-99	26-Feb-99	26-Feb-99	26-Feb-99
20-Mar-00	20-Mar-00	20-Mar-00	20-Mar-00	20-Mar-00	20-Mar-00	20-Mar-00	20-Mar-00	20-Mar-00	20-Mar-00
11-Feb-01	11-Feb-01	11-Feb-01	11-Feb-01	11-Feb-01	11-Feb-01	11-Feb-01	11-Feb-01	11-Feb-01	**12-Feb-01**
31-Dec-01	**2-Jan-02**	**30-Dec-01**	31-Dec-01	**30-Dec-01**	31-Dec-01	31-Dec-01	**30-Dec-01**	**30-Dec-01**	**1-Jan-02**
—	—	**17-Feb-02**	**18-Feb-02**	—	—	—	—	**17-Feb-02**	—
18-Jan-03	18-Jan-03	18-Jan-03		18-Jan-03	18-Jan-03	18-Jan-03	18-Jan-03	18-Jan-03	18-Jan-03
5-Jan-04	**7-Jan-04**	5-Jan-04		5-Jan-04	**6-Jan-04**	5-Jan-04	5-Jan-04	5-Jan-04	**6-Jan-04**
21-Jan-06	21-Jan-06	21-Jan-06		21-Jan-06	21-Jan-06	21-Jan-06	21-Jan-06	21-Jan-06	21-Jan-06
24-Feb-07	24-Feb-07	24-Feb-07		24-Feb-07	24-Feb-07	24-Feb-07	24-Feb-07	24-Feb-07	24-Feb-07
22-Feb-08	22-Feb-08	22-Feb-08		22-Feb-08	22-Feb-08	22-Feb-08	22-Feb-08	22-Feb-08	22-Feb-08
24-Jan-09	24-Jan-09	24-Jan-09		24-Jan-09	24-Jan-09	24-Jan-09	24-Jan-09	24-Jan-09	24-Jan-09
9-Feb-10	9-Feb-10	9-Feb-10		9-Feb-10	9-Feb-10	9-Feb-10	9-Feb-10	9-Feb-10	9-Feb-10
24-Mar-10	24-Mar-10	24-Mar-10		24-Mar-10	24-Mar-10	24-Mar-10	24-Mar-10	24-Mar-10	24-Mar-10

Table A6.5: *Classification of SSW events into splits and displacements adapted from the method described in Seviour et al. (2013). D and S denote displacement and split events, respectively. U denotes unclassifiable events. Bold text highlights disagreement from the "common" classification. Asterisks indicate that there was substantial disagreement on the classification of the 15-Dec-98, 20-Mar-00, and 09-Feb-10 events.*

Shared Dates	common	NCEP R1	CFSR	ERA-40	ERA-Interim	JRA-25	JRA-55	MERRA	MERRA-2	NCEP-R2
30-Jan-58	D	D		D			D			
17-Jan-60	S	S		**D**			S			
29-Jan-63	S	S		S			S			
17-Dec-65	D	D		D			D			
23-Feb-66	D	D		D			**U**			
07-Jan-68	S	S		S			S			
28-Nov-68	D	D		D			D			
13-Mar-69	S	**U**		S			S			
02-Jan-70	S	S		S			S			
18-Jan-71	S	S		S			S			
20-Mar-71	D	D		D			D			
31-Jan-73	S	S		S			S			
09-Jan-77	S	S		S			S			
22-Feb-79	S	S	S	S	S	S	S	S		S
29-Feb-80	D	D	D	D	D	D	D	D	D	D
04-Mar-81	D	D	D	D	D	D	D	D	D	D
04-Dec-81	U	U	U	U	U	U	U	U	U	U
24-Feb-84	D	D	D	D	D	D	D	D	D	D
01-Jan-85	S	S	S	S	S	S	S	S	S	S
23-Jan-87	D	D	D	D	D	D	D	D	D	D
08-Dec-87	S	S	S	S	S	S	S	S	S	S
14-Mar-88	S	S	S	S	S	S	S	S	S	S
21-Feb-89	S	**D**	S	S	S	S	S	S	S	S
15-Dec-98	D*	D	**U**	D	D	**U**	D	D	**U**	D
26-Feb-99	S	S	S	S	S	S	S	S	S	S
20-Mar-00	U*	U	**D**	U	**D**	U	**D**	U	U	U
11-Feb-01	S	S	S	S	S	S	S	S	S	S
31-Dec-01	S	S	S	S	S	S	S	S	S	S
18-Jan-03	S	S	S		S	S	S	S	S	S
05-Jan-04	D	D	D		D	D	D	D	D	D
21-Jan-06	D	D	D		D	D	D	D	D	D
24-Feb-07	D	D	D		D	D	D	D	D	D
22-Feb-08	D	D	D		D	D	D	D	D	D
24-Jan-09	S	S	S		S	S	S	S	S	S
09-Feb-10	U*	**S**	U		**D**	**D**	**S**	U	**D**	**S**
24-Mar-10	D	D	D		D	D	D	D	D	D

Table A6.6: *Classification of major SSW events into splits and displacements using the Shibata technique (Ayarzaguena et al., 2019). S and D denote split and displacement events, respectively.*

Shared Dates	common	NCEP R1	CFSR	ERA-40	ERA-Interim	JRA-25	JRA-55	MERRA	MERRA-2	NCEP-R2
30-Jan-58	S	S		S			S			
17-Jan-60	D	D		D			**S**			
29-Jan-63	S	S		S			S			
17-Dec-65	D	**S**		D			D			
23-Feb-66	D	D		D			D			
07-Jan-68	S	S		S			S			
28-Nov-68	D	D		D			D			
13-Mar-69	D	D		D			D			
02-Jan-70	D	D		D			D			
18-Jan-71	S	S		S			S			
20-Mar-71	D	D		D			D			
31-Jan-73	S	S		S			S			
09-Jan-77	D	D		D			D			
22-Feb-79	S	S	S	S	S	S	S	S		S
29-Feb-80	D	D	D	D	D	D	D	D	D	D
04-Mar-81	D	D	D	D	D	D	D	D	D	D
04-Dec-81	D	D	D	D	D	D	D	D	D	**S**
24-Feb-84	D	D	D	D	D	D	D	D	D	D
01-Jan-85	S	S	S	**D**	S	**D**	S	S	S	S
23-Jan-87	D	D	D	D	D	D	D	D	D	D
08-Dec-87	S	**D**	S	S	**D**	S	S	S	S	S
14-Mar-88	S	S	S	S	S	S	S	S	S	S
21-Feb-89	S	S	S	S	S	S	S	S	S	S
15-Dec-98	S	S	S	S	S	S	S	S	S	S
26-Feb-99	S	S	S	S	S	S	S	S	S	S
20-Mar-00	D	D	D	D	D	D	D	D	**S**	D
11-Feb-01	D	D	D	D	D	D	D	D	**S**	D
31-Dec-01	D	D	D	D	D	D	D	D	D	D
18-Jan-03	S	S	S		S	S	S	S	S	**D**
05-Jan-04	D	D	D		D	D	D	D	D	D
21-Jan-06	D	D	D		D	D	D	D	D	D
24-Feb-07	D	D	D		D	D	D	D	D	D
22-Feb-08	D	D	D		D	D	D	D	D	D
24-Jan-09	S	S	S		S	S	S	S	S	S
09-Feb-10	S	S	S		S	S	S	S	S	S
24-Mar-10	D	D	D		D	D	D	D	D	D

Table A6.7: *Classification of major SSW events into splits and displacements using the Lehtonen and Karpechko (2016) method. S and D denote split and displacement events, respectively.*

Shared Dates	common	NCEP R1	CFSR	ERA-40	ERA-Interim	JRA-25	JRA-55	MERRA	MERRA-2	NCEP-R2
30-Jan-58	S	S		S			S			
17-Jan-60	D	D		D			**S**			
29-Jan-63	D	D		**S**			D			
17-Dec-65	D	D		D			D			
23-Feb-66	S	S		S			S			
7-Jan-68	S	S		S			S			
28-Nov-68	D	D		D			D			
13-Mar-69	D	D		D			D			
2-Jan-70	D	D		D			**S**			
18-Jan-71	S	S		S			S			
20-Mar-71	D	D		D			D			
31-Jan-73	S	S		S			S			
9-Jan-77	S	S		S			S			
22-Feb-79	S	S	S	S	S	S	S	S		S
29-Feb-80	D	D	D	D	D	D	D	D	D	D
4-Mar-81	D	D	D	D	D	D	D	D	D	D
4-Dec-81	D	D	D	D	D	D	D	D	D	D
24-Feb-84	D	D	D	D	D	D	D	D	D	D
1-Jan-85	S	S	S	S	S	S	S	S	S	S
23-Jan-87	D	D	D	D	D	D	D	D	D	D
8-Dec-87	S	S	S	S	S	S	S	S	S	S
14-Mar-88	S	S	S	S	S	S	S	S	S	S
21-Feb-89	S	S	S	S	S	S	S	S	S	S
15-Dec-98	D	D	D	D	D	D	D	D	D	D
26-Feb-99	S	S	S	S	S	S	S	S	S	S
20-Mar-00	D	D	D	D	D	D	D	D	D	D
11-Feb-01	S	S	**D**	S	**D**	S	S	S	S	S
31-Dec-01	D	D	D	D	D	D	D	D	D	D
18-Jan-03	S	S	S	S	S	S	S	S	S	**D**
5-Jan-04	D	D	D		D	D	D	D	D	D
21-Jan-06	D	D	D		D	D	D	D	D	D
24-Feb-07	D	D	D		D	D	D	D	D	D
22-Feb-08	D	D	D		D	D	D	D	D	D
24-Jan-09	S	S	S		S	S	S	S	S	S
9-Feb-10	S	S	S		S	S	S	S	S	S
24-Mar-10	D	D	D		D	D	D	D	D	D

Major abbreviations and terms

20CR v2/v2c	20th Century Reanalysis of NOAA and CIRES
AMIP	Atmospheric Model Intercomparison Project
CESM	Community Earth System Model
CFSR	Climate Forecast System Reanalysis of the NCEP
CFSv2	Climate Forecast System version 2
DOE	Department of Energy
ECHAM	ECMWF-HAMburg model
EMAC	ECHAM/MESSy Atmospheric Chemistry model
ECMWF	European Centre for Medium-Range Weather Forecasts
ENSO	El Niño-Southern Oscillation
EP (Flux)	Eliassen-Palm Flux
ERA-20C	ECMWF 20th century reanalysis
ERA-40	ECMWF 40-year reanalysis
ERA-Interim	ECMWF interim reanalysis
HF	Heat flux
JRA-55	Japanese 55-year Reanalysis
JRA-55C	Japanese 55-year Reanalysis assimilating Conventional observations only
JSPS	Japan Society for the Promotion of Science
MERRA	Modern Era Retrospective-Analysis for Research and Applications
MERRA-2	Modern Era Retrospective-Analysis for Research and Applications, Version 2
MESSy	Modular Earth Submodel System
MSLP	mean sea-level pressure
NAM	Northern Annular Mode
NAO	North Atlantic Oscillation
NCAR	National Center for Atmospheric Research
NCEP	National Centers for Environmental Prediction of the NOAA
NCEP-CPC	National Centers for Environmental Prediction, Climate Prediction Cente
NCEP R2	Reanalysis 2 of the NCEP and DOE
NCEP R1	Reanalysis 1 of the NCEP and NCAR
NDJF	*November-December-January-February*
NH	Northern Hemisphere
NOAA	National Oceanic and Atmospheric Administration
PNJ	polar night jet
QBO	Quasi-biennial Oscillation
REM	Reanalysis ensemble mean
SAM	Southern Annular Mode
SFW	Stratospheric Final Warming
SH	Southern Hemisphere
SPARC	Stratosphere-troposphere Processes And their Role in Climate
SSW	Sudden Stratospheric Warming
S-RIP	SPARC Reanalysis Intercomparison Project
WMO	World Meteorological Organization
W1	Wavenumber 1
W2	Wavenumber 2