ADVANCING EARTH AND SPACE SCIENCE

Geophysical Research Letters

RESEARCH LETTER
10.1029/2021GL095063

Key Points:

- Amplitude and skill of predicted North Atlantic Oscillation (NAO) improve significantly by subsampling of ensemble of existing seasonal prediction systems
- Amplified NAO variability leads to significant improvement in predicting the upcoming winter temperature anomalies in the Northern Hemisphere

Supporting Information:

Supporting Information may be found in the online version of this article.

Correspondence to:
M. Dobrynin,
mikhail.dobrynin@dwd.de

Citation:
Dobrynin, M., Düsterhus, A., Fröhlich, K., Athanasiadis, P., Ruggieri, P., Müller, W. A., & Baehr, J. (2022). Hidden potential in predicting wintertime temperature anomalies in the Northern Hemisphere. *Geophysical Research Letters*, *49*, e2021GL095063. https://doi.org/10.1029/2021GL095063

Received 9 JUL 2021
Accepted 27 SEP 2022

Hidden Potential in Predicting Wintertime Temperature Anomalies in the Northern Hemisphere

Mikhail Dobrynin[1,2], André Düsterhus[3], Kristina Fröhlich[1], Panos Athanasiadis[4], Paolo Ruggieri[4,5], Wolfgang A. Müller[6], and Johanna Baehr[1]

[1]Deutscher Wetterdienst (DWD), Hamburg, Germany, [2]Center for Earth System Research and Sustainability (CEN), Institute of Oceanography, Universität Hamburg, Hamburg, Germany, [3]ICARUS, Department of Geography, Maynooth University, Maynooth, Ireland, [4]Department of Physics and Astronomy, University of Bologna, Bologna, Italy, [5]CMCC - Centro Euro-Mediterraneo sui Cambiamenti Climatici, Bologna, Italy, [6]Max Planck Institute for Meteorology, Hamburg, Germany

Abstract Variability of the North Atlantic Oscillation (NAO) drives wintertime temperature anomalies in the Northern Hemisphere. Dynamical seasonal prediction systems can skilfully predict the winter NAO. However, prediction of the NAO-dependent air temperature anomalies remains elusive, partially due to the low variability of predicted NAO. Here, we demonstrate a hidden potential of a multi-model ensemble of operational seasonal prediction systems for predicting wintertime temperature by increasing the variability of predicted NAO. We identify and subsample those ensemble members which are close to NAO index statistically estimated from initial autumn conditions. In our novel multi-model approach, the correlation prediction skill for wintertime Central Europe temperature is improved from 0.25 to 0.66, accompanied by an increased winter NAO prediction skill of 0.9. Thereby, temperature anomalies can be skilfully predicted for the upcoming winter over a large part of the Northern Hemisphere through increased variability and skill of predicted NAO.

Plain Language Summary Wintertime temperature in the Northern Hemisphere is regulated by the variations of atmospheric pressure, represented by the so-called North Atlantic Oscillation (NAO). The NAO's phase—negative or positive—is associated with the pathways of cold and warm air masses leading to cold or warm winters in Europe. While the NAO phase can be predicted well, predictions of the NAO-dependent air temperature remain elusive. Specifically, it is challenging to predict the strength of the NAO, the most important requirement for the accurate prediction of wintertime temperature. Here, we improve wintertime temperature prediction by increasing the strength of the predicted NAO. We use observation based autumn Northern Hemisphere ocean and air temperature, as well as ice and snow cover for statistical estimation of the first guess NAO for the upcoming winter. Then, we sub-select only those simulations from the multi-model ensemble, which are consistent with our first guess NAO. As a result, based on these selected members, the wintertime temperature prediction is substantially improved over a large part of the Northern Hemisphere.

1. Introduction

The North Atlantic sector has an important impact on weather regimes and the development of wintertime temperature anomalies in Europe and North America (Hertig & Jacobeit, 2014; Vautard, 1990). While ocean and atmosphere act on different time scales, they are both important for the formation of specific winter conditions (Cassou et al., 2004; Rodwell et al., 1999). The large-scale coupled ocean-atmosphere dynamics is well represented by the variability of sea level pressure (SLP) over the North Atlantic, known as the North Atlantic Oscillation (NAO). The winter NAO regimes impact the European wintertime weather not only in terms of the seasonally averaged values of temperature or precipitation (Hurrell, 1995; Hurrell et al., 2003; Thompson et al., 2003), but also in terms of the occurrence of extreme weather conditions (Jung et al., 2011a; Maidens et al., 2013). For example, the extremely cold winter 2009/2010 in Northern and Western Europe was attributed to the record persistence of the negative NAO phase (Cattiaux et al., 2010).

While ensemble-based dynamical seasonal prediction systems (hereafter SPSs) are known to skilfully predict the winter NAO index for a season ahead (Athanasiadis et al., 2017; O'Reilly et al., 2017), they are less successful in the prediction of the NAO-dependent temperature anomalies over the North-Atlantic sector. Increasing ensemble size, on the one hand, improves the prediction skill of the NAO (Butler et al., 2016). On the other hand, this improvement is limited by the ability of models to accurately reproduce the sources of the NAO predictability

(Årthun et al., 2017; Jung et al., 2011b). Recently, a multi-model approach demonstrates an ability to increase the NAO prediction skill by combining several prediction systems into one large ensemble (Athanasiadis et al., 2017). However, for already large ensembles, with about 30–40 members, a further increase of the ensemble size does not only demonstrate any potential for a further significant increase in the prediction skill of the winter NAO but also tends to suppress the variability of predicted NAO index. This can be partly attributed to well-known underestimation of the signal-to-noise ratio in prediction systems (Mayer et al., 2021; Scaife & Smith, 2018) which leads to an underestimation of predicted variability in the ensemble mean. In turn, the strength of the winter NAO phase directly impacts the formation of temperature anomalies, both for positive and negative NAO phases (Heape et al., 2013). Therefore, the low amplitude of the predicted ensemble mean NAO phase decouples the NAO from the formation of temperature anomaly and will produce only weakly pronounced wintertime temperature anomalies.

Here, we demonstrate a hidden potential of existing SPSs in skillful predicting the wintertime temperature anomalies in the Northern Hemisphere by increasing the variability of predicted NAO using a multi-model ensemble subsampling approach. Instead of following the traditional practice of averaging all ensemble members, we make use of the intrinsic memory of the Earth system, analyzing initial autumn conditions to identify ensemble members with well-established relationships between initial autumn conditions and the winter NAO (Dobrynin et al., 2018). Only these ensemble members are considered afterward in a subsampled ensemble mean, resulting in increased variability and prediction skill of the winter NAO index. We make a step forward from the NAO index prediction and predict wintertime temperature anomalies in the Northern Hemisphere using the well-predicted winter NAO index as a criterion for subsampling of a large dynamical ensemble. This reinforces the link between the NAO and temperature anomalies and significantly improves the prediction skill of temperature in the Northern Hemisphere.

2. Prediction Systems, Data, and Methods

2.1. Copernicus Climate Change Service Multi-Model Ensemble

In this study, we use a multi-model ensemble built from five SPSs contributing to Copernicus Climate Change Service (C3S) (hereafter C3S ensemble). The C3S ensemble covers the period from 1994 to 2014 and consists of 138 members provided by the Deutsche Wetterdienst (DWD GCFS2.0-v20171123, 30 members), UK Met Office (UKMO HadGEM3-GC2.0-v20150825, 28 members), European Centre for Medium-Range Weather Forecasts (ECMWF System4, 25 members), Meteo France (System6-v20170501, 15 members), and Centro Euro-Mediterraneo sui Cambiamenti Climatici (CMCC-CM2-v20160423, 40 members). All members are combined in one ensemble of 138 members without implementation of a bias correction procedure.

2.2. NAO Index

The NAO index is calculated using an empirical orthogonal function (EOF) analysis (Barnston & Livezey, 1987). For all systems and for the ERA-Interim, seasonal (DJF) means of SLP are calculated from monthly means prior to the EOF analysis. The region of SLP data is limited to the latitude range 20°N–90°N and to the longitude range 90°W to 60°E. The EOF is calculated for every SPS from all respective ensemble members merged along the time axis into one vector. This approach of EOF calculation allows us to represent the entire ensemble in one ensemble-common EOF pattern. Further, taking into account a relatively short period of hindcasts, this approach is more reliable than conducting the EOF calculation for individual ensemble members separately. The first principal component of SLP is then decomposed back to the number of ensemble members, building an individual time series for each ensemble member. The first principal component of SLP represents the NAO index (Kutzbach, 1970). All NAO indices are normalized by their respective standard deviations. The ERA-Interim NAO index is used as a reference for comparisons with other systems.

2.3. Predictors of the NAO

We use monthly mean data of SLP and 2-m air temperature (T2m) provided by the C3S ensemble. Additionally, SLP, T2m, 100 hPa level air temperature (T100), sea surface temperature in the North Atlantic (SST), Arctic sea ice concentration (SIC) and snow cover in Eurasia (SNC) data are used from the ERA-Interim reanalysis (Dee

10.1029/2021GL095063

et al., 2011). While seasonal, averaged over December, January, and February (DJF), means of SLP and T2m are used for the evaluation of model results, October T100, SST and SNC, and September SIC means represent the autumn predictors of the winter NAO index. All four autumn predictors were previously identified and reported as robust and significant in terms of correlation with the winter NAO: T100 by Domeisen et al. (2015) and Butler and Polvani (2011); SST by Czaja and Frankignoul (2002) and Wang et al. (2017); SNC by Cohen and Jones (2011) and Peings et al. (2013); and SIC by Strong et al. (2009) and Sun et al. (2015). Additionally, in support of predictor choice, in our previous study (Dobrynin et al., 2016) we demonstrated results of subsampling cross-validation, using different sets of selected periods from 1981 to 2016. Moreover, in our recent study on subsampling using other, in addition to NAO, modes of SLP variability (see Supporting Information P.1. in Dalelane et al., 2020), we show that, indeed, those predictors are robust also for an extended period of analysis from 1958 until now.

Generally, the robustness of predictors depends on the actual climate state characterized by a combination of the El Niño-Southern Oscillation, the Arctic oscillation, and the NAO. Some of their phases are modulated by decadal or even by multi-decadal processes and might favorite a predictor and predictability for a particular period (e.g., O'Reilly et al., 2017; Weisheimer et al., 2017). For example, Kolstad and Screen (2019) reported a strong variability of correlation between wintertime NAO and autumn ice conditions in the Arctic under pre-industrial and historical climate. Detailed discussion on the robustness of predictors for the future climate state, specifically with reduced or in absence of sea ice in the Arctic, is out of the scope of this study. However, it is likely that the polar regions will keep their role in the modulation of atmospheric circulation, but the link (predictor) might need reconsideration. Considering the possible non-stationarity of predictors, we use a running training period, covering the full available until the year of the forecast period. This approach keeps predictors dynamic and adaptive to the actual climate state.

Originally, autumn predictors were provided by an assimilation simulation used for hindcast initialization. Since assimilation simulations are not available for all C3S SPSs, in this study we use October T100, SST and SNC, and September SIC from ERA-Interim as predictors of first-guess of the next DJF winter season NAO index for ensemble subsampling as adopted from Dobrynin et al. (2018).

2.4. Subsampling of the C3S Multi-Model Ensemble

Here, we use two approaches for subsampling of the C3S multi-model ensemble in real forecast test: random and teleconnection-based. For both approaches, the range of ensemble sizes was varied from 3 to 138 for a period of real forecast test from 2001 to 2014. In the first random statistical approach, 1,000 samples of members (combinations) for each given ensemble size were analyzed. In the second approach, we use a teleconnection-based subsampling technique, suggested by Dobrynin et al. (2018) and generalized by Düsterhus (2020), selecting only ensemble members with well-represented links between the autumn NAO predictors and the winter NAO index. This requires a statistical estimation of the first-guess NAO value, therefore it can be considered as a statistical-dynamical approach. We construct a first-guess DJF NAO index from the ERA-Interim de-trended time series of the area-weighted mean over regions with significant positive correlations between each autumn predictor and DJF NAO (Dobrynin et al., 2018). The significance of correlation is calculated using a bootstrapping approach (e.g., Efron & Tibshirani, 1994) with 500 samples at a given confidence level. We use training periods from 1994 until the year previous to forecasted year. Thereby, we calculate sets of four first-guess NAO values for subsampling of the C3S multi-model ensemble.

We perform a sensitivity analysis aiming at dependency of the NAO skill to ensemble size for both approaches. By using all four predictors in teleconnection-based approaches, it is likely that one ensemble member will be selected by more than one predictor. We use each selected ensemble member only once and remove duplicates from subsampled ensemble. This, in turn, means, that the number of selected members can be different from year to year, even with a fixed number of members selected by each predictor. Therefore, in the sensitivity tests of the NAO skill versus ensemble size, we use only one SST-predictor in the teleconnection-based approach, as a guaranty of the constant number of selected member for consistency between random and teleconnection-based approaches.

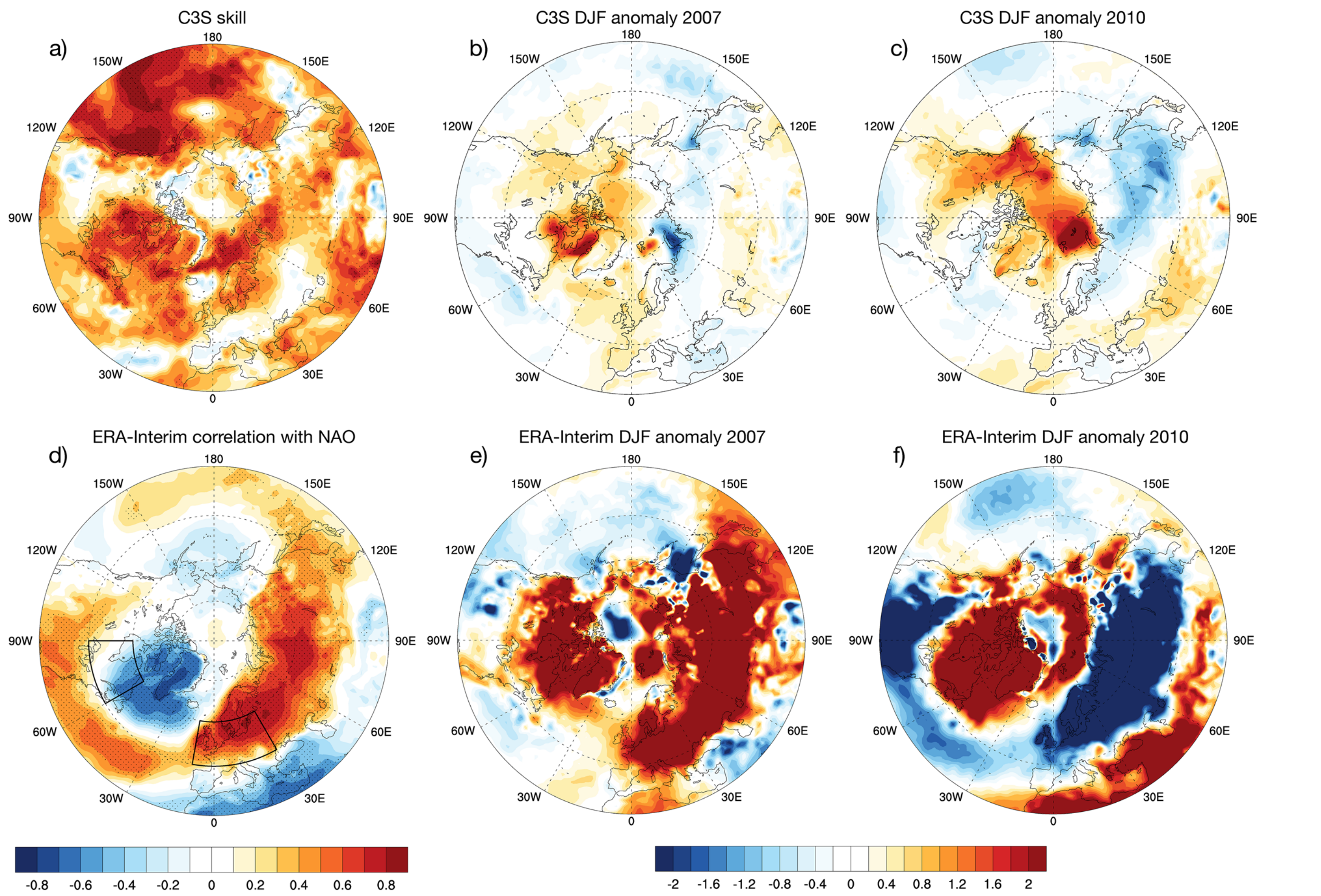


Figure 1. Prediction skill of the Copernicus Climate Change Service (C3S) ensemble and anomalies of wintertime temperature. (a) C3S ensemble prediction skill of 2-m temperature calculated for a period from 1994 to 2014 as compared to ERA-Interim; (b and c) December, January, and February (DJF) anomalies of 2-m temperature for a strong positive (2007) and negative (2010) North Atlantic Oscillation (NAO) phase as calculated from C3S ensemble; (d) correlation map between DJF 2-m temperature and NAO index in ERA-Interim, two regions of specific interest Central Europe (45N–60N, 10W–30 E) and Eastern Canada (45N–60N, 90W–60W) are shown; (e and f) same as (b and c) but from ERA-Interim. Regions that are significant at the 95% confidence level are indicated by dots on the maps in the left column.

The subsampling technique was also applied for individual C3S models. For this, the number of selected members per predictor was limited to 13, 8, 10, 5, and 9 members for CMCC, ECMWF, DWD, Meteo France, and UKMO system respectively.

2.5. Results Evaluation

Results of SPSs are evaluated over two periods: a hindcast from 1994 to 2014, and a forecast period from 2001 to 2014. Prediction skill for both periods is calculated as the correlation coefficient between the respective ensemble mean and the ERA-interim. For the hindcast period in multi-model ensemble the prediction skill for DJF NAO and T2m is calculated. For each model separately, only prediction skill for DJF NAO is calculated. For the forecast period, we mimic a real forecast calculating the NAO index and T2m anomalies individually for each year. Values of the NAO index and T2m for each particular year are then combined into time series. T2m anomalies for Northern Hemisphere and area-weighted regional mean anomalies for two regions shown in Figure 1d Central Europe (45N–60N, 10W–30E) and Eastern Canada (45N–60N, 90W–60W) are calculated by subtracting a mean value of T2m over a period from 1994 until 2014 or until each particular year in a real forecast test, depending on the end of the forecast period.

For comparison between statistical and statistical-dynamical subsampling methods, we calculated the NAO index as a mean value over four ERA-Interim predictors. We mimic a real statistical forecast for three periods: from 1985 to 2014, with a training period starting from 1979 and until the year previous to the forecasted year, from 1985 to 1999 starting from 1979, and from 2001 to 2014 starting from 1979. Additionally, we calculated the first-guess NAO index for the real statistical forecast test for 2001 to 2014 starting from 1994, which is directly comparable to a dynamical ensemble.

3. C3S Multi-Model Ensemble Prediction of Air Temperature

Prediction skill of the C3S ensemble for 2-m air temperature in the Northern Hemisphere demonstrates high skill in the North Pacific sector, moderate skill in the eastern part of North America and in the North Atlantic sector, and low skill in Europe (Figure 1a). The prediction skill for the winter NAO is represented by a correlation of 0.39 between the C3S ensemble mean (hereafter C3S-mean) and the ERA-Interim NAO index. The effect of change of winter NAO phase on temperature (hereafter temperature response) is well known and can be demonstrated by a correlation between the DJF temperature and NAO index. A dipole structure with a negative correlation in the North Atlantic sector and positive correlation over Eurasia (Figure 1d) highlights areas where cold and warm temperature anomalies can be formed depending on the NAO phase.

However, despite a moderate NAO prediction skill, it appears that for the strong positive and negative NAO states in 2007 and 2010 the ensemble mean temperature anomalies are similar in terms of weakly pronounced amplitude (Figures 1b and 1c) in regions where a strong effect on temperature is expected, for example, in Central Europe. This can be mainly attributed to two reasons. The first reason is the reduced ability of model systems to simulate temperature response dependent on NAO phase by individual ensemble members. The second reason is the fact that ensemble mean tends to suppress the variability, comparing to an individual ensemble member or the observation. The first reason might be overcome by further model development and initialization methods for prediction systems. The second reason highlights the need to increase the amplitude of ensemble mean temperature anomaly in response to NAO phases by reducing ensemble size and at the same time the noise, as suggested in this study.

Comparing to ERA-Interim (Figure 1d), the temperature response of the C3S ensemble (Figure S1f in Supporting Information S1) has a similar dipole structure combining all individual models (Figures S1a–S1e in Supporting Information S1). However, the negative correlation in the North Atlantic sector and positive correlation over Eurasia is underestimated. Simultaneously, a positive correlation over North America and the Pacific Ocean is overestimated. Overall, the well-pronounced temperature response in the C3S ensemble demonstrates a potential for forming temperature anomalies following changes of the NAO phase.

4. Skill and Variability Estimated From Subsampling Approaches

The C3S ensemble underestimates the inter-annual variability of the NAO index, calculated as a standard deviation (hereafter STD) of the ensemble mean (0.22). The NAO STD tends to decrease with an increase of the ensemble size (Figure 2a, gray dash line). Therefore, the full range of variability will not be covered even by the large multi-model ensemble C3S. On the contrary, individual members from each SPSs reproduce very well the full range of the ERA-Interim NAO index (Figure 2b). Thus, possible improvement in the variability and prediction skill of the NAO index and wintertime temperature can be achieved by ensemble subsampling, that is, considering only a part of the entire ensemble. We analyze the prediction skill and variability of the NAO and temperature depending on ensemble subsampling size for both random and teleconnection-based subsampling approaches, in the real forecast test from 2001 to 2014.

4.1. Random Versus Teleconnection-Based Subsampling Approach

Random and teleconnection-based subsampling approaches have two different goals. While the random approach provides an estimation of a possible change of the prediction skill and variability arising from increasing of ensemble size only, the teleconnection-based approach demonstrates an added value of including of initial conditions analysis into ensemble subsampling. For both regions of specific interest Central Europe and Eastern

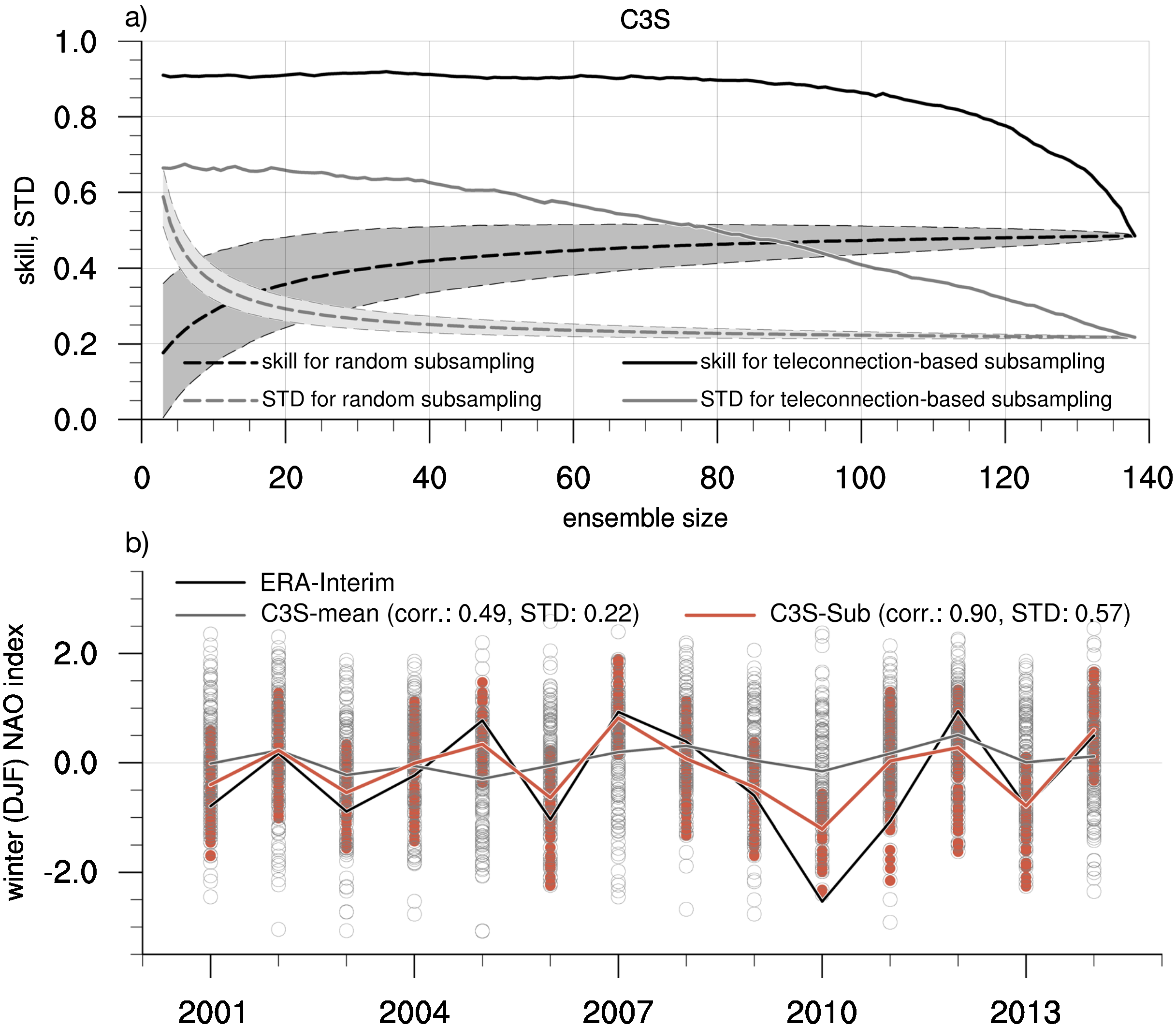


Figure 2. Prediction skill, variability and subsampling of the multi-model ensemble Copernicus Climate Change Service (C3S) for the North Atlantic Oscillation (NAO) index in a real forecast test from 2001 to 2014 (a) prediction skill (black lines) (b) and variability denoted as standard deviation (standard deviation [STD], gray lines) calculated for the C3S ensemble using two approaches: random selection of ensemble members (dashed lines) and NAO teleconnection-based subsampling (as in Dobrynin et al. (2018), but with SST predictor only, solid lines); (b) subsampling of the C3S ensemble for the winter NAO using all predictors (orange line) comparing to the C3S ensemble means (gray lines) and the ERA-Interim (black lines). The shaded area represents the sampling error for the random selection approach. Open circles denote each C3S ensemble member, filled circles indicate 46 subsampled due to NAO teleconnection-based approach ensemble members.

Canada (shown in Figure 1d), we analyze the time series of the DJF NAO and wintertime averaged 2-m air temperature, mimicking the real forecast for a period from 2001 to 2014.

The prediction skill of the winter NAO of the full 138-member C3S ensemble in a random subsampling approach follows a logarithmic-like behavior with a rapid increase of prediction skill from about 0.20 for 3-member ensemble to 0.40 for about one-third of the ensemble size (Figure 2a, black dash line). Afterward, the added value of the remaining ensemble members is limited to 0.09. This results in a skill of 0.49 for the full C3S ensemble for a period from 2001 to 2014. In contrast, the teleconnection-based subsampling approach demonstrates a stable high level of prediction skill of about 0.90 starting from a 3-member ensemble to an about 70-member ensemble (Figure 2a, black solid line). Afterward, the skill is decreasing down to the C3S ensemble mean value of 0.49. The range of the NAO prediction skill varies from 0.17 to 0.48 when all systems are individually considered (Figure S5 in Supporting Information S1). The subsampling improves NAO prediction skill for each individual system to a range from 0.85 to 0.90 even for the systems with initially low NAO skill (Figures S5 and S6 in Supporting Information S1).

Variability of the winter NAO index, denoted as the STD of the ensemble mean, in both approaches decreases with an increase of the ensemble size. However, while in random subsampling approach STD decreases by half within 20 ensemble members from 0.6 to 0.3 (Figure 2a, gray dash line), the teleconnection-based subsampling provides a stable high, more than 0.6, level of STD for 50 ensemble members (Figure 2a, gray solid line).

For wintertime averaged 2-m air temperature, the random subsampling approach demonstrates an increase of prediction skill as a function of ensemble size, similar to the winter NAO (Figure 3a, dash lines). Notable, the rapid growth of skill is also limited to about one-third of the ensemble size for both regions, but it results in a different ensemble mean prediction skill of 0.25 for Central Europe and 0.69 for Eastern Canada. The teleconnection-based subsampling for the air temperature uses the same members as selected for the winter NAO, therefore a clear difference appears between the prediction skill for Central Europe and Eastern Canada as for a region of strong and weak NAO impact respectively. For Eastern Canada, the high level of prediction skill of about 0.7 can be achieved already by a small ensemble size. This skill demonstrates low sensitivity to variations of prediction skill of the winter NAO and stays on the same level as for the full C3S ensemble mean (Figure 3a, blue solid line).

In contrast, for air temperature over Central Europe, the prediction skill tends to follow a decrease of the NAO prediction skill starting from about two-thirds of the ensemble size (Figure 3a, red solid line).

Note that, high skill is not the only criterion for selecting the optimal number of subsampled members. The more important reason is that a well-filled probability density function (PDF) of the subsampled ensemble is required for stabilization of prediction skill and ensemble mean statistics. A smoother ensemble PDF such as created by randomly increasing the ensemble members, is not necessarily beneficial for increasing skill. In contrast to this, subsampling does not act as a random manipulation of the PDF, but more like a filter, for example, as a band-pass filter. As such, it does not necessarily decrease the overall smoothness of the PDF, but selects the relevant sections of the PDF. Therefore, high 2-m temperature skill for 3 members and high skill variability up to about 30 members (Figure 3a, red solid line) can be explained by stabilization of skill by filling ensemble PDF with more members toward an optimal number of subsampled members. This has been further demonstrated and investigated by Düsterhus (2020), who showed that a filtered model PDF, within a system similar to the subsampling method applied in this manuscript, can be beneficial for higher prediction skill.

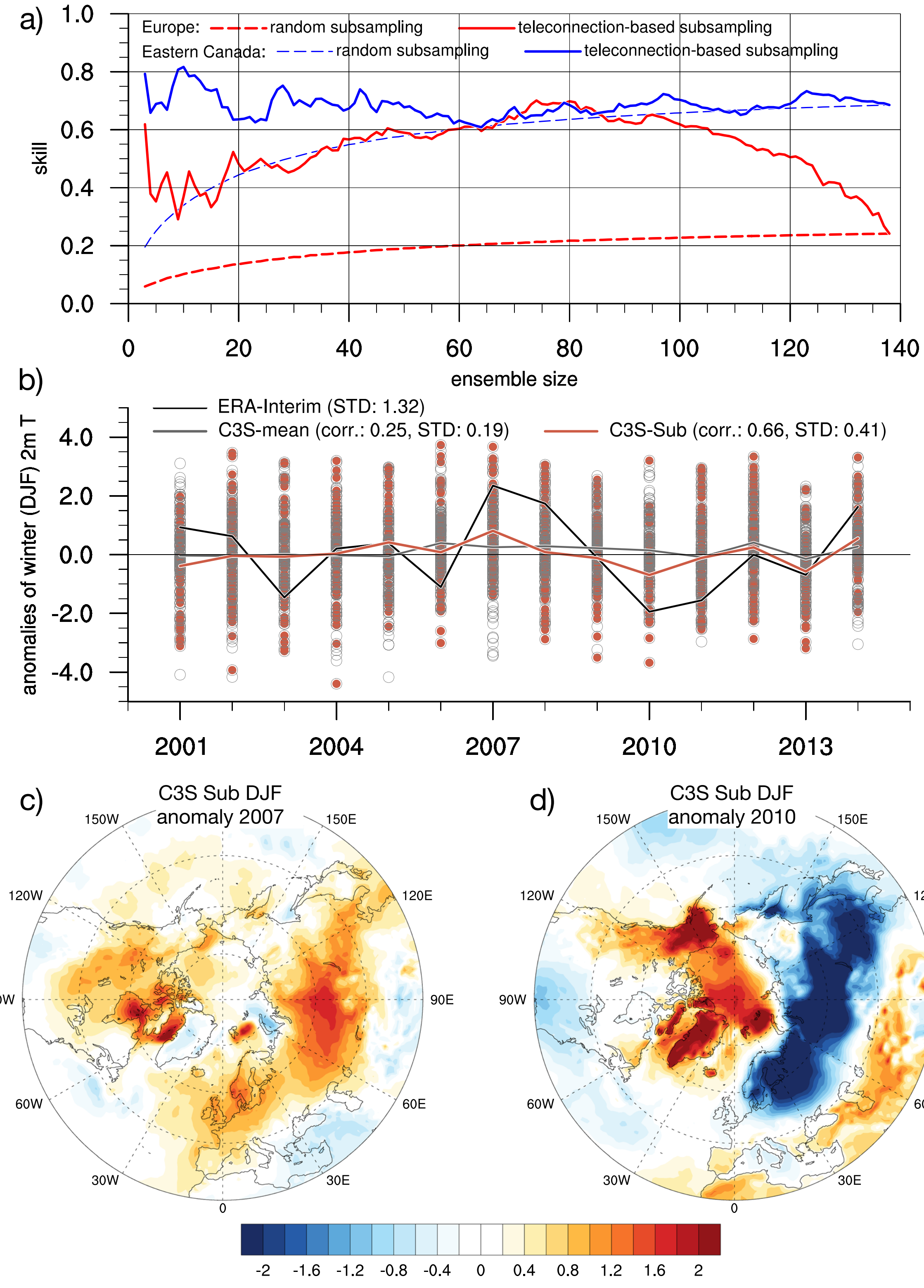
4.2. Implementation of Teleconection-Based Subsampling Approach for Predicting of Air Temperature in Central Europe

We focus now on Central Europe and analyze the prediction skill of regionally averaged air temperature anomalies in a real forecast test using the teleconnection-based subsampling approach for a period from 2001 to 2014. The number of selected ensemble members is limited to one-third of the C3S ensemble size, which is 46 members.

The subsampled 46-member C3S ensemble shows a significant increase both in NAO prediction skill from 0.49 to 0.90 and in the variability (STD) of the ensemble mean NAO index from 0.22 to 0.57 (Figure 2b). Following the increase of the NAO skill and variability, the air temperature skill is increased from 0.25 to a significant value of 0.66 (Figure 3b). The variability (STD) of the air temperature is also improved from 0.19 to 0.41. Corrections of the NAO phases due to subsampling are most notable for years with relatively strong NAO phase, such as for example, in 2005–2007 and 2010 (Figure 2b). In a more general context, the teleconnection-based subsampling approach significantly improves the C3S ensemble prediction skill of the SLP and air temperature over an essential part of the Northern Hemisphere (Figure S2 in Supporting Information S1). For the air temperature, the areas with mostly improved prediction skill (up to 0.8) are located in Eurasia (Figure S2 in Supporting Information S1). Over these areas, a better representation of the wintertime temperature anomalies related to NAO phases can be expected.

4.3. Statistical Versus Statistical-Dynamical Prediction

For comparison to the dynamical subsampled C3S ensemble, we calculate statistical first-guess NAO prediction from all four NAO predictors based on the ERA-Interim only (Figures S3 and S4 in Supporting Information S1). It appears that the length of the training period (TP, i.e., number of years before forecast year) affects the NAO prediction skill. For example, for a short TP of 6–20 years starting from 1979 and for a following forecast period



a)
Europe: random subsampling teleconnection-based subsampling
Eastern Canada: random subsampling teleconnection-based subsampling
skill
ensemble size
b)
ERA-Interim (STD: 1.32)
C3S-mean (corr.: 0.25, STD: 0.19)
C3S-Sub (corr.: 0.66, STD: 0.41)
anomalies of winter (DJF) 2m T
2001
2004
2007
2010
2013
c)
C3S Sub DJF
anomaly 2007
d)
C3S Sub DJF
anomaly 2010

Figure 3.

from 1985 to 1999 (Figure S4b in Supporting Information S1), the NAO skill is 0.91, while for the full forecast period from 1985 to 2014 with a TP of 6–35 years the value drops to 0.86 (Figure S4a in Supporting Information S1). For a short forecast period from 2001 to 2014 with a long TP of 22–35 years starting from 1979, the NAO prediction skill is 0.82 (Figure S4c in Supporting Information S1). With a short TP of 7–20 years starting from 1994 (not shown here), the NAO skill is 0.92 – higher than from dynamical subsampled C3S ensemble for the same period. This can be partly attributed to equal consideration of all systems within the C3S ensemble in one multi-model ensemble, independently of performance in predicting winter NAO for each individual model. Analyzing C3S's models individually, it appears that the subsampling has a different level of improvement of the winter NAO prediction skill for less and more skillful models (Figure S5 in Supporting Information S1). We notice that the final skill is partly determined by the initial model skill and that a higher prediction skill can be achieved for a more skillful system and such high skill cannot be achieved for a less skillful system due to subsampling (Figures S5 and S6 in Supporting Information S1). Most likely a combination of, for example, more skillful or systems with similar ensemble size, will have an effect on the NAO prediction skill of dynamical subsampled C3S ensemble (not shown here).

4.4. Improved Prediction of Wintertime Temperature Anomalies

Finally, we calculated wintertime temperature anomalies for two selected years: 2007 with a strong positive NAO phase, and 2010 with a strong negative phase from the subsampled C3S ensemble. As opposite to the C3S ensemble mean (Figures 1b and 1c), the C3S subsampled mean predicts the temperature anomalies with a clear characteristic structure for a positive NAO phase in 2007 and negative NAO phase in 2010 (Figures 3c and 3d). Note, that the area affected by better prediction of the NAO covers not only the North Atlantic sector, but also an essential part of Eurasia. Predicted temperature anomalies have a similar structure as compared to the ERA-Interim anomalies (Figures 1e and 1f). However, the exact prediction of the values of temperature anomaly at local scales remains challenging.

5. Conclusions

In summary, we found that the existing C3S operational prediction systems, being combined in a multi-model subsampled ensemble, can skilfully predict winter temperature anomalies in Central Europe and over an essential part of the Northern Hemisphere for a season ahead. Moreover, the C3S subsampled ensemble can provide a very high NAO prediction skill of 0.90. This leads us to the conclusion that the existing operational prediction systems do not fully use the potential coming from the large numbers of ensemble members in the prediction of wintertime temperature. Following a traditional ensemble mean approach, all C3S systems suppress the variability of predicted winter NAO index and temperature. From our analysis, we conclude that even a substantial increase of the ensemble size will not automatically improve the prediction skill and especially the variability of the NAO and temperature. Instead, the implementation of the NAO teleconnection-based subsampling approach to existing ensembles improves significantly the prediction skill and variability of the winter NAO index and temperature in the Northern Hemisphere. Notable, and of high importance for applications of subsampling on a hemisphere scale, is the conclusion that the improvement of temperature anomalies is focused on the regions of strong NAO impact, such as Central Europe. Simultaneously, regions of weak NAO impact, such as Eastern Canada, are not affected by subsampling. Moreover, our subsampling approach, being developed for the improvement of seasonal prediction of existing prediction systems, highlights also a need for further model development, rethinking ensemble generation, and initialization methods toward better representation of the NAO in individual ensemble members. This can be beneficial for an improved representation of ensemble mean variability of NAO and related parameters, keeping a realistic ensemble size.

Figure 3. Prediction skill and subsampling of Copernicus Climate Change Service (C3S) ensemble for the wintertime temperature in a real forecast test from 2001 to 2014. (a) prediction skill calculated for the C3S ensemble for two regional means in Central Europe (red) and in the Eastern Canada (blue) using two approaches: random selection of ensemble members (dashed lines) and North Atlantic Oscillation (NAO) teleconnection-based subsampling (as in Dobrynin et al. (2018), but with SST predictor only, solid lines); (b) subsampling of the C3S ensemble in Central Europe using all predictors (orange line) comparing to the C3S ensemble means (gray lines) and the ERA-Interim (black lines). Open circles denote each C3S ensemble member, filled circles indicate 46 subsampled due to NAO teleconnection-based approach ensemble members; (c–d) DJF anomalies of 2-m temperature for a strong positive (2007) and negative (2010) NAO phase as calculated from subsampled C3S ensemble.

Data Availability Statement

Seasonal forecasts, used in this study, provided by the Deutsche Wetterdienst, UK Met Office, European Centre for Medium-Range Weather Forecasts, Meteo France, and Centro Euro-Mediterraneo sui Cambiamenti Climatici for the period from 1994 to 2014 are available from Copernicus Climate Change Service (https://cds.climate.copernicus.eu/cdsapp#!/dataset/seasonal-original-single-levels?tab=overview). ERA-Interim data are available from ECMWF's at www.ecmwf.int/en/forecasts/datasets.

Acknowledgments
This work was funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) under Germany's Excellence Strategy—EXC 2037 "CLICCS—Climate, Climatic Change, and Society"— Project Number: 390683824, contribution to the Center for Earth System Research and Sustainability (CEN) of Universität Hamburg. A.D. is also supported by A4 (Aigéin, Aeráid, agus athrú Atlantaigh), funded by the Marine Institute (grant: PBA/CC/18/01). Work of K.F. is supported by the Copernicus C3S 433 DWD lot2 agreement. P.R. was supported by the Blue-Action project (European Union's Horizon 2020 research and innovation programme, grant: 727852). Open access funding enabled and organized by Projekt DEAL.

References

Årthun, M., Eldevik, T., Viste, E., Drange, H., Furevik, T., Johnson, H. L., & Keenlyside, N. S. (2017). Skillful prediction of northern climate provided by the ocean. *Nature Communications*, *8*(1), ncomms15875. https://doi.org/10.1038/ncomms15875

Athanasiadis, P. J., Bellucci, A., Scaife, A. A., Hermanson, L., Materia, S., Sanna, A., et al. (2017). A multisystem view of wintertime NAO seasonal predictions. *Journal of Climate*, *30*(4), 1461–1475. https://doi.org/10.1175/jcli-d-16-0153.1

Barnston, A. G., & Livezey, R. E. (1987). Classification, seasonality and persistence of low-frequency atmospheric circulation patterns. *Monthly Weather Review*, *115*(6), 1083–1126. https://doi.org/10.1175/1520-0493(1987)115<1083:csapol>2.0.co;2

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

Butler, A. H., & Polvani, L. M. (2011). El Nino, La Nina, and stratospheric sudden warmings: A re-evaluation in light of the observational record. *Geophysical Research Letters*, *38*(13), GL048084. https://doi.org/10.1029/2011gl048084

Cassou, C., Terray, L., Hurrell, J. W., & Deser, C. (2004). North Atlantic winter climate regimes: Spatial asymmetry, stationarity with time, and oceanic forcing. *Journal of Climate*, *17*(5), 1055–1068. https://doi.org/10.1175/1520-0442(2004)017<1055:nawcrs>2.0.co;2

Cattiaux, J., Vautard, R., Cassou, C., Yiou, P., Masson-Delmotte, V., & Codron, F. (2010). Winter 2010 in Europe: A cold extreme in a warming climate. *Geophysical Research Letters*, *37*(20). https://doi.org/10.1029/2010gl044613

Cohen, J., & Jones, J. (2011). A new index for more accurate winter predictions. *Geophysical Research Letters*, *38*(21). https://doi.org/10.1029/2011GL049626/full

Czaja, A., & Frankignoul, C. (2002). Observed impact of Atlantic SST anomalies on the North Atlantic Oscillation. *Journal of Climate*, *15*(6), 606–623. https://doi.org/10.1175/1520-0442(2002)015<0606:oioasa>2.0.co;2

Dalelane, C., Dobrynin, M., & Fröhlich, K. (2020). Seasonal forecasts of winter temperature improved by higher-order modes of mean sea level pressure variability in the North Atlantic sector. *Geophysical Research Letters*, *47*(16), e2020GL088717. https://doi.org/10.1029/2020GL088717

Dee, D., Uppala, S., Simmons, A., 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

Dobrynin, M., Domeisen, D. I., Müller, W. A., Bell, L., Brune, S., Bunzel, F., et al. (2016). Improved seasonal prediction of winter NAO through ensemble sub-sampling. *EGU General Assembly Conference Abstracts*, *18*, 4319.

Dobrynin, M., Domeisen, D. I., Müller, W. A., Bell, L., Brune, S., Bunzel, F., et al. (2018). Improved teleconnection-based dynamical seasonal predictions of boreal winter. *Geophysical Research Letters*, *45*(8), 3605–3614. https://doi.org/10.1002/2018GL077209

Domeisen, D. I. V., Butler, A. H., Fröhlich, K., Bittner, M., Müller, W. A., & Baehr, J. (2015). Seasonal predictability over Europe arising from El Nino and stratospheric variability in the MPI-ESM seasonal prediction system. *Journal of Climate*, *1*(28), 256–271. https://doi.org/10.1175/jcli-d-14-00207.1

Düsterhus, A. (2020). Seasonal statistical–dynamical prediction of the North Atlantic Oscillation by probabilistic post-processing and its evaluation. *Nonlinear Processes in Geophysics*, *27*(1), 121–131. https://doi.org/10.5194/npg-27-121-2020

Efron, B., & Tibshirani, R. J. (1994). *An introduction to the bootstrap*. CRC Press.

Heape, R., Hirschi, J., & Sinha, B. (2013). Asymmetric response of European pressure and temperature anomalies to NAO positive and NAO negative winters. *Weather*, *68*(3), 73–80. https://doi.org/10.1002/wea.2068

Hertig, E., & Jacobeit, J. (2014). Variability of weather regimes in the North Atlantic-European area: Past and future. *Atmospheric Science Letters*, *15*(4), 314–320. https://doi.org/10.1002/asl2.505

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., Kushnir, Y., Ottersen, G., & Visbeck, M. (2003). *The North Atlantic Oscillation: Climate significance and environmental impact* (Vol. 134)). American Geophysical Union.

Jung, T., Vitart, F., Ferranti, L., & Morcrette, J.-J. (2011a). Origin and predictability of the extreme negative NAO winter of 2009/10. *Geophysical Research Letters*, *38*(7). https://doi.org/10.1029/2011gl046786

Jung, T., Vitart, F., Ferranti, L., & Morcrette, J.-J. (2011b). Origin and predictability of the extreme negative NAO winter of 2009/10. *Geophysical Research Letters*, *38*(7). https://doi.org/10.1029/2011GL046786

Kolstad, E. W., & Screen, J. A. (2019). Nonstationary relationship between autumn Arctic sea ice and the winter North Atlantic Oscillation. *Geophysical Research Letters*, *46*(13), 7583–7591. https://doi.org/10.1029/2019GL083059

Kutzbach, J. E. (1970). Large-scale features of monthly mean Northern Hemisphere anomaly maps of sea-level pressure. *Monthly Weather Review*, *98*(9), 708–716. https://doi.org/10.1175/1520-0493(1970)098<0708:lsfomm>2.3.co;2

Maidens, A., Arribas, A., Scaife, A. A., MacLachlan, C., Peterson, D., & Knight, J. (2013). The influence of surface forcings on prediction of the North Atlantic Oscillation regime of winter 2010/11. *Monthly Weather Review*, *141*(11), 3801–3813. https://doi.org/10.1175/mwr-d-13-00033.1

Mayer, B., Düsterhus, A., & Baehr, J. (2021). When does the Lorenz 1963 model exhibit the signal-to-noise paradox? *Geophysical Research Letters*, *48*(4), e2020GL089283. https://doi.org/10.1029/2020GL089283

O'Reilly, C. H., Heatley, J., MacLeod, D., Weisheimer, A., Palmer, T. N., Schaller, N., & Woollings, T. (2017). Variability in seasonal forecast skill of Northern Hemisphere winters over the 20th century. *Geophysical Research Letters*, *44*(11), 5729–5738. https://doi.org/10.1002/2017gl073736

Peings, Y., Brun, E., Mauvais, V., & Douville, H. (2013). How stationary is the relationship between Siberian snow and Arctic Oscillation over the 20th century? *Geophysical Research Letters*, *40*(1), 183–188. https://doi.org/10.1029/2012gl054083

10.1029/2021GL095063

Rodwell, M., Rowell, D., & Folland, C. (1999). Oceanic forcing of the wintertime North Atlantic Oscillation and European climate. *Nature*, *398*(6725), 320–323. https://doi.org/10.1038/18648

Scaife, A. A., & Smith, D. (2018). A signal-to-noise paradox in climate science. *npj Climate and Atmospheric Science*, *1*(1), 28. https://doi.org/10.1038/s41612-018-0038-4

Strong, C., Magnusdottir, G., & Stern, H. (2009). Observed feedback between winter sea ice and the North Atlantic Oscillation. *Journal of Climate*, *22*(22), 6021–6032. https://doi.org/10.1175/2009jcli3100.1

Sun, L., Deser, C., & Tomas, R. A. (2015). Mechanisms of stratospheric and tropospheric circulation response to projected Arctic sea ice Loss. *Journal of Climate*, *28*(19), 7824–7845. https://doi.org/10.1175/jcli-d-15-0169.1

Thompson, D., Lee, S., & Baldwin, M. (2003). Atmospheric processes governing the northern hemisphere annular mode/North Atlantic Oscillation. *Geophysical Monograph*.

Vautard, R. (1990). Multiple weather regimes over the North Atlantic: Analysis of precursors and successors. *Monthly Weather Review*, *118*(10), 2056–2081. https://doi.org/10.1175/1520-0493(1990)118<2056:mwrotn>2.0.co;2

Wang, L., Ting, M., & Kushner, P. (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., Schaller, N., O'Reilly, C., MacLeod, D. A., & Palmer, T. (2017). Atmospheric seasonal forecasts of the twentieth century: Multi-decadal variability in predictive skill of the winter North Atlantic Oscillation (NAO) and their potential value for extreme event attribution. *Quarterly Journal of the Royal Meteorological Society*, *143*(703), 917–926. https://doi.org/10.1002/qj.2976