Hidden potential in predicting wintertime temperature anomalies in the Northern Hemisphere

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

[1]Deutscher Wetterdienst
[2]ICARUS
[3]CMCC
[4]University of Bologna
[5]Max Planck Institute for Meteorology
[6]Universität Hamburg, Center for Earth System Research and Sustainability

November 22, 2022

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 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.

Hidden potential in predicting wintertime temperature anomalies in the Northern Hemisphere

Mikhail Dobrynin[1,2], André Düsterhus[3], Kristina Fröhlich[1], Panos Athanasiadis[5], Paolo Ruggieri[5,6], Wolfgang A. Müller[7], Johanna Baehr[1]

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

Key Points:

- Temperature anomalies can be skilfully predicted for the upcoming winter through increased variability and skill of predicted NAO
- Skilful prediction of temperature anomalies in the Northern Hemisphere for upcoming winter

Corresponding author: Mikhail Dobrynin, `mikhail.dobrynin@dwd.de`

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 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

Accurate prediction of wintertime temperature anomalies in the Northern Hemisphere is closely connected to the ability of a dynamical prediction system to predict the North Atlantic Oscillation (NAO). While ensemble-based dynamical seasonal prediction systems have been shown to skilfully predict the winter NAO, the prediction for the NAO-dependent anomalies of the air temperature remains elusive. One of the main reasons is that the high correlation prediction skill, commonly used as a measure of prediction quality for the NAO, represents only a part of real NAO behavior, namely a good timing of the NAO phases. However, as we show in this study, the strength of the predicted NAO phase is the most important characteristic for the accurate prediction of wintertime temperature anomalies. Here, we demonstrate a hidden potential of existing operational seasonal prediction systems in predicting wintertime temperature by increasing the strength of the predicted NAO phase. We use a novel multi-model subsampling approach for the identification and subsampling of ensemble members, which are close to NAO index estimated from analysis of initial autumn conditions. We show that temperature anomalies can be skilfully predicted for the upcoming winter over a large part of the Northern Hemisphere.

1 Introduction

In the Northern Hemisphere, the development of wintertime temperature anomalies is governed mainly by large-scale weather regimes in the North Atlantic sector (Vautard, 1990; Hertig & Jacobeit, 2014). While ocean and atmosphere act on different time scales, they are both important for the formation of specific winter conditions (Rodwell et al., 1999; Cassou et al., 2004). 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 (Scaife et al., 2008; Jung et al., 2011a; Maidens et al., 2013) such as the anomalies of wintertime air temperature.

While ensemble-based dynamical seasonal prediction systems (hereafter SPSs) are known to skilfully predict the winter NAO index for a season ahead (Scaife et al., 2014; O'Reilly et al., 2017; Athanasiadis 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 (Jung et al., 2011b; Årthun et al., 2017; Scaife et al., 2017). 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 the predicted NAO index. This can be partly attributed to well-known underestimation of the signal-to-noise ratio in prediction systems (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 skilful predicting the wintertime temperature anomalies in the Northern Hemisphere by increasing the variability of predicted NAO using a multi-model subsampling approach. Instead of following the traditional practice of averaging all ensemble members, we make use of the intrinsic memory of the Earth system, analysing 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 enforces 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, 30 members), UK Met Office (UKMO, 28 members), European Centre for Medium-Range Weather Forecasts (ECMWF, 25 members), Meteo France (15 members), and Centro Euro-Mediterraneo sui Cambiamenti Climatici (CMCC, 40 members). All members are combined in one ensemble of 138 members without implementation of a bias correction procedure.

We use monthly mean data of sea level pressure (SLP) and 2-meter 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 et al., 2011). While averaged over December, January and February (DJF) monthly mean SLP and T2m data are used for the evaluation of model results, October T100, SST and SNC, and September SIC represent the autumn predictors of the winter NAO index. Originally, autumn predictors were provided by an assimilation simulation used for hindcast initialisation. 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 winter season DJF NAO index for ensemble subsampling as adopted from Dobrynin et al. (2018).

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 prior to the EOF analysis. The region of SLP data is limited to the latitude range 20°N to 90°N and to the longitude range 90°W to 60°E. The EOF is calculated in every system from a vector, where all ensemble members are merged over the entire time period. This approach of EOF calculation allows us to represent the entire ensemble in one 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 normalised by their respective standard deviations. The ERA-Interim NAO index is used as a reference for comparisons with other systems.

2.3 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, we use the range of ensemble sizes from 3 to 138 for a period of real forecast test from 2001 to 2014. In the first random statistical approach, we use 1000 samples (combinations) for each given ensemble size and then average them. In the second approach, we use a teleconnection-based subsampling technique (Dobrynin et al., 2018) 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 statistical-dynamical approach. We construct a first-guess DJF NAO index from the de-trended time series of area-weighted mean over regions with significant positive correlations between each autumn predictor and DJF NAO (Dobrynin et al., 2018). 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. For reasons of consistency, keeping the number of selected member constant for each year, only one, the SST - predictor of the NAO, is used here for the analysis of skill and variability depending on the ensemble size. In contrast, for final analysis of prediction skill and variability of the NAO index and T2m anomalies, all four predictors are used for subsampling. 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.

2.4 Results evaluation

Results of SPSs are evaluated over two periods. First, for each model separately and for multi-model ensemble the DJF NAO prediction skill is calculated for the full period of hindcast from 1994 to 2014 as the correlation coefficient between the ensemble mean and ERA-interim. T2m anomaly correlation coefficient (ACC) is calculated for the multi-model ensemble mean for the same period. Second, we mimic a real forecast test for a period from 2001 to 2014 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 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 four 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. Also, 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-meter air temperature in the Northern Hemisphere demonstrates high skill in the North Pacific sector, less skill in the eastern part of North America and in the North Atlantic sector, and low skill in Europe (Fig. 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 (Fig. 1d) highlights areas where cold and warm temperature anomalies can be formed depending on the NAO phase.

However, despite a moderate NAO prediction skill, comparing the C3S ensemble mean predicted anomalies of temperature, it appears that for the strong positive and negative NAO states in 2007 and 2010, the temperature anomalies are similar in terms of weakly pronounced amplitude (Fig. 1b and c) in regions where a strong effect on temperature is expected. Comparing to ERA-Interim (Fig. 1d), the temperature response of the C3S ensemble (Fig. S1f) has a similar dipole structure combining all individual models (Fig. S1a–e). 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. Thereby, 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 comparing to 1.00 for ERA-Interim NAO). The NAO STD tends to decrease with an increase of the ensemble size (Fig. 2a, grey 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 (Fig. 2b). Thus, possible improvement in the variability and prediction skill of the NAO index and wintertime temperature can be achieved by ensemble subsampling, i.e.

considering only a part of the entire ensemble. We analyse 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 teleconection-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. We select two regions for the air temperature analysis, Central Europe (45N–60N, 10W–30E), known as a region of strong NAO impact, and Eastern Canada (45N–60N, 90W–60W) as a region with a weaker NAO impact (Fig. 1d). For both regions, we analyse the time series of the DJF NAO and wintertime averaged 2-meter 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 behaviour 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 (Fig. 2a, black dash line). Afterwards, 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 (Fig. 2a, black solid line). Afterwards, the skill is decreasing down to the C3S ensemble mean value of 0.49.

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 factor of 2 within 20 ensemble members from 0.6 to 0.3 (Fig. 2a, grey dash line), the teleconnection-based subsampling provides a stable high, more than 0.6, level of STD for 50 ensemble members (Fig. 2a, grey solid line).

For wintertime averaged 2-meter air temperature, the random subsampling approach demonstrates an increase of prediction skill as a function of ensemble size, similar to the winter NAO (Fig. 3a, dash lines). Notably, 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 small ensemble size and the skill is not affected by the changing of the prediction skill of the winter NAO staying on the same level as for the full C3S ensemble mean (Fig. 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 (Fig. 3a, red solid line).

4.2 Teleconection-based subsampling approach for predicting of air temperature in Central Europe

We analyse now the prediction skill for the winter NAO and air temperature anomalies in Central Europe in a real forecast test using the teleconnection-based subsampling approach (Dobrynin et al., 2018) for a period from 2001 to 2014 (see Methods). We limit the number of selected ensemble members to one-third of the C3S ensemble size, which is 46 members. The subsampled 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 (Fig. 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 (Fig. 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 strong NAO phase such as for example in 2005-2007 and 2010. In a more general context, the teleconnection-based subsampling approach significantly improves the C3S ensemble prediction skill of the sea level pressure and air temperature over an essential part of the Northern Hemisphere (Fig. S2). For the air temperature, the areas with mostly improved prediction skill (up to 0.8) are located in Eurasia (Fig. S2). 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 (Fig. S3). 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 to 20 years starting from 1979 and for a following forecast period from 1985 to 1999, the NAO skill is 0.91, while for the full forecast period from 1985 to 2014 with a TP of 6 to 35 years the value drops to 0.86 (Fig. S3). For a short forecast period from 2001 to 2014 with a long TP of 22 to 35 years starting from 1979, the NAO prediction skill is 0.82 (Fig. S3). With a short TP of 7 to 20 years starting from 1994, the NAO skill is 0.92 – higher as 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 this study, we consider the C3S models as one multi-model ensemble. Considering C3S models individually, it appears that the subsampling has a different level of improvement of the winter NAO prediction skill for less and more skilful models (Fig. S4). For example, in the real forecast test from 2001 to 2014, for the DWD system this improvement is from 0.48 to 0.90 and for the ECMWF system from 0.17 to 0.85 before and after subsampling respectively. Part of the difference in improvement can be explained due to the fact that improvement for correlations is harder to gain the higher the actual correlation values are. However, we note that most likely a higher prediction skill can be achieved for a more skilful system and such high skill cannot be achieved for a less skilful system due to subsampling (Fig. S4). Most likely a combination of, for example, more skilful 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 (Fig. 1b and c), 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 (Fig. 3c and d). 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 (Fig. 1e and f). 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. Moreover, our subsampling approach, being developed for the improvement of seasonal prediction of existing prediction systems, highlights also a need for a rethinking of ensemble generation methods in general, for better NAO prediction from each ensemble member keeping a realistic ensemble size. A reduction of noise introduced by a large number of ensemble members is necessary to increase the variability of predicted NAO and avoid the decoupling of NAO from the formation of wintertime temperature anomalies.

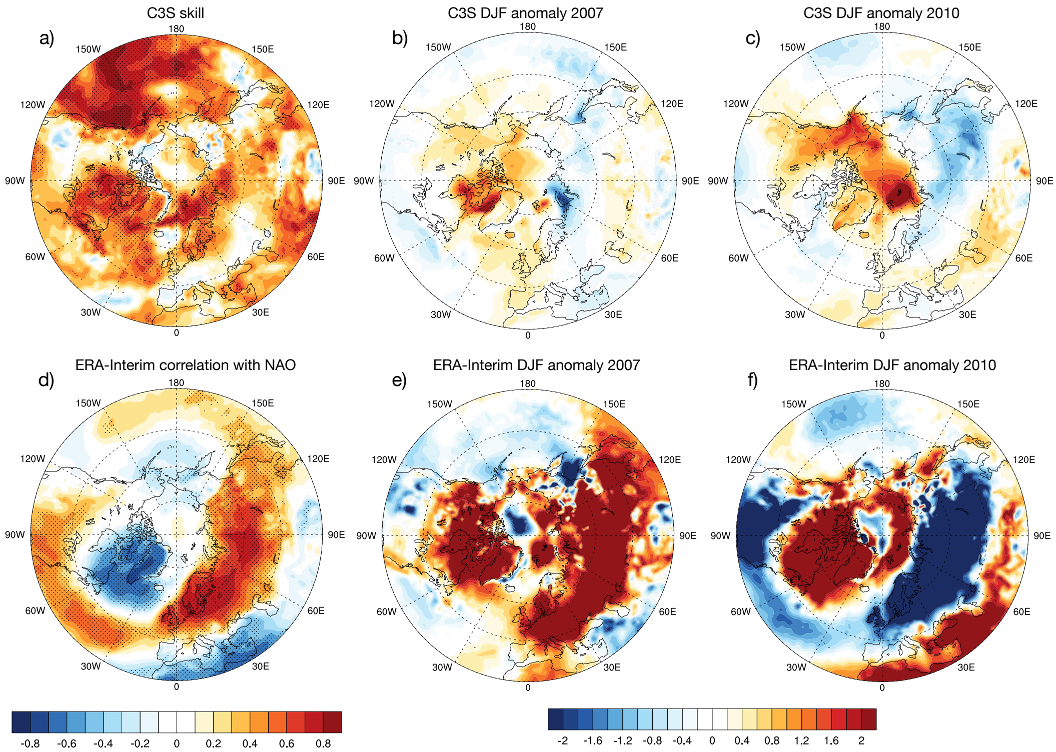


Figure 1. Prediction skill of the C3S ensemble and anomalies of wintertime temperature. a) C3S ensemble prediction skill of 2-meter temperature calculated for a period from 1994 to 2014 as compared to ERA-Interim; b) and c) DJF anomalies of 2-meter temperature for a strong positive (2007) and negative (2010) NAO phase as calculated from C3S ensemble; d) correlation map between DJF 2-meter temperature and NAO index in ERA-Interim; 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.

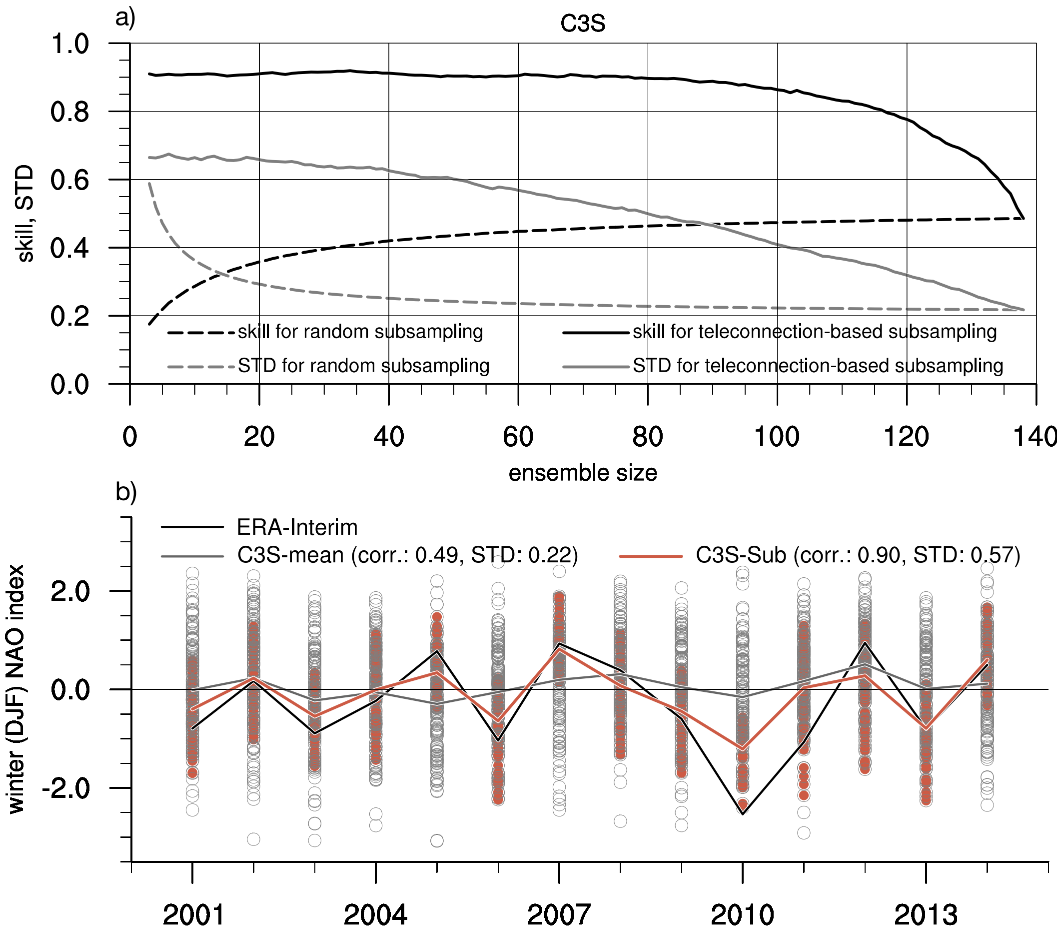


Figure 2. Prediction skill, variability and subsampling of the multi-model ensemble C3S for the NAO index in a real forecast test from 2001 to 2014 a) prediction skill (black lines) and variability denoted as standard deviation (STD, grey lines) calculated for the C3S ensemble using two approaches: random selection of ensemble members (dashed lines) and NAO teleconnection-based subsampling (Dobrynin et al., 2018) (solid lines); b) subsampling of the C3S ensemble for the winter NAO (orange line) comparing to the C3S ensemble means (grey lines) and the the ERA-Interim (black lines). Open circles denote each C3S ensemble member, filled circles indicate subsampled due to NAO teleconnection-based approach ensemble members.

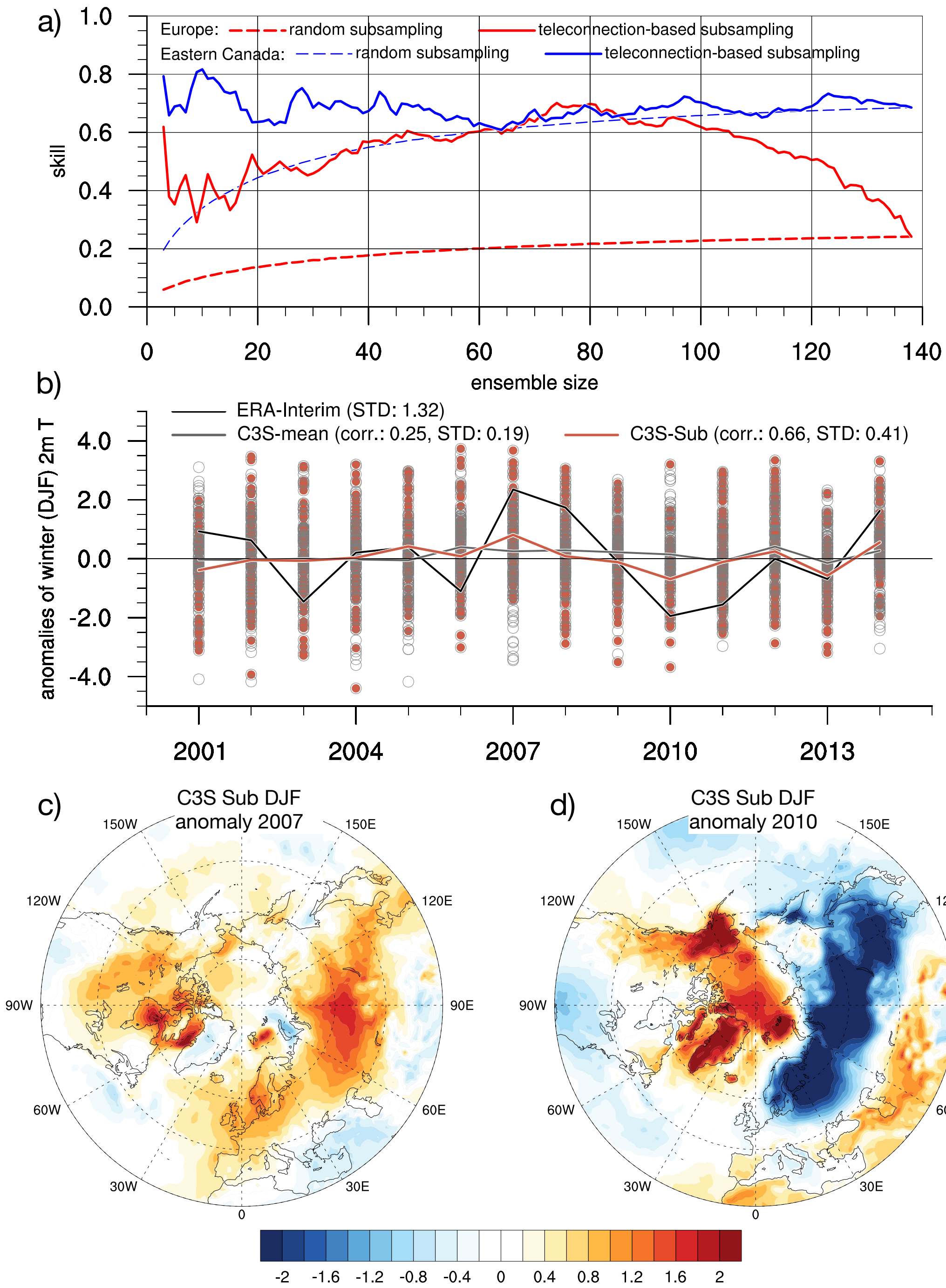


Figure 3. Prediction skill and subsampling of C3S ensemble for the wintertime temperature in a real forecast test from 2001 to 2014. a) prediction skill calculated for the C3S ensemble for two regional means in Central Europe (red) and in the Eastern Canada (blue) using two approaches: random selection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

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 and the European Regional Development fund (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).

Data availability

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-monthly-single-levels?tab=form`). ERA-Interim data are available from ECMWF’s at `www.ecmwf.int/en/forecasts/datasets`.

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*, ncomms15875.

Athanasiadis, P. J., Bellucci, A., Scaife, A. A., Hermanson, L., Materia, S., Sanna, A., … Gualdi, S. (2017). A multisystem view of wintertime NAO seasonal predictions. *Journal of Climate*, *30*(4), 1461–1475.

Barnston, A. G., & Livezey, R. E. (1987). Classification, seasonality and persistence of low-frequency atmospheric circulation patterns. *Monthly weather review*, *115*(6), 1083–1126.

Butler, A. H., Arribas, A., Athanassiadou, M., Baehr, J., Calvo, N., Charlton-Perez, A., … others (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.

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.

Dee, D., Uppala, S., Simmons, A., Berrisford, P., Poli, P., Kobayashi, S., ... others (2011). The ERA-Interim reanalysis: Configuration and performance of the data assimilation system. *Quarterly Journal of the Royal Meteorological Society*, *137*(656), 553–597.

Dobrynin, M., Domeisen, D. I., Müller, W. A., Bell, L., Brune, S., Bunzel, F., ... Baehr, J. (2018, 2). Improved teleconnection-based dynamical seasonal predictions of boreal winter. *Geophysical Research Letters*, *45*. doi: 10.1002/2018GL077209

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.

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.

Hurrell, J. W. (1995). Decadal trends in the North Atlantic Oscillation: regional temperatures and precipitation. *Science*, *269*(5224), 676–679.

Hurrell, J. W., Kushnir, Y., Ottersen, G., & Visbeck, M. (Eds.). (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).

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). Retrieved from `https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2011GL046786` doi: 10.1029/2011GL046786

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.

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.

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.

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

Scaife, A. A., Arribas, A., Blockley, E., Brookshaw, A., Clark, R. T., Dunstone, N., ... Williams, A. (2014, Apr). Skillful long-range prediction of European and North American winters. *Geophysical Research Letters*, *41*(7), 2514–2519. doi: 10.1002/2014GL059637

Scaife, A. A., Comer, R. E., Dunstone, N. J., Knight, J. R., Smith, D. M., MacLachlan, C., ... others (2017). Tropical rainfall, Rossby waves and regional winter climate predictions. *Quarterly Journal of the Royal Meteorological Society*, *143*(702), 1–11.

Scaife, A. A., Folland, C. K., Alexander, L. V., Moberg, A., & Knight, J. R. (2008). European Climate Extremes and the North Atlantic Oscillation. *J. Climate*, *21*, 72–83.

Scaife, A. A., & Smith, D. (2018). A signal-to-noise paradox in climate science. *npj Climate and Atmospheric Science*, *1*(1), 28.

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.

Hidden potential in predicting wintertime temperature anomalies in the Northern Hemisphere

Mikhail Dobrynin[1,2], André Düsterhus[3], Kristina Fröhlich[1], Panos Athanasiadis[5], Paolo Ruggieri[5,6], Wolfgang A. Müller[7], Johanna Baehr[1]

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

Key Points:

- Temperature anomalies can be skilfully predicted for the upcoming winter through increased variability and skill of predicted NAO
- Skilful prediction of temperature anomalies in the Northern Hemisphere for upcoming winter

Corresponding author: Mikhail Dobrynin, `mikhail.dobrynin@dwd.de`

Supplementary information

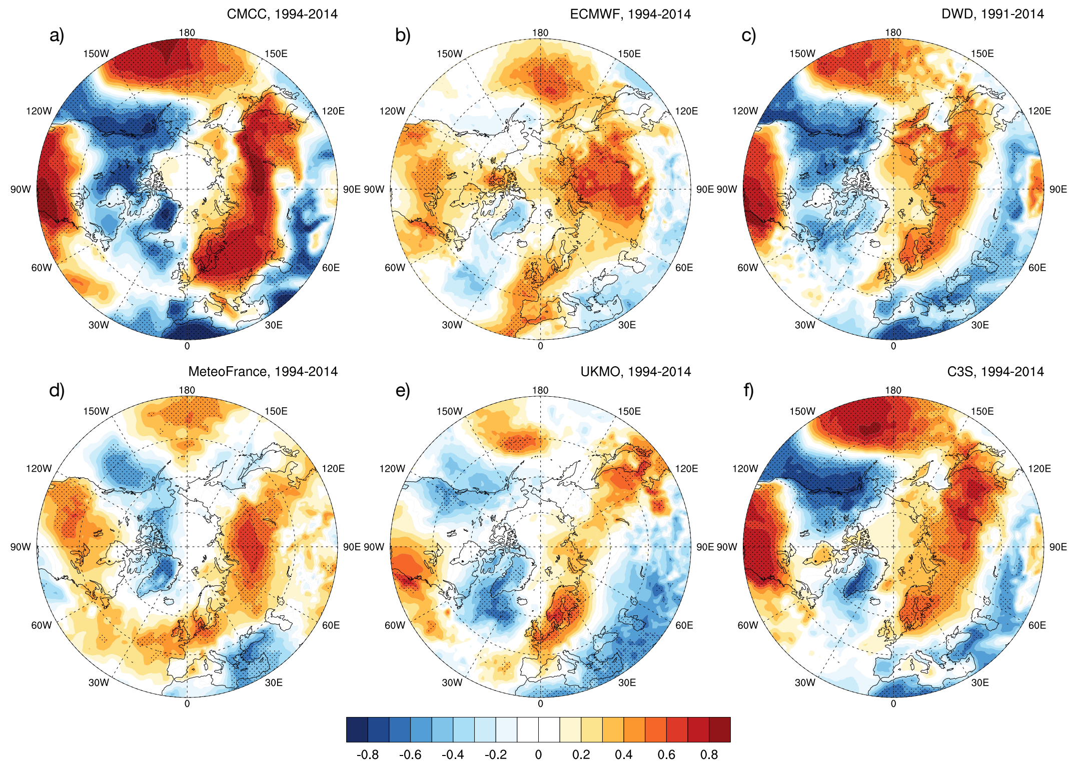
Figure S1 Response of wintertime (DJF) 2-meter temperature on the NAO variability.

Figure S5 Normalised winter (DJF) NAO index.

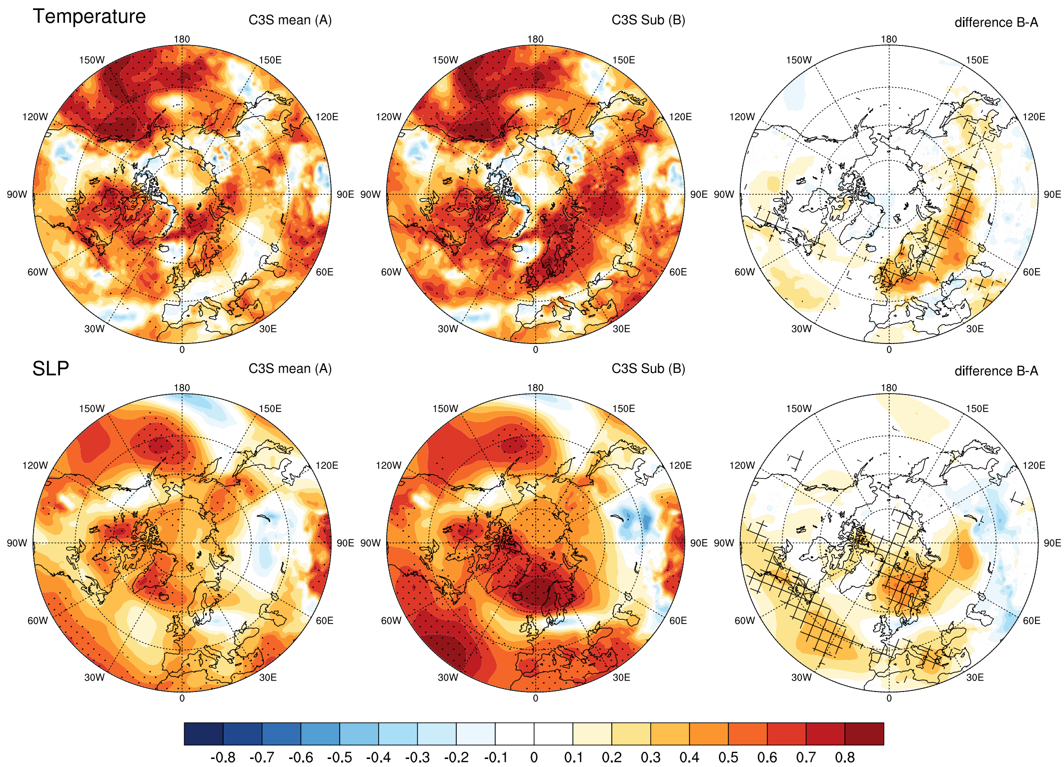
Figure S3 ERA-Interim first-guess of the winter NAO index.

Figure S2 Subsampling of the multi-model ensemble C3S.

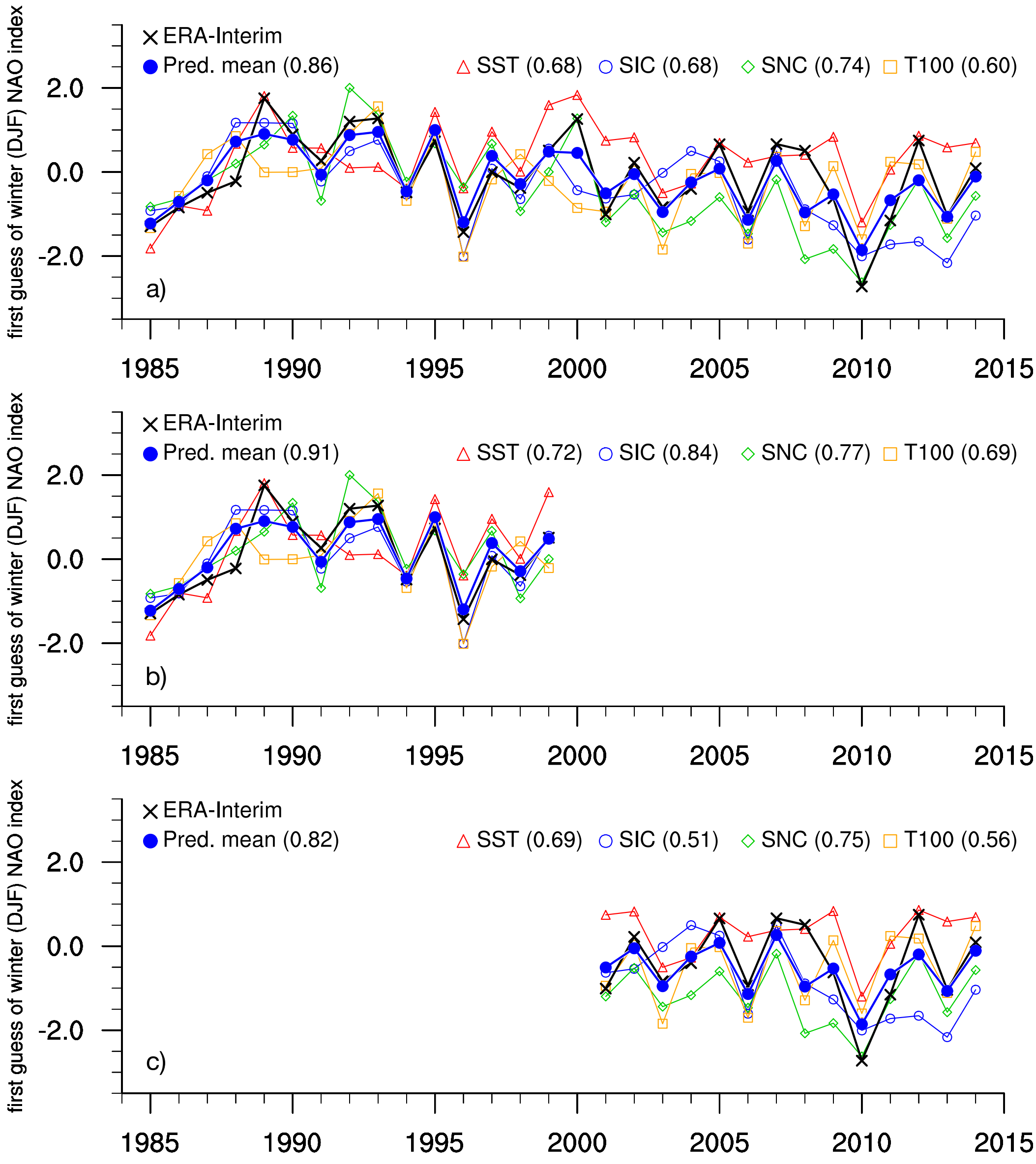
Figure S4 Subsampling of individual models from C3S.



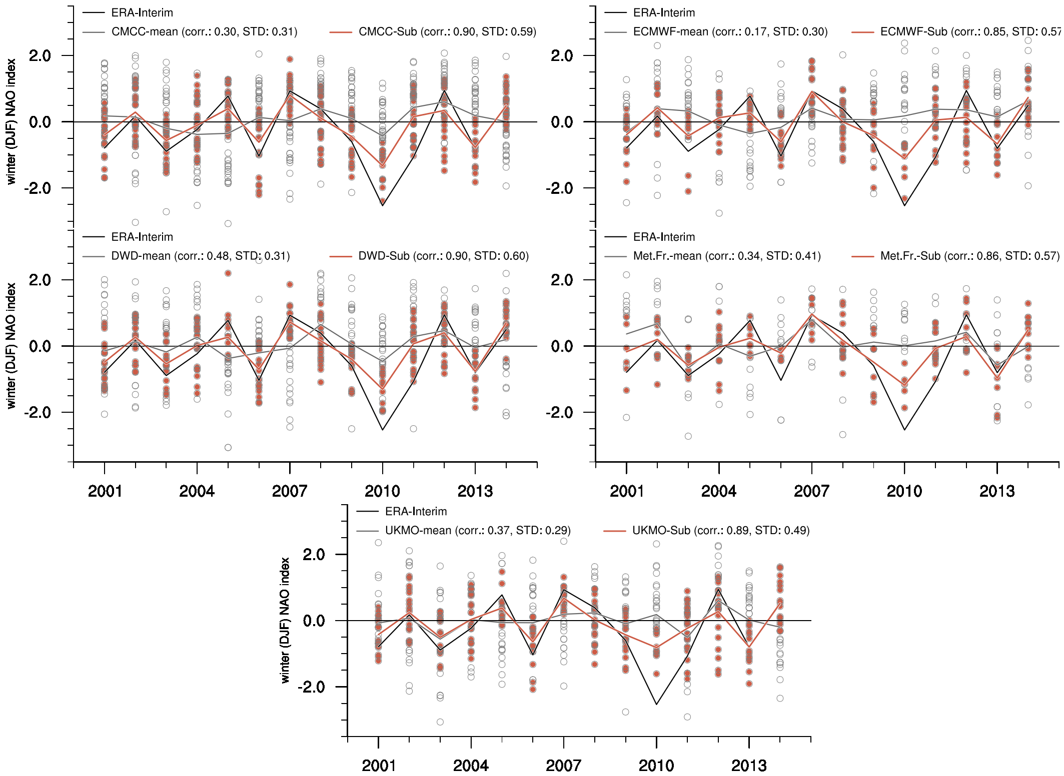
Supplementary Figure S1. Response of wintertime (DJF) 2-meter temperature on the NAO variability. Correlation between 2-meter DJF temperature and NAO index for individual models (a–e), and for multi-model ensemble C3S (f).



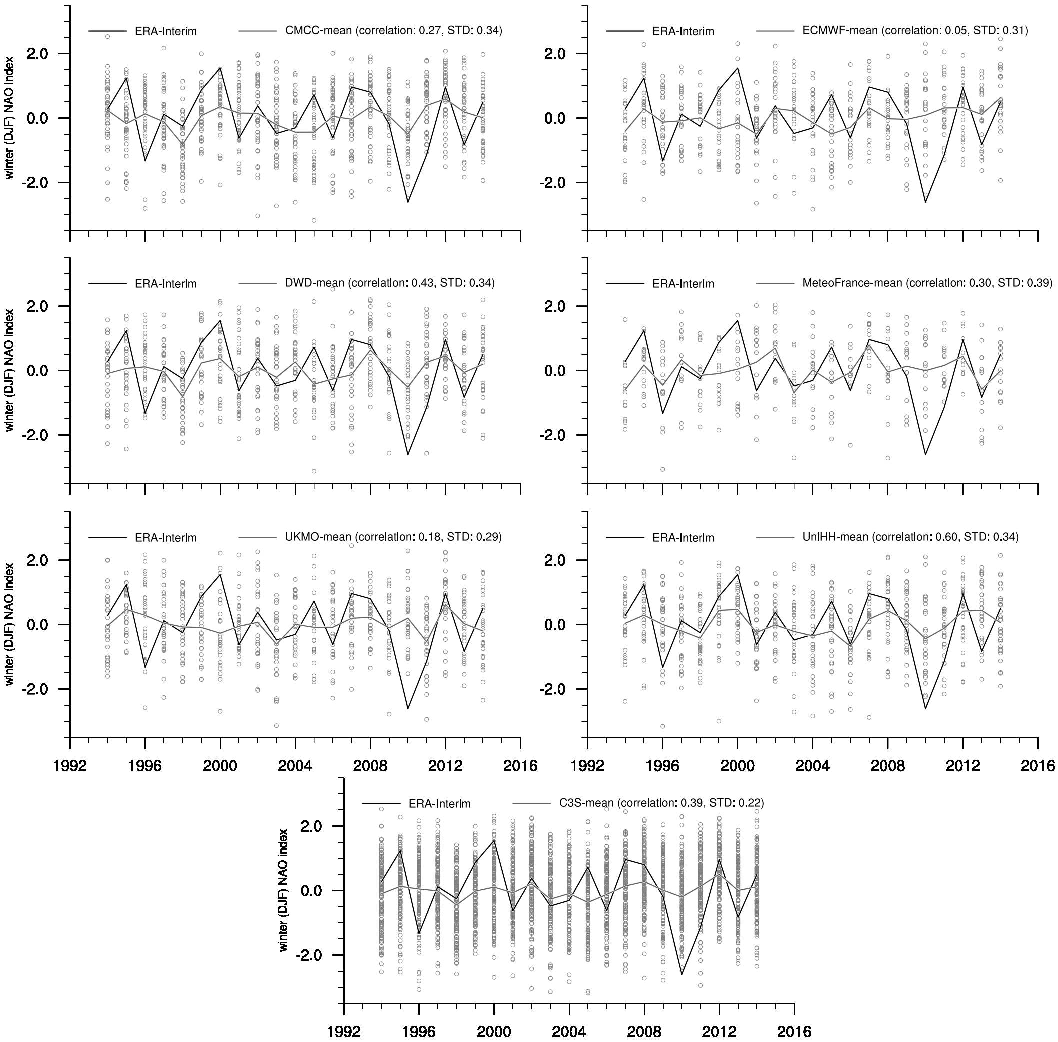
Supplementary Figure S2. Subsampling of the multi-model ensemble C3S. Anomalies correlation coefficient between C3S ensemble mean (left column) and C3S subsampled mean (middle column) and the ERA-Interim calculated for 2-meter temperature (upper panel) and sea level pressure (lower panel) in a real forecast test from 2001 to 2014. Regions that are significant at the 95% confidence level are indicated by dots on the maps in the left and middle column. Hashing on the differences plots (right column) indicates regions that became significant after subsampling.



Supplementary Figure S3. ERA-Interim first-guess of the winter NAO index. Statistically predicted from the ERA-Interim predictors winter NAO index for three periods: a) from 1985 to 2014, b) 14 years similar to real test period starting from 1985 to 1999, and c) real forecast test from 2001 to 2014. The training period from 1979 until the year previous to forecast year was used for all plots. Values of significant at the 95% confidence level correlation between the NAO index and each autumn predictor sea surface temperature in the North Atlantic (SST), Arctic sea ice concentration (SIC), snow cover in Eurasia (SNC) and 100 hPa level air temperature (T100), and for mean predictor (Pred. mean) are given in parentheses.



Supplementary Figure S4. Subsampling of individual models from C3S. NAO prediction skill of individual C3S models for the NAO index in a real forecast test from 2001 to 2014 as calculated from subsampled ensemble (orange line) and from ensemble mean (gray line) compared to the EAR-Interim (black line). Open circles denote each C3S ensemble member, filled circles indicate subsampled due to NAO teleconnection-based approach ensemble members. 13, 8, 10, 5, and 9 members for each predictor were selected for CMCC, ECMWF, DWD, Meteo France, and UKMO system respectevelly.



Supplementary Figure S5. Normalised winter (DJF) NAO index. Prediction skill of the winter NAO index is calculated as a correlation between the ERA-Interim NAO and the ensemble mean for each prediction system, and for the C3S ensemble mean. Correlations and standard deviation of ensemble means (STD) are given in parentheses. The range of the NAO prediction skill varies from 0.05 to 0.43 when all systems are individually considered. In general, all models individually and as a multi-model ensemble underestimate the variability of the NAO index calculated as a standard deviation (hereafter STD) of the ensemble mean. The NAO STD varies from 0.29 to 0.39 for individual models comparing to 1.00 for ERA-Interim NAO. The Universität Hamburg (UniHH-30) seasonal prediction system is not a part of the multi-model C3S ensemble and shown here for information only.