Earth Syst. Dynam., 17, 1299–1314, 2026
https://doi.org/10.5194/esd-17-1299-2026

Earth System Dynamics
Open Access
EGU

Research article

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], **and 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: Rémy Bonnet (bonnet@cerfacs.fr)

Received: 11 September 2025 – Discussion started: 24 September 2025
Revised: 23 April 2026 – Accepted: 7 July 2026 – Published: 22 September 2026

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 historical simulations in 5-, 10-, and 15-year temperature forecasts of summer and winter temperature over Europe. It can also provide larger forecast added value than decadal prediction, for example for land summer temperature over WCE, using surface temperature as predictor. 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 10–20 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 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 information about climate risk evolution over the near-term future. First, the so-called non-initialized ensembles of climate projections provide a seamless evolution from the historical period to the end of the 21st century under different socio-economic scenarios. As such, they encompass the full range of uncertainty including that related to internal climate variability. 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 (Meehl et al., 2021). 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 important that climate information be seamless across timescales, ensuring consistency between historical observations, near-term predictions, and long-term projections (e.g. Nissan et al., 2019; Befort et al., 2022).

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 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). Building on the method from Mahmood et al. (2021), Acosta Navarro et al. (2025) show that this type of constraining method has the potential to fill the gap between seasonal to multi-annual timescales by delivering seamless climate information while having a very low computational cost. 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, Sect. 3 evaluates the blending method, and Sect. 4 presents a summary and discussion of the results.

2 Data and Method

2.1 Datasets

We use 163 non-initialized transient historical simulations (HIST hereafter) from CMIP6 (Eyring et al., 2016) and 92 initialized decadal hindcasts (DEC 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

Table 1. CMIP6 models and associated number of 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).

Model (historical)	Number of members	Model (hindcast)	Simulations
CNRM-CM6-1	30	CNRM-ESM2-1 (5-yr)	25 r(1-15)i1p1f2 r(1-10)i1p1f2
EC-Earth3	15	EC-Earth3 (10-yr)	16 r(1-10i1p1f1) r(6-10)i2p1f1)
MIROC6	50	MIROC6 (10-yr)	10 r(1-10i1p1f1)
MRI-ESM2-0	12	MRI-ESM2-0 (5-yr)	10 r(1-10i1p1f1)
NorCPM1	30	NorCPM1 (10-yr)	20 r(1-10i1p1f1) r(1-10i2p1f1)
IPSL-CM6A-LR	26	IPSL-CM6A-LR (10-yr)	10 r(1-10i1p1f1)

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 were first regridded to the CNRM-CM6-1 atmospheric grid, and only grid points with at least 70 % land fraction were then considered. This choice has been motivated to keep enough grid points along the coast.

A lead-time 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.

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 and Dong, 2012, Qasmi et al., 2020). To estimate the evolution of the AMV, we define the AMV index as the average low-pass filtered annual 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). A Lanczos low-pass filter with a 10-year cutoff period and 11 weights is used for low-pass filtering.

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 15–40° W, 50–60° N region. A Lanczos low-pass filter with a 10-year cutoff period and 11 weights is used for low-pass filtering.

The third index is the 9-year average sea surface temperature pattern correlation at the global scale (GSST). 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.Mahmood et al., 2022).

The fourth index captures the evolution of the winter (December–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). A Lanczos low-pass filter with a 10-year cutoff period and 11 weights is used for low-pass filtering.

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).

2.3 Blending method protocol

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 of the summer land surface temperature average over lead years 1–5 in the WCE region, as defined by the IPCC (Iturbide et al., 2020), starting in 1983 (Fig. 1).

In a preprocessing step, the drift from the hindcast is removed using the method described in Sect. 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-year forecast window [1983–1987] relative to the period 1966–2000.

Two different types of datasets are built. First, $\mathrm{HIST_{OBS}}$, 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, the observational constraint OBS is based on the AMV index (see Sect. 2.3) and $Y = 20$ years, namely 1963–1982. 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 1900–2010 period. The simulations are ranked based on the sum of the two normalized scores and $N = 30$ best simulations are retained.

The second dataset $\mathrm{BLEND_{OBS}}$ is based on a double constraint: the one from observations as for $\mathrm{HIST_{OBS}}$ and a new one from DEC that are available over the forecast period (Fig. 1b). First, 50 simulations that best match the observation index over the $Y = 20$ years before the forecast are selected, applying the same method used for $\mathrm{HIST_{OBS}}$. Then, 30 simulations out of 50 that show the lowest absolute error with respect to the 5-year hindcast ensemble mean surface temperature over the region of interest, are retained. This second step in $\mathrm{BLEND_{OBS}}$ 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 50 members for the first selection and 30 for the second, as this setup provides a relatively strong constraint on HIST while preserving ensemble spread.

We chose to keep the same number of simulations for $\mathrm{HIST_{OBS}}$ and $\mathrm{BLEND_{OBS}}$ to ensure a fair comparison between the two approaches.

Decadal prediction performance is evaluated by comparing the distribution of all ensemble forecast dataset available (Fig. 1c). In this illustrative example, HIST treated here as the benchmark ensemble predicts a slight cooling with a substantial uncertainty assessed by the spread, namely $-0.09 \pm 0.81\,°\mathrm{C}$ (ensemble mean ±5–95th percentile range). DEC has an ensemble mean closer to the observations ($-0.43\,°\mathrm{C}$) than HIST, namely $-0.28 \pm 1.12\,°\mathrm{C}$. The temperature forecast from $\mathrm{HIST_{OBS}}$, $-0.19 \pm 0.72\,°\mathrm{C}$, shows a decrease in spread and a closer ensemble mean to the observation in comparison to HIST, while $\mathrm{BLEND_{OBS}}$, $-0.23 \pm 51\,°\mathrm{C}$, reduces the uncertainty and has an ensemble mean even closer to the observation.

Finally, we introduce three additional ensemble forecasts that are useful for evaluation purposes. First, $\mathrm{HIST_{TAS}}$, 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, $\mathrm{HIST_{Hindcast}}$, which derives from the selection of the 30 simulations from HIST that are closest to the 5-year [1983–1987] hindcast ensemble-mean surface temperature. This allows us to assess the added value using only the second step of the $\mathrm{BLEND_{OBS}}$ dataset.

In this study, we test these methods to predict summer (June–August; JJA) and winter (December–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 $\mathrm{BLEND_{OBS}}$ dataset is produced for each region, since the second constraint is based on the average temperature over the region of interest. The four climate indices described in Sect. 2.2 are tested to constrain the historical simulations, resulting in the $\mathrm{HIST_{OBS}}$ and $\mathrm{BLEND_{OBS}}$ 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 $\mathrm{HIST_{TAS}}$, $\mathrm{HIST_{AMV}}$, $\mathrm{HIST_{SPG}}$, $\mathrm{HIST_{NAO}}$, $\mathrm{HIST_{GSST}}$ and $\mathrm{BLEND_{TAS}}$, $\mathrm{BLEND_{AMV}}$, $\mathrm{BLEND_{SPG}}$, $\mathrm{BLEND_{NAO}}$, $\mathrm{BLEND_{GSST}}$ and $\mathrm{HIST_{Hindcast}}$.

2.4 Evaluation metrics

An important point is to evaluate the added value of BLEND developed in this study compared to previous approaches (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 this end, we perform a retrospective evaluation of temperature forecasts averaged over the three regions of interest. BLEND is applied each year over the 1966–2000 period. For each initialization year, we evaluate winter and summer tem-

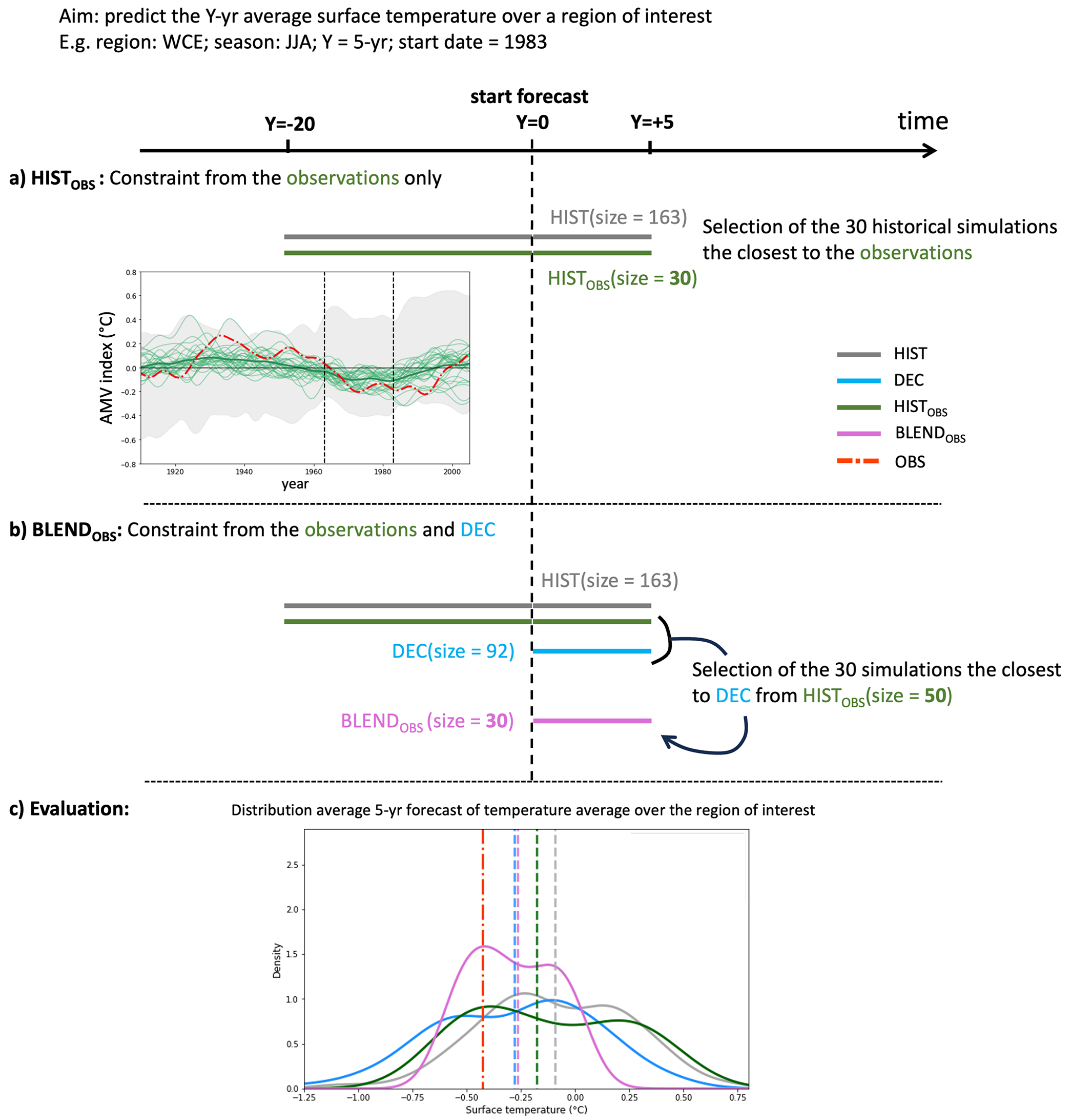

Figure 1. Diagram illustrating the concept and two types of dataset derived from the blending method. For both the historical (HIST) (Table 1) and hindcast (DEC) (Table 2) simulations, the datasets combine all the models (see Sect. 2.1). In the development of HIST$_{OBS}$ **(a)** the AMV index (see Sect. 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 green. In the development of BLEND$_{OBS}$ **(b)**, a first selection of 50 members is then refined to 30 members. The histogram and the ensemble mean (vertical dashed line) of average 5-year forecast surface temperature from the different datasets are evaluated against the ERA5 reanalysis (red dash dotted line) **(c)**.

perature forecasts averaged over 1–5, 1–10, and 1–15-year lead times. As some hindcasts simulations start in January, the first winter is computed using only January and February.

Because the 1966 forecast relies on processed data (e.g., filtering, a 20-year observation-constraining period) starting in 1940 – the initial year of the ERA5 product – the evaluation period begins in 1966.

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 winter and summer temperature forecasts averaged over 1–5, 1–10 and 1–15-year lead times. 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. We use HIST as the reference to assess whether the forecast provides added value beyond external forcing.

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: $\mathrm{RPSS} = 1 - (\mathrm{RPS_{BLEND}}/\mathrm{RPS_{HIST}})$. 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 Cos 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 ($\mathrm{HIST_{OBS}}$ and $\mathrm{BLEND_{OBS}}$) and the historical ensemble. 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 $j = 1$–n years of the evaluation period (1967–2000), following Murphy (1988). It is defined as $\mathrm{MSSS} = 1 - (\mathrm{MSE_{BLEND}}/\mathrm{MSE_{HIST}})$, with the MSE given by

$$\mathrm{MSE} = \frac{1}{N}\sum\nolimits_{j=1}^{n}(F_j - O_j)^2 \qquad (1)$$

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) 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 added value of the forecasts derived from the BLEND method is assessed using a resampling approach. From the 193 HIST members, we randomly select subsets of 30 members (i.e., the same size as the $\mathrm{HIST_{OBS}}$ and $\mathrm{BLEND_{OBS}}$ subsets). This procedure is repeated 1000 times, yielding 1000 subsets of 30 members. The scores are then computed for each subset to obtain a distribution representing the range of scores expected from a 30-member ensemble drawn from HIST due to sampling variability. A subset derived from BLEND is considered significant when its score exceeds the 95th percentile of the resampled score distribution, corresponding to a significance level of 5 %. For the spatial evaluation (Sect. 3.2), only 500 drawings are held due to computational constraints.

For the ensemble spread and the absolute errors, a subset derived from BLEND is considered significant when its median absolute error or ensemble spread, calculated over the evaluation period, is lower than the 5th percentile of the resampled score distribution, corresponding to a significance level of 5 %.

These tests assess whether forecasts derived from the BLEND method significantly outperform forecasts that would be obtained from randomly drawn 30-member subsets of HIST.

3 Results

3.1 Regional performance of BLEND over Europe in winter and summer

A large part of $\mathrm{HIST_{OBS}}$ forecasts all exhibit a significant pronounced reduction in spread relative to HIST and DEC during the testing period (1966–2000), independently of region and season and regardless of the observational index OBS used for the constraint (Fig. 2). The relatively similar results across all $\mathrm{HIST_{OBS}}$ datasets suggest that the selected observational indices are all relevant for constraining the surface temperatures simulated in HIST over Europe. A larger reduction in spread is obtained in all the $\mathrm{BLEND_{OBS}}$ forecasts compared to $\mathrm{HIST_{OBS}}$ forecasts, particularly at the 5-year lead time.

A reduction in both spread and absolute error (Figs. 2 and 3) is observed for the 10- and 15-year forecasts compared with the 5-year forecasts. This is likely due to the stronger influence of external forcing at longer timescales, as well as to the averaging out of higher-frequency internal climate variability in the 10- and 15-year forecasts relative to the 5-year forecasts.

$\mathrm{HIST_{Hindcast}}$ shows the largest spread reduction, as expected by construction. The large pool of historical simulations increases the likelihood of finding good analogues of the hindcast ensemble mean, resulting in a narrower spread. This spread reduction is lost at longer lead times (10- and 15-year forecasts), because the analogues are selected solely based on 5-year hindcasts (Sect. 2.3). This decrease in spread is only associated with a significant decrease in absolute errors for 10-year forecasts over NEU in DJF and 5-year forecasts over WCE in DJF.

The spread and absolute errors in DEC are overall relatively close to the one in HIST. This absence of clear spread and error reduction in DEC may be due to several factors. One possible explanation is poor hindcast performance, potentially of structural origin – for instance related to initialization shocks that can quasi-systematically trigger El Niño events during the first forecast year, as well as to a negative NAO-type mean bias, as reported for the CNRM-CM5 decadal prediction system by Sanchez-Gomez et al. (2016).

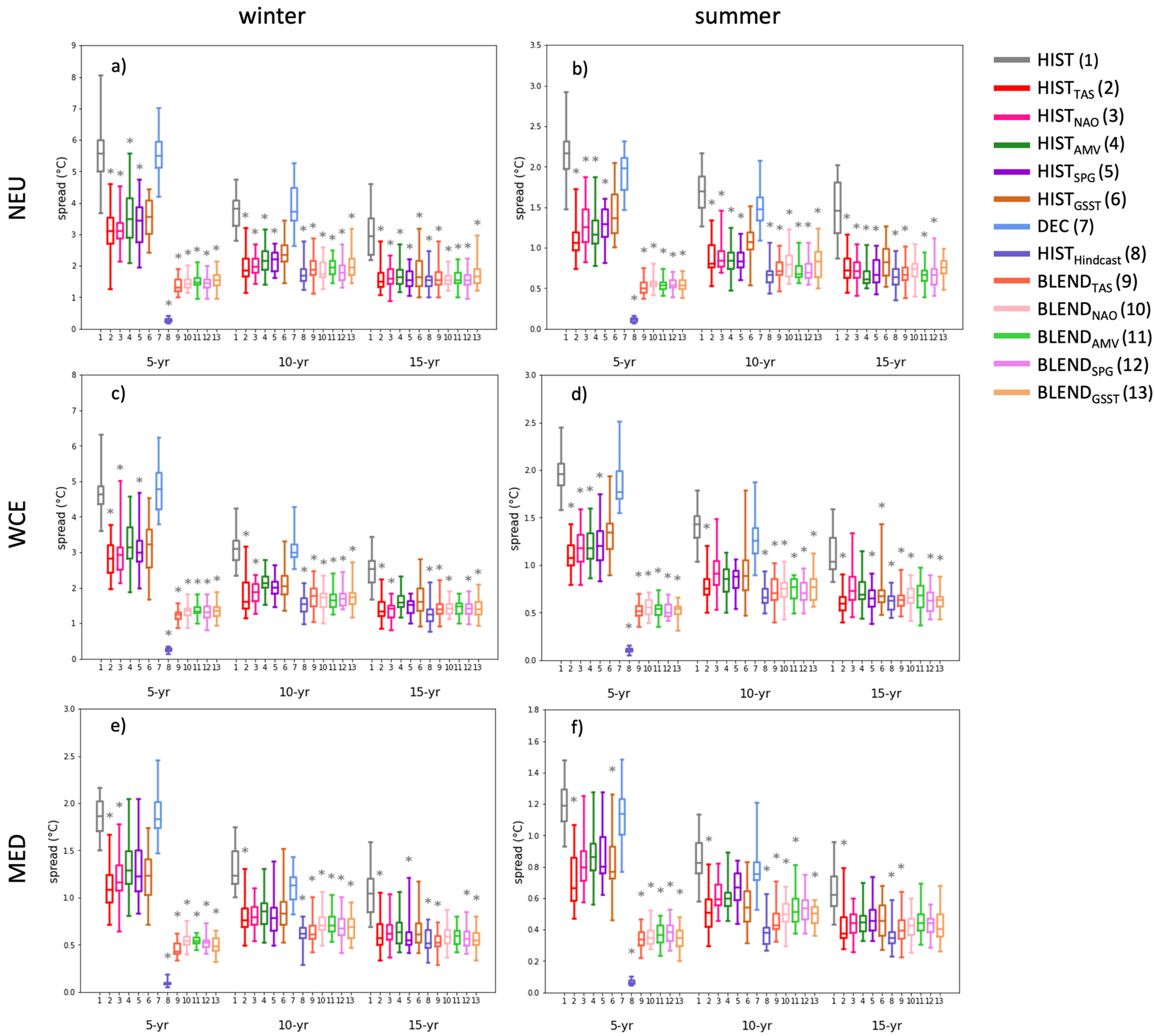

Figure 2. Boxplots of the spread of the average surface temperature for 5-, 10-, and 15-year forecasts in winter **(a, c, e)** and summer **(b, d, f)** over the NEU **(a, b)**, WCE **(c, d)**, and MED **(e, f)** regions. The spread is defined as the difference between the minimum and maximum and is calculated for each year of the retrospective evaluation period (1966–2000). The boxplots display the minimum, 25th percentile, median, 75th percentile, and maximum. Stars indicate cases where the median of the subsets derived from the method is significantly lower than those obtained from randomly drawn subsets of HIST (gray) (p value < 0.05; see Sect. 2.4).

These effects may overshadow the added value of ocean initialization in DEC, thereby degrading hindcast quality at all lead times. Alternatively, the limited spread reduction in DEC compared to HIST may reflect an intrinsic, or "true," climate origin. In particular, decadal variability in European surface temperature is relatively weak compared to the strong chaotic atmospheric variability operating at intraseasonal to interannual timescales, meaning that the predictable signal may be masked by noise. Finally, model deficiencies in the forecasting system may also contribute to this behavior.

Depending on the region and season, some subsets derived from BLEND provide significantly lower absolute errors in the ensemble mean. Over WCE and NEU in winter, 10-year forecasts from BLEND$_{AMV}$ provide a significant decrease in absolute errors (Fig. 3a and c), while also significantly reducing the ensemble spread (Fig. 2a and c). For summer, HIST$_{TAS}$ and HIST$_{AMV}$ show significant reductions in absolute error for 5-year over both WCE and NEU, as well as for 10-year forecasts for HIST$_{TAS}$ and 15-year forecasts for HIST$_{AMV}$ over NEU (Fig. 3b and d). BLEND$_{TAS}$ provides

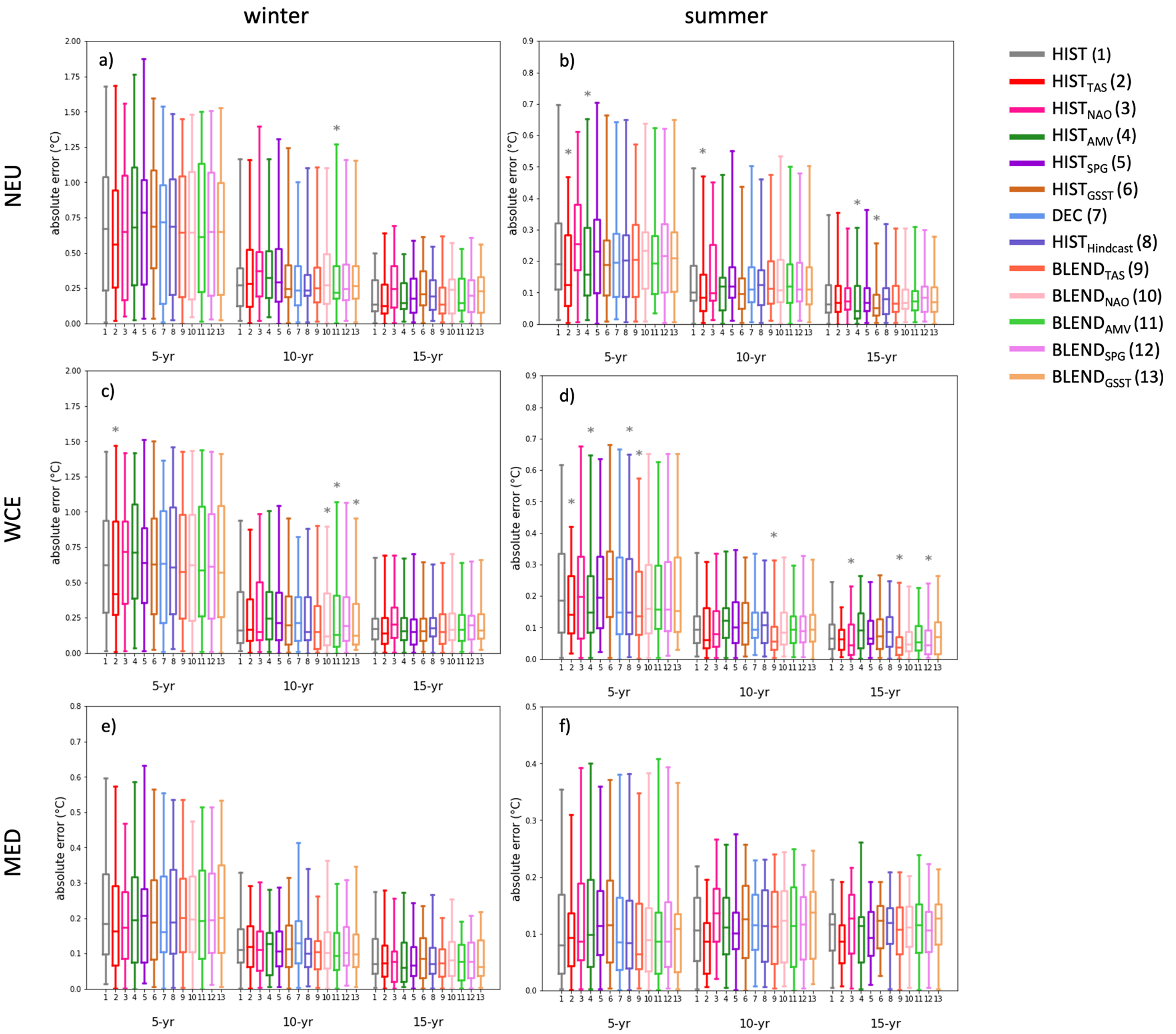

Figure 3. Boxplots of the absolute error of the average surface temperature for 5-, 10-, and 15-year forecasts in winter **(a, c, e)** and summer **(b, d, f)** over the NEU **(a, b)**, WCE **(c, d)**, and MED **(e, f)** regions. The absolute error is calculated each year of the testing period (1966–2000) between the observed surface temperature from ERA5 (Hersbach et al., 2020) and the ensemble mean of the different dataset described Sects. 2.1 and 2.3. The boxplots display the minimum, 25th percentile, median, 75th percentile, and maximum. Stars indicate cases where the median of the subsets derived from the method is significantly lower than those obtained from randomly drawn subsets of HIST (p value < 0.05; see Sect. 2.4).

significant reductions in absolute error for 5-, 10- and 15-year forecasts of summer temperature over NEU (Fig. 3b and d).

It is important to note that a lower median absolute error in the HIST$_{OBS}$ and BLEND$_{OBS}$ ensemble means compared to HIST does not imply that the observations lie 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 in comparison to the historical ensemble mean.

Over MED, none of the BLEND-derived subsets show a significant decrease in the absolute error median for any forecast horizons on both winter and summer (Fig. 3e and f). Only a few BLEND-derived subsets also show a significant reduction in the ensemble spread, particularly for the 15-year forecasts in winter (Fig. 2f). This lack of significance in comparison to randomly drawn 30-member subsets of HIST is likely due to the large influence of the external forcing over this region in summer, especially for 15-year forecasts that limits the added value of BLEND.

The added value of BLEND is further evidenced by the MSSS, RPSS, and CRPSS skill scores, using HIST as the reference (Fig. 4). In particular, the probabilistic scores (RPSS and CRPSS) are sensitive to ensemble spread, meaning that artificial improvements resulting from overconfident spread reduction would be penalized by these metrics.

For NEU in winter, $\mathrm{HIST_{TAS}}$ shows significant added values for the 5-year forecast in terms of MSSS and CRPSS, while $\mathrm{HIST_{Hindcast}}$ achieves the highest scores for the 10-year forecast (Fig. 4a). For the 15-year forecast, $\mathrm{BLEND_{TAS}}$ yields higher RPSS and CRPSS than HIST, although the improvement is not statistically significant. $\mathrm{HIST_{NAO}}$ has overall very low scores for the three metrics, 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 models is shifted southward (Eyring et al., 2021), or to an underestimation of the decadal NAO variability (Bonnet et al., 2024). Another possible explanation is that the 20-year period used prior to the forecast to select members is suboptimal and potentially too long given the decorrelation timescale of the NAO.

For NEU in summer, only $\mathrm{HIST_{TAS}}$ show significant 5- and 10-year forecasts, as well as $\mathrm{HIST_{GSST}}$ and $\mathrm{BLEND_{GSST}}$ for 15-year forecasts (Fig. 4b). Nevertheless, a lot of the BLEND-derived subsets show positive MSSS and CRPSS scores, especially 10- and 15-year forecasts, therefore providing a more accurate forecast in comparison to the reference HIST. Notably, $\mathrm{BLEND_{TAS}}$ exhibits a marked reduction in skill compared with $\mathrm{HIST_{TAS}}$. This suggests that the additional constraint imposed by DEC – based on similarity with the DEC ensemble mean temperature forecast over NEU in JJA from $\mathrm{HIST_{TAS}}$– degrades predictive performance.

Over WCE in winter, only $\mathrm{HIST_{TAS}}$ provides significant 5-year forecasts. For 10-year forecasts, only $\mathrm{BLEND_{TAS}}$ and $\mathrm{HIST_{Hindcast}}$ provide significant 10-year forecasts depending on the score (Fig. 4c).

In summer, both $\mathrm{HIST_{TAS}}$ and $\mathrm{BLEND_{TAS}}$ provide large significant forecast improvements over WCE in comparison to HIST, for the three forecast horizons (Fig. 4d). Scores from the $\mathrm{BLEND_{OBS}}$ datasets are generally higher than those from the corresponding $\mathrm{HIST_{OBS}}$ datasets, highlighting the benefit of combining observations with decadal predictions for this region in summer. Both $\mathrm{BLEND_{AMV}}$ and $\mathrm{BLEND_{SPG}}$ also provide significant improvement in 10- and 15-year forecasts regarding CRPSS (Fig. 4d).

Over MED in winter, $\mathrm{HIST_{NAO}}$ and $\mathrm{HIST_{TAS}}$ provide significant 5-year forecast improvement (Fig. 4e). $\mathrm{BLEND_{TAS}}$ and $\mathrm{BLEND_{AMV}}$ show substantial improvement for 10-year forecasts in terms of MSSS and CRPSS, but no improvement in RPSS (Fig. 4e). These discrepancies highlight that results may vary considerably depending on the evaluation metric used, reinforcing the importance of employing multiple complementary metrics to achieve a more robust assessment. As for other regions, a lot of the BLEND-derived subsets show positive MSSS and CRPSS scores, although not significant. Therefore, although they do not show a clear added value in comparison to randomly drawn 30-member subsets of HIST, they still provide a more accurate forecast in comparison to the reference HIST.

Over MED in summer, several BLEND-derived subsets provide significant improvement in 5-, 10-, and 15-year forecasts in terms of RPSS (Fig. 4f). However, only $\mathrm{HIST_{TAS}}$ shows significant improvement for 5- and 10-year forecasts in terms of MSSS, and for 15-year forecasts in terms of CRPSS. As noted above, this highlights that results can vary substantially depending on the evaluation metric used. The fact that $\mathrm{HIST_{TAS}}$ do not show clear significant added values for 10-year forecasts in terms of RPSS and CRPSS suggests that the improvement in the ensemble mean does not translate into a comparable improvement in the full forecast distribution. This indicates that while $\mathrm{HIST_{TAS}}$ may better capture the central tendency (i.e., deterministic predictability), their 15-year forecasts may be overconfident.

These results are summarized in Figs. S1–S3 in the Supplement. Overall, the results are quite heterogeneous: in some cases, BLEND provides no clear added value relative to HIST – for example, for the 15-year forecast over NEU in winter. In other cases, however, BLEND shows substantial added values in comparison to HIST. This is particularly evident over WCE in summer, where the $\mathrm{HIST_{TAS}}$ and $\mathrm{BLEND_{TAS}}$ dataset shows large improvements for the 5-,10-, and 15-year forecasts. Both $\mathrm{BLEND_{AMV}}$ and $\mathrm{BLEND_{SPG}}$ also provide some significant improvement in 10- and 15-year forecasts. This is also the case for the $\mathrm{HIST_{TAS}}$ dataset, which provides some improvements for summer forecasts over NEU and MED.

Finally, this is also the case for winter forecast temperature over MED, although the BLEND-derived dataset that shows improvement differs depending on the forecast horizon. The $\mathrm{HIST_{TAS}}$ and $\mathrm{HIST_{NAO}}$ and $\mathrm{HIST_{AMV}}$ provide large improvement for 5-year forecasts (Fig. S1), whereas $\mathrm{BLEND_{TAS}}$ and $\mathrm{BLEND_{AMV}}$ show significant improvement for 10-year forecasts (Fig. S2). 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 in modulating the decadal variability of MED winter surface temperatures.

The fact that $\mathrm{HIST_{TAS}}$ provides a significant reduction in absolute errors across several regions may be due either to the use of regional surface temperature inherently capturing the associated low-frequency variability, or to the possibility that the observed surface temperature lies outside, or at the edge of, the distribution of historical simulations, meaning that selecting simulations from this part of the distribution would systematically lead to improved predictions.

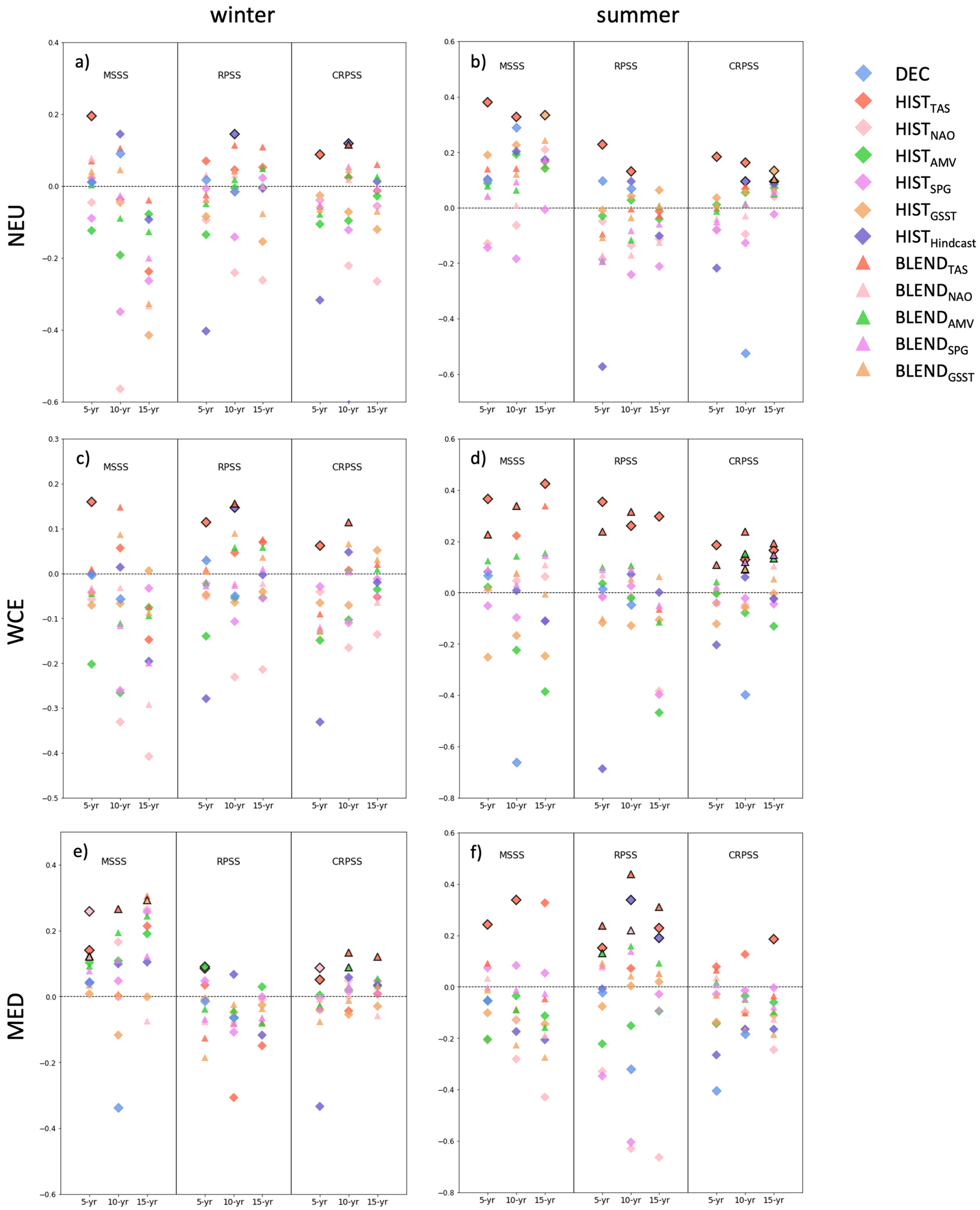

Figure 4. MSSS, RPSS, and CRPSS (see Sect. 2.4) calculated from the 5-, 10-, and 15-year time series of surface temperature forecasts for winter **(a, c, e)** and summer **(b, d, f)** over NEU **(a, b)**, WCE **(c, d)**, and MED **(e, f)**. Scores are shown for the hindcast dataset (see Sect. 2.1) as well as for the dataset derived from BLEND (see Sect. 2.3). HIST is used as the reference; therefore, positive values indicate an improvement relative to HIST. Black edges indicate scores significantly better than those obtained from randomly drawn subsets of HIST (see Sect. 2.4). The scores are calculated over the 1966–2000 period, using the ensemble mean for MSSS and the full ensembles for RPSS and CRPSS.

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

3.2 Spatial evaluation: two case studies

We now evaluate BLEND spatially to assess its added value relative to HIST and DEC across regions. We focus on two case studies: the 10-year forecasts of temperature over WCE in JJA and over MED in DJF, as the results from the evaluation in the previous section show an overall forecast improvement for several BLEND-derived subsets. Specifically, $\mathrm{HIST_{TAS}}$ and $\mathrm{BLEND_{TAS}}$ are evaluated for summer temperature over WCE, while $\mathrm{HIST_{AMV}}$ and $\mathrm{BLEND_{AMV}}$ are evaluated for winter temperature over MED. As a reminder, the second constraint in BLEND is based on the average temperature over the region of interest – here WCE or MED – from the ensemble mean of DEC (see Sect. 2.3).

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 for HIST, DEC, and the BLEND-derived datasets (Figs. 5 and 6a–d). To better identify skill improvements, we therefore use the MSSS and CRPSS metrics (Sect. 2.4).

For the 10-year forecasts of summer temperature, DEC shows slightly positive residual correlations over parts of WCE and NEU and provides forecast improvements relative to HIST across large portions of WCE, as well as parts of MED, in terms of MSSS (Fig. 5e and h). However, these improvements are much more limited regarding CRPSS, with quite small positive values mainly confined to France and parts of MED (Fig. 5h). $\mathrm{HIST_{TAS}}$ and $\mathrm{BLEND_{TAS}}$ largely outperform HIST across much of WCE according to both MSSS (Fig. 5f and g) and CRPSS (Fig. 5i and j). In some areas of WCE, such as Germany and Poland, $\mathrm{BLEND_{TAS}}$ provides greater improvement in the 10-year summer temperature forecasts than $\mathrm{HIST_{TAS}}$.

Interestingly, $\mathrm{HIST_{TAS}}$ also provides substantial forecast improvements over parts of NEU as well as portions of MED (Fig. 5f, i, and j). This suggests that selecting simulations that are closest to the observed summer temperature over WCE yields improved forecasts not only for WCE itself but also for some neighboring regions. This behavior may reflect coherent regional surface temperature variability. Alternatively, as for WCE temperatures, the observed surface temperature may lie outside, or at the edge of, the distribution of historical simulations. In that case, selecting simulations from this part of the distribution would systematically improve predictive skill.

As in summer, strong positive ACC values are visible across much of Europe for HIST and DEC, as well as for $\mathrm{HIST_{AMV}}$ and $\mathrm{BLEND_{AMV}}$ (Fig. 6a–d). However, compared with summer temperatures, parts of the Mediterranean region – particularly Italy, the Balkans, Greece, Turkey, and portions of the Maghreb – show smaller and non-significant ACC values for all datasets.

Although DEC shows slightly positive residual correlation over some parts of MED, it provides only a limited added value relative to HIST over MED and even underperforms over large parts of the region according to both MSSS and CRPSS. This result is consistent with the poor performance observed for the 10-year forecasts of winter temperature averaged over MED (Fig. 4e).

$\mathrm{HIST_{AMV}}$ shows large significant residual correlation over Spain for 10-year forecasts, but rather limited over the rest of MED, as well as other European regions. Some forecast improvements relative to HIST are visible over Spain and parts of the Maghreb in terms of MSSS, but this improvement is no longer visible when considering the CRPSS, particularly over the Maghreb. This suggests that although the $\mathrm{HIST_{AMV}}$ ensemble mean improves upon HIST, the benefit largely disappears once the ensemble spread is taken into account, indicating limited improvement in probabilistic forecast skill.

$\mathrm{BLEND_{AMV}}$ shows larger residual correlation over a large part of the MED regions, with some significant correlation over Spain and a small part of the Maghreb. $\mathrm{BLEND_{AMV}}$ also provides forecast improvement relative to HIST, mainly over parts of the Maghreb for both MSSS and CRPSS, but shows no significant improvement over Spain, although both scores remain positive. The second step of the method improves forecast quality over parts of the Maghreb, as $\mathrm{BLEND_{AMV}}$ exhibits higher CRPSS than $\mathrm{HIST_{AMV}}$. However, forecast skill deteriorates over NEU and WCE relative to HIST, whereas DEC and $\mathrm{HIST_{AMV}}$ provide some added values over NEU. This indicates that the second, region-specific selection step in BLEND can improve consistency with the targeted region while reducing performance elsewhere, as it prioritizes similarity with the regional hindcast signal rather than large-scale temperature coherence.

These results show that the substantial added value of BLEND relative to HIST and DEC for 10-year forecasts of temperature averaged over WCE in summer (Fig. S2), is also visible regionally. This result is more contrasted over MED in winter, where $\mathrm{BLEND_{AMV}}$ provides only limited forecast improvements regionally. These results highlight that performing such spatial evaluations is important before applying the method when focusing on specific regions.

4 Discussion and conclusion

In this study, we introduced 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

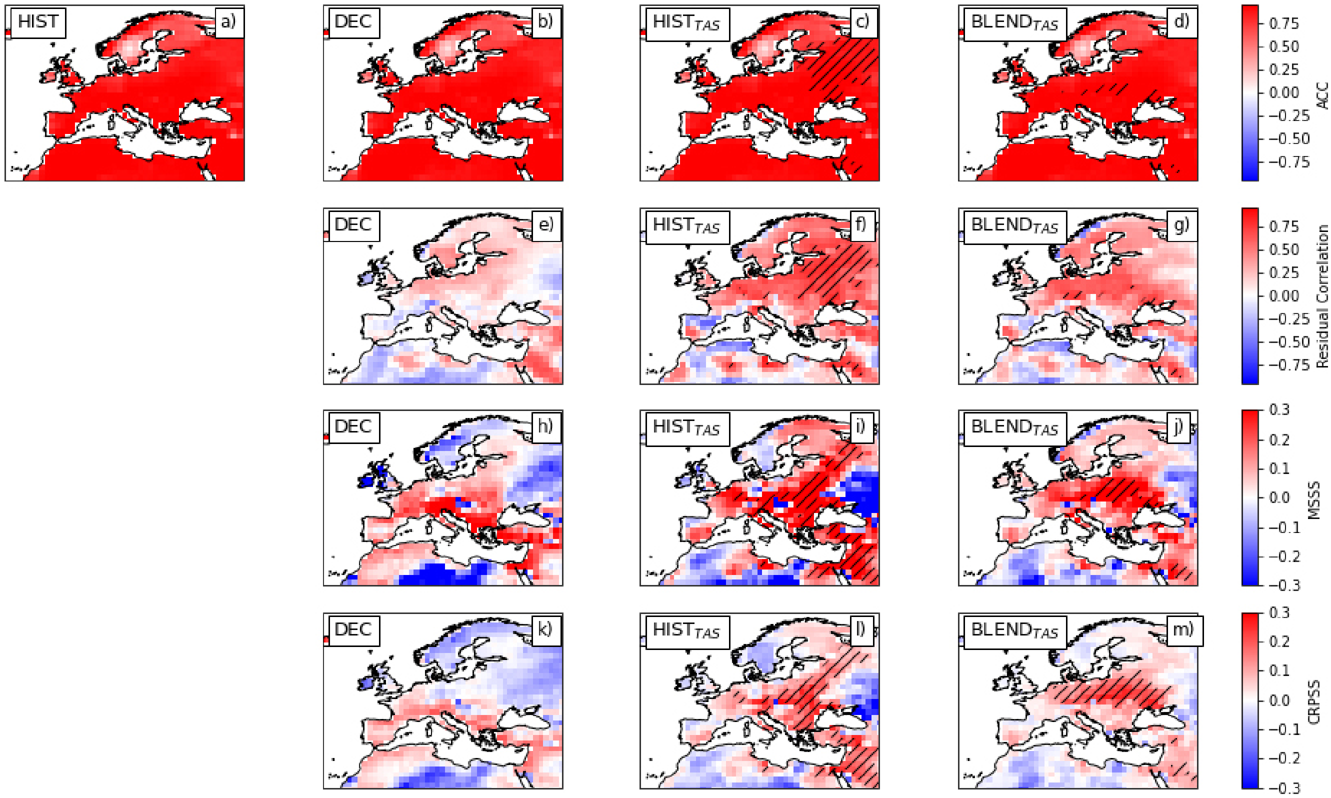

Figure 5. ACC **(a–d)**, Residual correlation **(e–g)**, MSSS **(h–j)**, and CRPSS **(k–m)** calculated from 10-year forecasts of summer surface temperature over the evaluation period 1966–2000 for HIST **(a)**, DEC **(b, e, h, k)**, HIST$_{\mathrm{TAS}}$ **(c, f, i, l)**, and BLEND$_{\mathrm{TAS}}$ **(d, g, j, m)**. For BLEND$_{\mathrm{TAS}}$, the summer surface temperature averaged over WCE is used for the second-step selection. Hatched regions indicate significant improvement in comparison to randomly drawn subsets of HIST (see Sect. 2.4).

affects 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 1966 to 2000.

The performance of the BLEND method depends on the quality of decadal prediction systems, which can still be affected by drift even after applying a lead-time-dependent climatological correction, potentially leading to a suboptimal selection of members. It also depends on the ability of models to correctly capture teleconnections between the climate indices used as predictors and the variable of interest, which may be limited in some regions. Nevertheless, our results show that it can provide substantial improvements of 5-, 10-, and 15-year winter and summer temperature forecasts over Europe relative to HIST, with reduced uncertainty relative to the historical or hindcast ensembles. This added value is also visible regionally, as illustrated in the case study (see Sect. 3.2), where BLEND$_{\mathrm{TAS}}$ approaches show large improvements in 10-year forecasts of summer temperature over WCE in comparison to HIST.

These large improvements in the 5-, 10-, and 15-year forecasts from BLEND in some regions in both winter and summer, relative to the historical ensemble mean – which reflects only externally forced responses – suggest that the method captures part of the internal climate variability. They may also indicate that the method improves the representation of the forced signal.

Although some subsets derived from the blending method showed consistent added values compared to the historical ensemble across the different scores tested, our results also reveal a strong sensitivity to the choice of the skill 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 context 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 an absence of skill. Indeed, the historical ensemble already benefits from some skill because of signals arising 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

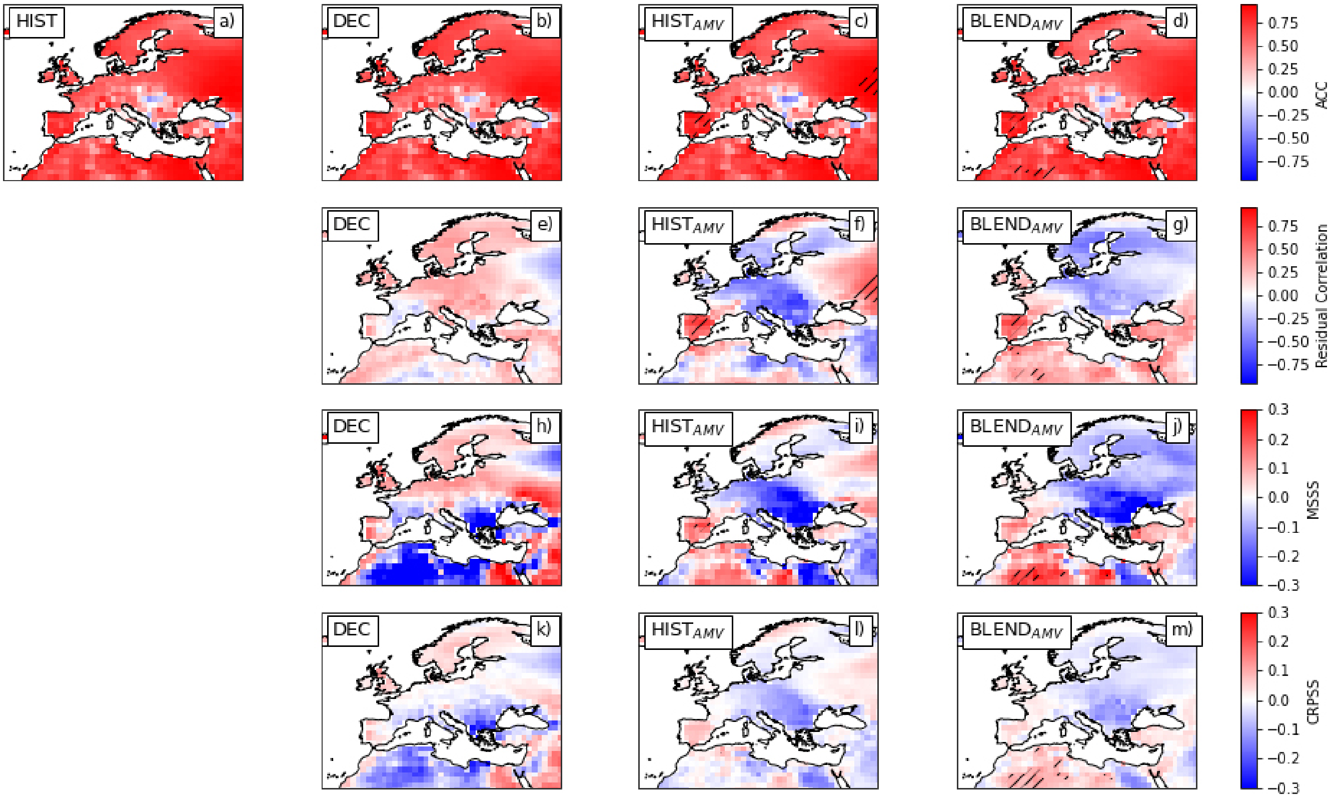

Figure 6. ACC **(a–d)**, Residual correlation **(e–g)**, MSSS **(h–j)**, and CRPSS **(k–m)** calculated from 10-year forecasts of winter surface temperature over the evaluation period 1966–2000 for HIST **(a)**, DEC **(b, e, h, k)**, HIST$_{AMV}$ **(c, f, i, l)**, and BLEND$_{AMV}$ **(d, g, j, m)**. For BLEND$_{AMV}$, the winter surface temperature averaged over MED is used for the second-step selection. Hatched regions indicate significant improvement in comparison to randomly drawn subsets of HIST (see Sect. 2.4).

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 large-scale 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).

This blending method is particularly relevant for climate impact studies and downscaling, as it delivers seamless climate information from the historical period to the end of the 21st century while ensuring consistency across timescales, from weeks to decades. Although the influence of emission pathway choice is relatively limited in the near term (e.g., Liné et al., 2024), consistency in emission pathways should be maintained when applying the method over longer timescales, and blending simulations from different pathways is therefore discouraged.

The implementation of this method requires preliminary work to identify the appropriate predictors, which depend 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 availability. All CMIP6 data are available through the Earth System Grid Federation. ERSSTv5 (https://psl.noaa.gov/data/gridded/data.noaa.ersst.v5.html, last access: April 2026) 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://doi.org/10.24381/cds.adbb2d47, Hersbach et al., 2023).

Supplement. The supplement related to this article is available online at https://doi.org/10.5194/esd-17-1299-2026-supplement.

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 interests. The contact author has declared that none of the authors has any competing interests.

Disclaimer. Views and opinions expressed are 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.

Acknowledgements. The authors gratefully acknowledge the support of Horizon Europe project Impetus4Change (I4C, grant id. 101081555).

Financial support. This research has been supported by the HORIZON EUROPE Climate, Energy and Mobility (grant no. 101081555).

Review statement. This paper was edited by Karin van der Wiel and reviewed by two anonymous referees.

References

Acosta Navarro, J. C., Aranyossy, A., De Luca, P., Donat, M. G., Hrast Essenfelder, A., Mahmood, R., Toreti, A., and Volpi, D.: Seamless seasonal to multi-annual predictions of temperature and Standardized Precipitation Index by constraining transient climate model simulations, Earth Syst. Dynam., 16, 1723–1737, https://doi.org/10.5194/esd-16-1723-2025, 2025.

Alfieri, L., Pappenberger, F., Wetterhall, F., Haiden, T., Richardson, D., and Salamon, P.: Evaluation of ensemble streamflow predictions in Europe, J. Hydrol., 517, 913–922, https://doi.org/10.1016/j.jhydrol.2014.06.035, 2014.

Ambaum, M. H., Hoskins, B. J., and Stephenson, D. B.: Arctic oscillation or North Atlantic oscillation?, J. Climate, 14, 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., and Tibaldi, S.: Decadal predictability of North Atlantic blocking and the NAO, NPJ Climate and Atmospheric Science, 3, 20, https://doi.org/10.1038/s41612-020-0120-6, 2020.

Baker, L. H., Shaffrey, L. C., Sutton, R. T., Weisheimer, A., and Scaife, A. A.: An intercomparison of skill and overconfidence/underconfidence of the wintertime North Atlantic Oscillation in multimodel seasonal forecasts, Geophys. Res. Lett., 45, 7808–7817, https://doi.org/10.1029/2018GL078838, 2018.

Befort, D. J., O'Reilly, C. H., and Weisheimer, A.: Constraining projections using decadal predictions, Geophys. Res. Lett., 47, e2020GL087900, https://doi.org/10.1029/2020GL087900, 2020.

Befort, D., Brunner, L., Borchert, L., O'Reilly, C., Mignot, J., Ballinger, A., Hegerl, G., Murphy, J., and Weisheimer, A.: Combination of decadal predictions and climate projections in time: Challenges and potential solutions, Geophys. Res. Lett., 49, e2022GL098568, https://doi.org/10.1029/2022GL098568, 2022.

Boer, G. J., Smith, D. M., Cassou, C., Doblas-Reyes, F., Danabasoglu, G., Kirtman, B., Kushnir, Y., Kimoto, M., Meehl, G. A., Msadek, R., Mueller, W. A., Taylor, K. E., Zwiers, F., Rixen, M., Ruprich-Robert, Y., and Eade, R.: The Decadal Climate Prediction Project (DCPP) contribution to CMIP6, Geosci. Model Dev., 9, 3751–3777, https://doi.org/10.5194/gmd-9-3751-2016, 2016.

Bonnet, R., Swingedouw, D., Gastineau, G., Boucher, O., Deshayes, J., Hourdin, F., Mignot, J., Servonnat, J., and Sima, A.: Increased risk of near term global warming due to a recent AMOC weakening, Nat. Commun., 12, 6108, https://doi.org/10.1038/s41467-021-26370-0, 2021.

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

Brune, S., Düsterhus, A., Pohlmann, H., Müller, W. A., and Baehr, J.: Time dependency of the prediction skill for the North Atlantic subpolar gyre in initialized decadal hindcasts, Clim. Dynam., 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., and Caltabiano, N.: Decadal climate variability and predictability: Challenges and opportunities, B. Am. Meteorol. Soc., 99, 479–490, https://doi.org/10.1175/BAMS-D-16-0286.1, 2018.

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

Dommenget, D. and Latif, M.: A cautionary note on the interpretation of EOFs, J. Climate, 15, 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., and 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., and Trimble, P. J.: The Atlantic multidecadal oscillation and its relation to rainfall and river flows in the continental US, Geophys. Res. Lett., 28, 2077–2080, https://doi.org/10.1029/2000GL012745, 2001.

Eyring, V., Bony, S., Meehl, G. A., Senior, C. A., Stevens, B., Stouffer, R. J., and Taylor, K. E.: Overview of the Coupled Model Intercomparison Project Phase 6 (CMIP6) experimental design and organization, Geosci. Model Dev., 9, 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., Cassou, C., Durack, P. J., Kosaka, Y., McGregor, S., Min, S., Morgenstern, O., and Sun, Y.: Human Influence on the Climate System, in: Climate Change 2021: The Physical Science Basis. Contribution of Working Group I to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change, edited by: Masson-Delmotte, V., Zhai, P., Pirani, A., Connors, S. L., Péan, C., Berger, S., Caud, N., Chen, Y., Goldfarb, L., Gomis, M. I., Huang, M., Leitzell, K., Lonnoy, E., Matthews, J. B. R., Maycock, T. K., Waterfield, T., Yelekçi, O., Yu, R., and Zhou, B., Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA, 423–552, https://doi.org/10.1017/9781009157896.005, 2021.

García-Serrano, J., Guemas, V., and Doblas-Reyes, F. J.: Added-value from initialization in predictions of Atlantic multi-decadal variability, Clim. Dynam., 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., Kharin, V., Merryfield, W., Deser, C., Mason, S. J., Kirtman, B. P., Msadek, R., Sutton, R., Hawkins, E., Fricker, T., Hegerl, G., Ferro, C. A. T., Stephenson, D. B., Meehl, G. A., Stockdale, T., Burgman, R., Greene, A. M., Kushnir, Y., Newman, M., Carton, J., Fukumori, I., and Delworth, T.: A verification framework for interannual-to-decadal predictions experiments, Clim. Dynam., 40, 245–272, https://doi.org/10.1007/s00382-012-1481-2, 2013.

Hermanson, L., Eade, R., Robinson, N. H., Dunstone, N. J., Andrews, M. B., Knight, J. R., Scaife, A. A., and Smith, D. M.: Forecast cooling of the Atlantic subpolar gyre and associated impacts, Geophys. Res. Lett., 41, 5167–5174, https://doi.org/10.1002/2014GL060420, 2014.

Hersbach, H., Bell, B., Berrisford, P., Hirahara, S., Horányi, A., Muñoz-Sabater, J., Nicolas, J., Peubey, C., Radu, R., Schepers, D., Simmons, A., Soci, C., Abdalla, S., Abellan, X., Balsamo, G., Bechtold, P., Biavati, G., Bidlot, J., Bonavita, M., De Chiara, G., Dahlgren, P., Dee, D., Diamantakis, M., Dragani, R., Flemming, J., Forbes, R., Fuentes, M., Geer, A., Haimberger, L., Healy, S., Hogan, R. J., Hólm, E., Janisková, M., Keeley, S., Laloyaux, P., Lopez, P., Lupu, C., Radnoti, G., de Rosnay, P., Rozum, I., Vamborg, F., Villaume, S., and Thépaut, J.-N. : The ERA5 global reanalysis, Q. J. Roy. Meteor. Soc., 146, 1999–2049, https://doi.org/10.1002/qj.3803, 2020.

Hersbach, H., Bell, B., Berrisford, P., Biavati, G., Horányi, A., Muñoz Sabater, J., Nicolas, J., Peubey, C., Radu, R., Rozum, I., Schepers, D., Simmons, A., Soci, C., Dee, D., and Thépaut, J.-N.: ERA5 hourly data on single levels from 1940 to present, Copernicus Climate Change Service (C3S) Climate Data Store (CDS) [data set], https://doi.org/10.24381/cds.adbb2d47, 2023.

Huang, B., Thorne, P. W., Banzon, V. F., Boyer, T., Chepurin, G., Lawrimore, J. H., Menne, M. J., Smith, T. M., Vose, R. S., and Zhang, H.-M.: Extended Reconstructed Sea Surface Temperature, Version 5 (ERSSTv5): Upgrades, Validations, and Intercomparisons, J. Climate, 30, 8179–8205, https://doi.org/10.1175/JCLI-D-16-0836.1, 2017.

Hurrell, J. W., Kushnir, Y., Ottersen, G., and 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. and Hegerl, G.: Role of the North Atlantic Oscillation in decadal temperature trends, Environ. Res. Lett., 12, 114010, https://doi.org/10.1088/1748-9326/aa9152, 2017.

Iturbide, M., Gutiérrez, J. M., Alves, L. M., Bedia, J., Cerezo-Mota, R., Cimadevilla, E., Cofiño, A. S., Di Luca, A., Faria, S. H., Gorodetskaya, I. V., Hauser, M., Herrera, S., Hennessy, K., Hewitt, H. T., Jones, R. G., Krakovska, S., Manzanas, R., Martínez-Castro, D., Narisma, G. T., Nurhati, I. S., Pinto, I., Seneviratne, S. I., van den Hurk, B., and Vera, C. S.: An update of IPCC climate reference regions for subcontinental analysis of climate model data: definition and aggregated datasets, Earth Syst. Sci. Data, 12, 2959–2970, https://doi.org/10.5194/essd-12-2959-2020, 2020.

Kushnir, Y., Scaife, A. A., Arritt, R., Balsamo, G., Boer, G., Doblas-Reyes, F., Hawkins, E., Kimoto, M., Kolli, R. K., Kumar, A., Matei, D., Matthes, K., Müller, W. A., O'Kane, T., Perlwitz, J., Power, S., Raphael, M., Shimpo, A., Smith, D., Tuma, M., and Wu, B.: Towards operational predictions of the near-term climate, Nat. Clim. Change, 9, 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., Knutti, R., and Hawkins, E.: Partitioning climate projection uncertainty with multiple large ensembles and CMIP5/6, Earth Syst. Dynam., 11, 491–508, https://doi.org/10.5194/esd-11-491-2020, 2020.

Liné, A., Cassou, C., Msadek, R., and Parey, S.: Modulation of Northern Europe near-term anthropogenic warming and wettening assessed through internal variability storylines, NPJ Climate and Atmospheric Science, 7, 272, https://doi.org/10.1038/s41612-024-00759-2, 2024.

Mahmood, R., Donat, M. G., Ortega, P., Doblas-Reyes, F. J., and Ruprich-Robert, Y.: Constraining decadal variability yields skillful projections of near-term climate change, Geophys. Res. Lett., 48, e2021GL094915, https://doi.org/10.1029/2021GL094915, 2021.

Mahmood, R., Donat, M. G., Ortega, P., Doblas-Reyes, F. J., Delgado-Torres, C., Samsó, M., and 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 Syst. Dynam., 13, 1437–1450, https://doi.org/10.5194/esd-13-1437-2022, 2022.

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

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

Meehl, G. A., Richter, J. H., Teng, H., Capotondi, A., Cobb, K., Doblas-Reyes, F., Donat, M. G., England, M. H., Fyfe, J. C., Han, W., Kim, H., Kirtman, B. P., Kushnir, Y., Lovenduski, N. S., Mann, M. E., Merryfield, W. J., Nieves, V., Pegion, K., Rosenbloom, N., Sanchez, S. C., Scaife, A. A., Smith, D., Subramanian, A. C., Sun, L., Thompson, D., Ummenhofer, C. C., and

Xie, S.-P.: Initialized Earth System prediction from subseasonal to decadal timescales, Nat. Rev. Earth Environ., 2, 340–357, https://doi.org/10.1038/s43017-021-00155-x, 2021.

Menary, M. B., Mignot, J., and Robson, J.: Skilful decadal predictions of subpolar North Atlantic SSTs using CMIP model-analogues, Environ. Res. Lett., 16, 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, Mon. Weather Rev., 116, 2417–2424, https://doi.org/10.1175/1520-0493(1988)116<2417:SSBOTM>2.0.CO;2, 1988.

Nissan, H., Goddard, L., de Perez, E. C., Furlow, J., Baethgen, W., Thomson, M. C., and Mason, S. J.:. On the use and misuse of climate change projections in international development, WIRes Clim. Change, 10, e579, https://doi.org/10.1002/wcc.579, 2019.

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

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

Sanchez-Gomez, E., Cassou, C., Ruprich-Robert, Y., Fernandez, E., and Terray, L.: Drift dynamics in a coupled model initialized for decadal forecasts, Clim. Dynam., 46, 1819–1840, https://doi.org/10.1007/s00382-015-2678-y, 2016.

Schlesinger, M. E. and Ramankutty, N.: An oscillation in the global climate system of period 65–70 years, Nature, 367, 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., Delworth, T., Doblas-Reyes, F. J., Dunstone, N. J., Hermanson, L., Kharin, V., Kimoto, M., Merryfield, W. J., Mochizuki, T., Müller, W. A., Pohlmann, H., Yeager, S., and Yang, X.: Robust skill of decadal climate predictions, Npj Climate and Atmospheric Science, 2, 13, https://doi.org/10.1038/s41612-019-0071-y, 2019.

Solaraju-Murali, B., Caron, L. P., Gonzalez-Reviriego, N., and Doblas-Reyes, F. J.: Multi-year prediction of European summer drought conditions for the agricultural sector, Environ. Res. Lett., 14, 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., and 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, Clim. Dynam., 27, 401–420, https://doi.org/10.1007/s00382-006-0140-x, 2006.

Sutton, R. T. and Dong, B.: Atlantic Ocean influence on a shift in European climate in the 1990s, Nat. Geosci., 5, 788–792, https://doi.org/10.1038/ngeo1595, 2012.

Trenberth, K. E. and Shea, D. J.: Atlantic hurricanes and natural variability in 2005, Geophys. Res. Lett., 33, L12704, https://doi.org/10.1029/2006GL026894, 2006.

Wilks, D. S.: Forecast verification, in: International Geophysics, Vol. 100, 301–394, Academic Press, https://doi.org/10.1016/B978-0-12-385022-5.00008-7, 2011.

Yeager, S. G. and 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., Karspeck, A. R., Lindsay, K., Long, M. C., Teng, H., and 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, B. Am. Meteorol. Soc., 99, 1867–1886, https://doi.org/10.1175/BAMS-D-17-0098.1, 2018.

Zhang, R., Delworth, T. L., and Held, I. M.: Can the Atlantic Ocean drive the observed multidecadal variability in Northern Hemisphere mean temperature?, Geophys. Res. Lett., 34, https://doi.org/10.1029/2006GL028683, 2007.