

The figures in this document provide additional details for the analysis of seasonal re-forecast skill in experiments REF, SMM, S5D discussed in the paper.

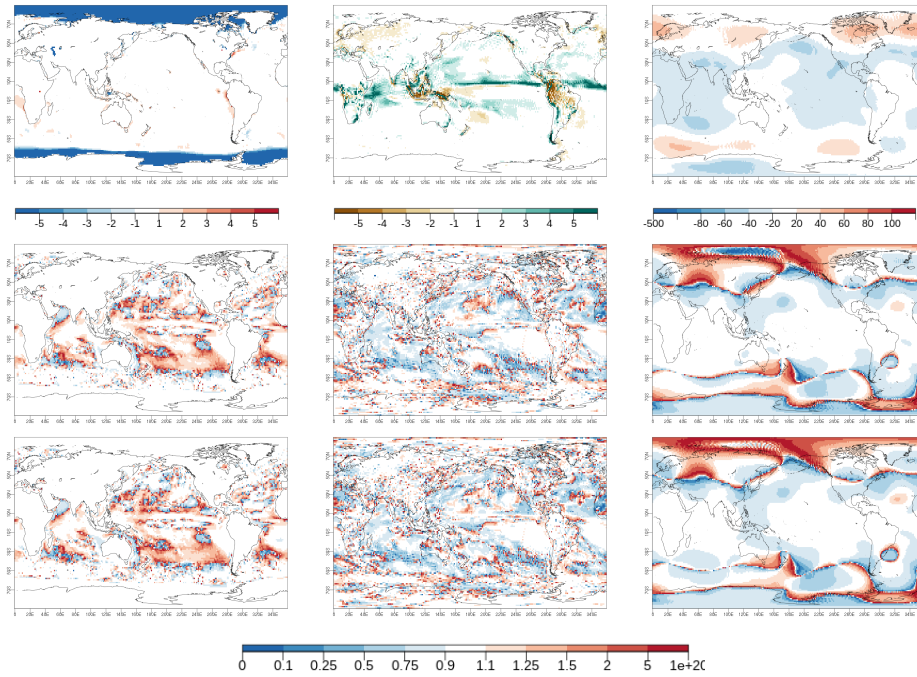


Figure S1: November bias (top row) for REF experiment SST, precipitation and Z500 (from left to right); corresponding relative absolute bias in experiments SMM and S5D (second and bottom rows, respectively). Bias is computed with respect to ERA-Interim for SST and GPCP for precipitation. Areas in blue indicate where bias is lower with respect to REF, whereas areas in shades of red show where bias is increased, regardless of the sign of the bias.

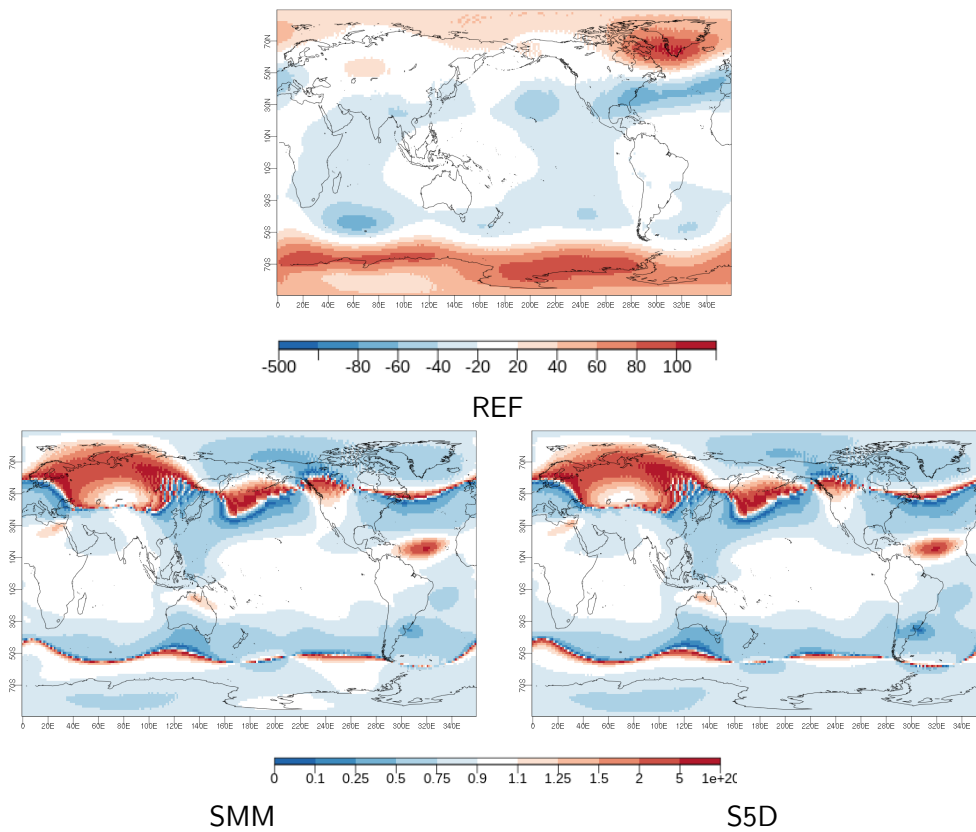


Figure S2: REF bias (top) for DJF 500 hPa geopotential height over the globe with respect to ERA-Interim data. SMM (bottom left) and S5D (bottom right) relative absolute bias with respect to REF.

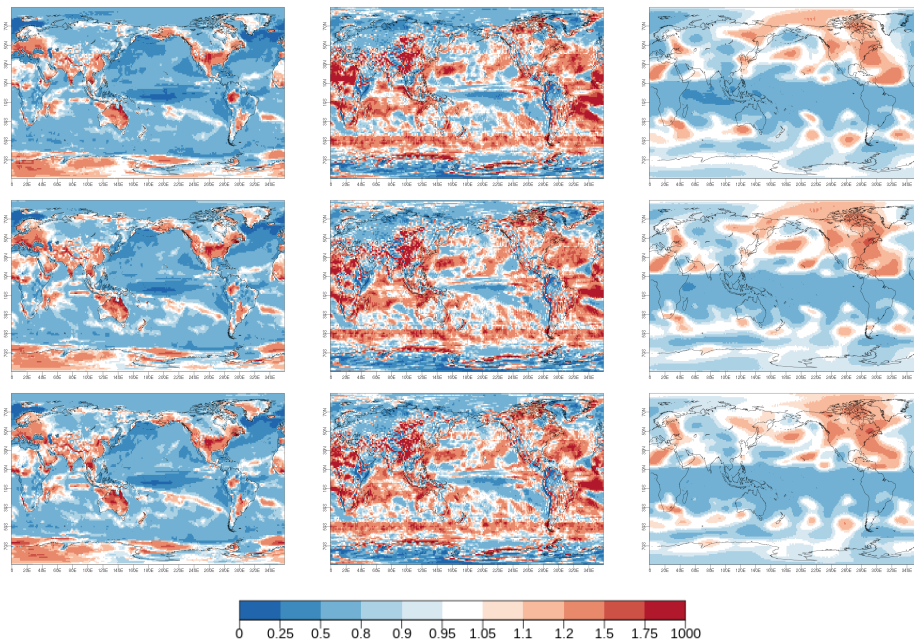


Figure S3: DJF spread-skill ratio for REF, SMM and S5D experiments (top to bottom rows, respectively) near-surface air temperature, precipitation and Z500 (from left to right). Spread is computed as the standard deviation around the ensemble mean, and skill is the ensemble mean root mean square error (RMSE). Areas in blue indicate where spread is lower than RMSE, whereas areas in shades of red show where spread is greater than RMSE.

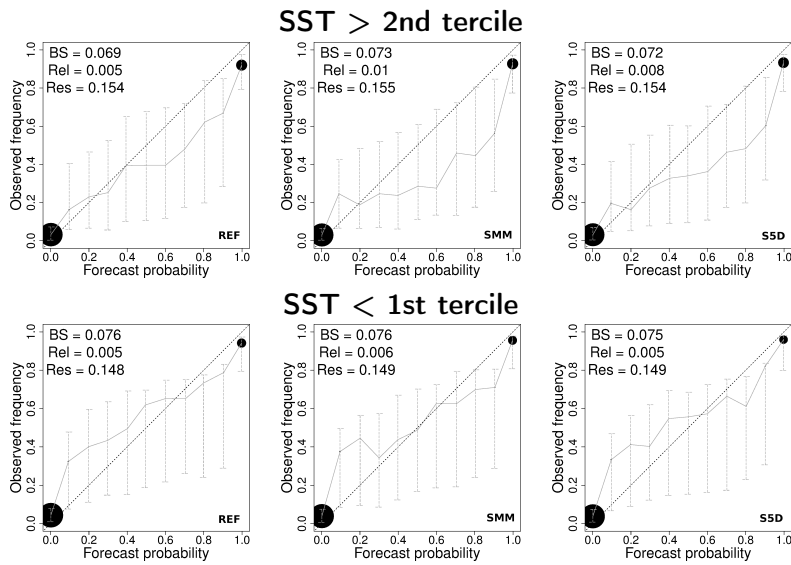


Figure S4: DJF Niño 3.4 reliability diagrams for SST exceeding the second tercile (top row) and remaining below the first tercile of reference climatology (bottom row), for experiments REF, SMM and S5D (from left to right). The Brier Score and reliability (Rel) and resolution (Res) decomposition is shown at the top left corner of each diagram. Dashed lines show the confidence intervals estimated by a bootstrapping method over the re-forecast period. The diameter of the circles is proportional to the size of the population bin for the different forecast probabilities.

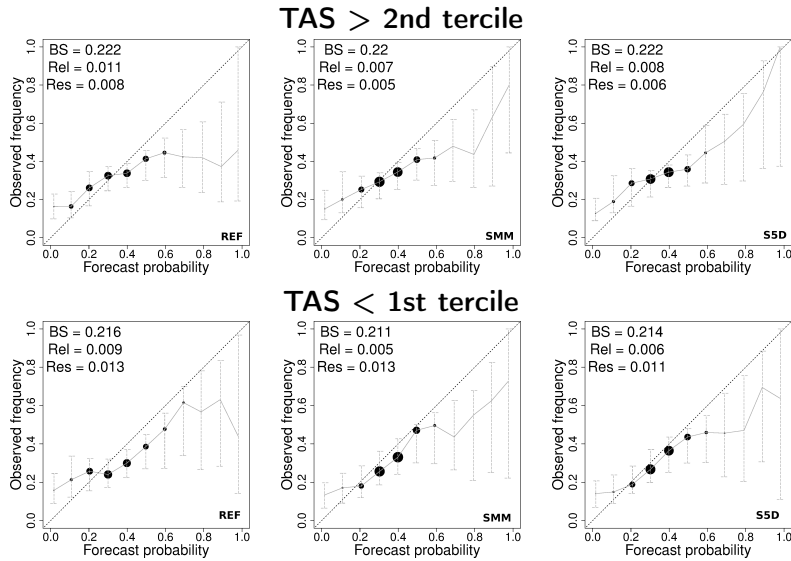


Figure S5: Same as figure S4 but for near-surface temperature over land in Europe verified against CRU data.

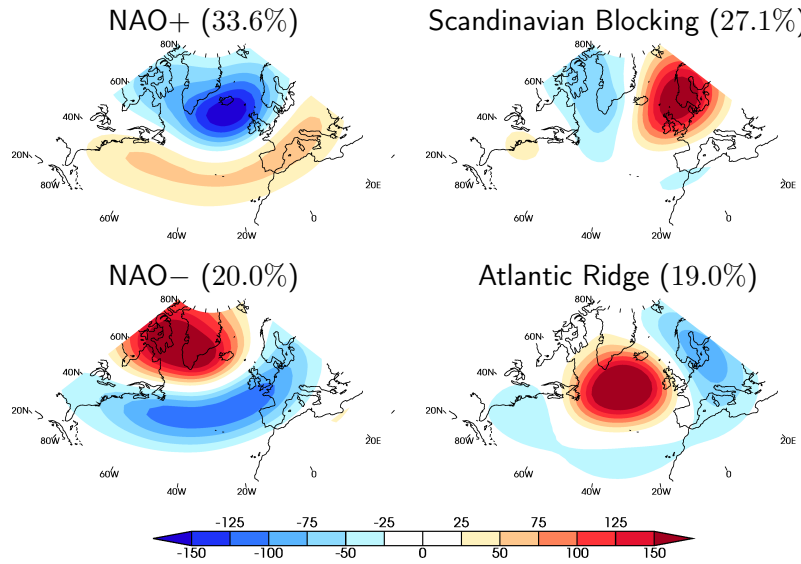


Figure S6: 500 hPa geopotential height anomalies for the four North Atlantic weather regimes computed with ERA-Interim reanalysis data for DJF 1979–2012

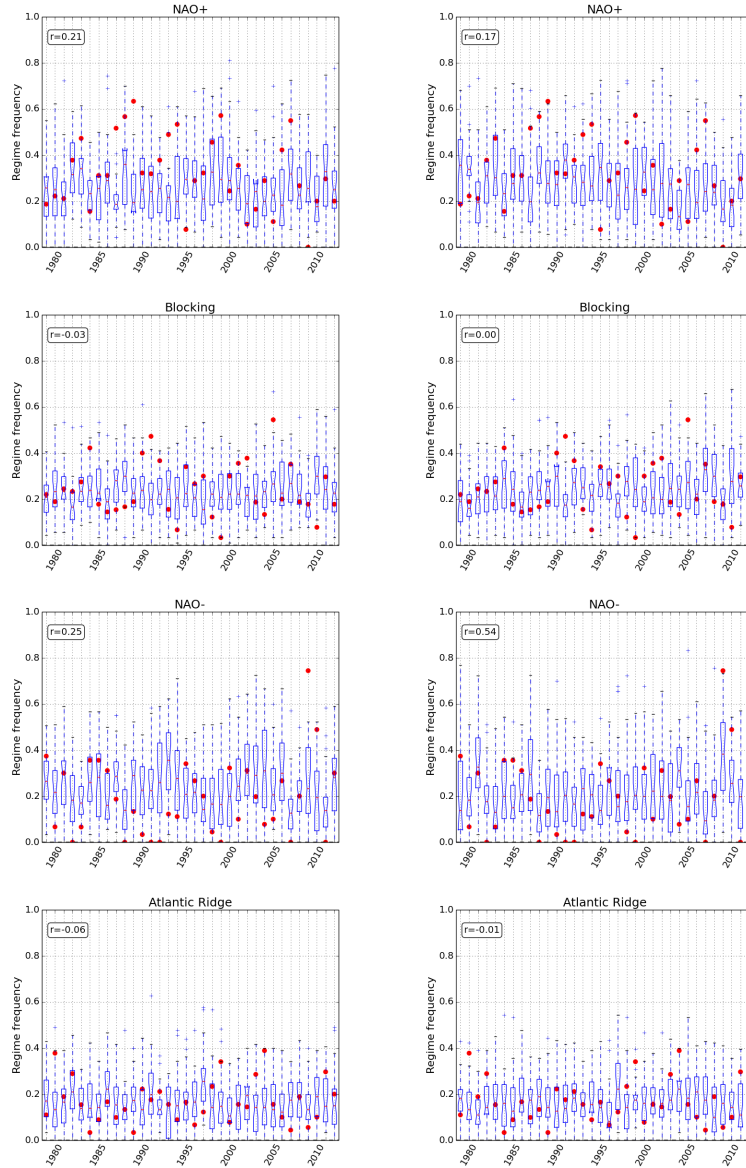


Figure S7: Boxplots for re-forecast ensemble DJF frequencies of each of the four North Atlantic weather regimes over the 1979–2012 time period in experiments REF (left) and S5D (right), and corresponding ERA-Interim frequency (red dots). Boxes show the interquartile range of the 30 ensemble members, while whiskers extend to a maximum of $\pm\sigma$. The correlation between ensemble mean and ERA-Interim frequencies is shown in the top left corner of each plot.