https://doi.org/10.5194/egusphere-2025-4463
Preprint. Discussion started: 24 September 2025

Reduction of uncertainty in near-term climate forecast by combining observations and decadal predictions

Rémy Bonnet[1], Julien Boé[1], Emilia Sanchez-Gomez[2,3], Christophe Cassou[4]

[1]CECI, Université de Toulouse, CERFACS/CNRS/IRD, Toulouse, France.
[2]Météo-France, Direction des Services Météorologiques, Toulouse, France.
[3]Météo-France, CNRS, Univ. Toulouse, CNRM, Toulouse, France.
[4]LMD-IPSL, CNRS, Ecole Normale Superieure, PSL Research University, Paris, France

Correspondence to: R. Bonnet (remy.bonnet@cerfacs.fr)

Abstract. The implementation of adaptation policies requires seamless relevant information about near-term climate evolution, which remains highly uncertain due to the strong influence of internal variability. The recent development of approaches to improve near-term climate information by selecting members from large ensembles—based on their agreement with either observed or predicted sea surface temperature patterns—have shown promising results across timescales from weeks to decades. Here, we propose a new method to provide climate forecasts over Europe by combining information from both observations and decadal predictions through a two-stage member selection from ensembles of climate simulations. Several predictors are tested as observational metrics based on their influence on the European climate variability at annual to decadal timescale. A retrospective evaluation over Europe demonstrates the added value of this method in reducing the spread of uncertainty stemming from both internal climate variability and model uncertainty. This method can outperform both historical simulations and decadal prediction in 5- 10- and 15-year temperature forecasts of winter MED, as well as summer NEU and WCE. Significant skill improvements are visible for 10- and 15-year forecasts of winter Mediterranean surface temperature over land, when using the North Atlantic Oscillation or the Atlantic Multidecadal Variability as predictors in the first selection. The optimal predictor varies by region and should be evaluated on a case-by-case basis. This improved regional climate information supports more targeted adaptation strategies for the coming decades.

1 Introduction

In the context of ongoing climate change, the implementation of adaptation policies requires relevant and actionable information about climate risk evolution over the coming years-to-decades. The near-term future (i.e. the next 10 to 20 years) represents a relevant timescale for strategic decision-making, as this temporal horizon aligns with the planning framework of a large portion of stakeholders in climate-vulnerable sectors (e.g. agriculture) (Kushnir et al., 2019). While the future emission pathways dominate the uncertainty affecting long-term projections at the global scale, internal climate variability associated with the chaotic or aleatoric nature of the climate system, is the leading source of uncertainty in near-term future change at regional scale (Lehner et al., 2020). Reducing the uncertainty related to internal climate variability over the next decades and

https://doi.org/10.5194/egusphere-2025-4463
Preprint. Discussion started: 24 September 2025

providing an objective and reliable estimate of the related modulation of anthropologically-forced changes, is therefore of primary interest.

Multiple lines of evidence are available to provide such information. First, the so-called non-initialized ensembles of climate projections provide a seamless evolution from the historical period to the end of the 21th century as a function of socio-economic scenario. In that case, they encompass the full range of uncertainty including the one related to internal climate variability that is explicitly resolved. Second, so-called initialized decadal predictions aim to specifically reduce this uncertainty by initializing the climate model simulations from estimates of the observed state of the climate system, including the ocean, atmosphere, and other relevant components. This initialization aims to phase the temporal evolution of the simulated and observed modes of climate variability. However, the predictive skill of initialized decadal forecasts often fades out after a few years, showing limited added value over non-initialized projections except over specific regions and for some persistent variables, and they are usually limited to 5–10 years (Yeager et al., 2018). Decadal forecasts are also subject to drift due to the so-called initialisation shock explained by mismatch between biased models and assimilated observational estimates (Sanchez-Gomez et al., 2016). Third, raw observations can be used to provide information to constrain the climate evolution over the next decades. For example, Bonnet et al. (2021) apply an objective selection of members from large ensembles of simulations using observed proxies of the Atlantic meridional overturning circulation (AMOC) in order to narrow the range of possibilities associated with the internal variability of near-term change of AMOC whose effect is long-lasting, and global mean surface temperature. Similarly, Liné et al (2024) proposed a storyline approach in a perfect model framework to partition raw uncertainties of climate change over Northern Europe before 2040 as a function of the combined phase of AMOC and the North Atlantic Oscillation (NAO).

Combining all the sources of information —observations, initialized decadal predictions, and non-initialized climate projections— to deliver the most robust climate information at near term, with reduced uncertainty around the most likely evolution of internal variability, remains a significant challenge (Cassou et al. 2018). Yet, for effective decision-making and long-term adaptation planning, it is crucial that climate information be seamless across timescales, ensuring consistency between historical observations, near-term predictions, and long-term projections.

To address this challenge, several methods have recently been developed to incorporate information from the observed climate state or from decadal predictions within large ensembles of non-initialized transient climate simulations in order to constrain aspects of internal climate variability. Some studies explore this idea by developing methods based on the subselection of non-initialized climate projections from large ensembles based on their agreement with sea surface temperature (SST) evolution Befort et al. (2020) or with SST patterns Mahmood et al. (2021) assessed from initialized decadal predictions. They highlight the added-value of these methods in comparison to the full ensemble of simulations beyond the time period covered by decadal predictions. Mahmood et al. (2022) proposed a similar method to constrain non-initialized climate simulations, but using observed SST patterns instead of information taken from decadal prediction. Their method shows skill levels comparable to state-of-the-art decadal prediction systems for 10-year forecasts. Donat et al. (2024) provide a consistent evaluation of these different approaches and highlight that a selection of non-initialized members based on observations or

https://doi.org/10.5194/egusphere-2025-4463
Preprint. Discussion started: 24 September 2025

decadal predictions significantly enhances the skill of 10 and 20-year projections for near-surface temperatures in some regions, including Europe, with the selection based on decadal predictions having the largest added value in terms of probabilistic skill. Similarly, Cos et al. (2024) provide a comparison of these methods to predict near-term mediterranean summer temperature but show instead heterogeneous improvements in comparison to the full non-initialized climate simulations from the Coupled Model Intercomparison Project Phase 6 (CMIP6). Other methods derive seasonal to decadal climate predictions by constraining large climate model ensembles using analogue approaches. Menary et al. (2021) for example developed a new analog-method to derive a skillful decadal forecast of the subpolar North Atlantic SSTs in comparison to climate prediction system by selecting 35-years analog of the observed spatial averaged evolution from CMIP5 and CMIP6 archives.

To date, proposed methods have relied on information from either observations or decadal prediction. In this study, we explore the potential benefits of blending the three sources of information available —observations, initialized decadal predictions, and non-initialized climate projections— to provide relevant and seamless information about near-term climate change with reduced uncertainty related to internal climate variability. This new "blending" method is based on the constraint of non-initialized climate projections from large ensembles, using information from both observations and decadal predictions. The performance of datasets derived from the blending method in predicting winter and summer surface temperatures over Europe will be evaluated using a retrospective assessment framework, as in Cos et al. (2024).

This paper is organized as follows: Section 2 details the data and methods, Section 3 evaluates the blending method, and Section 4 presents a summary and discussion of the results.

2 Data and Method

2.2 Datasets

We use 163 non-initialized transient historical simulations (historical simulations hereafter) from CMIP6 (Eyring et al., 2016) and 92 initialized decadal hindcasts (hindcast hereafter) from CMIP6/DCPP Component A (Boer et al., 2016), based on large ensembles from six models (see Table 1). The historical simulations start from the atmospheric, oceanic and land surface initial conditions of a preindustrial simulation and are forced with estimates of anthropogenic and natural forcings from 1850 to 2014. Hindcast simulations are initialized each year from 1960 with best estimates of the observed climate state— including ocean, atmosphere, and other components, and run for five or ten years depending on the model. For simplicity, all data have been regridded to the CNRM-CM6-1 atmospheric grid that is our in-house model and only the land grid points with at least 70% of land are considered. This choice has been motivated to keep enough grid points along the coast.

A lead-dependent drift correction is applied to the hindcast simulations prior to their use in order to remove the mean drift caused by the initialization shock (Boer et al., 2016). In practice, for each model, the drift is estimated as the ensemble mean over the period of interest as a function of each lead time. This drift is then subtracted from each hindcast year at the corresponding lead time to correct for mean biases.

https://doi.org/10.5194/egusphere-2025-4463
Preprint. Discussion started: 24 September 2025

Model (historical)	Number of members	Model (hindcast)	Number of members
CNRM-CM6-1	30	CNRM-ESM2-1 (5-yr)	25
EC-Earth3	15	EC-Earth3 (10-yr)	15
MIROC6	50	MIROC6 (10-yr)	10
MRI-ESM2-0	12	MRI-ESM2-0 (5-yr)	12
NorCPM1	30	NorCPM1 (10-yr)	20
IPSL-CM6A-LR	26	IPSL-CM6A-LR (10-yr)	10

Table 1: CMIP6 models and associated number of DECK historical simulations used in this study (left) and CMIP6 models and the associated number and list of hindcast simulations, as well as their time length, used in this study (right).

2.2 Climate indices

As described in the next section, the blending method developed in this study consists, in a first step, in selecting the non-initialized historical simulations that most closely match few pre-determined observed climate indices before the forecast start date. Based on literature, we select four climate indices representing large-scale phenomena that drive climate predictability of surface temperature over Europe from decadal to multidecadal timescales (García-Serrano et al., 2015; Smith et al., 2019).

The first index is the Atlantic Multidecadal Variability (AMV) index, which describe the evolution of the leading mode of multidecadal variability in the North Atlantic Ocean (Schlesinger and Ramankutty, 1994; Enfield et al., 2001; Yeager and Robson, 2017). The AMV is characterised by basin-wide SST. It has been linked to many observed low-frequency global and regional climate variations, including the Northern Hemisphere temperature (Zhang et al., 2007) and the European precipitation and temperature (Sutton & Dong, 2012, Qasmi et al., 2020). To estimate the evolution of the AMV, we define the AMV index as the average SST over the North Atlantic (0–60° N, 80°W–0° E) after the removal of the externally forced signal following Trenberth and Shea, (2006).

The second index describes the evolution of the North Atlantic subpolar gyre (NASPG), which is a key part of the SST decadal variability in the North Atlantic and has been linked to the European climate (e.g. Hermanson et al., 2014). The NASPG is of particular interest as skillful predictions of up to a decade can be achieved over this region (e.g., Matei et al., 2012; Brune et al., 2018; Robson et al., 2018). In this study, we define the SPG index as the average SST over the 15°W–40°W, 50°N–60°N region.

https://doi.org/10.5194/egusphere-2025-4463
Preprint. Discussion started: 24 September 2025

The third index is the 9-year average sea surface temperature (SST) pattern correlation at the global scale. This index has been proposed in previous studies to constrain low-frequency internal climate variability in surface temperature by selecting the simulations that best match the observed SST spatial pattern of sea surface temperature based on spatial correlations at global scale (e.g. Befort et al. 2020; Mahmood et al. 2022).

The fourth index captures the evolution of the winter (December to February) North Atlantic Oscillation (NAO), which is the dominant mode of atmospheric circulation variability in the North Atlantic sector. Winter NAO exerts a strong influence on European weather and climate (e.g. Hurrell et al., 2003) and shows predictability over several years in advance (Smith et al., 2019; Athanasiadis et al., 2020). The NAO index is defined as the difference in area-averaged mean sea level pressure (MSLP) between a southern box (20–55° N, 90°W–60° E) and a northern box (55–90° N, 90°W–60° E) in the North Atlantic (Stephenson et al., 2006; Baker et al., 2018). We choose this regional index because it is less sensitive to modest differences in NAO centres of action between the observations and the CMIP6 models than the station-based index (Hurrell et al., 2003; Stephenson et al., 2006). Another benefit of this index is that it is less affected by issues of interpretability that occur when a mathematically constructed empirical orthogonal function (EOF)-based index is used (Ambaum et al., 2001; Dommenget and Latif, 2002; Stephenson et al., 2006).

The oceanic SST indices are evaluated against the NOAA Extended Reconstructed SST V5 (ERSSTv5; Huang et al., 2017) observed dataset. The NAO index is evaluated against the ERA5 reanalysis (Hersbach et al. 2020). A Lanczos low-pass filter with a cutoff frequency of 1/10 years and 11 weights is applied to all indices to retain only the low-frequency variations, except for the third one.

2.3 Blending method protocole

The goal of the blending method developed here (BLEND hereafter) is to make the best use of the different sources of available information to provide the most robust and actionable forecast possible, of a variable of interest over a specific region, through reduction of aleatoric and structural uncertainty due to internal climate variability and climate models, respectively. The method is illustrated here using a case study : the forecast at leadtime [1-5]-yr of winter land surface temperature in the Mediterranean region, as defined in the IPCC (Iturbide et al., 2020), starting in 1977 (Fig. 1).

In a preprocessing step, the drift from the hindcast is removed using the method described in section 2.1. Then the temperature anomalies for the observations, the 163 historical simulations (HIST) and the 92 hindcasts (DEC) are all computed over the 5-yr forecast window [1977-1981] relative to the period 1967-2000.

Two different types of dataset are built. First, HIST_{OBS}, which consists in a selection of simulations from HIST that best-fit the evolution of a given observational metric over *a calibration period* of Y years preceding the start of the forecast period (Fig. 1a). In this example, we consider the AMV index as observational constraint (see section 2.3) and Y=20 years, namely 1957-1976. To assess the similarity between the historical simulations and the observations, the RMSE and the correlation scores are computed over time, on an annual basis from the temperature anomalies calculated over the full 1900–

https://doi.org/10.5194/egusphere-2025-4463
Preprint. Discussion started: 24 September 2025

2010 period. The simulations are ranked based on the sum of the two normalized scores and N = 20 best simulations are retained.

The second dataset $BLEND_{OBS}$ is based on a double constraint: the one from observations as for $HIST_{obs}$ and a new one from DEC that are available over the forecast period (Fig. 1b). First, 30 simulations that best match the observation index over the Y=20 years before the forecast are selected, applying the same method used for $HIST_{OBS}$. Then, 20 simulations out of 30 that show the lowest absolute error with respect to the hindcast ensemble mean surface temperature over the region of interest, are retained. This second step in $BLEND_{obs}$ is applied in order to take full advantage from all the different sources of available information. We tested various combinations of member selection and chose to retain 30 members for the first selection and 20 for the second, as this setup provides a relatively strong constraint on HIST while preserving ensemble spread.

Decadal prediction performance is evaluated by comparing the distribution of all ensemble forecast dataset available (Fig. 1c). In our illustrative example, HIST treated here as the benchmark ensemble predict a warming with a substantial uncertainty assessed by the spread, namely -0.16 ± 0.62 °C (ensemble mean ± 90th percentile range). DEC shows a close ensemble mean to HIST, but with a reduced uncertainty, namely -0.16 ± 0.6 °C. The temperature forecast from $HIST_{OBS}$, -0.12 ± 0.52°C, shows a decrease in spread and a closer ensemble mean to the observation (0.09°C), while $BLEND_{OBS}$, -0.01 ± 0.23°C, significantly reduces the uncertainty and has an ensemble mean even closer to the observation.

Finally, we introduce two additional ensemble forecasts that are useful for evaluation purposes. First, $HIST_{TAS}$, which derives from the first step of BLEND and uses the average surface temperature over the region of interest as observational metric. This allows us to assess whether using only the variable we aim to predict is sufficient to constrain the historical simulations. Second, $HIST_{hindcast}$, which derives from the selection of the 20 simulations from HIST that are closest to the 5-year [1977-1981] hindcast ensemble-mean surface temperature. This allows us to assess the added value using only the second step of the $BLEND_{OBS}$ dataset.

In this study, we test this method to predict summer (June to August; JJA) and winter (December to February; DJF) surface temperature over the 3 European IPCC reference regions: Northern Europe (NEU), West Central Europe (WCE) and the Mediterranean (MED) (Iturbide et al., 2020). Therefore, a $BLEND_{OBS}$ dataset is produced for each region, as they use the averaged temperature over the region of interest in the second constraint. This is also the case for $HIST_{TAS}$ and $HIST_{Hindcast}$. The four climate indices described in section 2.2 are tested to constrain the historical simulations, resulting in the $HIST_{OBS}$ and $BLEND_{OBS}$ experiments. Note that for the observational metrics based on spatial global SST, we used Y= 9 years instead of 20 years, as used in previous studies (Befort et al. 2020; Mahmood et al. 2022). Therefore, we will evaluate HIST and DEC against $HIST_{AMV}$, $HIST_{SPG}$, $HIST_{NAO}$, $HIST_{GSST}$ and $BLEND_{AMV}$, $BLEND_{SPG}$, $BLEND_{NAO}$, $BLEND_{GSST}$, $HIST_{TAS}$ and $HIST_{Hindcast}$.

https://doi.org/10.5194/egusphere-2025-4463
Preprint. Discussion started: 24 September 2025

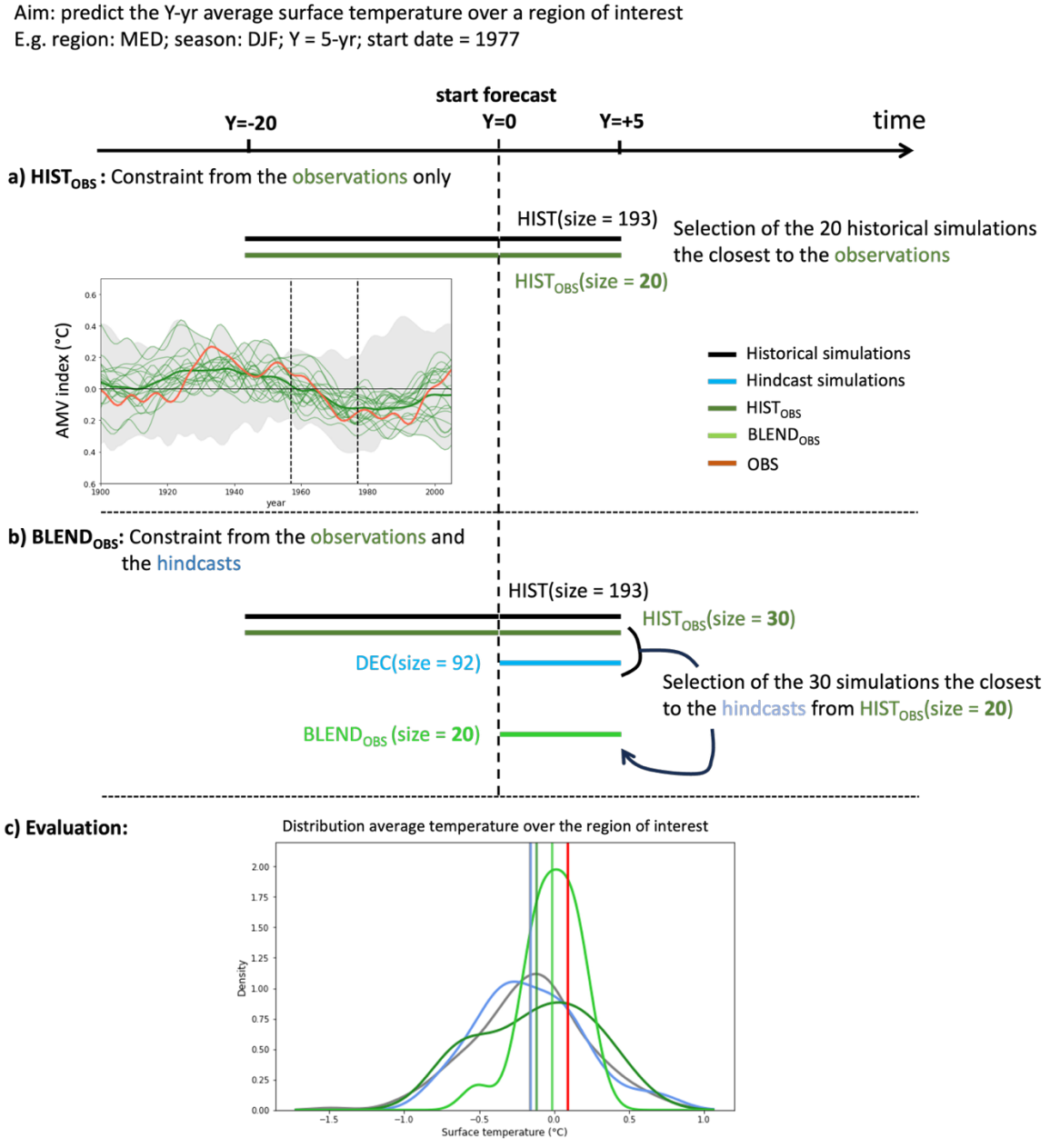

Figure 1. Diagram illustrating the concept and two types of dataset derived from the blending method. The data used for the historical simulations and the hindcast simulations mixed all the models together (see section 2.1) are described in Tables 1 and 2. In the development of HistOBS (a) the AMV index (see section 2.2) from the historical simulations (minimum and maximum in gray) is compared to the ERSSTv5 observational dataset (green line), with the selection of the best 20 members in dark green. In the development of BlendOBS (b), a first selection of 30 members is then refined to 20 members. The histogram and the ensemble mean (vertical line) of surface temperature forecasts from the different datasets are evaluated against the ERA5 reanalysis (red line) (c).

https://doi.org/10.5194/egusphere-2025-4463
Preprint. Discussion started: 24 September 2025

2.4 Evaluation metrics

An important point is to evaluate the added value of BLEND developed in this study compared to traditional studies (Donat et al., 2024). This added value can differ as a function of the user's needs and objectives. It can be a reduction of the spread in comparison to the full ensemble of historical simulations, or a reduction of the error of the ensemble mean corresponding to the most likely outcome. Several metrics are compared in this study, in order to evaluate the uncertainties in the score used for the evaluation and to cover a wide range of potential users.

To do that, we applied a retrospective evaluation. BLEND is applied each year over the 1967-2000 period, so that for each year we have an ensemble forecast of the temperature over the next 5, 10 and 15 years and because the 1967 forecast uses data processed (filtering, 20-yr length of the obs-constraining period, etc) from 1940, the initial date of the ERA5 product. For each forecast horizon, we compute the spread, defined as the differences between the maximum and minimum values of the ensemble forecasts, and the absolute error between the ensemble mean forecast obtained from our method and the observations.

Then, we use three different scores to evaluate the 1-to-5, 1-to-10 and 1-to-15 years mean temperature forecasts over the 1967-2000 period. The first two scores are probabilistic: the ranked probability skill score (RPSS; Wilks, 2011) and the Continuous Ranked Probability Skill Score (CRPSS). They indicate the skill of a forecast against a reference forecast, with positive values indicating better skill than the reference.

The RPSS is derived from the relative difference between the ranked probability score (RPS) of one forecast derived from the blending method and the historical ensemble defined as the reference: RPSS = 1 - (RPS_{BLEND} / RPS_{HIST}). The RPS quantifies the squared cumulative probability error for categorical events. Each ensemble forecast is divided into three equiprobable categories, as in Mahmood et al. (2022) and Coz et al. (2024), computing the terciles separately for observations and simulations to avoid the biases impact in mean and variance.

As for the RPSS, the CRPSS is derived from the relative difference between the Continuous Ranked Probability Score (CRPS) of one forecast derived from the method ($HIST_{OBS}$ and $BLEND_{OBS}$) and the historical ensemble (Hersbach 2000). The CRPS measures the integrated squared difference between the forecast cumulative distribution function (CDF) and the observed CDF and is widely used in evaluating probabilistic forecasts (e.g. Goddard et al., 2013; Alfieri et al., 2014).

The third score considered in this assessment is the mean squared skill score (MSSS), based on the mean squared error (MSE) between a set of paired forecasts, F_j, and observations, O_j, over j = 1 to n years of the evaluation period (1967-2000), following Murphy (1988). It is defined as MSSS = 1 - (MSE_{BLEND} / MSE_{HIST}), with the MSE given by

$$MSE = \frac{1}{N} \sum_{j=1}^{n} (F_j - O_j)^2 \quad (1)$$

with F the forecast from HIST, DEC or BLEND and O the observations. A positive MSSS indicates that the test forecast outperforms the reference forecast (here the historical ensemble), while a negative MSSS indicates lower skill relative to the reference.

Finally, we compute the difference in the temporal anomaly correlation coefficient (ACC) of the ensemble mean historical and constrained simulations obtained by BLEND against the observations. The residual correlation (ResCor) (Smith et al., 2019)

https://doi.org/10.5194/egusphere-2025-4463
Preprint. Discussion started: 24 September 2025

is then computed to assess whether the constrained ensembles capture any part of the observed internal variability that is not already explained by the ensemble mean of the historical simulations, which describe the forced response. The observational reference and the constrained ensemble are regressed against the historical simulations and their residuals correlated against each other. The statistical significance of ResCor is computed using bootstrap (using n=1000 resampling) (Good, 2005).

3 Results

3.1 Regional performance of BLEND over Europe in winter and summer

$HIST_{obs}$ exhibits a pronounced reduction in spread relative to HIST and DEC during the testing period (1967–2000), independently of regions and seasons and whatever the observational index used for constraint (Fig. 2). The relatively similar temperature trajectories in all $HIST_{obs}$ suggest that all selected indices in observations are relevant to constrain surface temperatures simulated in HIST over Europe. A larger reduction in spread is obtained in $BLEND_{obs}$ compared to $HIST_{obs}$, particularly at 5-year leadtime.

$HIST_{hindcast}$ shows the largest spread reduction as expected by construction. The large pool of historical simulations increases the likelihood of finding good analogs of hindcast ensemble mean, resulting in a narrower spread. The reduction in spread is lost at greater leadtime for 10-year and 15-year forecasts, as the analogues are selected solely based on 5-year hindcasts. (section 2.3).

The spread in DEC is comparable or slightly lower than in HIST. This small added value in some region such as WCE is consistent with earlier studies that evidenced weak previsibility over Europe for 2m-temperature, which may be due to poor hindcast performance and could have a structural origin due for instance, to the initialization shock, which quasi-systematically triggers El Niño events in the first year of the forecast and also negative NAO type mean bias (Sanchez-Gomez et al., 2016). All these may overshadow any add values from ocean initialization in DEC, thereby affecting the hindcasts quality at any leadtime. DEC poor scores could also have an intrinsic, i.e., “true” climate origin because of the relatively weak decadal variability in surface temperature over Europe compared to the strong chaotic atmospheric variability at intraseasonal to interannual timescale and that is therefore drawn in the noise. This could be finally due to model errors in forecasting.

Regarding absolute errors in ensemble means, $HIST_{obs}$ and $BLEND_{obs}$ are overall relatively close to HIST, especially in winter (Fig. 3). Although absolute errors from DEC are also close to HIST, there are regions, such as MED in winter or WCE in summer, for which 5-yr forecast temperature from DEC shows a slight reduction in absolute errors in comparison to HIST. This highlights the potential benefit from the second constraint of BLEND based on the hindcast ensemble mean of 5-yr surface temperature over the region of interest (Fig. 3d, e).

For the WCE and MED regions in summer, the $HIST_{TAS}$ shows the lowest median absolute error for the 5-, 10-, and 15-year mean-temperature forecasts (Fig. 3d, f). This could be due either to the fact that using surface temperature over these regions inherently captures the associated low-frequency variability, or to the possibility that the observed surface temperature

https://doi.org/10.5194/egusphere-2025-4463
Preprint. Discussion started: 24 September 2025

lies outside or at the edge of the distribution of historical simulations, meaning that selecting simulations from this part of the distribution would consistently serve as a good predictor.

It is important to note that the overall lower absolute error in the median of $HIST_{obs}$ and $BLEND_{obs}$ ensemble mean compared to HIST does not imply that the observations fall within the spread of $HIST_{obs}$ and $BLEND_{obs}$. Nevertheless, it highlights the added value of using ensemble mean from datasets derived from BLEND developed here over the near-term future, especially for the 10- and 15-year forecasts, in comparison to the historical ensemble mean.

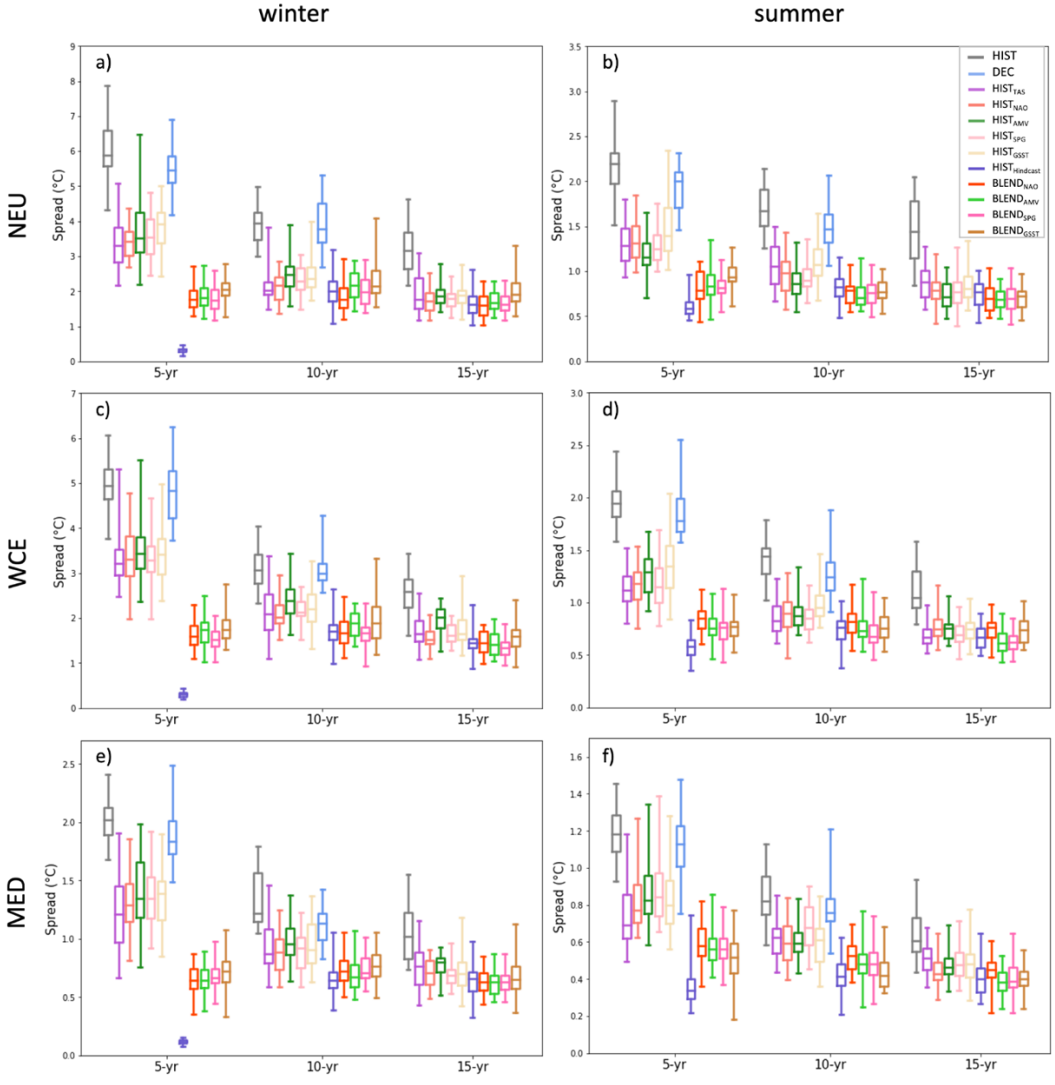

Figure 2: Boxplots of the spread of average surface temperature for 5, 10 and 15 years forecast in (a, c, e) winter and (b, d, f) summer over the NEU (a, b), WCE (b, c) and MED (e, f) regions. The spread is defined as the difference between the minimum and the maximum and is calculated for each year of the retrospective evaluation period (1967-2000). The boxplots are defined with the minimum, 25th percentile, median, 75th percentile and maximum.

https://doi.org/10.5194/egusphere-2025-4463
Preprint. Discussion started: 24 September 2025

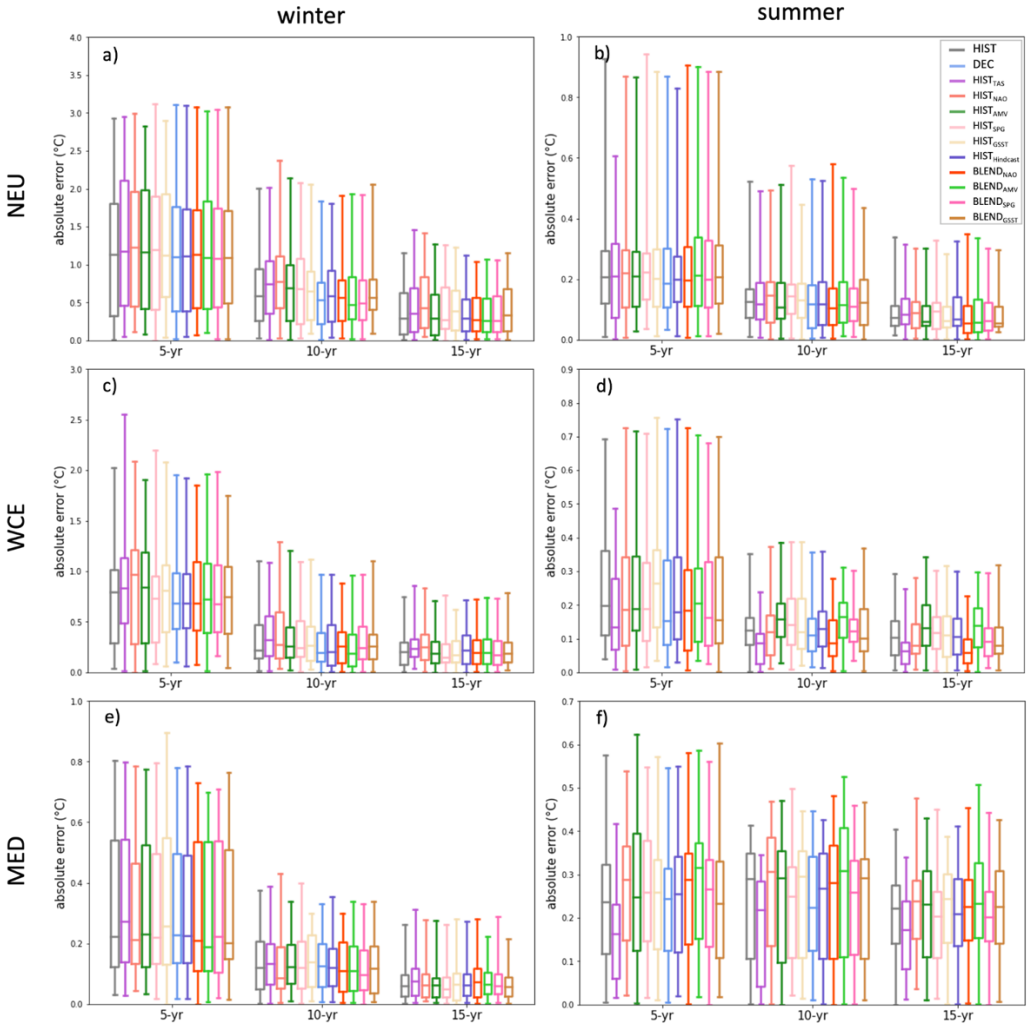

Figure 3: Boxplots of the surface temperature absolute error for 5, 10 and 15 years forecasts for (a, c, e) winter and (b, d, f) summer over the NEU (a, b), WCE (c, d) and MED (e, f) regions. The absolute error is calculated each year of the testing period (1967-2000) between the observed surface temperature from ERA5 (Hersbach et al. 2020) and the ensemble mean of the different dataset described section 2.1 and 2.3. The distribution of these absolute errors is represented by the boxplots. The boxplots are defined with the minimum, 25th percentile, median, 75th percentile and maximum.

The added value of BLEND is further evidenced from MSSS, RPSS and CRPSS, skill scores (Fig. 4). For winter NEU, DEC, $BLEND_{NAO}$ and $BLEND_{AMV}$ provide better 10 and 15-year forecasts assessed from the three scores, although it is less pronounced for the RPSS (Fig. 4a). $HIST_{NAO}$ has the lowest scores, despite a clear relationship has been established at decadal timescales between the atmospheric circulation index and surface temperature over Northern Europe (e.g. Iles and Hegerl, 2017). This may be explained by structural models underestimation of the teleconnection over NEU, potentially due to the fact that the spatial pattern of NAO in CMIP-class of models is southward shifted and has underestimated power in the decadal frequency band of variability (e.g. Eyring et al., 2021; Bonnet et al., 2024). Another possible explanation is that the

https://doi.org/10.5194/egusphere-2025-4463
Preprint. Discussion started: 24 September 2025

20-year period used prior to the forecast to select members is suboptimal—possibly too long considering the decorrelation time-scale of the NAO. The large forecast improvement gained from the second subselection—based on hindcast surface-temperature anomalies over winter NEU—highlights the value of adding this step to take full advantage of all available information.

For winter WCE, $BLEND_{NAO}$, and $BLEND_{AMV}$ 10-year forecasts perform better than HIST when looking at the MSSS and CRPSS (Fig. 4c). DEC achieves the second-best MSSS score for 10-year average predictions, whereas it shows a slight improvement in RPSS and no improvement in CRPSS compared to HIST. This suggests that DEC outperforms HIST in predicting the target variable when considering the ensemble mean. However, its lower CRPSS reflects reduced reliability — the extent to which forecasts are statistically consistent with observed outcomes— in representing the full distribution. This suggests that while DEC may better capture the central tendency (i.e., deterministic predictability), HIST offers more reliable probabilistic information. This could be due to the smaller ensemble size from DEC in comparison to HIST. For the 15-year forecast, the MSSS and CRPSS scores show an improvement for $HIST_{GSST}$. Therefore, the GSST predictor (see sect. 2.2) seems to integrate the large-scale drivers of the climate predictability of winter surface temperature over WCE at 15-years timescale.

For winter in the MED region, the $BLEND_{AMV}$, $BLEND_{SPG}$ and $BLEND_{NAO}$ significantly improve the 10- and 15-year forecasts compared to both HIST and DEC across the scores (Fig. 4e). In contrast, $HIST_{TAS}$ shows a marked deterioration in performance. For the 5-year forecast, $HIST_{NAO}$ performs the best among all the dataset investigated here. DEC also shows notable improvement in 5-year forecast performance across all three scores compared to HIST.

For summer NEU, skills vary depending on the evaluation score used. DEC shows substantial added value for 5- and 10-year forecasts according to the MSSS, but performs worse in terms of probabilistic scores at 10-year forecast (Fig. 4b). $HIST_{TAS}$ performs the best for the three scores for the 5-year forecast. Although a large portion of the 10-year forecasts from $HIST_{obs}$ and $BLEND_{obs}$ outperform HIST in terms of MSSS and CRPSS, this is not observed for RPSS, where only $BLEND_{GSST}$ and $BLEND_{SPG}$ produce better forecasts than HIST. Regarding MSSS, all forecasts derived from BLEND yield more accurate 5-year predictions than HIST.

These discrepancies highlight that results can vary significantly depending on the evaluation metric used, which reinforces the importance of using several complementary metrics to obtain a more robust assessment.

For summer in the WCE and MED regions, $HIST_{TAS}$ shows a large improvement in forecasts at all leadtime compared to HIST and DEC across all three scores, except for RPSS in MED at the 10 and 15-year lead times (Fig. 4d, f). $BLEND_{NAO}$ also demonstrates significant improvement for summer in WCE, again across all three scores, and surpasses DEC at the 10-year forecast horizon. On the opposite, $HIST_{AMV}$ and $BLEND_{AMV}$ have poorer predictability than the historical ensemble considering the forecast time series, which is consistent with the increase in the ensemble mean absolute error regarding each year individually (Fig. 3d).

For summer MED, the $BLEND_{AMV}$ shows a decrease in scores in comparison to $HIST_{AMV}$, which means that the constraint by DEC, based on the similarity with the DEC ensemble mean temperature forecast over MED from $HIST_{AMV}$,

https://doi.org/10.5194/egusphere-2025-4463
Preprint. Discussion started: 24 September 2025

deteriorates the prediction. The lower performance of the other $BLEND_{OBS}$ forecasts compared to $HIST_{OBS}$ suggests that the DEC ensemble mean may provide limited predictive skill over this region—an issue that can be further exacerbated when the predictor used offers poor climate predictability of summer surface temperatures at decadal to multidecadal timescales.

These results are summarized in Figs S1, S2 and S3. In some cases, such as 10-year temperature forecasts of winter MED, the blending method developed in this study provides substantial added value when the AMV, SPG or NAO are used (Fig. S2). Indeed, $BLEND_{NAO}$, $BLEND_{AMV}$ and $BLEND_{SPG}$ show a reduced spread compared to both HIST and DEC, while also lowering the error relative to ensemble mean (Figs. 2e and 3e). These ensemble forecasts also more accurately represent the temperature evolution over the evaluation period across all three tested scores (Figs. 4e, 5c, and 6c). Therefore, BLEND appears to capture part of the low-frequency internal variability of winter temperatures in MED. This is consistent with previous studies (e.g., Mariotti and Dell'Aquila, 2012) that emphasize the role of the AMV and NAO in modulating the decadal variability of MED winter surface temperatures. However, this added value is sensitive to the forecast time horizon, with a noticeable decrease, particularly at the 5-year lead time for $BLEND_{NAO}$ (Fig. S1).

BLEND also provides strong added value over HIST and DEC for 5, 10 and 15-year summer temperature forecasts over WCE, using the $BLEND_{SPG}$ and $BLEND_{NAO}$ approaches (Figs. S1, S2, S3).

In some cases, such as winter over NEU and summer over MED, BLEND provides little to no added value compared to HIST and DEC for 5-year forecasts (Fig. S1). However, at 10 and 15-year forecasts, $BLEND_{AMV}$, $BLEND_{SPG}$ and $BLEND_{NAO}$ show substantial added value over HIST for winter NEU (Figs. S2 and S3).

For winter WCE, BLEND provides overall no or small added value for 5-year forecasts (Fig. S2). Some added value can be found with $BLEND_{NAO}$ at 10- and 15-year forecasts, depending on the score of interest (Figs. S3 and S4). For summer WCE, $BLEND_{NAO}$ also provides good added value at 5- 10- and 15-year forecasts (Figs. S2, S3 and S4).

The evaluations presented here highlight that tailoring the choice of observational predictors according to region and forecast horizon is key to improving forecast performance.

https://doi.org/10.5194/egusphere-2025-4463
Preprint. Discussion started: 24 September 2025

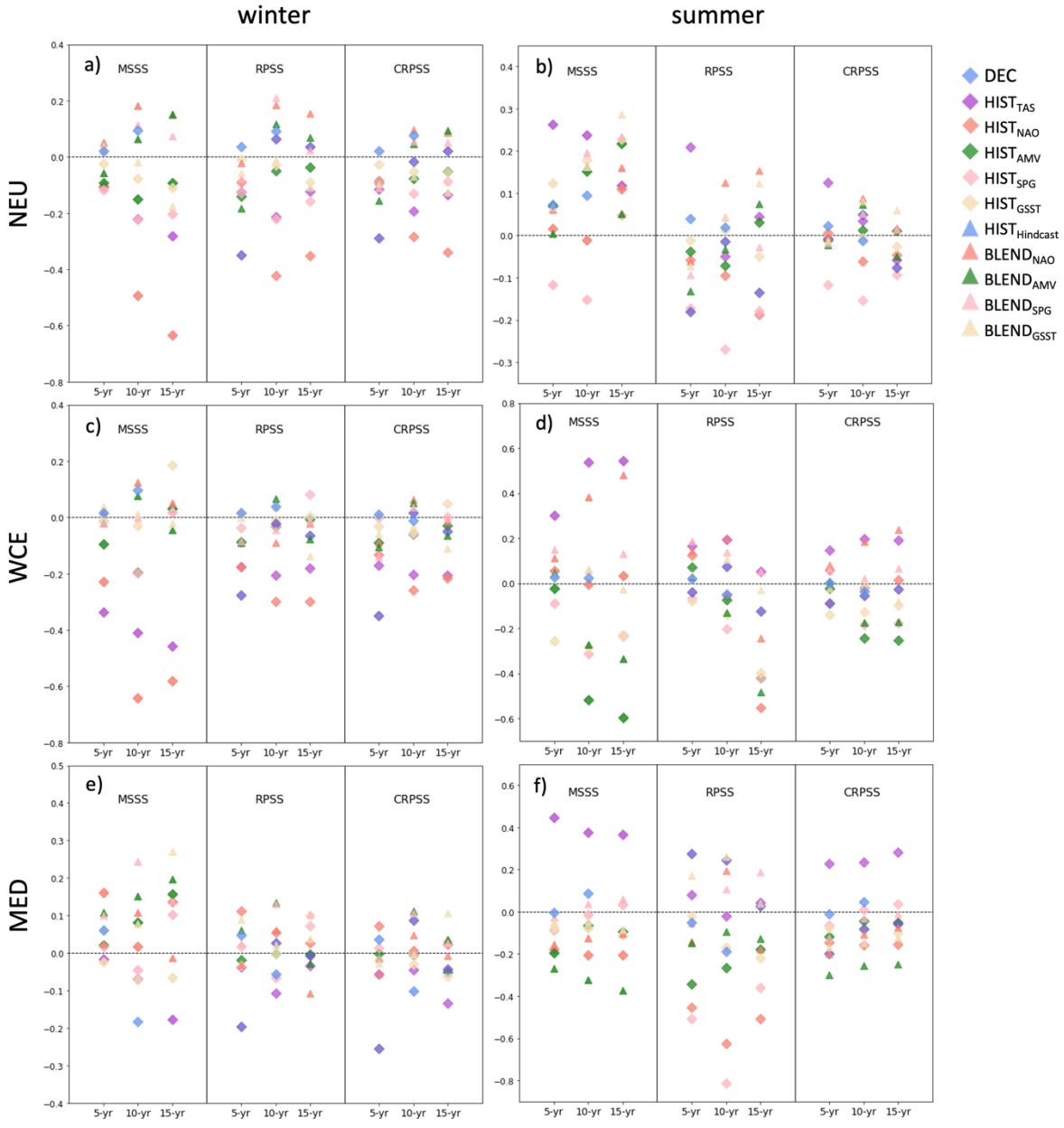

Figure 4: MSSS, RPSS and CRPSS (see section 2.4) calculated from the 5, 10 and 15 years time series of (a, c, e) winter and (b, d, f) summer forecast of surface temperature over (a, b) NEU, (c, d) WCE and (e, f) MED from the historical and hindcasts dataset (see section 2.1), as well as the dataset derived from BLEND (see section 2.3). The scores are calculated over the 1967-2000 period, using the ensemble mean for the MSSS and the whole ensembles for the RPSS and CRPSS.

3.2 Wintertime spatial evaluation: a case study for the MED region

We now evaluate BLEND at the spatial scale for winter temperatures, to study its added value in comparison to HIST and DEC at the regional level. We focus on evaluating BLEND$_{NAO}$ and BLEND$_{AMV}$ that provide added value at this season

https://doi.org/10.5194/egusphere-2025-4463
Preprint. Discussion started: 24 September 2025

over the Mediterranean region. As a reminder, the second constraint in BLEND is based on the averaged temperature over the region of interest, here MED, from the ensemble mean of DEC.

Due to the presence of anthropogenically-forced warming trends, which largely dominate the forecast signal, the anomaly correlation coefficient (ACC) between observed temperature and forecast ensembles is very high across much of Europe whatever the datasets (not shown). To better identify skill improvements, we use residual correlations, as described in section 2.4 (Smith et al., 2019). DEC shows some significant added value over parts of Northern Europe with respect to HIST, but only limited improvement over the Mediterranean region (Supplementary Fig. S1a). Skill is slightly improved in the Mediterranean with $HIST_{Hindcast}$, particularly over the western Maghreb (Supplementary Fig. S1b).

$HIST_{AMV}$ provides significant skill improvement over Spain, France, and some parts of the Maghreb and the Western Middle East (Supplementary Fig. S2c). This skill in ACC is even enhanced over the Mediterranean region for $BLEND_{AMV}$ (Supplementary Fig. S2d). $HIST_{NAO}$ also provides some significant added skill over the Mediterranean region, but with a pronounced decrease of skill over South Western Europe (Supplementary Fig. S2e). As for $BLEND_{AMV}$, $BLEND_{NAO}$ shows a significant ACC skill increase over a large part of the Mediterranean region (Supplementary Fig. S2f).

Considering both MSSS and CRPSS, DEC provides strong added value over a large part of Northern Europe with respect to HIST, but a poorer 10-year forecast of winter surface temperature over most of the Mediterranean region, except over Spain, parts of the Maghreb and the Middle East (Figs. 5a and 6a), in line with the results for average temperature (Fig. 4c). $HIST_{Hindcast}$ also shows added value over Northern Europe, but to a lesser extent than DEC (Figs. 5b and 6b). Over MED, $HIST_{Hindcast}$ shows more contrasted results, with small added value for the CRPSS and some small improvements regarding MSSS, especially over Spain and parts of the Western Maghreb.

$HIST_{AMV}$ yields better 10-year forecasts of winter surface temperature over Spain and France than HIST, but offers little and heterogenous added value over the rest of the Mediterranean region, with some improvements over the Magrheb in comparison to HIST (Figs. 5c and 6c). $HIST_{NAO}$ shows skill improvements in the western Middle East but exhibits poorer forecast performance over much of southwestern Europe (Figs. 5e and 6e). In contrast, $BLEND_{AMV}$ and $BLEND_{NAO}$ show substantial improvements in 10-year forecasts of winter surface temperature across the Mediterranean region for both MSSS and CRPSS, with clear added value compared to both the historical and hindcast ensembles (Figs. 5d, f and 6d, f).

$HIST_{AMV}$ and $HIST_{NAO}$ provide poorer 10-year forecast performance regarding most of Northern Europe in comparison to HIST and DEC for both MSSS and CRPSS (Figs. 5c,e and 6c,e). The 10-year forecast performance is further degraded for $BLEND_{AMV}$, as the second constraint in the blending method relies on the similarity of the regional average temperature—here over MED—derived from DEC (Figs. 5d and 6d). For $BLEND_{NAO}$, this second constraint improves the forecast quality in comparison to $HIST_{NAO}$ over a large part of Europe (Figs. 5f and 6f).

The results for $BLEND_{AMV}$ and $BLEND_{NAO}$ highlight the added value of the blending method developed in this study in providing improved 10-year forecasts of winter temperature over significant areas in the Mediterranean region, compared to both the historical and hindcast ensembles.

https://doi.org/10.5194/egusphere-2025-4463
Preprint. Discussion started: 24 September 2025

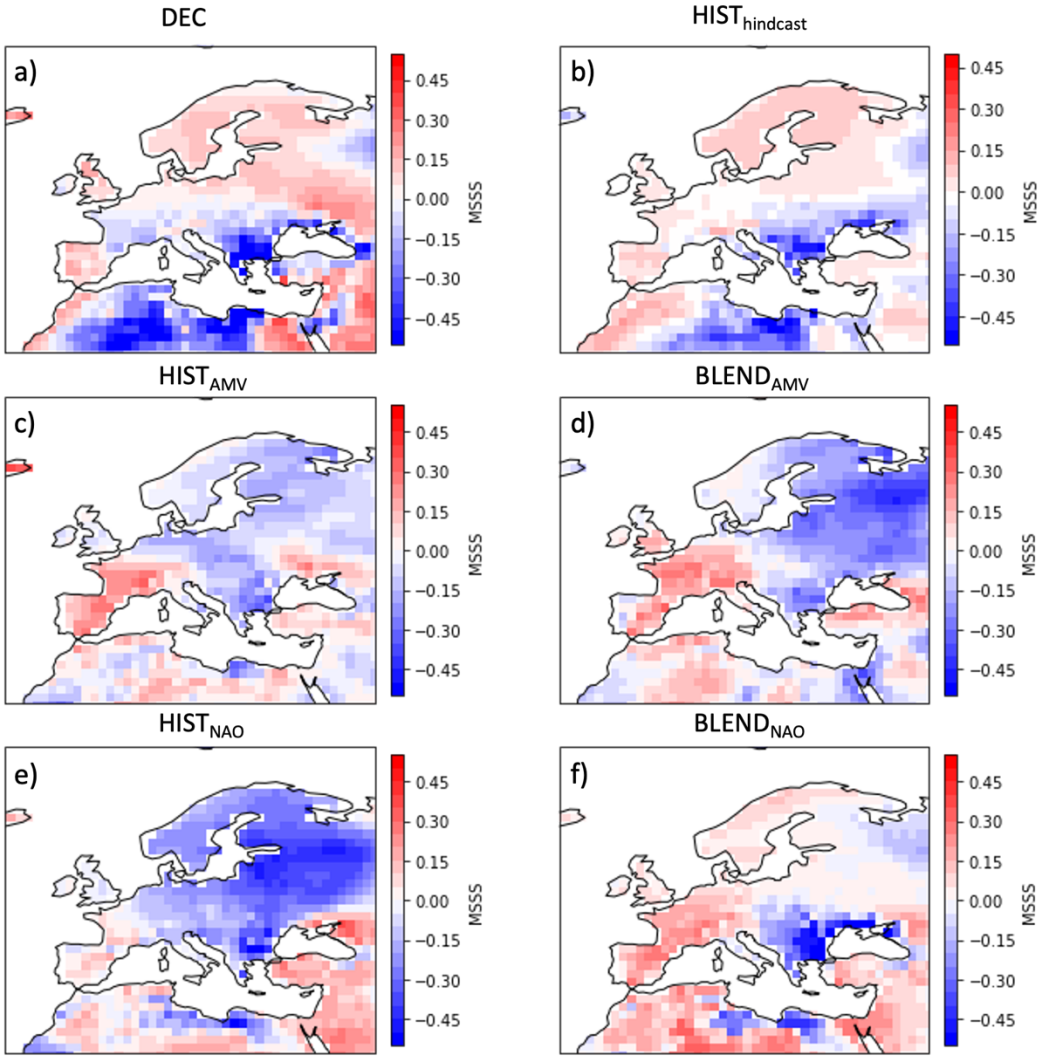

Figure 5: MSSS calculated from the time series of 10 years forecast of winter surface temperature over the evaluation period (1967-2000) for (a) the hindcasts dataset (see section 2.1), (b) $HIST_{Hindcast}$, (c) $HIST_{AMV}$, (d) $BLEND_{AMV}$, (e) $HIST_{NAO}$ and (f) $BLEND_{NAO}$. The winter surface temperature averaged over the Mediterranean region is used for the second step selection in $BLEND_{AMV}$ and $BLEND_{NAO}$.

https://doi.org/10.5194/egusphere-2025-4463
Preprint. Discussion started: 24 September 2025

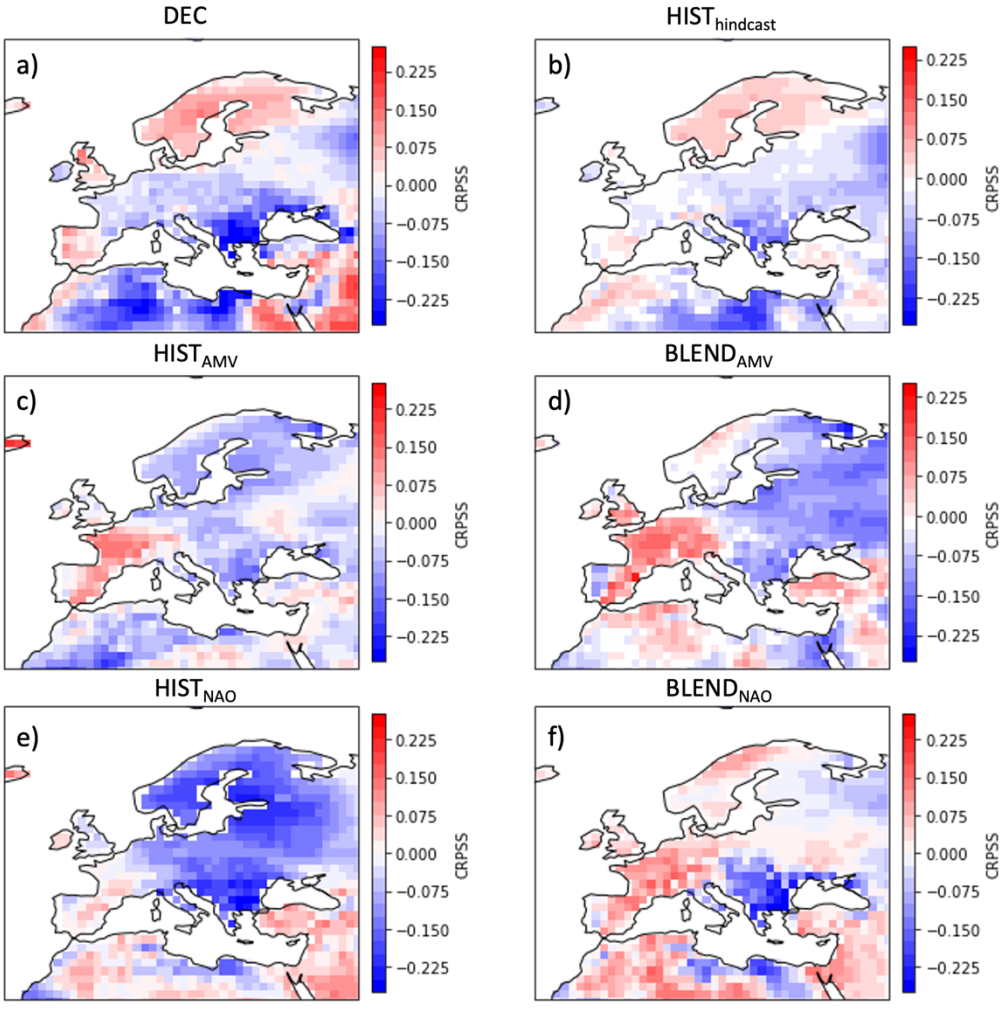

Figure 6: CRPSS calculated from the time series of 10 years forecast of winter surface temperature over the evaluation period (1967-2000) for (a) the hindcasts dataset (see section 2.1), (b) $HIST_{Hindcast}$, (c) $HIST_{AMV}$, (d) $BLEND_{AMV}$, (e) $HIST_{NAO}$ and (f) $BLEND_{NAO}$. The winter surface temperature averaged over the Mediterranean region is used for the second step selection in $BLEND_{AMV}$ and $BLEND_{NAO}$.

4. Discussion and conclusion

In this study, we introduce a novel blending method that, for the first time, combines information from both observations and decadal predictions—whereas previous approaches relied on only one or the other—to provide seamless and relevant climate information at near-term. By selecting a subset of non-initialized HIST simulations that are closest to both observations and decadal forecasts, the blending method avoids by construction the issue of model drift that typically affects

https://doi.org/10.5194/egusphere-2025-4463
Preprint. Discussion started: 24 September 2025

decadal predictions, while still incorporating their information. A retrospective evaluation was conducted to assess the quality of temperature predictions at different time horizons, every year from 1967 to 2000.

Although this method depends on the quality of decadal prediction systems and the ability of models to simulate teleconnections between relevant climate indices and the variable of interest—which can be limited in some regions—our results show that it can provide substantial added value for winter and summer temperature forecasts over Europe, with reduced uncertainty relative to the historical or hindcast ensembles. The predictors tested in BLEND provide overall a good added value for 5-10 and 15-year forecasts of winter temperature over MED, as well as for summer temperature over NEU (Figs. S2, S3 and S4). This added value is also visible at spatial scale, as illustrated in the case study (see section 3.2), where $BLEND_{AMV}$ and $BLEND_{NAO}$ approaches yield overall more skillful 10-year forecasts of winter temperature over MED than either the historical or hindcast ensemble alone.

The important improvements in 5- 10- and 15 years forecast from BLEND in some regions in winter and summer in comparison to the historical ensemble mean—which reflects only externally forced responses—suggest that the method captures part of the internal decadal temperature variability.

Although some subset derived from the blending method showed consistent added value compared to the historical ensemble across the different scores tested, our results also reveal a strong sensitivity to the choice of the score, which can lead to contrasting conclusions. Therefore, careful consideration must be given to the selection of performance metrics when assessing the added value of any method, ensuring they align with the specific scientific or decision-making question being addressed. It is important to note that although some forecast datasets produced by our blending method show lower skill compared to the historical ensemble, this does not imply they lack skill entirely. Indeed, the historical ensemble already benefits from some skill because of signals that come from external forcing.

In this case study, we used historical ensembles from six global coupled climate models and their corresponding decadal prediction systems. Therefore, the reduction in spread in the forecasts derived from our blending method results from a reduction of the uncertainty related to internal climate variability, as well as the uncertainty arising from differences amongst climate models. One way to quantify the reduction in uncertainty associated only with internal climate variability would be to use a single large ensemble of climate simulations and hindcasts from the prediction system based on the same model, and choosing one member as the observations. In this study, we chose to apply the method directly with observations, using multiple historical ensembles and hindcasts from the corresponding decadal prediction systems, in order to assess the blending method's ability under real-world conditions.

The advantage of the framework proposed here is that it can be easily applied to other regions or variables of interest, provided that the variable exhibits internal multi-annual to decadal variability and associated drivers. The benefit of applying the method at more regional scales and context-relevant variables is that end-users in different sectors typically need climate predictions tailored to specific regions, forecast ranges, periods, and/or seasons, rather than relying on key global or large-scale indices (e.g., Solaraju-Murali et al. 2019).

https://doi.org/10.5194/egusphere-2025-4463
Preprint. Discussion started: 24 September 2025

The application of this blending method could be useful for climate impact studies and downscaling, as it provides seamless climate information across timescales from the historical period to the near-term future, while reducing the range of uncertainties. However, it requires preliminary work to identify the appropriate predictors, which depends on the specific variable(s) and region of interest. Tests assessing the method's ability to provide valuable near-term climate information on extremes, such as heatwaves and droughts, would be of particular interest.

Data statement. All CMIP6 data are available through the Earth System Grid Federation. ERSSTv5 (https://psl.noaa.gov/data/gridded/data.noaa.ersst.v5.html) is provided by the NOAA Earth System Research Laboratory Physical Sciences Division (PSD), Boulder, Colorado, USA, from their website. ERA5 is available from the C3S store (https://cds.climate.copernicus.eu/datasets/reanalysis-era5-single-levels?tab=download).

Author contributions. RB, JB, ESG and CC designed the study and developed the method. RB processed the data, performed the calculations of the indices and the analyses. RB prepared the paper with contributions from all co-authors.

Competing interest. The authors declare no competing of interest.

Acknowledgment. This work was supported by the European Union through the HORIZON-EUROPE IMPETUS4CHANGE project under Grant Agreement 101081555. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union. Neither the European Union nor the granting authority can be held responsible for them.

References

Alfieri, L., Pappenberger, F., Wetterhall, F., Haiden, T., Richardson, D., & Salamon, P.: Evaluation of ensemble streamflow predictions in Europe, Journal of Hydrology, *517*, 913-922, https://doi.org/10.1016/j.jhydrol.2014.06.035, 2014.

Ambaum, M. H., Hoskins, B. J., & Stephenson, D. B.: Arctic oscillation or North Atlantic oscillation?, Journal of Climate, *14*(16), 3495-3507, https://doi.org/10.1175/1520-0442(2001)014<3495:AOONAO>2.0.CO;2, 2001.

Athanasiadis, P. J., Yeager, S., Kwon, Y. O., Bellucci, A., Smith, D. W., & Tibaldi, S.: Decadal predictability of North Atlantic blocking and the NAO, NPJ Climate and Atmospheric Science, *3*(1), 20, https://doi.org/10.1038/s41612-020-0120-6, 2020.

Baker, L. H., Shaffrey, L. C., Sutton, R. T., Weisheimer, A., & Scaife, A. A.: An intercomparison of skill and overconfidence/underconfidence of the wintertime North Atlantic Oscillation in multimodel seasonal forecasts, Geophysical Research Letters, *45*(15), 7808-7817, https://doi.org/10.1029/2018GL078838, 2018.

Befort, D. J., O'Reilly, C. H., & Weisheimer, A.: Constraining projections using decadal predictions, Geophysical Research Letters, 47(18), e2020GL087900, https://doi.org/10.1029/2020GL087900, 2020.

https://doi.org/10.5194/egusphere-2025-4463
Preprint. Discussion started: 24 September 2025

Boer, G. J., Smith, D. M., Cassou, C., Doblas-Reyes, F., Danabasoglu, G., Kirtman, B., ... & Eade, R.: The decadal climate prediction project (DCPP) contribution to CMIP6, Geoscientific Model Development, 9(10), 3751-3777, https://doi.org/10.5194/gmd-9-3751-2016, 2016.

Bonnet, R., Swingedouw, D., Gastineau, G., Boucher, O., Deshayes, J., Hourdin, F., ... & Sima, A.: Increased risk of near term global warming due to a recent AMOC weakening, Nature communications, 12(1), 6108, https://doi.org/10.1038/s41467-021-26370-0, 2021.

Bonnet, R., McKenna, C. M., & Maycock, A. C.: Model spread in multidecadal North Atlantic Oscillation variability connected to stratosphere–troposphere coupling, Weather and Climate Dynamics, 5(3), 913-926, https://doi.org/10.5194/wcd-5-913-2024, 2024.

Brune, S., Düsterhus, A., Pohlmann, H., Müller, W. A., & Baehr, J.: Time dependency of the prediction skill for the North Atlantic subpolar gyre in initialized decadal hindcasts, Climate Dynamics, 51, 1947-1970, https://doi.org/10.1007/s00382-017-3991-4, 2018.

Cassou, C., Kushnir, Y., Hawkins, E., Pirani, A., Kucharski, F., Kang, I. S., & Caltabiano, N.: Decadal climate variability and predictability: Challenges and opportunities. Bulletin of the American Meteorological Society, 99(3), 479-490, https://doi.org/10.1175/BAMS-D-16-0286.1, 2018.

Cos, P., Marcos-Matamoros, R., Donat, M., Mahmood, R., & Doblas-Reyes, F. J.: Near-term Mediterranean summer temperature climate projections: a comparison of constraining methods, Journal of Climate, 37(17), 4367-4388. https://doi.org/10.1175/JCLI-D-23-0494.1, 2024.

Dommenget, D., & Latif, M.: A cautionary note on the interpretation of EOFs. Journal of climate, 15(2), 216-225, https://doi.org/10.1175/1520-0442(2002)015<0216:ACNOTI>2.0.CO;2, 2002.

Donat, M., Mahmood, R., Cos, J., Ortega, P., & Doblas-Reyes, F. J.: Improving the forecast quality of near-term climate projections by constraining internal variability based on decadal predictions and observations, Environmental Research: Climate, https://doi.org/10.1088/2752-5295/ad5463, 2024.

Enfield, D. B., Mestas-Nuñez, A. M., & Trimble, P. J. : The Atlantic multidecadal oscillation and its relation to rainfall and river flows in the continental US, Geophysical research letters, 28(10), 2077-2080, https://doi.org/10.1029/2000GL012745, 2001.

Eyring, V., Bony, S., Meehl, G. A., Senior, C. A., Stevens, B., Stouffer, R. J., & Taylor, K. E.: Overview of the Coupled Model Intercomparison Project Phase 6 (CMIP6) experimental design and organization, Geoscientific Model Development, 9(5), 1937-1958, https://doi.org/10.5194/gmd-9-1937-2016, 2016.

Eyring, V., Gillett, N. P., Achuta Rao, K. M., Barimalala, R., Barreiro Parrillo, M., Bellouin, N., ... & Sun, Y.: Human influence on the climate system (chapter 3), https://doi.org/10.1017/9781009157896.005, 2021.

https://doi.org/10.5194/egusphere-2025-4463
Preprint. Discussion started: 24 September 2025

García-Serrano, J., Guemas, V., & Doblas-Reyes, F. J.: Added-value from initialization in predictions of Atlantic multi-decadal variability, *Climate Dynamics*, *44*, 2539-2555, https://doi.org/10.1007/s00382-014-2370-7, 2015.

Goddard, L., Kumar, A., Solomon, A., Smith, D., Boer, G., Gonzalez, P., ... & Delworth, T.: A verification framework for interannual-to-decadal predictions experiments, Climate Dynamics, 40, 245-272, https://doi.org/10.1007/s00382-012-1481-2, 2013.

Hegerl, G. C., Ballinger, A. P., Booth, B. B., Borchert, L. F., Brunner, L., Donat, M. G., ... & Weisheimer, A.: Toward consistent observational constraints in climate predictions and projections, Frontiers in Climate, 3, 678109, https://doi.org/10.3389/fclim.2021.678109, 2021.

Hermanson, L., Eade, R., Robinson, N. H., Dunstone, N. J., Andrews, M. B., Knight, J. R., ... & Smith, D. M.: Forecast cooling of the Atlantic subpolar gyre and associated impacts, Geophysical research letters, 41(14), 5167-5174, https://doi.org/10.1002/2014GL060420, 2014.

Hersbach, H., Bell, B., Berrisford, P., Hirahara, S., Horányi, A., Muñoz-Sabater, J., ... & Thépaut, J. N.: The ERA5 global reanalysis, Quarterly Journal of the Royal Meteorological Society, 146(730), 1999-2049, https://doi.org/10.1002/qj.3803, 2020.

Huang, B., Thorne, P. W., Banzon, V. F., Boyer, T., Chepurin, G., Lawrimore, J. H., ... & Zhang, H. M.: Extended reconstructed sea surface temperature, version 5 (ERSSTv5): upgrades, validations, and intercomparisons, Journal of Climate, 30(20), 8179-8205, https://doi.org/10.1175/JCLI-D-16-0836.1, 2017.

Hurrell, J. W., Kushnir, Y., Ottersen, G., & Visbeck, M.: An overview of the North Atlantic oscillation, Geophysical Monograph-American Geophysical Union, 134, 1-36, https://doi.org/10.1029/134GM01, 2003.

Iles, C., & Hegerl, G.: Role of the North Atlantic Oscillation in decadal temperature trends, Environmental Research Letters, 12(11), 114010, https://doi.org/10.1088/1748-9326/aa9152, 2017.

Kushnir, Y., Scaife, A. A., Arritt, R., Balsamo, G., Boer, G., Doblas-Reyes, F., ... & Wu, B.: Towards operational predictions of the near-term climate, Nature Climate Change, 9(2), 94-101, https://doi.org/10.1038/s41558-018-0359-7, 2019.

Lehner, F., Deser, C., Maher, N., Marotzke, J., Fischer, E. M., Brunner, L., ... & Hawkins, E.: Partitioning climate projection uncertainty with multiple large ensembles and CMIP5/6, Earth System Dynamics, 11(2), 491-508, https://doi.org/10.5194/esd-11-491-2020, 2020.

Liné, A., Cassou, C., Msadek, R., & Parey, S.: Modulation of Northern Europe near-term anthropogenic warming and wettening assessed through internal variability storylines, npj climate and atmospheric science, 7(1), 272, https://doi.org/10.1038/s41612-024-00759-2, 2024.

Mahmood, R., Donat, M. G., Ortega, P., Doblas-Reyes, F. J., & Ruprich-Robert, Y.: Constraining decadal variability yields skillful projections of near-term climate change, Geophysical Research Letters, 48(24), e2021GL094915, https://doi.org/10.1029/2021GL094915, 2021.

https://doi.org/10.5194/egusphere-2025-4463
Preprint. Discussion started: 24 September 2025

Mahmood, R., Donat, M. G., Ortega, P., Doblas-Reyes, F. J., Delgado-Torres, C., Samsó, M., & Bretonnière, P. A.: Constraining low-frequency variability in climate projections to predict climate on decadal to multi-decadal timescales–a poor man's initialized prediction system, Earth System Dynamics, 13(4), 1437-1450, https://doi.org/10.5194/esd-13-1437-2022, 2022.

Mariotti, A., & Dell'Aquila, A.: Decadal climate variability in the Mediterranean region: roles of large-scale forcings and regional processes, Climate Dynamics, 38, 1129-1145, https://doi.org/10.1007/s00382-011-1056-7, 2012.

Matei, D., Pohlmann, H., Jungclaus, J., Müller, W., Haak, H., & Marotzke, J.: Two tales of initializing decadal climate prediction experiments with the ECHAM5/MPI-OM model, Journal of Climate, 25(24), 8502-8523, https://doi.org/10.1175/JCLI-D-11-00633.1, 2012.

Menary, M. B., Mignot, J., & Robson, J.: Skilful decadal predictions of subpolar North Atlantic SSTs using CMIP model-analogues, Environmental Research Letters, 16(6), 064090, https://doi.org/10.1088/1748-9326/ac06fb, 2021.

Murphy, A. H.: Skill scores based on the mean square error and their relationships to the correlation coefficient, Monthly weather review, 116(12), 2417-2424, https://doi.org/10.1175/1520-0493(1988)116<2417:SSBOTM>2.0.CO;2, 1988.

Qasmi, S., Cassou, C., & Boé, J.: Teleconnection processes linking the intensity of the Atlantic multidecadal variability to the climate impacts over Europe in boreal winter, Journal of Climate, 33(7), 2681-2700, https://doi.org/10.1175/JCLI-D-19-0428.1, 2020.

Robson, J., Polo, I., Hodson, D. L., Stevens, D. P., & Shaffrey, L. C.: Decadal prediction of the North Atlantic subpolar gyre in the HiGEM high-resolution climate model, Climate dynamics, 50, 921-937, https://doi.org/10.1007/s00382-017-3649-2, 2018.

Sanchez-Gomez, E., Cassou, C., Ruprich-Robert, Y., Fernandez, E., & Terray, L.: Drift dynamics in a coupled model initialized for decadal forecasts, Climate Dynamics, 46(5), 1819-1840, https://doi.org/10.1007/s00382-015-2678-y, 2016.

Schlesinger, M. E., & Ramankutty, N.: An oscillation in the global climate system of period 65–70 years, Nature, 367(6465), 723-726, https://doi.org/10.1038/367723a0, 1994.

Smith, D. M., Eade, R., Scaife, A. A., Caron, L. P., Danabasoglu, G., DelSole, T. M., ... & Yang, X.: Robust skill of decadal climate predictions, Npj Climate and Atmospheric Science, 2(1), 13, https://doi.org/10.1038/s41612-019-0071-y, 2019.

Solaraju-Murali, B., Caron, L. P., Gonzalez-Reviriego, N., & Doblas-Reyes, F. J.: Multi-year prediction of European summer drought conditions for the agricultural sector, Environmental Research Letters, 14(12), 124014, https://doi.org/ 10.1088/1748-9326/ab5043, 2019.

Stephenson, D. B., Pavan, V., Collins, M. M. J. M., Junge, M. M., Quadrelli, R., & Participating CMIP2 Modelling Groups: North Atlantic Oscillation response to transient greenhouse gas forcing and the impact on European winter climate: a CMIP2 multi-model assessment, Climate Dynamics, 27, 401-420, https://doi.org/10.1007/s00382-006-0140-x, 2006.

https://doi.org/10.5194/egusphere-2025-4463
Preprint. Discussion started: 24 September 2025

Sutton, R. T., & Dong, B.: Atlantic Ocean influence on a shift in European climate in the 1990s, Nature Geoscience, 5(11), 788-792, https://doi.org/10.1038/ngeo1595, 2012.

Wilks, D. S.: Forecast verification, In International geophysics (Vol. 100, pp. 301-394). Academic Press, https://doi.org/10.1016/B978-0-12-385022-5.00008-7, 2011.

Yeager, S. G., & Robson, J. I.: Recent progress in understanding and predicting Atlantic decadal climate variability, Current Climate Change Reports, 3, 112-127, https://doi.org/10.1007/s40641-017-0064-z, 2017.

Yeager, S. G., Danabasoglu, G., Rosenbloom, N. A., Strand, W., Bates, S. C., Meehl, G. A., ... & Lovenduski, N. S.: Predicting near-term changes in the earth system: a large ensemble of initialized decadal prediction simulations using the community earth system model, Bulletin of the American Meteorological Society, 99(9), 1867-1886, https://doi.org/10.1175/BAMS-D-17-0098.1, 2018.

Zhang, R., Delworth, T. L., & Held, I. M.: Can the Atlantic Ocean drive the observed multidecadal variability in Northern Hemisphere mean temperature?, Geophysical Research Letters, 34(2), https://doi.org/10.1029/2006GL028683, 2007.