Key Points:
- Seasonal NAO skill is significantly enhanced to 0.83 by combining a dynamical and statistical forecast
- Enhanced NAO skill improves prediction of surface temperature, precipitation, and sea level pressure in Europe
- Teleconnection-based subsampling approach can be potentially used in operational seasonal prediction systems

Improved Teleconnection-Based Dynamical Seasonal Predictions of Boreal Winter

Mikhail Dobrynin[1] (ID), **Daniela I. V. Domeisen[2]** (ID), **Wolfgang A. Müller[3]** (ID), **Louisa Bell[3]**, **Sebastian Brune[1]**, **Felix Bunzel[3]** (ID), **André Düsterhus[1]** (ID), **Kristina Fröhlich[4]** (ID), **Holger Pohlmann[3]** (ID), and **Johanna Baehr[1]** (ID)

[1]Institute of Oceanography, Center for Earth System Research and Sustainability (CEN), Universität Hamburg, Hamburg, Germany, [2]Institute for Atmospheric and Climate Science, ETH Zürich, Zürich, Switzerland, [3]Max Planck Institute for Meteorology, Hamburg, Germany, [4]Deutscher Wetterdienst (DWD), Offenbach am Main, Germany

Abstract Climate and weather variability in the North Atlantic region is determined largely by the North Atlantic Oscillation (NAO). The potential for skillful seasonal forecasts of the winter NAO using an ensemble-based dynamical prediction system has only recently been demonstrated. Here we show that the winter predictability can be significantly improved by refining a dynamical ensemble through subsampling. We enhance prediction skill of surface temperature, precipitation, and sea level pressure over essential parts of the Northern Hemisphere by retaining only the ensemble members whose NAO state is close to a "first guess" NAO prediction based on a statistical analysis of the initial autumn state of the ocean, sea ice, land, and stratosphere. The correlation coefficient between the reforecasted and observation-based winter NAO is significantly increased from 0.49 to 0.83 over a reforecast period from 1982 to 2016, and from 0.42 to 0.86 for a forecast period from 2001 to 2017. Our novel approach represents a successful and robust alternative to further increasing the ensemble size, and potentially can be used in operational seasonal prediction systems.

Plain Language Summary Predicting Northern Hemisphere winter conditions, which are controlled largely by fluctuations in the pressure filed over the North Atlantic (North Atlantic Oscillation, NAO), for the next season is a major challenge. Most state-of-the-art seasonal prediction systems show a correlation between observed and predicted NAOs of less than 0.30. Our novel approach uses dynamical links (teleconnections) between the autumn state of sea surface temperature in the North Atlantic, Arctic sea ice, snow in Eurasia, and stratosphere temperature over the Northern Hemisphere as predictors of the NAO in the subsequent winter to subsample a dynamical reforecast ensemble. We select only the ensemble members that consistently reproduce winter NAO states that evolve in accordance with the autumn state of these predictors. As a result the winter NAO prediction skill increases to a correlation value of 0.83. Considering these well established NAO teleconnections in our Earth system model leads to an improved prediction skill of European winter conditions, that is, surface temperature, precipitation, and sea level pressure. Our results advance seasonal prediction of European weather to a level that is usually limited to tropical regions and are relevant for a variety of societal sectors, such as global and national economies and energy and water resources.

1. Introduction

Over the North Atlantic, the North Atlantic Oscillation (NAO) index is the leading mode of winter climate variability, and variability in the NAO has a widespread impact on changes in temperature and precipitation (Hurrell, 1995; Hurrell et al., 2003; Thompson et al., 2003), marine ecosystems (Drinkwater et al., 2003), and storm track location (Ulbrich & Christoph, 1999). Positive and negative NAO phases are characterized by a change in the position of the jet stream, which is in turn associated with a change in the storm tracks. During a negative NAO phase, the jet stream is located farther south over the European continent; thus, southern Europe experiences a wetter winter, whereas northern Europe experiences colder and drier weather (Hurrell et al., 2003). Anomalous persistence or variability in the NAO can have a significant impact on European weather, including extreme events (Jung et al., 2011; Maidens et al., 2013; Scaife et al., 2008). Such changes have an impact on society, including the economic sector (Marshall et al., 2001) and insurance

industry (Pinto et al., 2012), and they also affect renewable energy and water resources (Jerez et al., 2013; López-Moreno et al., 2007). Therefore, a robust seasonal prediction system that provides a sufficiently accurate NAO forecast would be beneficial for mitigating the potentially negative impacts of the increased occurrence of extreme weather conditions. Because of the intrinsic chaotic nature of the ocean-atmosphere system, the potential use of dynamical seasonal prediction systems to forecast the winter NAO index has not been considered feasible for a long time, although recent research has shown that the intrinsic predictability of the NAO may extend to around 20 days (Domeisen et al., 2017).

Only recent studies have demonstrated skillful dynamical seasonal predictions of the winter NAO (Butler et al., 2016; Müller et al., 2005; O'Reilly et al., 2017; Scaife et al., 2014; Weisheimer et al., 2017) with a prediction skill of about 0.5–0.6 depending on the period evaluated. These improvements are partially associated with the improved ability of seasonal prediction systems to simulate the sources of the NAO predictability, and it is expected that NAO prediction should be further improved by an increase in the ensemble size (Butler et al., 2016; Scaife et al., 2014). Unlike dynamical systems, existing statistical models depending on a set of predictors and evaluation period demonstrate a moderate to high (approximately 0.5–0.7) significant skill of winter NAO prediction (Hall et al., 2017; Wang et al., 2017).

Here we propose an alternative approach in which the NAO predictability is enhanced through ensemble subsampling. Instead of further increasing the ensemble size, we retain or reject individual ensemble members based on known and well-documented physical links (Smith et al., 2014) between the sea surface temperatures (SSTs) in the North Atlantic (Czaja & Frankignoul, 2002) and to some extent in the North Pacific (Wang et al., 2017), Arctic sea ice (Strong et al., 2009; Sun et al., 2015), Eurasian snow cover (Cohen & Jones, 2011; Peings et al., 2013), and stratospheric variability (Butler & Polvani, 2011; Domeisen et al., 2015; Scaife et al., 2016) in autumn and the state of the NAO in the subsequent winter.

2. Setup and Methods

2.1. MPI-ESM-MR-Based Seasonal Prediction System

The seasonal prediction system (Baehr et al., 2015) is based on the mixed resolution (MR) Coupled Model Intercomparison Project Phase 5 version of the Max Planck Institute for Meteorology Earth system model, that is, the MPI-ESM-MR (Giorgetta et al., 2013). The atmospheric part of MPI-ESM-MR (ECHAM6) has an approximate horizontal resolution of 200 km (1.875°) and 95 vertical layers up to 0.01 hPa. The ocean component the Max Planck Institute ocean model (MPIOM) has an approximate horizontal resolution of 40 km (0.4°) and 40 vertical layers.

For each year within 1982–2016, a 30-member ensemble of reforecasts (hereafter MR-30) is initialized on 1 November from an assimilation experiment where ERA-Interim data (Dee et al., 2011) is assimilated in the atmospheric model component, and ORA-S4 data (Balmaseda et al., 2013) and National Snow and Ice Data Center observations (Comiso, 1995) are assimilated in the ocean/sea ice component. Newtonian relaxation ("nudging") is used as an assimilation technique in full-field mode throughout the model space. Full-field nudging is used for vorticity, divergence, temperature, and surface pressure in the atmosphere and temperature, salinity, and sea ice concentration in the ocean. The initial conditions in the ocean are perturbed using bred vectors with a vertically varying norm (Baehr & Piontek, 2014). In the atmosphere each ensemble member has a slightly disturbed diffusion coefficient in the uppermost model layer.

2.2. NAO in the MPI-ESM-MR Seasonal Prediction System

The boreal winter (averaged over December, January, and February) NAO index is calculated using an empirical orthogonal function (EOF) analysis (Barnston & Livezey, 1987) from sea level pressure (SLP) in the Northern Hemisphere (NH). We calculated the NAO in the region limited according to the latitude range 20°N to 90°N and the longitude range 90°W to 60°E (Figure 1). The principal component of the leading EOF of SLP represents the NAO index (Kutzbach, 1970). The EOF-based December, January, and February NAO is calculated for the ERA-Interim data and for each of the MR-30 ensemble members. The ensemble mean NAO is calculated as the average over all ensemble members. Hereafter, we use the ERA-Interim NAO as a reference for comparisons with the MR-30 NAO. The NAO_{MR-30} and $NAO_{ERA-Interim}$ indices are normalized by their respective standard deviations.

2.3. Cross Validation and Results Verification

We apply a cross validation by leaving out every year and 3 years (hereafter leave-one-out and leave-three-out, respectively) centered on the year left-out for calculation of correlation coefficients over the reforecast period
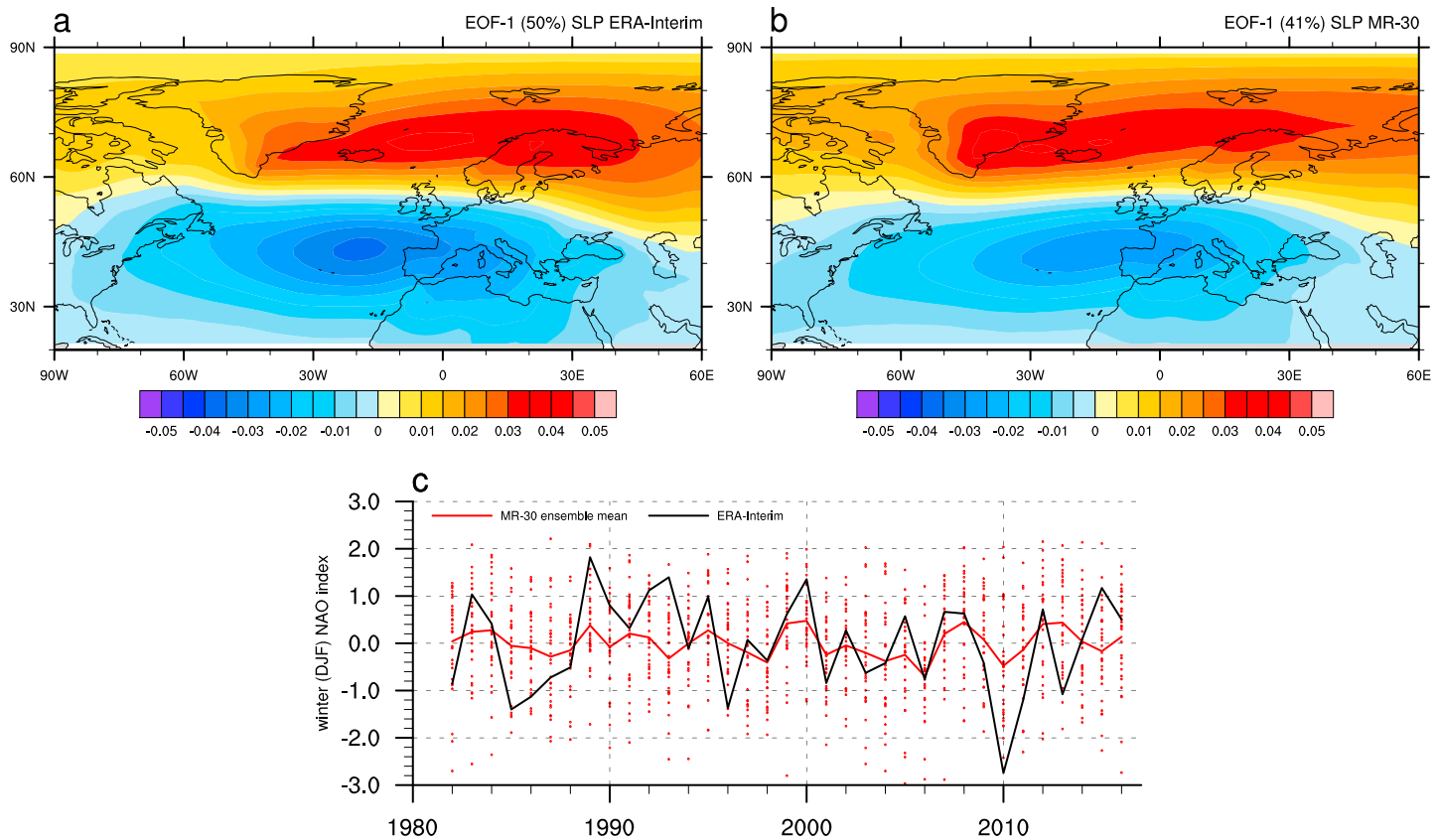
Figure 1. Winter (December, January, and February, DJF) North Atlantic Oscillation (NAO). The respective leading empirical orthogonal functions (EOF1) of sea level pressure (SLP) from the ERA-Interim reanalysis (a) and the MR-30 ensemble (b) calculated over the period from 1982 to 2016. (c) The normalized winter NAO index calculated from EOF1 for the ERA-Interim reanalysis (black line) and the MR-30 ensemble (red lines). Red dots denote the MR-30 ensemble members. EOF1 of the reforecast ensemble (MR-30) explains 41% of the SLP variance for the ensemble mean, whereas the EOF1 for the ERA-Interim reanalysis explains 50%. Although the overall patterns and locations of the zero line are consistent between the MR-30 and ERA-Interim, the exact locations of the center of both the negative and the positive SLP anomalies are shifted slightly westward in the MR-30 compared to ERA-Interim.

from 1982 to 2016. The prediction skill is represented by correlation values between the reforecasted and reanalyzed time series calculated as mean, minimum, and maximum values from cross validation (indicated in parentheses). Additionally, root-mean-square error (RMSE) and spread-error ratio are also calculated for every cross-validation iteration.

Leave-one-out cross validation is applied for correlation analysis between the ensemble mean NAO_{MR-30} and the $NAO_{ERA-Interim}$ time series over the reforecast period. The mean and range of the cross-validated correlation coefficient are 0.49 (0.42–0.57). They are significant at the 99% confidence level (Figure 1 and Table S1 in the supporting information).

Afterward, cross validation is applied in two steps, first, to the correlation analysis between autumn model parameters used as predictors (see section 3.1) and the winter NAO index (leave-one-out and leave-three-out), and second, to the NAO time series calculated for each step of cross validation applied to predictors (leave-one-out). The first step of cross validation is used to discard the impact of the NAO phase on the robustness of the predictors. In the second step, the NAO time series per left-out year are cross validated for estimation of the correlation range for the NAO index comparing to ERA-Interim. A bootstrapping test with 500 samples is applied to every cross-validation iteration for calculation of significance at a given confidence level.

We also perform a "real" forecast test where we split the reforecast period into a "training period" (1981–2000) and a real forecast period (2001–2017). For the real forecast we strictly only use the information from previous years.

The anomaly correlation coefficient (ACC) represents the prediction skill (Collins, 2002) for selected Earth system parameters such as surface temperature, total precipitation, and sea level pressure. The ACC is calculated for every cross-validation iteration between reforecasted and reanalyzed fields. In the manuscript ACCs only for the year 2016 left-out in cross validation are shown.

3. Methodology of Subsampling

For any given winter, some members of the MR-30 predict NAO states that are too positive and too negative (Figure 1c) suggesting that the ensemble spread might be too large. Yet we find that the ensemble spread in NAO_{MR-30} of 4.09 (3.73–4.15) hPa is comparable to the standard deviation of 4.66 hPa in $NAO_{ERA-Interim}$. For a reliable forecast, the ratio between the ensemble standard deviation to the RMSE (spread-error ratio) should equal 1, while we find a spread-error ratio of 1.16 (1.13–1.26) in NAO_{MR-30}, suggesting that the ensemble spread is too large compared with the reforecast error (Ho et al., 2013). To improve the spread-error ratio, we assess whether ensemble members could be retained or rejected by comparing the state of the climate in the NH in autumn (before the start of the reforecast) and the state of the NAO in the subsequent winter.

3.1. Selection of Autumn Predictors

As predictors, we use Earth system parameters for which dynamical links to the NAO are well established and documented in the literature. We investigate the autumn states of the ocean, sea ice, land surface, and stratosphere as preconditions (hereafter predictors P_I–P_{IV}) for the subsequent winter NAO. The first predictor is the SST (P_I) in the North Atlantic, which has previously been found to be related to the winter NAO (Bjerknes, 1964; Czaja & Frankignoul, 2002). SSTs are linked to the NAO through air-sea interaction, which suggests a lagged response in the atmospheric circulation to the thermal forcing of the upper ocean (Kushnir et al., 2002; Wang et al., 2004), which would allow the reconstruction of the winter NAO from autumn SSTs in the North Atlantic (Rodwell et al., 1999; Robertson et al., 2000). The second predictor is the sea ice volume (P_{II}) in the Arctic, which has an impact on the atmospheric circulation in the following season (Liptak & Strong, 2014) due to changes in the ocean-atmosphere heat fluxes and wind stress caused by sea ice anomalies. The third predictor is the snow depth (P_{III}) in Eurasia in the latitude range 40°N to 90°N and longitude range 50°E to 150°E. A lagged snow-atmosphere interaction mechanism may be the link between autumn snow cover over Eurasia and the NAO in the subsequent winter season (Bojariu & Gimeno, 2003; Cohen et al., 2014). The fourth predictor is the stratospheric temperature at 100 hPa (P_{IV}). Recent findings (Stockdale et al., 2015) have shown that initial atmospheric conditions, especially in the stratosphere, exert a stronger influence in seasonal forecasts than previously thought.

3.2. "First Guess" NAO Prediction for Full Re-Forecast

In our assimilation experiment we conduct a correlation analysis using the leave-one-out cross validation between the winter NAO and the autumn state of predictors P_I–P_{IV} over the reforecast period from 1982 to 2016. Through this correlation analysis, we identify for every predictor the regions where significant correlations with the winter NAO are found, which are also the regions for which teleconnections are well established in the literature and represented in a "free" model simulation (see Figure 2a and Table S1). For every predictor in any given year and for every left-out year in cross validation, we then average the state of the predictor over the respective area (Figure 2b) to make a simple first guess prediction of the winter NAO state (NAO(P)). Taking into account that a typical winter hindcast starts at 1 November, we select October as the last month before model initialization for comparison with the subsequent winter NAO for SST, snow depth, and stratospheric temperature and September for ice volume because the minimum ice cover in the Arctic occurs in September (Figure S1). We construct four predictors as detrended time series of area-weighted values over the regions that exhibit a significant positive correlation with the observed NAO time series. In leave-one-out cross validation the patterns of significant correlation between predictors and NAO vary and provide a set of cross-validated first guess values for the following subsampling. Note that the location of these significant regions varies from year to year depending on the NAO state. We use all regions computed for each cross-validation step and apply them to every year in the reforecast period. All predictors are calculated over the entire period of the assimilation experiment from 1982 to 2016 and normalized by their respective standard deviations. In our assimilation experiment, the mean and range (indicated in parentheses) of the leave-one-out cross-validated correlation values between the four predictors and the winter NAO index are 0.58 (0.48–0.63) for SST, 0.51 (0.47–0.58) for sea ice volume, 0.71 (0.61–0.74) for snow depth, and 0.55 (0.44–0.58) for stratospheric temperature (Figures S1a–S1d).
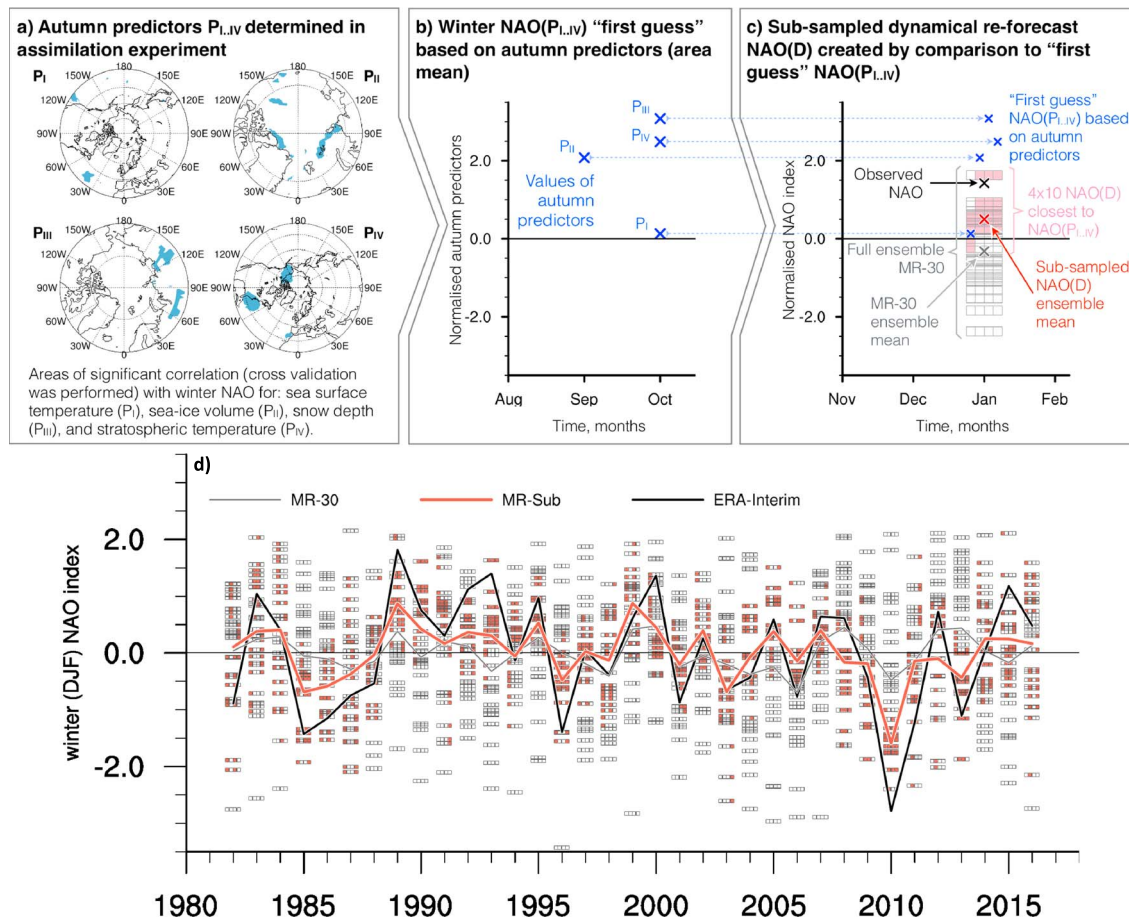
Figure 2. Schematic representation of the subsampling method for the MR-30 dynamical reforecast (a, b) and construction of the subsampled ensemble MR-Sub of the winter North Atlantic Oscillation (NAO) (d). (a) A correlation analysis (in cross-validation mode) based on the assimilation experiment between the winter (December, January, and February, DJF) NAO index and four autumn predictors: (P_I) SST, (P_{II}) sea ice volume, (P_{III}) snow depth, and (P_{IV}) stratospheric temperature. (b) Autumn predictors P_I–P_{IV} are defined as time series of a detrended mean over the region with a significant positive correlation at the 95% confidence level. For every autumn-winter pair, the autumn predictors project the "first guess" state of the NAO in the subsequent winter: NAO(P_I–P_{IV}) = P_I–P_{IV}. In (c) the dynamical seasonal prediction system provides an ensemble of the reforecast winter NAO(D) indicated by gray four-cell blocks. Each cell represents one of the four predictors (from left to right): SST, sea ice volume, snow depth, and stratospheric temperature. An orange-filled cell indicates that the respective ensemble member is selected for the new MR-Sub ensemble by one or more of the four predictors. For the period from 1982 to 2016 (d) the correlations at the 99% confidence level between the MR-30 (gray line), the MR-Sub ensemble mean NAO (red line), and the ERA-Interim NAO (black line) are 0.49 and 0.86, respectively. Here an example for the year 2016 left-out in cross validation is shown.

Based on a leave-one-out cross-validated correlation analysis between the four predictors and the winter NAO in our assimilation experiment, we assume that the autumn states of the ocean, sea ice, land surface, and stratosphere may serve as predictors of the winter NAO. Therefore, we use a linear relationship between the normalized predictor values P_I–P_{IV} and the normalized winter NAO(P_I–P_{IV}): NAO(P_I–P_{IV}) = P_I–P_{IV}.

We confirm the robustness of our approach by applying leave-three-out (centered on the year that is left-out) cross validation. The results (see Table S1) show a stable level of correlation for all predictors comparing leave-one-out and leave-three-out cross-validation. We also find similar levels of correlation between autumn predictors and winter NAO in a "free" uninitialized historical Coupled Model Intercomparison Project Phase 5 MPI-ESM-MR simulation (see Table S1).

3.3. First Guess NAO Prediction in Real Forecast

For a real forecast test we use a period from 2001 to 2017. For each forecast year, we derive the patterns for the predictors from all previous years starting from 1981. For example, the autumn predictors calculated over the period from 1981 to 2000 (training period) are used for the winter NAO forecast for 2001, a training period from 1981 to 2001 is used for the winter NAO forecast for 2002 and so on.

The patterns for the predictors estimated over only 20 years depend highly on the occurrence of long-lived or extreme positive or negative NAO phases in the preceding period. Therefore, we perform an additional perturbation of the predictors' variability by leaving 1 year out over the training period. We combine all significant regions for each predictor into one region, using only the points, where at least half of the years show significant correlation for the respective training period.

3.4. Implementation of Subsampling to the MR-30 Ensemble

We use a subsampling approach to calculate a new MR-Sub ensemble mean of the winter NAO, which is identical for the reforecast and for the real forecast test. From the full MR-30 ensemble we select the ensemble members in the dynamical reforecast ensemble whose NAO state in winter is closest to our first guess NAO prediction. For every predictor in any given year and for every left-out year(s) in cross validation, we individually select from the total 30 ensemble members the 10 members with the smallest difference between our first guess NAO prediction and the reforecasted NAO state in the respective ensemble member (Figure 2c). The four 10-member subsets (one per predictor) are then combined into one large ensemble (hereafter MR-Sub). Because each particular ensemble member can be selected by more than one predictor, duplicate entries are removed from the MR-Sub ensemble (Figure 2d). Therefore, the maximum size of the new subensemble can be 30 members (the full ensemble), and the minimum size is 10 members (if all predictors select the same ensemble members). This subsampling approach is applied to every autumn-winter pair over the period of the reforecast from 1982 to 2016 in leave-one-out and leave-three-out cross validation. Computed for every predictor's cross-validation step time series of the NAO are cross validated as well. The real forecast test is conducted from 2001 to 2017.

## 4. Results of the Subsampling Implementation

4.1. Reforecast From 1982 to 2016

As a result, over the reforecast period from 1982 to 2016, the MR-Sub shows a high cross-validated mean correlation coefficient calculated with the $NAO_{ERA-Interim}$ of 0.83 (0.77–0.87) at the 99% significance level. Prediction skill varies due to change of location of significant regions for each predictor in every predictors' cross-validation step (see supporting information Table S1). Moreover, the computed NAO time series also demonstrate a stable correlation level of 0.84 (0.79–0.87) in leave-one-out cross-validation (see supporting information Table S1 and Figure S5). The size of MR-Sub varies from 10 to 30 members over the period of the simulation. Test studies that include a fixed number of the 10 "best" selected ensemble members for each year and all 40 selected members, including those selected by more than one predictor (not shown), do not indicate a notable impact on our results and conclusions. The increase in the number of selected ensemble members demonstrates an increase of mean cross-validated correlation between NAO_{MR-Sub} and $NAO_{ERA-Interim}$ from 0.83 (0.77–0.87) for 10 members to 0.86 (0.79–0.89) for 15 members and a decrease to 0.75 (0.71–0.78) for 25 members.

All four predictors $P_I–P_{IV}$ contribute to the improved representation of the winter NAO in MR-Sub (Figures 2d and S2). Therefore, correlation for MR-Sub is higher than for any single predictor-based first guess NAO(P), with a maximum of mean cross-validated correlation coefficient of 0.75 (0.65–0.78) for P_{III} and a minimum of 0.5 (0.47–0.57) for P_{II}, compared with the $NAO_{ERA-Interim}$ (Figure S1 and Table S1). The mean and range of cross-validated spread-error ratio and RMSE of the NAO_{MR-Sub} relative to the $NAO_{ERA-Interim}$ are reduced to 1.01 (0.98–1.07) and 2.90 (2.79–2.94) hPa, respectively, compared with the values of 1.16 (1.13–1.26) and 4.09 (3.73–4.15) hPa for the full ensemble MR-30. We confirm the robustness of predictors by the same level of correlation with the NAO in leave-three-out cross validation (see section 2 and Table S1 for more details).

For the subsampled ensemble the correlation between NAO_{MR-Sub} and $NAO_{ERA-Interim}$ has the same value of 0.83 for both leave-one-out and leave-three-out cross validations (Table S1). Uncertainty in terms of RMSE increases to 3.06 hPa for leave-three-out compared to 2.90 hPa for leave-one-out cross validation. The subsampled ensemble became underdispersed (spread-error ratio < 1) in leave-three-out cross validation, most likely because of an insufficient spread of the initial ensemble which is considerably shorted by removing of 3 years centered on the year that is left-out.

In contrast to the first guess NAO prediction and possible other statistical approaches, ensemble subsampling allows us to build a new subsampled ensemble mean for any simulated field from once selected ensemble members with a reasonable representation of NAO teleconnections. For surface temperature, we find
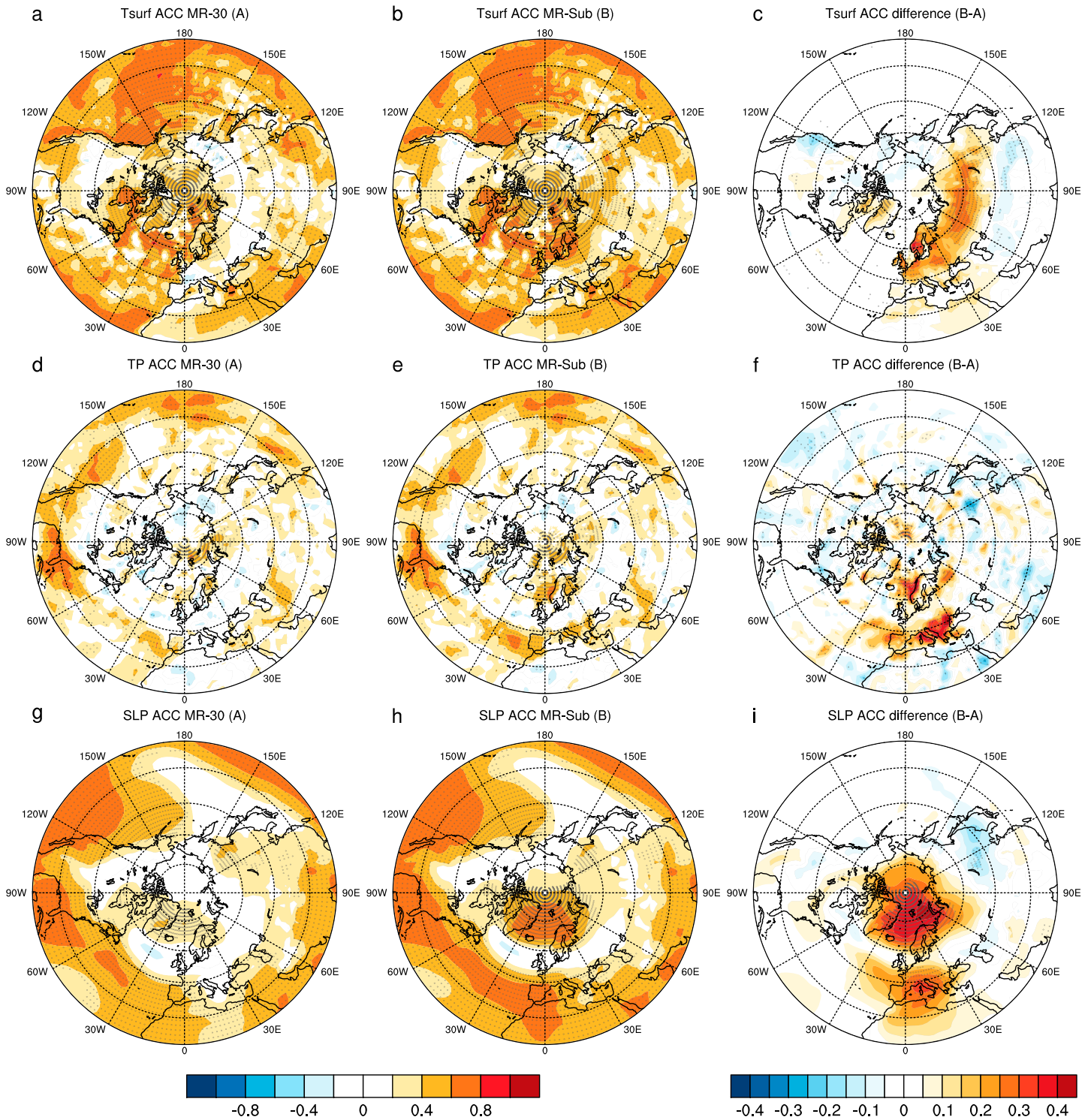
Figure 3. Improvement of the prediction of the selected Earth system parameters due to more accurate prediction of the winter NAO. Anomaly correlation coefficient (ACC) calculated for the MR-30 (left column) and MR-Sub (middle column), and differences (right column) relative to the ERA-Interim reanalysis for surface temperature (a–c), total precipitation (d–f), and sea level pressure (g–i). Regions that are significant at the 95% confidence level are indicated by stippling on the maps. Here an example for the year 2016 left-out in cross validation is shown.
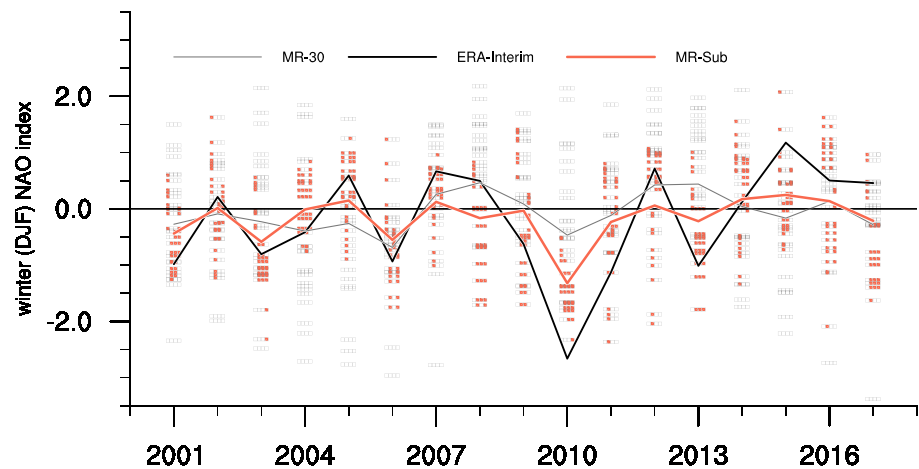
Figure 4. "Real" forecast test of the winter North Atlantic Oscillation (NAO) for the period from 2001 to 2017. The forecasted winter NAO values are calculated separately for each year and then merged into a single time series. The correlations between the MR-30 (gray line), the MR-Sub ensemble mean NAO (red line), and the ERA-Interim NAO (black line) are 0.42 and 0.86, respectively (significant at the 99% confidence level). Similar to Figure 2 each cell in gray four-cell blocks represents one of the four predictors. Ten ensemble members for each predictor are used in subsampling. DJF = December, January, and February.

a significant increase of prediction skill (ACC relative to ERA-Interim) in MR-Sub compared to MR-30 over Eurasia, the Scandinavian Peninsula, and northern Europe (Figures 3a–3c). The ACC for surface temperature from MR-Sub is now significant in Eurasia. The ACC for total precipitation (Figures 3d–3f) in northern and southern Europe shows an increase in MR-Sub. For the sea level pressure, we find a significant increase in the ACC, particularly in the polar regions between 60°N and 90°N and the longitude range from 30°W to 130°E. The maximum ACC for SLP in MR-Sub is located over the Barents Sea (Figures 3g–3i). The ACC for the Mediterranean region is increased to approximately 0.5. A small decrease in the ACC for SLP and surface temperature occurs in Eurasia south of 45°N and in North America. The ACC for SLP in the North Atlantic slightly decreases between 50°N–60°N and 30°W–0°E. All predictors contribute to an increase of the ACC, demonstrating the added value compared to the ACC calculated for the full MR-30 ensemble (Figure S2).

Note that the overall improvement (with only small local degradation) of the reforecast presents the same regional distribution as the theoretically "perfect" NAO reforecast, which was conducted in reference to the observed NAO. For this hypothetical analysis, we select from the MR-30 the 10 members that best resemble the observed winter NAO. This analysis represents the level of theoretically achievable reforecast skill, with the correlation coefficient between this "perfect" ensemble and the ERA-Interim data for the winter NAO at 0.97 (Figure S3). However, more importantly, the ACC analysis exhibits the level of improvement that can be expected from a "perfect" NAO forecast because an increase in the ACC for surface temperature, total precipitation, and SLP has an equivalent but much more pronounced structure compared with that from the MR-Sub ensemble (Figures 3 and S4). In addition, a pronounced decrease in the ACC also occurs in Eurasia for SLP and total precipitation and in North America for SLP and surface temperature. Therefore, the increased ACC for temperature, total precipitation, and SLP that is observed in MR-Sub is representative of the improvement that can be expected from an improved NAO reforecast.

4.2. Real Forecast Test From 2001 to 2017

Similar to the reforecast, we compute the winter NAO forecast as a real forecast test for the period from 2001 to 2017. During this period the MR-30 ensemble mean exhibits a lower prediction skill for the NAO than for the full period (0.42 and 0.49, respectively). However, when applying subsampling based on predictors calculated using all previous years for each real forecast year, the prediction skill increases from 0.42 to 0.86 (Figure 4). The increase in skill can be attributed to the correction of the NAO phases due to subsampling for example for 2005, 2013, or 2015. The RMSE of the real forecast NAO_{MR-Sub} relative to the $NAO_{ERA-Interim}$ is reduced to 3.06 hPa compared with the values of 4.09 hPa for the full ensemble MR-30.

5. Summary and Conclusions

In summary, we find enhanced seasonal prediction skill of the boreal winter through a combination of a dynamical forecast ensemble and a proper statistical method, and a refinement of a dynamical ensemble due to subsampling. The reforecast skill expressed by the cross-validated correlation between observed and reforecasted winter NAO is significantly improved from 0.49 (0.42–0.57) to 0.83 (0.77–0.87), when the connections between the autumn state of the ocean, sea ice, land, and stratosphere and the subsequent winter NAO are considered. For the real forecast test from 2001 to 2017 the prediction skill of the winter NAO is increased from 0.42 for the full ensemble mean to 0.86 for the subsampled ensemble mean. As a result of a better representation of the winter NAO, the prediction skill for the winter surface temperature, total precipitation, and SLP is improved for considerable parts of the NH. The robustness of our approach is confirmed by using leave-one-out and leave-three-out cross validation in the correlation analysis (see Table S1).

The presented subsampling approach makes use of a simple statistical prediction though fully maintaining the advantage of a dynamical model. We find that a large ensemble is needed for subsampling in order to have an ensemble spread which is comparable to observed variability. Since entire ensemble members are rejected or included in the analysis, the remaining fields are dynamically self-consistent in space and time and can be analyzed for any desired variable the dynamical model provides. Since only observations from the initial state are taken into account for the selection of the dynamical ensemble members, the presented approach is potentially applicable to current operational prediction systems. In our study, the selection of ensemble members is tailored to the NAO. Provided that a physical mechanism that links the initial state and the forecasted variable can be formulated as a first guess prediction, our subsampling of a dynamical ensemble prediction can be formulated for any variable, index, region, or time scale.

References

Baehr, J., Fröhlich, K., Botzet, M., Domeisen, D., Kornblueh, L., Notz, D., et al. (2015). The prediction of surface temperature in the new seasonal prediction system based on the MPI-ESM coupled climate model. *Climate Dynamics*, *44*(9-10), 2723–2735.

Baehr, J., & Piontek, R. (2014). Ensemble initialization of the oceanic component of a coupled model through bred vectors at seasonal-to-interannual timescales. *Geoscientific Model Development*, *7*(1), 453–461.

Balmaseda, M. A., Mogensen, K., & Weaver, A. T. (2013). Evaluation of the ECMWF ocean reanalysis system ORAS4. *Quarterly Journal of the Royal Meteorological Society*, *139*(674), 1132–1161.

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

Bjerknes, J. (1964). Atlantic air-sea interaction. *Advances in Geophysics*, *10*(1), 1–82.

Bojariu, R., & Gimeno, L. (2003). The role of snow cover fluctuations in multiannual NAO persistence. *Geophysical Research Letters*, *30*(4), 1156. https://doi.org/10.1029/2002GL015651

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.

Butler, A. H., & Polvani, L. M. (2011). El Niño, La Niña, and stratospheric sudden warmings: A re-evaluation in light of the observational record. *Geophysical Research Letters*, *38*, L13807. https://doi.org/10.1029/2011GL048084

Cohen, J., & Jones, J. (2011). A new index for more accurate winter predictions. *Geophysical Research Letters*, *38*, L21701. https://doi.org/10.1029/2011GL049626

Cohen, J., Screen, J. A., Furtado, J. C., Barlow, M., Whittleston, D., Coumou, D., et al. (2014). Recent Arctic amplification and extreme mid-latitude weather. *Nature Geoscience*, *7*(9), 627–637.

Collins, M. (2002). Climate predictability on interannual to decadal time scales: The initial value problem. *Climate Dynamics*, *19*(8), 671–692.

Comiso, J. C. (1995). *SSM/I sea ice concentrations using the bootstrap algorithm* (Vol. 1380). Greenbelt, MD: National Aeronautics and Space Administration, Goddard Space Flight Center.

Czaja, A., & Frankignoul, C. (2002). Observed impact of Atlantic SST anomalies on the North Atlantic Oscillation. *Journal of Climate*, *15*(6), 606–623.

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.

Domeisen, D. I., Badin, G., & Koszalka, I. M. (2017). How predictable are the Arctic and North Atlantic Oscillations? Exploring the variability and predictability of the Northern hemisphere. *Journal of Climate*, *31*(3), 997–1014.

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

Drinkwater, K. F., Belgrano, A., Borja, A., Conversi, A., Edwards, M., Greene, C. H., et al. (2003). The response of marine ecosystems to climate variability associated with the North Atlantic Oscillation. In J. W. Hurrell, Y. Kushnir, G. Ottersen, & M. Visbeck (Eds.), *The North Atlantic Oscillation: Climatic significance and environmental impact, Geophysical Monograph Series* (Vol. 134, pp. 211–234). Washington, DC: American Geophysical Union.

Giorgetta, M. A., Jungclaus, J., Reick, C. H., Legutke, S., Bader, J., Böttinger, M., et al. (2013). Climate and carbon cycle changes from 1850 to 2100 in MPI-ESM simulations for the Coupled Model Intercomparison Project Phase 5. *Journal of Advances in Modeling Earth Systems*, *5*, 572–597. https://doi.org/10.1002/jame.20038

Hall, R. J., Scaife, A. A., Hanna, E., Jones, J. M., & Erdélyi, R. (2017). Simple statistical probabilistic forecasts of the winter NAO. *Weather and Forecasting*, *32*(4), 1585–1601.

Ho, C. K., Hawkins, E., Shaffrey, L., Bröcker, J., Hermanson, L., Murphy, J. M., et al. (2013). Examining reliability of seasonal to decadal sea surface temperature forecasts: The role of ensemble dispersion. *Geophysical Research Letters*, *40*, 5770–5775. https://doi.org/10.1002/2013GL057630

Hurrell, J., Kushnir, Y., Ottersen, G., & Visbeck, M. (Eds.) (2003). *The North Atlantic Oscillation: Climatic significance and environmental impact, Geophysical Monograph Series* (Vol. 134). Washington, DC: American Geophysical Union.

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

Jerez, S., Trigo, R., Vicente-Serrano, S. M., Pozo-Vázquez, D., Lorente-Plazas, R., Lorenzo-Lacruz, J., et al. (2013). The impact of the North Atlantic Oscillation on renewable energy resources in southwestern Europe. *Journal of Applied Meteorology and Climatology*, *52*(10), 2204–2225.

Jung, T., Vitart, F., Ferranti, L., & Morcrette, J.-J. (2011). Origin and predictability of the extreme negative NAO winter of 2009/10. *Geophysical Research Letters*, *38*, L07701. https://doi.org/10.1029/2011GL046786

Kushnir, Y., Robinson, W., Bladé, I., Hall, N., Peng, S., & Sutton, R. (2002). Atmospheric GCM response to extratropical SST anomalies: Synthesis and evaluation*. *Journal of Climate*, *15*(16), 2233–2256.

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.

Liptak, J., & Strong, C. (2014). The winter atmospheric response to sea ice anomalies in the Barents Sea. *Journal of Climate*, *27*(2), 914–924. https://doi.org/10.1175/JCLI-D-13-00186.1

López-Moreno, J. I., Beguería, S., Vicente-Serrano, S. M., & García-Ruiz, J. M. (2007). Influence of the North Atlantic Oscillation on water resources in central Iberia: Precipitation, streamflow anomalies, and reservoir management strategies. *Water Resources Research*, *43*, W09411.

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.

Marshall, J., Kushnir, Y., Battisti, D., Chang, P., Czaja, A., Dickson, R., et al. (2001). North Atlantic climate variability: Phenomena, impacts and mechanisms. *International Journal of Climatology*, *21*(15), 1863–1898.

Müller, W., Appenzeller, C., & Schär, C. (2005). Probabilistic seasonal prediction of the winter North Atlantic Oscillation and its impact on near surface temperature. *Climate Dynamics*, *24*(2–3), 213–226.

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*, 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*, 183–188. https://doi.org/10.1029/2012GL054083

Pinto, J. G., Karremann, M. K., Born, K., Della-Marta, P. M., & Klawa, M. (2012). Loss potentials associated with European windstorms under future climate conditions. *Climate Research*, *54*(1), 1–20.

Robertson, A. W., Mechoso, C. R., & Kim, Y.-J. (2000). The influence of Atlantic sea surface temperature anomalies on the North Atlantic Oscillation. *Journal of Climate*, *13*(1), 122–138.

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., Karpechko, A. Y., Baldwin, M., Brookshaw, A., Butler, A., Eade, R., et al. (2016). Seasonal winter forecasts and the stratosphere. *Atmospheric Science Letters*, *17*(1), 51–56.

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

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

Smith, D. M., Scaife, A. A., Eade, R., & Knight, J. R. (2014). Seasonal to decadal prediction of the winter North Atlantic Oscillation: Emerging capability and future prospects. *Quarterly Journal of the Royal Meteorological Society*, *142*(695), 611–617.

Stockdale, T. N., Molteni, F., & Ferranti, L. (2015). Atmospheric initial conditions and the predictability of the Arctic Oscillation. *Geophysical Research Letters*, *42*, 1173–1179. https://doi.org/10.1002/2014GL062681

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.

Sun, L., Deser, C., & Tomas, R. A. (2015). Mechanisms of stratospheric and tropospheric circulation response to projected Arctic Sea Ice Loss. *J Climate*, *28*, 7824–7845.

Thompson, D., Lee, S., & Baldwin, M. (2003). Atmospheric processes governing the Northern Hemisphere annular mode/North Atlantic Oscillation. *Geophysical Monograph*, *134*, 81–112.

Ulbrich, U., & Christoph, M. (1999). A shift of the NAO and increasing storm track activity over Europe due to anthropogenic greenhouse gas forcing. *Climate dynamics*, *15*(7), 551–559.

Wang, L., Ting, M., & Kushner, P. (2017). A robust empirical seasonal prediction of winter NAO and surface climate. *Scientific Reports*, *7*(1), 279.

Wang, W., Anderson, B. T., Kaufmann, R. K., & Myneni, R. B. (2004). The relation between the North Atlantic Oscillation and SSTs in the North Atlantic basin. *Journal of Climate*, *17*(24), 4752–4759.

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.