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

Clementine Dalelane, Mikhail Dobrynin, Kristina Fröhlich

Deutscher Wetterdienst (DWD)

Contents of this file

Backtesting

Backtesting is a causal type of cross-validation, in which the forecasting procedure is applied only to data prior to the forecast time. In order to reforecast a circulation index at time t, we perform the whole procedure of selection of significant predictors and least squares estimation on the data for $s = 1 \ldots t-1$.

We note that the Principal Component Analysis is not included in the backtesting mode. That is because PCA is essentially a transformation of coordinates with the objective to aggregate as much variance as possible into a small number of directions. A unique "true" set of principal components does not exist. Furthermore the estimation procedure is not affected by small changes in the principal components as long as the correlation between the predictors and the indices is preserved. So we use fixed circulation indices calculated from the whole time series of SLP fields to investigate the properties of the proposed subselection.

Structural Similarity Index

To summarize the performance of the hindcasts, we introduce the Structural Similarity Index (SSIM), a concept developed in the context of image processing (Wang et al., 2004). The SSIM is used to measure the similarity between two images, in our case the similarity between the hindcast and assimilation fields. It combines three important aspects of spatial goodness-of-fit, which in climatological forecast validation are usually measured and assessed separately: mean, variance and correlation.

Let x and y be the two fields to compare, c_1 and c_2 small constants. Then

$$SSIM(x,y) = \frac{(2\mu_x\mu_y + c_1)(2\sigma_{xy} + c_2)}{(\mu_x^2 + \mu_y^2 + c_1)(\sigma_x^2 + \sigma_y^2 + c_2)}$$

where μ_x, μ_y are the respective spatial means, σ_x^2, σ_y^2 the spatial variances and σ_{xy} is the spatial covariance.

SSIM satisfies the non-negativity, identity of indiscernibles, and symmetry properties. The resultant SSIM index is a decimal value between -1 and 1, and value 1 is only reached for two identical fields and therefore indicates perfect structural similarity. A value of 0 indicates no structural similarity.

As the SSIM includes terms of mean and variance, it is improved by linear bias-adjustment (rescaling), although this does not alter the ACC.

SSIM values of the full, subselected and rescaled-subselected ensemble hindcasts corrsponding to the winter seasons 2008/09, 2009/10 and 2015/16 for the variables TAS, SLP and PR, calculated over certain regions, are listed in Table S1. As the smoothness and spatial correlation of these climatological parameters are very different, we choose larger/smaller regions for the spatial average, which are nevertheless all centered over

Germany as it constitutes the natural target for the German Meteorological Service. For TAS this is the region between 10°W-30°E and 35°N-65°N, for SLP 50°W-47°E and 23°N-85°N, and for PR 6°-16.5°E and 46.75°N-56°N.

References

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

Selection by Machine Learning Procedures

A *weighted subselection* is realized by the application of a radial Epanechnikov-kernel to the (Eigenvalue-weighted) Euclidian distances between statistically predicted and dynamically hindcasted circulation-index vectors x_s and x_d. Let W be the diagonal matrix of Eigenvalues of the circulation indices obtained from the principal component analysis, then the Epanechnikov-kernel with bandwidth h is defined as

$$K_h(x_s, x_d) = \frac{3}{4h}\left(1 - \frac{\|x_s - x_d\|_{2W}^2}{h^2}\right), \text{ where } \|x_s - x_d\|_{2W}^2 = x_s^T W^{1/2} x_d$$

The weighted subensemble is realized by the weighted sum of all hindcast ensemble members. Best results where obtained with three circulation indices and a bandwidth of $h = 87$ (see Table S2).

The *clustering of circulation indices* allows for nonlinear interdependencies between the four circulation indices, apart from linear orthogonality imposed by PCA.

To obtain an unsupervised classification of the vectors of circulation indices, a K-means algorithm (with Egenvalue-weighted Eucliadian distance) is applied to the circultion-index vectors $x_a^1 \ldots x_a^T$ for all winters from the assimilation run. The algorithm is initialized by intermediate index values. Subsequently, the statistically predicted and the dynamically hindcasted vectors for a specified winter are assigned to their nearest cluster. The subensemble is then composed of those hindcast members that pertain to the cluster indicated by the statistical prediction. The resulting improvements w.r.t SSIM(TAS), achieved for three index vectors and five clusters, are listed in table S2.

Algorithm of clustered selection

1. The circulation-index vectors from assimilation $x_a^1 ... x_a^T$ are clustered, clusters $C_1 ... C_K$ are obtained

2. The statistically predicted index vector x_s for the specified winter is assigned to the nearest cluster C_k – this is the staistically predicted cluster

3. The hindcasted index vectors $x_d^1 ... x_d^{30}$ for the specified winter are assigned each to their nearest cluster

4. The subensemble is composed of those hindcast members that fall into the staistically predicted cluster C_k

To improve the stratification of the clusters w.r.t. some target variable (TAS in our case), in *semi-supervised clustering* the training sample is augmented by the values of the target variable assumed in the training sample $y_a^1 ... y_a^T$, such that $\tilde{x}_a^t = (x_a^t, y_a^t)$. The clustering procedure is otherwise identical to the unsupervised clustering. In the classification of a statistical prediction x_s^t, the target value is of course unknown and the assignment is based on the circulation indices only. In our case, we generate the target variables from the TAS field by principal component analysis. The resulting scores of the leading TAS PCs are introduced as target variables into the K-means algorithm, in addition to the circulation indices resulting from the PCA of the SLP fields. The hindcasts are processed analogously to the un-supervised clustering (see results for the best parameter combination [4 circulation indices, 2 TAS PCs, 4 clusters] in Table S2).

Linear distcriminant analysis is a supervised classification procedure that optimally seperates two or more classes of objects on the basis of observable variables. The classification of the training sample has to be known in advance. We define three classes w.r.t. the same two TAS PCs used above. The algorithm finds the linear partition in the space of circulation indices that best predicts the given classification. According to the statistically predicted circulation indices, a class is selected along with the corresponding hindcasts. Best discrimination results where obtained using all 4 circulation indices (Table S2).

All selection results in Table S2 have been generated in the Backtesting mode.

Table S1. SSIM of hindcasted to assimilation anomalies in selected winters of full/subselected/subselected+rescaled ensemble

	DJF 2008/09	DJF 2009/10	DJF 2015/16
TAS	0.01/0.24/0.30	0.27/0.75/0.82	0.42/0.62/0.70
SLP	-0.09/0.51/0.54	0.08/0.32/0.39	0.25/0.61/0.68
PR	0.01/0.25/0.24	0.30/0.35/0.37	0.18/0.54/0.69

Table S2. Average SSIM of hindcasted to assimilation anomalies during winters from 2003/04 through 2017/18

	full	Sub	SubR	SubWR	ClAnaR	sClAnaR	DisAnaR
TAS	0.12	0.29	0.31	0.36	0.34	0.33	0.38
SLP	0.11	0.34	0.37	0.45	0.40	0.33	0.41
PR	0.06	0.10	0.10	0.07	0.17	0.12	0.12

Selection types: full-full ensemble, Sub-subselection of 8 best members, SubR-rescaled subselection of 8 best members, SubWR-rescaled weighted subselection, ClAnaR-rescaled subselection according to unsupervised cluster analysis, sClAnaR-rescaled subselection according to semi-supervised cluster analysis, DisAnaR-rescaled subselection according to discriminant analysis

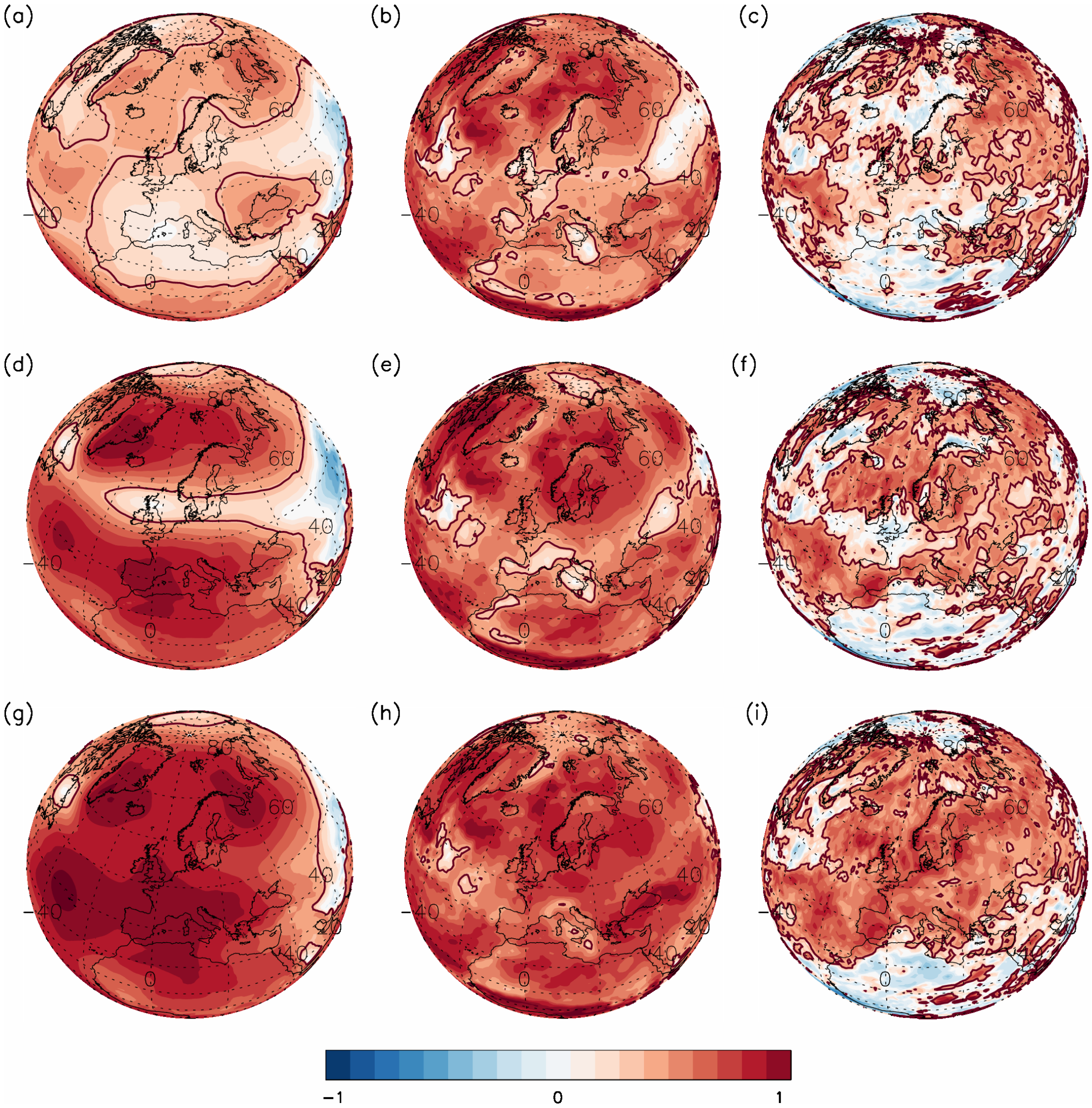

Figure S1. Anomaly correlation coefficients between ensemble means and assimilation. 1st row: complete ensemble, 2nd row: subselection for perfect NAO, 3rd row: subselection for perfect NAO, SCAN, EA/WR and EA. Left column: SLP, center column: TAS, right column: PR. Regions, where the ACC is significantly positive to the 95% level (critical value 0.271), are contoured in dark red.