Seasonal Forecasts of Winter Temperature Improved by Higher-Order Modes of Mean Sea Level Pressure Variability in the North Atlantic Sector

Clementine Dalelane1, Mikhail Dobrynin1, and Kristina Fröhlich1

1Deutscher Wetterdienst (DWD), Offenbach, Germany

Abstract  The variability of the sea level pressure in the North Atlantic sector is the most important driver of weather and climate in Europe. The main mode of this variability, the North Atlantic Oscillation (NAO), explains up to 50\% of the total variance. Other modes, known as the Scandinavian index, East Atlantic, and East Atlantic/West Russian pattern, complement the variability of the sea level pressure, thereby influencing the European climate. It has been shown previously that a seasonal prediction system with enhanced winter NAO skill due to ensemble subsampling entails an improved prediction of the surface climate variables as well. Here, we show that a refined subselection procedure that accounts both for the NAO index and for the three additional modes of sea level pressure variability is able to further increase the prediction skill of wintertime mean sea level pressure, near-surface temperature, and precipitation across Europe.

Plain Language Summary  Atmospheric winter conditions in Europe are primarily controlled by the varying pressure field over the North Atlantic, influencing temperature and precipitation in Europe. Current seasonal forecasts of European winter climate, though highly desirable for society and economy, are as yet not fully reliable. There exist a number of autumn predictors, such as sea surface and stratospheric temperature, Eurasian snow depth, and Arctic sea ice, that impact on the upcoming pressure regimes in a predictable way. The present dynamical seasonal forecast systems respond still too weakly to these known seasonal predictors. But the relationship is reproduced quite well by means of statistics. In combination, statistical and dynamical forecasts have the potential to improve forecasts of the North Atlantic pressure conditions and thereby affected variables like temperature and precipitation in Europe considerably. We extend an existing hybrid seasonal forecast procedure by considering more modes of the Atlantic pressure regimes than just the North Atlantic Oscillation. In this way, we are able to improve the forecasts for temperature and precipitation over wider regions in Europe.

1. Introduction

Seasonal prediction is a field of active research with several meteorological institutions worldwide issuing such seasonal forecasts to support environmental and economic decisions of a wide range of user groups. To date, the greatest success of such dynamical ensemble forecast systems is the prediction of El Niño Southern Oscillation (ENSO) several months ahead, which is the most important mode of interannual variability of the global climate influencing atmospheric phenomena around the world. In general, the skill of seasonal forecasts is satisfactory in the tropics, whereas prediction of northern midlatitude seasonal climate remains challenging, as recently evaluated by Baker, Shaffrey, Sutton, et al. (2018). They show that the anomaly correlation coefficient (ACC) used to measure the prediction skill of mean sea level pressure (SLP) in a multi-model ensemble is low and not significant over most of the North Atlantic-European sector in most of the analyzed models.

Cohen et al. (2019) argue that new statistical techniques can increase the accuracy of seasonal forecasts and advocate the development of hybrid dynamical-statistical forecasts to produce more robust seasonal predictions. Hybrid forecasts based on circulation specification were presented, for example, by Baker, Shaffrey, and Scaife (2018) and Dobrynin et al. (2018).

In boreal winter, European weather and climate is dominated by the zonal propagation of planetary and synoptic-scale waves. This large-scale circulation is an extremely high-dimensional phenomenon in real
space. The technique of principal component analysis (PCA), applied to the evolving SLP field, is one way to describe the states of this phenomenon in a sparse manner. The first principal component (PC) of SLP corresponds closely to the North Atlantic Oscillation (NAO) index, the importance of which for wintertime temperature, wind, and precipitation anomalies in the North Atlantic-European sector has been known for long time (Hurrell, 1995; Hurrell et al., 2003; Thompson et al., 2003). However, despite its importance, it would be misleading to consider the NAO in isolation. Although PCs are orthogonal by construction, the components are interwoven nonlinearly, and every PC represents just one aspect of the whole circulation.

We therefore extend our notion of SLP variability considering three further modes of variability (second, third, and fourth PC) in addition to the NAO index. These modes, hence called circulation indices, correspond to the Scandinavian index (SCAN), the East Atlantic/West Russian (EA/WR), and the East Atlantic (EA) pattern (although the denomination differs between authors, Barnston & Livezey, 1987). Together these indices explain about 80% of SLP variability. We set aside the inclusion of even more circulation indices, as their identification in short time series is complicated by stochastic noise.

Comas-Bru and McDermott (2014) show that higher-order circulation indices modulate the relation between NAO and European climate by shifting the NAO dipole in the South-West/North-East direction or rotating it in a clockwise/anticlockwise movement, impacting differently on temperature and precipitation. Moreover, Vihma et al. (2018) explore the effects of large-scale atmospheric patterns besides NAO on European winter temperatures. Among other things, they show that cold spells over Europe are associated with a combination of negative NAO and positive SCAN.

Compared to ensemble mean prediction, Dobrynin et al. (2018) reported significant improvements in the seasonal prediction of surface temperature (TAS) and precipitation (PR) over a large area mostly in northern Eurasia: On the basis of an accurate statistical prediction of the NAO index, “good” dynamical forecast members are selected from the forecast ensemble. But as the NAO index explains no more than 50% of the SLP variance, even a perfect prediction of the winter NAO will not improve the seasonal prediction of temperature and precipitation beyond certain limits (Dobrynin et al., 2018). The objective of the present paper is to explore possible improvements facilitated by the specification of all four leading circulation indices in the Euro-Atlantic sector (NAO, SCAN, EA/WR, and EA).

To produce the mentioned accurate prediction of the NAO index, Dobrynin et al. (2018) developed a statistical estimator of the mean winter NAO index with a correlation of around 0.8 by taking into account autumn states of slowly varying boundary conditions of the ocean and atmosphere: Arctic sea ice thickness (SIT), sea surface temperature (SST), snow depth (SND) in Eurasia, and stratospheric temperature in 100 hPa; see also Hall et al. (2017) and Wang et al. (2017). Similarly, Iglesias et al. (2014) and Oss et al. (2018) predict the seasonal evolution of the EA pattern based on SST. Rust et al. (2015) identify a linear relationship between temperature in Europe and several circulation indices, which allows the isochronic prediction of temperature anomalies given those indices.

We are going to broaden the approach of Dobrynin et al. (2018) by including the above mentioned predictor fields in four multiple linear regressions to predict each of the four considered circulation indices. These fields have been corroborated as physically meaningful drivers of the Euro-Atlantic SLP variability independently using causal network methods by Kretschmer et al. (2016). We show that an ensemble selection technique similar to Dobrynin et al. (2018), applied to the hindcasts of the operational seasonal forecast model of the German Meteorological Service GCFS2.0, accounting for four circulation indices, leads to substantial improvement in the forecasts of SLP, TAS, and PR in the North Atlantic-European sector.

2. Data

We use data from the operational German Climate Forecast System, version 2 (GCFS2.0). GCFS2.0 is based on the MPI-ESM-HR (Mauritsen et al., 2018; Müller et al., 2018) with a horizontal resolution corresponding to 0.9° in the atmosphere and an ocean resolution of nominally 0.4°. In cooperation, Universität Hamburg (UHH), Max-Planck-Institute for Meteorology (MPI), and Deutscher Wetterdienst (DWD) have developed the seasonal prediction system GCFS, issuing operational seasonal forecasts once a month since 2016, starting on the first day of each month covering the upcoming 6 months. The first month is discarded as spin up.
The forecasts (both retrospective and real time) are initialized with the state of the climate system inferred from the assimilation run, which starts in 1980, using a continuous full-field nudging for ocean, sea ice, and atmosphere (Baehr et al., 2015). ERA-Interim vorticity, divergence, temperature, and SLP are used for the atmosphere; ORAS5 sea ice, temperature, and salinity are used for the ocean and sea ice model. An ensemble consisting of 30 hindcast members is created by varying initial conditions in the ocean and a changing parameter in the atmosphere.

For each of the 12 forecasts per year, a hindcast data set (retrospective forecasts) consisting of 30 members per start date is provided to derive the model climate, error metrics, and skill scores. In GCFS2.0, hindcast data cover the monthly starting dates from 1990 through 2017. The present study concentrates on hindcasts starting in November, which is when the upcoming boreal winter (December, January, February [DJF]) is routinely forecasted.

As a complement to the assimilation run of the GCFS2.0 seasonal forecast system, we will also need the assimilation of the decadal prediction system developed in the MiKlip project (Pohlmann et al., 2019) because it extends 21 years farther into the past (autumn 1958–present). This system facilitates a slightly different initialization method compared to the seasonal prediction system. The atmosphere is nudged with ERA40 reanalysis full-field data until 1979 and ERA-Interim reanalysis data from 1980 onward. The ocean is nudged with ORAS4 reanalysis anomalies during the whole duration of the simulation. The sea ice is nudged with NSIDC sea ice concentration anomalies from 1980 until present.

3. Methods

We adopt the idea of Dobrynin et al. (2018) to predict the NAO index of the upcoming winter (DJF) based on four predictors, autumn SIT, SND, SST, and stratospheric temperature at 100 hPa (TA100), from the assimilation run of GCFS2.0. The actual values of the predictors are calculated as an area-weighted mean of monthly grid cell values, taking into account only grid cells that show a significant correlation to the NAO index. We construct a multiple linear regression estimator for the NAO index that takes all four predictors into account simultaneously.

Multiple linear regression estimators for the three other circulation indices (SCAN, EA/WR, and EA) are constructed analogously to the NAO prediction. The literature on driving conditions influencing these indices is rather sparse. However, as already mentioned above, the large-scale circulation in the North Atlantic-European sector is a complex interaction of many factors. Boundary fields like the chosen predictors do not impact exclusively on one or another circulation index, but the whole system, exerting a greater or lesser influence on all components. For these reasons, we use the same predictors for SCAN, EW/WR, and EA as are proposed for the NAO in Dobrynin et al. (2018).

After having predicted the four circulation indices statistically, in the second step we select the “best” members from the dynamical hindcast ensemble. "Best" is defined here in terms of the Euclidean distance between a dynamical hindcast member’s vector of indices (see section 4.1) and our statistically predicted index vector. The “best” members are selected to build a subensemble. The new seasonal hindcasts for SLP, TAS, PR, and so forth are based only on the subensemble instead of the complete dynamical hindcast ensemble.

3.1. Predictors and Regression

The dynamical seasonal hindcasts for DJF are initialized on 1 November. We therefore take the October monthly mean values of SST, SND, and TA100 from the assimilation run as predictors, as this is the latest information known when the integration starts. For SIT, we use the September monthly mean, because it reflects the annual minimum sea ice extension (Dobrynin et al., 2018).

The correlation between the predictor values and the circulation indices, as well from the assimilation run, is calculated on grid cell basis. Grid cells, which show a significant positive correlation, are combined to an area-weighted sum, as well as grid cells with significant negative correlation. Consequently, each predictor can contribute two exogenous variables to the multi-linear regression. Before entering the regression, the area-weighted sums are centered and detrended.

The performance of the proposed estimation procedure is evaluated in section 4.2 in the so-called backtesting mode (see supporting information), a realistic cross-validation setting, where the prediction at a given
time is based exclusively on information from its past. In the backtesting mode, we find a high annual vari-
ation of the regions, where grid cells with significant correlations between SST and the circulation indices are detected. As a remedy, we replace the assimilation time series of SST and SLP (for the calculation of circulation indices) from GCFS2.0 by the respective time series from the latest MiKlip assimilation, which start as early as 1958 (for a discussion see the supporting information). The MiKlip assimilation is utilized exclusively to detect the significant grid cells. For the calculation of the predictor values, we return to the GCFS2.0 assimilation time series.

An ordinary least squares algorithm is performed to estimate the regression coefficients. In order to avoid overfitting, the combination of predictor variables is selected so as to minimize the mean squared error (MSE) of the predicted index in the backtesting mode (see supporting information), using a maximum of four predictors.

3.2. Subselection

The subselection of members from the dynamical seasonal hindcast ensemble is based on the statistically predicted circulation indices. To compute the circulation indices realized by each ensemble member, we use the PCs calculated from the assimilation SLP fields. It is very probable that, when applying a PCA to the union of all dynamical seasonal hindcast ensembles, the PCs will not coincide with the GCFS2.0 assimilation. However, for a meaningful comparison between statistically predicted circulation index and its counterpart in a dynamical forecast run, the indices have to refer to the same PC pattern. We therefore project the dynamical forecast members onto the patterns from the assimilation.

We can now fix the number of circulation indices to be included in the subselection (only one index [NAO], or more than one up to four). The Euclidean distance is calculated between the index vectors of the dynamical hindcast ensemble members and the vector of statistically predicted indices for a given winter. The Euclidean distance is weighted by the eigenvalues of the PCs to emphasize the importance of the respective circulation index. Subsequently, the members with the smallest distance to the statistical prediction are selected to build the subensemble. This can be a fixed number of members for every year (like in section 4.3) or a varying number depending on elaborate criteria (see section 3.3 and supporting information). We reiterate the post processing for this subensemble, like generating the ensemble mean, tercile probabilities, and error measures for variables of interest like TAS and PR as we have done before on the complete ensemble.

3.3. Selection by Machine Learning Procedures

Further refinements of the subselection that make use of various machine learning procedures are conceivable. We would like to name but a few details and results of which are described in the supporting information. A most obvious refinement would be the weighted mean of the hindcast members according to their proximity to the statistically predicted circulation indices. More sophisticated, a clustering of the vectors of circulation indices would allow for nonlinear interdependencies between the four circulation indices, apart from linear orthogonality imposed by PCA. To improve the achieved stratification of the clusters with respect to TAS (or any other selected parameter), a semi-supervised clustering algorithm or a discriminant analysis could be applied.

4. Results

4.1. Circulation Indices

In this section, we examine our assumption that the seasonal hindcast skill benefits from the inclusion of further circulation indices in the ensemble subselection procedure of Dobrynin et al. (2018). To this end, we repeat their perfect-NAO prediction experiment and compare it to an analog perfect-circulation indices experiment.

The assimilation run from the current seasonal forecast model GCFS2.0 starts in 1980; hindcasts were provided for start dates in 1990–2017. We consider seasonal means for winter (DJF), such that our time series starts in winter 1980/1981 and runs through winter 2017/2018, a total of 38 time steps. In order to calculate the winter circulation indices, singular value decomposition is applied to the area-weighted non-standardized anomalies of seasonal SLP over the North Atlantic-European sector (20–85°N and 90°W–60°E) (Figure 1). Note that a subsequent standardization of the indices does not affect our computations.
Likewise, the ensemble members of the seasonal hindcast ensembles are projected onto the same PCs extracted from the assimilation to calculate the respective circulation indices.

Now, we select those members from the hindcast ensembles, which reproduce the true circulation indices from the assimilation run most closely—first only for the NAO index, after that for NAO, SCAN, EA/WR, and EA indices. The forecast skill of the full and of the two subensembles is plotted in Figure S1 in the supporting information.

The improvement in ACCs for winter SLP in the Euro-Atlantic sector when selecting for all four indices, taken over time at each point separately, is strong. In particular, the zonal band of low SLP predictability between 50°N and 60°N, which stands out in the perfect-NAO-only ensemble, is completely recovered in the four-indices ensemble. The ACCs for TAS and PR show considerable improvements, too (Figure S1).
We therefore conclude that the subselection for more than one circulation index is worthwhile—as long as we are able to construct reliable predictors for them.

4.2. Regression

We evaluate the whole estimation and subselection procedure in the backtesting mode, as this is the most realistic setting possible in view of prospective operationalization (see supporting information) and the most challenging at the same time. In the following, we will evaluate our predictions and the resulting hindcast skill against the assimilation run of GCFS2.0. We choose the assimilation run over the obvious alternative ERA-Interim for the following reasons: The GCFS assimilation and ERA-Interim are both model assimilations, but the GCFS assimilation was produced with the same model as the hindcasts as opposed to ERA-Interim. The hindcasts will therefore be more similar to the GCFS assimilation than to the ERA-Interim data, regardless of the skill of the hindcasts. Here, we aim to evaluate the relative differences in skill generated by the subselection, so for the moment we set aside model differences between GCFS and ERA-Interim.

The selected predictors and respective correlations between assimilated and statistically predicted circulation indices (as described in section 3.1) are listed in Table 1, along with the correlation of the full ensemble mean indices for comparison. The choice of the evaluation period 2003/2004–2017/2018 is discussed in detail in the supporting information. We note that all statistical estimators perform quite well w.r.t. MSE and correlation (see Figure 1 and Table 1).

4.3. Subselection

To evaluate whether the subselection leads to an improvement in the seasonal hindcast, we first analyze the ACCs between the ensemble mean of the two hindcasts (subensemble vs. complete ensemble) and the GCFS2.0 assimilation values:

\[
ACC = \frac{\sum_{t=1}^{T} (f_t - \bar{f})(o_t - \bar{o})}{\sqrt{\sum_{t=1}^{T} (f_t - \bar{f})^2} \sqrt{\sum_{t=1}^{T} (o_t - \bar{o})^2}}
\]

where \(f_1 \ldots f_T\) are the ensemble mean hindcasts and \(o_1 \ldots o_T\) the values from the assimilation, and \(\bar{f}\) and \(\bar{o}\) are the respective averages over the same time period. As suggested by the perfect-prediction experiment in section 4.1, the ACC for the hindcast fields for SLP, TAS, and PR increases in most regions with each additional circulation index included. Varying the number of selected hindcasts between 4 and 20, we obtained the highest increases in ACC for four-indices subensembles of eight members (see Figure 2 for comparison to eight-member NAO-only subensemble and full ensemble mean).

The area-weighted average ACC over Europe (target region: 10°W–30°E and 35–65°N) for SLP is calculated for the full/NAO-only/four-indices (sub)ensembles: 0.24/0.63/0.73. Analogous mean ACCs for TAS amount to 0.41/0.63/0.73 and for PR to 0.22/0.33/0.41.

Besides the evaluation of the ensemble mean, we consider the \(L_2\) ranked probability score (RPS) for terciles to evaluate the subsampled hindcasts probabilistically (Figure S2 in the supporting information):

\[
RPS = \frac{1}{T} \sum_{t=1}^{T} \sum_{i=1}^{3} \left( \frac{1}{M} \sum_{m=1}^{M} I(f^m_t < q_i(f)) - I(o_t < q_i(o)) \right)^2
\]
where this time \( (f_1^{T}, \ldots, f_M^{T}) \) are the \( T \) hindcast ensembles, comprising \( M \) members each, with their respective terciles \( q_1(f), q_2(f), q_3(f) \), and \( o_1 \) \( \ldots \) \( o_T \) are again the assimilation values with terciles \( q_1(o), q_2(o), q_3(o) \), and \( I \) is the indicator function. The area-weighted mean RPS in the target region for the full/NAO-only/four-indices (sub)ensembles are as follows: SLP, 0.53/0.46/0.43; TAS, 0.49/0.43/0.41; and PR, 0.60/0.45/0.45. We note that the improvement caused by subselection in general is high; the difference between NAO-only and all indices is less pronounced. This could in part be due to the small ensemble size of the subensembles. But to increase the size of the subensembles, a considerably higher number of full ensemble members would be necessary to select from.

4.4. Spatial Evaluation of Individual Hindcasts
To further explore the improvement in our temperature hindcasts obtained by subselecting for circulation indices, we compare the individual hindcasts for winter seasons 2008/2009, 2009/2010, and 2015/2016 with the respective GCFS2.0 assimilation in Figure 3. Winters 2009/2010 and 2015/2016 represent distinctive atmospheric conditions showing unusual values in their circulation indices (2009/2010: very low NAO and low EA, 2015/2016: high NAO and very high SCAN), whereas winter 2008/2009 shows average values in

![Figure 2. Anomaly correlation coefficients between hindcast ensemble means and assimilation for winters 2003/2004–2017/2018. First row: complete ensemble, second row: subselection for NAO, third row: subselection for NAO, SCAN, EA/WR, and EA. Left column: SLP, center column: TAS, right column: PR. Regions, where the ACC is significantly positive to the 95% level (critical value 0.441), are contoured in dark red.](image-url)
all four indices. We find that the assimilation values are poorly reflected in the full ensemble mean indices, except for EA/WR in 2008/2009, but they are estimated well by our statistical procedure (Figure 1).

The assimilation temperature anomalies in the three selected winters are quite pronounced. In contrast, the hindcasts anomalies for 2009/2010 and 2015/2016 from the full ensemble appear pale (we concentrate on 10°W–30°E and 35–65°N, a region that constitutes a natural target for the German Meteorological Service; see Figure 3). For 2008/2009 the full ensemble mean hindcast fails completely to capture the generalized cold anomaly. After subselection, in 2009/2010 the spatial pattern of anomalies is very well reproduced, and also, the warm hindcast anomalies for 2015/2016 are increased and much closer to the analyzed ones. For winter 2008/2009, the subselected forecast shows a cold anomaly reversing the sign of the full ensemble hindcast. However, all subselected anomalies are still weakly pronounced in amplitude compared to the assimilation run (Figure 3).

A quantification of the goodness of fit of all individual full and subselected hindcasts by means of the Structural Similarity Index (SSIM, proposed in Wang et al., 2004) is discussed in the supporting information.
5. Summary and Discussion

We have constructed an ensemble selection procedure based on the statistical prediction of the four leading PCs of SLP in the North Atlantic-European sector, which leads to a substantial improvement of seasonal hindcast skill for winter (DJF) hindcasts of SLP, TAS, and PR compared to the full ensemble mean hindcasts. This method is evaluated in the backtesting mode, with average anomaly correlation over Europe for SLP, TAS, and PR of 0.73, 0.58, and 0.41, respectively. The statistical predictions rely solely on the autumn states of four drivers of atmospheric circulation, which are known at the time the dynamical model integration starts. The procedure is therefore fully applicable to operational forecasts. It furthermore offers plenty of refinement possibilities.

The presented subsampling method is tailored to improve the seasonal hindcasts in winter over Europe only. Skill over other regions and seasons is thus possibly degraded. Nonetheless, an analog approach aiming at other regions and seasons is conceivable.

We have to assume that the relationships between the predictors, the circulation indices, and the seasonal climate that we exploit in our subselection might be subject to climate variability as well as climate change. In the long run, strategies accounting for such non-stationarity have to be developed.

Data Availability Statement

Hindcast data used for this study are available online (under https://cds.climate.copernicus.eu/cdsapp#!/dataset/seasonal-monthly-single-levels?tab=form).

Acknowledgments

The authors would like to thank Klaus Pankatz for providing the MiKlip assimilation data and the corresponding description and Johanna Baehr and two anonymous reviewers for valuable comments on the manuscript.

References

Baker, L. H., Shaffrey, L. C., & Scaife, A. A. (2018). Improved seasonal prediction of UK regional precipitation using atmospheric circulation. *International Journal of Climatology*, 38, e437–e453.
Baker, L. H., Shaffrey, L. C., Sutton, R. T., Weisheimer, A., & Scaife, A. A. (2018). An intercomparison of skill and overconfidence/underconfidence of the wintertime North Atlantic Oscillation in multimodel seasonal forecasts. *Geophysical Research Letters*, 45, 7808–7817. https://doi.org/10.1029/2018GL078838
Barnston, A. G., & Livezey, R. E. (1987). Classification, seasonality and persistence of low-frequency atmospheric circulation patterns. *Monthly Weather Review*, 115(6), 1083–1126.
Cohen, J., Coumou, D., Hwang, J., Mackey, L., Orenstein, P., Totz, S., & Tziperman, E. (2019). S2S reboots: An argument for greater inclusion of machine learning in subseasonal to seasonal forecasts. *Wiley Interdisciplinary Reviews: Climate Change*, 10(2), e00567.
Comas-Bru, L., & McDermott, F. (2014). Impacts of the EA and SCA patterns on the European twentieth century NAO-winter climate relationship. *Quarterly Journal of the Royal Meteorological Society*, 140(679), 354–363.
Dobrynin, M., Domeisen, D. I. V., Müller, W. A., Pohlmann, H., & Baehr, J. (2018). Improved teleconnection-based dynamical seasonal predictions of boreal winter. *Geophysical Research Letters*, 45, 3605–3614. https://doi.org/10.1002/2018GL077209
Hall, R. J., Scaife, A. A., Hanna, E., Jones, J. M., & Ebdly, R. (2017). Simple statistical probabilistic forecasts of the winter NAO. *Weather and Forecasting*, 32(4), 1585–1601.
Hurrell, J. W. (1995). Decadal trends in the North Atlantic Oscillation: Regional temperatures and precipitation. *Science*, 269(5224), 676–679.
Hurrell, J., Kushnir, Y., Ottersen, G., & Visbeck, M. E. (2003). The North Atlantic Oscillation: Climatic significance and environmental impact. *Geophysical Monograph Series* (Vol. 134). Washington, DC: American Geophysical Union.
Iglesias, I., Lorenzo, M. N., & Taboada, I. (2014). Seasonal predictability of the East Atlantic pattern from sea surface temperatures. *PLoS ONE*, 9(1), e86439.
Kretschmer, M., Coumou, D., Donges, J. F., & Runde, J. (2016). Using causal effect networks to analyze different arctic drivers of midlatitude winter circulation. *Journal of Climate*, 29(11), 4069–4081.
Mauritzen, T., Bader, J., Becker, T., Behrens, J., Bittner, M., Brokopf, R., et al. (2018). Developments in the MPI-M Earth System Model version 1.2 (MPI-ESM1.2) and its response to increasing CO2. *Journal of Advances in Modeling Earth Systems*, 11, 998–1038. https://doi.org/10.1002/2018MS001400
Müller, W. A., Jungclaus, J. H., Mauritzen, T., Baehr, J., Bittner, M., Budich, R., et al. (2018). A higher-resolution version of the Max Planck Institute Earth System Model (MPI-ESM1.2-HR). *Journal of Advances in Modeling Earth Systems*, 10, 1383–1413. https://doi.org/10.1029/2017MS001217
Oss, A., Sutton, R., Shaffrey, L., & Dong, B. (2018). Observational evidence of European summer weather patterns predictable from spring. *Proceedings of the National Academy of Sciences of the United States of America*, 115(1), 59–63.
Pohlmann, H., Müller, W. A., Bittner, M., Hettrich, S., Modali, K., Pankatz, K., & Marotzke, I. (2019). Realistic quasi-biennial oscillation variability in historical and decadal hindcast simulations using CMIP6 forcing. *Geophysical Research Letters*, 46, 14,118–14,125. https://doi.org/10.1029/2019GL084878
Rust, H. W., Richling, A., Bissoli, P., & Ulbrich, U. (2015). Linking teleconnection patterns to European temperature—A multiple linear regression model. *Meteorologische Zeitschrift*, 24(4), 411–423.
Thompson, D., Lee, S., & Baldwin, M. (2003). *Atmospheric processes governing the Northern Hemisphere annular mode/North Atlantic Oscillation*. Geophysical Monograph (Vol. 134, pp. 81–112). Washington, DC: American Geophysical Union.
Vihma, T., Graversen, R., Chen, L., Handorf, D., Skific, N., Francis, J. A., et al. (2018). Effects of the tropospheric large-scale circulation on European winter temperatures during the period of amplified Arctic warming. International Journal of Climatology, 40(1), 509–529.

Wang, Z., Bovik, A. C., Sheikh, H. R., & Simoncelli, E. P. (2004). Image quality assessment: From error visibility to structural similarity. IEEE Transactions on Image Processing, 13(4), 600–612.

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