

Article

Verification and Bias Adjustment of ECMWF SEAS5 Seasonal Forecasts over Europe for Climate Service Applications

Alice Crespi ^{1,*}, Marcello Petitta ^{1,2}, Paola Marson ³, Christian Viel ³ and Lucas Grigis ³¹ Institute for Earth Observation, Eurac Research, 39100 Bolzano, Italy; marcello.petitta@eurac.edu² Laboratory Climate Modelling, SSPT-MET-CLIM, ENEA, 00123 Rome, Italy³ Directorate of Climatology and Climate Services, Météo-France, 31057 Toulouse, France; paola.marson@gmail.com (P.M.); christian.viel@meteo.fr (C.V.); lucas.grigis@gmail.com (L.G.)

* Correspondence: alice.crespi@eurac.edu

Abstract: This work discusses the ability of a bias-adjustment method using empirical quantile mapping to improve the skills of seasonal forecasts over Europe for three key climate variables, i.e., temperature, precipitation and wind speed. In particular, the suitability of the approach to be integrated in climate services and to provide tailored predictions for local applications was evaluated. The workflow was defined in order to allow a flexible implementation and applicability while providing accurate results. The scheme adjusted monthly quantities from the seasonal forecasting system SEAS5 of the European Centre for Medium-Range Forecasts (ECMWF) by using ERA5 reanalysis as reference. Raw and adjusted forecasts were verified through several metrics analyzing different aspects of forecast skills. The applied method reduced model biases for all variables and seasons even though more limited improvements were obtained for precipitation. In order to further assess the benefits and limitations of the procedure, the results were compared with those obtained by the ADAMONT method, which calibrates daily quantities by empirical quantile mapping conditioned by weather regimes. The comparable performances demonstrated the overall suitability of the proposed method to provide end users with calibrated predictions of monthly and seasonal quantities.

Keywords: seasonal forecasts; bias adjustment; verification; climate services

Citation: Crespi, A.; Petitta, M.; Marson, P.; Viel, C.; Grigis, L. Verification and Bias Adjustment of ECMWF SEAS5 Seasonal Forecasts over Europe for Climate Service Applications. *Climate* **2021**, *9*, 181. <https://doi.org/10.3390/cli9120181>

Academic Editor: Chaofan Li

Received: 28 October 2021
Accepted: 7 December 2021
Published: 10 December 2021

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2021 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Seasonal climate forecasting systems are primary tools to derive predictions of the seasonal climatic conditions several months in advance and, due to recent improvements in forecasting, they are gaining relevance as support to decision-making processes in a wide range of sectors, such as energy, agriculture, water and risk management [1,2]. Several centers worldwide, such as the National Centers for Environmental Prediction (NCEP) and the European Center for Medium-Range Weather Forecasts (ECMWF), provide seasonal climate predictions using fully coupled ocean-atmosphere general circulation models (GCMs). However, the effective spatial resolutions of global models, in the order of 100–300 km, are too coarse to provide suitable information for the regional and local scales generally required by sectoral applications. Such scale mismatch can result in systematic errors when model simulations are compared to observations, often preventing their direct usage by end users [3,4].

In order to improve the local-scale representativeness of predictions and to provide tailored information supporting decision-making processes, both dynamical and statistical downscaling techniques as well as bias-adjustment schemes have been developed. Dynamical downscaling implies the use of regional climate models (RCMs), which run on finer spatial resolutions, generally in the order of 10–20 km, over a limited domain and are initialized and driven at the boundaries by GCM outputs [5]. Dynamical downscaling is computationally demanding, and the downscaled predictions can still be affected by biases

and can thus require additional post-processing [6,7]. Statistical downscaling includes a large range of techniques of different complexities that are based on the relationship between large-scale climate predictors and local-scale observed predictands [8]. It is beneficial in a wide range of applications since it is less computationally demanding than dynamical methods and has been found to perform comparably in most cases [2,3]. However, the choice of predictors is crucial, and it can significantly affect the variability and accuracy of results. The use of large sets of predictors could increase result uncertainty and reduce their interpretability [9]. Bias-adjustment methods are post-processing techniques that compare coarse model predictions with reference fields over a calibration period and derive the proper corrections in order to match the statistical properties of model outputs with those of local climatological values [10]. They include corrections of the mean and more complex adjustments of the distribution. These methods can be used in combination with downscaling or be applied directly to GCM outputs [11,12]. Bias adjustment was originally introduced to post-process climate model projections (see e.g., [13]) and has been recently tested and applied in the context of seasonal forecasts [14–16]. It provides the advantage of being easily applicable and adaptable to different types of variables and temporal resolutions. However, most inter-comparison and evaluation studies to date have focused only on temperature or precipitation, while fewer works have discussed the bias adjustment of seasonal predictions for other climate variables, such as wind speed [17,18].

The integration of post-processing schemes that enhance predictions and require relatively low computational costs could be beneficial for climate services together with the access to robust climate information for end users [19]. In recent years, international initiatives have been developed in order to simplify the retrieval and processing of climate information, including seasonal predictions. For instance, the Copernicus Climate Change Service (C3S) provides access through the Climate Data Store (CDS, <https://cds.climate.copernicus.eu/#!/home>, accessed on 26 October 2021) to a wide archive of global and European climate data, most of which are already targeted to the requirements of sectoral applications. However, efforts are still needed in the context of seasonal forecasts to provide users with technical tools and information on forecast performances and to tailor predictions on application domains. The bias removal is still an essential requirement for users when forecasts are included as input in impact models or when predicted values are used in the assessment of critical threshold exceedances supporting the decision-making process in specific sectors. However, it is important to select suitable reference datasets for estimating adjustment in order to avoid potential misrepresentation and errors, as discussed in detail in Ehret et al. (2012) and Maraun (2016) [20,21]. Bias-adjustment schemes require the use of reference datasets of a proper temporal length (at least 10 years) for ensuring a robust estimation of correction factors. In addition, the quality of reference data, either observations or reanalyses, influences the calibrated results and must be carefully ensured. In some cases, post-processing may compromise the physical consistency of climate variables and lead to unrealistic values (e.g., relative humidity above 100% or minimum temperature greater than maximum temperature) and requires a final check of outcomes. Finally, the main hypothesis on which bias-adjustment procedures are based is that data can always be described by the same distribution and the biases remain stable. In the case of seasonal forecasts, this assumption is valid, whereas for climate projections, more specific approaches are needed [22].

Several international projects have recently been undertaken with the aim of fostering the use of forecast information for the improvement of sectoral applications. In particular, the EU H2020 project SECLI-FIRM (The Added Value of Seasonal Climate Forecasts for Integrated Risk Management Decisions) focused on assessing the impact of improved climate forecasts on operational planning for specific sectors, such as renewable energy production (<http://www.secli-firm.eu/>, accessed on 26 October 2021). The SECLI-FIRM project was characterized by nine case studies in which different applications of seasonal forecast for energy were developed. The project stakeholders required the definition of a general approach for bias adjusting and downscaling seasonal forecast data. In this

framework, we implemented a post-processing method implying the bias adjustment of seasonal forecasts and tested it over Europe. The scheme was chosen based on the need to define a suitable compromise between calibration accuracy and its flexibility to be adapted and run in several end-user applications. The method was applied and validated on ECMWF's global forecasting system, SEAS5 [23], over Europe for 2-m air temperature, precipitation and 10-m wind speed, using ERA5 as reference. Since most applications require only monthly or longer aggregated predictions, the calibration was applied directly to monthly forecasts. The performances of the presented approach were compared with those provided by the ADAMONT statistical scheme, which was also included in the SECLIFIRM framework. ADAMONT was developed by Météo France and performs a daily bias adjustment of model data conditioned by atmospheric patterns [24]. The intercomparison was conducted to derive further insights into the features of the proposed monthly forecast calibration using a bias-adjustment approach applied to a finer temporal resolution and including additional assumptions, such as the potential influence of atmospheric drivers on the distribution of climate variables and model biases.

2. Materials and Methods

2.1. Data

ECMWF SEAS5 seasonal forecasts used for this study were retrieved from CDS C3S. We focused on monthly aggregated reforecasts (or hindcasts) of 2-m mean air temperature, total precipitation and 10-m wind speed over the period 1993–2016. The dataset covered the global surface on a $1^\circ \times 1^\circ$ regular grid and each forecast was composed of 25 members (i.e., independent realizations of the forecast) and 6 lead times (i.e., the predictions provided for the month of initialization and the following 5 months).

The reference data used for the bias adjustment and for the skill assessment of forecasts were derived from the fifth generation of the ECMWF global reanalysis, ERA5 [25]. ERA5 spans the period 1950–present with hourly temporal resolution and is provided on a regular $0.25^\circ \times 0.25^\circ$ grid. The monthly aggregates of ERA5 for the three variables were derived from the CDS C3S service for the same period spanned by the forecasts. ERA5 global fields were cropped over an extended European area (26.5° N– 72.5° N, 22° W– 45.5° E), which was used as target domain for the assessment of both raw and calibrated ECMWF SEAS5 seasonal predictions. In this study, the reanalysis was used as it is the best alternative to observations over the large European domain considered. This allowed us to obtain a more reliable assessment of ECMWF SEAS5 forecasts and calibration methods, which is independent from the heterogeneity in space and time of in-situ data availability. Moreover, it improved the replicability of the methodology over other regions where no or scarce observations are available.

2.2. The Bias Adjustment

The proposed post-processing scheme is a two-step procedure combining the spatial disaggregation of the $1^\circ \times 1^\circ$ forecast fields to the target ERA5 grid and the empirical quantile mapping (QM). The method will be hereafter called B-QM.

The forecast spatial disaggregation was performed by means of a bilinear interpolation separately applied to each monthly prediction, ensemble member and lead time.

The QM adopted for the bias adjustment is a widely applied method to post-process climate model simulations that reduces the mismatch between the coarser model outputs and the spatial scales of interest [26]. QM adjusts the modeled values to the reference data by matching the cumulative density function (CDF) of the simulations at each target location. More specifically, modeled and reference distributions are matched by establishing a quantile-dependent correction function that translates simulated quantiles into their reference counterparts. This function is then used to translate the modeled time series into bias-adjusted values with a distribution representative of the reference data, which is ERA5 in this case. QM was applied separately for each month and lead time. The transfer functions were obtained for each 0.25-grid cell from the entire forecast ensemble over the

period 1993–2016, i.e., 25 realizations times the forecast instances for each month and then applied to each individual member. In order to avoid overfitting due to the small sample size of monthly values included in the calibration, the quantile adjustment was computed by considering deciles instead of centiles and applied by linearly interpolating the empirical distribution. Negative values in precipitation and wind speed, if any, were set to zero before QM and a wet-day correction equalizing the fraction of days with precipitation between the observed and the modelled data was applied.

The QM adjustment was performed under a leave-one-year-out (LOYO) cross-validation scheme in order to avoid artificial skills in result assessment, which can be particularly relevant for samples of small sizes [16]. The implemented QM scheme was based on the R package *qmap* [27].

2.3. The Skill Assessment

The skills of ECMWF SEAS5 seasonal predictions over Europe were assessed before and after the bias adjustment using ERA5 as reference. The evaluation was performed over the 1993–2016 period for seasonal aggregates for winter (December to February, DJF), spring (March to May, MAM), summer (June to August, JJA) and autumn (September to November, SON) for a one-month lead time, e.g., the forecasts for JJA were initialized in May. In order to allow a more direct comparison, unadjusted forecasts were verified on the spatially disaggregated fields at 0.25° resolution.

Temperature and precipitation forecasts were assessed over land areas only, while the evaluation of wind speed was extended over the sea grid points due to the relevance of offshore wind, especially for the renewable energy sector. Moreover, the obtained skills over Europe for each variable were grouped by sub-regions based on the IPCC European sub-regional classification [28]. This evaluation was designed to better identify the areas most prone to low forecasting system performance and to highlight their seasonal dependencies.

To further assess the robustness of B-QM, the skills were compared to those of SEAS5 predictions corrected using the ADAMONT method [24]. ADAMONT was originally introduced to adjust climate model projections and then adapted to process seasonal forecasts. It performs forecast adjustment on a daily basis by applying a QM conditioned by weather regimes. The weather regimes were based on the classification of daily large-scale recurrent states in the circulation over a wide box spanning North Atlantic and Europe. The patterns were identified by grouping together 4 similar daily fields to create clusters of different fields. The classification was made by clustering the daily mean sea-level pressure (MSLP) anomalies (compared to the 1981–2010 monthly climatology) of ERA5. The motivation for using the weather regimes is the impact that different regimes of circulation can have on the distribution of environmental variables at the surface, as well as that of forecast biases. After the classification of daily fields, QM was separately calculated for groups of days to which the same weather regime is attributed. The bias adjustment in ADAMONT was applied directly to the original forecast grid without any previous spatial interpolation to the target grid. ERA5 was used as reference, making the outcomes comparable with those of B-QM.

The intercomparison was performed on temperature and precipitation aggregates for DJF and JJA for the one-month lead time over the 1993–2016 period for the European subdomain (35.5° N–59.5° N and 10.5° W–19.5° E) covered by ADAMONT. The analysis did not include wind speed, since the two sets of calibrated forecasts were not directly comparable. ADAMONT uses daily wind speed derived by averaging 6-h values, which are, in turn, the mean of 6-h *u* and *v* components, while B-QM uses monthly mean wind speed directly from both seasonal forecasts and ERA5. Such difference in the processing and retrieval of monthly wind speed prevented the equal comparison and validation of the two datasets.

2.4. The Verification Metrics

Deterministic and probabilistic metrics were used to measure both the performances of the forecasted ensemble mean and the event representativeness in the forecasted distribution.

Mean error (ME) and mean absolute error (MAE) report the accuracy of the ensemble mean predictions, i.e., the deviation from the reference fields:

$$ME = \frac{1}{N} \sum_{i=1}^N (fcst_i - ref_i) \quad (1)$$

$$MAE = \frac{1}{N} \sum_{i=1}^N |fcst_i - ref_i| \quad (2)$$

where $fcst_i$ is the ensemble mean prediction for the temporal instance i , ref_i is the corresponding ERA5 value and N is the total number of forecasted instances.

The Pearson correlation (CORR) assesses the strength of association between the interannual time series of the ensemble mean forecast and the reference:

$$CORR = \frac{\sum (fcst - \overline{fcst}) (\overline{ref} - ref)}{\sqrt{\sum (fcst - \overline{fcst})^2} \sqrt{\sum (\overline{ref} - ref)^2}} \quad (3)$$

The spread to error ratio (SPR) is a measure of the forecast reliability and quantifies the ability of the ensemble forecast to represent the forecast error in a statistical sense:

$$SPR = \frac{\sigma}{RMSE} \quad (4)$$

where σ is the intra-ensemble standard deviation and RMSE is the root mean squared error of the ensemble mean forecast.

The ranked probability score (RPS) assesses the ability of forecasts to predict the category the reference falls into. Both forecast and reference are separated into M categories, in this case tercile-based categories, and the squared difference between the CDFs of forecast and reference is calculated:

$$RPS = \frac{1}{M-1} \sum_{m=1}^M (CDF_{fcst,m} - CDF_{ref,m})^2 \quad (5)$$

The Continuous Ranked Probability Score (CRPS) is the continuous version of RCP and accounts for the integrated squared difference between the cumulative distribution functions of prediction and reference for a continuous variable. CRPS corresponds to MAE for the deterministic forecast.

Based on these scores, the corresponding skills (RPSS and CRPSS) of the calibrated forecast with respect to those of a reference forecast were derived. More specifically, the added value of the forecasting system was estimated with respect to a climatological forecast derived from the reanalysis:

$$RPSS = 1 - \frac{RPS}{RPS_{clim}} \quad (6)$$

$$CRPSS = 1 - \frac{CRPS}{CRPS_{clim}} \quad (7)$$

The significance of derived RPSS and CRPSS was computed using the standard error of the skill score estimated by the propagation of error and a 5% significant level was used to identify the statistically significant skill improvements.

In addition, the relative operating characteristic skill score (ROCSS) was computed to verify the ability of tercile-based categorical forecasts to discriminate between alternative outcomes with respect to a climatological forecast.

All metrics are summarized in Table 1 together with the corresponding range of possible values and main rules for their interpretation.

Table 1. Summary of forecast verification metrics used in this study, reporting the range of possible values, the score representing a perfect forecast and the main guidelines for result interpretation. In the cases of ME and MAE no interpretation notes are needed.

Metric	Range	Perfect Score	Interpretation
ME	$(-\infty; +\infty)$	0	–
MAE	$[0; +\infty)$	0	–
CORR	$[-1; +1]$	1	>0 correlation <0 anti-correlation
SPR	$(-\infty; +\infty)$	1	>1 overdispersion <1 overconfidence
RPSS/CRPSS/ROCSS	$(-\infty; 1]$	1	>0 skill gain <0 skill degradation

3. Results

3.1. ECMWF SEAS5 over Europe

The ensemble ME for mean temperature was mostly negative over Europe for MAM and DJF predictions with average values of -1.4 and -1.6 °C, respectively, revealing an overall temperature underestimation in predictions. The greatest discrepancies were located over Norway, Turkey and in lower elevation areas close to the Alps. Conversely, systematic overestimations of up to $+8$ °C were obtained in all seasons over the main European mountainous areas, especially the Alps, Pyrenees, Carpathians and near Caucasus (Figure 1a).

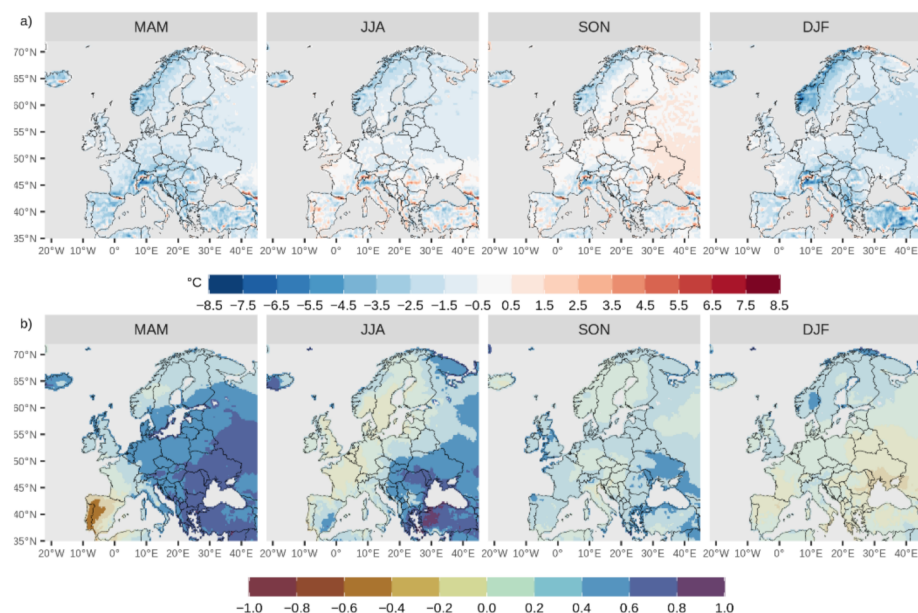


Figure 1. Spatial distribution of (a) ME and (b) CORR for unadjusted ECMWF SEAS5 ensemble forecasts of 2-m mean air temperature (one-month lead time) over Europe (land only) with respect to ERA5 reanalysis over 1993–2016. Significant correlation (p -value < 0.05) is reported by unshaded colors.

The regional dependency of the temperature bias was evidenced by comparing ME distributions of European sub-regions (Figure 2). The spread of forecast errors was larger and mostly negative for Alpine regions and southern areas, while more pronounced seasonal differences and sharper ME distributions were reported for the other sub-regions. CORR showed an overall agreement between forecasts and ERA5 in all seasons, except for DJF, when anticorrelation was mostly reported (Figure 1b). The highest and most significant correlation was obtained in MAM and JJA on the eastern part of the domain, particularly in the areas surrounding the Black Sea, and in Iceland. However, strong anticorrelated behavior was observed over the Iberian Peninsula in MAM and no correlation over the Atlantic coast was obtained for JJA forecasts. As for ME, the CORR seasonal distributions split by sub-regions highlighted the spatial and seasonal dependencies of forecast skills and the gradual correlation decrease in DJF predictions for all areas (Figure S1). The spatial distribution of correlation coefficients was largely in agreement with those of ROCSS values for the first and third terciles (Figure 3). The best performances were shown in spring and summer also in terms of discrimination, with highly positive ROCSS for both terciles.

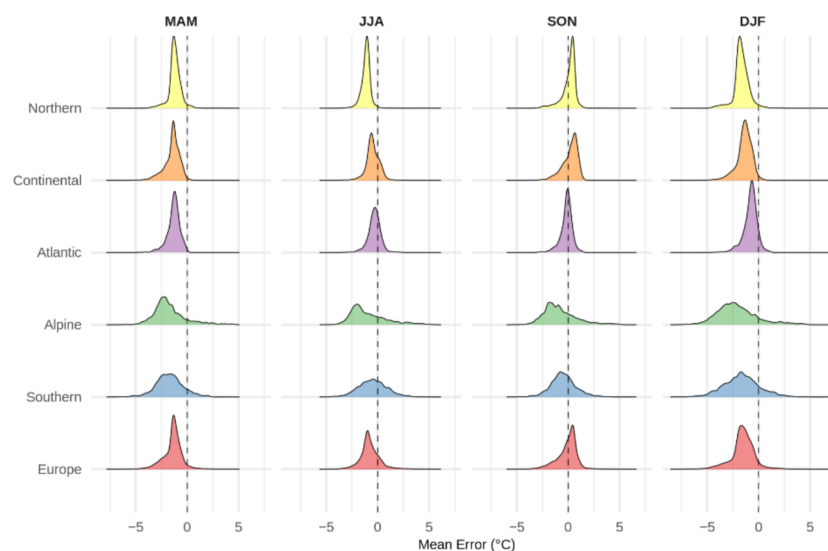


Figure 2. ME distributions of unadjusted ECMWF SEAS5 ensemble seasonal forecasts of 2-m mean air temperature (one-month lead time) for European sub-regions and whole domain. ME was computed with respect to ERA5 over 1993–2016.

ECMWF SEAS5 underestimated ERA5 precipitation over most areas in all seasons with ME values locally below -100 mm. The greatest discrepancies were observed for JJA predictions in the Mediterranean region, and in correspondence of the main mountainous areas (Figure 4a). As for temperature, error distributions exhibited a larger spread for Alpine and southern sub-regions with mostly negative values. However, it is worth noting that in relative terms some relevant overestimations were locally observed throughout the domain, with positive tails in ME distributions up to $+200\%$ in summer for all sub-regions (Figure 5). Precipitation predictions exhibited lower interannual correlation with ERA5 in all seasons except for MAM (Figure 4b). The low skills of precipitation forecasts as well as the absence of a specific spatial dependency were also confirmed by the sub-regional CORR distributions (Figure S2).

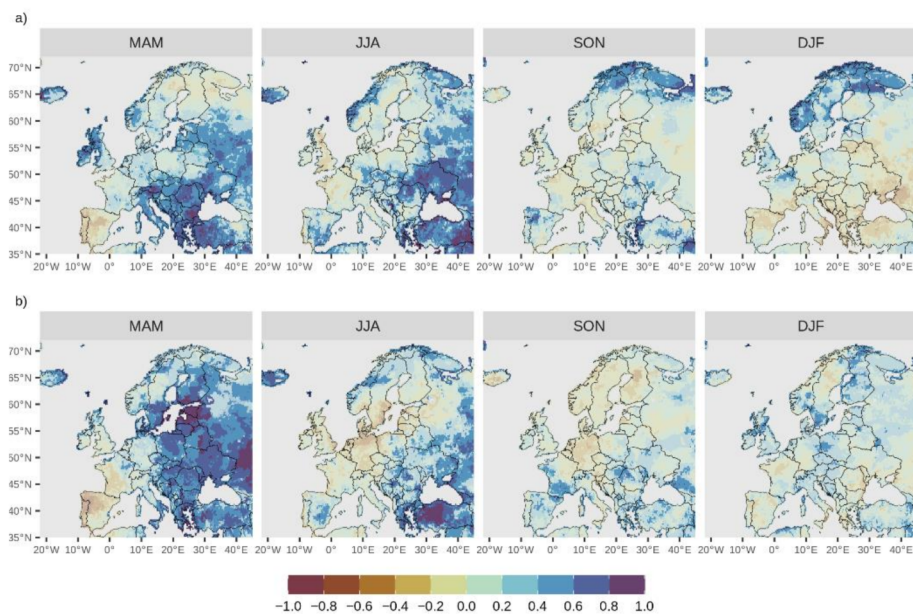


Figure 3. Spatial distribution of ROCSS for (a) first and (b) third tercile for unadjusted ECMWF SEAS5 ensemble seasonal forecasts of 2-m mean air temperature (one-month lead time) over Europe (land only) with respect to the climatological forecast over 1993–2016. Unshaded colors indicate grid cells with significant positive scores.

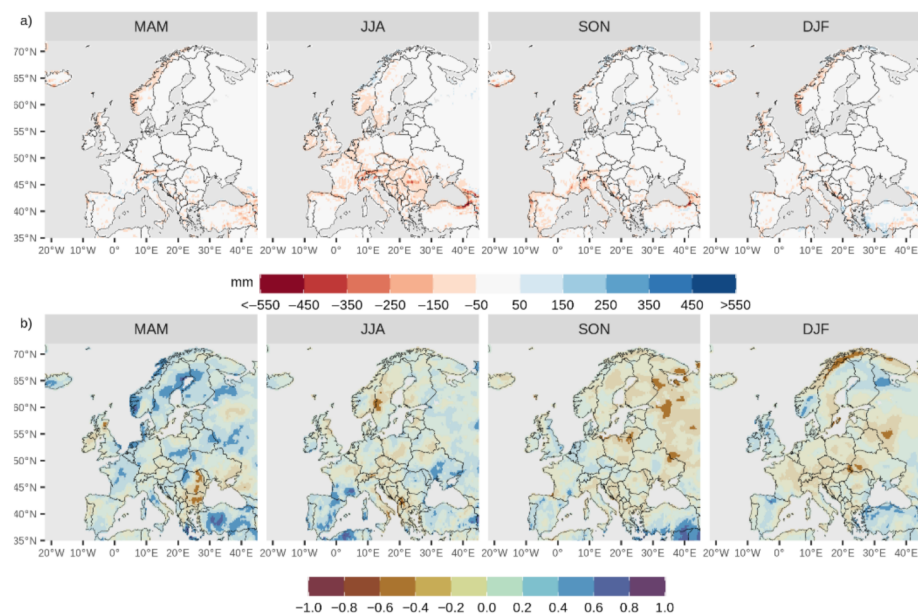


Figure 4. Spatial distribution of (a) ME and (b) CORR for unadjusted ECMWF SEAS5 ensemble forecasts of precipitation (one-month lead time) over Europe (land only) with respect to ERA5 reanalysis over 1993–2016. Significant correlation (p -value < 0.05) is reported by unshaded colors.

ME reported an overall overestimation of wind speed in all seasons and throughout the domain, especially for sea areas with values up to 3 m/s (Figure 6a). The error distributions were rather invariant over seasons. They were sharper for northern and continental Europe and, in most cases, had a longer right tail, especially in the Atlantic and southern sub-regions (Figure 7). The highest discrepancies with respect to the reanalysis were localized along the coastline with both positive and negative biases. The correlation patterns showed a relevant spatial heterogeneity. Positive and significant correlation coefficients were obtained mostly in the north in MAM and over central Europe and the Atlantic area in

DJF, while negative coefficients were highlighted in SON (Figure 6b). The high spatial variability of correlation coefficients for wind speed forecasts was also reflected in the sub-regional distributions (Figure S3).

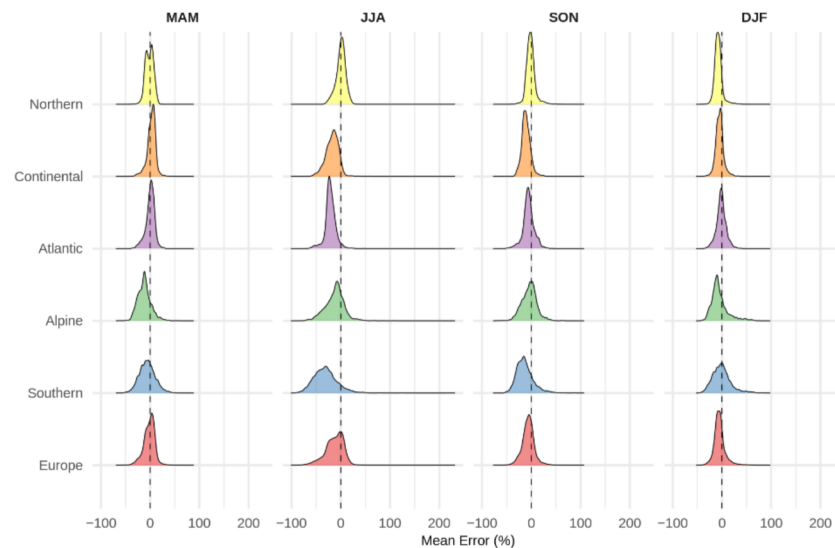


Figure 5. Relative ME distributions of unadjusted ECMWF SEAS5 ensemble seasonal forecasts of precipitation (one-month lead time) for European sub-regions and over the whole domain. ME was computed with respect to ERA5 over 1993–2016.

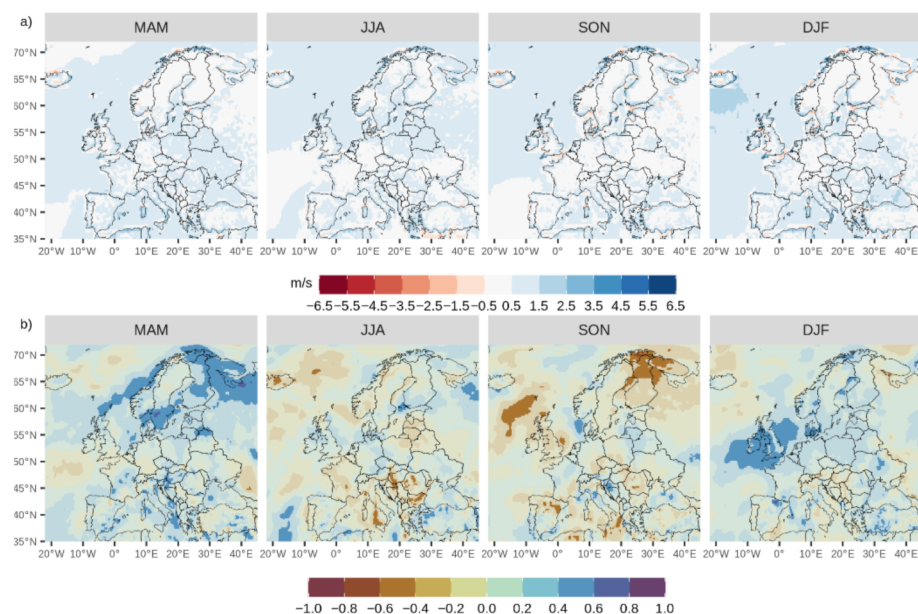


Figure 6. Spatial distribution of (a) ME and (b) CORR for unadjusted ECMWF SEAS5 ensemble forecasts of 10-m wind speed (one-month lead time) over Europe with respect to ERA5 reanalysis over 1993–2016. Significant correlation (p -value < 0.05) is reported by unshaded colors.

ROCSS distribution is largely similar to CORR patterns, as it was for temperature. In the case of precipitation and wind speed, the spatial distribution of ROCSS in all seasons exhibited negative and positive scores heterogeneously distributed throughout Europe, confirming the lack of coherent regions of skill gain and loss in the forecasting system. Since ROCSS for these variables did not add further information with respect to CORR, figures are not shown here and are instead reported in the Supplementary Material (Figures S4 and S5).

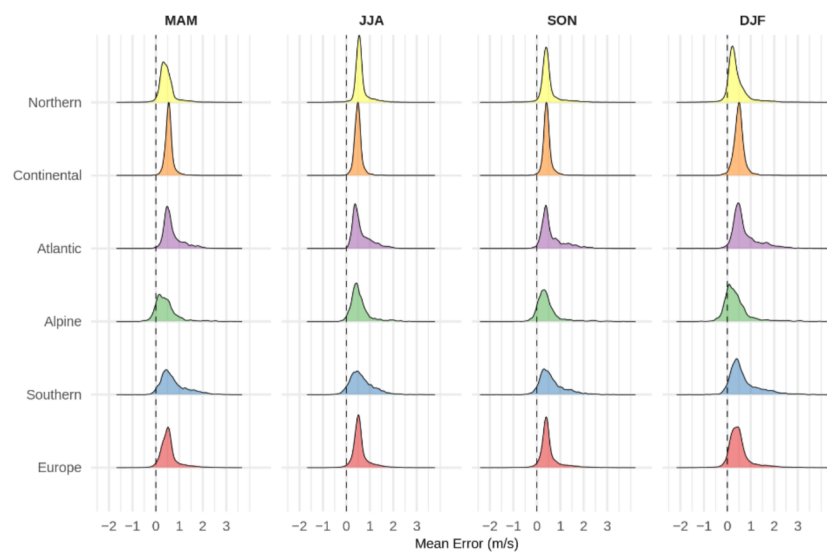


Figure 7. ME distributions of unadjusted ECMWF SEAS5 ensemble seasonal forecasts of 10-m wind speed (one-month lead time) for European sub-regions and whole domain. ME was computed with respect to ERA5 over 1993–2016.

3.2. Bias-Adjusted ECMWF SEAS5 over Europe

B-QM was found to remarkably reduce ECMWF SEAS5 forecast bias throughout Europe in all seasons and for all variables (not shown). The resulting ensemble ME was within ± 0.1 °C for temperature, ± 5 mm for precipitation and ± 0.1 m/s for wind speed. In all cases, significant improvements in the raw forecasts were obtained for bias-dependent metrics, while bias-insensitive scores such as RPSS and ROCSS did not reveal remarkable differences in skill between raw and adjusted predictions (not shown). In the following, only skill scores of bias-adjusted forecasts computed with respect to the climatological forecast were reported in order to provide a more comprehensive characterization of the features of final calibrated predictions.

The ensemble MAE for temperature showed the lowest values in JJA when absolute errors were mostly below 1 °C. The forecast skills gradually decreased in SON and DJF, especially in northern and north-eastern parts of the domain, where MAE of winter predictions exceeded 2 °C in Finland and Russia (Figure 8a). SPR distribution reported overdispersion in spring temperature predictions in eastern Europe, especially north of the Black Sea between Ukraine and Russia. This signal was already present in the uncalibrated predictions, suggesting the high internal variability of original ECMWF SEAS5 ensemble forecasts for spring temperature in the area. In contrast, the highest tendency to overconfidence is reported for the Iberian Peninsula, where the lowest SPR values were depicted (Figure 8b). The calibrated predictions showed greater accuracy with respect to the climatological forecast, especially in MAM and JJA in the south-eastern part of the European domain with positive and significant values for both CRPSS and RPSS (Figure S6). Skill gain in autumn and winter forecasts was more limited, with positive CRPSS only in Turkey and northern Europe (Figure S6a).

The B-QM ability to reduce the forecast errors of original forecasted fields was lower for precipitation than for temperature. The best results were obtained in MAM with MAE within 50 mm over most of the domain, whereas autumn and winter forecasts exhibited a MAE locally exceeding 200 mm. The largest errors were mainly recorded over the Alps, the Norwegian coast and the northern Atlantic coast of the Iberian Peninsula (Figure 9a). The SPR did not show specific spatial patterns and reported overdispersion in northern and Atlantic Europe (Figure 9b). The spatial distributions of CRPSS and RPSS were more heterogeneous than for temperature forecasts, with only small improvements in MAM with CRPSS within 0.2 in all cases. The negative scores throughout Europe, especially in

SON and DJF, suggested a general degradation of the relative accuracy of predictions with respect to the climatological forecast that was present in the forecasting system and not effectively adjusted by the calibration (Figure S7).

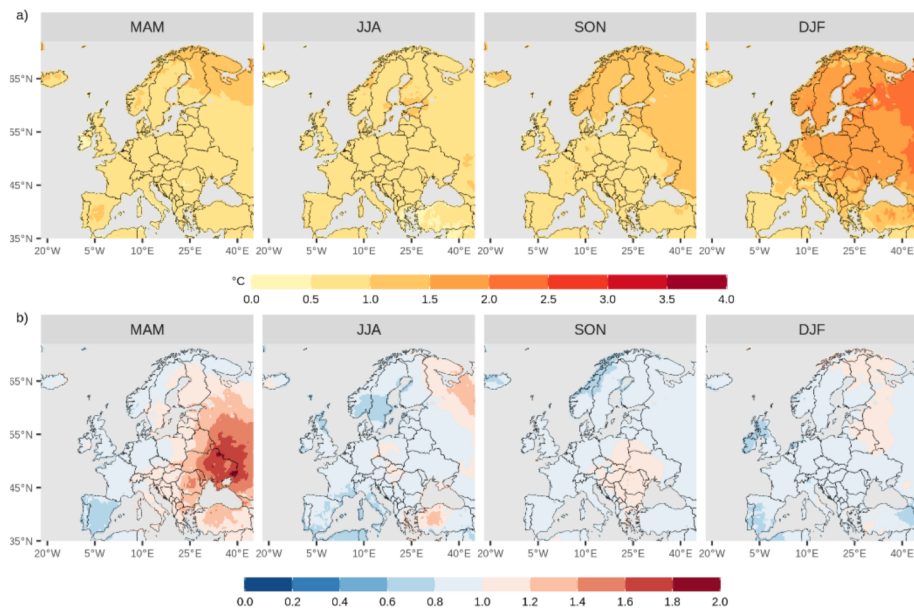


Figure 8. Spatial distribution of (a) MAE and (b) SPR for calibrated ECMWF SEAS5 ensemble seasonal forecasts of 2-m mean air temperature (one-month lead time) over Europe (land only) with respect to ERA5 reanalysis over 1993–2016.

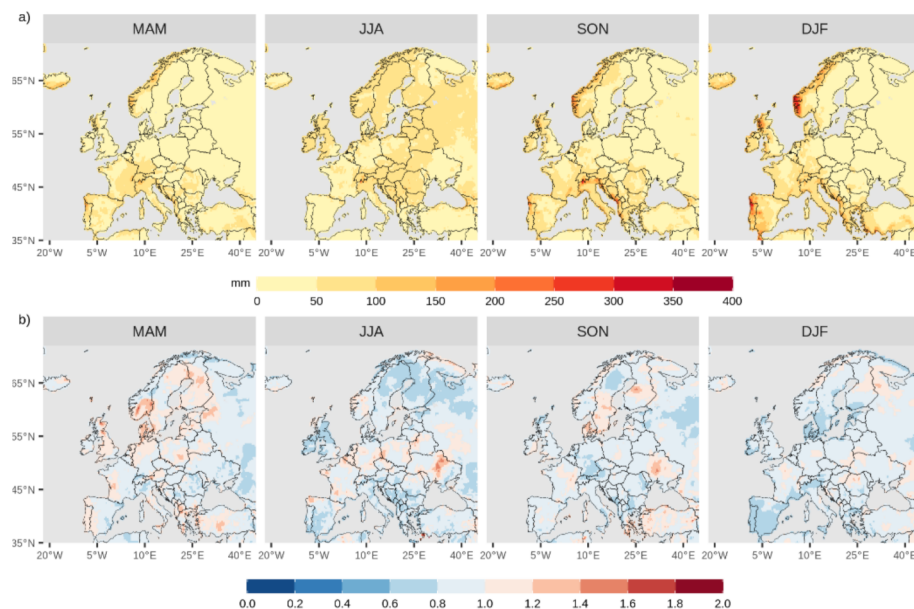


Figure 9. Spatial distribution of (a) MAE and (b) SPR for calibrated ECMWF SEAS5 ensemble seasonal forecasts of precipitation (one-month lead time) over Europe (land only) with respect to ERA5 reanalysis over 1993–2016.

B-QM for 10-m wind speed reported the highest MAE values over the sea in all seasons. The best performances were obtained for summer predictions with MAE below 0.4 m/s throughout the domain, while the greatest errors were obtained for winter forecasts (Figure 10a). It is interesting to note that the areas affected by the largest MAE were localized over the northern coasts of Ireland and the United Kingdom (UK), the western

coast of Norway and south to the Mediterranean coast of France with errors exceeding 1 m/s. The higher errors in such areas could be partly explained by the occurrence of specific weather conditions, such as low pressure systems in northern Atlantic and mistral phenomena in the northwestern Mediterranean [29]. This reduced the ability of the forecasting system to capture the local wind-speed variability, which was only partially improved by B-QM. SPR reported overdispersion mainly over the Atlantic coastline and in the northernmost part of Scandinavia, especially in MAM and SON, while no specific signal was depicted in JJA and DJF (Figure 10b). CRPSS was positive over most of the domain in MAM and statistical significance was obtained over the offshore area in front of Portugal, all along the Norwegian coastline and part of Fennoscandia and Russia (Figure S8a). Positive skills in prediction accuracy were also observed in DJF over the northern Atlantic including northern France and internal coasts of Ireland and the UK. Similar patterns with lower skill scores are depicted by RPSS. An overall skill degradation of tercile-based SEAS5 forecasts was particularly clear in JJA and SON throughout Europe (Figure S8b).

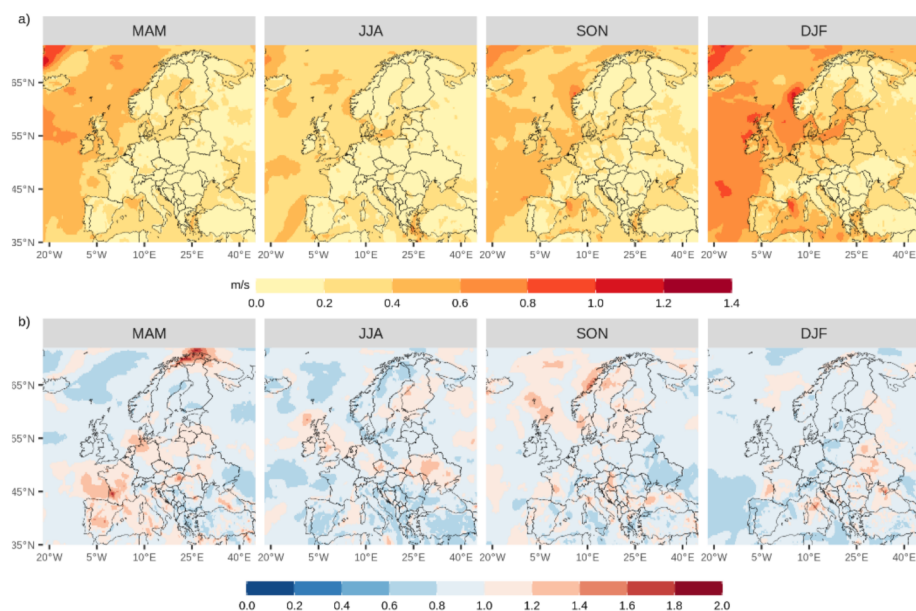


Figure 10. Spatial distribution of (a) MAE and (b) SPR for calibrated ECMWF SEAS5 ensemble seasonal forecasts of 10-m wind speed (one-month lead time) over Europe with respect to ERA5 reanalysis over 1993–2016.

3.3. The Comparison with ADAMONT

Seasonal temperature and precipitation hindcasts calibrated by ADAMONT and by the monthly B-QM scheme showed very similar skills for both JJA and DJF when compared to ERA5. The comparison here discussed was based on the one-month lead time only, but the same results were obtained by increasing the lead time (not shown). The validation was performed on adjusted forecasts without the LOYO approach, since it was not implemented for the available ADAMONT outputs.

All deterministic metrics for temperature exhibited comparable values and the same spatial patterns for both seasons. Small differences were only observed in ME distribution (Figure 11a): the bias in ADAMONT fields was locally more pronounced and its spatial pattern was more spread over the domain. Winter temperatures were slightly underestimated (within 0.5 °C) in ADAMONT throughout the domain, while summer underestimates occurred most frequently in the southern part of the domain. Such regional dependency of ADAMONT bias could be partly explained by the unique QM adjustment that was equally performed to all grid points assigned to the same weather regime. Such differences disappeared when MAE was considered, suggesting similar performances of the two methods in reducing the discrepancy of raw seasonal temperature forecasts from

ERA5 throughout the domain. The overall agreement of the two calibrated sets was also confirmed by the SPR distribution showing similar seasonal patterns. In both datasets, the metric for JJA highlighted overdispersion over the eastern portion of the domain, which was slightly more pronounced for ADAMONT. This may suggest a larger spread of the ensemble forecast (Figure 12a). In both cases, the CRPSS reported positive skills for JJA temperature forecasts in the eastern domain and in the south-east of Spain, while in winter only limited skills were observed in the northern part of the study area (Figure S9a).

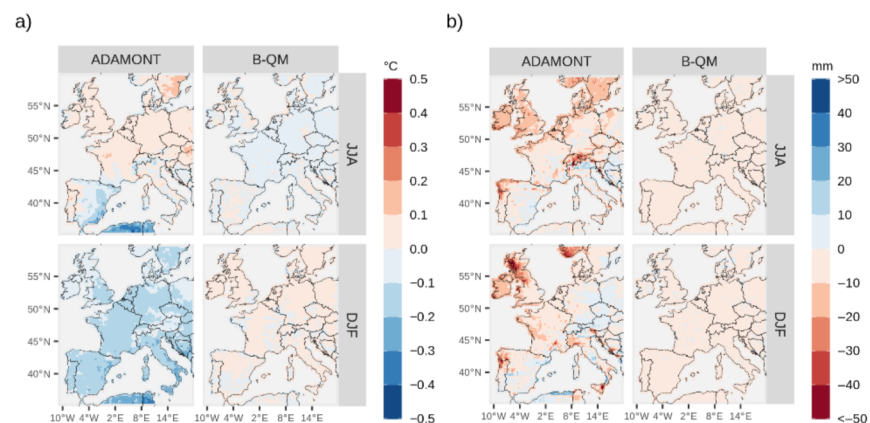


Figure 11. Spatial distribution of ME for (a) 2-m mean air temperature and (b) precipitation in JJA and DJF for calibrated seasonal forecasts (one-month lead time) by ADAMONT and B-QM over 1993–2016.

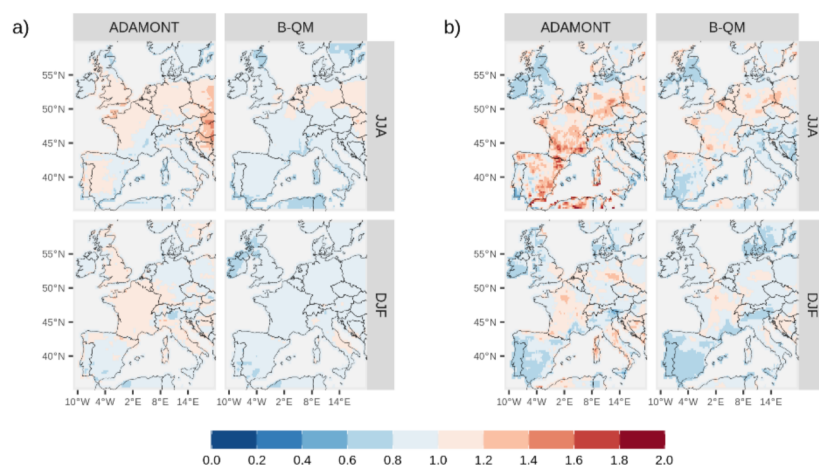


Figure 12. Spatial distribution of SPR for (a) 2-m mean air temperature and (b) precipitation in JJA and DJF for calibrated seasonal forecasts (one-month lead time) by ADAMONT and B-QM over 1993–2016.

The comparison of precipitation forecasts provided analogue outcomes, even though existing discrepancies were emphasized. The negative bias of ADAMONT was more pronounced in JJA over the northernmost part of the domain, the northern side of the Alps and along the Atlantic coast (Figure 11b). DJF bias was mostly negative over the UK and southern Norway with a ME below -50 mm, while very localized precipitation overestimations were highlighted in the Mediterranean region. However, MAE and all other deterministic metrics were comparable between the two products throughout the domain, suggesting similar absolute residual errors and skills of the calibration procedures to reproduce ERA5 reference (not shown). Moreover, the spatial SPR distributions were similar, with more localized overdispersion in JJA for ADAMONT, especially in southern France and eastern Spain (Figure 12b). As for temperature, the larger SPR could be

explained by a wider spread of the ensemble-calibrated forecast. The spatial discontinuities in the ADAMONT fields could be mainly due to the fact that the quantile mapping is applied directly to the raw forecasts without performing any spatial interpolation onto the target grid. The skill scores reported no relevant improvements with respect to the climatological forecast for either calibrated dataset, particularly in DJF (Figure S9b). The scores also confirmed the greater prediction difficulty for this variable, which was not reduced by any bias-adjustment procedure.

4. Discussion

The comparison between unadjusted ECMWF SEAS5 seasonal forecasts and ERA5 demonstrated the need to adjust the model bias in order to improve the local representativeness of predictions. Relevant discrepancies in raw forecasts with respect to the reference were observed throughout Europe for all variables and in all seasons. The largest underestimations occurred in spring and winter temperatures (up to 8 °C) and in summer precipitation (up to 500 mm), while wind speed was mostly overestimated, especially over the sea. The B-QM approach allowed for the overall improvement of the agreement of seasonal forecasts with the ERA5 reference. The adjustment was proven to be particularly effective at calibrating temperature and wind-speed predictions, while more limited improvements were observed for precipitation, especially over European mountain areas where the residual MAE could still exceed 300 mm. Even though the scope of this work was to test the benefits and limitations of the standard QM approach, alternative versions of this procedure were considered. More specifically, parametric QM techniques using specific distribution functions, such as gamma, double gamma and generalized Pareto distributions to calibrate precipitation data from climate models were proposed and tested in the framework of the VALUE initiative and could represent suitable options for improving the ability of post-processing precipitation forecasts [8].

For all variables, no relevant added value was found in terms of skill gain for bias-independent scores with respect to the use of unadjusted forecasts. However, significant improvements in CRPSS were obtained for calibrated predictions, even though the skills remained, in most cases, comparable to those provided by a climatological forecast. Positive scores were mainly obtained for spring and summer temperatures, as well as spring wind-speed predictions. These outcomes are in agreement with previous studies evaluating post-processing techniques for seasonal forecasts and confirm that bias-adjustment schemes do not significantly modify the skills of raw forecasts beyond the adjustment of systematic biases. Other statistical downscaling methods modelling the contribution of atmospheric predictors, such as Perfect Prognosis, were found to have a more relevant impact on forecast skills [3,16,30]. Nevertheless, the choice of the most suitable calibration technique is strictly dependent on the scope and type of application. The effective bias removal of B-QM without worsening the skills of original forecasts still represents a meaningful achievement whenever predictions are integrated in end-user applications focusing on the mean properties of the forecast ensemble. Alternative downscaling procedures should be evaluated and adopted if end users need to focus on the probabilistic skills and tune them to the target scales of the analysis.

The overall agreement shown by the inter-comparison of B-QM and ADAMONT for seasonally aggregated predictions of temperature and precipitation suggested that the B-QM calibration, directly applying standard QM on monthly forecasts, represents a suitable alternative to derive tailored data for applications requiring monthly or seasonal quantities. Small observed discrepancies in residual bias distributions highlighted the effects of different calibration settings. In particular, the regional dependency of ME spatial patterns of ADAMONT fields could be partly due to the customized QM based on weather regimes that inflate the same correction to all grid points assigned to the same cluster. Moreover, the application of QM to raw forecasts without any previous spatial interpolation to the target grid of the reanalysis can lead to spatial discontinuities in the resulting fields. However, the unique quantile adjustment based on the same weather

regime applied by ADAMONT to all variables is expected to better preserve the consistency between forecasted parameters, which could be particularly relevant when they are used to feed impact models.

The aim of the present work was not to establish the best method for calibrating seasonal forecasts, but rather to verify whether the proposed scheme can provide reasonable outputs, to identify the most critical variables and regions for the adjustment performance, and to provide end users with alternatives for processing seasonal forecasts by choosing the approach that best suits their needs. ADAMONT is planned to be used by Météo France as an operational service to provide seasonal forecasts of daily variables, while B-QM was proposed in the framework of the SECLI-FIRM project as an alternative method for downscaling and bias adjusting seasonal forecasts of monthly quantities. In this study, B-QM was applied and evaluated on hindcasts only; however, the same procedure can be applied to calibrate operational seasonal forecasts without substantial changes to the methodology. Due to the very similar performances of B-QM and ADAMONT, there are no specific reasons to prefer one method to the other when seasonal forecasts are required as monthly or seasonal aggregations. End users and industrial players can benefit from both approaches when managing and extracting calibrated information to add value to their businesses.

The considered schemes were proven to be effective in providing medium spatial resolution data, but they require further testing for finer resolutions, e.g., at the kilometer scale, to better verify their suitability for tailoring meaningful predictions for local applications. The same downscaling approaches can also be applied by replacing reanalysis with observation data in order to improve the representativeness of bias-adjusted fields. However, this evaluation is strongly dependent on the availability of accurate reference datasets at fine spatial scales.

Moreover, the ability of emerging alternative deep learning-based approaches, such as random forests and convolutional neural networks, in downscaling seasonal forecasts needs to be further investigated in forthcoming studies, and B-QM could represent a benchmark in the outcome evaluation. Deep learning-based techniques could also provide effective tools to bridge the gaps among short-term, sub-seasonal and seasonal forecasts and further enhance their integration into innovative climate services.

Supplementary Materials: The following are available online at <https://www.mdpi.com/article/10.3390/cli9120181/s1>, Figure S1, CORR distribution of unadjusted ECMWF SEAS5 ensemble seasonal forecasts of 2-m mean air temperature (one-month lead time) for European sub-regions and whole domain (bottom row). CORR was computed with respect to ERA5 over 1993–2016. Figure S2, CORR distribution of unadjusted ECMWF SEAS5 ensemble seasonal forecasts of precipitation (one-month lead time) for European sub-regions and whole domain (bottom row). CORR was computed with respect to ERA5 over 1993–2016. Figure S3, CORR distribution of unadjusted ECMWF SEAS5 ensemble seasonal forecasts of 10-m wind speed (one-month lead time) for European sub-regions and whole domain (bottom row). CORR was computed with respect to ERA5 over 1993–2016. Figure S4, Spatial distribution of ROCSS for a) first and b) third tercile for unadjusted ECMWF SEAS5 ensemble seasonal forecasts of precipitation (one-month lead time) over Europe (land only) with respect to the climatological forecast over 1993–2016. Unshaded colors indicate grid cells with significant positive scores. Figure S5, Spatial distribution of ROCSS for a) first and b) third tercile for unadjusted ECMWF SEAS5 ensemble seasonal forecasts of 10-m wind speed (one-month lead time) over Europe with respect to climatological forecast over 1993–2016. Unshaded colors indicate grid cells with significant positive scores. Figure S6, Spatial distribution of a) CRPSS and b) RPSS for calibrated ECMWF SEAS5 ensemble seasonal forecasts of 2-m mean air temperature (one-month lead time) over Europe (land only) with respect to the climatological forecast over 1993–2016. Points indicate grid cells with significant positive scores. Figure S7, Spatial distribution of a) CRPSS and b) RPSS for calibrated ECMWF SEAS5 ensemble seasonal forecasts of precipitation (one-month lead time) over Europe (land only) with respect to the climatological forecast over 1993–2016. Points indicate grid cells with significant positive scores. Figure S8, Spatial distribution of a) CRPSS and b) RPSS for calibrated ECMWF SEAS5 ensemble seasonal forecasts of 10-m wind speed (one-month lead time)

over Europe with respect to climatological forecast over 1993–2016. Points indicate grid cells with significant positive scores. Figure S9, Spatial distribution of CRPSS for a) mean air temperature and b) precipitation in JJA and DJF for calibrated seasonal forecasts (one-month lead time) by ADAMONT and B-QM over 1993–2016 compared to the climatological forecast. Points indicate grid cells with significant positive scores.

Author Contributions: Conceptualization, A.C. and M.P.; methodology, A.C.; validation, A.C., M.P., P.M. and C.V.; data curation, A.C., M.P., P.M. and L.G.; writing—original draft preparation, A.C.; writing—review and editing, all; visualization, A.C. All authors have read and agreed to the published version of the manuscript.

Funding: This research has received funding from the European Commission Horizon 2020 Framework Program SECLI-FIRM (grant no. 776868). The contents of this publication are the sole responsibility of the authors and do not necessarily reflect the opinion of the European Union.

Data Availability Statement: All data used in this study are openly available in the Climate Data Store of the Copernicus Climate Change Service (<https://cds.climate.copernicus.eu/#!/home>, accessed on 26 October 2021).

Acknowledgments: This research has received funding from the European Union’s Horizon 2020 research and innovation program under grant agreement No 776868 (SECLI-FIRM project).

Conflicts of Interest: The authors declare no conflict of interest.

References

- Orlov, A.; Sillmann, J.; Vigo, I. Better seasonal forecasts for the renewable energy industry. *Nat. Energy* **2020**, *5*, 108–110. [[CrossRef](#)]
- Han, E.; Ines, A.V.M. Downscaling probabilistic seasonal climate forecasts for decision support in agriculture: A comparison of parametric and non-parametric approach. *Clim. Risk Manag.* **2017**, *18*, 51–65. [[CrossRef](#)]
- Manzanas, R.; Gutiérrez, J.M.; Fernández, J.; van Meijgaard, E.; Calmanti, S.; Magariño, M.E.; Cofiño, A.S.; Herrera, S. Dynamical and statistical downscaling of seasonal temperature forecasts in Europe: Added value for user applications. *Clim. Serv.* **2018**, *9*, 44–56. [[CrossRef](#)]
- Tian, D.; Martinez, C.J.; Graham, W.D.; Hwang, S. Statistical Downscaling Multimodel Forecasts for Seasonal Precipitation and Surface Temperature over the Southeastern United States. *J. Clim.* **2014**, *27*, 8384–8411. [[CrossRef](#)]
- Laprise, R.; de Elía, R.; Caya, D.; Biner, S.; Lucas-Picher, P.; Diaconescu, E.; Leduc, M.; Alexandru, A.; Separovic, L.; Canadian Network for Regional Climate Modelling and Diagnostic. Challenging some tenets of Regional Climate Modelling. *Meteorol. Atmos. Phys.* **2008**, *100*, 3–22. [[CrossRef](#)]
- Hwang, S.; Graham, W.D.; Hernández, J.L.; Martinez, C.; Jones, J.W.; Adams, A. Quantitative spatiotemporal evaluation of dynamically downscaled MM5 precipitation predictions over the Tampa Bay region, Florida. *J. Hydrometeor.* **2011**, *12*, 1447–1464. [[CrossRef](#)]
- Tiwari, P.R.; Kar, S.C.; Mohanty, U.C.; Dey, S.; Sinha, P.; Shekhar, M.S.; Sokhi, R.S. Comparison of statistical and dynamical downscaling methods for seasonal-scale winter precipitation predictions over north India. *Int. J. Climatol.* **2019**, *39*, 1504–1516. [[CrossRef](#)]
- Gutiérrez, J.M.; Maraun, D.; Widmann, M.; Huth, R.; Hertig, E.; Benestad, R.; Roessler, O.; Wibig, J.; Wilcke, R.; Kotlarski, S.; et al. An intercomparison of a large ensemble of statistical downscaling methods over Europe: Results from the VALUE perfect predictor cross-validation experiment. *Int. J. Climatol.* **2019**, *39*, 3750–3785. [[CrossRef](#)]
- Maraun, D.; Huth, R.; Gutiérrez, J.M.; San Martín, D.; Dubrovsky, M.; Fischer, A.; Hertig, E.; Soares, P.M.M.; Bartholy, J.; Pongrácz, R.; et al. The VALUE perfect predictor experiment: Evaluation of temporal variability. *Int. J. Climatol.* **2019**, *39*, 3786–3818. [[CrossRef](#)]
- Cannon, A.J.; Piani, C.; Sippel, S. Chapter 5—Bias correction of climate model output for impact models. In *Climate Extremes and Their Implications for Impact and Risk Assessment*; Sillmann, J., Sippel, S., Russo, S., Eds.; Elsevier: Amsterdam, The Netherlands, 2020; pp. 77–104. [[CrossRef](#)]
- Volosciuk, C.; Maraun, D.; Vrac, M.; Widmann, M. A combined statistical bias correction and stochastic downscaling method for precipitation. *Hydrol. Earth Syst. Sci.* **2017**, *21*, 1693–1719. [[CrossRef](#)]
- Casanueva, A.; Herrera, S.; Iturbide, M.; Lange, S.; Jury, M.; Dosio, A.; Maraun, D.; Gutiérrez, J.M. Testing bias adjustment methods for regional climate change applications under observational uncertainty and resolution mismatch. *Atmos. Sci. Lett.* **2020**, *21*, e978. [[CrossRef](#)]
- Piani, C.; Haerter, J.O.; Coppola, E. Statistical bias correction for daily precipitation in regional climate models over Europe. *Theor. Appl. Climatol.* **2010**, *99*, 187–192. [[CrossRef](#)]
- Marcos, R.; Llasat, M.C.; Quintana-Seguí, P.; Turco, M. Use of bias correction techniques to improve seasonal forecasts for reservoirs—A case-study in northwestern Mediterranean. *Sci. Total Environ.* **2018**, *610–611*, 64–74. [[CrossRef](#)]

15. Crochemore, L.; Ramos, M.-H.; Pappenberger, F. Bias correcting precipitation forecasts to improve the skill of seasonal streamflow forecasts. *Hydrol. Earth Syst. Sci.* **2016**, *20*, 3601–3618. [[CrossRef](#)]
16. Manzanas, R.; Gutiérrez, J.M.; Bhend, J.; Hemri, S.; Doblas-Reyes, F.J.; Torralba, V.; Penabad, E.; Brookshaw, A. Bias adjustment and ensemble recalibration methods for seasonal forecasting: A comprehensive intercomparison using the C3S dataset. *Clim. Dyn.* **2019**, *53*, 1287–1305. [[CrossRef](#)]
17. Ramon, J.; Lledó, L.; Bretonnière, P.-A.; Samsó, M.; Doblas-Reyes, F.J. A perfect prognosis downscaling methodology for seasonal prediction of local-scale wind speeds. *Environ. Res. Lett.* **2021**, *16*, 5. [[CrossRef](#)]
18. Hyvärinen, O.; Venäläinen, A.; Vajda, A. Bias-adjusted seasonal forecasts of soil moisture for forestry applications in Finland. *Adv. Sci. Res.* **2020**, *17*, 23–27. [[CrossRef](#)]
19. Buontempo, C.; Hewitt, C.D.; Doblas-Reyes, F.J.; Dessai, S. Climate service development, delivery and use in Europe at monthly to inter-annual timescales. *Clim. Risk Manag.* **2014**, *6*, 1–5. [[CrossRef](#)]
20. Ehret, U.; Zehe, E.; Wulfmeyer, V.; Warrach-Sagi, K.; Liebert, J. Should we apply bias correction to global and regional climate model data? *Hydrol. Earth Syst. Sci.* **2012**, *16*, 3391–3404. [[CrossRef](#)]
21. Maraun, D. Bias Correcting Climate Change Simulations—A Critical Review. *Curr. Clim. Chang. Rep.* **2016**, *2*, 211. [[CrossRef](#)]
22. Cannon, A.J. Multivariate quantile mapping bias correction: An N -dimensional probability density function transform for climate model simulations of multiple variables. *Clim. Dyn.* **2018**, *50*, 31–49. [[CrossRef](#)]
23. Johnson, S.J.; Stockdale, T.N.; Ferranti, L.; Balmaseda, M.A.; Molteni, F.; Magnusson, L.; Tietsche, S.; Decremmer, D.; Weisheimer, A.; Balsamo, G.; et al. SEAS5: The new ECMWF seasonal forecast system. *Geosci. Model Dev.* **2019**, *12*, 1087–1117. [[CrossRef](#)]
24. Verfaillie, D.; Déqué, M.; Morin, S.; Lafaysse, M. The method ADAMONT v1.0 for statistical adjustment of climate projections applicable to energy balance land surface models. *Geosci. Model Dev.* **2017**, *10*, 4257–4283. [[CrossRef](#)]
25. Hersbach, H.; Bell, B.; Berrisford, P.; Hirahara, S.; Horányi, A.; Muñoz-Sabater, J.; Nicolas, J.; Peubey, C.; Radu, R.; Schepers, D.; et al. The ERA5 global reanalysis. *Q. J. R. Meteorol. Soc.* **2020**, *146*, 1999–2049. [[CrossRef](#)]
26. Cannon, A.J.; Sobie, S.R.; Murdock, T.Q. Bias Correction of GCM Precipitation by Quantile Mapping: How Well Do Methods Preserve Changes in Quantiles and Extremes? *J. Clim.* **2015**, *28*, 6938–6959. [[CrossRef](#)]
27. Gudmundsson, L.; Bremnes, J.B.; Haugen, J.E.; Engen-Skaugen, T. Technical Note: Downscaling RCM precipitation to the station scale using statistical transformations—A comparison of methods. *Hydrol. Earth Syst. Sci.* **2012**, *16*, 3383–3390. [[CrossRef](#)]
28. Metzger, M.J. *Sub-Regional Classification of the IPCC Europe Region*; University of Edinburgh: Edinburgh, UK, 2018. [[CrossRef](#)]
29. Obermann, A.; Bastin, S.; Belamari, S.; Conte, D.; Gaertner, M.A.; Li, L.; Ahrens, B. Mistral and Tramontane wind speed and wind direction patterns in regional climate simulations. *Clim. Dyn.* **2018**, *51*, 1059–1076. [[CrossRef](#)]
30. Manzanas, R.; Lucero, A.; Weisheimer, A.; Gutiérrez, J.M. Can bias correction and statistical downscaling methods improve the skill of seasonal precipitation forecasts? *Clim. Dyn.* **2018**, *50*, 1161–1176. [[CrossRef](#)]