Tracing North Atlantic Oscillation Forecast Errors to Stratospheric Origins

ERIK W. KOLSTAD

NORCE Norwegian Research Center, Bjerknes Center for Climate Research, Bergen, Norway

C. OLE WULFF AND DANIELA I. V. DOMEISEN

Institute for Atmospheric and Climate Science, ETH Zurich, Zurich, Switzerland

TIM WOOLLINGS

Department of Physics, University of Oxford, Oxford, United Kingdom

(Manuscript received 16 April 2020, in final form 2 July 2020)

ABSTRACT

The North Atlantic Oscillation (NAO) is the main driver of weather variability in parts of Eurasia, Greenland, North America, and North Africa on a range of time scales. Successful extended-range NAO predictions would equate to improved predictions of precipitation and temperature in these regions. It has become clear that the NAO is influenced by the stratosphere, but because this downward coupling is not fully reproduced by all forecast models the potential for improved NAO forecasts has not been fully realized. Here, an analysis of 21 winters of subseasonal forecast data from the European Centre for Medium-Range Weather Forecasts monthly forecasting system is presented. By dividing the forecasts into clusters according to their errors in North Atlantic Ocean sea level pressure 15–30 days into the forecasts, we identify relationships between these errors and the state of the stratospheric polar vortex when the forecasts were initialized. A key finding is that the model overestimates the persistence of both the negative NAO response following a weak polar vortex and the positive NAO response following a strong polar vortex. A case in point is the sudden stratospheric warming in early 2019, which was followed by five consecutive weeks of an overestimation of the negative NAO regime. A consequence on the ground was temperature predictions for northern Europe that were too cold. Another important finding is that the model appears to misrepresent the gradual downward impact of stratospheric vortex anomalies. This result suggests that an improved representation and prediction of stratosphere–troposphere coupling in models might yield substantial benefits for extended-range weather forecasting in the Northern Hemisphere midlatitudes.

KEYWORDS: North Atlantic Oscillation; Stratosphere–troposphere coupling; Climate prediction; Seasonal forecasting

1. Introduction

The North Atlantic Oscillation (NAO), which was described qualitatively in the latter half of the 18th century and more quantitatively using weather observations about 100 years later (e.g., van Loon and Rogers 1978; Greatbatch 2000; Hurrell et al. 2001; Visbeck et al. 2001; Wanner et al. 2001, and references therein), is the dominant driver of atmospheric variability over the

extratropical North Atlantic Ocean. Usually defined as a meridional seesaw pattern of atmospheric pressure between a northern region near Iceland and a southern area roughly between the Azores and Portugal, the NAO varies on daily to decadal time scales (Thompson and Wallace 2000; Woollings et al. 2014) and affects weather in North Atlantic coastal regions and beyond. The negative phase of the NAO is characterized by a blocking high in the northern part of the North Atlantic (Crocì-Maspoli et al. 2007) and has wide-reaching impacts on the ground, including anomalously cold and dry weather in northern Europe and warm, wet anomalies over the Mediterranean region. At the other end of the scale, the positive NAO phase is typically accompanied

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

Corresponding author: Erik W. Kolstad, ekol@norce-research.no

by lower-than-normal pressure near Iceland and a jet stream curving northward into northern Europe, bringing warm and stormy weather (Hurrell and Deser 2010).

Because the NAO is so important for the weather in large parts of the Northern Hemisphere's midlatitudes, a substantial body of research has been devoted to NAO forecasting in recent years. Even though an increase in NAO prediction skill does not necessarily translate to an enhanced ability to forecast surface variables (Scaife et al. 2014; Athanasiadis et al. 2017), the NAO is often used as a benchmark for dynamical prediction models.

While it has emerged that the NAO has a potentially predictable deterministic component on time scales of 2–3 weeks (Domeisen et al. 2018; Zhang et al. 2019), there is also significant skill in predicting the NAO on subseasonal (Vitart 2014), seasonal (Weisheimer et al. 2020), annual (Dunstone et al. 2016) and decadal (Smith et al. 2019) time scales, although the models lay bare a marked underconfidence in predicting the NAO at seasonal and longer lead times (Baker et al. 2018).

To enhance the NAO forecast skill, it is important to understand the mechanisms that dictate its variability. The fluctuations of the NAO between positive, neutral and negative phases are partly generated by internal atmospheric variability (DeWeaver and Nigam 2000), but they also respond to remote forcing (e.g., Marshall et al. 2001; Osborn 2004; Cui et al. 2014). In particular, the stratosphere has been shown to strongly affect NAO variability (Thompson and Wallace 1998). Periods during which the stratospheric polar vortex is unusually weak, sudden stratospheric warmings (SSWs; Butler et al. 2017), tend to coincide with a nudging of the NAO toward its negative phase (Charlton-Perez et al. 2018) and an increased persistence of this phase (Domeisen 2019). On the ground, this yields a higher likelihood of e.g., cold-air outbreaks in northern Europe (Kolstad et al. 2010). At the other end of the spectrum, strong vortex events are frequently followed by a positive phase of the NAO (Baldwin and Dunkerton 2001).

As a result of its influence on the NAO, the stratosphere has been shown to contribute skill in the North Atlantic–European region on subseasonal (Tripathi et al. 2015; Butler et al. 2019; Choi and Son 2019; Domeisen et al. 2020c) to seasonal (Sigmond et al. 2013; Scaife et al. 2016; Dobrynin et al. 2018; O'Reilly et al. 2019; Nie et al. 2019) time scales. But despite the demonstrated increase in NAO skill after SSW events, not all stratospheric events contribute to this skill. Although there is variation between SSW events (Domeisen et al. 2020a), only roughly two thirds of SSWs are followed by a persistent negative NAO response (Charlton-Perez et al. 2018; Domeisen 2019), while the remaining third are followed by a positive NAO response (Afargan-Gerstman and Domeisen 2020).

The purpose of the study presented here was to investigate North Atlantic SLP errors in the European Centre for Medium-Range Weather Forecasts (ECMWF) monthly forecasting system for lead times on the subseasonal time scale (15–30 days). When we divided the errors into three clusters, it became evident that two of the clusters projected onto too-negative or too-positive NAO conditions. By relating these errors to the state of the stratospheric polar vortex when the model forecasts were initialized, we identified a weak but general relationship between NAO errors and the initial vortex state. The largest NAO errors were found to arise when the model failed to capture the diversity of tropospheric responses to anomalous stratospheric vortex states, instead overpredicting the “standard” NAO response to both SSWs and strong vortex events. This prediction bias negatively impacts the models' prediction skill after anomalous stratospheric vortex states, which could otherwise have greatly contributed to improved subseasonal prediction skill of the winter troposphere.

2. Data and methods

a. Data sources

The model system used here is the ECMWF monthly forecasting system, which has two components: a real-time forecasting system with 51 ensemble members, and hindcasts with 11 members. Each ensemble member is a 46-day coupled ocean–atmosphere integration. The real-time forecasts are initialized on each Monday and Thursday, and we used all the 30 ensemble forecasts that were initialized between 19 November 2018 and 28 February 2019. The corresponding hindcasts were initialized on the same dates each year from 1998/99 to 2017/18, giving us a total of 21 winters when combining the hindcasts and forecasts, for a total of $30 \times 21 = 630$ predictions. Because we only consider ensemble means, we do not distinguish between hindcasts and forecasts and from now on refer to them simply as “forecasts.” Unless otherwise specified, we used time-averaged values for the forecast lead times between 15 and 30 days.

The model version investigated here is CY45R1, which became operational on 6 June 2018. The horizontal grid spacing of the model is 16 km at lead times of up to 15 days and 32 km thereafter. The forecasting system is a high-top model (model top around 0.01 hPa) and exhibits high resolution in the stratosphere (38 levels at 100 hPa and above), which is among the best resolutions among subseasonal prediction systems that provide stratospheric data (Domeisen et al. 2020b). The data were downloaded from the S2S database (Vitart et al. 2017), where instantaneous data for 0000 UTC on

FIG. 1. Average forecast error (forecast minus reanalysis) for clusters (a) A, (b) B, and (c) C, averaged for lead times from 15 to 30 days. The unit is standard deviations (σ), because both the forecasts and the reanalysis were standardized separately. The number of forecasts in each cluster is given in the titles, with the total number of forecasts in parentheses. The stippling indicates errors that are significantly different from zero at the 5% level (details are in the text).

each day were available at a resolution of 1.5° . The following variables were downloaded: SLP, 2-m temperature, zonal wind at 10 hPa, and geopotential height at 50, 100, and 200 hPa. The forecasts are compared with ERA5 reanalysis data (Hersbach et al. 2020), from which the same variables were downloaded on the same 1.5° grid as the forecasts.

Prior to performing the analysis described below, we converted both reanalysis and forecast values to dimensionless anomalies by first subtracting the datewise climatological mean and then dividing by the datewise climatological standard deviation. For the forecasts, we applied this standardization to the ensemble means for each lead time.

b. Methods

1) CLUSTERING

For each of the 630 forecasts, the average SLP for lead times between 15 and 30 days was computed for each grid point inside a subregion of the North Atlantic (25° – 75° N, 80° W– 10° E). From ERA5 data, the average SLP for the same days was also calculated, and then 630 maps of the differences between the two (forecast minus reanalysis) were compiled. As mentioned, the differences were evaluated using standardized and dimensionless SLP values. The *k*-means (MacQueen 1967) implementation in the Scikit-learn Python library (Pedregosa et al. 2011) was then used to divide the difference maps into clusters. The number of clusters is a subjective choice, but it had little bearing on the results presented here. Many alternatives were tried, ranging from 2 to 10 clusters. Two clusters with oppositely signed SLP errors with NAO-like

spatial structures always emerged, and the spatial signatures of these did not change appreciably with the number of total clusters. This is probably because the NAO explains a large share of the SLP variance in the North Atlantic. The results presented here are based on three clusters. As will be shown below (Fig. 1), this yielded two largely symmetric clusters that projected onto an NAO-like pattern and a third cluster without a distinct spatial signature.

2) STRATOSPHERIC SUDDEN WARMINGS

The major SSW central dates in the study period were identified using the zonal mean of ERA5 10-hPa zonal wind at 60° N. Henceforth, we refer to this wind speed (m s^{-1}) as $\bar{u}$ and its dimensionless standardized version, which was used to neutralize the seasonal cycle, as $\bar{u}_\sigma$. Following Charlton and Polvani (2007), a central date is the first day that $\bar{u} < 0$. To avoid recording the same event twice, we also required that $\bar{u} > 0$ for 20 consecutive dates before the central date. The SSW central dates during the study period from November 1998 to February 2019 are listed in Table 1, along with the central dates in Butler et al. (2017). Note that the two missing central dates, numbers 29 and 40 in Butler et al. (2017), both occurred in March, that is, outside our seasonal scope.

3) NAO AND NAM INDEX

The NAO index was calculated as follows. For each date, the area-averaged SLP was computed for two regions: all grid points within 500 km of Stykkisholmur in Iceland, and all grid points within 500 km of Lisbon, Portugal. This yielded time series of SLP for each region.

TABLE 1. The major SSW central dates derived here, and the corresponding number in Table 1 of [Butler et al. \(2017\)](#).

Butler et al. No.	ERA5 date
27	16 Dec 1998
28	26 Feb 1999
30	12 Feb 2001
31	31 Dec 2001
32	18 Feb 2002
33	18 Jan 2003
34	5 Jan 2004
35	21 Jan 2006
36	24 Feb 2007
37	22 Feb 2008
38	25 Jan 2009
39	9 Feb 2010
41	7 Jan 2013
—	12 Feb 2018
—	2 Jan 2019

As in [Hurrell \(1995\)](#), these time series were standardized separately for each region to avoid domination by the “northern” time series, which has a considerably higher variance than the “southern” time series. Then the difference between the southern and the northern time series was calculated to form a new time series, and the standardized version of this was used as our NAO index.

To assess the forecast errors lower in the stratosphere and near the tropopause level, we calculated an indicator of the Northern Hemisphere Annular Mode (NAM), which is related to the strength of the polar vortex and the NAO index, at 50, 100, and 200 hPa. We computed a NAM index as -1 times the standardized area-averaged geopotential height anomalies north of 65°N . This index is equivalent to the NAM index obtained with more complex methods based on principal component analysis ([Baldwin and Thompson 2009](#)).

4) SIGNIFICANCE TESTING

As we computed composites for 16-day periods and the ECMWF forecasts were initialized twice a week, many of the composite periods overlapped. To account for the resulting autocorrelation, we used a two-sided bootstrapping approach to calculate the statistical significance of the composites in [Figs. 1, 2, 4a, and 4b](#) (note that [Fig. 4](#) is discussed in more detail below). The null hypothesis was that each set of cluster members within a given winter could just as well have occurred in any of the 21 winters. To test this, 10 000 synthetic composites were calculated as follows. For cluster x , the set of forecast initialization dates belonging to winter y were retained, but y was replaced by a random winter out of the 21 possible winters (with replacement). As an example, assume a hypothetical cluster of 10 forecasts. Further assume that five of these forecasts were initialized in the winter of 2000/1 and the other five were initialized in 2012/13. Each of the 10 000 synthetic composites would consist of five forecasts from a random winter between 1998/99 and 2018/19 (initialized on the same days and months as the ones in 2000/1) and five forecasts from another random winter in the same time interval (initialized on the same days and months as the ones in 2018/19). This procedure ensured that each of the synthetic composites for x had the same intraseasonal autocorrelation. If the actual composite value was lower than the 2.5th or greater than the 97.5th percentile of the set of synthetic composites, we defined the value to be statistically significant at the 5% level.

Bootstrapping was also used to find out how many forecasts each bin in [Fig. 4c](#) might have contained by chance if they had been independent of $\bar{u}_{\sigma}$. We created a set of 10 000 forecasts for which the distribution of the 30

FIG. 2. The two components of the mean forecast errors shown in [Fig. 1](#), averaged for (a),(b) cluster A and (c),(d) cluster B for (left) forecasts and (right) reanalysis values. The unit is standard deviations (σ), and the stippling indicates anomalies that are significantly different from zero at the 5% level (details are in the text).

FIG. 3. Time series of the average NAO index for (a) cluster A and (b) cluster B for lead times of 1–30 days, calculated for the forecasts (thin lines) and the reanalysis (thick lines). The zero line of the y axis is accentuated by a horizontal thin, dashed line. The circles indicate lead times for which the difference between the forecast and the reanalysis composites is significantly different from zero at the 5% level.

forecasts into clusters for each year was swapped with the distribution from a random year out of the possible 21 years (with replacement). This method retained the autocorrelation of the original distribution while at the same time the cluster distribution was randomized. We defined the number of forecasts in a cluster to be significantly different from random if it was lower than the 2.5th percentile or higher than the 97.5th percentile of this random set.

A similar method was used to calculate the significance of the composite average temperature anomalies in Figs. 8b and 9b (discussed in more detail below), where a set of 10 000 sequences consisting of 11 consecutive forecasts were drawn randomly from the forecasts. The anomalies were defined to be statistically significantly different from random if they were either less than the 2.5th or greater than the 97.5th percentiles of this set.

3. Results

a. Forecast error clusters

As described in the previous section, the 630 North Atlantic SLP error maps (forecast minus reanalysis), averaged for lead times between 15 and 30 days, were divided into three clusters. In Fig. 1, the mean forecast errors are shown for each of the clusters, named A, B, and C. The errors in clusters A and B project onto the typical spatial pattern of the NAO. Figure 1a shows that the forecasts in cluster A are in a too-negative phase of the NAO, with too low pressure in a wide belt stretching from the subtropics to the Iberian Peninsula and too high pressure over Greenland and Iceland. Although the clustering algorithm only considered grid points inside the outlined region, the forecast errors are significant

and positive over most of the polar cap north of 60°N. Significant and negative errors also occur over large parts of the Pacific. Figure 1b shows that the NAO in the cluster-B forecasts is too positive. Overall, there is substantial symmetry between the errors in clusters A and B, although the errors in cluster A cover a larger geographical region than the errors in cluster B. The total number of forecasts in A and B is 355 (56% of the 630 forecasts). The average of the forecast errors in cluster C is small, as shown in Fig. 1c, indicating that there is no systematic error pattern in this set of forecasts. The remainder of the study is mainly focused on clusters A and B.

The SLP anomalies in the forecasts and the reanalysis corresponding to the forecast errors shown in Fig. 1 are now studied separately. The mean forecast SLP anomaly for cluster A has a distinctly NAO-negative signature after 15–30 days (Fig. 2a), but what actually occurred during said period was that the NAO was weakly positive (Fig. 2b). In cluster B, the mean forecast SLP anomaly corresponds to a positive NAO (Fig. 2c), while the mean reanalysis anomaly is that of a negative NAO (Fig. 2d). Hence, on average, these clusters comprise forecasts in which the model predicted the incorrect sign of the NAO.

We now show that the divergence of the NAO phase between forecasts and observations is clear already from a few days into the forecasts. In Fig. 3, the NAO index averaged over all members in each cluster is shown for each lead time between 1 and 30 days. Over the first few days in the cluster-A cases, the mean NAO index is decreasing in both the reanalysis and the forecasts (Fig. 3a). After about five days, the mean NAO index in the reanalysis starts to increase and eventually becomes positive, while the mean forecast NAO index continues to decrease and stays negative throughout the period. The differences between the forecast and the

FIG. 4. Average zonal wind anomalies at 10 hPa on the dates that the forecasts in (a) cluster A and (b) cluster B were initialized. The unit is standard deviations (σ), and the stippling indicates anomalies that are significantly different from zero at the 5% level (details are in the text). The dashed lines show 60°N. (c) The percentage of forecasts in each cluster in five equally sized bins, organized according to the initial vortex strength ($\bar{u}_\sigma$). The thick outlines indicate percentages that are significantly (at the 5% level) different from what could be expected by chance.

reanalysis is significant from one week and onward. For the forecasts in cluster B, there is a similar divergence between the reanalysis and the forecast NAO index (Fig. 3b). In both clusters, the forecast model tends to persist and enhance the initial phase of the NAO, while in the reanalysis, the NAO index switches sign on average between 11 and 14 days after initialization.

The cluster analysis identified sets of forecasts with large ensemble mean error in the NAO in days 15–30. Importantly, the emergence of consistent signals early in the forecast period, as shown in Fig. 3, reveals that these are not simply random errors, as might be expected in a system with low intrinsic predictability. Instead, there is a consistent dynamical setup in these cases that suggests a systematic bias in the model under these situations. Further evidence of the common structures within these two clusters is presented in the following section.

b. Linkages to initial vortex states

We now consider the state of the stratospheric polar vortex on the initialization dates of the forecasts in clusters A and B. Figure 4a shows the mean 10-hPa zonal wind anomalies on the 168 initialization dates in cluster A. The significant and negative mean anomalies in a wide belt on both sides of 60°N (the usual reference latitude for the polar vortex at 10 hPa) show that, on

average, the polar vortex was weaker than normal on these dates. On the initialization dates of the 187 forecasts in cluster B, the average vortex was slightly stronger than normal, but the composite average zonal wind anomalies are not statistically significant (Fig. 4b).

The histogram in Fig. 4c is an alternative visualization of the link between the vortex states on initialization and the SLP forecast errors. The 630 forecasts were divided into five equally sized bins according to $\bar{u}_\sigma$ (the standardization was done to account for the seasonal cycle of $\bar{u}$). Each bin contains 126 forecasts, and the height of each bar indicates the percentage of the total number of forecasts in each cluster.

First, we note that the cluster-C percentages, which are shown here to contrast with the ones for clusters A and B, do not significantly deviate from what could have occurred by chance. By contrast, 53 (32%) of the 168 cluster-A forecasts belong to the bin with the weakest initial vortex states. This fraction is significantly different from what could be expected by chance. The weak vortex bin also contains significantly fewer cluster-B forecasts than a randomly distributed set. The number of cluster-A members in each bin decreases gradually with increasing vortex strength. The last bin, which consists of the forecasts with strong initial vortex states, only contains 16 (10%) of the cluster-A members, that

is, less than one-third of the number of cluster-A forecasts in the weak vortex bin. This is significantly fewer than expected by chance. The fourth bin, which consists of forecasts that were initialized when the vortex was between average and strong, contains significantly more cluster-B forecasts than expected by chance.

There is clearly a relationship between North Atlantic SLP errors and the polar vortex when the initial state of the vortex is either weak or strong, although a large share of the 630 SLP forecasts do not appear to be directly linked to the initial vortex state. We now take a small detour from the clusters and relate the initial vortex strength to subsequent errors in the NAO index, which is the most familiar indicator of SLP variance in the extratropical North Atlantic. In Fig. 5, all the 630 NAO forecast errors at lead times of 15–30 days are plotted against $\bar{u}_\sigma$ when the forecasts were initialized. The correlation coefficient is 0.25, and this is statistically significant at the 5% level, according to a bootstrapping test that takes the intraseasonal autocorrelation of the time series into account. The correlation is even significant for the cluster-C members alone (the gray circles in the scatterplot). This is a powerful result, which shows that there is a *general* relationship between NAO forecast errors and the initial vortex state.

c. Downward impact

The results in the previous section made it clear that the initial state of the stratospheric polar vortex is linked to SLP forecast errors in the North Atlantic region between 15 and 30 days into the forecasts. It is possible that the NAO forecast errors can be explained by an erroneous representation of the downward impact of stratospheric vortex anomalies. To test this, we calculated zonal wind anomalies at 10 hPa and geopotential height anomalies at 50, 100, and 200 hPa in both the model and the reanalysis during the period from 15 to 30 days after the forecasts in each cluster had been initialized.

As we already know from Fig. 4a, the initial polar vortex is weaker than normal in the cluster-A forecasts (Fig. 6a). With increasing lead time, the negative $\bar{u}_\sigma$ anomalies diminish in magnitude. It is remarkable how closely the forecast $\bar{u}_\sigma$ anomalies follow the reanalysis. The differences are nonsignificant at all lead times. The same applies to the $\bar{u}_\sigma$ anomalies for the cluster-B cases (Fig. 6a). In summary, this shows that the mean zonal wind forecast at 10 hPa is successful for both cluster A and cluster B at lead times of up to 30 days.

Figure 6c shows how the cluster-A NAM index forecast errors evolve with increasing lead time. Already 8 days into the forecast, the error becomes significant at all the three levels. At lead times longer than about 10 days, the errors are consistently smaller at 50 hPa

FIG. 5. NAO forecast errors 15–30 days after initialization plotted against ERA5 zonal mean zonal wind at 10 hPa and 60°N ($\bar{u}_\sigma$) at the time of initialization. Each point is colored by the cluster into which its respective forecast's subseasonal SLP error falls, using the same color scheme as in Fig. 4c. The r value in the top-left corner indicates the correlation between the NAO error and $\bar{u}_\sigma$ for all 630 forecasts.

than at 100 hPa, and smaller at 100 hPa than at 200 hPa. As was seen for the NAO index in Fig. 3a, the 200-hPa NAM index in the reanalysis gradually becomes less negative with increasing lead time, whereas the forecast NAM index first becomes more negative and then stays negative. Similar evolutions, although less pronounced, are seen for the NAM index at 50 and 100 hPa. In the cluster-B cases shown in Fig. 6d, the forecast and reanalysis-based NAM index values also diverge after about 10 days, and the mean forecast error grows with decreasing height, just as it did in the cluster-A cases.

d. Persistence of forecast errors

A timeline of how the forecasts are distributed among the clusters is shown in Fig. 7. The forecasts have a tendency to belong to the same cluster for multiple subsequent initialization dates. Cluster B has the highest persistence, with 64% of cluster-B forecasts immediately followed by forecasts in the same cluster. The corresponding fractions for cluster A and cluster C are 57% and 59%, respectively. Note, however, that these percentages should only be interpreted relative to each other, since some persistence is expected merely due to the fact that we consider 16-day means of forecasts that are separated by only 3 or 4 days. Figure 7 also shows the dates on which $\bar{u} < 0$, along with the major SSW central

FIG. 6. Mean evolution of standardized anomalies of (a),(b) $\bar{u}_\sigma$ at 10 hPa and 60°N and (c),(d) the NAM index at three levels in the lower stratosphere for composite evolutions averaged over the forecasts in clusters (left) A and (right) B, the two main clusters of SLP errors discussed in the text. Thick solid lines show the composite mean of ERA5, and thin lines show that of the forecasts. The zero line is indicated by a dashed line in all panels. (In (c) and (d), the NAM evolution is shown at the 50-, 100-, and 200-hPa levels. In all panels, the circles indicate where the difference between the forecasts and ERA5 is statistically significant at the 5% level.

dates. Of the total number of dates in the analysis period, $\bar{u}$ was negative on 8%. To identify an equal number of weak and strong vortex dates, we defined strong vortex dates as the dates on which $\bar{u}_\sigma$ was higher than its overall 92nd percentile. These dates are also marked in Fig. 7.

Roughly one-half (23 of 50) of the forecasts that were initialized when $\bar{u} < 0$ belong to cluster A, even though the cluster-A cases only constitute about 27% of the total number of forecasts. The major SSW in the beginning of January 2019 was followed by 11 consecutive forecasts in cluster A (corresponding to more than 5 weeks, as there are two forecasts per week). Another example of long cluster-A sequences following an SSW occur after the SSW in January 2009.

Cluster B makes up 17 of the 47 forecasts (36%) that were initialized on strong vortex dates. But as the cluster-B forecasts make up 30% of the total, this is not remarkable. It is more revealing that only 4 of the 47 strong vortex forecasts belong in cluster A. Two sequences of 10 or more consecutive cluster-B forecasts occur in 2012/13 and 2006/7, but these are not linked to strong initial vortex states. The long run with mainly cluster-B forecasts early in the winter of 2015/16 is, however, associated with a long period of strong vortex dates.

In the next section, the 2018/19 and 2015/16 winters are used to illustrate how cluster-A and cluster-B SLP forecast errors relate to regional temperature forecast errors.

e. Links to temperature forecast errors

To show how the forecast errors evolved during the 2018/19 winter, we calculated the NAO index for the forecast and the reanalysis, averaged for lead times of 15–30 days after each of the forecast initialization dates. These NAO index time series are shown in Fig. 8a, along with the standardized $\bar{u}_\sigma$ anomaly on initialization. The major SSW central date (2 January 2019) is indicated by a vertical line. The shaded area denoting the difference between the forecast and the reanalysis-derived NAO index shows that the forecasts were very good prior to the SSW, but after this the forecast NAO index was far too low. As expected, the too-negative NAO forecasts belong to cluster A, as indicated by the blue circles along the bottom of the graph.

The average 2-m temperature anomalies 15–30 days into the 11 cluster-A forecasts are shown in the map in Fig. 8b, and the geographical pattern clearly resembles the typical temperature anomaly pattern accompanying

FIG. 7. The circles show to which of the three clusters each of the 630 forecasts belongs. The initialization date of each forecast is indicated on the x axis, and the corresponding winter seasons are indicated on the y axis. The star symbols indicate the SSW central dates in the period, and the vertical lines show dates for which the zonal mean zonal wind at 60°N was below zero. The plus signs indicate dates for which $\bar{u}_\sigma$ was higher than its 92nd percentile (see the text for details).

the negative phase of the NAO, which consists of anomalously cold temperatures in northern Europe and on the U.S. East Coast, in tandem with anomalously warm temperatures in North Africa and in the northwest Atlantic (e.g., van Loon and Rogers 1978; Domeisen et al. 2020c).

To provide some context of just how large the forecast errors were in this case in a specific region, we calculated the area-averaged 2-m temperature errors at lead times between 15 and 30 days inside the northern European region outlined in Fig. 8b. The first 10 of the 11 cluster-A forecasts after the major SSW were too cold in this region. The largest error occurred for the forecast initialized on 24 January 2019, when the forecast was 2.2σ colder than the reanalysis. To put this in perspective, this was the third-largest negative northern European temperature forecast error following major SSWs in the study period. The two even larger errors were found after the 5 January 2004 SSW, when the forecast initialized on 17 January had a temperature error of -2.4σ 15–30 days into the forecast, and after the 31 December 2001 SSW, when the forecast initialized on 7 January 2002 was 2.3σ colder than the reanalysis for the same lead times.

In Fig. 7, several series of consecutive cluster-B forecasts were evident, and some of these could be linked to strong initial vortex states. In Fig. 9a, the NAO forecast error during the winter of 2015/16, during which one such sequence occurred, is shown. The bars representing $\bar{u}_\sigma$ show that except for a few days in February, the initial stratospheric vortex was strong throughout the winter.

Most of the forecasts up until the one initialized on 31 December 2015, belong in cluster B. The forecast NAO index was consistently too high for the cluster-B members. The average 2-m temperature anomaly for the cluster-B forecasts (including the one initialized on 4 February 2016) is shown in Fig. 9b. These forecasts were too warm in northern Europe, with average errors of more than 2 standard deviations in Finland.

4. Summary and discussion

In this paper, we have investigated the sources of subseasonal forecast errors over the North Atlantic sector. A cluster analysis identified cases of significant discrepancies between the predicted and observed sign of the NAO. While such errors might be expected to occur randomly in forecasts of a complex system with limited predictability, common dynamical structures were seen in both the troposphere and stratosphere in these cases, revealing systematic model biases in stratosphere–troposphere coupling in certain situations. It is worth noting that the results presented here were based on model data from only one leading prediction center (ECMWF). Further study is needed to determine if other subseasonal ensemble systems exhibit similar characteristics, in particular low-top models. Butler et al. (2020) identified considerable diversity in the bias of the NAM response among S2S prediction systems for the 2018 and 2019 SSW events, some of which can be traced back to the

FIG. 8. (a) The NAO index calculated from the forecasts (dashed line) and the reanalysis (solid line) are shown for each forecast initialization date during the 2018/19 winter, as indicated on the x axis. The cluster to which each forecast belongs is indicated with circles, using the same color scheme as in Fig. 7. The NAO index was averaged for lead times of 15–30 days. The unit is standard deviations (σ). The blue or red areas indicate that the forecast NAO index is lower or higher, respectively, than in the reanalysis. The gray bars show standardized $\bar{u}_\sigma$ values on each forecast initialization date, and the SSW central date on 2 Jan 2019 is indicated with a vertical dashed line. (b) The mean 2-m temperature anomalies 15–30 days into the cluster-A forecasts during the same winter. The stippling indicates values that are significantly different from zero at the 5% level, calculated using a bootstrapping approach that retained the autocorrelation of the sequence. The region used to calculate northern European temperature errors is outlined with dashed lines.

stratosphere. We emphasize that even though our results are statistically significant, there is still much diversity in the relationship between NAO errors and the initial state of the polar vortex (Fig. 5).

A cluster analysis enabled the cases exhibiting errors in the NAO forecast to be isolated. These cases exhibit a remarkable persistence of forecast errors, which are seen to emerge in forecast week two and persist over the

FIG. 9. As in Fig. 8, but during the 2015/16 winter and (b) shows forecasts for cluster B.

subsequent two weeks (see Fig. 3). Forecast errors can also persist for several consecutive forecasts, as highlighted by the long series of too-negative NAO forecasts during the winter of 2018/19 (see Fig. 8). This suggests that the emerging errors in forecasts from the preceding week or two may provide useful information on likely errors for today's forecasts.

The overestimation of the negative NAO response in the troposphere likely results from an overestimation of the tendency of weak vortex states to be followed by typical negative NAO phases. As shown for the example of the 2019 SSW, the negative NAO response was severely overestimated and resulted in much too cold temperature forecasts for northern Europe. Rao et al. (2019) found that the onset of the January 2019 SSW event was forecast in the ECMWF model 25 days before the event, which is notably longer than for most other SSW events (Domeisen et al. 2020c). Nevertheless, the downward impact was not correctly forecast by the model, highlighting that the question of forecasting SSWs and their downward coupling are separate problems. We mention here that a follow-up study on how the forecast errors were distributed vertically for specific events, including both SSWs and strong vortex events, would be worthwhile for enhancing our understanding of the mechanisms that give rise to the forecast errors near the surface.

In a recent paper related to the one presented here, Choi and Son (2019) used the ECMWF model to investigate composites for high and low SLP forecast skill north of 20°N. Their high skill composite, which corresponded to enhanced skill over parts of the North Atlantic and North Pacific, was associated with a 10-hPa geopotential height dipole anomaly pattern. As we did not fully account for zonally asymmetric flow, our study is not directly comparable, but it is worth noting that Choi and Son (2019) showed that different asymmetries with respect to the wave phasing in the stratospheric flow can have an impact on tropospheric prediction skill.

In another related study, Matsueda and Palmer (2018) investigated the flow-dependent predictability of the North Atlantic/European region in the THORPEX Interactive Grand Global Ensemble (TIGGE) and found a general overestimation of the frequency of the negative NAO regime. They related this to too frequent transitions from zonal to “wavier” regimes like the negative NAO, rather than an overestimation of its persistence. This is not necessarily at odds with the overly persistent negative NAO cases reported here. Their focus was the shorter 10–15 day time scale, on which stratospheric dynamics are likely to play a weaker role than in our forecasts. Also note that Matsueda and Palmer (2018) used a four regime system, and so some of the transitions between these four may occur

within one of the larger clusters used here. Thus, the overly frequent negative NAO transitions in the forecasts of Matsueda and Palmer (2018) likely manifest as enhanced persistence of the negative NAO cases in our analysis.

We emphasize that the origin of tropospheric NAO biases is not limited to the stratosphere. North Atlantic winter variability can also be influenced by the tropics (Scaife et al. 2017; Jiménez-Esteve and Domeisen 2018), by the Arctic (e.g., Kolstad and Årthun 2018; Blackport and Screen 2019), by the upstream circulation in the North Pacific (Drouard et al. 2013, 2015), and by the underlying ocean (Visbeck 2002).

5. Conclusions

Stratospheric variability is widely recognized as an important source of predictability on subseasonal time scales. The results presented here highlight that the diversity in the tropospheric response to stratospheric forcing is a particular challenge for forecast models.

Subseasonal prediction lies in an intermediate band of time scales between those dominated by initial condition information, as in weather forecasting, and those dominated by boundary condition information, as in near-term climate predictions. Weather forecasts have traditionally suffered from a lack of diversity or spread in model ensembles (e.g., Buizza 1997), whereas near-term climate predictions of the NAO suffer from the opposite problem of having too-large internal variability within ensembles (e.g., Eade et al. 2014). The forecast errors presented here could provide examples of the former of these issues on the stratospheric weather time scale, as the model appears to be constrained too strongly by the polar vortex state in the initial conditions.

Overcoming the challenge of allowing for diversity in dynamical stratosphere–troposphere coupling could significantly improve the skill of subseasonal predictions for events such as cold temperature extremes in northern Europe and storm track anomalies in central and southern Europe in winter.

Acknowledgments. Author Kolstad was supported by the Research Council of Norway through the Seasonal Forecasting Engine project (Grant 270733). Support to authors Domeisen and Wulff from the Swiss National Science Foundation through project PP00P2_170523 is gratefully acknowledged. Author Woollings acknowledges support from the Research Council of Norway, Grant 310391.

Data availability statement. This work is based on S2S data. S2S is a joint initiative of the World Weather Research Program (WWRP) and the World Climate Research Program (WCRP). The S2S database is hosted at ECMWF as

an extension of the TIGGE database and the S2S data are available online (<https://apps.ecmwf.int/datasets/data/s2s>).

REFERENCES

- Afargan-Gerstman, H., and D. I. V. Domeisen, 2020: Pacific modulation of the North Atlantic storm track response to sudden stratospheric warming events. *Geophys. Res. Lett.*, **47**, e2019GL085007, <https://doi.org/10.1029/2019GL085007>.
- Athanasiadis, P. J., and Coauthors, 2017: A multisystem view of wintertime NAO seasonal predictions. *J. Climate*, **30**, 1461–1475, <https://doi.org/10.1175/JCLI-D-16-0153.1>.
- Baker, L. H., L. C. Shaffrey, R. T. Sutton, A. Weisheimer, and A. A. Scaife, 2018: An intercomparison of skill and overconfidence/underconfidence of the wintertime North Atlantic Oscillation in multimodel seasonal forecasts. *Geophys. Res. Lett.*, **45**, 7808–7817, <https://doi.org/10.1029/2018GL078838>.
- Baldwin, M. P., and T. J. Dunkerton, 2001: Stratospheric harbingers of anomalous weather regimes. *Science*, **294**, 581–584, <https://doi.org/10.1126/science.1063315>.
- , and D. W. J. Thompson, 2009: A critical comparison of stratosphere–troposphere coupling indices. *Quart. J. Roy. Meteor. Soc.*, **135**, 1661–1672, <https://doi.org/10.1002/qj.479>.
- Blackport, R., and J. A. Screen, 2019: Influence of Arctic Sea ice loss in autumn compared to that in winter on the atmospheric circulation. *Geophys. Res. Lett.*, **46**, 2213–2221, <https://doi.org/10.1029/2018GL081469>.
- Buizza, R., 1997: Potential forecast skill of ensemble prediction and spread and skill distributions of the ECMWF ensemble prediction system. *Mon. Wea. Rev.*, **125**, 99–119, [https://doi.org/10.1175/1520-0493\(1997\)125<0099:PFSOEP>2.0.CO;2](https://doi.org/10.1175/1520-0493(1997)125<0099:PFSOEP>2.0.CO;2).
- Butler, A., and Coauthors, 2019: Sub-seasonal predictability and the stratosphere. *The Gap between Weather and Climate Forecasting*, A. W. Robertson and F. Vitart, Eds., Elsevier, 223–241, <https://doi.org/10.1016/B978-0-12-811714-9.00011-5>.
- , J. P. Sjöberg, D. J. Seidel, and K. H. Rosenlof, 2017: A sudden stratospheric warming compendium. *Earth Syst. Sci. Data*, **9**, 63–76, <https://doi.org/10.5194/essd-9-63-2017>.
- , Z. D. Lawrence, S. H. Lee, S. P. Lillo, and C. S. Long, 2020: Differences between the 2018 and 2019 stratospheric polar vortex split events. *Quart. J. Roy. Meteor. Soc.*, <https://doi.org/10.1002/qj.3858>, in press.
- Charlton, A. J., and L. M. Polvani, 2007: A new look at stratospheric sudden warmings. Part I: Climatology and modeling benchmarks. *J. Climate*, **20**, 449–469, <https://doi.org/10.1175/JCLI3996.1>.
- Charlton-Perez, A. J., L. Ferranti, and R. W. Lee, 2018: The influence of the stratospheric state on North Atlantic weather regimes. *Quart. J. Roy. Meteor. Soc.*, **144**, 1140–1151, <https://doi.org/10.1002/qj.3280>.
- Choi, J., and S.-W. Son, 2019: Stratospheric initial condition for skillful surface prediction in the ECMWF model. *Geophys. Res. Lett.*, **46**, 12 556–12 564, <https://doi.org/10.1029/2019GL083519>.
- Croci-Maspoli, M., C. Schierz, and H. C. Davies, 2007: Atmospheric blocking: Space-time links to the NAO and PNA. *Climate Dyn.*, **29**, 713–725, <https://doi.org/10.1007/s00382-007-0259-4>.
- Cui, X., Y. Gao, J. Sun, D. Guo, S. Li, and O. M. Johannessen, 2014: Role of natural external forcing factors in modulating the Indian summer monsoon rainfall, the winter North Atlantic Oscillation and their relationship on inter-decadal timescale. *Climate Dyn.*, **43**, 2283–2295, <https://doi.org/10.1007/S00382-014-2053-4>.
- DeWeaver, E., and S. Nigam, 2000: Zonal-eddy dynamics of the North Atlantic Oscillation. *J. Climate*, **13**, 3893–3914, [https://doi.org/10.1175/1520-0442\(2000\)013<3893:ZEDOTN>2.0.CO;2](https://doi.org/10.1175/1520-0442(2000)013<3893:ZEDOTN>2.0.CO;2).
- Dobrynin, M., and Coauthors, 2018: Improved teleconnection-based dynamical seasonal predictions of boreal winter. *Geophys. Res. Lett.*, **45**, 3605–3614, <https://doi.org/10.1002/2018GL077209>.
- Domeisen, D. I. V., 2019: Estimating the frequency of sudden stratospheric warming events from surface observations of the North Atlantic Oscillation. *J. Geophys. Res. Atmos.*, **124**, 3180–3194, <https://doi.org/10.1029/2018JD030077>.
- , G. Badin, and I. M. Koszalka, 2018: How predictable Are the Arctic and North Atlantic Oscillations? Exploring the variability and predictability of the Northern Hemisphere. *J. Climate*, **31**, 997–1014, <https://doi.org/10.1175/JCLI-D-17-0226.1>.
- , C. M. Grams, and L. Papritz, 2020a: The role of North Atlantic–European weather regimes in the surface impact of sudden stratospheric warming events. *Wea. Climate Dyn.*, **1**, 373–388, <https://doi.org/10.5194/wcd-1-373-2020>.
- , and Coauthors, 2020b: The role of the stratosphere in sub-seasonal to seasonal prediction: 1. Predictability of the stratosphere. *J. Geophys. Res. Atmos.*, **125**, e2019JD030920, <https://doi.org/10.1029/2019JD030920>.
- , and Coauthors, 2020c: The role of the stratosphere in sub-seasonal to seasonal prediction: 2. Predictability arising from stratosphere–troposphere coupling. *J. Geophys. Res. Atmos.*, **125**, e2019JD030923, <https://doi.org/10.1029/2019JD030923>.
- Drouard, M., G. Rivière, and P. Arbogast, 2013: The North Atlantic Oscillation response to large-scale Atmospheric anomalies in the northeastern Pacific. *J. Atmos. Sci.*, **70**, 2854–2874, <https://doi.org/10.1175/JAS-D-12-0351.1>.
- , —, and —, 2015: The link between the North Pacific climate variability and the North Atlantic Oscillation via downstream propagation of synoptic waves. *J. Climate*, **28**, 3957–3976, <https://doi.org/10.1175/JCLI-D-14-00552.1>.
- Dunstone, N., D. Smith, A. Scaife, L. Hermanson, R. Eade, N. Robinson, M. Andrews, and J. Knight, 2016: Skilful predictions of the winter North Atlantic Oscillation one year ahead. *Nat. Geosci.*, **9**, 809–814, <https://doi.org/10.1038/ngeo2824>.
- Eade, R., D. Smith, A. Scaife, E. Wallace, N. Dunstone, L. Hermanson, and N. Robinson, 2014: Do seasonal-to-decadal climate predictions underestimate the predictability of the real world? *Geophys. Res. Lett.*, **41**, 5620–5628, <https://doi.org/10.1002/2014GL061146>.
- Greatbatch, R. J., 2000: The North Atlantic Oscillation. *Stochastic Environ. Res. Risk Assess.*, **14**, 213–242, <https://doi.org/10.1007/s004770000047>.
- Hersbach, H., and Coauthors, 2020: The ERA5 global reanalysis. *Quart. J. Roy. Meteor. Soc.*, **146**, 1999–2049, <https://doi.org/10.1002/qj.3803>.
- Hurrell, J. W., 1995: Decadal trends in the North Atlantic Oscillation: Regional temperatures and precipitation. *Science*, **269**, 676–679, <https://doi.org/10.1126/science.269.5224.676>.
- , and C. Deser, 2010: North Atlantic climate variability: The role of the North Atlantic Oscillation. *J. Mar. Syst.*, **79**, 231–244, <https://doi.org/10.1016/j.jmarsys.2009.11.002>.
- , Y. Kushnir, and M. Visbeck, 2001: The North Atlantic Oscillation. *Science*, **291**, 603–605, <https://doi.org/10.1126/science.1058761>.
- Jiménez-Esteve, B., and D. I. V. Domeisen, 2018: The tropospheric pathway of the ENSO–North Atlantic teleconnection. *J. Climate*, **31**, 4563–4584, <https://doi.org/10.1175/JCLI-D-17-0716.1>.
- Kolstad, E. W., and M. Årthun, 2018: Seasonal prediction from Arctic sea surface temperatures: Opportunities and pitfalls. *J. Climate*, **31**, 8197–8210, <https://doi.org/10.1175/JCLI-D-18-0016.1>.
- , T. Breiteig, and A. A. Scaife, 2010: The association between stratospheric weak polar vortex events and cold air outbreaks

- in the Northern Hemisphere. *Quart. J. Roy. Meteor. Soc.*, **136**, 886–893, <https://doi.org/10.1002/qj.620>.
- MacQueen, J., 1967: Some methods for classification and analysis of multivariate observations. *Proc. Fifth Berkeley Symp. on Mathematical Statistics and Probability*, Vol. 1, Oakland, CA, University of California Press, 281–297.
- Marshall, J., and Coauthors, 2001: North Atlantic climate variability: Phenomena, impacts and mechanisms. *Int. J. Climatol.*, **21**, 1863–1898, <https://doi.org/10.1002/joc.693>.
- Matsueda, M., and T. N. Palmer, 2018: Estimates of flow-dependent predictability of wintertime Euro-Atlantic weather regimes in medium-range forecasts. *Quart. J. Roy. Meteor. Soc.*, **144**, 1012–1027, <https://doi.org/10.1002/qj.3265>.
- Nie, Y., A. A. Scaife, H.-L. Ren, R. E. Comer, M. B. Andrews, P. Davis, and N. Martin, 2019: Stratospheric initial conditions provide seasonal predictability of the North Atlantic and Arctic Oscillations. *Environ. Res. Lett.*, **14**, 034006, <https://doi.org/10.1088/1748-9326/AB0385>.
- Osborn, T. J., 2004: Simulating the winter North Atlantic Oscillation: The roles of internal variability and greenhouse gas forcing. *Climate Dyn.*, **22**, 605–623, <https://doi.org/10.1007/S00382-004-0405-1>.
- O'Reilly, C. H., A. Weisheimer, T. Woollings, L. J. Gray, and D. MacLeod, 2019: The importance of stratospheric initial conditions for winter North Atlantic Oscillation predictability and implications for the signal-to-noise paradox. *Quart. J. Roy. Meteor. Soc.*, **145**, 131–146, <https://doi.org/10.1002/qj.3413>.
- Pedregosa, F., and Coauthors, 2011: Scikit-learn: Machine learning in Python. *J. Mach. Learn. Res.*, **12**, 2825–2830.
- Rao, J., C. I. Garfinkel, H. Chen, and I. P. White, 2019: The 2019 new year stratospheric sudden warming and its real-time predictions in multiple S2S models. *J. Geophys. Res. Atmos.*, **124**, 111 55–111 74, <https://doi.org/10.1029/2019JD030826>.
- Scaife, A. A., and Coauthors, 2014: Skillful long-range prediction of European and North American winters. *Geophys. Res. Lett.*, **41**, 2514–2519, <https://doi.org/10.1002/2014GL059637>.
- , and Coauthors, 2016: Seasonal winter forecasts and the stratosphere. *Atmos. Sci. Lett.*, **17**, 51–56, <https://doi.org/10.1002/asl.598>.
- , and Coauthors, 2017: Tropical rainfall, Rossby waves and regional winter climate predictions. *Quart. J. Roy. Meteor. Soc.*, **143**, 1–11, <https://doi.org/10.1002/qj.2910>.
- Sigmond, M., J. F. Scinocca, V. V. Kharin, and T. G. Shepherd, 2013: Enhanced seasonal forecast skill following stratospheric sudden warmings. *Nat. Geosci.*, **6**, 98–102, <https://doi.org/10.1038/ngeo1698>.
- Smith, D., and Coauthors, 2019: Robust skill of decadal climate predictions. *npj Climate Atmos. Sci.*, **2**, 1–10, <https://doi.org/10.1038/s41612-019-0071-y>.
- Thompson, D., and J. M. Wallace, 1998: The Arctic oscillation signature in the wintertime geopotential height and temperature fields. *Geophys. Res. Lett.*, **25**, 1297–1300, <https://doi.org/10.1029/98GL00950>.
- , and —, 2000: Annular modes in the extratropical circulation. Part I: Month-to-month variability. *J. Climate*, **13**, 1000–1016, [https://doi.org/10.1175/1520-0442\(2000\)013<1000:AMITEC>2.0.CO;2](https://doi.org/10.1175/1520-0442(2000)013<1000:AMITEC>2.0.CO;2).
- Tripathi, O. P., and Coauthors, 2015: The predictability of the extratropical stratosphere on monthly time-scales and its impact on the skill of tropospheric forecasts. *Quart. J. Roy. Meteor. Soc.*, **141**, 987–1003, <https://doi.org/10.1002/qj.2432>.
- van Loon, H., and J. C. Rogers, 1978: The seesaw in winter temperatures between Greenland and northern Europe. Part I: General description. *Mon. Wea. Rev.*, **106**, 296–310, [https://doi.org/10.1175/1520-0493\(1978\)106<0296:TSIWTB>2.0.CO;2](https://doi.org/10.1175/1520-0493(1978)106<0296:TSIWTB>2.0.CO;2).
- Visbeck, M., 2002: The ocean's role in Atlantic climate variability. *Science*, **297**, 2223–2224, <https://doi.org/10.1126/science.1074029>.
- , J. W. Hurrell, L. Polvani, and H. M. Cullen, 2001: The North Atlantic Oscillation: Past, present, and future. *Proc. Natl. Acad. Sci. USA*, **98**, 12 876–12 877, <https://doi.org/10.1073/pnas.231391598>.
- Vitart, F., 2014: Evolution of ECMWF sub-seasonal forecast skill scores. *Quart. J. Roy. Meteor. Soc.*, **140**, 1889–1899, <https://doi.org/10.1002/qj.2256>.
- , and Coauthors, 2017: The subseasonal to seasonal (S2S) prediction project database. *Bull. Amer. Meteor. Soc.*, **98**, 163–173, <https://doi.org/10.1175/BAMS-D-16-0017.1>.
- Wanner, H., S. Brönnimann, C. Casty, D. Gyalistras, J. Luterbacher, C. Schmutz, D. B. Stephenson, and E. Xoplaki, 2001: North Atlantic Oscillation—Concepts and studies. *Surv. Geophys.*, **22**, 321–381, <https://doi.org/10.1023/A:1014217317898>.
- Weisheimer, A., D. J. Belfort, D. MacLeod, T. Palmer, C. O'Reilly, and K. Strømme, 2020: Seasonal forecasts of the 20th century. *Bull. Amer. Meteor. Soc.*, <https://doi.org/10.1175/BAMS-D-19-0019.1>, in press.
- Woollings, T., C. Franzke, D. L. R. Hodson, B. Dong, E. A. Barnes, C. C. Raible, and J. G. Pinto, 2014: Contrasting interannual and multidecadal NAO variability. *Climate Dyn.*, **45**, 539–556, <https://doi.org/10.1007/s00382-014-2237-y>.
- Zhang, F., Y. Q. Sun, L. Magnusson, R. Buizza, S.-J. Lin, J.-H. Chen, and K. Emanuel, 2019: What is the predictability limit of midlatitude weather? *J. Atmos. Sci.*, **76**, 1077–1091, <https://doi.org/10.1175/JAS-D-18-0269.1>.