Stratospheric Modulation of Arctic Oscillation Extremes as Represented by Extended-Range Ensemble Forecasts

Jonas Spaeth and Thomas Birner

Meteorological Institute Munich, Ludwig-Maximilians-University, Munich, Germany

Correspondence: Jonas Spaeth (jonas.spaeth@physik.uni-muenchen.de)

Abstract. The Arctic Oscillation (AO) describes a seesaw pattern of variations in atmospheric mass over the polar cap. It is by now well established that the AO pattern is in part determined by the state of the stratosphere. In particular, sudden stratospheric warmings (SSWs) are known to nudge the tropospheric circulation toward a more negative phase of the AO, which is associated with a more equatorward shifted jet and enhanced likelihood for blocking and cold air outbreaks in mid-latitudes. SSWs are

5   also thought to contribute to the occurrence of extreme AO events. However, statistically robust results about such extremes are difficult to obtain from observations or meteorological (re-)analyses due to the limited sample size of SSW events in the observational record (roughly 6 SSWs per decade). Here we exploit a large set of extended-range ensemble forecasts within the subseasonal-to-seasonal (S2S) framework to obtain an improved characterization of the modulation of AO extremes due to stratosphere-troposphere coupling. Specifically, we greatly boost the sample size of stratospheric events by using potential

10   SSWs (p-SSWs), i.e., SSWs that are predicted to occur in individual forecast ensemble members regardless of whether they actually occurred in the real atmosphere. For example, for the ECMWF S2S ensemble this gives us a total of 6101 p-SSW events for the period 1997-2021.

A standard lag-composite analysis around these p-SSWs validates our approach, i.e., the associated composite evolution of stratosphere-troposphere coupling matches the known evolution based on reanalyses data around real SSW events. Our

15   statistical analyses further reveal that following p-SSWs, relative to climatology: 1) persistently negative AO states (> 1 week duration) are 16% more likely, 2) the likelihood for extremely negative AO states ($< -3\sigma$) is enhanced by at least 35%, while that for extremely positive AO states ($> +3\sigma$) is reduced to almost zero, 3) a p-SSW preceding an extremely negative AO state within 4 weeks is causal for this AO extreme (in a statistical sense) up to a degree of 27%. A corresponding analysis relative to strong stratospheric vortex events reveals similar insights into the stratospheric modulation of positive AO extremes.

1 Introduction

20

Day-to-day variability of the northern extratropical hemispheric-scale circulation during winter is dominated by the so-called Northern Annular Mode (NAM, Thompson and Wallace, 1998). The surface manifestation of the NAM is often referred to as Arctic Oscillation (AO). This variability pattern primarily describes fluctuations of atmospheric mass over the polar cap with associated opposite fluctuations on its equatorward flank. In its positive phase the AO corresponds to increased mass

25   over the polar cap with associated strengthened pressure gradient across mid-latitudes that goes along with a stronger polar-

1

front/eddy-driven jet that is shifted poleward and more zonally aligned. Likewise, in its negative phase the jet is weakened, shifted equatorward and often more meriodonally distorted.

Although a single index cannot represent the entire extratropical weather, it indicates tendencies towards certain weather patterns, which in turn can also have strong local effects. Especially AO values that deviate considerably from 0 (the clima-
30   tological mean) are rare, by construction, and can often be associated with strong *local* weather extremes (Thompson and Wallace, 2001): For instance, the AO index was around -2.5 in winter 2009/10, which was accompanied by record cold snaps and snow fall over large parts of the United States, Europe and East Asia (Cohen et al., 2010). In winter 2019/20, extreme storminess over Central Europe occurred during a highly positive AO phase with wind gusts of up to 177 km/h being recorded over Germany (Haeseler et al., 2020). Furthermore, Kim et al. (2020) report increased likelihood of Siberian wildfires following
35   positive AO periods.

The AO can also be influenced by "external" weather patterns and one prominent teleconnection exists between the AO and the stratospheric polar vortex. The latter describes a strong westerly wind band around $60°$N extending over 10 hPa, which forms every year in winter (Waugh et al., 2017). Numerous studies show that, on average, a very strong polar vortex (SPV) is associated with a strengthened circumpolar flow in the troposphere - as indicated by a positive AO index (e.g., Baldwin
40   and Dunkerton, 2001; Lawrence et al., 2020; Rupp et al., 2021). The reverse is true for a weak polar vortex, with such events being a special case: The breaking of planetary waves in the stratosphere and the associated westward forcing can lead to a complete breakdown of the polar vortex. In these cases, the zonal mean zonal wind reverses and the climatologically dominant westerly winds are replaced by weak or moderate easterlies. During the vortex disruption, air masses converge in the center of the vortex and are forced to sink. The accompanying strong and rapid adiabatic heating is the reason that such extreme weak
45   vortex events are called sudden stratospheric warmings (SSWs, Baldwin et al., 2021). SSWs are observed about 6 times per decade and are, as described previously, associated with a negative AO index on average. On synoptic scales, SSWs have also been tight to subsequently favored occurrence of certain weather regimes over the North Atlantic (Domeisen et al., 2020c) and over North America (Lee et al., 2019).

Consistent with the local implications of a negative AO index, SSWs can for example lead to cold spells in Northern Europe
50   and increased storminess over Southern Europe (Domeisen and Butler, 2020, and references herein). Whether it is generally valid that SSWs, and also strong polar vortex events, lead to a subsequently more likely occurrence of AO extremes (and associated local extremes) is difficult to analyze because the statistical links are weak in each case, i.e., not each SSW/SPV event is followed by an AO extreme. Therefore, a very large sample of SSW and SPV events are needed. However, reanalyses data only cover about 40-70 years, depending on the data set, and thus about 30-40 SSWs - too few to robustly determine
55   conditional probabilities (e.g., given a stratospheric extreme event, how likely is a following tropospheric extreme event).

In order to allow for analyses of larger event sample sizes, past studies have used, for example, idealized model simulations (e.g., Hitchcock and Simpson, 2014; Jucker, 2016). Even though such models have proven to be useful to develop a qualitative and conceptual picture, they often show a weaker tropospheric response to stratospheric events compared to observational data (Gerber et al., 2009). In this study, we aim to improve the characterization of coupled stratospheric and tropospheric
60   circulation extremes using operational, state-of-the-art, extended-range forecasts. Relatively large ensembles, frequent model

initializations and the generation of hindcasts allows us to analyze a large set of predicted SSWs and SPV events (p-SSWs/ p-SPVs, see discussion in section 2). Although the vast majority of these p-SSWs did not materialize in the real atmosphere we show that they nevertheless provide reliable statistical information about stratosphere-troposphere coupling. Our analyses implicitly assume that each ensemble member corresponds to a possible real-atmospheric evolution. The diagnosed p-SSWs include false alarm events (see, e.g., Taguchi, 2020), which we assume are based on the same underlying physics as those SSWs that occurred in the real atmosphere. Furthermore, the individual evolution (related to forecast score) is arguably not relevant for statistical characterizations of circulation anomalies.

The analysis is thus based on the assumption that the forecast models simulate the observed variability of the AO sufficiently well, including its modulation due to stratospheric variability. High-top models, in particular, show realistic stratosphere-troposphere coupling (Domeisen et al., 2020a, b). However, due to the small sample size of observed events, it is generally difficult to conclude whether any discrepancies between model and observational data are due to model or sampling errors. For this study, we will show that the models agree with observations in established diagnostics that can be robustly derived from reanalyses, including, e.g., the frequency of SSWs, their seasonality and their average impact on the subsequent AO. Although our quantitative statistical analyses cannot be compared directly to observational data they may be considered as best estimate given the currently available observational record and modeling capabilities.

We focus on following research questions:

1. By how much do stratospheric polar vortex extremes increase the probability of persistently positive or negative AO phases?

2. By how much do stratospheric polar vortex extremes increase the probability of subsequent AO extremes?

3. How often can stratospheric polar vortex extremes be considered causal[1] for subsequent AO extremes?

The paper is organized as follows: Section 2 provides an overview of the extended-range forecasts used in this study. Section 3 defines stratospheric and tropospheric circulation extremes and presents basic event statistics. For SSWs, we validate in section 4 that the predicted events agree, in well-known diagnostics, with events that are identified in reanalysis data. This motivates section 5, where the probability of AO extremes following predicted SSWs is analyzed. Conversely, section 6 shows how often predicted AO extremes are preceded (and caused) by predicted SSWs. Section 7 reveals in a similar fashion the statistical relation between predicted strong polar vortex events and predicted positive AO extremes, before the key findings are discussed and summarized in section 8.

2 Description of extended-range ensemble forecasts

The subseasonal to seasonal (S2S) prediction project (Vitart et al., 2017) provides a collection of extended-range (up to 60 days lead time) ensemble forecasts from different weather services. Forecasts differ in terms of model specifications (e.g., spatial

---

[1]in a statistical sense, see section 6 for a discussion

Table 1. Dataset specifications.

	S2S ECMWF	S2S UKMO	ERA5
Type	Forecast	Forecast	Reanalysis
Vertical Res.	L91	L85	L137
Time Range	d 0-46	d 0-60	1979-2021
Realtime	51 member, 2 inits / week	4 member, daily inits	-
Hindcast	11 member, 2 inits / week, past 20y	7 member, 4 inits / month, 1993-2016	-
# Realtime Ens. Used	114	396	-
# Hindcast Ens. Used	2280	1173	-
# Individual Model Runs	30894	9795	-

resolution, parameterizations, maximum lead time). All forecast systems create hindcasts in addition to the realtime forecasts in order to calibrate the forecasts and to allow the construction of the model's climatology. For our application, the most relevant demand is an accurate representation of the stratosphere and in particular of stratosphere-troposphere coupling. Furthermore, a forecast model with a large number of hindcasts is beneficial, because it allows for more robust analyses by including multiple past years. Lastly, a large maximum lead time is needed as we want to identify extreme events in the forecasts and are then also interested in the time periods before and after the event.

We choose to use ECMWF and UKMO forecasts for this study, as these models best meet the above listed requirements. Importantly, both models have been demonstrated in previous studies to have a realistic representation of stratosphere-troposphere coupling (Domeisen et al., 2020a, b).

For the decision on which initialization dates to use for the analyses, a trade-off has to be made between having as large a sample as possible and the fact that the forecast models are updated about every 1-3 years. Since late 2016, the ECMWF model (CY43R1) has been running at a higher horizontal resolution. Therefore, to avoid mixing forecasts before and after 2016, forecasts from winter 2017/18 up to and including 2020/21 are analyzed. Note that a minor model version change occurred in 2019, where initial conditions for the hindcasts are then obtained from ERA5 instead of ERA-Interim. However, we do not expect this to be a major limitation for our analyses, as we are mostly interested in the overall statistical behavior of stratosphere-troposphere coupling, as opposed to single forecast performance.

We focus on Northern winter dynamics by analyzing forecasts initialized between mid-November (11/16) and end of February (02/22). For the four winter seasons, the ECMWF model thus features 114 real-time ensemble forecasts of 51 members each and 2280 ensemble hindcasts of 11 members each. This results in a total of 30894 individual model runs, all of which we refer to as "forecasts" for simplicity. For consistency, UKMO forecasts are used from the same initialization period, leading to 9795 forecasts available for this model. A summary of the key specifications of the forecasts is given in Table 1, along with details of the ERA5 data (Hersbach et al., 2020) used.

3 Event statistics of stratospheric and tropospheric circulation extremes

3.1 Data sets and overall methodology

115 Each of the forecasts from the total set of 30 894 ECMWF forecasts provides a 47-day time series of the evolution of the atmosphere (UKMO: 61 days). In this study, we define specific events and then scan each forecast for the occurrence of such an event. If there are multiple events of one event type within one forecast, only the first event is used. Note that, by definition, all identified events are predicted events, but each may or may not actually occur in the real atmosphere. To highlight this aspect, and to avoid confusion with actual real-atmospheric events, we will refer to events identified in the forecasts with

120 a "p"-prefix in the following, where "p" stands for "predicted" (alternatively, it could be thought of as "potential" for some aspects).

For both datasets, ECMWF and UKMO, all individual forecast runs are treated equally and independently. This assumption is violated especially for forecasts belonging to the same ensemble. In fact, at initialization time these forecasts agree *entirely* except for ensemble perturbations. The individual members diverge from each other only with increasing lead time, when the

125 predictability of the atmospheric flow gradually decreases. For this reason, we analyze only those events that occur at or after a forecast lead time of 10 days. It is assumed that initial condition-memory has sufficiently reduced by this point so that no two individual forecasts fully match, and the same is thus true for the evolution of the identified events. This ensures a degree of statistical independence. The use of hindcasts further guarantees sampling of different boundary conditions, such as due to the El-Niño-Southern-Oscillation, the Maddan-Julian-Osciallation or sea ice variations.

130 Furthermore, it is ensured that for each identified event both negative and positive lags can be considered. Due to the finite maximum lead time of each forecast, this demand is generally limited. For a predicted event that occurs early in the forecast (but after 10 days at the earliest), only a short period before the event can be examined, and the reverse is true for an event that occurs shortly before the end of the forecast. Therefore, to ensure a minimum common lag time that can be analyzed, events are additionally required to occur no later than 10 days before the end of the forecast. Consequently, events are allowed to

135 occur between day 10 and 36 for ECMWF forecasts and between day 10 and 50 for UKMO forecasts. Thus, for all events, the lag period ±10 days can be examined, but with increasingly longer positive and negative lag times, fewer and fewer events contribute to the composite.

Extreme events are defined that refer to exceptional anomalies in the stratospheric and tropospheric circulation, respectively. As a measure of the strength of the stratospheric polar vortex we use the zonally averaged zonal wind at 10hPa at 60°N,

140 hereafter referred to as u60.

3.2 Predicted SSWs

We define Sudden Stratospheric Warmings (p-SSWs), as days when u60 transitions from positive to negative, i.e., the polar vortex breaks down. As explained above, we do not include p-SSWs predicted within the first 10 days after forecast initialization. However, for p-SSWs, u60 is required to be solely positive within these first 10 days to ensure an intact westerly polar

145 vortex at the start of the forecast. Following this event definition, we identify 6 101 p-SSWs in the ECMWF and 2 716 p-SSWs
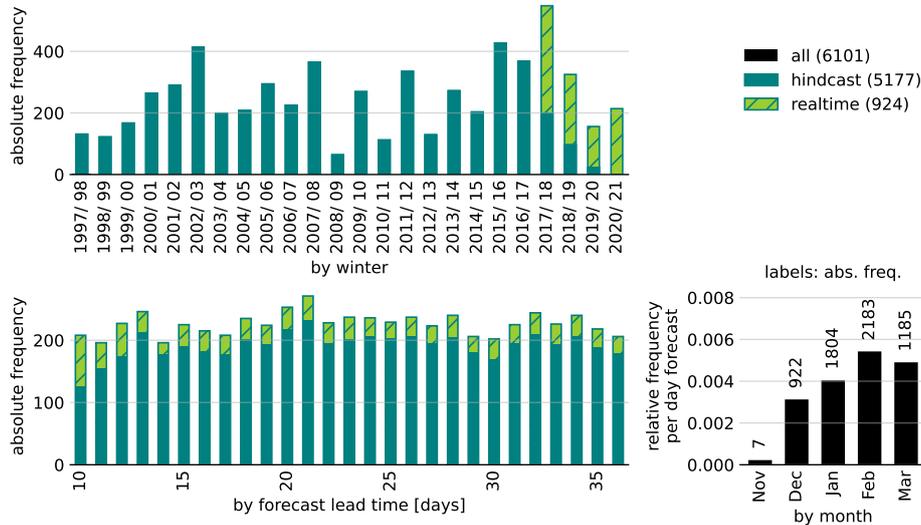
Figure 1. Distribution of analyzed p-SSWs in the ECMWF forecasts grouped by winter (top left), by forecast lead time (bottom left) and by month (bottom right).

in the UKMO model. As a sensitivity test we also applied our analyses to a subset of predicted *dynamical* SSWs (defined by a SSW where, in addition, u60 drops at least 20 ms^{-1} averaged over -5 to $+5$ days lag relative to the SSW central date). The results agree with some surface signatures being enhanced, but as differences to p-SSWs are only small, we here restrict the analyses to p-SSWs. Furthermore, previous literature has suggested that polar vortex displacement events could be predictable

150   at slightly longer lead times than displacement events. However, due to small sample sizes, the differences are generally not statistical significance and we therefore do.

     In Figure 1 we provide an overview about the distribution of ECMWF p-SSWs as a function of the year, forecast lead time and calendar month (see Fig. S1 for a corresponding analysis of UKMO forecasts). p-SSWs are found for all winter seasons considered (top left panel). The largest number of events is identified in the winter season 2017/18, which has also the most

155   forecasts (realtime 2017/18 plus hindcasts related to initializations from 2018/19 to 2020/21). Different factors lead to a highly varying number of events between the different years. These include internal dynamic variability, a slightly varying number of underlying forecasts, due to the realtime/ hindcast prediction setup, and the varying number of events per winter due to the evolution of the polar vortex of the real atmosphere in the respective winter, which determines the initial conditions of the forecasts.

160   A forecast that is initialized with a strong polar vortex tends to maintain a strong polar vortex and produces fewer SSWs compared to a forecast with an initially weak polar vortex. Moreover, forecasts that do not start with ten consecutive days of positive u60 are discarded by default. Thus, if the polar vortex in the real atmosphere is already easterly at the initialization time or is predicted to become easterly within the quasi-deterministic forecast range of ten days, such forecasts will not contribute any events to the analysis. This can be illustrated by the example of the 2009'th SSW (24 January 2009, see Butler et al., 2017).

165 The event had low predictability at lead times longer than 8 days (Karpechko, 2018). Moreover, the event was generally very rare due to the polar vortex being exceptionally strong between end of December 2008 and mid-January 2009. As a result, 2008/09 shows the lowest number of SSWs: In the first winter half up to initialization dates around mid-January, hardly any events were predicted due to the relatively strong polar vortex. Later, forecasts predicting the real atmosphere SSW only did so at less than +10 day lead time, such that those events are discarded. Later initializations up to mid-February are excluded,

170 because these do not predict persistent positive u60 within the first 10 days lead time, due to the preceding SSW. As a result, winter season 2008/ 09 contributes only 64 (UKMO: 22) p-SSWs to the analysis.

Based on the average number of 226 events per day lead time in the ECMWF model (cf. bottom left panel in Fig. 1), we estimate the corresponding probability of at least one SSW per winter season ($\approx$ 135 days from mid-November to end of March):

175 $$P(\text{SSW}) = 1 - \underbrace{\underbrace{\left(1 - \frac{226}{30894}\right)^{135}}_{\text{no SSW}}}_{\text{for 135 consecutive days}} = 63.3\%/\,\text{winter}$$

This is consistent with the number of observed SSWs in reanalyses that is roughly 6 per decade (Butler et al., 2015). While the rate of events per forecast day fluctuates only weakly in the ECMWF model, it moderately increases with lead time in the UKMO model (Fig. S1, bottom left panel). One might expect this to be due to the longer maximum lead time of the UKMO model (+60 days) compared to the ECMWF model (+46 days), which may allow more final-warming-like events. However,

180 we find that the trend is still apparent when all forecasts initialized in February are excluded from the analyses (not shown).

Consistent with reanalyses (e.g, Ayarzagüena et al., 2019) and across both, the ECMWF and the UKMO model, the p-SSW frequency shows a maximum in February (bottom right panel in Fig. 1).

3.3 Predicted strong vortex events

Past literature has identified stratosphere-troposphere coupling not only following SSWs, but also following strong polar vortex

185 events (SPVs, e.g., Baldwin and Dunkerton, 2001). However, the definition of a single event in these cases is somewhat more ambiguous, as there is no dynamically motivated threshold for u60, compared to $0\,\text{ms}^{-1}$ for SSWs. In addition, the dynamical changes in cases of a strong polar vortex are generally less abrupt, making it harder to pin down one particular central event day. For these reasons, we focus mainly on SSWs in this paper, however, we also provide a summary of the key results for SPV analyses in section 7. In these analyses, p-SPVs are defined as the first day on which u60 exceeds a threshold that, based on

190 percentiles, represents the "opposite" of the SSW threshold of $0\,\text{ms}^{-1}$. Depending on the model's climatology, this threshold describes approximately the 91st percentile of the u60 distribution and is around $47\,\text{ms}^{-1}$.

3.4 Predicted NAM1000 events

In the troposphere, extreme events are defined based on the Northern Annular Mode Index at 1000 hPa (short: NAM1000, or AO[2]). This is calculated by first area-weighting the geopotential field between 65 and 90°N by the cosine of latitude and then averaging over the entire polar cap. The NAM index then is the negative standardized anomaly of the obtained quantity. For technical details about the deseasonalization via the hindcasts, the reader is referred to appendix A. The positive phase of the NAM1000 describes a negative geopotential anomaly over the polar cap and a thereby induced enhanced circumpolar westerly circulation. Conversely, a negative NAM reflects a weaker westerly circulation, which is typically associated with a southward shift of the jet that is also zonally more distorted.

We define tropospheric extreme events as the first day when the NAM1000 falls below -3 (p-NAM1000$^-$ extreme) or exceeds $+3$ (p-NAM1000$^+$ extreme). After testing different thresholds, we opt for a threshold of 3 standard deviations because it represents a tradeoff between severity of event and sufficiently large resulting sample sizes.

4 Evaluation of stratosphere-troposphere coupling based on predicted SSWs

To provide a baseline for our more detailed statistical analyses in the following sections, we first evaluate the general behavior of stratosphere-troposphere coupling based on p-SSW events in the S2S data. To do so we focus on the lag-composite evolution of the NAM1000 index relative to p-SSWs compared to real-atmospheric SSWs from ERA5. In addition, we show the NAM index at 200hPa (short: NAM200) because the lower stratosphere has been found to play an important role in stratosphere-troposphere coupling (e.g, Karpechko et al., 2017; White et al., 2020).

Figure 2 shows the evolution of u60 (top), NAM200 (center) and NAM1000 (bottom) during SSWs, averaged over all events, separately for ECMWF and UKMO. In addition to the composite mean, the 33rd to 66th percentiles across all ECMWF events on the respective lag day are shown. By construction, 100% of all events (ECMWF: 6 101, UKMO: 2 716) contribute to lag days ±10. For larger positive or negative lags, some forecasts have reached their maximum forecast lead time or have not yet been initialized. Therefore, the number of events drops off, which makes the statistics less robust: For the ECMWF model, the number of contributing events falls below 20% for lags smaller than -31 and larger than $+31$ days (UKMO: smaller than -44 and larger than $+39$ days).

By construction, u60 transitions from westerly to easterly at lag 0. Anomalies of u60 are slightly positive ahead of -14 days lag, which we interpret as an indication for vortex preconditioning (McIntyre, 1982; Albers and Birner, 2014; Jucker and Reichler, 2018). The anomalies become negative within the second week prior to the event central date. The largest average negative anomalies occur only few days after the event central day. Afterwards, the vortex reestablishes and the average anomalies reach zero again after approximately 35 days. Consistent with, e.g., Baldwin and Dunkerton (2001), both NAM200 and NAM1000 are negative following the event. The shift of the NAM200 happens earlier (at lag day -11) and the timing aligns well with the weakening of the polar vortex at 10hPa. The NAM200 anomaly is also more pronounced (≈ -0.5) compared to

---

[2]We will use the notation "(p-)NAM1000" where we explain and refer to technical details. Due to better readability and more widespread usage in other literature, we use the term "AO" where we make generalized statements and in the conclusions. However, we note that both terms are interchangeable.
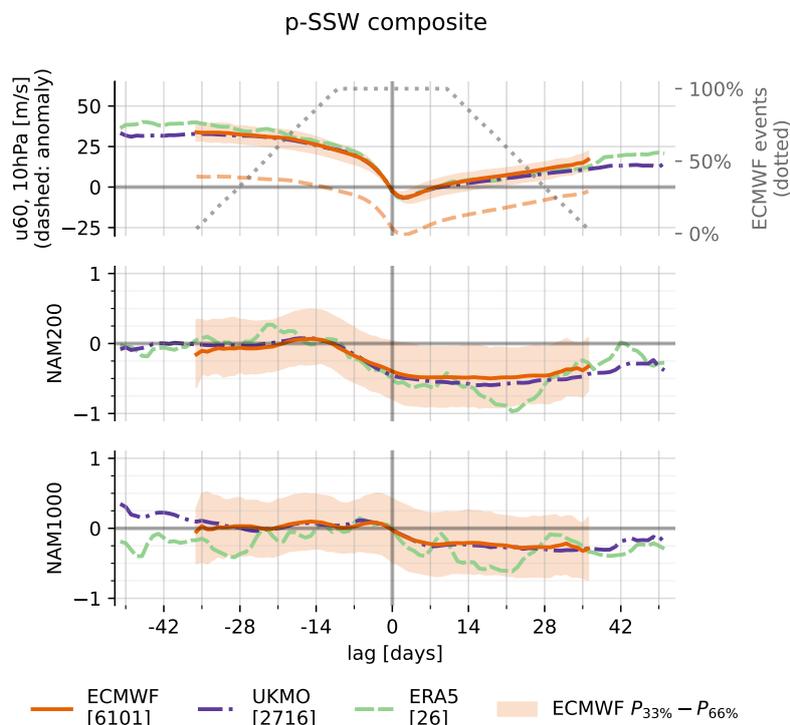
Figure 2. Lagged composite evolution of u60 (top panel), NAM200 (middle panel) and NAM1000 (bottom panel) relative to p-SSWs (ECMWF, UKMO) and SSWs (ERA5). It is presented the mean across all ECMWF events (orange, solid), the 33rd to 66th percentiles across all ECMWF events (orange, shaded), the mean of all UKMO events (purple, dash-dotted) and the mean across all ERA5 events (green, dashed). In the top panel further denoted are the average u60 anomalies (orange, dashed) and the relative number of contributing events to the composite in the ECMWF model (gray, dotted). Square brackets denote the total number of events, for each dataset.

the NAM1000 (≈ -0.3). Interestingly, the NAM1000 distribution is slightly shifted toward positive values in the week prior to the central date, which is also robust for other diagnostics like the 10th, 30th, 70th and 90th percentiles (not shown). At long

225 positive lag times, the NAM indices at 200 and 1000hPa are still negative (ECMWF: lag +36 days, UKMO: lag +51 days), but the trend goes to weaker positive values again.

Overall, the results are in agreement with ERA5 and previous literature and especially the evolution of u60 is remarkably similar. The negative NAM response at 200hPa and 1000hPa seems to be slightly stronger in the reanalysis, however, it is also noisier due to the smaller sample size.
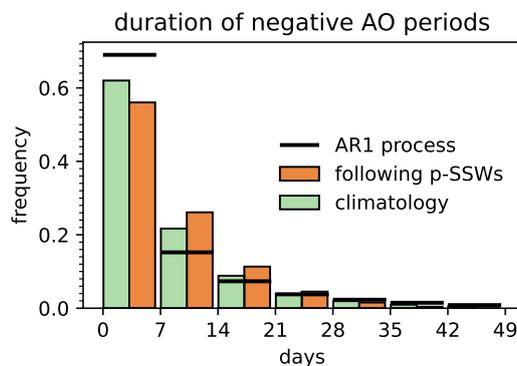
Figure 3. Histogram of the duration of negative NAM1000 periods, quantified by the number of consecutive days of NAM1000 < 0 and binned by 7-day chunks. Periods following ECMWF p-SSWs (orange bars, right half) are compared to the ECMWF model's climatology (green bars, left half) and a random first order auto-regressive model of the same 1-day-lag-autocorrelation as the NAM1000 in ERA5 (black, horizontal lines).

5   Predicted AO extremes following predicted SSWs

In the following, we will exploit the larger available sample size of p-SSW events to diagnose and quantify whether the shift of the average AO index towards negative values is caused by 1) more persistent negative AO phases and/or 2) an increased probability for AO^{-} extremes.

Figure 3 presents a histogram of the duration of predicted negative NAM1000 phases in the ECMWF S2S model, binned into 7 day chunks. The duration is defined as the number of consecutive days with negative NAM1000. The climatology serves as a reference including all $30\,894$ ECMWF forecast used for this study. With approximately 62%, most phases of negative NAM1000 are shorter than 8 days. As another reference, a first order autoregressive model was set up with zero mean and standard deviation of 1, which may serve as a baseline. Its 1-day-autocorrelation is chosen to match the ERA5 NAM1000 timeseries and for robustness, it is estimated by averaging the lag-1-autocorrelation and the square-root of the lag-2-autocorrelation, yielding 0.91. It turns out that the NAM1000 climatology shows short negative phases (≤ 7 days) less often and long negatives phases (≥ 8 days) more often compared to the AR1 process. This is an indication for the NAM1000 index having a slightly longer decorrelation timescale in the S2S model compared to ERA5; which apparently also overwhelms the effect of negative NAM1000 periods being cut off by the end of the forecast which introduces a bias towards shorter negative periods.

In addition, the diagnostic is presented for periods following p-SSWs. Here, the NAM1000 index is analyzed between lag day +1 relative to the event date and the maximum available lag time, which ranges between +10 and +36 days, depending on the forecast lead time when the event happens. Similar to the reference climatology, this diagnostic also underestimates the occurrence of long negative NAM1000 periods as the forecasts have finite maximum lead time. Nevertheless, periods following SSWs show a reduced frequency of shorter and an increased frequency of longer negative NAM1000 periods, compared to the
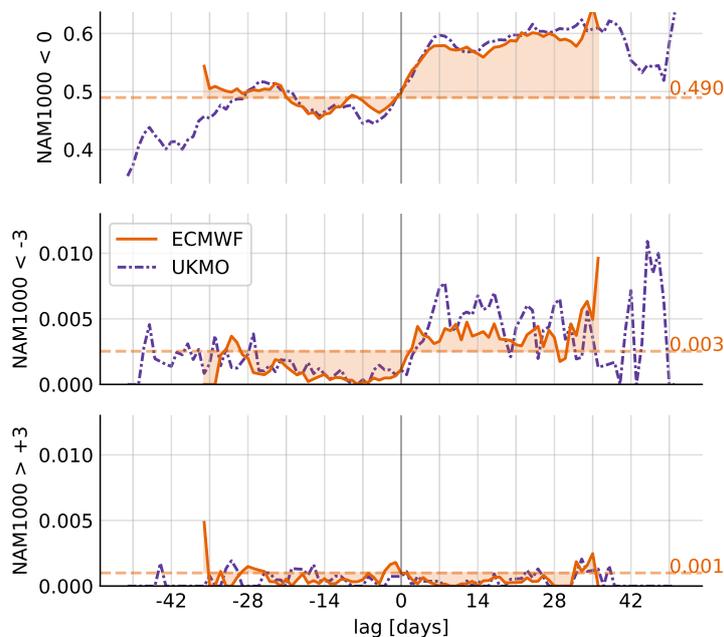
Figure 4. Daily probability of p-NAM1000 < 0 (top panel), p-NAM1000 < -3 (middle panel) and p-NAM1000 $> +3$ (bottom panel) relative to p-SSWs, quantified by the fraction of events fulfilling the respective condition. The statistics are presented for ECMWF events (orange, solid) and UKMO events (purple, dash-dotted) and compared to the ECMWF climatology (dashed horizontal lines). For each lag day, the number of events fulfilling the respective condition has been normalized by the total number of events contributing to the composite on that lag day.

250   climatology (and thus also to the AR1 process): For instance, 38% of negative NAM1000 periods are longer than 7 days in the climatology, whereas this probability rises to 44% following p-SSWs, which corresponds to a relative increase of 16% (UKMO: also 16%, not shown).

We now focus on p-NAM1000 extreme events and analyze to what extent p-SSWs contribute to an increased probability after such events. As the NAM1000 distribution shifts at positive lag times (see Fig. 2), also the daily statistical probability of

255   extreme NAM1000 values changes. This effect is quantified in Fig. 4. First, based on all available forecasts, the climatological likelihood is computed for negative (< 0), extremely negative (< -3) and extremely positive ($> +3$) p-NAM1000 events on any random day, by normalizing the days fulfilling the respective condition with the total number of available forecast days. The resulting probability baseline for the ECMWF forecasts is 49% for negative, 0.3% for extremely negative and 0.1% for extremely positive events, where the asymmetry is due to the negative skewness of the NAM1000 distribution (-0.13).

260   The fraction of events in the p-SSW composite that have negative NAM1000 values fluctuates around 50% at negative lags with only small deviations from the climatology. Within the first week following the event, this fraction increases and appears to saturate around 60%. Consequently, in the period following a p-SSW, a negative NAM1000 is, at each day, approximately
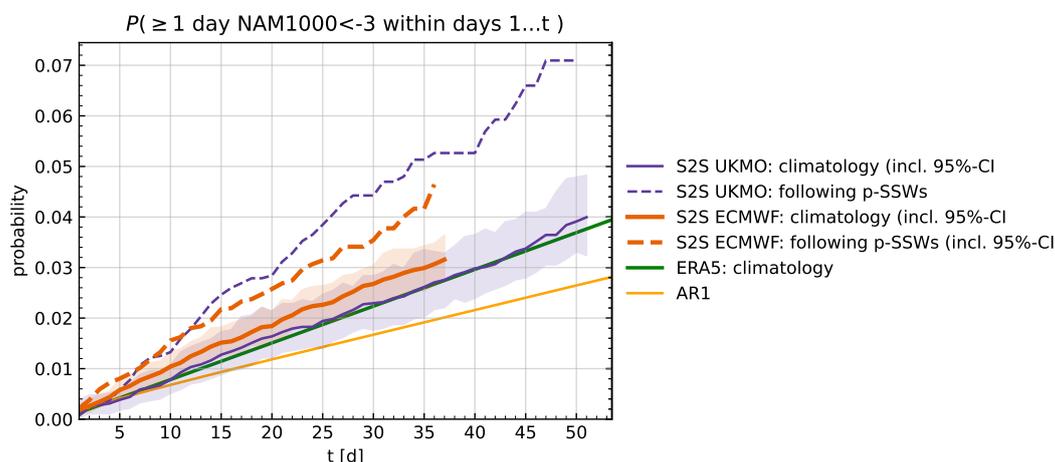
Figure 5. Probability of at least one day of p-NAM1000 < -3 within days 1 to t, where periods following p-SSWs (dashed) are compared to the model's climatology (solid + shading: mean incl. 95%-confidence-interval), separately for the ECMWF (orange) and the UKMO (purple) model. In addition, the climatology of ERA5 (solid, green) and a random first-order auto-regressive model of the same 1-day-autocorrelation (yellow) are presented.

50% more likely compared to a positive NAM1000 (60% vs. 40%). The results are consistent between ECMWF and UKMO during the ± 4 week period where the composites for both models consist of more than 30% of all events.

265   Extremely negative NAM1000 values in the dataset appear with a climatological probability that is similar to what would be expected for a (one-sided) 3-sigma-event of a standard normal distribution (0.27%). At negative lags, they occur overall less frequent compared to climatology. In contrast, around lag 0, the probability increases and persists at $\approx 0.40\%$ for more than four weeks. The increase appears to be larger in the UKMO model, however due to fewer events the diagnostic is also noisier. The fraction of events with extremely positive NAM1000 values is smaller compared to climatology throughout the entire lag

270 period. This is largely consistent between the models from ECMWF and UKMO.

     An altered probability of extreme NAM1000 events may be of higher socio-economic relevance than a small shift of the mean. However, the absolute daily probabilities of extremely negative NAM1000 events are still small even though the relative increase due to the p-SSWs is indeed considerable. In practice, the relevant question might not be how much the probability increases on only one specific day following a p-SSW. It may be more relevant to quantify the increased risk for an extreme

275 NAM1000 within a given time period that is due to a p-SSW.

     Figure 5 shows the probability P of at least one p-NAM1000$^-$ extreme between day 1 and day t as a function of t. We compare the period following p-SSWs with the climatology of the UKMO and the ECMWF model, the ERA5 climatology and an AR1 process of the same autocorrelation as the NAM1000 index in ERA5. For the ECMWF and the UKMO climatology, the probability was sampled for all the forecasts from lead time +10 days[3] to lead time +10+t days. For the sampling, bootstrapping

---

[3]day 10 is the first where we search for p-SSWs, however, this choice is arbitrary and the resulting climatology is not very sensitive to this choice

280 was applied where n random forecasts were picked and analyzed for p-NAM1000$^-$ events, providing also an estimate for the 95% confidence interval (with n being the number of forecasts in the p-SSW composite). For ERA5 and the AR1-process, the probability is sampled from all days t_0 of the time series to day $t_0 + t$, respectively. The probability for ECMWF and UKMO p-SSWs is computed between lag day 1 of the respective composite until lag day t. This is realized by computing the counter-event, i.e.: $1 -$ "no p-SSW between day 1 and day t"[4]. Naturally for all datasets, as t increases, also the probability P

285 increases, as the time window for finding at least one NAM1000$^-$ extreme gets wider.

The results show that p-SSWs are consistently leading to an increased integrated risk of extremely negative NAM1000 events. For example, the probability in the ECMWF forecasts of at least one NAM1000$^-$ extreme within 28 days following the event is 3.4%, compared to 2.6% for its climatology. Overall, p-SSWs seem to affect the probability more in the UKMO model, as the probability following p-SSWs is higher and the climatological baseline is also lower compared to the ECMWF

290 model. The baseline in ERA5 is slightly lower than in the ECMWF model, but agrees well with the UKMO climatology. All probabilities range considerably higher than the probability of a one-sided 3-sigma event for the AR1-process and as before this is a result of the negative skewness of the NAM1000 distribution.

Generally, all probabilities appear approximately linear in t, but it should be noted that the linear regime only holds for small enough t, as the probability will approach 1 and saturate in the limit of very large t. Furthermore, it is expected that

295 for much larger t (which cannot be displayed here, due to the finite maximum forecast lead time), the effect of a p-SSW increasing the subsequent extreme NAM1000$^-$ probability diminishes and the climatology will approach the one for p-SSWs. It has been tested and verified that the results do not change significantly when forecasts containing p-SSWs are left out for the computation of the climatology, suggesting that p-SSWs do increase but not dominate the total number of p-NAM1000$^-$ extreme events.

300 We have shown that the time-integrated probability for at least one day with NAM1000 < -3 is increased by a SSW. When displayed as a function of the period that is used to search for such p-NAM1000 events, the probability increase *relative to the climatological baseline* is roughly constant (e.g., 2% versus 1.5% $\hat{=} \times 1.33$ after 14 days, 3.4% versus 2.5% $\hat{=} \times 1.36$ after 28 days). This motivates the calculation of the average relative probability increase over time. The resulting factor provides an estimate for the question of how much p-SSWs increase the probability for p-NAM1000$^-$ extreme events – not limited

305 to a specific lag day, but time-integrated and thus independent of t. If the relative probability increase is around 0, negative p-NAM1000 extremes occur after p-SSWs with a similar frequency than climatology. If the increase is larger than 0, then p-SSWs lead to a higher probability of p-NAM1000$^-$ extremes.

Figure 6 summarizes this probability increase factor for different NAM1000 thresholds and for both S2S models. The estimated probability increase is computed by dividing P (as displayed in Fig. 5) by the corresponding climatological P

310 and averaging the obtained ratios over t. In Fig. 5, the probability curves were only displayed for NAM1000 < -3, whereas here, the procedure is repeated for different choices of the NAM1000 threshold. The results imply that the more negative the threshold, the greater the relative increase in probability. Even though the daily probability for negative p-NAM1000 values

---

[4]Special care must be taken when normalizing the events matching the condition by the total amount of events where the latter 1) must be subtracted by the number of events that have already matched the condition at earlier t's and 2) generally decreases for longer lag times.

Estimated Probability Increase of NAM1000 Extremes
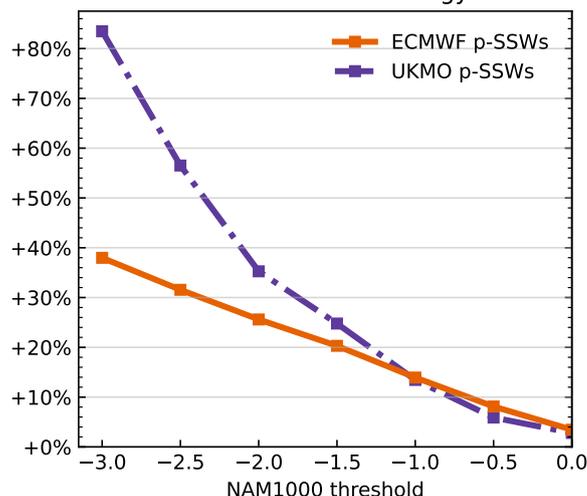Relative to Climatology

Figure 6. Estimated probability increase of negative p-NAM1000 extremes relative to the model's climatology, separately for the ECMWF (orange, solid) and the UKMO (purple, dash-dotted) model. The estimate is obtained by dividing the curve for p-SSWs in Fig. 5 through the respective model's climatology and averaging over t. In Fig. 5, this is only shown for p-NAM1000 < -3; here, we present the diagnostic as a function of this p-NAM1000 threshold.

was shown to be considerably increased following p-SSWs (see Fig. 4), the probability increase of finding at least one day of negative p-NAM1000 within a longer time-period is extremely high for both, climatology and following p-SSWs. Therefore

315   the relative effect of p-SSWs is necessarily small by this measure (around +3%). In contrast, for larger negative thresholds, the effect becomes stronger as those values are generally rare and even a small influence on the distribution matters. The estimated probability increase for NAM1000 $< 2, 2.5, 3$ as revealed by UKMO is stronger compared to the ECMWF model, consistent with Fig. 4. In particular, p-SSWs increase the probability of days with NAM1000 < -3 by $\approx 40\%$ in the ECMWF and even $\approx 80\%$ in the UKMO model.

320   **6   Predicted AO extremes preceded by predicted SSWs**

The last section focused on given p-SSWs and subsequent statistical signatures in AO$^-$ extremes: $P(\text{subsequent AO}^- \text{ extreme}|\text{SSW})$. It was shown that not every SSW is followed by an AO$^-$ extreme (in fact, the vast majority are *not*). However, the probability relative to the baseline is increased, i.e., a SSW makes an AO$^-$ extreme significantly more likely.

A remaining open question is: to what extent do SSWs *cause* the AO$^-$ extremes that are following them? Here, we interpret

325   "cause" in a statistical sense, which involves comparing the likelihood of occurrence of AO$^-$ extremes with and without a preceding SSW. A direct cause-effect relationship between, say, a single SSW and AO$^-$ extreme is difficult to determine,
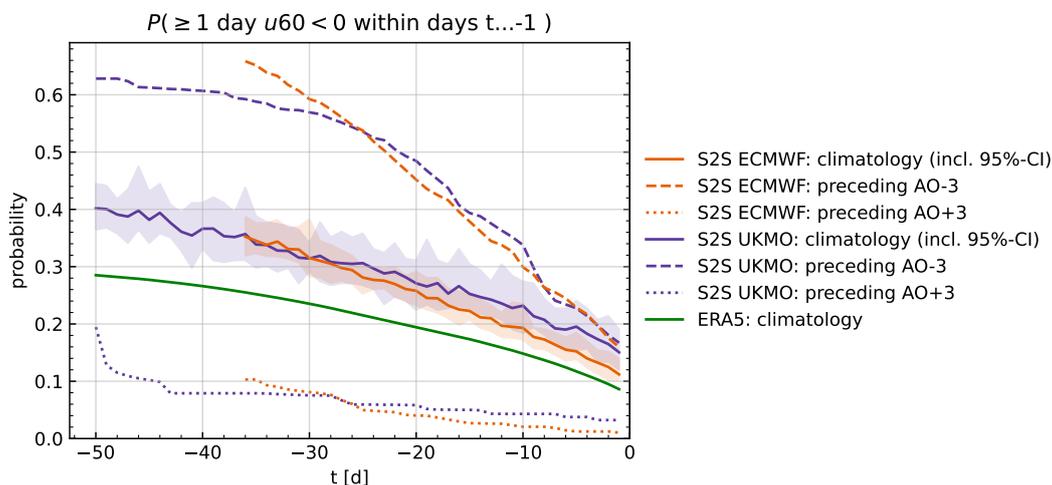
Figure 7. Probability of at least one day u60 < 0 within day t and day -1 relative to day zero, where day 0 is either a randomly sampled day in ERA5 or one of the S2S models (full lines), an AO$^-$ extreme event in the S2S models (dashed lines), or an AO$^+$ extreme event in the S2S models (dotted lines).

although mechanistic model experiments, e.g., by nudging the stratosphere to its observed evolution have been used with some success (e.g., Hitchcock and Simpson, 2014; Kautz et al., 2020).

330 An important required piece of information is the baseline climatology of the frequency by which any random day (i.e., regardless of its AO value) is preceded by a SSW, which provides an estimate of the expected chance occurrence of a SSW preceding an AO extreme. This baseline frequency is shown as a function of scanned time interval as full lines in Fig. 7 (comparing ERA5 and the S2S models). For example, the probability of SSW occurrence within 30 days preceding any random day is ≈ 0.24 in ERA5 and ≈ 0.32 in the S2S models. By subtracting this baseline frequency from the probability of p-SSW occurrence preceding p-NAM1000$^-$ extremes, we may then obtain estimates of causal SSW-AO extreme relationships. For
335 example, the probability of p-SSW occurrence within 30 days preceding a p-NAM1000$^-$ extreme is close to 0.6 in the ECMWF S2S model (orange dashed line in Fig. 7); by subtracting the baseline estimate of 0.32 we may conclude that, based on the ECMWF S2S ensemble and the time interval of 30 days, $\approx 28\%$ of all p-NAM1000$^-$ extremes are *caused* by a preceding p-SSW. Conversely, this implies that $\approx 72\%$ of all p-NAM1000$^-$ extremes are *not caused* by a preceding p-SSW (based on the 30 day time interval and the p-NAM1000$^-$ threshold).
340 Figure 8 summarizes the probabilities for AO$^-$ extremes that are either preceded by u60<0 by chance, caused by u60<0 and or not preceded by u60<0, as a function of the preceding period length and extended to different AO extreme thresholds. The chance occurrence is independent of the AO threshold, but it is slightly enhanced in the UKMO compared to the ECMWF model. In the limit of large preceding periods, which cannot be analyzed here due to the finite maximum forecast lead time, the probability for chance occurrences is expected to saturate at 1.
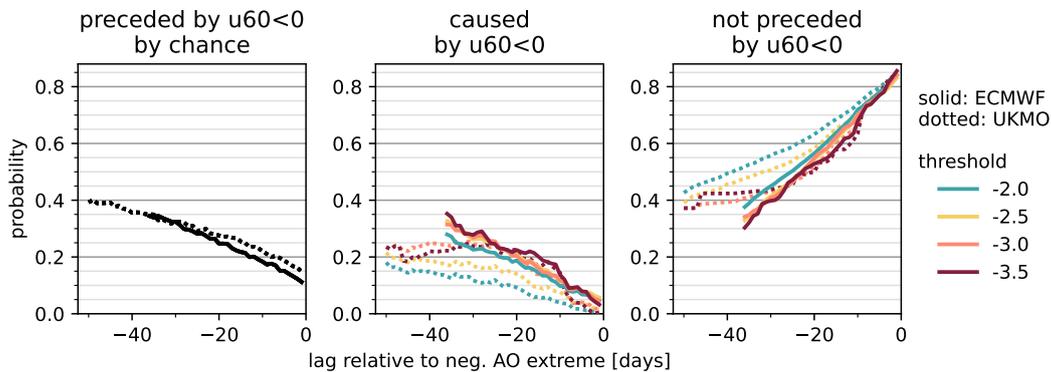
15

Figure 8. Probability for AO^- extremes to be preceded by u60<0 by chance (left panel; equal to "climatology" in Fig. 7), to be caused by u60 (center panel; equal to "preceding AO extreme" minus "climatology" in Fig. 7) or to be not preceded by u60<0 (right panel; equal to 1 minus "preceding AO extreme" in Fig. 7), as a function of the preceding time interval.

Based on ECMWF forecasts, probability that AO^- extremes are caused by preceding stratospheric easterlies (u60<0) increases for larger preceding periods. We find that the probability furthermore increases for stricter AO^- thresholds, e.g., within 28 days, 20% of AO< −2 events and 27% of AO< −3.5 events are caused by u60<0. Based on UKMO forecasts, the diagnostic shows a more pronounced sensitivity to the actual AO threshold, e.g., 14% of AO< −2 events, but 28% of AO< −3 events are caused by preceding u60<0 within 28 days. Furthermore, the probabilities seem to saturate in the UKMO model for preceding period lengths longer than about 30 days, which cannot be tested with the ECMWF model, due to the shorter maximum forecast lead time.

7 Predicted strong polar vortex events and related predicted, positive AO extremes

The previous section revealed that a significant fraction of AO^- extremes may be thought of as being caused by a preceding SSW (between $\sim 25-40\%$, depending on the threshold used). Here, we extend this analysis to the relationship between strong polar vortex events (SPVs) and AO^+ extremes.

The composite-mean evolution of p-SPVs (Fig. 9) reveals that u60 anomalies are of opposite sign, somewhat weaker in magnitude, but otherwise qualitatively similar to p-SSWs (lag 0: $\approx +20$ ms^{-1} for p-SPVs; ≈ -30 ms^{-1} for p-SSWs, cf. Fig. 2). It is observed that both S2S models agree very well in this respect. Moreover, for negative lags, there is little difference compared to a corresponding composite based on ERA5 data, but for positive lags, u60 is slightly stronger in ERA5. The NAM response at 200hPa and 1000hPa is qualitatively similar for p-SPVs and p-SSWs (with opposite sign), but the anomalies are again slightly weaker for p-SPVs, which is consistent with the weaker u60 anomalies (lag 21: +0.35 at 200hPa, +0.25 at 1000hPa). It is interesting that the NAM200 seems to react later to p-SPVs than to p-SSWs: While the index for p-SSWs starts to shift significantly to negative values already at lag −10 on average, a shift to positive NAM200 values for p-SPVs is observed only from lag −5 on. As with p-SSWs, the evolution of the NAM at 200hPa and 1000hPa relative to p-SPVs is less
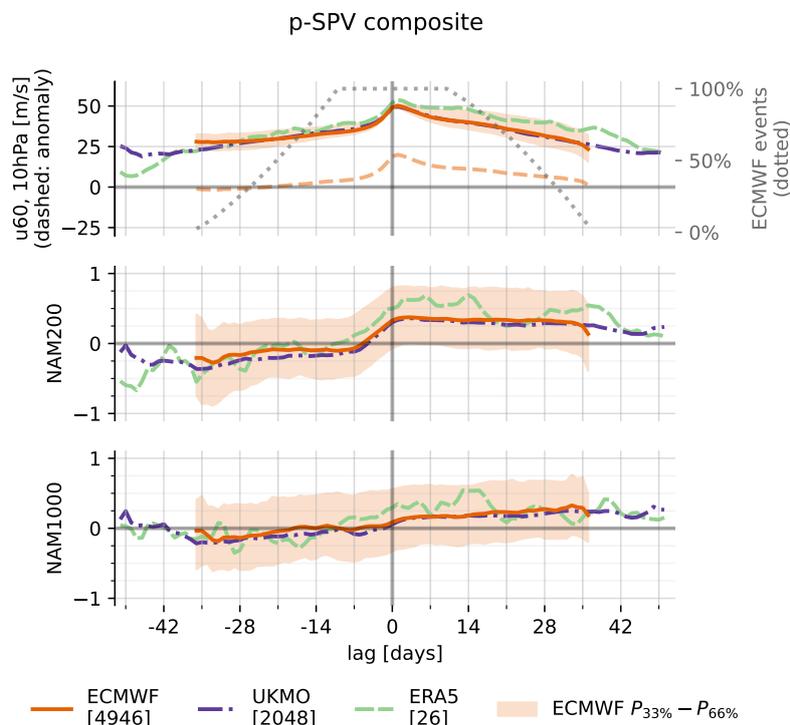
Figure 9. As in Fig. 2, for p-SPVs.

robust in ERA5 due to the smaller sample size, however, the anomalies tend to be slightly more pronounced than in the two S2S models. Overall, the composite-mean evolution of p-SPVs in the ECMWF and UKMO models appear to be consistent with real-atmosphere SPVs (as revealed by reanalysis data), as well as with previous studies (e.g., Baldwin and Dunkerton, 2001).

Following the same methodology as for p-SSWs, we use the large event sample sizes to quantify the statistical relation between p-SPVs and subsequent AO$^+$ extremes. First, we quantify the probability increase for an AO$^+$ extreme after a given p-SPV. Second, we analyze whether AO$^+$ extremes are more often preceded by p-SPVs than any random days, in order to compute the fraction of AO$^+$ extremes that my be considered to be *caused* by p-SPVs.

Figure 10 shows the probability increase of p-NAM1000$^+$ extremes after p-SPVs relative to climatology and as a function of the NAM1000 threshold, for both S2S models: $P($subsequent AO$^+$ extreme $|$ p-SPV$)$. For extreme thresholds of up to 2 standard deviations, the probability increase of positive NAM1000 extremes after p-SPVs is similar to the probability increase of negative NAM1000 extremes after p-SSWs ($\approx$30-40%). However, for larger thresholds, the probability increase gradually diminishes again.
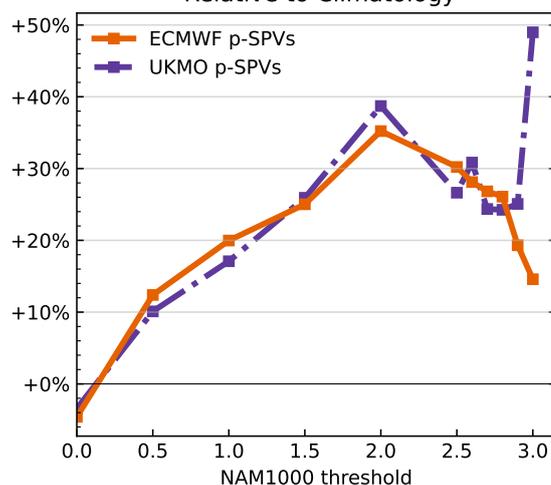
Figure 10. As in Fig. 6, for p-SPVs and subsequent positive NAM1000 extremes.

Figure 11 quantifies the fraction of positive NAM1000 extreme events that are preceded by a p-SPV. Following the same procedure as described for negative extremes and p-SSWs in section 6, it is observed that randomly sampled days are in about 34% of the cases preceded by a p-SPV within 28 days (left panel; UKMO: 29%), which serves as the baseline.

Ahead of positive NAM1000 extremes, p-SPV events are observed more often and the difference yields an estimate for the cases where the p-SPV is causal for the NAM1000 extreme (center panel). The results show that NAM1000 events of stricter thresholds are more often caused by a preceding p-SPV. For example, about 30% of NAM1000$> +2$ and 46% of NAM1000$> +3$ events are caused by a preceding p-SPV within 28 days (UKMO: 30% and 37%). The results further suggest that the causal influence of preceding p-SPVs starts to saturate for preceding periods longer than 30 to 40 days. Even though the agreement between the probabilities obtained via the ECMWF and via the UKMO model is not perfect, we highlight that the estimates are still relatively close, considering that the analyses refer to the extreme tails of the pdf and only small changes therein.

Finally, about 35% of NAM1000$> +2$ and 20% of NAM1000$> +3$ are not preceded by a p-SPV within 28 days. For longer periods, beyond 40 to 50 days, it appears that almost none of the positive NAM1000 events are *not* preceded by a p-SPV. This also explains the only moderate probability increase of strong NAM1000$^+$ extremes following p-SPVs in Fig. 10: In contrast to p-SSWs and subsequent NAM1000$^-$ extremes, the occurrence of NAM1000$^+$ extremes is in general already dominated by preceding p-SPVs. As a result, the increase in frequency is comparatively low relative to the climatology.

For more detailed analyses that apply the diagnostics presented in Fig. 3 and in Fig. 5 to positive AO extremes and p-SPVs, the reader is referred to the supplement.
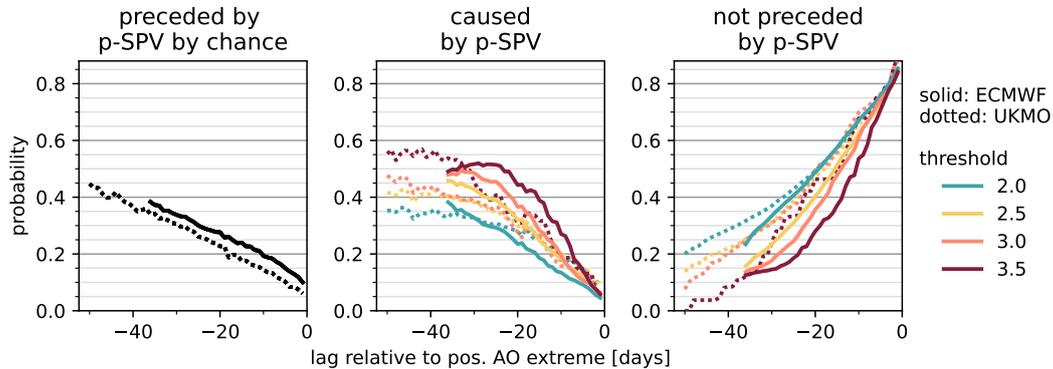
Figure 11. As in Fig. 8, for u60>47 ms^{-1} preceding a p-NAM1000^{+} extreme. Analysis based on ECMWF forecasts.

8 Conclusions

Our results based on a large number of extended-range ensemble forecasts provide further evidence for stratospheric modulation of large-scale weather patterns near the surface, broadly consistent with previous results (Domeisen and Butler, 2020, and references therein). Previous studies generally suffer from relatively small available sample sizes, which hampers estimation of robust statistical relationships between stratospheric and tropospheric extremes (= rare events). Here, by analyzing extended-range forecast periods around predicted extreme events (e.g., p-SSWs), we effectively boost the available sample size by more than a factor of 100 and are therefore in the position to obtain robust estimates in response to our research questions:

1. By how much do stratospheric polar vortex extremes increase the probability of persistently positive or negative AO phases?

   Climatologically, 38% of negative AO phases (days with consecutive NAM1000 < 0) are longer than than 7 days. Following p-SSWs, this is increased to 44%, which corresponds to a relative increase of 16%.

   Following p-SPVs, the probability for positive AO phases that last longer than 7 days is increased from 40% to 44%.

2. By how much do stratospheric polar vortex extremes increase the probability of subsequent AO extremes?

   p-SSWs are followed by AO^{-} extremes significantly more often than expected based on climatology. For instance, AO < −3 events are about 40% (ECMWF forecasts) to about 80% (UKMO forecasts) more likely following p-SSWs. However, the absolute probabilities are still low, i.e., only 3.5% of SSWs are followed by AO < −3 within four weeks, based on ECMWF forecasts (UKMO: 4.5%).

   Following p-SPVs, the probability of an AO > +3 is increased by 16% (ECMWF) to 55% (UKMO), relative to climatology. The probability increase is smaller compared to the increase of negative AO extremes following SSWs, which is a result of the AO^{+}-extreme-climatology itself being dominated by events that follow p-SPVs.

3. How often are AO extremes caused by stratospheric polar vortex extremes?

About one-third of AO < -3 events are, within five weeks and based on our statistical approach, caused by predicted, preceding stratospheric easterlies (u60 < 0). Another one-third of the events is, within five weeks, not preceded by u60 < 0.

420   Within five weeks, about 45% of AO $> +3$ events are caused by a preceding SPV event.

We note that we have used the term "causality" to describe an exceedance probability relative to a climatological baseline. However, this does not rule out the existence of common drivers. For example, Kretschmer et al. (2016) have used Causal Effect Networks to analyze linear pathways that influence the midlatitude winter circulation. They find, e.g., that the AO is correlated to the strength of the polar vortex and also to sea level pressure over the Ural mountains, where the latter is again

425   correlated (with 1 month lag) with the polar vortex strength. In contrast, we focused solely on the direct statistical relationship between extreme states of the stratospheric and tropospheric circulation, with the chosen event-based approach also revealing non-linear relationships.

In this study two forecast models (ECMWF, UKMO) were considered. Given quantitative disagreements in some of the diagnostics, analyses of additional models may help to make definitive quantitative statements.

430   Furthermore, future work should address the question, how much of the predicted surface impact following predicted stratospheric extremes, i.e., following p-SSWs and p-SPVs, can be explained by the NAM1000. Lastly, we conclude that the analysis of predicted events offers potential for improved statistical characterization of other atmospheric extreme events, provided that the forecast model is capable of truthfully representing the event of interest.

Data availability. Forecasts from the S2S archive can be found at https://apps.ecmwf.int/datasets/data/s2s. ERA5 data is available at https:

435   //cds.climate.copernicus.eu/cdsapp#!/dataset/reanalysis-era5-pressure-levels.

Appendix A: Deseasonalization of S2S Forecasts

In addition to realtime forecasts, all S2S forecasting systems create also hindcasts (or "reforecasts"), which allow the construction of the respective model's climatology. In the following, we describe the procedure[5] we applied to compute a climatology of a forecast that starts on some date d (month & day of month).

440   1. Compute the ensemble mean of the hindcasts (Fig. A1a).

2. Compute the inter-annual mean of the hindcast ensemble means. In case of the ECMWF forecasts for example, the hindcasts cover the past 20 years (see Fig. A1b).

---

[5]based on the ECMWF article "Re-forecast for medium and extended forecast range" (https://www.ecmwf.int/en/forecasts/documentation-and-support/extended-range/re-forecast-medium-and-extended-forecast-range, accessed on 23 Aug 2021).

3. Select all (inter-annually averaged) hindcasts that start within ± 14 days relative to the date d (the start of the forecast of interest. In case of the ECMWF model, this selection subsumes 9 (inter-annually averaged) hindcasts, since hindcasts are available for every Monday and Thursday (see Fig. A1c).

4. Average the hindcasts obtained in 3, such that the forecast valid time match (e.g., average forecasts for Feb 22, Feb 23, ... as opposed to matching forecast lead times, e.g., forecasts with lead time +4, +5, ..., see Fig. A1c).

5. Apply, to the resulting time-series, a 7-day running mean filter (Fig. A1d).

6. Due to the ±14 day window, the resulting time-series starts earlier than date d and covers a period that is longer than the forecast of interest. Cut the time-series at the beginning and at the end such that it matches the time-series of the forecast of interest. This gives the climatology (see Fig. A1d).
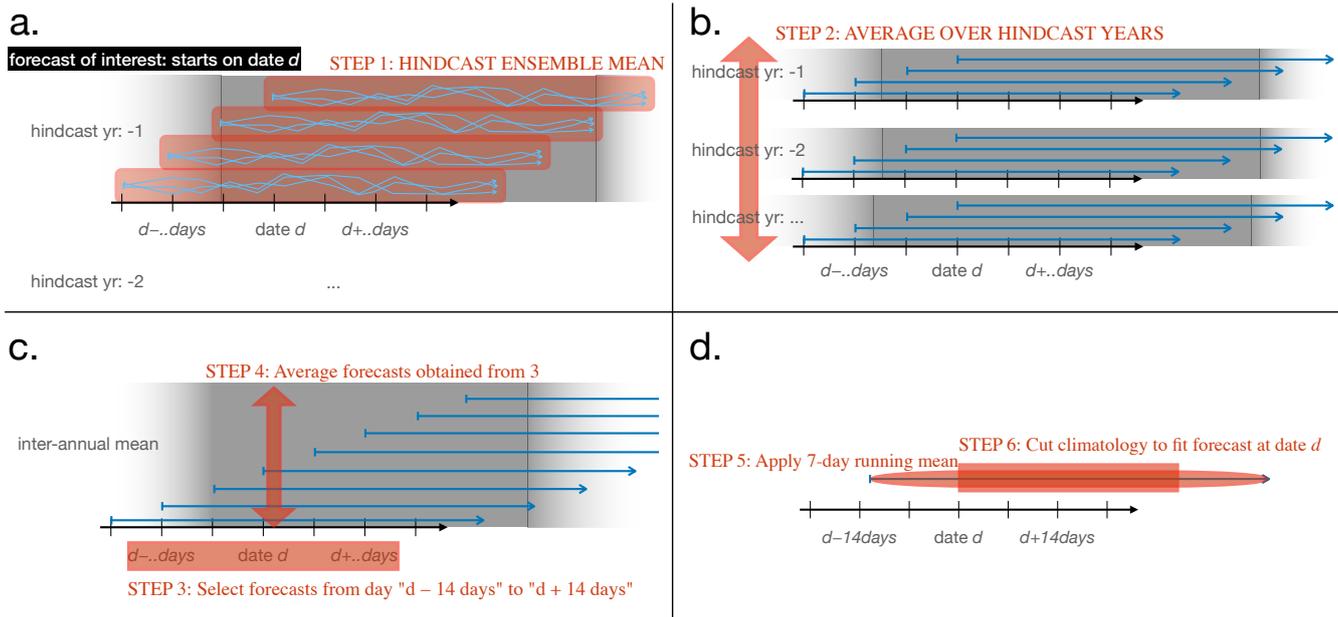

Figure A1. Schematic workflow for the computation of a climatology for a S2S forecast model, based on hindcasts. Gray planes illustrate that forecasts belong to the same hindcast year, where the axis from left to right denotes time and the axis from the front to the back.

Anomalies are obtained by subtracting the climatology from the raw field. Standardized anomalies can be computed by dividing the anomalies through a climatology standard deviation, that is computed as the climatology mean, but where

– *(ad step 1)* instead of the ensemble mean, the unperturbed control run is selected (or any other single ensemble member). Using the ensemble mean would result in a too small inter-annual standard deviation at long forecast lead times (see step 2), because at long lead times, the ensemble mean *always* tends to the climatological mean state.

– *(ad step 2)* instead of the inter-annual mean, the inter-annual standard deviation is computed.

The presented deseasonalization procedure comes with several implications, for example:

- The climatologies for realtime forecasts and for hindcasts are always based only on hindcasts.

460
- By computing anomalies from a climatology, model errors that are a function of the season, are mitigated.

- By computing anomalies from a climatology, model errors that are a function of the forecast lead time ("model drift"), are not mitigated, because the climatology averages information that stems from different forecast lead times (see step 4).

- In case of the ECMWF model, 9 hindcast ensembles / four-week-window · 20 years · 11 ensemble member = 1980
465   integrations contribute to the construction of one climatology.

Author contributions. JS performed the analyses under the guidance of TB. JS wrote the first draft of the manuscript. Both authors contributed to the interpretation of the results and improved the manuscript.

References

475 Albers, J. R. and Birner, T.: Vortex Preconditioning due to Planetary and Gravity Waves prior to Sudden Stratospheric Warmings, Journal of the Atmospheric Sciences, 71, 4028–4054, https://doi.org/10.1175/JAS-D-14-0026.1, 2014.

Ayarzagüena, B., Palmeiro, F. M., Barriopedro, D., Calvo, N., Langematz, U., and Shibata, K.: On the representation of major stratospheric warmings in reanalyses, Atmospheric Chemistry and Physics, 19, 9469–9484, https://doi.org/10.5194/acp-19-9469-2019, 2019.

Baldwin, M. and Dunkerton, T.: Stratospheric harbingers of anomalous weather regimes, Science, 294, 581–584, https://doi.org/10.1126/science.1063315, 2001.

Baldwin, M. P., Ayarzagüena, B., Birner, T., Butchart, N., Butler, A. H., Charlton-Perez, A. J., Domeisen, D. I., Garfinkel, C. I., Garny, H., Gerber, E. P., Hegglin, M. I., Langematz, U., and Pedatella, N. M.: Sudden Stratospheric Warmings, Reviews of Geophysics, 59, 1–37, https://doi.org/10.1029/2020RG000708, 2021.

Butler, A. H., Seidel, D. J., Hardiman, S. C., Butchart, N., Birner, T., and Match, A.: Defining sudden stratospheric warmings, Bulletin of the American Meteorological Society, 96, 1913–1928, https://doi.org/10.1175/BAMS-D-13-00173.1, 2015.

Butler, A. H., Sjoberg, J. P., Seidel, D. J., and Rosenlof, K. H.: A sudden stratospheric warming compendium, Earth System Science Data, 9, 63–76, https://doi.org/10.5194/essd-9-63-2017, 2017.

Cohen, J., Foster, J., Barlow, M., Saito, K., and Jones, J.: Winter 2009–2010: A case study of an extreme Arctic Oscillation event, Geophysical Research Letters, 37, https://doi.org/https://doi.org/10.1029/2010GL044256, 2010.

490 Domeisen, D. I., Butler, A. H., Charlton-Perez, A. J., Ayarzagüena, B., Baldwin, M. P., Dunn-Sigouin, E., Furtado, J. C., Garfinkel, C. I., Hitchcock, P., Karpechko, A. Y., Kim, H., Knight, J., Lang, A. L., Lim, E. P., Marshall, A., Roff, G., Schwartz, C., Simpson, I. R., Son, S. W., and Taguchi, M.: The Role of the Stratosphere in Subseasonal to Seasonal Prediction: 1. Predictability of the Stratosphere, Journal of Geophysical Research: Atmospheres, 125, 1–17, https://doi.org/10.1029/2019JD030920, 2020a.

Domeisen, D. I., Butler, A. H., Charlton-Perez, A. J., Ayarzagüena, B., Baldwin, M. P., Dunn-Sigouin, E., Furtado, J. C., Garfinkel, C. I., 495 Hitchcock, P., Karpechko, A. Y., Kim, H., Knight, J., Lang, A. L., Lim, E. P., Marshall, A., Roff, G., Schwartz, C., Simpson, I. R., Son, S. W., and Taguchi, M.: The Role of the Stratosphere in Subseasonal to Seasonal Prediction: 2. Predictability Arising From Stratosphere-Troposphere Coupling, Journal of Geophysical Research: Atmospheres, 125, 1–20, https://doi.org/10.1029/2019JD030923, 2020b.

Domeisen, D. I., Grams, C. M., and Papritz, L.: The role of North Atlantic-European weather regimes in the surface impact of sudden stratospheric warming events, Weather and Climate Dynamics Discussions, 2020, 1–24, https://doi.org/10.5194/wcd-2019-16, 2020c.

500 Domeisen, D. I. V. and Butler, A. H.: Stratospheric drivers of extreme events at the Earth's surface, Communications Earth & Environment, 1, 59, https://doi.org/10.1038/s43247-020-00060-z, 2020.

Gerber, E. P., Orbe, C., and Polvani, L. M.: Stratospheric influence on the tropospheric circulation revealed by idealized ensemble forecasts, Geophysical Research Letters, 36, https://doi.org/10.1029/2009GL040913, 2009.

Haeseler, S., Bissolli, P., Daßler, J., Zins, V., and Kreis, A.: Orkantief SABINE löst am 9 ./ 10. Februar 2020 eine 505 schwere Sturmlage über Europa aus, DWD (Hrsg.), pp. 1–9, https://www.dwd.de/DE/leistungen/besondereereignisse/stuerme/ 20200213{_}orkantief{_}sabine{_}europa.html, 2020.

Hersbach, H., Bell, B., Berrisford, P., Hirahara, S., Horányi, A., Muñoz-Sabater, J., Nicolas, J., Peubey, C., Radu, R., Schepers, D., Sim-mons, A., Soci, C., Abdalla, S., Abellan, X., Balsamo, G., Bechtold, P., Biavati, G., Bidlot, J., Bonavita, M., De Chiara, G., Dahlgren, P., Dee, D., Diamantakis, M., Dragani, R., Flemming, J., Forbes, R., Fuentes, M., Geer, A., Haimberger, L., Healy, S., Hogan, R. J., 510 Hólm, E., Janisková, M., Keeley, S., Laloyaux, P., Lopez, P., Lupu, C., Radnoti, G., de Rosnay, P., Rozum, I., Vamborg, F., Vil-

laume, S., and Thépaut, J.-N.: The ERA5 global reanalysis, Quarterly Journal of the Royal Meteorological Society, 146, 1999–2049, https://doi.org/https://doi.org/10.1002/qj.3803, 2020.

Hitchcock, P. and Simpson, I. R.: The Downward Influence of Stratospheric Sudden Warmings, Journal of the Atmospheric Sciences, 71, 3856–3876, https://doi.org/10.1175/JAS-D-14-0012.1, 2014.

515   Jucker, M.: Are sudden stratospheric warmings generic? Insights from an idealized GCM, Journal of the Atmospheric Sciences, 73, 5061–5080, https://doi.org/10.1175/JAS-D-15-0353.1, 2016.

Jucker, M. and Reichler, T.: Dynamical Precursors for Statistical Prediction of Stratospheric Sudden Warming Events, Geophysical Research Letters, 45, 13,124–13,132, https://doi.org/10.1029/2018GL080691, 2018.

Karpechko, A. Y.: Predictability of sudden stratospheric warmings in the ECMWF extended-range forecast system, Monthly Weather Review, 520   146, 1063–1075, https://doi.org/10.1175/MWR-D-17-0317.1, 2018.

Karpechko, A. Y., Hitchcock, P., Peters, D. H., and Schneidereit, A.: Predictability of downward propagation of major sudden stratospheric warmings, Quarterly Journal of the Royal Meteorological Society, https://doi.org/10.1002/qj.3017, 2017.

Kautz, L. A., Polichtchouk, I., Birner, T., Garny, H., and Pinto, J. G.: Enhanced extended-range predictability of the 2018 late-winter Eurasian cold spell due to the stratosphere, Quarterly Journal of the Royal Meteorological Society, 146, 1040–1055, https://doi.org/10.1002/qj.3724, 525   2020.

Kim, J.-S., Kug, J.-S., Jeong, S.-J., Park, H., and Schaepman-Strub, G.: Extensive fires in southeastern Siberian permafrost linked to preceding Arctic Oscillation, Science Advances, 6, https://doi.org/10.1126/sciadv.aax3308, 2020.

Kretschmer, M., Coumou, D., Donges, J. F., and Runge, J.: Using causal effect networks to analyze different arctic drivers of midlatitude winter circulation, Journal of Climate, 29, 4069–4081, https://doi.org/10.1175/JCLI-D-15-0654.1, 2016.

530   Lawrence, Z. D., Perlwitz, J., Butler, A. H., Manney, G. L., Newman, P. A., Lee, S. H., and Nash, E. R.: The Remarkably Strong Arctic Stratospheric Polar Vortex of Winter 2020: Links to Record-Breaking Arctic Oscillation and Ozone Loss, Earth and Space Science Open Archive, p. 27, https://doi.org/10.1002/essoar.10503356.1, 2020.

Lee, S. H., Furtado, J. C., and Charlton-Perez, A. J.: Wintertime North American Weather Regimes and the Arctic Stratospheric Polar Vortex, Geophysical Research Letters, 46, 14 892–14 900, https://doi.org/10.1029/2019GL085592, 2019.

535   McIntyre, M.: How Well do we Understand the Dynamics of Stratospheric Warmings, Journal of the Meteorological Society of Japan, 60, 37–65, 1982.

Rupp, P., Loeffel, S., Garny, H., Chen, X., Pinto, J. G., and Birner, T.: Potential links between tropospheric and stratospheric circulation extremes during early 2020, Earth and Space Science Open Archive, p. 37, https://doi.org/10.1002/essoar.10507772.1, 2021.

Taguchi, M.: A study of false alarms of a major sudden stratospheric warming by real-time subseasonal-to-seasonal forecasts for the 540   2017/2018 Northern Winter, Atmosphere, 11, https://doi.org/10.3390/ATMOS11080875, 2020.

Thompson, D. W. and Wallace, J. M.: The Arctic oscillation signature in the wintertime geopotential height and temperature fields, Geophysical Research Letters, 25, 1297–1300, https://doi.org/10.1029/98GL00950, 1998.

Thompson, D. W. J. and Wallace, J. M.: Regional Climate Impacts of the Northern Hemisphere Annular Mode, Science, 293, 85–89, https://doi.org/10.1126/science.1058958, 2001.

545   Vitart, F., Ardilouze, C., Bonet, A., Brookshaw, A., Chen, M., Codorean, C., Déqué, M., Ferranti, L., Fucile, E., Fuentes, M., Hendon, H. H., Hodgson, J., Kang, H. S., Kumar, A., Lin, H., Liu, G., Liu, X., Malguzzi, P., Mallas, I., Manoussakis, M., Mastrangelo, D., MacLachlan, C., McLean, P., Minami, A., Mladek, R., Nakazawa, T., Najm, S., Nie, Y., Rixen, M., Robertson, A. W., Ruti, P., Sun, C., Takaya, Y., Tolstykh, M., Venuti, F., Waliser, D., Woolnough, S., Wu, T., Won, D. J., Xiao, H., Zaripov, R., and Zhang, L.: The subseasonal to seasonal (S2S)

prediction project database, Bulletin of the American Meteorological Society, 98, 163–173, https://doi.org/10.1175/BAMS-D-16-0017.1,
550   2017.

Waugh, D. W., Sobel, A. H., and Polvani, L. M.: What is the polar vortex and how does it influence weather?, Bulletin of the American
Meteorological Society, 98, 37–44, https://doi.org/10.1175/BAMS-D-15-00212.1, 2017.

White, I. P., Garfinkel, C. I., Gerber, E. P., Jucker, M., Hitchcock, P., and Rao, J.: The Generic Nature of the Tropospheric Response to Sudden
Stratospheric Warmings, Journal of Climate, 33, 5589–5610, https://doi.org/10.1175/JCLI-D-19-0697.1, 2020.