Attribution of NAO Predictive Skill beyond Two Weeks in Boreal Winter

Lantao Sun[1*], Judith Perlwitz[2], Jadwiga (Yaga) Richter[3],

Martin Hoerling[2], James W. Hurrell[1]

[1] Department of Atmospheric Science, Colorado State University, Fort Collins, CO

[2]NOAA Physical Sciences Laboratory, Boulder, CO

[3]National Center for Atmospheric Research, Boulder, CO

Submitted to *Geophysical Research Letters*

Aug 19, 2020

[*]Corresponding author: Lantao Sun (lantao.sun@colostate.edu)

Key Points:

- Week 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 attributed partly to ENSO

Key words: NAO predictive skill, attribution, stratosphere, ENSO

Abstract

The week 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 six weeks is higher following extreme stratospheric polar vortex conditions (weak and strong vortex events) compared to neutral states. Enhanced week 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

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 six weeks following extreme weak or strong states of the stratospheric polar vortex compared to stratospheric neutral vortex states. Enhanced week 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 (White et al. 2017; Smith et al. 2019; Merryfield et al. 2020; Mariotti et al. 2020; Meehl et al. 2020). Winter climate over Europe and eastern North America is significantly affected by the variability of the NAO (Hurrell, 1995; 1996; Hurrell and 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., Hurrell et al. 2004; Hoerling 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., Cassou 2008; Newman and Sardeshmukh, 2008; Scaife et al., 2014; Brunet et al. 2015). Evidence suggests that stratospheric processes and associated variability are important (Tripathi et al., 2015a; Scaife et al., 2016; Wang et al. 2017; Jia et al. 2017; O'Reilly et al. 2018; Xiang et al. 2019; Nie et al. 2019), including the contribution of the stratospheric pathway associated with the El Niño-Southern Oscillation (ENSO) phenomenon (Domeisen et al., 2015; Butler et al., 2016; Domeisen et al. 2018).

Extreme stratospheric polar vortex states (i.e., weak and strong vortex events) are followed by anomalous near-surface NAO conditions (Baldwin and Dunkerton, 2001) implying enhanced

NAO prediction skill could occur as forecasts-of-opportunity contingent on initial stratospheric states (Albers and Newman 2019). For instance, Sigmond et al. (2013) demonstrated that the skill of NAO forecasts averaged over 15-60-day periods is substantially enhanced when forecasts are initialized at the onset time of weak stratospheric polar vortex events. Tripathi et al. (2015b) 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 (WWRP/WCRP, 2018; Domeisen et al. 2020a, 2020b).

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. Surface boundary-forced NAO predictability can be inferred from uninitialized Atmosphere Model Intercomparison Project (AMIP) simulations prescribed with observed SST and sea ice conditions. By comparing the skill in a climate prediction system with its corresponding AMIP simulations, the relative importance of atmospheric (e.g., extreme stratospheric vortex states) and lower 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 isolate and quantify boundary forced predictability.

2. Model reforecasts and AMIP simulations

a. 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., 2002). The ocean and sea ice initial conditions come 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.

The CESM reforecast ensembles are generated using the random field initialization method (Magnusson et al., 2009). There are 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 improve the 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 4-member CFS reforecast ensemble conducted every day for 1999-2010 and a 11-member ECMWF reforecast ensemble conducted twice per week for 1996-2015, both from the international S2S database (Vitart et al., 2016; 2017). Recognizing that the models have somewhat

different hindcast periods, all the findings presented here have been re-calculated using their common period of 1999-2010 and the results were not materially different.

b. NAO skill evaluation

We use the bias correction method used in SubX, where the daily anomaly for each variable is obtained by removing the daily climatology at different time lags (Pegion et al., 2019). The predictive skill is evaluated by the anomaly correlation coefficient (ACC) that has been commonly used in S2S forecasts (e.g., Tripathi et al., 2015b). Namely,

$$ACC = \frac{\sum x'_{model} x'_{obs}}{\sqrt{\sum (x'^2_{model} \sum x'^2_{obs})}}$$

where x'_{model} and x'_{obs} represent the weekly or monthly anomaly for models and observations (ERA-Interim reanalysis; Dee et al., 2002), respectively. We also calculate the root-mean-square error (RMSE) skill score between the model and observations (Wilks, 1995) and find smaller values but with roughly the same pattern as ACC (not shown).

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 (day 15-42 average). Discrete weekly skill for lead times of 1- 6-weeks is also diagnosed. Similar to other studies (e.g., Johansson, 2007; Butler et al., 2016) we conduct Empirical Orthogonal Function (EOF) analysis for ERA-Interim monthly (November-March) SLP anomalies over the Atlantic sector (20^{o}N-80^{o}N; 90^{o}W-40^{o}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 10000 times with replacement and obtaining the 5% and 95% of the ACC significance levels (Mudelsee 2010).

140 The role of initial persistence in NAO predictability is evaluated by the correlation of the
141 NAO anomaly relative to its week 0 value (day -3 to day 3 average), and the resulting monthly
142 persistence skill is compared to that from the initialized dynamical models. Also, to isolate the
143 role of surface lower boundary forcing, we utilize 50-member free-running atmosphere model
144 simulations with CAM5 forced by observed greenhouse gases, SST and sea ice conditions (Hurrell
145 et al., 2008). The ensemble-mean anomaly for the uninitialized AMIP forecast is obtained by
146 removing the daily long-term climatology and then calculating the ACC against observations.
147 Again, the analysis is of monthly mean variability, and the simulation skill is compared to the
148 week 3-6 skill from initialized forecasts and from simple persistence.

149 To explore the role of the stratosphere in surface NAO predictability, we subsample the winter
150 forecast into three categories based on the initial stratospheric zonal-mean zonal winds at 10 hPa
151 and 60°N. Similar to Tripathi et al., (2015b), we first calculate the probability density function
152 (PDF) of the 1999-2015 10-hPa 60°N November-March zonal-mean zonal wind distribution. The
153 weak (strong) stratospheric polar vortex events can be defined for forecasts whose initial zonal
154 wind at 10-hPa 60°N is below 10% of the PDF (3.55 m/s) or above 80% of the PDF (40.79 m/s).
155 Note that slightly different thresholds are used so that the number of extreme events (60 weak and
156 62 strong vortex events) are comparable, but the results are not sensitive to the exact choice of
157 thresholds. Figure S1 shows the composite of standardized polar-cap geopotential height and SLP
158 anomalies for weak and strong vortex events, indicating that the observed downward coupling of
159 extreme polar vortex events and their surface NAO features are both well captured in CESM.
160 Neutral vortex events are defined for forecasts whose initial zonal wind at 10-hPa 60°N is between
161 30% and 70% of the PDF (17.18 m/s – 35.74 m/s; 154 cases). The predictive skills 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 week 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, 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 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 Supplemental Figure S2).

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., Kumar and Hoerling 1995; Butler et al. 2016; Athanasiadis et al. 2020; 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 1000-times bootstrapping to calculate the 5% and 95% statistical significance level of the NAO skill. Figure 1b shows the week 3-6 monthly NAO skill in CESM as a function of ensemble size. In agreement

185 with those earlier studies, the NAO skill approximately doubles (0.27 versus 0.51) when the
186 ensemble size increases from 1 to 20, but it is close to saturation at 15 ensemble members (e.g.,
187 there is a less than 0.01 increase from 15 to 20 members).

188 Thus, when accounting for the differences in ensemble size of the systems diagnosed herein,
189 the lower NAO skill in the smaller CFS ensemble is reconcilable with sampling alone; its skill
190 overlaps with that of CESM when the latter's ensemble is sub-sampled to match the CFS
191 population. That said, the lower skill with the CFS model might also be related to the fact that the
192 stratospheric polar vortex is poorly predicted and stratosphere-troposphere coupling is too weak
193 (Miller and Wang 2019). Evidence to support this argument for degraded skill is presented in the
194 next section where we demonstrate an appreciable sensitivity of skill to stratospheric conditions
195 overall.

196 As a prelude to a more detailed analysis of the attributable causes (sources) for week 3-6
197 NAO prediction skill, the initialized prediction skill of CESM is compared to that resulting from
198 simple persistence, and that arising from lower boundary forcing alone with no recourse to the
199 initial weather state. Summarized in Supplemental Figure S3, the monthly NAO skill in CESM
200 hindcasts (~0.5) is found to be considerably larger than in both persistence forecasts (~0.3) and the
201 simulation skill arising solely from effects of prescribed surface lower-boundary forcing (~0.05).
202 It is interesting to note from their NH distributions of SLP skill (Supplemental Fig. S3, left) that
203 the AMIP simulations have more than double the correlation skill of persistence averaged over the
204 NH overall. But persistence skill is quite high over the two NAO centers of variability in the
205 Atlantic basin, while AMIP skill is virtually absent over the NAO's northern node. This overall
206 estimate of skill sources, while informative, obscures a considerable conditionality of the skill by
207 each source.

208

209 *3.2. Sensitivity to the stratospheric initial states*

210 To explore the role of initial stratospheric states on NAO prediction skill, we subsample the
211 CESM winter reforecasts into weak (60), neutral (154) and strong (62) polar vortex events. Figure
212 2a shows the monthly time series of 10 hPa zonal winds averaged along 60°N, the extreme values
213 of which are used to identify strong (blue dots) and weak (red dots) vortex events during 1999-
214 2015. The weekly NAO skill using this stratification is displayed in Figure 2b. In agreement with
215 earlier studies, the NAO skill is generally higher in weeks 3–6 when initialized from extreme
216 stratospheric polar vortex conditions compared to neutral states. Especially striking is the sustained
217 high skill emerging from initial weak polar vortex conditions, having correlations above 0.5
218 through week 6. By week 6, the skill improvement relative to neutral states (0.23) is appreciable
219 and statistically significant, indicating that NAO skill is substantially enhanced when initialized
220 during weak vortex events, a result based on CESM that affirms prior findings based on other
221 forecast systems (Sigmond et al. 2013; Tripathi et al. 2015; Domeisen et al. 2020b).

222 The results indicate that NAO skill when forecasts begin from strong and weak vortex events
223 is comparable - and is moderately elevated compared to neutral vortex states during weeks 3 and
224 4. However, the skill decreases quickly in weeks 5 and 6 for strong vortex initializations, so that
225 the improvement relative to neutral states becomes marginal (Figure 2b). This skill progression
226 can be explained by a difference between CESM and observed NAO life cycles during strong
227 vortex events. In weeks 5-6 following strong vortex events, the canonical positive NAO pattern is
228 evident in CESM but not in observations (see Supplemental Figure S1b). We speculate that the
229 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. 2013).

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 versus 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 two weeks (Figure 3b). The high skill for weak polar vortex initial states is not an outcome of particularly high persistence of the NAO during such states. Rather, NAO persistence skill following weak vortex initial conditions is significantly lower than for strong vortex states. There is also statistically significant asymmetry in AMIP simulation skill when binned according to reforecast polar vortex intensity. It is important to 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. Yet, the results of Fig. 3a can only be understood if those 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. Note that the absence of

253 persistence skill implies a fundamentally nonlinear dynamical evolution in the forecasts that
254 cannot be represented via simple persistence methods alone. By contrast, both persistence and
255 lower boundary forcing contribute to monthly NAO skill when the initial polar vortex is strong.
256 An explanation for the role of ocean forcing is provided by an analysis that stratifies the
257 occurrences of polar vortex events according to ENSO phase. The Niño 3.4 index in relation to
258 weak and strong polar vortex events is shown in Fig. 4a. The relationship is not linear and is
259 complicated. For instance, many strong vortex events appear to happen in either El Niño or La
260 Niña years (e.g., 1999/2000, 2007/2008, 2009/2010, 2010/2011, 2015/2016), though this is less
261 evident for weak and neutral events. This is evident from the histograms of Niño3.4 SST states
262 stratified according to polar vortex intensity (Fig. 4b). The spread of the distribution for strong
263 vortex states is wide and highly non-Gaussian (blue curve), especially compared to that for weak
264 polar vortex states (red curve). It suggests that the NAO predictive skill following strong vortex
265 events is influenced by ENSO, agreeing with the recent findings of Xiang et al. (2019) who
266 identified ENSO as the most predictable subseasonal mode and found that its influence in the
267 North Atlantic resembles the NAO pattern. However, further research is needed to determine why
268 many strong vortex events coincide with both El Niño and La Niña years.
269 The lower panel of Fig. 4 provides a spatial view of the monthly SLP skill patterns related to
270 effects of initial stratospheric upper air conditions and ENSO lower boundary conditions. Here
271 we subsample the winter cases into extreme stratospheric polar vortex conditions (both weak and
272 strong), ENSO cases (both El Niño and La Niña, defined by Niño3.4 index above/below +/-0.5,
273 respectively), and neutral stratosphere-ENSO cases (Figure 4b). Note that the ENSO cases have
274 been excluded from the composites of extreme stratospheric vortex states, and extreme polar
275 vortex events have been excluded from the ENSO cases. Immediately apparent is that SLP skill

for neutral stratosphere and neutral ENSO cases is negligibly low (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 and the polar vortex. 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. These results demonstrate that both extreme stratospheric polar vortex states and extreme lower boundary forcing linked to ENSO account for week 3-6 NAO predictive skill during boreal winter.

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 six 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 can be significantly 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 running 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 still 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.

Acknowledgements

This study was funded by NOAA's Climate Program Office (CPO) Modeling, Analysis, Predictions and Projections (MAPP). This work was supported by 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. We thank Yan Wang at NOAA's

322 Physical Sciences Laboratory for her help with downloading CFS and ECMWF hindcast data and
323 Dr. Zac Lawrence for helpful comments on an earlier version of this paper. The CESM reforecast
324 data is available to the public via Columbia University's International Research Institute for
325 Climate and Society (IRI) library at
326 http://iridl.ldeo.columbia.edu/SOURCES/.Models/.SubX/.CESM/. The CAM5 simulations are
327 available at PSL's Facility for Weather and Climate Assessments (FACTS) website at
328 https://www.esrl.noaa.gov/psl/repository/facts/. The CFS and ECMWF data is available at
329 ECMWF's S2S portal https://apps.ecmwf.int/datasets/data/s2s/levtype=sfc/origin=ecmf/type=cf/.

Reference

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. et al. (2020): Decadal predictability of North Atlantic blocking and the NAO. *npj Clim Atmos Sci*, 3, 20. https://doi.org/10.1038/s41612-020-0120-6.

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.

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

Butler, A. H., Arribas, A., Athanassiadou, M., Baehr, J., Calvo, N., Charlton-Perez, and co-authors (2016): The Climate-system Historical Forecast Project: do stratosphere-resolving models make better seasonal climate predictions in boreal winter? *Q.J.R. Meteorol. Soc.,* 142, 1413–1427. https://doi.org/10.1002/qj.2743

Butler, A. H., and L. M. Polvani, (2011): El Niño, La Niña, and strato- spheric sudden warmings: A reevaluation in light of the obser- vational record. *Geophys. Res. Lett.*, 38, L13807, doi:10.1029/2011GL048084.

Butler, A. H., L. M. Polvani and C. Deser, (2014): Separating the stratospheric and tropospheric pathways of El Niño–Southern Oscillation teleconnections. *Environ. Res. Lett.*, 9, 024014, doi:10.1088/ 1748-9326/9/2/024014.

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

Dee, D. P., Uppala, S. M., Simmons, A. J., Berrisford, P., Poli, P., Kobayashi, S., ….Vitart, F. (2011). The ERA-Interim reanalysis: Configuration and performance of the data assimilation system, *Quart. J. Roy. Meteor. Soc.*, 137, 553–597, https://doi.org/10.1002/qj.828

Dennis, J., Edwards, J., Evans, K., Guba, O., Lauritzen, P., Mirin, A., …Worley P. (2012). CAM-SE: A scalable spectral element dynamical core for the Community Atmosphere Model, *Int. J. High Perform. Comput. Appl.*, 26(1), 74–89. https://doi.org/10.1177/1094342011428142.

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

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., 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., C. I.Garfinkel, and A. H. Butler, (2018): The teleconnection of El Niño Southern Oscillation to the stratosphere. *Rev. Geophys.*, 57, 5–47, https://doi.org/10.1029/2018RG000596.

Feldstein, S.B., (2000): Teleconnections and ENSO, the timescale, power spectra, and climate noise properties. *J. Climate* 13, 4430–4440. https://doi.org/10.1175/1520-0442(2000)013%3C4430:TTPSAC%3E2.0.CO;2.

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

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

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

Hurrell, J. W., and C. Deser, (2009): North Atlantic climate variability: The role of the North Atlantic Oscillation. *J. Mar. Syst.*, 78, 28–41, doi: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. *J. Climate*, 21, 5145–53. https://doi.org/10.1175/2008JCLI2292.1

Hurrell, J. W., M. P. Hoerling, A. S. Phillips, and T. Xu, (2004): Twentieth century north Atlantic climate change. Part I: assessing determinism. *Clim. Dyn.*, 23, 371–389, doi:10.1007/s00382-004-0432-y.

Jia, L., and Coauthors, (2017): Seasonal Prediction Skill of Northern Extratropical Surface Temperature Driven by the Stratosphere. *J. Climate*, 30, 4463–4475, https://doi.org/10.1175/JCLI-D-16-0475.1.

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.

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

Kumar, A., and M. P. Hoerling, 1995: Prospects and limitation of seasonal atmospheric GCM predictions. *Bull. Amer. Metor. Soc.*, 76, 335-345.

L'Heureux, M. L., Tippett, M. K., Kumar, A., Butler, A. H., Ciasto, L. M., Ding, Q., ... Johnson, N. C. (2017): Strong relations between ENSO and the Arctic Oscillation in the North American Multimodel Ensemble. *Geophysical Research Letters*, 44, 11,654–11,662. https://doi.org/10.1002/2017GL074854.

Magnusson, Linus, Jonas Nycander & Erland Källén (2009): Flow-dependent versus flow-independent initial perturbations for ensemble prediction, *Tellus A: Dynamic Meteorology and Oceanography*, 61:2, 194-209, DOI: 10.1111/j.1600-0870.2008.00385.x.

Mariotti, A., and Coauthors, (2020): Windows of Opportunity for Skillful Forecasts Subseasonal to Seasonal and Beyond. *Bull. Amer. Meteor. Soc.*, 101, E608–E625, https://doi.org/10.1175/BAMS-D-18-0326.1.

Meehl, G. A., and Coauthors, (2020): Initialized Earth system prediction from subseasonal to decadal timescales, *Nature Reviews Earth and Environment*, in review.

Merryfield, W. J., and Coauthors, (2020): Current and emerging developments in subseasonal to decadal prediction. *Bull. Amer. Meteor. Soc.*, doi: https://doi.org/10.1175/BAMS-D-19-0037.1.

Miller, D. E., and Z. Wang, (2019): Assessing Seasonal Predictability Sources and Windows of High Predictability in the Climate Forecast System, Version 2. *J. 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. First edition. Springer, Dordrecht, 474pp

Neale, R. B. and Coauthors (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, 4326. https://doi.org/10.1175/2008JCLI2118.1

Nie, Y., A. A. Scaife, H-L Ren, R. E Comer, M. B Andrews, P. Davis and N. Marin, (2019): Stratospheric initial conditions provide seasonal predictability of the North Atlantic and Arctic Oscillations. *Environ. Res. Lett.* 14 034006.

O'Reilly CH, Weisheimer A, Woollings T, Gray LJ, MacLeod D. (2019): The importance of stratospheric initial conditions for winter North Atlantic Oscillation predictability and implications for the signal-to-noise paradox. *Q J R Meteorol Soc*. 145:131–146. https://doi.org/10.1002/qj.3413.

Pegion, K., and Coauthors, (2019): The Subseasonal Experiment (SubX): A multi-model subseasonal prediction experiment. *Bull. Amer. Meteor. Soc*., 100, 2043–2060, https://doi.org/ 10.1175/BAMS-D-18-0270.1.

Polvani, L., Sun, L., Butler, A. H., Richter, J. H., & Deser, C. (2017). Distinguishing stratospheric sudden warmings from ENSO as key drivers of wintertime climate variability over the North Atlantic and Eurasia, *J. Climate*, 30(6), 1959-1969. https://doi.org/10.1175/JCLI-D-16-0277.1

Reichler, T., Kim, J., Manzini, E. et al. (2012): A stratospheric connection to Atlantic climate variability. Nature Geosci 5, 783–787. https://doi.org/10.1038/ngeo1586

Richter, J. H., C. Deser, and L. Sun, (2015): Effects of stratospheric variability on El Niño teleconnections. *Environ. Res. Lett*., 10, 124021, doi:10.1088/1748-9326/10/12/124021.

Richter, J. H., K. Pegion, L. Sun, H. Kim, J. M. Caron, A. Glanville, S. Yeager, W. Kim and A. Tawfik, (2020): Subseasonal prediction with CESM1 and the role of stratospheric variability on subseasonal prediction skill, *Weather and Forecasting*, In review.

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., S. Moorthi, X. Wu, J. Wang, and Coauthors, (2014): The NCEP Climate Forecast System Version 2. *Journal of Climate*, 27, 2185–2208, doi:10.1175/JCLI-D-12-00823.1.

Scaife, A. A., et al. (2014), Skillful long-range prediction of European and North American winters, *Geophys. Res. Lett*., 41, 2514–2519, doi: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., doi:10.1002/asl.598.

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

Sigmond, M., J. Scinocca, V. Kharin, and T. Shepherd, (2013): Enhanced seasonal forecast skill following stratospheric sudden warmings. *Nat. Geosci*., 6, 98–102, doi:10.1038/ngeo1698.

Smith, D. M., and Coauthors, (2019): Robust skill of decadal climate predictions. *npj Clim. Atmos*. Sci., 2, 1–10, doi:10.1038/s41612-019-0071-y.

Smith, D.M., Scaife, A.A., Eade, R. et al. (2020): North Atlantic climate far more predictable than models imply. *Nature*, 583, 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., Harrison, R. G., Jackson, D. R., Kim, B.-M., Kuroda, Y., Lang, A., Mahmood, S., Mizuta, R., Roff, G., Sigmond, M. and Son, S.-W. (2015a) 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). pp. 987-1003. doi: 10.1002/qj.2432

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

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

Vitart, F. 2014: Evolution of ECMWF sub-seasonal forecast skill scores. *Quart. J. Roy. Meteor. Soc*, 140, 1889-1899.

Vitart F and Coauthors 2016. The sub-seasonal to seasonal prediction (S2S) project database. *Bull. Am. Meteorol. Soc*. 98: 163–173. https://doi.org/10.1175/BAMS-D-16-0017.1.

Vitart, F. and Coauthors (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. Sub-seasonal to Seasonal Prediction Project, 2012: bridging the gap between weather and climate. *WMO Bull*. 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.

White, Christopher J. et al., (2017): Potential applications of subseasonal-to-seasonal (S2S) predictions , *Met. Apps*, 24: 315-325. doi:10.1002/met.1654.

Wilks, D. S. (1995), Statistical Methods in the Atmospheric Sciences: An Introduction, 467 pp., Academic, San Diego, Calif.

WWRP/WCRP, 2018: WWRP/WCRP Sub-seasonal to Seasonal Prediction Project (S2S) Phase II Proposal.

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. https://doi.org/10.1029/2018GL081314.

List of Figures

Figure 1: (a) Sea-level pressure week 3-6 predictive skill evaluated by anomaly correlation coefficient (ACC) during November-March (NDJFM) 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 ensemble size. (b) Week 3-6 North Atlantic Oscillation (NAO) predictive skill as a function of ensemble size. Shading denotes the 5% and 95% statistical level of the CESM reforecast by conducting bootstrapping. Blue (red) error bars indicate the 5% and 95% statistical level of CFS and ECMWF reforecasts, respectively.

Figure 2: (a) Observed Nov-March zonal-mean zonal wind at 10hPa and 60^{o}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 level based on 10000-time bootstrapping.

Figure 3: (a) Sea-level pressure week 3-6 predictive skill for weak, neutral and strong polar vortex events. Values on the left corner denote the average of the ACC averaged over NH extratropics. (b) As in (a), but for the NAO skill in CESM, persistence and uninitialized AMIP simulations. Error bars indicate the 5% and 95% statistical levels by bootstrapping.

Figure 4: (a) Niño3.4 time series and the weak (red dots) and strong (blue dots) polar vortex events and the probability density function of Niño3.4 index for the weak (red), neutral (gray shading) and strong (blue) polar vortex events. In (a) the daily Niño3.4 index is smoothed by 31 days running mean. (b) sea-level pressure predictive skill for initial extreme stratospheric polar vortex events (bottom left; combining weak and strong events), neutral stratosphere-ENSO events (bottom middle) and ENSO events. Values on the left corner denote the skill averaged over NH extratropics.

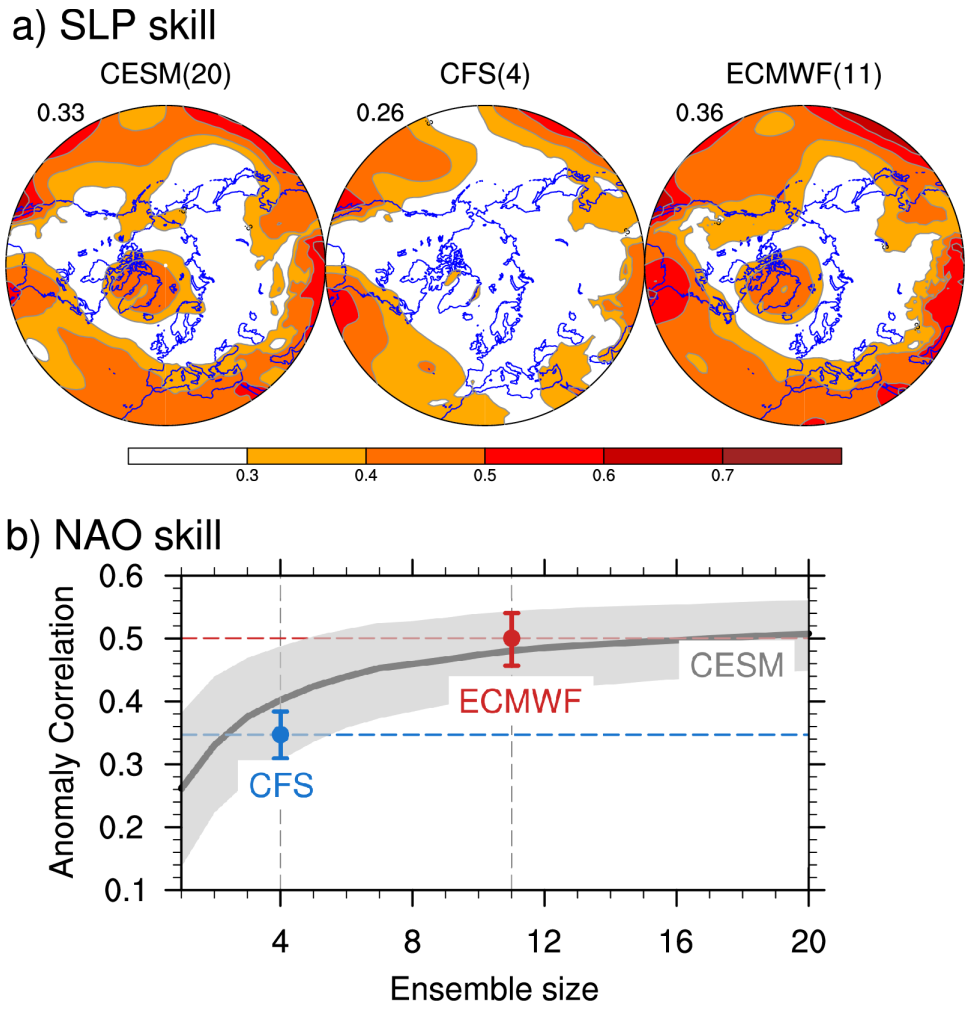

Figure 1: (a) Sea-level pressure week 3-6 predictive skill evaluated by anomaly correlation coefficient (ACC) during November-March (NDJFM) for CESM, CFS and ECMWF reforecasts. Values on the left corner denote the skill averaged over Northern hemisphere (NH) extratropics. Numbers in the brackets of each model show the ensemble size. (b) Week 3-6 North Atlantic Oscillation (NAO) predictive skill as a function of ensemble size. Shading denotes the 5% and 95% statistical level of the CESM reforecast by conducting bootstrapping. Blue (red) error bars indicate the 5% and 95% statistical level of CFS and ECMWF reforecasts, respectively.

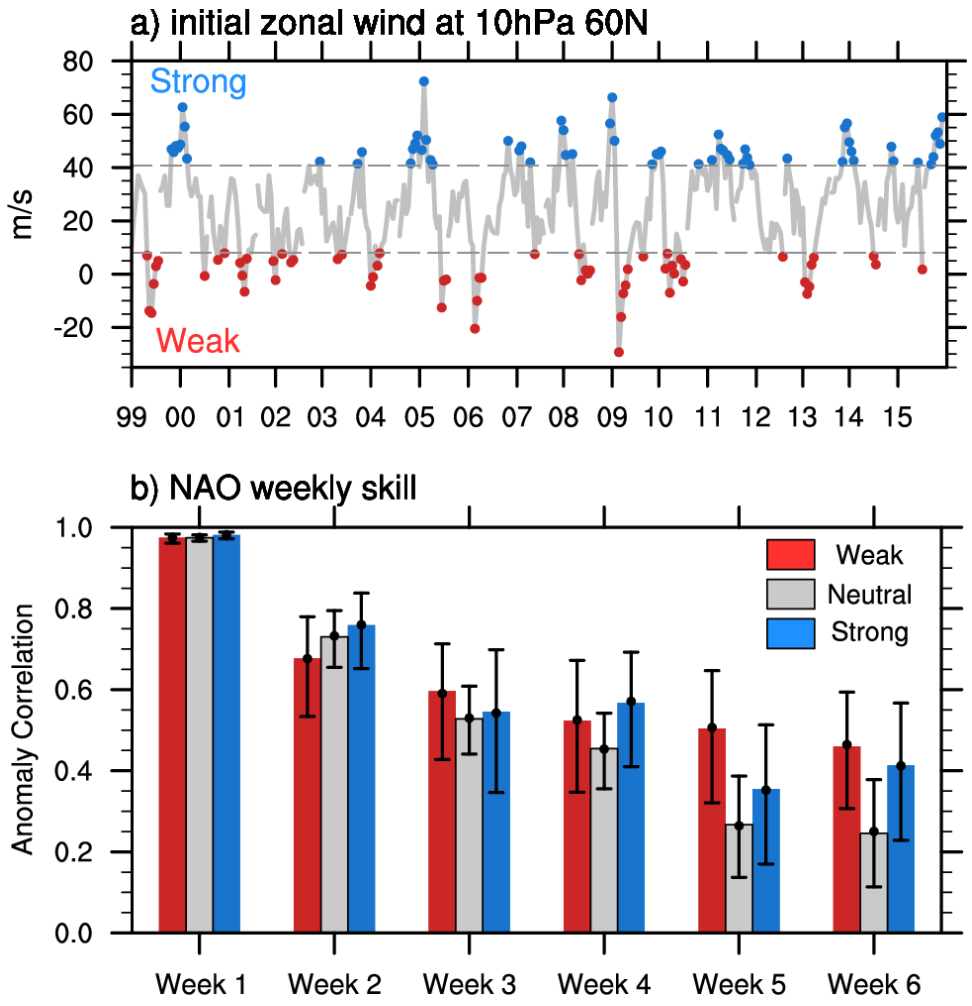

Figure 2: (a) Observed Nov-March zonal-mean zonal wind at 10hPa and 60°N and the weak (red), strong (blue) polar vortex CESM reforecast cases. (b) Weekly NAO skill for the weak (red), strong (blue) stratospheric polar vortex events and neutral polar vortex conditions (gray). Error bars indicate the 5% and 95% statistical level based on 1000-time bootstrapping.

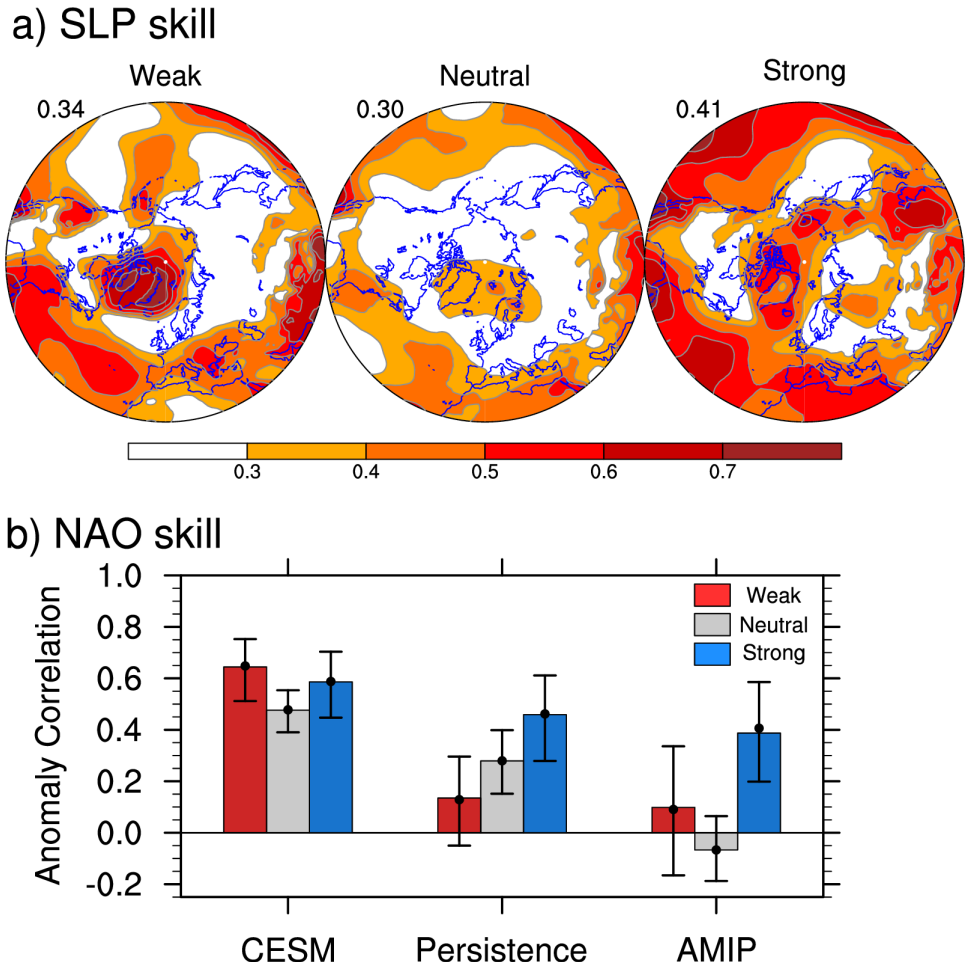

Figure 3: (a) Sea-level pressure week 3-6 predictive skill for weak, neutral and strong polar vortex events. Values on the left corner denote the average of the ACC averaged over NH extratropics. (b) As in (a), but for the NAO predictive skill in CESM, persistence and uninitialized AMIP simulations. Error bars indicate the 5% and 95% statistical levels by bootstrapping.

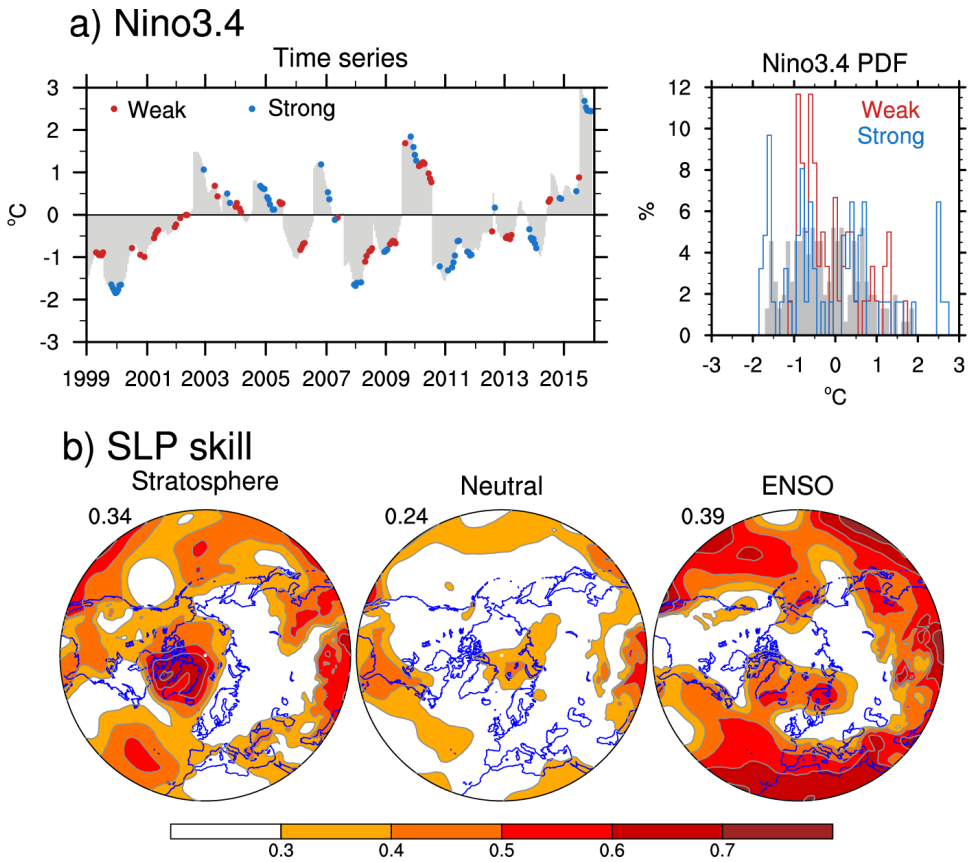

Figure 4: (a) Niño3.4 time series and the weak (red dots) and strong (blue dots) polar vortex events and the probability density function of Niño3.4 index for the weak (red), neutral (gray shading) and strong (blue) polar vortex events. In (a) the daily Niño3.4 index is smoothed by 31 days running mean. (b) sea-level pressure predictive skill for initial extreme stratospheric polar vortex events (bottom left; combining weak and strong events), neutral stratosphere-ENSO events (bottom middle) and ENSO events. Values on the left corner denote the skill averaged over NH extratropics.