Evaluating Bias-Correction Methods for Seasonal Dynamical Precipitation Forecasts

Article *in* Journal of Hydrometeorology · June 2022

DOI: 10.1175/JHM-D-22-0049.1

2 authors:

Saeed Golian
SRK consulting

Conor Murphy
National University of Ireland, Maynooth

Evaluating Bias-Correction Methods for Seasonal Dynamical Precipitation Forecasts

SAEED GOLIAN[a] AND CONOR MURPHY[a]

[a] *Irish Climate Analysis and Research Units, Department of Geography, Maynooth University, Kildare, Ireland*

(Manuscript received 18 April 2022, in final form 13 June 2022)

ABSTRACT: Seasonal forecasting of climatological variables is important for water and climatic-related decision-making. Dynamical models provide seasonal forecasts up to one year in advance, but direct outputs from these models need to be bias-corrected prior to application by end users. Here, five bias-correction methods are applied to precipitation hindcasts from ECMWF's fifth generation seasonal forecast system (SEAS5). We apply each method in two distinct ways; first to the ensemble mean and second to individual ensemble members, before deriving an ensemble mean. The performance of bias-correction methods in both schemes is assessed relative to the simple average of raw ensemble members as a benchmark. Results show that in general, bias correction of individual ensemble members before deriving an ensemble mean (scheme 2) is most skillful for more frequent precipitation values while bias correction of the ensemble mean (scheme 1) performed better for extreme high and low precipitation values. Irrespective of application scheme, all bias-correction methods improved precipitation hindcasts compared to the benchmark method for lead times up to 6 months, with the best performance obtained at one month lead time in winter.

KEYWORDS: Precipitation; Bias; Probabilistic Quantitative Precipitation Forecasting (PQPF); Seasonal forecasting; General circulation models

1. Introduction

Seasonal precipitation forecasts have high utility in informing decision-making in diverse areas including in agricultural decision-making (e.g., Calanca et al. 2011), private insurance companies (e.g., Osgood et al. 2008), water authorities (e.g., Baker et al. 2019), and other sectors to assist preparation for probable future weather conditions (Blench 1999). However, due to uncertainties related to model structure, boundary conditions, and input datasets, outputs from dynamical forecasting systems contain systematic and random model errors. Changing the initial conditions or even the formulation of the model, e.g., perturbing the values of model parameters, can produce ensemble forecasts for the same time period. In this way, uncertainties in dynamical models can be taken into account (Troccoli 2010). The advantages of using dynamical seasonal forecasts have been shown in many studies. For example, Arnal et al. (2018) showed that using ensemble hindcasts from the European Centre for Medium-Range Weather Forecasts (ECMWF) System 4 (SEAS4) is more skillful in streamflow forecasting compared to the ensemble streamflow prediction (ESP) forecasting approach in some catchments in Europe and for certain seasons, especially in winter. Golian et al. (2022) also showed how a hybrid statistical–dynamical method can improve precipitation forecast skill in winter and summer in Ireland.

While ensemble forecasting provides more complete information on possible future weather conditions compared to a single forecast (Atger 2001), their accuracy and reliability are not usually good for variables close to Earth's surface such as precipitation (Buizza 2018). To increase the skill of seasonal weather prediction systems, Stockdale et al. (2010) reviewed and explained the requirements for these systems, including the initial condition, high-quality models of the ocean–atmosphere–land system, and required data for validation of seasonal forecasting systems. The raw ensemble forecasts from dynamical models usually do not provide reliable and accurate forecasts, especially for longer lead times. Qian et al. (2020) showed that when using the ensemble mean of the dynamical models, they perform worse than a statistical method which employed regression relationships between sea surface temperature (SST) and precipitation. They showed that the accuracy of raw dynamical model outputs sharply decreased with lead times longer than 1 month. This implies the importance of using postprocessing and bias-correction methods to increase the skill of dynamical model outputs for decision-making. Bias correction is referred to the process of adjusting biased simulated data to observations (Reiter et al. 2016).

Different methods have been developed for bias correction and downscaling dynamical models (e.g., Bhatti et al. 2016; Moghim and Bras 2017; Maity et al. 2019; Yang et al. 2020; Kim et al. 2021). Two widely used bias-correction methods are linear scaling and distribution mapping (Crochemore et al. 2016). Ghimire et al. (2019) applied eight bias-correction methods to rainfall forecasts from three global climate models (GCMs) at monthly and annual time scales to improve hydrological simulations at multiple time scales. While all methods improved the accuracy of forecasts, linear scaling and empirical quantile mapping showed better performance compared to other methods, i.e., parametric quantile mapping methods with scaling function. Mendez et al. (2020) compared six bias-correction methods to adjust the precipitation outputs of five dynamical models over Costa Rica and found that empirical

Supplemental information related to this paper is available at the Journals Online website: https://doi.org/10.1175/JHM-D-22-0049.s1.

Corresponding author: Saeed Golian, saeed.golian@mu.ie

DOI: 10.1175/JHM-D-22-0049.1

quantile mapping (EQM) and the delta method (DT) outperformed the other approaches including linear scaling, gamma quantile mapping, power transformation of precipitation, and gamma–Pareto quantile mapping in enhancing the accuracy of dynamical model predictions.

Despite previous research, there are few comprehensive studies demonstrating how bias-correction methods perform over different lead times for seasonal forecasting. For example, Crochemore et al. (2016) showed how bias-correcting precipitation forecasts can improve the skill of streamflow forecasts up to 3-month lead time. While their results revealed that bias correction generally improved the skill of forecasts, they did not clearly show the effect of lead time on performance of bias-correction methods. In another study, Monhart et al. (2018) assessed the two bias-correction methods at different lead times (5–32 days) and seasons and found improved skill for bias-corrected temperature in all seasons except spring. In terms of ensemble forecast outputs from dynamical models, the majority of studies apply bias-correction methods to ensemble members individually and then average them or use them separately (e.g., Ratri el al. 2019; Crespi et al. 2021; Lorenz et al. 2021), while there is little research to compare how bias-correction methods perform when applied to individual ensembles members or when applied to the ensemble mean.

This study aims to assess the performance of different bias-correction methods for application to precipitation forecasts from ECMWF SEAS5 for 44 Irish catchments over the period 1981–2016. Specifically, we address the following three questions: 1) Which bias-correction method is most effective at improving forecast skill? 2) Is bias-correction skill a function of lead time? 3) Is there a significant difference between the performance of bias-correction methods when applied directly to the ensemble mean or alternatively when applied to individual ensembles first and then averaged?

The remainder of the paper is organized as follows. In section 2, a brief description of datasets, bias-correction methods, and evaluation criteria employed is presented. Section 3 contains results and discussion, and finally, conclusions are briefly presented in section 4.

2. Data and methods

a. Study catchments and data

Figure 1 shows the location of the 44 study catchments across Ireland. These catchments were selected as they have good quality data and provide a representative sample of Ireland's diverse hydrological and climatological conditions, with good spatial coverage. They have also been employed in previous efforts to develop seasonal hydrological forecasting techniques (Donegan et al. 2021; Foran Quinn et al. 2021). Table 1 summarizes the key characteristics of each catchment.

For precipitation, monthly hindcasts from ECMWF's fifth generation long-range seasonal forecasting system (SEAS5) with up to 6 months lead time (LT) were downloaded from the ECMWF Meteorological Archival and Retrieval System (MARS) system (https://www.ecmwf.int/en/computing/software/ecmwf-web-api)

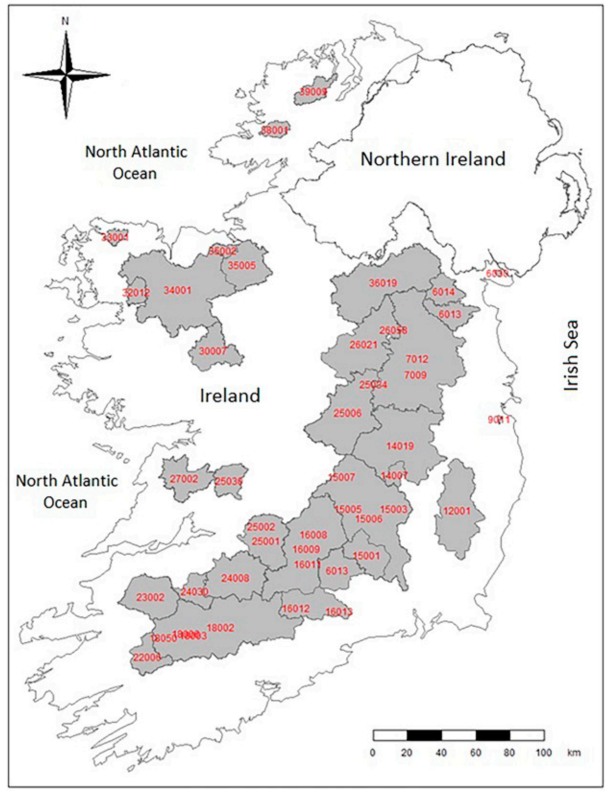

FIG. 1. The location of study catchments (after Golian et al. 2021).

for the period 1981–2016. SEAS5 consists of 25 ensemble members initialized on the first of the month. The monthly values were used to calculate seasonal forecasts for winter (DJF) and summer (JJA) as the wettest and driest seasons in Ireland, respectively, and also for spring (MAM) and autumn (SON) seasons. SEAS5 data were downloaded at 0.125° spatial resolution using the ECMWF web application programming interface (API) tools and then averaged over each catchment by overlaying the shapefile of catchments on precipitation maps. Seasonal forecasts are compared with observed precipitation values for each catchment derived from a national gridded precipitation dataset produced by Met Éireann (Ireland's national meteorological service) (Walsh 2012). The number of rain gauges varies from year to year, with approximately 550 rain gauge locations used by Walsh (2012). Data were quality controlled (QC) and missing data were filled using three methods, namely, weighted ratios of nearby stations, weighted spatial regression and finally spatial interpolation method. For more details, readers are referred to Walsh (2012).

b. Bias-correction methods

Following collation of observed and forecast data for concurrent periods, different bias-correction methods were applied (Fig. 2). We employ five methods, namely, linear scaling (Scaling), quantile mapping based on empirical distribution (EQM), quantile mapping based on the gamma distribution (GQM), quantile delta mapping (QDM), and ordinary least squares (OLS) regression. The OLS method is applied only to

TABLE 1. The characteristics of selected catchments for this study.

Station	River	Location	Area (km^2)	Mean elevation (m)	Mean annual precipitation (mm)	Mean annual temperature (°C)
6013	Dee	Charleville	309.15	83.92	882.95	9.27
6014	Glyde	Tallanstown	270.38	84.53	920.38	9.27
6030	Big	Ballygoly	10.40	256.44	1152.69	8.52
7009	Boyne	Navan Weir	1658.19	84.77	892.05	9.30
7012	Boyne	Slane Castle	2460.27	90.96	911.39	9.25
9011	Slang	Frankfort	5.46	101.91	972.36	9.82
12001	Slaney	Scarawalsh	1030.75	160.60	1094.06	9.17
14007	Stradbally	Derrybrock	118.59	134.91	900.68	9.14
14019	Barrow	Levitstown	1697.28	93.90	869.21	9.34
15001	Kings	Annamult	444.35	118.43	964.21	9.41
15003	Dinin	Dinin Br.	299.17	208.31	1015.51	8.76
15005	Erkina	Durrow Ft. Br.	379.37	126.65	896.77	9.25
15006	Nore	Brownsbarn	2418.27	136.60	967.75	9.22
15007	Nore	Kilbricken	339.76	168.78	1114.96	8.96
16008	Suir	New Bridge	1090.25	138.01	1033.20	9.30
16009	Suir	Caher Park	1582.69	139.35	1074.94	9.32
16011	Suir	Clonmel	2143.67	148.99	1103.13	9.29
16012	Tar	Tar Br.	229.63	195.91	1271.01	9.11
16013	Nire	Fourmilewater	93.58	284.81	1313.72	8.47
18002	Blackwater	Ballyduff	2333.69	165.62	1256.67	9.43
18003	Blackwater	Killavullen	1256.70	181.07	1342.32	9.40
18006	Blackwater	CSET Mallow	1054.78	187.84	1366.37	9.36
18050	Blackwater	Duarrigle	248.83	210.71	1574.50	9.27
22006	Flesk (Laune)	Flesk	328.81	233.29	1819.37	9.21
23002	Feale	Listowel	646.85	195.81	1441.98	9.31
24008	Maigue	Castleroberts	806.04	96.12	979.78	9.77
24030	Deel	Danganbeg	258.88	119.62	1055.29	9.75
25001	Mulkear	Annacotty	647.56	152.49	1172.67	9.26
25002	Newport	Barrington's Br.	221.61	188.91	1263.04	8.98
25006	Brosna	Ferbane	1162.76	88.40	924.53	9.33
25030	Graney	Scarriff	280.02	136.00	1207.74	9.31
25034	L. Ennell trib	Rochfort	10.77	109.64	989.29	9.10
26021	Inny	Ballymahon	1098.78	89.74	975.83	9.18
26058	Inny Upper	Ballinrink Br.	59.98	118.97	1001.93	9.02
27002	Fergus	Ballycorey	564.27	70.06	1327.55	9.80
30007	Clare	Ballygaddy	469.90	75.00	1131.77	9.38
32012	Newport	Newport Weir	146.16	133.40	1674.39	9.10
33001	Glenamoy	Glenamoy	76.12	108.58	1587.81	9.19
34001	Moy	Rahans	1974.76	81.22	1317.81	9.31
35002	Owenbeg	Billa Br.	88.82	183.24	1562.77	8.60
35005	Ballysadare	Ballysadare	639.66	99.71	1263.21	9.13
36019	Erne	Belturbet	1491.76	106.49	1034.89	9.03
38001	Owenea	Clonconwal Ford	111.25	184.91	1866.51	8.38
39009	Fern O/L	Aghawoney	206.83	139.82	1560.59	8.67

selected ensemble members while the other four bias-correction methods are applied to both the ensemble mean (scheme 1) and each (×25) ensemble members (scheme 2) to evaluate if there is a significant difference between methods of deployment (Fig. 2). Like many data-driven models, for OLS method we only apply the best combination of predictors (here ensembles) derived from a feature selection method to reduce the computational cost, enhance model generalization, and increase model performance [section 2b(5)]. The results of these bias-correction methods are compared with the raw ensemble mean (Mean_ens) for each season to evaluate the skill of different methods in improving forecast accuracy. The following subsections provide a brief overview of each bias-correction method.

1) LINEAR SCALING METHOD (SCALING)

This is the simplest bias-correction approach in which a monthly correction factor is applied to the forecasted precipitation data using the following equations

$$P_{m,\mathrm{rev}} = P_{m,\mathrm{orig}} \times \frac{\overline{P}_{m,\mathrm{obs}}}{\overline{P}_{m,\mathrm{orig}}}, \tag{1}$$

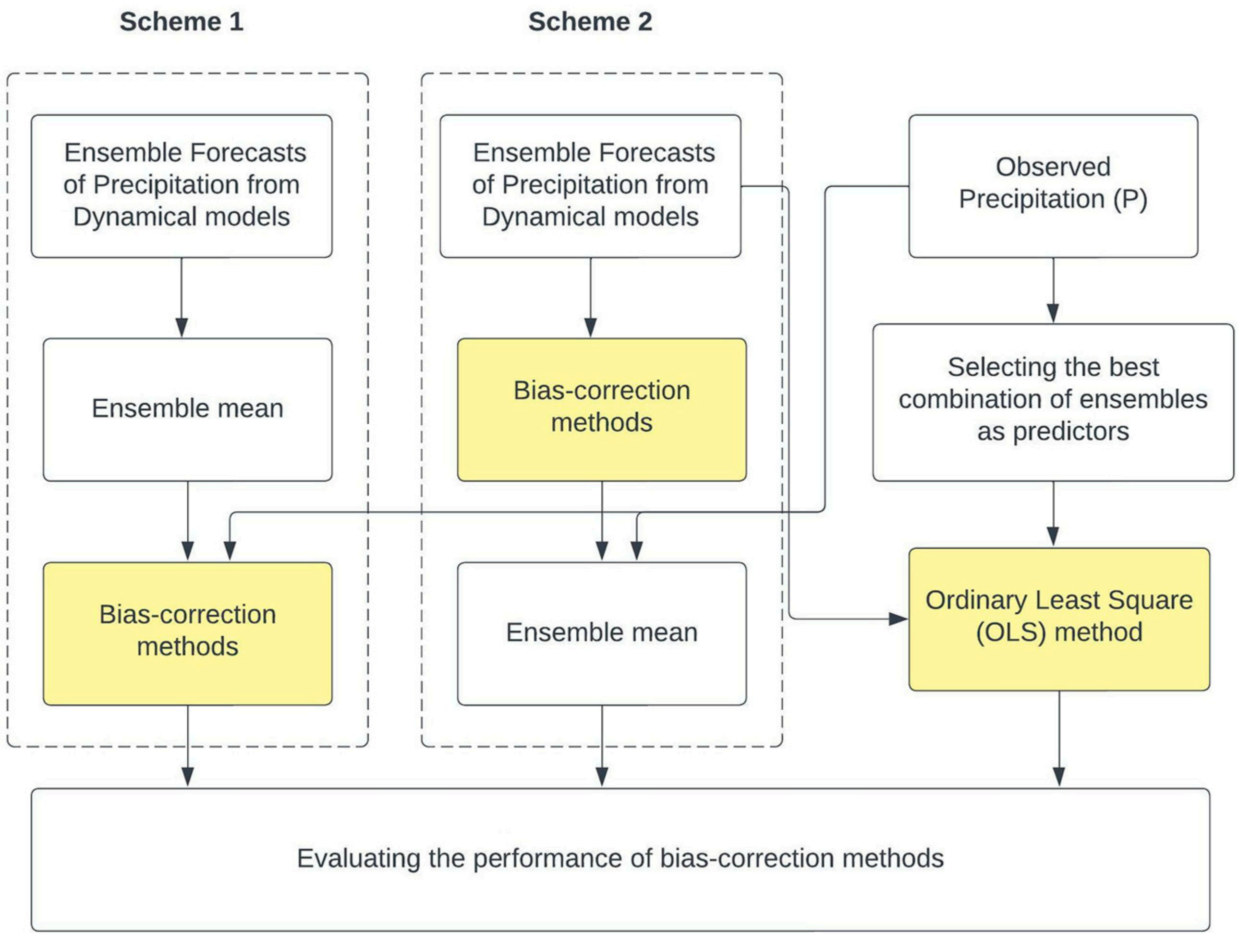

FIG. 2. Overview of study design adopted for evaluating bias-correction methods and approaches to implementation.

where $P_{m,\mathrm{rev}}$ is revised precipitation for month m, $P_{m,\mathrm{orig}}$ is the average precipitation of the raw ensembles from dynamical models for month m, $\overline{P}_{m,\mathrm{obs}}$ is the average observed precipitation for month m, and $\overline{P}_{m,\mathrm{orig}}$ is the average precipitation from dynamical models for month m. For precipitation multiplicative correction factors are applied (Teutschbein and Seibert 2012).

2) EMPIRICAL DISTRIBUTION QUANTILE MAPPING (EQM)

This method is based on the empirical cumulative distribution function (ecdf) and can be applied to both wet and dry spells. In the EQM method, both the frequency of precipitation occurrences along with standard deviations can be corrected simultaneously (Luo et al. 2018) using the following equations:

$$P_{m,\mathrm{rev}} = \mathrm{ecdf}^{-1}[\mathrm{ecdf}_{m,\mathrm{orig}}(P_{m,\mathrm{orig}})], \tag{2}$$

where ecdf^{-1} represents the inverse of the ecdf.

3) GAMMA DISTRIBUTION QUANTILE MAPPING (GQM)

It has been shown in some studies that the gamma distribution is an appropriate distribution function for climatological variables, especially precipitation (e.g., Piani et al. 2010). The probability distribution function has the following equation:

$$f_\gamma(x|\alpha,\beta) = x^{\alpha-1} \times \frac{1}{\beta^\alpha \times \Gamma(\alpha)} e^{-x/\beta};\ x \ge 0;\ \alpha,\beta > 0, \tag{3}$$

where α and β are shape and scale parameters, respectively, and $\Gamma(\cdot)$ is the gamma function. The bias-corrected precipitation is then calculated by the cumulative gamma distribution function using following equation:

$$P_{m,\mathrm{rev}} = F_\gamma^{-1}[F_\gamma(P_{m,\mathrm{orig}} \mid \alpha_{m,\mathrm{orig}}, \beta_{m,\mathrm{orig}})|\alpha_{m,\mathrm{obs}}, \beta_{m,\mathrm{obs}}] \tag{4}$$

where F_γ and F_γ^{-1} are the cumulative gamma distribution and its inverse form, respectively; $\alpha_{m,\mathrm{orig}}$ and $\beta_{m,\mathrm{orig}}$ are gamma parameters of original (raw) ensemble data; and $\alpha_{m,\mathrm{obs}}$ and $\beta_{m,\mathrm{obs}}$ are gamma parameters from observed precipitation for month m.

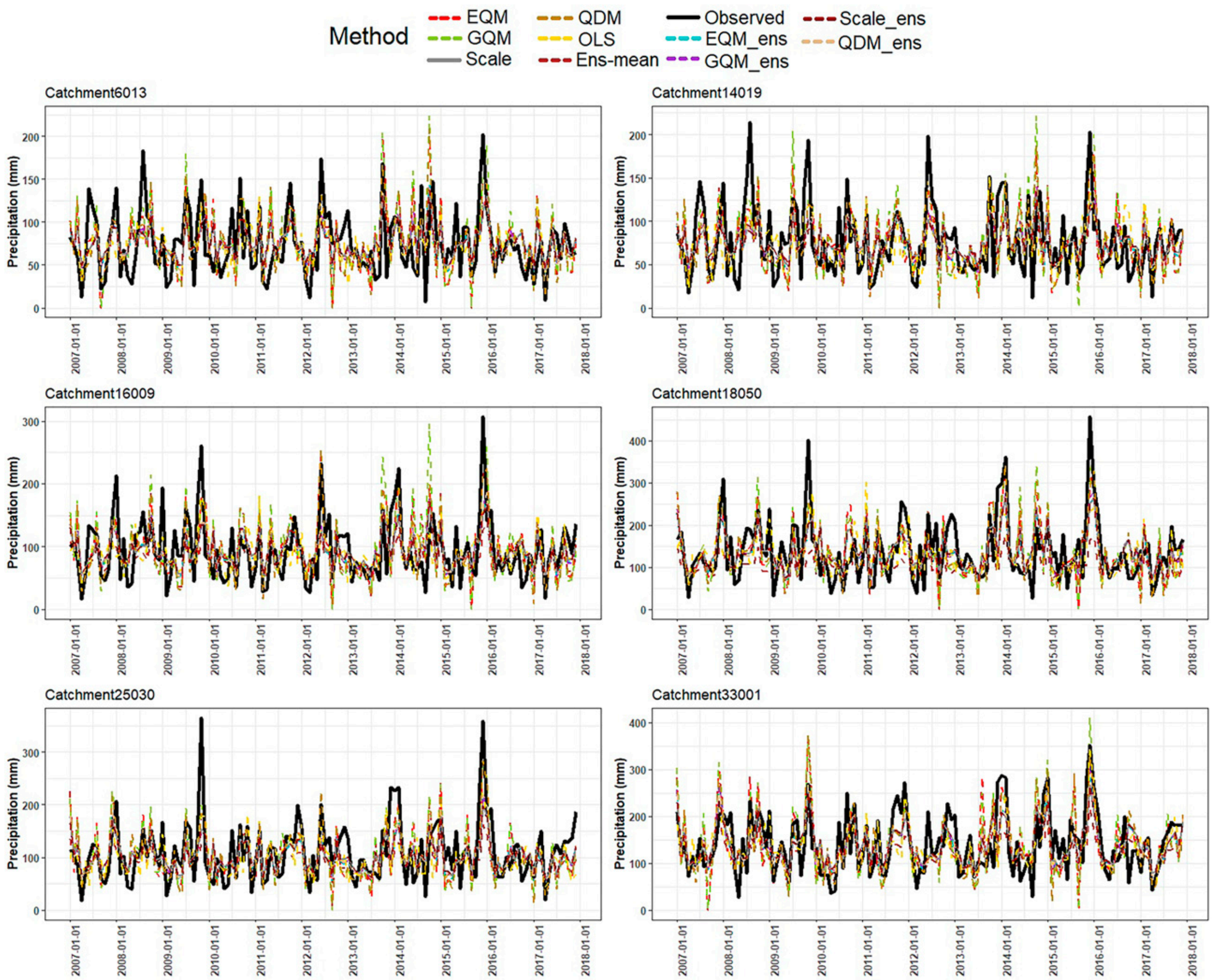

FIG. 3. Time series of forecasted precipitation with 1-month lead time from different bias-correction methods and raw ensemble mean and observed precipitation for six sample catchments.

4) Quantile delta mapping (QDM)

QDM, preserves the relative changes in quantiles of modeled precipitation while at the same time the algorithm seeks to correct the systematic biases in quantiles of the modeled precipitation with respect to observed values (Cannon et al. 2015). The relative change term at month m [$\Delta_{\rm orig}(m)$] is calculated first [Eq. (5)] and then bias-corrected precipitation is derived using Eqs. (6) and (7):

$$\Delta_{\rm orig}(m) = \frac{P_{\rm orig,test}(m)}{F^{-1}_{\rm orig,cal}[F^{m}_{\rm orig,test}\{P_{\rm orig,test}(m)\}]}, \qquad (5)$$

$$P_{\rm rev,cal}(m) = F^{-1}_{\rm m,obs}[F^{(m)}_{\rm orig,test}\{P_{\rm orig,test}(m)\}], \qquad (6)$$

$$P_{\rm rev,test}(m) = P_{\rm rev,cal}(m)\Delta_{\rm orig}(m), \qquad (7)$$

where $P_{\rm orig,test}(m)$ is the average precipitation ensemble for month m, $F^{-1}_{\rm orig,cal}$ is the inverse cumulative distribution function of the of raw data from dynamical models for calibration (baseline) period, $F^{m}_{\rm orig,test}$ is the cumulative distribution function of raw data from dynamical models for test period, $P_{\rm rev,cal}$ is the bias-corrected data for the calibration (baseline) period, and finally, $P_{\rm rev,test}(m)$ is the bias-corrected precipitation at month m of the test (validation) period.

5) Ordinary least squares (OLS) combination method

OLS is a regression-based method which combines individual forecast ensembles linearly to predict target (observed) values using the following equation:

$$P_{\rm obs}(m) = \beta_0 + \sum_{i=1}^{E}\beta_i P^{(m)}_{{\rm orig},i} + \varepsilon, \qquad (10)$$

where $P_{\rm obs}(m)$ is the target (observed) precipitation at month m, $P^{(m)}_{{\rm orig},i}$ are ith ensemble predicted precipitation from the dynamical model, E is total number of ensembles (E = 25

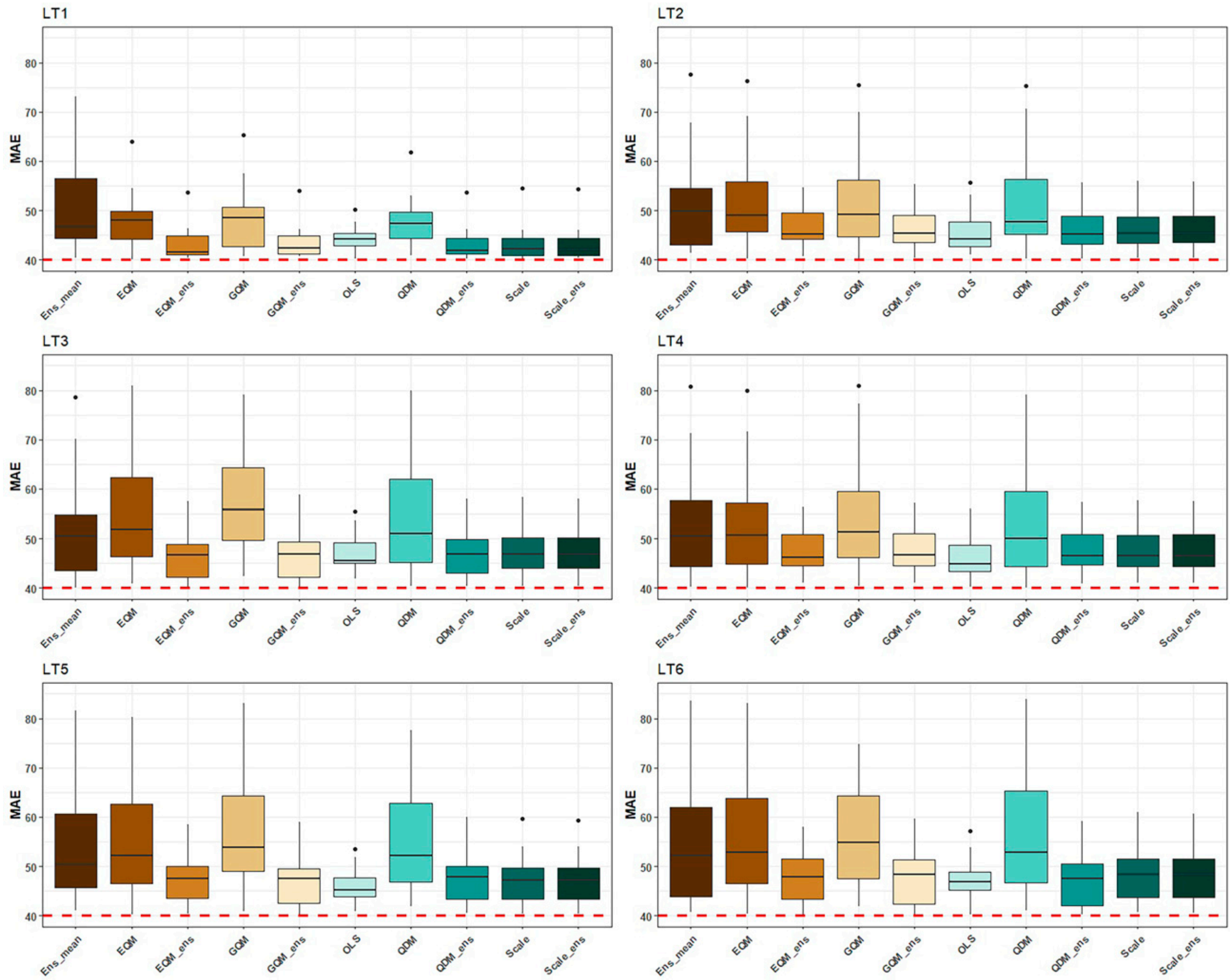

FIG. 4. Mean absolute error (MAE) for different bias-correction methods over 44 study catchments. Bias-correction methods applied on individual ensemble members first and then averaged have the "_ens" extension on their names. The red dashed line shows the value for MAE = 40.0 to make comparison easier.

for SEAS5 model), β_i are coefficients of the regression model, and ε is the residual (error) term. OLS estimates the coefficients in a way to minimize the sum of squared errors. Unlike the schemes 1 and 2 where all ensembles are used (in scheme 1, average of all ensembles are calculated first and then bias-corrected, and in scheme 2, all individual ensembles are bias-corrected first and then averaged), in the OLS method only those selected individual ensembles are utilized to develop the regression equation to calculate the bias-corrected precipitation. An exhaustive search method (Golian et al. 2022) was used to select the best combination of predictors (here ensemble members) to be used in the OLS method.

c. Evaluation of methods

Based on bias-corrected precipitation with different lead times and the observed precipitation datasets, various statistics are used to evaluate the performance of different bias-correction methods. We estimate parameters of bias-correction methods over the calibration period (1981–2006) and applied the models to the validation period (2007–16). These criteria include the correlation coefficient, mean absolute error (MAE), and coefficient of variation (CV). The ideal value for the correlation coefficient and MAE is 1 and 0, respectively, while for CV the closer the bias-corrected CV to the observed, the better. In addition to deriving these performance criteria for the entire observed and bias-corrected time series for the validation period, we also derive them for seasonal precipitation series. Finally, we also compare the performance of different methods with observed precipitation in terms of high, medium, and low precipitation values, which are defined as precipitation with a probability of exceedance of 5%, 50%, and 95%, respectively.

3. Results and discussion

a. Performance of bias-correction methods and schemes

The methods described in section 2 were applied to bias correct precipitation from SEAS5. These methods include the

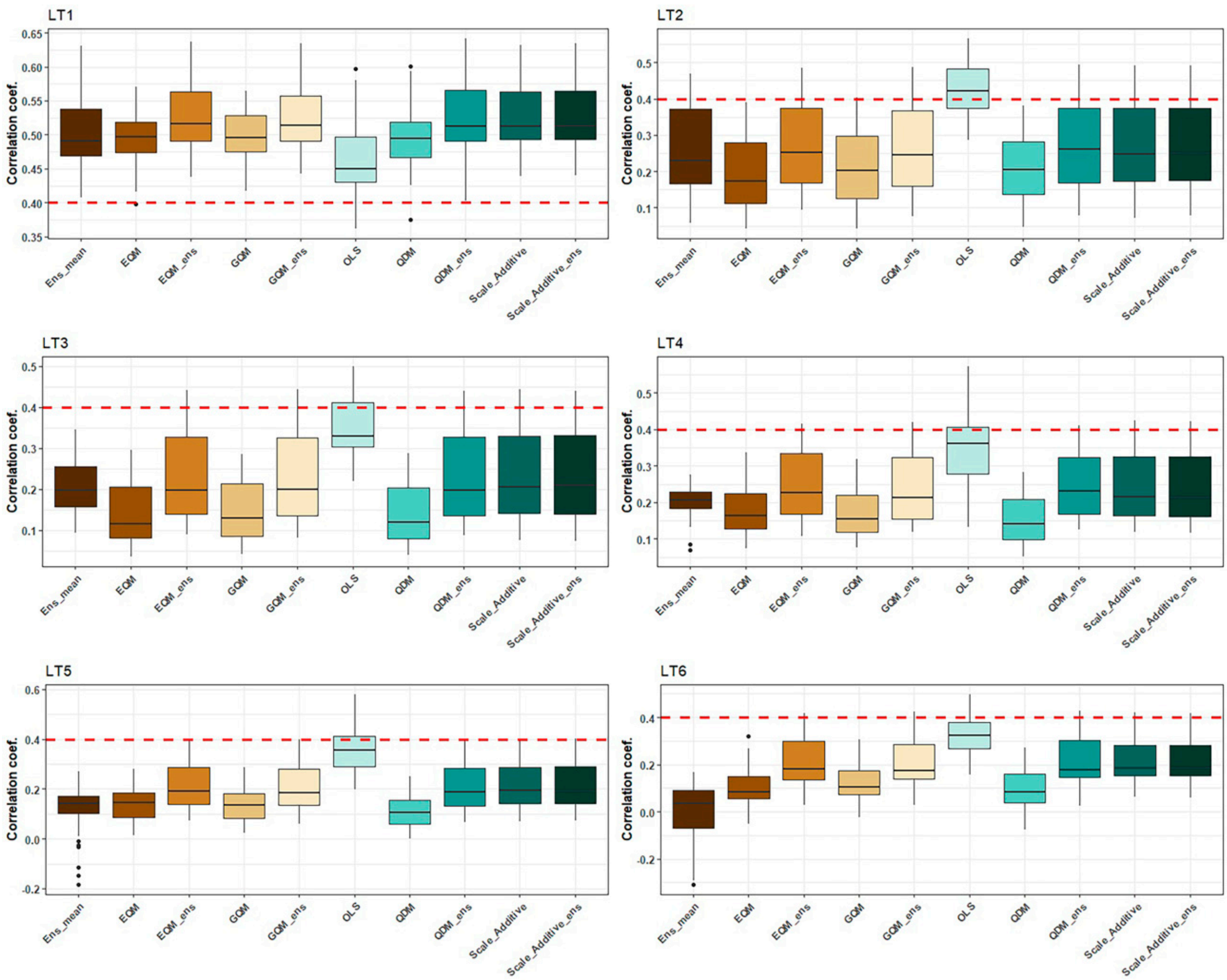

FIG. 5. Correlation coefficient between observed and bias-corrected precipitation over different lead times. Bias-correction methods applied on individual ensemble members first and then averaged are shown with "_ens" extension on their names. The red dashed line shows the value of correlation = 0.4 for the sake of convenience in their comparison.

scaling method (Scale), EQM, GQM, and QDM, which all were applied to mean ensemble. The same methods were also applied to individual ensembles first and then averaged. We also used the OLS method with selected ensemble members while the raw ensemble mean precipitation without bias correction (Ens_mean) was used as the benchmark method.

Figure 3 shows the time series of different bias-corrected precipitation series with 1-month lead time for six sample catchments as an example. It can be seen that some methods, e.g., EQM_ens, GQM_ens, and QDM_ens provide more accurate precipitation when compared to observed precipitation, especially for more frequent precipitation values.

Using observed data, mean absolute error (MAE) and the coefficient of variation are employed to measure the accuracy of bias-correction methods. Figures 4–6 show the performance of different methods. The dashed red line is plotted as a scale to compare the performance of different methods more conveniently. From Fig. 4, almost all methods improve the MAE compared to the benchmark (Ens_mean) for 1-month lead time (LT1), but over other LTs, GQM followed by QDM and EQM have the worst performance, which shows the weaker performance of these methods in bias-correcting of the bulk of precipitation values, i.e., values around the mean.

Applying bias-correction methods to individual ensemble members first and then calculating the average precipitation improves the performance of all methods, especially GQM, QDM, and EQM. For LT1, QDM_ens, Scale, OLS, and other ensemble-based methods have the best performance. For other LTs, OLS followed by scale and ensemble-based methods outperform other methods. However, there is no specific pattern as to which bias-correction method systematically performs better over different lead times. There have not been many studies on the relationship between lead time and performance of bias-correction methods at seasonal time scales. Li et al. (2019) examined the efficiency of bias-correction methods up to 11 days. Crespi et al. (2021) evaluated the performance of different bias-correction methods in

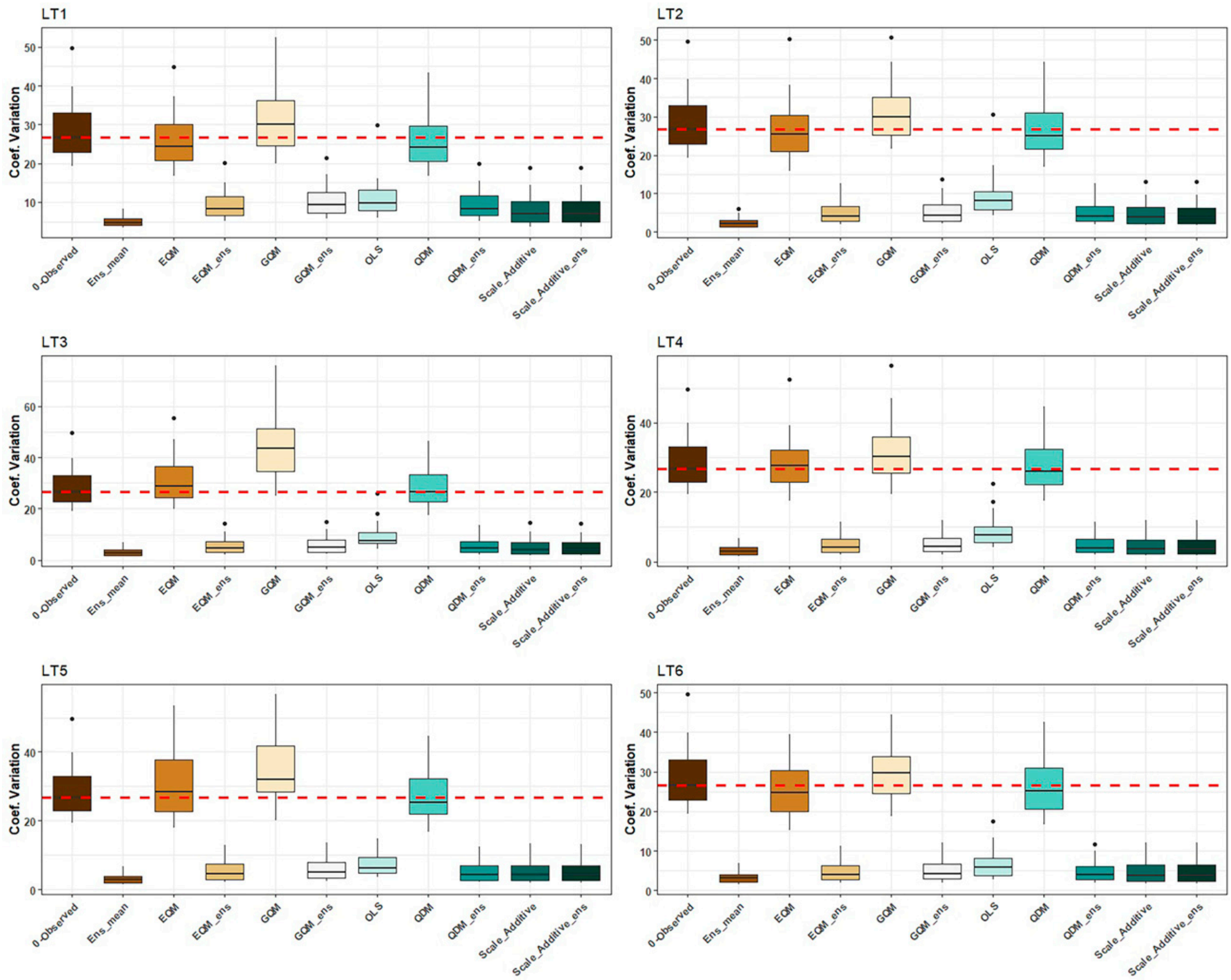

FIG. 6. Coefficient of variation (CV) for different bias-correction methods over all study catchments. The red dashed line shows the median CV for observed data.

improving precipitation and temperature from the SEAS5 model with 1-month lead time, but did not examine other lead times.

Using Spearman's correlation coefficient between observed and bias-corrected precipitation with different lead times, it can be seen from Fig. 5 that all methods show very similar performance for LT1, but for longer LTs, OLS outperforms other methods, while QDM has the worst performance. The correlation score of the raw ensemble mean (Ens_mean) decreases as lead time increases from 1 to 6 months. In general, except for 1-month lead time, correlation values are very small. Gubler et al. (2020) derived similar correlation coefficients for precipitation forecasts with 1-month lead time over South America with correlation values less than 0.5 for most regions. Crespi et al. (2021) also showed very low positive and negative correlations between reference precipitation, i.e., ERA5 and forecasted precipitation from SEAS5 with 1-month lead time for most parts of Europe (correlation spanning from −0.4 to 0.4 for most regions), without clear spatial dependency.

From Fig. 6 the CV is closer to observed values for those bias-correction methods which are applied to the mean ensemble. EQM and QDM followed by GQM outperform other methods. These methods better preserve the relative dispersion of bias-corrected precipitation around the mean. It was also shown by Son et al. (2017) that nonparametric bias-correction methods, e.g., EQM, provided the best results in increasing the accuracy of GloSea5 precipitation forecasts over South Korea. The reason can be related to the fact that when bias-correction methods are applied to individual ensemble-members first and then averaged, the resulting values have a tendency toward average precipitation. Some ensemble members tend to have much higher/lower values and when they are averaged, the result preserves the mean (average) statistics. In this way, the relative dispersion of data points around the mean is lower, resulting in lower CV values, as shown in Fig. 6.

In general, bias-correction methods performed better for 1-month lead time (LT1) compared to other LTs based on correlation and MAE criteria. They revealed the worst

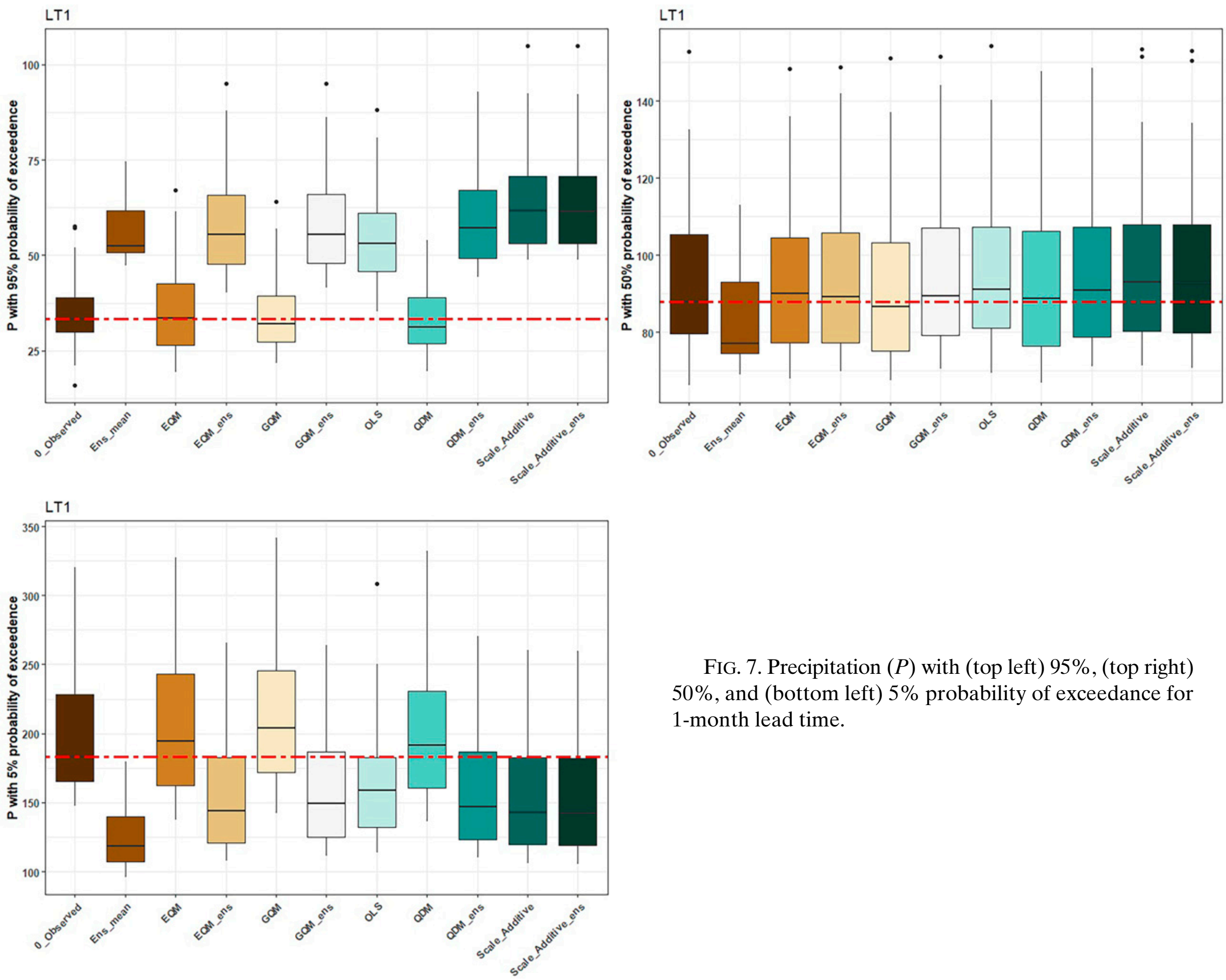

FIG. 7. Precipitation (P) with (top left) 95%, (top right) 50%, and (bottom left) 5% probability of exceedance for 1-month lead time.

performance for LT6, but there is no specific rule/pattern for other LTs. This can be related to the fact that the performance of precipitation hindcast from SEAS5 model is much better for LT1 compared to other lead times (e.g., Fig. S1 in the online supplemental material for correlation coefficient). Zhao et al. (2017) also showed that at longer lead times, the efficiency of applying bias-correction postprocessing of forecasts decreases compared to the case with 0-month lead time. Also, compared to the Ens_mean, it can be seen that some methods have similar or even worse performance based on correlation coefficient and MAE, e.g., EQM, GQM, and QDM for LT1. In another study, Crochemore et al. (2016) showed that resulting biases vary more with the calendar month of the forecast horizon than with lead time.

b. *Assessment of bias-correction methods for extreme precipitation values*

To examine if bias-correction methods and schemes can preserve extreme precipitation values, the precipitation with 5% and 95% exceedance probabilities were calculated from bias-corrected precipitation time series for the test period and results compared to the observed values over different catchments and lead times. Results are illustrated in Figs. 7 and 8 for 1- and 6-month lead times as examples. For both extreme conditions, i.e., low and high precipitation, ordinary quantile-based methods, i.e., EQM, GQM, and QDM applied to ensemble mean outperformed those bias-correction methods applied to individual ensemble members first and then averaged. Also, ordinary EQM, GQM, and QDM outperformed the other methods, i.e., OLS and scale methods for all lead times. When applying bias-correction methods to individual ensembles first and then averaging them, this averaging tends to neutralize the extreme information which is inherent in some ensemble members, with averaging of individual members tending toward median/mean values.

c. *Performance of bias-correction methods at seasonal scales*

Finally, the performance of different bias-correction methods across seasons (winter, DJF; spring, MAM; summer, JJA; and autumn, SON) is shown in Fig. 9 in term of the median MAE for all study catchments. EQM, GQM, and QDM

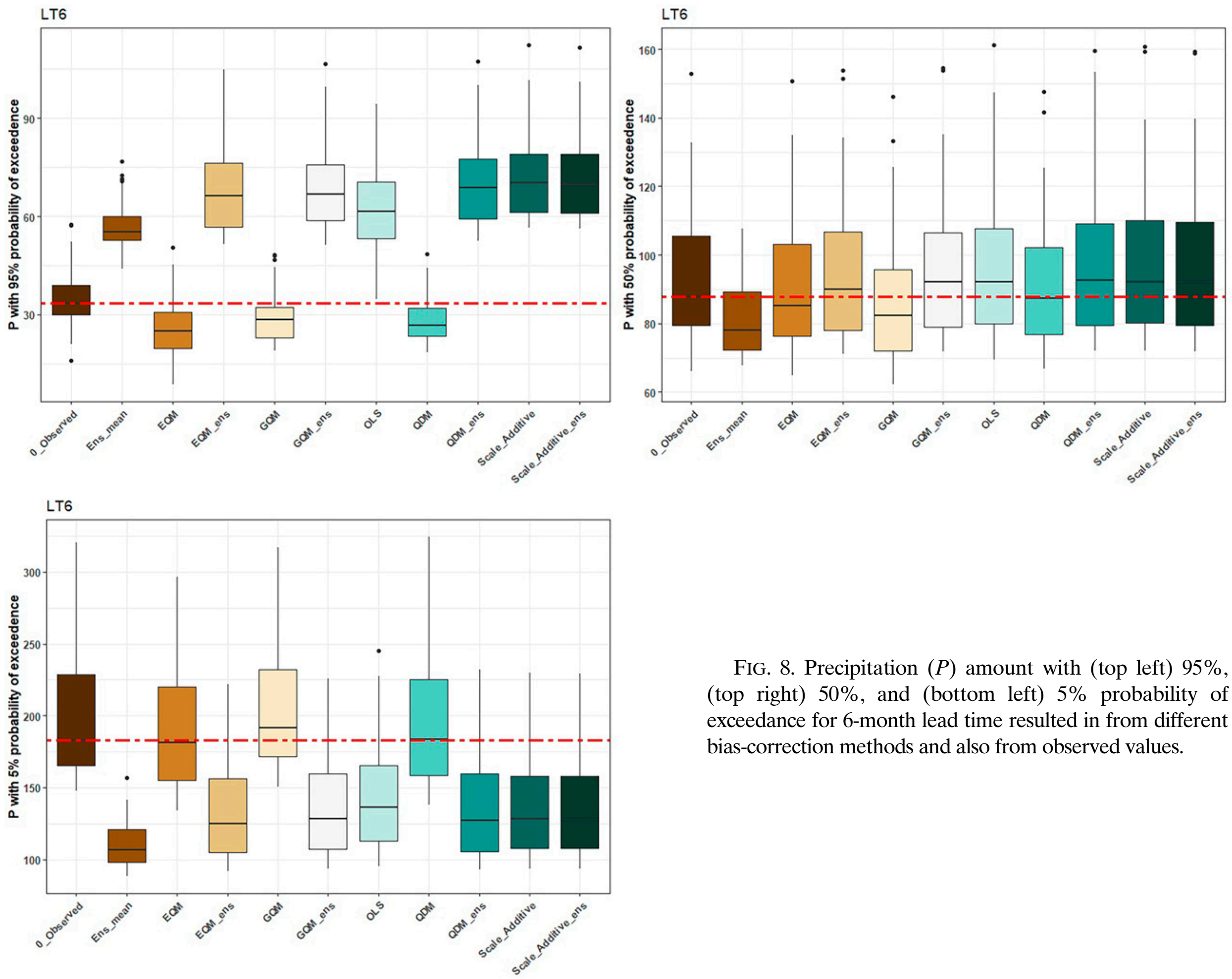

FIG. 8. Precipitation (P) amount with (top left) 95%, (top right) 50%, and (bottom left) 5% probability of exceedance for 6-month lead time resulted in from different bias-correction methods and also from observed values.

applied to the ensemble mean in winter and summer; QDM in spring; and GQM followed by EQM and QDM in autumn had the worst performance. In spring and summer, the MAE values for all bias-correction methods are smaller than and perform more closely over different lead times compared to winter and autumn. This is because MAE is related to precipitation and therefore smaller MAE is expected in drier seasons. Overall, for all seasons, the performance of bias-correction methods is better for LT1 compared to other LTs. Based on the correlation coefficient (Fig. 10), all methods perform better for LT1, while the performance of bias-correction methods is better in winter compared to other seasons. Tong et al. (2021) also showed that the effects of bias correction are season dependent, performing better in the wet season in their study region. This might be related to the fact that dynamical models provide less accurate precipitation forecasts in spring and summer than wet seasons, i.e., winter and autumn due to difficulties in modeling convective rainfall (Lenderink et al. 2007) and this weaker performance is transferred to bias-correction methods over these seasons too. Zarei et al. (2021) also showed that their bias-correction methods, i.e., quantile mapping and random forests revealed better performance in improving precipitation forecasts in winter and autumn compared to summer and spring.

In general, while it has been shown by previous studies that bias-correction methods based on quantile mapping can improve precipitation estimates from regional climate models (e.g., Jakob Themeßl et al. 2011; Enayati et al. 2021), our results show that regression based methods, e.g., OLS in our study can perform as well as or better in improving the accuracy of precipitation from dynamical climate models based on correlation and MAE criteria. However, this resulted in poor performance in correcting extreme high and low precipitation, i.e., precipitation with 5% and 95% exceedance probabilities. Also, while most climatemodels provide ensembles of forecasts/hindcasts as their outputs, our study showed that applying bias-correction methods to individual members first and then averaging the results can further improve the outputs of dynamical models in terms of correlation and MAE criteria. However, when it comes to extreme high or low precipitation, quantile-based methods applied to the ensemble mean provides more skillful results

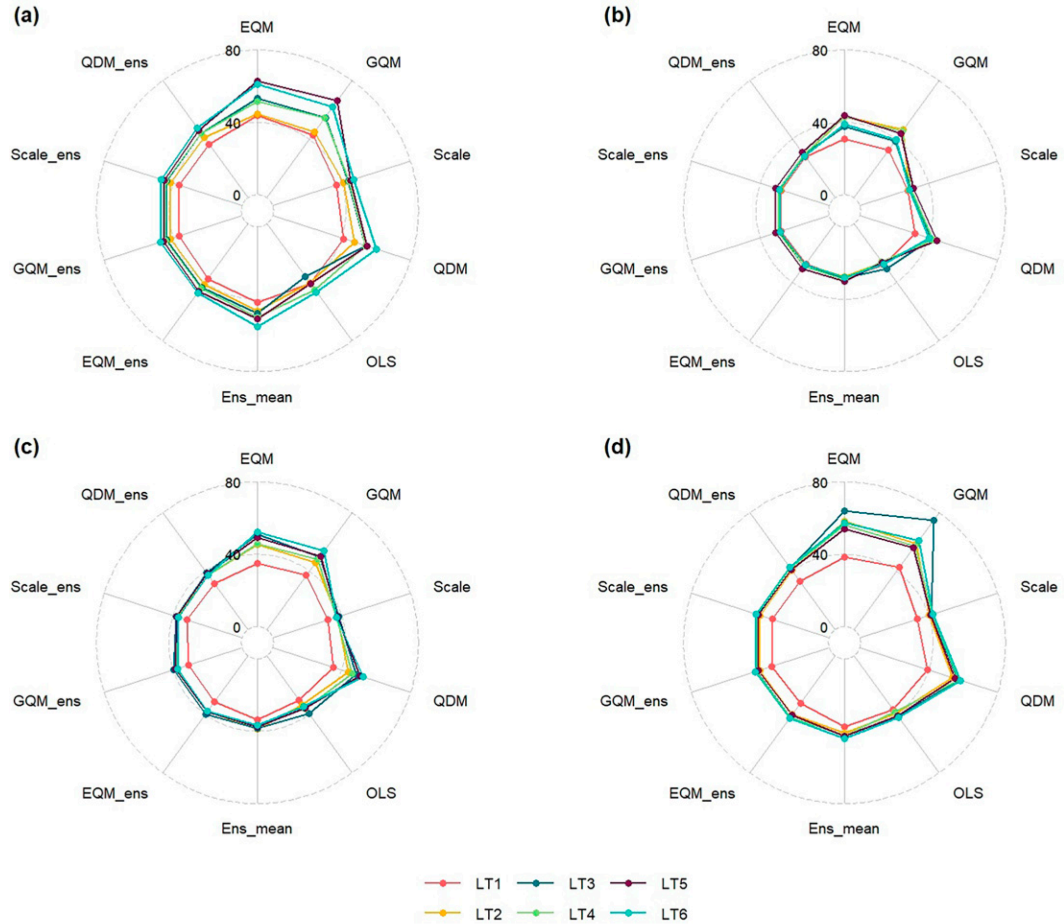

FIG. 9. Median MAE across all study catchments for different lead times and bias-correction methods in (a) winter, (b) spring, (c) summer, and (d) autumn.

compared to the same methods applied to individual ensembles first and then averaged.

4. Conclusions

We compared the performance of five different bias-correction methods in improving the accuracy of precipitation forecasts from the European Center for Medium Range Forecasts (ECMWF) System 5 (SEAS5) with 1–6-month lead times. Bias-correction methods were applied to (i) the ensemble mean and (ii) individual ensemble members that were then averaged, to examine differences between bias-correction methods under both schemes. Using multiple evaluation criteria, the performance of bias-correction methods were evaluated monthly and seasonally. Applying bias correction to individual ensemble members (scheme 2) and the OLS method applied to selected ensembles provide the best performance in terms of correlation and MAE for most precipitation values distributed around the mean. However, for extreme precipitation application of simple methods like EQM, QDM, and GQM to the ensemble mean (scheme 1) are more skillful. All methods perform better over 1-month lead time (LT1) compared to other lead times. Bias-correction performance is best in winter relative to other seasons.

Bias-correction methods perform more similarly over different lead times in drier seasons, i.e., spring and summer compared with wetter winter and autumn seasons. It is difficult to identify a single best-performing bias-correction method for a specific lead time and season.

Given that the ensemble-based method, i.e., applying the bias-correction methods to individual ensembles first and then calculate the average of corrected precipitation improves the accuracy of hindcasted precipitation from SEAS5 based on correlation and MAE criteria, we conclude that this ensemble bias-correction scheme is useful for the bulk of precipitation values, e.g., values around the mean, but not for extreme values. For low and high precipitation, ordinary quantile-based bias-correction methods can directly be applied to the ensemble mean. Hence, developing a hybrid method based on both simple and ensemble-based bias-correction methods may be an interesting subject for future research.

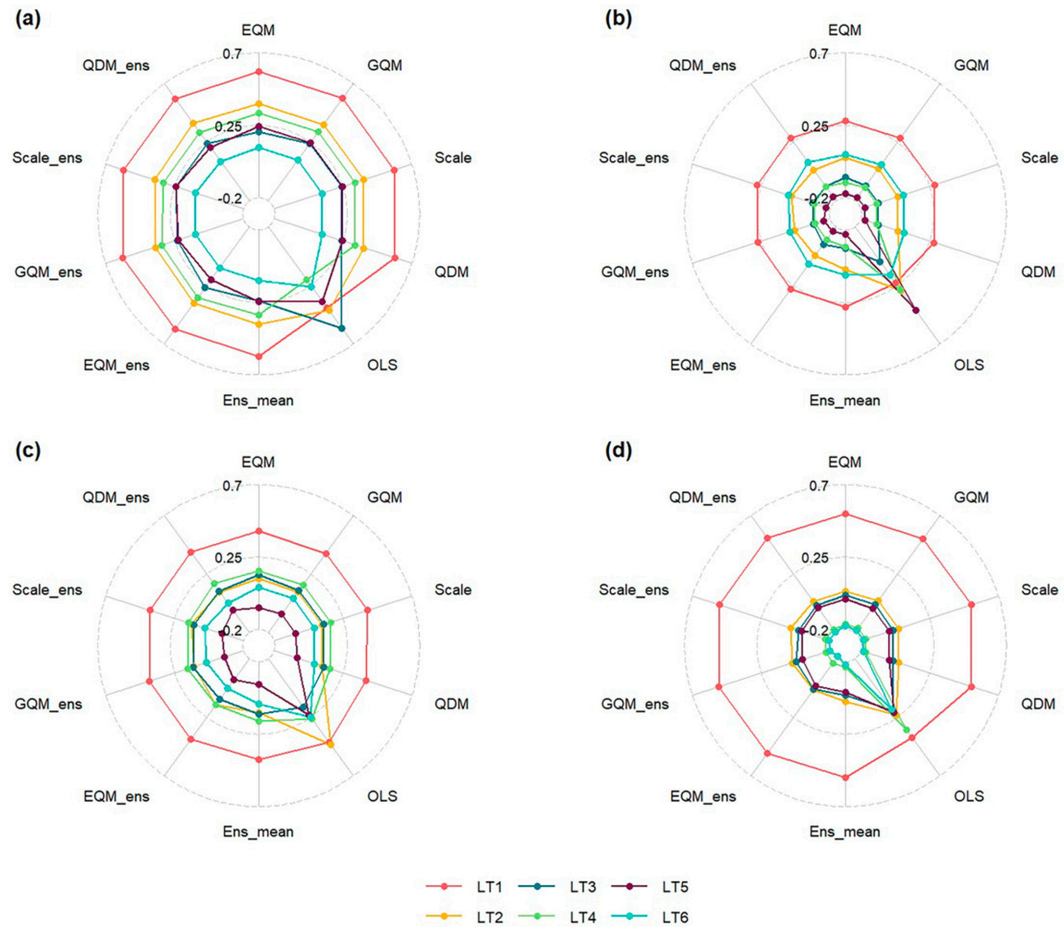

FIG. 10. The average correlation coefficient between observed and bias-corrected precipitations over all study catchments for different lead times and bias-correction methods in (a) winter, (b) spring, (c) summer, and (d) autumn.

Acknowledgments. C.M. was funded by a Science Foundation Ireland Career Development Award (Grant SFI/17/CDA/4783).

Data availability statement. All seasonal hindcasts from SEAS5 used in this study are openly available from the ECMWF website at https://www.ecmwf.int/en/computing/software/ecmwf-web-api.

REFERENCES

Arnal, L., H. L. Cloke, E. Stephens, F. Wetterhall, C. Prudhomme, J. Neumann, B. Krzeminski, and F. Pappenberger, 2018: Skilful seasonal forecasts of streamflow over Europe? *Hydrol. Earth Syst. Sci.*, **22**, 2057–2072, https://doi.org/10.5194/hess-22-2057-2018.

Atger, F., 2001: Verification of intense precipitation forecasts from single models and ensemble prediction systems. *Nonlinear Processes Geophys.*, **8**, 401–417, https://doi.org/10.5194/npg-8-401-2001.

Baker, S. A., A. W. Wood, and B. Rajagopalan, 2019: Developing subseasonal to seasonal climate forecast products for hydrology and water management. *J. Amer. Water Resour. Assoc.*, **55**, 1024–1037, https://doi.org/10.1111/1752-1688.12746.

Bhatti, H. A., T. Rientjes, A. T. Haile, E. Habib, and W. Verhoef, 2016: Evaluation of bias correction method for satellite-based rainfall data. *Sensors*, **16**, 884, https://doi.org/10.3390/s16060884.

Blench, R., 1999: Seasonal climatic forecasting: WHO can use it and how should it be disseminated? *National Resource Perspectives*, No. 47, Overseas Development Institute, London, United Kingdom, 4 pp., https://cdn.odi.org/media/documents/2871.pdf.

Buizza, R., 2018: Ensemble forecasting and the need for calibration. *Statistical Postprocessing of Ensemble Forecasts*, 1st ed. S. Vannitsem et al., Eds., Elsevier, 15–48.

Calanca, P., D. Bolius, A. P. Weigel, and M. A. Liniger, 2011: Application of long-range weather forecasts to agricultural decision problems in Europe. *J. Agric. Sci.*, **149**, 15–22, https://doi.org/10.1017/S0021859610000729.

Cannon, A. J., S. R. Sobie, and T. Q. Murdock, 2015: Bias correction of GCM precipitation by quantile mapping: How well do methods preserve changes in quantiles and extremes? *J. Climate*, **28**, 6938–6959, https://doi.org/10.1175/JCLI-D-14-00754.1.

Crespi, A., M. Petitta, P. Marson, C. Viel, and L. Grigis, 2021: Verification and bias adjustment of ECMWF SEAS5 seasonal forecasts over Europe for climate service applications. *Climate*, **9**, 181, https://doi.org/10.3390/cli9120181.

Crochemore, L., M.-H. Ramos, and F. Pappenberger, 2016: Bias correcting precipitation forecasts to improve the skill of seasonal streamflow forecasts. *Hydrol. Earth Syst. Sci.*, **20**, 3601–3618, https://doi.org/10.5194/hess-20-3601-2016.

Donegan, S., and Coauthors, 2021: Conditioning ensemble streamflow prediction with the North Atlantic Oscillation improves skill at longer lead times. *Hydrol. Earth Syst. Sci.*, **25**, 4159–4183, https://doi.org/10.5194/hess-25-4159-2021.

Enayati, M., O. Bozorg-Haddad, J. Bazrafshan, S. Hejabi, and X. Chu, 2021: Bias correction capabilities of quantile mapping methods for rainfall and temperature variables. *J. Water Climate Change*, **12**, 401–419, https://doi.org/10.2166/wcc.2020.261.

Foran Quinn, D., C. Murphy, R. L. Wilby, T. Matthews, C. Broderick, S. Golian, S. Donegan, and S. Harrigan, 2021: Benchmarking seasonal forecasting skill using river flow persistence in Irish catchments. *Hydrol. Sci. J.*, **66**, 672–688, https://doi.org/10.1080/02626667.2021.1874612.

Ghimire, U., G. Srinivasan, and A. Agarwal, 2019: Assessment of rainfall bias correction techniques for improved hydrological simulation. *Int. J. Climatol.*, **39**, 2386–2399, https://doi.org/10.1002/joc.5959.

Golian, S., C. Murphy, and H. Meresa, 2021: Regionalization of hydrological models for flow estimation in ungauged catchments in Ireland. *J. Hydrol. Reg. Stud.*, **36**, 100859, https://doi.org/10.1016/j.ejrh.2021.100859.

——, ——, R. L. Wilby, T. Matthews, S. Donegan, D. F. Quinn, and S. Harrigan, 2022: Dynamical–statistical seasonal forecasts of winter and summer precipitation for the Island of Ireland. *Int. J. Climatol.*, https://doi.org/10.1002/joc.7557, in press.

Gubler, S., and Coauthors, 2020: Assessment of ECMWF SEAS5 seasonal forecast performance over South America. *Wea. Forecasting*, **35**, 561–584, https://doi.org/10.1175/WAF-D-19-0106.1.

Jakob Themeßl, M., A. Gobiet, and A. Leuprecht, 2011: Empirical-statistical downscaling and error correction of daily precipitation from regional climate models. *Int. J. Climatol.*, **31**, 1530–1544, https://doi.org/10.1002/joc.2168.

Kim, H., Y. G. Ham, Y. S. Joo, and S. W. Son, 2021: Deep learning for bias correction of MJO prediction. *Nat. Commun.*, **12**, 3087, https://doi.org/10.1038/s41467-021-23406-3.

Lenderink, G., A. Buishand, and W. van Deursen, 2007: Estimates of future discharges of the river Rhine using two scenario methodologies: Direct versus delta approach. *Hydrol. Earth Syst. Sci.*, **11**, 1145–1159, https://doi.org/10.5194/hess-11-1145-2007.

Li, W., J. Chen, L. Li, H. Chen, B. Liu, C.-Y. Xu, and X. Li, 2019: Evaluation and bias correction of S2S precipitation for hydrological extremes. *J. Hydrometeor.*, **20**, 1887–1906, https://doi.org/10.1175/JHM-D-19-0042.1.

Lorenz, C., T. C. Portele, P. Laux, and H. Kunstmann, 2021: Bias-corrected and spatially disaggregated seasonal forecasts: A long-term reference forecast product for the water sector in semi-arid regions. *Earth Syst. Sci. Data*, **13**, 2701–2722, https://doi.org/10.5194/essd-13-2701-2021.

Luo, M., T. Liu, F. Meng, Y. Duan, A. Frankl, A. Bao, and P. De Maeyer, 2018: Comparing bias correction methods used in downscaling precipitation and temperature from regional climate models: A case study from the Kaidu River basin in western China. *Water*, **10**, 1046, https://doi.org/10.3390/w10081046.

Maity, R., M. Suman, P. Laux, and H. Kunstmann, 2019: Bias correction of zero-inflated RCM precipitation fields: A copula-based scheme for both mean and extreme conditions. *J. Hydrometeor.*, **20**, 595–611, https://doi.org/10.1175/JHM-D-18-0126.1.

Mendez, M., B. Maathuis, D. Hein-Griggs, and L.-F. Alvarado-Gamboa, 2020: Performance evaluation of bias correction methods for climate change monthly precipitation projections over Costa Rica. *Water*, **12**, 482, https://doi.org/10.3390/w12020482.

Moghim, S., and R. L. Bras, 2017: Bias correction of climate modeled temperature and precipitation using artificial neural networks. *J. Hydrometeor.*, **18**, 1867–1884, https://doi.org/10.1175/JHM-D-16-0247.1.

Monhart, S., C. Spirig, J. Bhend, K. Bogner, C. Schär, and M. A. Liniger, 2018: Skill of subseasonal forecasts in Europe: Effect of bias correction and downscaling using surface observations. *J. Geophys. Res. Atmos.*, **123**, 7999–8016, https://doi.org/10.1029/2017JD027923.

Osgood, D. E., P. Suarez, J. Hansen, M. Carriquiry, and A. Mishra, 2008: Integrating seasonal forecasts and insurance for adaptation among subsistence farmers: The case of Malawi. Policy Research Working Paper WPS 4651, World Bank Group, 30 pp., https://documents.worldbank.org/en/publication/documents-reports/documentdetail/893941468046487650/integrating-seasonal-forecasts-and-insurance-for-adaptation-among-subsistence-farmers-the-case-of-malawi.

Piani, C., G. P. Weedon, M. Best, S. M. Gomes, P. Viterbo, S. Hagemann, and J. O. Haerter, 2010: Statistical bias correction of global simulated daily precipitation and temperature for the application of hydrological models. *J. Hydrol.*, **395**, 199–215, https://doi.org/10.1016/j.jhydrol.2010.10.024.

Qian, S., J. Chen, X. Li, C.-Y. Xu, S. Guo, H. Chen, and X. Wu, 2020: Seasonal rainfall forecasting for the Yangtze River basin using statistical and dynamical models. *Int. J. Climatol.*, **40**, 361–377, https://doi.org/10.1002/joc.6216.

Ratri, D. N., K. Whan, and M. Schmeits, 2019: A comparative verification of raw and bias-corrected ECMWF seasonal ensemble precipitation reforecasts in Java (Indonesia). *J. Appl. Meteor. Climatol.*, **58**, 1709–1723, https://doi.org/10.1175/JAMC-D-18-0210.1.

Reiter, P., O. Gutjahr, L. Schefczyk, G. Heinemann, and M. Casper, 2016: Bias correction of ENSEMBLES precipitation data with focus on the effect of the length of the calibration period. *Meteor. Z.*, **25**, 85–96, https://doi.org/10.1127/metz/2015/0714.

Son, C., J. Song, S. Kim, and Y. Cho, 2017: A selection of optimal method for bias-correction in Global Seasonal Forecast System version 5 (GloSea5). *J. Korea Water Resour. Assoc.*, **50**, 551–562, https://doi.org/10.3741/JKWRA.2017.50.8.551.

Stockdale, T. N., and Coauthors, 2010: Understanding and predicting seasonal-to-interannual climate variability-the producer perspective. *Procedia Environ. Sci.*, **1**, 55–80, https://doi.org/10.1016/j.proenv.2010.09.006.

Teutschbein, C., and J. Seibert, 2012: Bias correction of regional climate model simulations for hydrological climate-change impact studies: Review and evaluation of different methods. *J. Hydrol.*, **456**, 12–29, https://doi.org/10.1016/j.jhydrol.2012.05.052.

Tong, Y., X. Gao, Z. Han, Y. Xu, Y. Xu, and F. Giorgi, 2021: Bias correction of temperature and precipitation over China for RCM simulations using the QM and QDM methods. *Climate Dyn.*, **57**, 1425–1443, https://doi.org/10.1007/s00382-020-05447-4.

Troccoli, A., 2010: Seasonal climate forecasting. *Meteor. Appl.*, **17**, 251–268, https://doi.org/10.1002/met.184.

Walsh, S., 2012. New long-term rainfall averages for Ireland. *13th Annual Irish National Hydrology Conf.*, Tullamore, Ireland, International Hydrological Programme, 3–12.

Yang, C., H. Yuan, and X. Su, 2020: Bias correction of ensemble precipitation forecasts in the improvement of summer streamflow prediction skill. *J. Hydrol.*, **588**, 124955, https://doi.org/10.1016/j.jhydrol.2020.124955.

Zarei, M., M. Najarchi, and R. Mastouri, 2021: Bias correction of global ensemble precipitation forecasts by Random Forest method. *Earth Sci. Inf.*, **14**, 677–689, https://doi.org/10.1007/s12145-021-00577-7.

Zhao, T., J. C. Bennett, Q. J. Wang, A. Schepen, A. W. Wood, D. E. Robertson, and M.-H. Ramos, 2017: How suitable is quantile mapping for postprocessing GCM precipitation forecasts? *J. Climate*, **30**, 3185–3196, https://doi.org/10.1175/JCLI-D-16-0652.1.