Geophysical Research Letters

RESEARCH LETTER
10.1029/2020GL090451

Attribution of NAO Predictive Skill Beyond 2 Weeks in Boreal Winter

Lantao Sun[1], Judith Perlwitz[2], Jadwiga H. Richter[3], Martin P. Hoerling[2], and James W. Hurrell[1]

[1]Department of Atmospheric Science, Colorado State University, Fort Collins, CO, USA, [2]NOAA Physical Sciences Laboratory, Boulder, CO, USA, [3]National Center for Atmospheric Research, Boulder, CO, USA

Key Points:
- Weeks 3–6 NAO predictive skill is attributed to stratospheric polar vortex conditions and ocean lower boundary forcing
- Enhanced NAO skill following weak vortex events results primarily from stratospheric coupling to the troposphere
- Enhanced NAO skill following strong vortex events can be partly attributed to ENSO

Supporting Information:
- Supporting Information S1

Correspondence to:
L. Sun,
lantao.sun@colostate.edu

Citation:
Sun, L., Perlwitz, J., Richter, J. H., Hoerling, M. P., & Hurrell, J. W. (2020). Attribution of NAO predictive skill beyond 2 weeks in boreal winter. *Geophysical Research Letters*, *47*, e2020GL090451. https://doi.org/10.1029/2020GL090451

Received 19 AUG 2020
Accepted 3 NOV 2020
Accepted article online 9 NOV 2020

Abstract Weeks 3–6 averaged winter North Atlantic Oscillation (NAO) predictive skill in a state-of-the-art coupled climate prediction system is attributed to two principle sources: upper and lower boundary conditions linked to the stratosphere and El Niño-Southern Oscillation (ENSO), respectively. A 20-member ensemble of 45-day reforecasts over 1999–2015 is utilized, together with uninitialized simulations with the atmospheric component of the prediction system forced with observed radiative forcing and lower boundary conditions. NAO forecast skill for lead times out to 6 weeks is higher following extreme stratospheric polar vortex conditions (weak and strong vortex events) compared to neutral states. Enhanced weeks 3–6 NAO predictive skill for weak vortex events results primarily from stratospheric downward coupling to the troposphere, while enhanced skill for strong vortex events can be partly attributed to lower boundary forcing related to the ENSO phenomenon. Implications for forecast system development and improvement are discussed.

Plain Language Summary Winter climate over Europe and eastern North America is significantly affected by variability of the North Atlantic Oscillation (NAO). In this study, we quantify the NAO predictive skill for lead times of 3–6 weeks and attribute it to two main sources: the stratosphere and the El Niño-Southern Oscillation (ENSO) phenomenon. This is done by contrasting ensembles of 45-day reforecasts over 1999–2015 with the corresponding uninitialized atmosphere model simulations forced with observed radiative forcing, sea-surface temperature, and sea ice conditions. We find that the model is able to better predict the NAO up to 3 weeks following extreme weak or strong states of the stratospheric polar vortex compared to stratospheric neutral vortex states. Enhanced weeks 3–6 NAO predictive skill for weak vortex events results primarily from stratospheric coupling to the troposphere, while enhanced skill for strong vortex events can be attributed in part to lower boundary forcing related to ENSO. These results have implications for forecast model development and improvement.

1. Introduction

Forecasting of North Atlantic Oscillation (NAO) variability on subseasonal-to-seasonal (S2S) and seasonal-to-decadal (S2D) timescales has recently received much attention due to its potential to provide enormous social-economic benefits (Mariotti et al., 2020; Merryfield et al., 2020; Smith et al., 2019; White et al., 2017). Winter climate over Europe and eastern North America is significantly affected by the variability of the NAO (Hurrell, 1995, 1996; Hurrell & Deser, 2009). The temporal evolution of the NAO has been suggested to be primarily a stochastic process with a fundamental timescale of ~10 days (Feldstein, 2000) implying limited predictability beyond weather time scales (e.g., Johansson, 2007). Yet, at longer timescales, a fraction of NAO variability has been shown to be forced by changes in sea-surface temperature (SST) and sea ice (e.g., Hoerling et al., 2004; Hurrell et al., 2004; Screen, 2017). These low frequency NAO variations, therefore, are likely more than just statistical remnants of energetic high-frequency atmospheric fluctuations and could be predictable if they are driven by predictable changes in boundary forcing.

Recent forecast system assessments have suggested that there may be enhanced predictability of the NAO on S2S timescales (Riddle et al., 2013; Scaife et al., 2014). Efforts have been made to understand the source of this enhanced predictability (e.g., Brunet et al., 2015; Cassou, 2008; Newman & Sardeshmukh, 2008; Scaife et al., 2014). Evidence suggests that stratospheric processes and associated variability are important (Jia et al., 2017; Nie et al., 2019; O'Reilly et al., 2019; Scaife et al., 2016; Tripathi, Baldwin, et al., 2015; Wang et al., 2017; Xiang et al., 2019), including the contribution of the stratospheric pathway associated

with the El Niño-Southern Oscillation (ENSO) phenomenon (Butler et al., 2016; Domeisen et al., 2015, 2018).

Extreme stratospheric polar vortex states (i.e., weak and strong vortex events) are followed by anomalous near-surface NAO conditions (Baldwin & Dunkerton, 2001) implying enhanced NAO prediction skill could occur as forecasts-of-opportunity contingent on initial stratospheric states (Albers & Newman, 2019). For instance, Sigmond et al. (2013) demonstrated that the skill of NAO forecasts averaged over 15- to 60-day periods is substantially enhanced when forecasts are initialized at the onset time of weak stratospheric polar vortex events. Tripathi, Charlton-Perez, et al. (2015) found that initialization based on anomalous vortex events improves the skill of NAO predictions up to week 4, which has also been supported by multiple S2S model assessments (Domeisen et al., 2020b; WWRP/WCRP, 2018). Kolstad et al. (2020) revealed that errors in subseasonal NAO forecasts can be traced to the state of the stratospheric polar vortex when the forecasts were initialized.

In these aforementioned studies, the role of the stratosphere is typically assessed by categorizing the individual forecasts based on initial stratospheric polar vortex states. However, this method cannot rule out the possibility that enhanced NAO predictive skill may also partly come from other sources, for instance, lower boundary forcing. Separating different sources of predictability is difficult, especially when the sample size is small. By comparing the skill in a subseasonal prediction system with its corresponding uninitialized AMIP simulations forced with observed SST and sea ice conditions, the relative importance of atmospheric (e.g., extreme stratospheric vortex states) and oceanic boundary sources of NAO predictability can be identified.

In this study, we attribute boreal winter NAO predictive skill for lead times of 3–6 weeks to distinct physical sources. Reforecasts using a coupled model are analyzed to quantify skill dependency on the initial stratospheric conditions. Parallel uninitialized simulations employing the same modeling system are analyzed to identify potential predictors in lower boundary forcing.

2. Model Reforecasts and AMIP Simulations

2.1. Model Reforecast

The reforecasts are conducted with the Community Earth System Model Version 1 (CESM; Hurrell et al., 2013) for 1999–2015. The daily output follows the protocol of the Subseasonal Experiment project (SubX; Pegion et al., 2019). The forecasts start every Wednesday and run for 45 days. The atmospheric initial conditions are based on ERA-Interim reanalysis (Dee et al., 2011). The ocean and sea ice initial conditions are utilized from a forced ocean-sea-ice simulation that employs adjusted Japanese 55-year reanalysis atmospheric state fields and fluxes (Tsujino et al., 2018) as surface boundary conditions. Details can be found in the Supporting Information Text S1 and Richter et al. (2020).

The CESM reforecasts are composited of two 10-member ensembles with the same initial conditions but 30 (default) and 46 vertical levels in its atmospheric component—Community Atmosphere Model version 5 (CAM5; Neale et al., 2012; Richter et al., 2015). Richter et al. (2020) compared these two CESM reforecasts and found that improved representation of the stratosphere can improve the predictive skill of the stratosphere but does not materially change tropospheric NAO predictive skill. Therefore, we combine the two 10-member ensembles to yield a 20-member ensemble for detailed analysis. Our conclusions hold when we analyze the 10-member ensembles separately.

In addition, hindcasts with the National Centers for Environmental Prediction (NCEP) coupled forecast system model version 2 (CFS; Saha et al., 2014) and European Centre for Medium-Range Weather Forecasts (ECMWF; Vitart, 2014) are compared to those using CESM. This includes a four-member CFS reforecast ensemble initialized every day for 1999–2010 and a 11-member ECMWF reforecast ensemble initialized twice per week for 1996–2015, both from the international S2S database (Vitart et al., 2012, 2017). Recognizing that the models have somewhat different hindcast periods, all the findings presented here have been recalculated using their common period of 1999–2010.

2.2. NAO Skill Evaluation

The predictive skill is evaluated by the anomaly correlation coefficient (ACC) that has been commonly used in S2S forecasts (e.g., Tripathi, Charlton-Perez, et al., 2015). Namely,

$$ACC = \frac{\sum_{t=1}^{T}(X_{mod} - C_{mod})(X_{obs} - C_{obs})}{\sqrt{\sum_{t=1}^{T}(X_{mod} - C_{mod})^2 \cdot \sum_{t=1}^{T}(X_{obs} - C_{mod})^2}},$$

where X_{mod} and X_{obs} represent the weekly or monthly values for model ensemble-mean and reanalysis (ERA-Interim; Dee et al., 2011), respectively. C_{mod} is the lead-time dependent model climatology and has the same period of time as the reanalysis climatology C_{obs}. T is the number of events over which the ACC is calculated.

In this study, we diagnose the NAO forecast skill of sea-level pressure (SLP) based on initialization during November–March for lead times of 3–6 weeks (days 15–42 average). Discrete weekly skill for lead times of 1–6 weeks is also diagnosed. Similar to other studies (e.g., Butler et al., 2016; Johansson, 2007), we conduct Empirical Orthogonal Function (EOF) analysis for ERA-Interim monthly (November–March) SLP anomalies over the Atlantic sector (20°N-80°N; 90°W-40°E) and treat the leading EOF regression as the NAO pattern for both observations and models. The NAO index is then calculated by projecting the daily SLP anomaly from model and reanalysis data onto the leading EOF pattern and using it for ACC diagnostics. To quantify the confidence interval in the predictive skill of the NAO, we apply a bootstrapping method by resampling 1,000 times with replacement and obtaining the 5% and 95% of the ACC significance levels (Mudelsee, 2010).

Initial state persistence as a simple baseline source for NAO skill is evaluated from the correlation of the NAO anomaly relative to its week 0 value (day −3 to day 3 average), and the resulting monthly persistence skill is compared to that from the initialized dynamical models. Also, to isolate the role of surface boundary forcing, we utilize 50-member atmosphere model simulations with CAM5 forced by observed greenhouse gases, SST, and sea ice conditions (Hurrell et al., 2008). The ensemble-mean anomaly for these uninitialized AMIP simulations is obtained by removing the daily long-term climatology and then calculating the ACC against observations. Again, the analysis is of monthly mean variability, and the simulation skill is compared to weeks 3–6 skill from initialized forecasts and from simple persistence.

To explore the role of the stratosphere in NAO predictability, we subsample the winter forecast into three categories based on the initial stratospheric zonal-mean zonal winds at 10 hPa and 60°N. Similar to Tripathi, Charlton-Perez, et al. (2015), we first calculate the probability density function (PDF) of the 1999–2015 10-hPa 60°N November–March zonal-mean zonal wind distribution. The weak (strong) stratospheric polar vortex events are defined for forecasts whose initial zonal wind at 10-hPa 60°N is below 15% of the PDF (3.55 m/s) or above 80% of the PDF (40.79 m/s) resulting in 60 weak and 62 strong vortex events. Figure S1 shows the composite of standardized polar-cap geopotential height and SLP anomalies for weak and strong vortex events, indicating that the observed downward coupling of extreme polar vortex events and their surface NAO features are both well captured in CESM. Neutral vortex events are defined for forecasts when the 10-hPa 60°N zonal winds (in reanalysis) are between 15% and 80% of the PDF for lead times of 1–5 days. Compared to the common definition based on initial winds, our method further excludes extreme polar vortex events occurring in the first 5 days that the majority of the model ensemble members are able to predict (Domeisen et al., 2020a; Richter et al., 2020). The predictive skill of persistence and uninitialized AMIP simulations can be evaluated based on the same stratospheric events as in the CESM reforecasts.

3. Results

3.1. Winter NAO Predictive Skill

The Northern Hemisphere (NH) monthly SLP hindcast skill from weeks 3 to 6 in boreal winter (November–March) for CESM (1999–2015), CFS (1999–2010), and ECMWF (1996–2015) models is presented in Figure 1a. The common feature among the three reforecasts is overall greater skill in lower latitudes than higher latitudes, with a secondary skill maximum centered over Greenland especially in CESM and ECMWF forecasts. When averaged over NH extratropics (northward of 23.5°N), ECMWF has the highest skill (0.36), followed by CESM (0.33) and CFS (0.26). In the Atlantic basin, high skill regions over Greenland and near the Azores project onto the NAO centers of action, suggesting that this leading mode of climate variability has particularly enhanced forecast skill (Johansson, 2007). CESM and

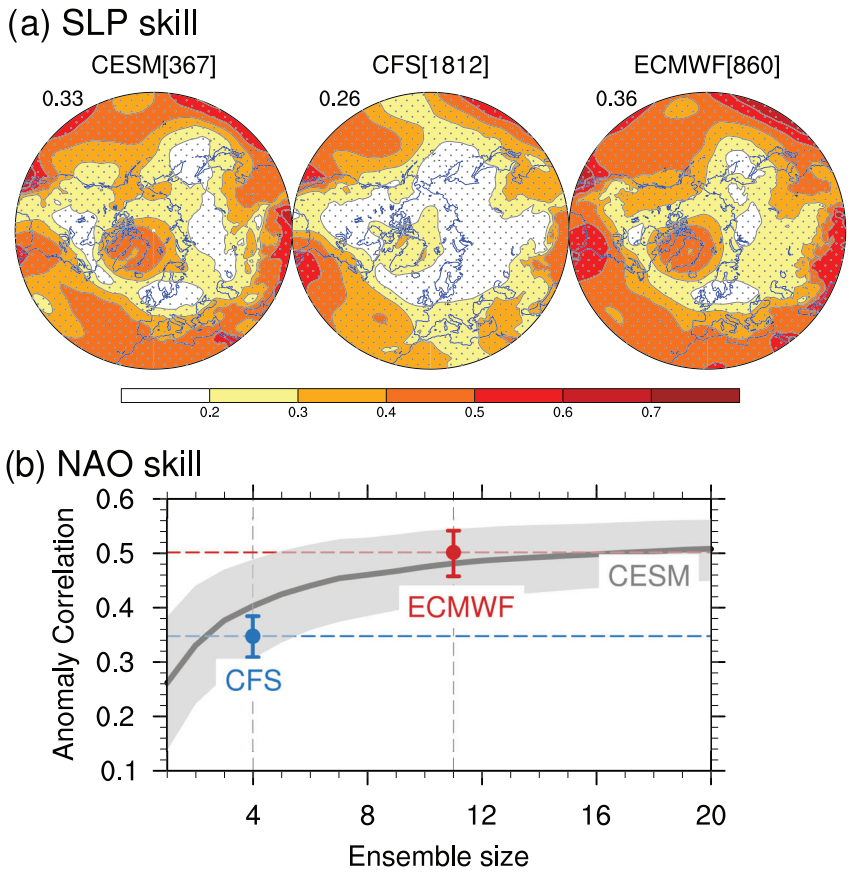

Figure 1. (a) Sea-level pressure weeks 3–6 predictive skill evaluated by anomaly correlation coefficient (ACC) during November–March for CESM, CFS, and ECMWF reforecasts. Values on the left corner denote the skill averaged over Northern hemisphere extratropics. Numbers in the brackets of each model show the sample size. Stippling indicates where values are different from zero at the 95% significance level based on bootstrapping. (b) Weeks 3–6 North Atlantic Oscillation (NAO) predictive skill as a function of ensemble size. Shading, blue and red error bars indicate the 5% and 95% statistical significance levels by conducting bootstrapping for CESM, CFS and ECMWF reforecasts, respectively.

ECMWF monthly SLP skill implies greater NAO predictability than indicated by CFS, a contrast among these systems that is evident already by week 2 and persists until week 6 (see Figure S2). We also assess the model performance using the spread-error ratio (Weisheimer et al., 2011). We find a much better spread-to-error ratio in CESM and ECMWF compared to CFS (Figure S3).

Some of the model differences in skill are a function of ensemble size rather than a fundamental model bias that might plague a particular prediction system, and it has been shown that skill declines with diminishing model ensemble size (e.g., Athanasiadis et al., 2020; Butler et al., 2016; Kumar & Hoerling, 1995; Smith et al., 2020). To explore this further, we randomly select N ensembles (N from 1 to 20) from the CESM reforecast and then conduct 1,000-times bootstrapping to calculate the 5% and 95% statistical significance level of the NAO skill. Figure 1b shows weeks 3–6 monthly NAO skill in CESM as a function of ensemble size. In agreement with those earlier studies, the NAO skill approximately doubles (0.27 vs. 0.51) when the ensemble size increases from 1 to 20, but it is close to saturation at 15 ensemble members (e.g., there is a less than 0.01 increase from 15 to 20 members).

Thus, when accounting for the differences in ensemble size of the systems diagnosed herein, the lower NAO skill in the smaller CFS ensemble is reconcilable with sampling alone; its skill overlaps with that of CESM when the latter's ensemble is subsampled to match the CFS population. That said, the lower skill with the CFS model might also be related to the not well-represented connection between tropical Pacific SST and NAO, poorly predicted stratospheric polar vortex variability and too weak stratosphere-troposphere

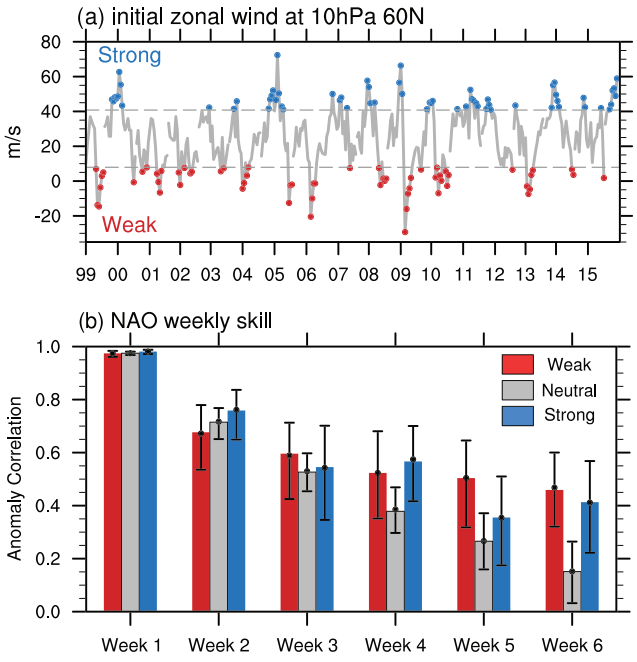

Figure 2. (a) Observed November–March zonal-mean zonal wind at 10 hPa and 60°N and the weak (red), strong (blue) polar vortex CESM reforecast cases. (b) Weekly NAO skill for the weak (red), strong (blue) stratospheric polar vortex events and neutral polar vortex conditions (gray). Error bars indicate the 5% and 95% statistical significance levels.

coupling (Miller & Wang, 2019). Evidence to support this argument for degraded skill is presented in the next section where we demonstrate an appreciable sensitivity of skill to initial stratospheric conditions overall.

As a prelude to a more detailed analysis of the attributable causes (sources) for weeks 3–6 NAO prediction skill, the initialized prediction skill of CESM is compared to that resulting from simple persistence and that arising from lower boundary forcing alone with no recourse to the initial weather state. Summarized in Figure S4, the monthly NAO skill in CESM hindcasts (~0.5) is found to be considerably larger than in both persistence forecasts (~0.3) and the simulation skill arising solely from effects of prescribed surface lower boundary forcing (~0.05). It is interesting to note from their NH distributions of SLP skill (Figure S4, left) that the AMIP simulations have more than double the correlation skill of persistence averaged over the NH overall. But persistence skill is quite high over the two NAO centers of variability in the Atlantic basin, while AMIP skill is virtually absent over the NAO's northern node. This overall estimate of skill sources, while informative, obscures a considerable conditionality of the skill by each source.

3.2. Sensitivity to the Stratospheric Initial States

To explore the role of initial stratospheric states on NAO prediction skill, we subsample the CESM winter reforecasts into weak, neutral, and strong polar vortex events. Figure 2a shows the monthly time series of 10-hPa zonal winds averaged along 60°N, the extreme values of which are used to identify strong (blue dots) and weak (red dots) vortex events during 1999–2015. The weekly NAO skill using this stratification is displayed in Figure 2b. In agreement with earlier studies, the NAO skill is generally higher in weeks 3–6 when initialized from extreme stratospheric polar vortex conditions compared to neutral states. Whereas NAO skill initialized from neutral polar vortex states declines roughly exponentially, a comparatively sustained high skill occurs when initialized from extreme polar vortex state with correlations above 0.5 through week 6 for weak vortex initial conditions in particular. By week 6, this skill improvement relative to neutral states (0.15) is appreciable and statistically significant, indicating that NAO skill is substantially enhanced when initialized during weak vortex events, a result based on CESM that affirms prior findings based on other forecast systems (Domeisen et al., 2020b; Sigmond et al., 2013; Tripathi, Charlton-Perez, et al., 2015). The NAO skill for strong polar vortex events is also elevated compared to neutral vortex states especially after week 4. However, the skill is not as high as for weak vortex events in weeks 5 and 6. This is likely due to a difference between CESM and observed NAO life cycles during strong vortex events. In weeks 5–6 following strong vortex events, the canonical positive NAO pattern is evident in CESM but not in observations (see Figure S1b). We speculate that the lack of positive NAO skill in weeks 5–6 for these strong vortex cases might simply be due to sampling rather than due to a fundamentally shorter time scale of stratosphere-troposphere interactions during strong versus weak vortex environments (c.f., Figure 2 of Reichler et al., 2012).

The CESM SLP and NAO skill averaged for forecast lead times of 3–6 weeks under weak, neutral, and strong polar vortex events is presented in Figure 3 and further elucidates the sensitivity to the stratospheric initial conditions. The higher NAO skill for weak vortex cases originates from a superior forecast performance over the northern center of NAO variability. There the monthly SLP correlation skill is near 0.7 for weak vortex conditions compared to slightly below 0.5 for the strong vortex cases. Interestingly, the NH average SLP skill is higher for the strong vortex environments (0.41 vs. 0.34) and tends to also be higher in most locations, though with the clear regional exception being over the far North Atlantic.

The NAO skill for persistence forecasts and AMIP simulations provides further insight on the attributable causes for NAO prediction skill beyond 2 weeks (Figure 3b). The persistence skill for weak vortex events is significantly lower than for strong vortex events, which is possibly related to the initial NAO state during the onset of extreme polar vortex events (Domeisen et al., 2020). There is also statistically significant asymmetry in AMIP simulation skill when binned according to reforecast polar vortex intensity. It is important to

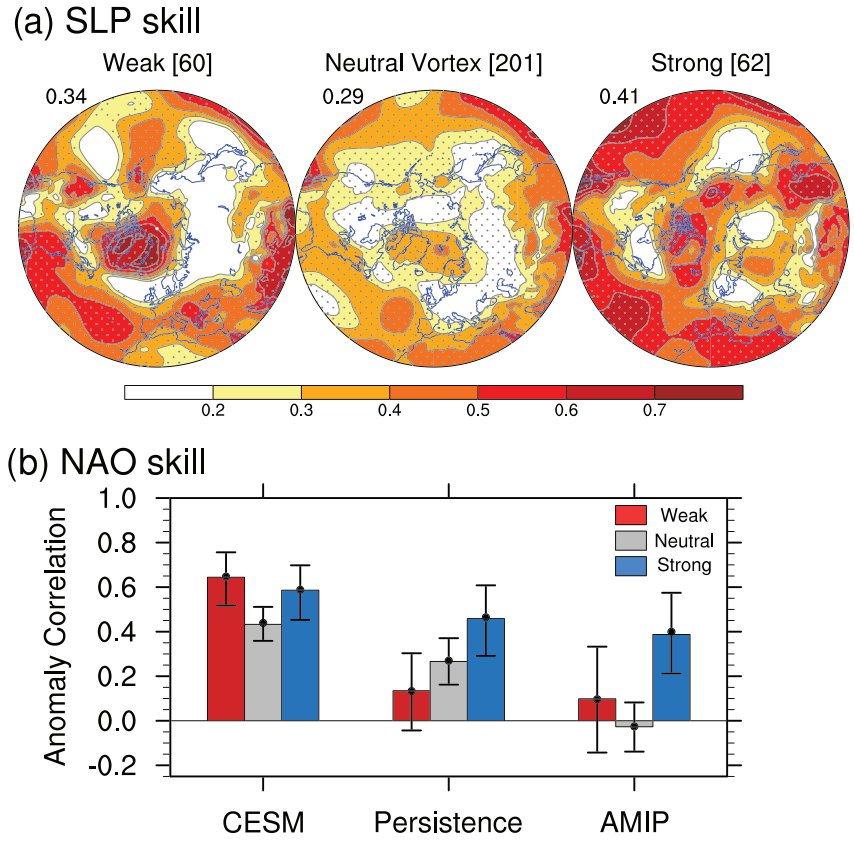

Figure 3. (a) As in Figure 1a, but for the CESM weak, neutral, and strong vortex events. (b) As in (a), but for the NAO skill in CESM, persistence and uninitialized AMIP simulations. Error bars indicate the 5% and 95% statistical significance levels.

note that the AMIP runs are constrained only by variability in surface boundary conditions and have no explicit synchronicity with the actual temporal variability of the polar vortex. Thus, the results of Figure 3a strongly suggest that boundary forcings themselves force the NAO variability for the strong events.

Considering these two factors in aggregate reveals an absence of skill during weak vortex states in both persistence and AMIP and that the considerable monthly skill of CESM predictions is mostly due to the importance of the initial stratospheric state. By contrast, both persistence and lower boundary forcing contribute to monthly NAO skill when the initial polar vortex is strong.

An explanation for the role of ocean forcing is provided by an analysis that stratifies the occurrences of polar vortex events according to ENSO phase. The Niño3.4 index, based on Optimum Interpolation (OI) v2 SST analyses (Reynolds et al., 2007), is shown in Figure 4a in relation to weak and strong polar vortex events. The relationship is not linear and is complicated. For instance, many strong vortex events appear to happen in either El Niño or La Niña years (e.g., 1999/2000, 2007/2008, 2009/2010, 2010/2011, & 2015/2016), though such an association is less evident for weak and neutral events. These relationships are summarized via histograms of Niño3.4 SST states stratified according to polar vortex intensity (Figure 4b). The spread of the distribution for strong vortex states is wide and highly non-Gaussian (blue curve), especially compared to that for weak polar vortex states (red curve). It suggests that the NAO predictive skill following strong vortex events is influenced by ENSO, agreeing with the recent findings of Xiang et al. (2019) who identified ENSO as the most predictable subseasonal mode and found that its influence in the North Atlantic resembles the NAO pattern. However, further research is needed to determine why many strong vortex events coincide with both El Niño and La Niña years.

The lower panel of Figure 4 provides a spatial view of the monthly SLP skill patterns related to effects of initial stratospheric conditions and ENSO boundary conditions. Here, we subsample the winter cases into

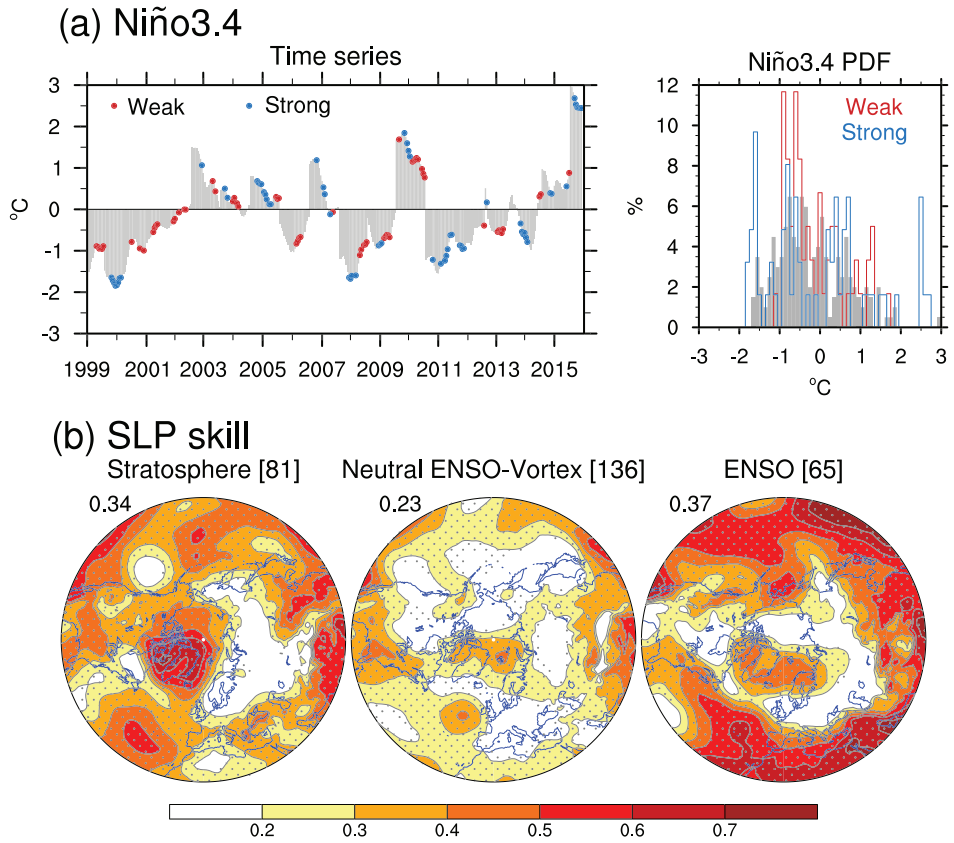

Figure 4. (a) Niño 3.4 time series and the weak (red dots) and strong (blue dots) polar vortex events and the probability density function of Niño 3.4 index for the weak (red), neutral (gray shading), and strong (blue) polar vortex events. In (a), the daily Niño 3.4 index is smoothed by 31 days running mean. (b) As in Figure 3a, but for initial extreme stratospheric polar vortex events (bottom left; combining weak and strong events), neutral stratosphere-ENSO events (bottom middle), and ENSO events.

extreme stratospheric polar vortex conditions (both weak and strong), ENSO cases (both El Niño and La Niña, defined by Niño 3.4 index above/below +1/−1, respectively), and neutral stratosphere-neutral ENSO cases (Figure 4b). Note that the ENSO cases have been excluded from the composites of extreme stratospheric vortex states, and extreme polar vortex events have been excluded from the ENSO cases. Immediately apparent is that monthly SLP skill for neutral stratosphere and neutral ENSO cases is low overall (Figure 4b middle panel) and has little amplitude in the NAO regional centers of variability. By contrast, SLP skill is comparatively high for extreme states of ENSO (right panel) and for extreme states of the polar vortex (left panel). While the initial atmospheric state is an especially effective skill source over the northern node of the NAO, it is also clear that ENSO likewise projects onto centers of skill that align with the NAO dipole. Moreover, when the AMIP simulation skill is subsampled for the same extreme ENSO states, the resulting skill map (Figure S5) closely resembles that of Figure 4b, confirming an ENSO-attributable skill source for the NAO in initialized CESM1 (Figure S6). Overall, our conclusions regarding forecast performance hold when analyses are repeated using a mean square error skill score (see Supporting Information Text S2 and Figure S7).

4. Summary and Discussion

In this study, we explore sources of subseasonal predictive skill of the boreal winter NAO. By utilizing ensembles of 45-day reforecasts during 1999–2015 with a state-of-the-art prediction system, in combination with uninitialized AMIP simulations, we attribute NAO predictive skill for lead times of 3–6 weeks. We find two principle sources of skill: upper and lower boundary constraints linked to the stratosphere and ENSO, respectively.

Specifically, we find that the winter NAO skill for forecast lead times out to 6 weeks is higher following extreme stratospheric polar vortex conditions (weak and strong vortex events) compared to neutral states. We further show that the enhanced NAO predictive skill for weak vortex events results primarily from stratospheric downward coupling, while the enhanced skill for strong vortex events can be attributed, in part, to lower boundary forcing associated with ENSO.

We also find that NAO prediction skill is appreciably reduced if the forecast ensemble size is too small. While a similar result has been determined for seasonal and decadal prediction systems, it may become more important for the subseasonal forecast due to compromise between ensemble size and initialization frequency. Our finding suggests that given fixed computational resources, the ensemble size should be no less than 10; otherwise, NAO predictive skill will be substantially lower than achievable levels.

Our findings come with a caveat that CESM and most other similar subseasonal forecast systems (e.g., S2S Prediction Project and SubX) cover only the most recent 20–30 years. Since NAO predictability has been found to vary at decadal and multidecadal timescale (Weisheimer et al., 2018), the degree to which our results are specific to the period analyzed is an open question.

Our findings also have important implications for forecast system development and improvement. Specifically, our results indicate that NAO boreal winter forecast skill depends on not only the representation of stratospheric processes in the forecast model, but also on the ENSO evolution during the model testing period. Thus, for NAO predictive skill as benchmark for model improvement, the analysis of subseasonal reforecasts in combination with AMIP simulations will provide insight as to whether NAO predictive skill can be improved by focusing on better representation of stratospheric processes and related variability in a specific prediction system or through better simulation of ENSO, if not both.

Data Availability Statement

The CESM reforecast data are available to the public via Columbia University's International Research Institute for Climate and Society (IRI) library at http://iridl.ldeo.columbia.edu/SOURCES/.Models/.SubX/.CESM/. The CAM5 simulations are available at PSL's Facility for Weather and Climate Assessments (FACTS) website at https://www.esrl.noaa.gov/psl/repository/facts/. The CFS and ECMWF data are available at ECMWF's S2S portal https://apps.ecmwf.int/datasets/data/s2s/levtype=sfc/origin=ecmf/type=cf/. OI SST v2 Niño3.4 index is downloaded from the KNMI climate explorer at https://climexp.knmi.nl/getindices.cgi?WMO=NCEPData/nino34_daily&STATION=NINO3.4&TYPE=i&id=someone@somewhere&NPERYEAR=366.

Acknowledgments

This work has been conducted as part of the NOAA's Modeling, Analysis, Predictions and Projections (MAPP) program S2S prediction Task Force and supported by NOAA's Climate Program Office's MAPP program. We also acknowledge support from the National Center for Atmospheric Research, which is a major facility sponsored by the National Science Foundation under Cooperative Agreement No. 1852977. Portions of this study were supported by the Regional and Global Model Analysis (RGMA) component of the Earth and Environmental System Modeling Program of the U.S. Department of Energy's Office of Biological & Environmental Research (BER) via National Science Foundation IA 1844590. L.S. is supported by Jim Hurrell's presidential chair startup funding at CSU. We thank Yan Wang at NOAA's Physical Sciences Laboratory for her help with downloading CFS and ECMWF hindcast data and Zac Lawrence (NOAA/PSL), Daniela Domeisen (ETH Zurich), and one anonymous reviewer for helpful comments on an earlier version of this paper.

References

Albers, J. R., & Newman, M. (2019). A priori identification of skillful extratropical subseasonal forecasts. *Geophysical Research Letters*, *46*, 12,527–12,536. https://doi.org/10.1029/2019GL085270

Athanasiadis, P. J., Yeager, S., Kwon, Y., Bellucci, A., Smith, D. W., & Tibaldi, S. (2020). Decadal predictability of North Atlantic blocking and the NAO. *npj Climate and Atmospheric Science*, *3*(1), 20. https://doi.org/10.1038/s41612-020-0120-6

Baldwin, M. P., & Dunkerton, T. J. (2001). Stratospheric harbingers of anomalous weather regimes. *Science*, *294*(5542), 581–584. https://doi.org/10.1126/science.1063315

Brunet, G., Jones, S., & Ruti, P. M. (Eds.). (2015). *Seamless prediction of the Earth system: From minutes to months (WMO-1156)* (pp. 1–483). Retrieved from https://library.wmo.int/pmb_ged/wmo_1156_en.pdf

Butler, A. H., Arribas, A., Athanassiadou, M., Baehr, J., Calvo, N., Charlton-Perez, A., et al. (2016). The Climate-system Historical Forecast Project: Do stratosphere-resolving models make better seasonal climate predictions in boreal winter? *Quarterly Journal of the Royal Meteorological Society*, *142*(696), 1413–1427. https://doi.org/10.1002/qj.2743

Cassou, C. (2008). Intraseasonal interaction between the Madden–Julian Oscillation and the North Atlantic Oscillation. *Nature*, *455*(7212), 523–527. https://doi.org/10.1038/nature07286

Dee, D. P., Uppala, S. M., Simmons, A. J., Berrisford, P., Poli, P., Kobayashi, S., et al. (2011). The ERA-Interim reanalysis: Configuration and performance of the data assimilation system. *Quarterly Journal of the Royal Meteorological Society*, *137*(656), 553–597. https://doi.org/10.1002/qj.828

Domeisen, D. I. V., Butler, A. H., Charlton-Perez, A. J., Ayarzagüena, B., Baldwin, M. P., Dunn-Sigouin, E., et al. (2020b). The role of the stratosphere in subseasonal to seasonal prediction: 2. Predictability arising from stratosphere-troposphere coupling. *Journal of Geophysical Research: Atmospheres*, *125*, e2019JD030923. https://doi.org/10.1029/2019JD030923

Domeisen, D. I. V., Butler, A. H., Charlton-Perez, A. J., Ayarzagüena, B., Baldwin, M. P., Dunn-Sigouin, E., et al. (2020a). The role of the stratosphere in subseasonal to seasonal prediction: 1. Predictability of the stratosphere. *Journal of Geophysical Research: Atmospheres*, *125*, e2019JD030920. https://doi.org/10.1029/2019JD030920

Domeisen, D. I. V., Butler, A. H., Fröhlich, K., Bittner, M., Müller, W., & Baehr, J. (2015). Seasonal predictability over Europe arising from El Niño and stratospheric variability in the MPI-ESM seasonal prediction system. *Journal of Climate*, *28*, 256–271. https://doi.org/10.1175/JCLI-D-14-00207.1

Domeisen, D. I. V., Garfinkel, C. I., & Butler, A. H. (2018). The teleconnection of El Niño Southern Oscillation to the stratosphere. *Reviews of Geophysics*, *57*, 5–47. https://doi.org/10.1029/2018RG000596

Domeisen, D. I. V., Grams, C. M., & Papritz, L. (2020). The role of North Atlantic-European weather regimes in the surface impact of sudden stratospheric warming events. *Weather and Climate Dynamics*, *1*(2), 373–388. http://doi.org/10.5194/wcd-1-373-2020

Feldstein, S. B. (2000). Teleconnections and ENSO, the timescale, power spectra, and climate noise properties. *Journal of Climate*, *13*, 4430–4440. https://doi.org/10.1175/1520-0442(2000)013<4430:ttpsac>2.0.co;2

Hoerling, M. P., Hurrell, J. W., Xu, T., Bates, G. T., & Phillips, A. S. (2004). Twentieth century North Atlantic climate change. Part II: Understanding the effect of Indian Ocean warming. *Climate Dynamics*, *23*, 391–405. https://doi.org/10.1007/s00382-004-0433-x

Hurrell, J. W. (1995). Decadal trends in the North Atlantic Oscillation: Regional temperatures and precipitation. *Science*, *269*(5224), 676–679. https://doi.org/10.1126/science.269.5224.676

Hurrell, J. W. (1996). Influence of variations in extratropical wintertime teleconnections on Northern Hemisphere temperature. *Geophysical Research Letters*, *23*(6), 665–668. https://doi.org/10.1029/96GL00459

Hurrell, J. W., & Deser, C. (2009). North Atlantic climate variability: The role of the North Atlantic Oscillation. *Journal of Marine Systems*, *78*, 28–41. https://doi.org/10.1016/j.jmarsys

Hurrell, J. W., Hack, J. J., Shea, D., Caron, J. M., & Rosinski, J. (2008). A new sea surface temperature and sea ice boundary dataset for the community atmosphere model. *Journal of Climate*, *21*, 5145–5153. https://doi.org/10.1175/2008JCLI2292.1

Hurrell, J. W., Hoerling, M. P., Phillips, A. S., & Xu, T. (2004). Twentieth century North Atlantic climate change. Part I: Assessing determinism. *Climate Dynamics*, *23*, 371–389. https://doi.org/10.1007/s00382-004-0432-y

Hurrell, J. W., Holland, M. M., Gent, P. R., Ghan, S., Kay, J. E., Kushner, P. J., et al. (2013). The Community Earth System Model: A framework for collaborative research. *Bulletin of the American Meteorological Society*, *94*(9), 1339–1360. https://doi.org/10.1175/bams-d-12-00121.1

Jia, L., Yang, X., Vecchi, G., Gudgel, R., Delworth, T., Fueglistaler, S., et al. (2017). Seasonal prediction skill of northern extratropical surface temperature driven by the stratosphere. *Journal of Climate*, *30*, 4463–4475. https://doi.org/10.1175/JCLI-D-16-0475.1

Johansson, A. (2007). Prediction skill of the NAO and PNA from daily to seasonal timescale. *Journal of Climate*, *20*, 1957–1975. https://doi.org/10.1175/JCLI4072.1

Kolstad, E. W., Wulff, C. O., Domeisen, D., & Woolings, T. (2020). Tracing North Atlantic Oscillation forecast errors to stratospheric origins. *Journal of Climate*, *33*(21), 9145–9157. http://doi.org/10.1175/JCLI-D-20-0270.1

Kumar, A., & Hoerling, M. P. (1995). Prospects and limitation of seasonal atmospheric GCM predictions. *Bulletin of the American Meteorological Society*, *76*(3), 335–345. http://doi.org/10.1175/1520-0477(1995)076<0335:PALOSA>2.0.CO;2

Mariotti, A., Baggett, C., Barnes, E. A., Becker, E., Butler, A., Collins, D. C., et al. (2020). Windows of opportunity for skillful forecasts subseasonal to seasonal and beyond. *Bulletin of the American Meteorological Society*, *101*, E608–E625. https://doi.org/10.1175/BAMS-D-18-0326.1

Merryfield, W. J., Baehr, J., Batté, L., Becker, E. J., Butler, A. H., Coelho, C. A., et al. (2020). Current and emerging developments in subseasonal to decadal prediction. *Bulletin of the American Meteorological Society*, *101*(6), E869–E896. https://doi.org/10.1175/BAMS-D-19-0037.1

Miller, D. E., & Wang, Z. (2019). Assessing seasonal predictability sources and windows of high predictability in the climate forecast system, version 2. *Journal of Climate*, *32*, 1307–1326. https://doi.org/10.1175/JCLI-D-18-0389.1

Mudelsee, M. (2010). *Climate time series analysis: Classical statistical and bootstrap methods* (1st ed., p. 474). Dordrecht: Springer.

Neale, R. B., Chen, C. C., Gettelman, A., Lauritzen, P. H., Park, S., Williamson, D. L., ... & Marsh, D. (2012). Description of the NCAR community atmosphere model (CAM 5.0). NCAR, *Technical Note*, NCAR, TN 486. (http://www.cesm.ucar.edu/models/cesm1.0/cam/docs/description/cam5_desc.pdf).

Newman, M., & Sardeshmukh, P. D. (2008). Tropical and stratospheric influences on extratropical short-term climate variability. *Journal of Climate*, *21*(17), 4326–4347. https://doi.org/10.1175/2008JCLI2118.1

Nie, Y., Scaife, A. A., Ren, H.-L., Comer, R. E., Andrews, M. B., Davis, P., & Marin, N. (2019). Stratospheric initial conditions provide seasonal predictability of the North Atlantic and Arctic Oscillations. *Environmental Research Letters*, *14*(3), 034006. https://doi.org/10.1088/1748-9326/ab0385

O'Reilly, C. H., Weisheimer, A., Woollings, T., Gray, L. J., & MacLeod, D. (2019). The importance of stratospheric initial conditions for winter North Atlantic Oscillation predictability and implications for the signal-to-noise paradox. *Quarterly Journal of the Royal Meteorological Society*, *145*(718), 131–146. https://doi.org/10.1002/qj.3413

Pegion, K., Kirtman, B. P., Becker, E., Collins, D. C., LaJoie, E., Burgman, R., et al. (2019). The Subseasonal Experiment (SubX): A multi-model subseasonal prediction experiment. *Bulletin of the American Meteorological Society*, *100*, 2043–2060. https://doi.org/10.1175/BAMS-D-18-0270.1

Reichler, T., Kim, J., Manzini, E., & Kröger, J. (2012). A stratospheric connection to Atlantic climate variability. *Nature Geoscience*, *5*(11), 783–787. https://doi.org/10.1038/ngeo1586

Reynolds, R. W., Smith, T. M., Liu, C., Chelton, D. B., Casey, K. S., & Schlax, M. G. (2007). Daily high-resolution-blended analyses for sea surface temperature. *Journal of Climate*, *20*(22), 5473–5496. https://doi.org/10.1175/2007JCLI1824.1

Richter, J. H., Deser, C., & Sun, L. (2015). Effects of stratospheric variability on El Niño teleconnections. *Environmental Research Letters*, *10*(12), 124021. https://doi.org/10.1088/1748-9326/10/12/124021

Richter, J. H., Pegion, K., Sun, L., Kim, H., Caron, J. M., Glanville, A., et al. (2020). Subseasonal prediction with and without a well-represented stratosphere in CESM1. *Weather Forecasting*. https://doi.org/10.1175/WAF-D-20-0029.1

Riddle, E. E., Stoner, M. B., Johnson, N. C., L'Heureux, M. L., Collins, D. C., & Feldstein, S. B. (2013). The impact of the MJO on clusters of wintertime circulation anomalies over the North American region. *Climate Dynamics*, *40*, 1749–1766. https://doi.org/10.1007/s00382-012-1493-y

Saha, S., Moorthi, S., Wu, X., Wang, J., Nadiga, S., Tripp, P., et al. (2014). The NCEP Climate Forecast System Version 2. *Journal of Climate*, *27*, 2185–2208. https://doi.org/10.1175/JCLI-D-12-00823.1

Scaife, A. A., Arribas, A., Blockley, E., Brookshaw, A., Clark, R. T., Dunstone, N., et al. (2014). Skillful long-range prediction of European and North American winters. *Geophysical Research Letters*, *41*, 2514–2519. https://doi.org/10.1002/2014GL059637

Scaife, A. A., Karpechko, A. Y., Baldwin, M. P., Brookshaw, A., Butler, A. H., Eade, R., et al. (2016). Seasonal winter forecasts and the stratosphere. *Atmospheric Science Letters*, *17*(1), 51–56. https://doi.org/10.1002/asl.598

Screen, J. (2017). The missing Northern European winter cooling response to Arctic Sea ice loss. *Nature Communications*, *8*, 14603. https://doi.org/10.1038/ncomms14603

Sigmond, M., Scinocca, J., Kharin, V., & Shepherd, T. (2013). Enhanced seasonal forecast skill following stratospheric sudden warmings. *Nature Geoscience*, *6*(2), 98–102. https://doi.org/10.1038/ngeo1698

Smith, D. M., Eade, R., Scaife, A. A., Caron, L. P., Danabasoglu, G., DelSole, T. M., et al. (2019). Robust skill of decadal climate predictions. *npj Climate and Atmospheric Science*, *2*, 1–10. https://doi.org/10.1038/s41612-019-0071-y

Smith, D. M., Scaife, A. A., Eade, R., Athanasiadis, P., Bellucci, A., Bethke, I., et al. (2020). North Atlantic climate far more predictable than models imply. *Nature*, *583*(7818), 796–800. https://doi.org/10.1038/s41586-020-2525-0

Tripathi, O. P., Baldwin, M., Charlton-Perez, A., Charron, M., Eckermann, S. D., Gerber, E., et al. (2015). Review: The predictability of the extra-tropical stratosphere on monthly timescales and its impact on the skill of tropospheric forecasts. *Quarterly Journal of the Royal Meteorological Society*, *141*(689), 987–1003. https://doi.org/10.1002/qj.2432

Tripathi, O. P., Charlton-Perez, A., Sigmond, M., & Vitart, F. (2015). Enhanced long-range forecast skill in boreal winter following stratospheric strong vortex conditions. *Environmental Research Letters*, *10*(10), 104007. https://doi.org/10.1088/1748-9326/10/10/104007

Tsujino, H., Urakawa, S., Nakano, H., Small, R. J., Kim, W. M., Yeager, S. G., et al. (2018). JRA-55 based surface dataset for driving ocean-sea-ice models. *Ocean Modelling*, *130*, 79–139. https://doi.org/10.1016/j.ocemod.2018.07.002

Vitart, F. (2014). Evolution of ECMWF sub-seasonal forecast skill scores. *Quarterly Journal of the Royal Meteorological Society*, *140*(683), 1889–1899. https://doi.org/10.1002/qj.2256

Vitart, F., Ardilouze, C., Bonet, A., Brookshaw, A., Chen, M., Codorean, C., et al. (2017). 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

Vitart, F., Robertson, A., & Anderson, D. (2012). Sub-seasonal to Seasonal Prediction Project: Bridging the gap between weather and climate. *WMO Bulletin*, *61*, 23–28.

Wang, L., Ting, M., & Kushner, P. J. (2017). A robust empirical seasonal prediction of winter NAO and surface climate. *Scientific Reports*, *7*(1), 279. https://doi.org/10.1038/s41598-017-00353-y

Weisheimer, A., Decremer, D., Macleod, D., O'Reilly, C., Stockdale, T. N., Johnson, S., & Palmer, T. N. (2018). How confident are predictability estimates of the winter North Atlantic Oscillation? *Quarterly Journal of the Royal Meteorological Society*, *145*(S1), 140–159.

Weisheimer, A., Palmer, T. N., & Doblas-Reyes, F. J. (2011). Assessment of representations of model uncertainty in monthly and seasonal forecast ensembles. *Geophysical Research Letters*, *38*, L16703. https://doi.org/10.1029/2011GL048123

White, C. J., Carlsen, H., Robertson, A. W., Klein, R. J., Lazo, J. K., Kumar, A., et al. (2017). Potential applications of subseasonal-to-seasonal (S2S) predictions. *Meteorological Applications*, *24*(3), 315–325. https://doi.org/10.1002/met.1654

WWRP/WCRP (2018). *WWRP/WCRP Sub-seasonal to Seasonal Prediction Project (S2S) Phase II Proposal. WWRP 2018–4*, WCRP Report No. 11/2018. Available online at http://s2sprediction.net/file/documents_reports/P2_Pro.pdf

Xiang, B., Lin, S.-J., Zhao, M., Johnson, N. C., Yang, X., & Jiang, X. (2019). Subseasonal week 3–5 surface air temperature prediction during boreal wintertime in a GFDL model. *Geophysical Research Letters*, *46*, 416–425. https://doi.org/10.1029/2018GL081314