Statistical Assessment and Augmentation of European Centre for Medium-Range Weather Forecasts Monthly Precipitation Forecast (SEASonal Prediction of Precipitation)

Nasseri, Mohsen; Schoups, Gerrit; Taheri, Mercedeh

DOI
10.1002/joc.8723

Publication date
2025

Document Version
Final published version

Published in
International Journal of Climatology

Citation (APA)
Nasseri, M., Schoups, G., & Taheri, M. (2025). Statistical Assessment and Augmentation of European Centre for Medium-Range Weather Forecasts Monthly Precipitation Forecast (SEASonal Prediction of Precipitation). *International Journal of Climatology*, *45*(3), Article e8723. https://doi.org/10.1002/joc.8723

International Journal of Climatology

International Journal of Climatology RMetS

RESEARCH ARTICLE

Statistical Assessment and Augmentation of European Centre for Medium-Range Weather Forecasts Monthly Precipitation Forecast (SEASonal Prediction of Precipitation)

Mohsen Nasseri[1] | Gerrit Schoups[2] | Mercedeh Taheri[3]

[1]School of Civil Engineering, College of Engineering, University of Tehran, Tehran, Iran | [2]Department of Water Management, Faculty of Civil Engineering and Geosciences, Delft University of Technology, Delft, The Netherlands | [3]Department of Civil Engineering, University of Ottawa, Ottawa, Ontario, Canada

Correspondence: Mohsen Nasseri (mnasseri@ut.ac.ir)

Received: 12 October 2023 | **Revised:** 3 December 2024 | **Accepted:** 4 December 2024

Funding: The first author was supported by Visitor's Travel Grant 040.11.731 from the Dutch Research Council (NWO).

Keywords: augmentation method | bias correction | flexible Least Square | ordinary Least Square | Q-Q regression | SEAS5

ABSTRACT

Accurate prediction of precipitation is of paramount importance for effective planning of future water resources. In this study, we focused on the improvement and evaluation of the European Centre for Medium-Range Weather Forecasts (ECMWF) fifth-generation ensemble-based seasonal precipitation prediction product, designated (SEASonal prediction of precipitation (SEAS5)). Three selected linear regression methods, namely ordinary least squares (OLS), flexible least squares (FLS) and the quantile-quantile (Q-Q) methods, were used to develop a correction procedure. The watershed of Lake Urmia was selected as a case study. The application of these augmentation methods has yielded encouraging results, demonstrating an improvement in the statistical metrics of SEAS5 precipitation forecasts for the first and second-coming months. However, all linear projection methods improve the performance of the SEAS5 products. The Q-Q method has shown the highest efficiency among the methods, playing a significant role in improving the accuracy of the hindcast precipitation. A variety of statistics (deterministic, forecast skill and uncertainty scores) were used to evaluate the effectiveness of both the raw and enhanced SEAS5 products. These analyses provide a comprehensive understanding of the performance of the SEAS5 product in its original form and after augmentation. The results highlight the potential of the linear projection method (specifically Q-Q method) to improve the accuracy of hindcast precipitation and provide valuable insights for water resource planning in the study area.

1 | Introduction

Precipitation is of vital importance in the context of water resources management and regional development programmes. In addition to temperature, precipitation is a significant hydroclimatological factor that is influenced by climate change in various geographical regions and countries (Sarojini, Stott, and Black 2016). The significance of precipitation is particularly evident in semiarid regions with substantial agricultural sectors or water-dependent units (Dorward et al. 2019). In such regions, stakeholders may enhance the certainty of their decisions through the use of reliable precipitation prediction models (Ratri, Whan, and Schmeits 2019). Consequently, accurately predicting the spatiotemporal patterns of precipitation for the near future, such as seasonal predictions, has significant potential for regional water resources

planning and management (Gbangou et al. 2019, 2020). Furthermore, the estimation of available and renewable water resources based on calibrated water balance models necessitates the availability of accurate precipitation estimates. The accuracy of precipitation measurements can be obtained from two principal sources including rain gauge records and calibrated remotely sensed (RS) datasets.

However, it should be noted that all global and regional climatological datasets (including RS data, general circulation models, reanalysis information and predictions) are subject to random and systematic biases. It is imperative that these biases are addressed to guarantee the dependable utilisation of such information in the context of hydrological simulations and water resources management (Ehret et al. 2012; Schepen, Wang, and Everingham 2016; Worku et al. 2020). In the existing literature, a variety of statistical and probabilistic techniques have been employed to reduce the bias and scale of predicted and simulated precipitation datasets. These methods can be grouped into different frameworks, including scale factor methods (Fang et al. 2015; Zhang et al. 2020), quantile regression and mapping (Gudmundsson et al. 2012; Ogutu et al. 2016; Berg, Donnelly, and Gustafsson 2018; Grillakis, Koutroulis, and Tsanis 2018; Ratri, Whan, and Schmeits 2019; Gbangou et al. 2019, 2020), linear projection methods (Fang et al. 2015; Wang et al. 2019) and other approaches (Wang et al. 2019).

The fifth generation of ensemble-based SEASonal prediction of precipitation (SEAS5) datasets from the European Centre for Medium-Range Weather Forecasts (ECMWF), which is widely recognised as a climatological prediction model, also contains biases that need to be addressed before it can be employed in hydroclimatological applications (Johnson et al. 2019; Ratri, Whan, and Schmeits 2019; Manzanas et al. 2019; Schick, Rössler, and Weingartner 2019; Chevuturi et al. 2021; Ratri, Whan, and Schmeits 2021; Golian and Murphy 2022; Ratri et al. 2023). SEAS5 represents an enhanced iteration of the ECMWF System 4 seasonal forecasts, which were operational in 2011. Extensively studied, evaluated, analysed, augmented, bias-corrected and implemented in the literature (Ogutu et al. 2016; Crochemore et al. 2017; Lucatero et al. 2018; Johnson et al. 2019; Bergman et al. 2019; Gbangou et al. 2019, 2020; Mori et al. 2021), this predecessor has informed the development of SEAS5.

In this study, we present a systematic framework that employs linear approaches for bias reduction and scale-based augmentation with the objective of calibrating SEAS5 products. The proposed augmentation method is based on the three-month-ahead predictions of monthly precipitation from the SEAS5 model. To this end, three linear regression methods using ordinary least square (OLS), flexible least square (FLS) and quantile-quantile (Q-Q) methods, are employed for the purpose of augmenting the hindcasted precipitation products. To assess the efficacy of these methodologies, the authors selected the Urmia Lake Watershed (ULW) as a case study, a region grappling with persistent water scarcity, drought and rising salinity in its central lake (Ghajarnia, Liaghat, and Arasteh 2015; Hosseini-Moghari et al. 2018; Taheri et al. 2019; Dehghanipour et al. 2019; Nasseri, Schoups, and Taheri 2022).

The following is a description of the structure of the article. Section 2 outlines the materials and methods employed, including the case study, the SEAS5 dataset (from ECMWF) and the implemented linear regression techniques (including OLS, FLS and Q-Q methods). Section 3 presents the evaluation metrics employed, including similarity/dissimilarity values, metrics for comparing distributions, indicators for assessing prediction uncertainty and prediction skills. The modelling procedure, modelling results and conclusions are presented in Sections 4, 5 and 6, respectively.

2 | Materials and Methods

2.1 | The Ground Precipitation Information

The ULW is situated in the northwestern region of Iran and is regarded as a significant watershed area, encompassing an estimated 51,761 km^2, representing 3.2% of Iran's total landmass. Furthermore, the basin is the fifth largest in Iran in terms of watershed area. Figure 1 depicts the location of the ULW and other major watersheds in Iran. The ULW is an archetypal endorheic watershed, wherein the river network amasses the streamflow from the entire watershed and subsequently channels it into the central lake, Urmia Lake. The increasing hydrological drought and salinity of the lake over the past decade have rendered it a prominent case study in international water and environmental research (Hamidi-Razi et al. 2019; Habibi, Babaeian, and Schöner 2021; Ghazi, Dutt, and Torabi Haghighi 2023).

The ULW is characterised by the presence of distinct climatic regions. Figure 1 presents the climatological classes of the watershed based on the Köppen-Geiger climate index (Kottek et al. 2006; Peel, Finlayson, and McMahon 2007). The principal climatic zone within the ULW is classified as Bsk, denoting an arid steppe cold arid climate. Additionally, snowy classes (Dfd and Dsf) represent significant regions within the watershed. Moreover, the central region and areas in proximity to Urmia Lake demonstrate a warm steppe, humid and cold climatic profile.

The observed precipitation dataset employed in the present study is derived from recorded monthly precipitation data collected at 87 rain gauge stations from March 1993 to March 2017 with a maximum of 20% missed information during the computational period (total samples: $87 \times 288 = 25{,}056$, missed value: 2637). The location of each station is indicated in Figure 1. The accuracy of the recorded data is 0.1 mm, and the network is managed by the Iranian Ministry of Power. It is noteworthy that the number of measuring stations is relatively high in the vicinity of the lake, whereas there is a paucity of stations in the eastern part of the ULW.

2.2 | Seasonal Forecast of ECMWF (SEAS5)

This research employs the SEAS5 product, which represents the fifth generation of ensemble-based seasonal precipitation prediction. The SEAS5 data are obtained from the Forecasted Global Gridded Precipitation values (FGGPs), which were released by the ECMWF (Copernicus Climate Change Service,

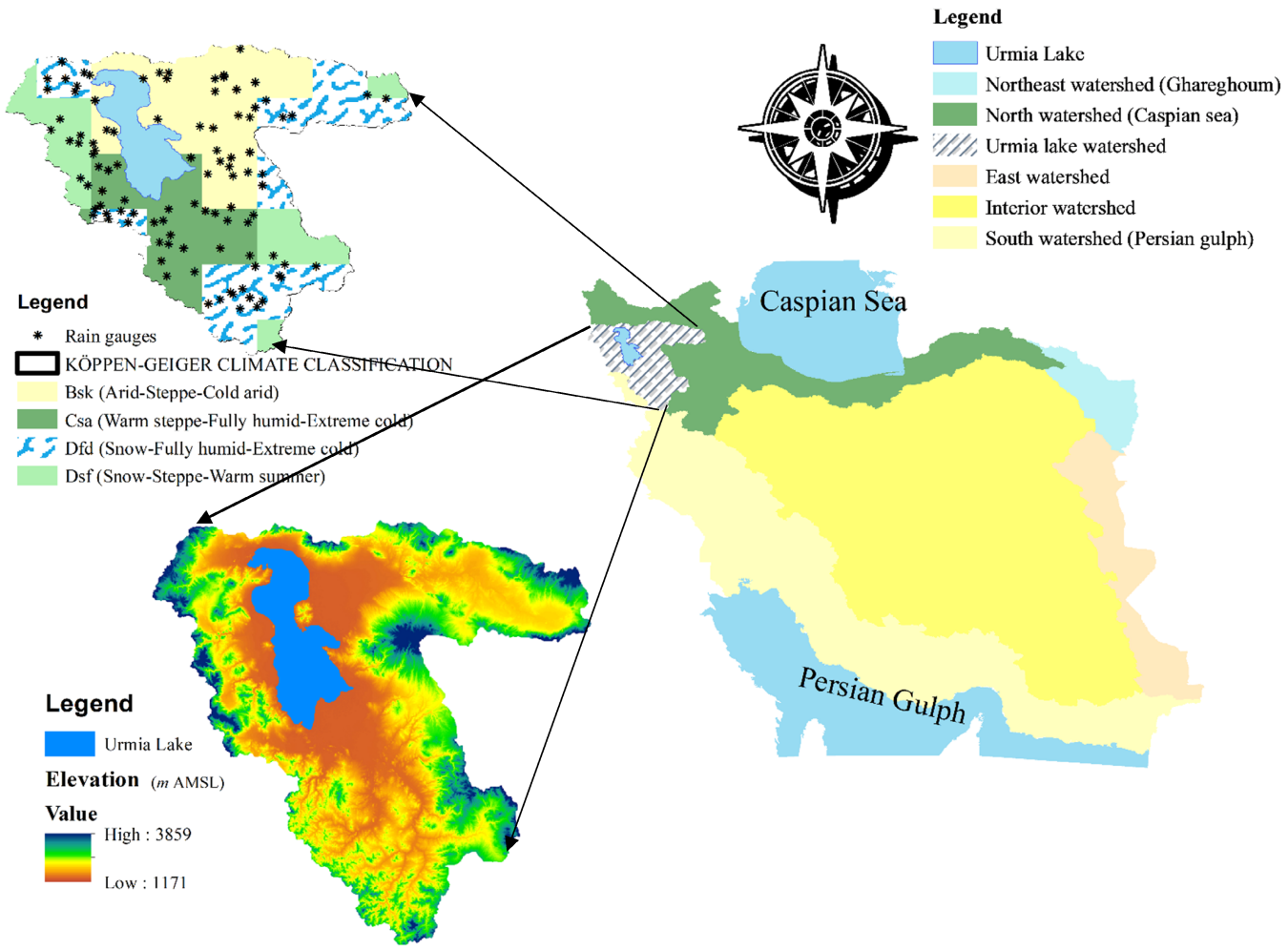

FIGURE 1 | Location of Iran's watersheds, climatological classification, rain gauge network and DEM (metre AMSL) pattern of ULW. [Colour figure can be viewed at wileyonlinelibrary.com]

Climate Data Store 2018; Johnson et al. 2019). The SEAS5 product has a spatial resolution of 1°×1°. The forecasts are issued on the first day of each month and provide predictions for climatological parameters up to 6 months in advance.

In addition to the forward-looking seasonal forecasts, SEAS5 includes a set of backward-looking seasonal forecasts, which are referred to as hindcasts or reforecasts. These utilise historical information to assess the effectiveness of the forecast framework. In this study, the SEAS5 dataset comprises 25 ensemble members, whereas the operational SEAS5 ensemble comprises 51 ensemble members (Johnson et al. 2019; Ratri, Whan, and Schmeits 2019). The SEAS5 data have been employed in the period between March 1993 and March 2017, which align with the selected temporal span of ground information of ULW. For further detailed information regarding SEAS5 and its properties, readers are encouraged to refer to the supporting documentation available on the ECMWF website (https://www.ecmwf.int/en/forecasts/documentation-and-support). The SEAS5 precipitation products were in GRIB format, containing various data records, including the time of issue, the number of ensembles, precipitation values and other relevant information. The dataset has been converted from GRIB to a text file using the GDAL Python library.

2.3 | Statistical Augmentation Methods

This section presents the three selected linear regression approaches, which are employed to augment the SEAS5 seasonal monthly precipitation forecast. These methods are employed as dynamic calibration approaches with the objective of enhancing the forecasted precipitation. This is achieved by comparing the results obtained with ground precipitation information via the assessment of their time series similarity or dissimilarity. These approaches are classified as temporal projection techniques and employ two distinct linear methods and Q-Q projection methods to scale and bias-correct the FGGPs values.

2.3.1 | Linear Regression Using OLS

Calibrated linear regression via OLS is the most prevalent regression method employed in the analysis of correlated climatological predictor(s) and predictand(s), utilising scaling and bias correction (slope and offset values) techniques. The aforementioned vectorised formula is presented in Equation (1):

$$\overline{Y}_{obs} = \overline{Y}\left(= [a \quad mL] \times \left[\overline{1} \quad \overline{Y}_{Model}\right]\right) + \overline{\varepsilon} \tag{1}$$

when $\overline{Y}_{obs}, \overline{Y}, a, mL, \overline{Y}_{Model}$ and $\overline{\varepsilon}$ are in situ values (observed values), calibrated dataset, offset (bias), slope (scale), estimated attribute with a model (such as FGGPs) and residual between projected and observed information, respectively. The symbol ($^{-}$) is corresponded to the vector value. It is important to note that the linear regression using OLS was calibrated recursively. For each new sample, all previous information was utilised to calibrate the model parameters (slope and offset), after which the new sample was evaluated using these parameters.

2.3.2 | Linear Regression Using FLS

The linear regression approach (typically calibrated using OLS) employs fixed coefficients for scale (slope) and bias (intercept) based on a specific set of predictors (independent variables) and predictands (dependent variables). This means that the slope and intercept remain fixed throughout the analysis, regardless of any changes in the data or context after calibration.

In contrast, the FLS method introduces a more adaptable calibration procedure for linear regression. This technique provides a dynamic computational framework that allows the regression parameters (including slope and intercept values) to vary during simulations. Such flexibility is particularly advantageous in scenarios where relationships between variables may change over time or in response to new data.

The FLS method was first proposed by Kalaba and Tesfatsion in a series of studies from 1988 to 1989 (Kalaba and Tesfatsion 1988a, 1988b). Their work aimed to extend the traditional linear regression framework by permitting the linear coefficients (scaling and bias values) to adjust dynamically. This adaptability enables the model to better capture the evolving nature of relationships in the data, allowing for a more accurate representation of underlying trends as new samples are introduced or conditions change. While OLS is a method that adjusts constant coefficients in a linear regression equation (Equation 1), FLS calibrates the linear regression by concurrently minimising the following equations:

$$\overline{Y}^{t}_{obs} - \left[a^{t} \quad mL^{t}\right] \times \left[\overline{1} \quad \overline{Y}^{t}_{Model}\right] \approx \overline{\varepsilon}^{t} \tag{2}$$

$$\left[a^{t+1} \quad mL^{t+1}\right] - \left[a^{t} \quad mL^{t}\right] \approx \overline{\sigma}^{t} \tag{3}$$

where t is corresponded to time or number of observed sets. One of the most important specifications of FLS is its theoretical independency to residual distribution in Equation (2), and this is one of the most important advantages of using FLS for time series simulation. To calibrate Equations (2) and (3) together, OLS has been used as below:

$$\text{Min} = \sum_{t=1}^{n} \left(\overline{\varepsilon}^{t}\right)^{2} + \mu \sum_{t=1}^{n} \left(\overline{\sigma}^{t}\right)^{2} \tag{4}$$

The first and second terms of Equation (4) ensure both the linearity and parametric stability inherent in linear regression models. In this context, the parameter μ serves as a scalar value that must be defined by the user prior to the minimisation of Equation (4). As μ approaches zero, the model grants the slopes and offsets a high degree of freedom, allowing their values to adapt readily as new samples are incorporated into the analysis. Conversely, when μ approaches positive infinity, the results produced by the FLS method converge toward those obtained from the OLS method. This behaviour illustrates the transition from a highly flexible model to a more rigid one, akin to traditional linear regression.

Kalaba and Tesfatsion (1989a) further contributed to this field by proposing two methodologies (batch and sequential approaches) to estimate the parameter sets within their proposed linear regression framework. These methods provide practical solutions for optimising the regression parameters based on the specific characteristics of the data.

In the existing literature, various time series have been simulated using the FLS framework. Notable examples include Rao (1995), who successfully employed the FLS methodology to simulate hydrological time series and assess the method's effectiveness. Additionally, other studies, such as those by Montana, Triantafyllopoulos, and Tsagaris (2009) and Alptekin et al. (2018), have also explored this framework in different contexts. For a comprehensive understanding of the FLS method, readers are encouraged to refer to the foundational works of Kalaba and Tesfatsion (1989a, 1989b). In addition, Lucchetti and Valentini (2024) has compared various linear regression methods with variable parameters with the FLS method.

2.3.3 | Linear Regression Using Q-Q Transformation

Matching the distributions of two variables is a widely used technique for scaling and reducing bias in hydroclimatological datasets (Tareghian and Rasmussen 2013; Fang et al. 2015; Hassanzadeh et al. 2019). Gudmundsson et al. (2012) introduced an experimental approach for distribution matching that involves creating a transformation through linear regression between the empirical percentiles of observed and modelled values. The proposed framework, known as Q-Q regression, resembles traditional linear regression; however, the input–output pairs $Y_{obs}, \overline{Y}_{Model}$ are matched based on their probabilities of occurrence, which are determined using a nonparametric empirical distribution method (Gudmundsson et al. 2012).

It is crucial to recognise that when a new sample is added to the dataset, the slope and offset coefficients computed in the Q-Q regression method must be recalibrated. This necessity arises because the probability of occurrence for each sample will change, consequently altering their ranks within the input–output matrix. The linear model parameters (offset and slope) have been calibrated recursively as in the linear regression process.

3 | Evaluation Metrics

In the current manuscript, we introduce and apply four types of evaluation metrics. The first type comprises similarity/

dissimilarity statistics, which assess the correspondence between two datasets (observed and computed values). The second type focuses on evaluating the similarity of the cumulative distributions of two random variables, specifically the observed values and FGGPs in this study. The third type is used to assess the forecast skill score. Finally, the fourth type of metric measures the efficiency of the uncertainty bounds derived from the ensemble members of SEAS5 (the hindcast model) in comparison to the actual precipitation values. The following section provides a brief description of these metrics.

3.1 | Similarity/Dissimilarity Metrics

In this study, two dissimilarity metrics, including bias (B) and root mean square error (RMSE), have been employed to quantify the differences between the observed and SEAS5 precipitation values. When the observed and forecasted precipitation values are closely aligned, these metrics tend to approach zero. The formulas for B and RMSE are defined by the following relationships,

$$B = \frac{\sum_{i=1}^{n} \mathrm{For}(x_i) - \sum_{i=1}^{n} \mathrm{Obs}(x_i)}{n} \tag{5}$$

$$\mathrm{RMSE} = \sqrt{\frac{\sum_{i=1}^{n} \left(\mathrm{Obs}(x_i) - \mathrm{For}(x_i)\right)^2}{n}} \tag{6}$$

where $\mathrm{Obs}(x_i)$, $\mathrm{For}(x_i)$ and n are observed and forecasted values at x_i and total number of observations, respectively. i states the month number of the estimated values. The third metric is the Kling–Gupta efficiency (KGE) statistic (Knoben, Freer, and Woods 2019). KGE is a multifactorial similarity measure that assesses the relationship between observed and forecasted precipitation values. This indicator incorporates three well-known statistics in its formulation, with a range of [−∞, 1]. A perfect KGE value of 1 indicates the highest level of compatibility between the observed and forecasted vectors. The formula for KGE is presented as follows:

$$\mathrm{KGE} = 1 - \sqrt{(R-1)^2 + \left(\frac{\sigma_{\mathrm{For}}}{\sigma_{\mathrm{Obs}}} - 1\right)^2 + \left(\frac{\mu_{\mathrm{For}}}{\mu_{\mathrm{Obs}}} - 1\right)^2} \tag{7}$$

where R is the correlation coefficient between the two vector (e.g., observed and forecasted precipitation) values, and σ and μ represent the standard deviation and mean values of the vectors.

3.2 | Comparison of Distributions

The compatibility of distributions between two random variables is a pivotal element in statistical comparison. In 1984–1985, Székely proposed the energy distance (ED) as a metric for comparing the distributions of two random vectors. This distance-based function quantifies the stochastic dissimilarities between the two vectors. Theoretically, two random vectors are considered to have the same statistical distribution if and only if their ED value is zero. An increase in ED values indicates a divergence in the statistical distributions of the two vectors (Rizzo and Székely 2016). The ED is calculated using the following equation:

$$\mathrm{ED}^2 = 2\int_{-\infty}^{+\infty} (F(x) - G(x))^2 dx \tag{8}$$

where F and G are the cumulative distribution of observed and forecasted precipitation values. For more details of its computational procedure, the readers are addressed to Rizzo and Székely (2016).

3.3 | Forecast Skill Scores

In this section, seven commonly used metrics for weather forecast systems have been selected and applied for the purpose of assessing the SEAS5 dataset over the watershed. These metrics are based on the information provided in Table 1, which presents a 2×2 contingency matrix used to calculate the efficiency metrics of the forecast (Roebber 2009). In the table, the variables H, F, M and Z represent the number of correct alarms, false alarms, missed alarms and correct negatives, respectively. The efficiency metrics calculated from this table include Probability of Detection (POD), False Alarm Ratio (FAR), Frequency Bias Index (FBI), Critical Success Index (CSI), Bias Indicator (BI) and Heidke Skill Score (HSS) (Roebber 2009; Ghajarnia, Liaghat, and Arasteh 2015; Zeng et al. 2018). Based on Table 1, it is necessary to set a threshold (ε) to differentiate between months with and without precipitation. In the present study, the threshold was set to 0.1 mm. The following section presents the mathematical forms of the aforementioned metrics:

$$\mathrm{POD} = \frac{H}{\mathrm{H} + M} \tag{9}$$

$$\mathrm{FAR} = \frac{F}{H + M} \tag{10}$$

$$\mathrm{FBI} = \frac{H + F}{H + M} \tag{11}$$

$$\mathrm{CSI} = \frac{H}{H + F + M} \tag{12}$$

$$\mathrm{BI} = \frac{POD}{1 - FAR} \tag{13}$$

$$\mathrm{HSS} = \frac{2 \times (HZ - MF)}{(H + M)(M + Z) + (H + F)(F + Z)} \tag{14}$$

TABLE 1 | Contingency matrix based on dichotomous forecasts.

	Observed	
Forecast	***Observed*** $> \epsilon$	***Observed*** $< \epsilon$
Forecast $> \epsilon$	H	F
Forecast $< \epsilon$	M	Z

Note: H, hits (event forecast to occur, but did); F, false alarm (event forecast to occur, but did not); M, misses (event forecast not to occur, but did); Z, correct negative (event forecast not to occur, but did not).

The POD metric describes the fraction of correct forecasts, with a range of [0, 1] and the optimal value equal to 1. The FAR indicates the fraction of false alarms in the predictions, with a range of [0, 1] and the optimal value equal to 0. The CSI represents the fraction of positive forecasts relative to the observed events, with a range of [0, 1] and the best value equal to 1. The BI has a range of [0, ∞], with the best value equal to 1. The HSS measures the accuracy of the forecast in comparison to that of random chance. The range is [−1, 1], with the optimal value equal to 1 and the no-skill value equal to 0. Readers interested in a more detailed examination of these forecast efficiency metrics and contingency statistics are encouraged to visit the webpage (https://www.cawcr.gov.au/projects/verification/#BSS).

3.4 | Uncertainty Metrics

This subsection presents the metrics for evaluating the uncertainty-based assessment of variables. The initial metric is the Plevel, which quantifies the proportion of observed precipitation values that fall within the uncertainty bounds of the forecasted precipitation (Xiong et al. 2009; Ahmadi, Nasseri, and Solomatine 2019; Ahmadi and Nasseri 2020).

$$\text{Plevel} = 100 \times \left(\frac{1}{N} \sum_{i=1}^{N} \delta_i \right), \delta_i = \begin{cases} 1 \; if \; y_i^- \le Y_i^{obs} \le y_i^+ \\ 0 \qquad\quad otherwise \end{cases} \tag{15}$$

where y_i^+ and y_i^- are the upper and lower uncertainty bounds of a CL, respectively, inferred from 25 ensemble members of SEAS5 model in the ith time step of the time series, and Y_i^{obs} is the observed precipitation values in that time step. N is the total number of observed values.

The Normalised Uncertainty Efficiency (NUE) is a metric employed for the assessment of uncertainty bounds (Nasseri et al. 2013; Nasseri, Ansari, and Zahraie 2014). The NUE incorporates two metrics including the Plevel and the Average Relative Interval Length (ARIL). The formulations of the statistics are as follows,

$$\text{ARIL} = \frac{1}{N} \times \left(\sum_{i=1}^{N} \frac{y_i^+ - y_i^-}{Y_i^{obs}} \right) \tag{16}$$

$$\text{NUE} = \frac{\text{Plevel}}{w \times \text{ARIL}} \tag{17}$$

where N, y_i^+, y_i^- and Y_i^{obs} are the same as the prior equation, and w is the scale factor of Plevel versus $ARIL$, and it is considered to be equal to 1 in this study.

4 | Modelling Procedure

This article aimed to present a novel dynamic calibration method with the objective of enhancing the precision of the SEAS5 forecast model in predicting seasonal monthly precipitation. In this section, the authors present the stepwise modelling procedure, which is depicted in Figure 2. This figure illustrates the flowchart of the modelling process.

- Step 1: Statistical combination. In the first step, the average, median and three confidence intervals (CIs) values have been extracted for each month throughout the simulation period. Given that SEAS5 utilises 25 ensemble members, a total of 25 values are available for each month and rain gauge station. Experimental cumulative distributions have been employed to calculate three CIs including [2.5%–97.5%], [5%–95%] and [25%–75%].
- Step 2: Statistical evaluation. In the second step, statistical metrics will be calculated between the observed and forecasted precipitation values obtained in the preceding step. These metrics encompass four types of evaluation

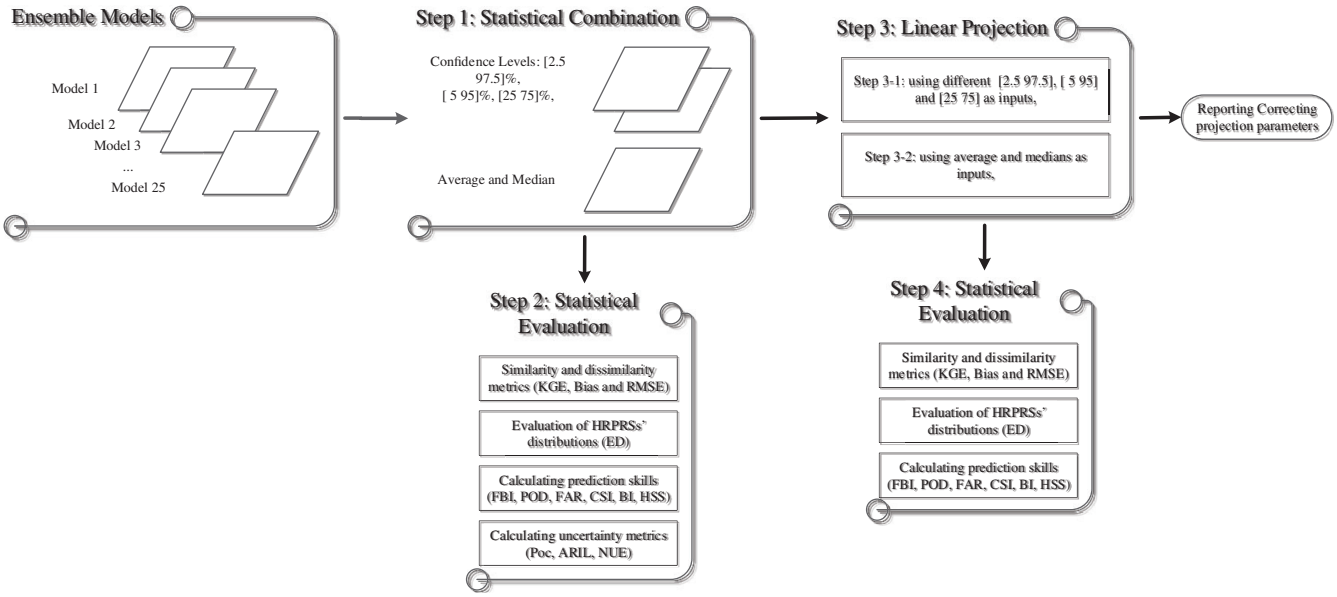

FIGURE 2 | Schematic flowchart of modelling procedure.

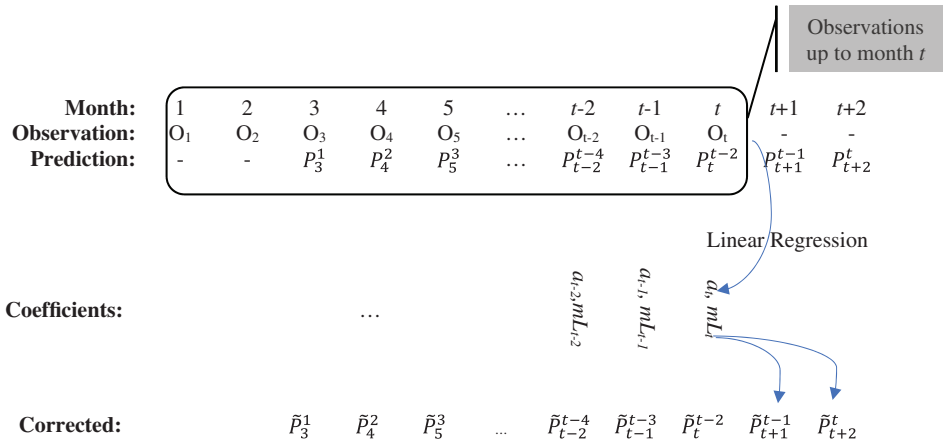

FIGURE 3 | Schematic procedure of the proposed augmentation approach. [Colour figure can be viewed at wileyonlinelibrary.com]

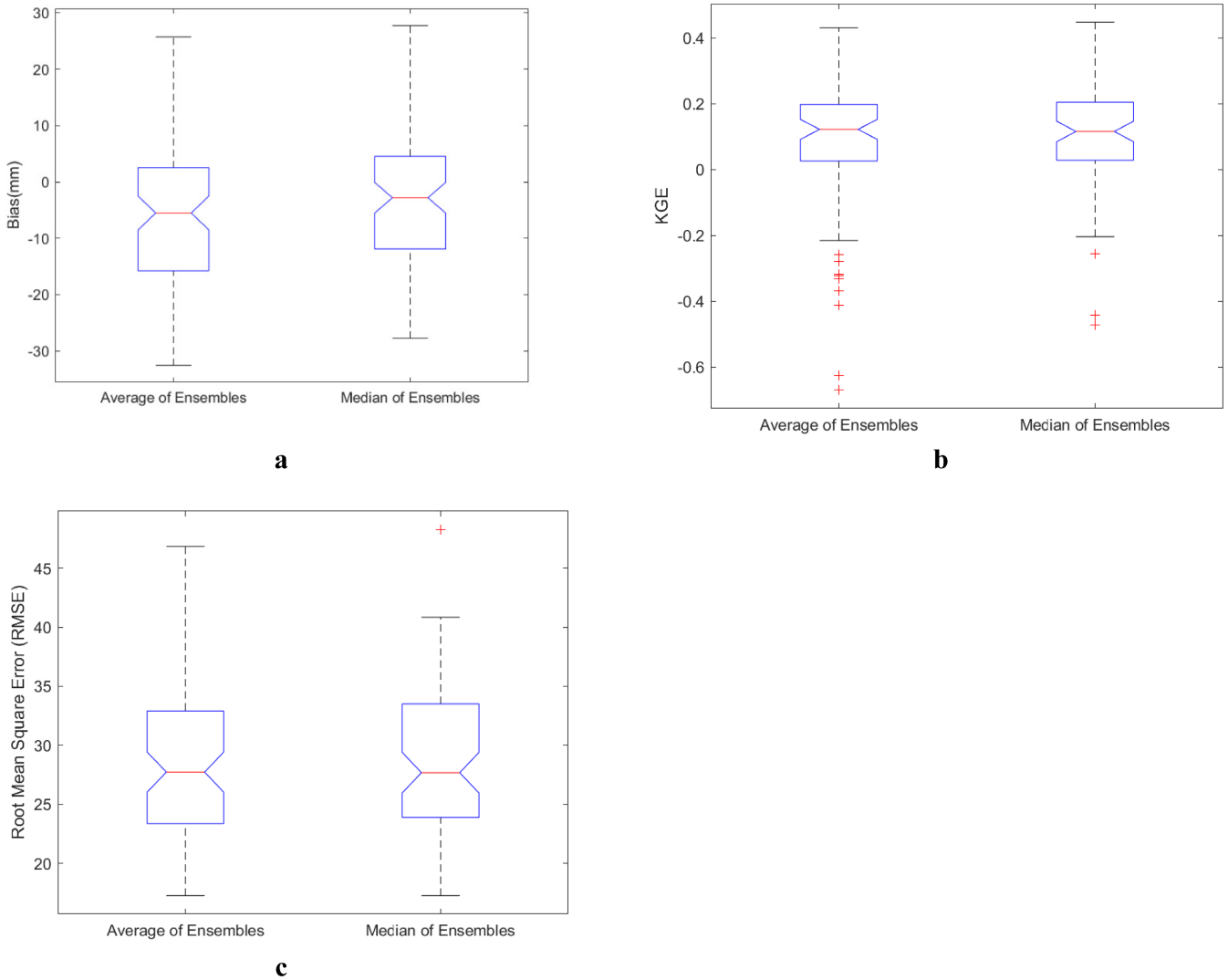

FIGURE 4 | Distributions of three statistical metrics based on 87 stations (a) bias (mm), (b) KGE and (c) RMSE (mm) between SEAS5 and observed precipitation values. [Colour figure can be viewed at wileyonlinelibrary.com]

indicators including similarity/dissimilarity, distribution compatibility, forecast skill scores and uncertainty metrics (see Section 3 for further details). The initial three types of statistics facilitate the selection of either the median or the mean of the ensemble values for the subsequent step (Step 3). Conversely, the uncertainty metrics facilitate the inference of the uncertainty behaviour exhibited by the SEAS5 ensemble members.

- Step 3: Linear projection. In the current stage, the dynamic evaluation of linear relationships calibrated using OLS, FLS and Q-Q regression methods will be conducted for each station. The calibration procedure for these methods is illustrated in Figure 3. The warmup time has been set to 12 (sample) month to initialised the regressions' parameters. In the figure, O_t represents observed monthly precipitation at the station of interest and P_t^{t-2} addressed precipitation forecasted at month $t-2$ to calculate for month t.

  By means of the both observed and forecasted precipitation values available up to month t, their offsets and slopes (a_t, mL_t) could be calculated using the above mentioned regression methods. Using the variables (a_t, mL_t), updated precipitations for month $t+1$ and $t+2$ would be estimated, recursively. Increasing time indicator to $t+1$, the both slopes and offsets must be calculated because of new added observed samples to the information pool. However, the updating sequence may be set more than one-by-one month, but in the current research it set to 1 month.

- Step 4: Statistical evaluation. Following the completion of the augmentation procedure, three selected statistical metrics (KGE, RMSE and B values) have been calculated for each station individually.

- Step 5: Selection of the most effective method. The last step, the most effective linear correction procedure based on the reported statistics (in Step 4) will be selected.

5 | Modelling Results

In line with the modelling approach depicted in Figure 2, the initial stage involved calculating three CL values of the ensemble members for each station. Following this, statistical metrics were computed to assess the original three-month ahead hindcast of the SEAS5 datasets. Applying linear regression techniques, their statistical effectiveness was evaluated to enable a comparison with the observed precipitation values. The subsequent sections will present the results of the analyses. It should be noted that the number of observed values is 22,419 (the total number of missed values in the sample is 2637).

5.1 | Statistical and Probabilistic Assessment of SEAS5 Forecast

5.1.1 | Similarities and Dissimilarities

Figure 4 illustrates the distributions of similarity and dissimilarity statistics between the SEAS5 hindcast and observed values for the median and average of the ensemble members using B, KGE and RMSE metrics as detailed in Section 3.1. Notably, there is no significant variation in the metrics between the average and median ensemble values compared to the observations. However, it is important to highlight that the median results tend to align more closely with the recorded precipitation values than the average results.

Figure 4a shows the distributions of bias values across the stations. It is clear that the maximum value of the ensemble median is lower than that of the ensemble average. Figure 4b presents the distributions of KGE values, where the median values not only surpass the average values but also display a narrower range compared to the ensemble average. Figure 4c illustrates the distributions of RMSE, indicating that the ensemble median values demonstrate greater efficiency than the ensemble average values.

5.1.2 | Evaluation of Distributions

Figure 5 presents two boxplots, each depicting the ED values for the average and median of the SEAS5 ensemble members. As illustrated in the figure, the median values of the seasonal forecast from SEAS5 display distributions that are more compatible with the observed values than the average values.

5.1.3 | Forecast Skill Score

Figure 6 illustrates the distributions of six forecast skill scores (including FBI, FAR, CSI, POD, BI and HSS) for both the median and average values. It is evident that the distributions of the FBI, FAR, CSI, POD and BI statistics are highly similar between the median and average ensembles. However, the distribution of the HSS statistic for the median ensembles shows a greater range than that of the average ensembles. The POD values in the median and average scenarios are close to 1 (its perfect value), which means that the SEAS5 hindcast has a near perfect performance in predicting monthly precipitation. While the perfect value for HSS is 1, it appears that the median ensembles outperform the averages, as indicated by the broader range of values.

5.1.4 | Uncertainty Assessment

To evaluate the probabilistic efficacy of the 25 ensembles included in SEAS5, three CIs are presented in Figure 7. The CIs shown in the figure consist of the following ranges: [2.5%,

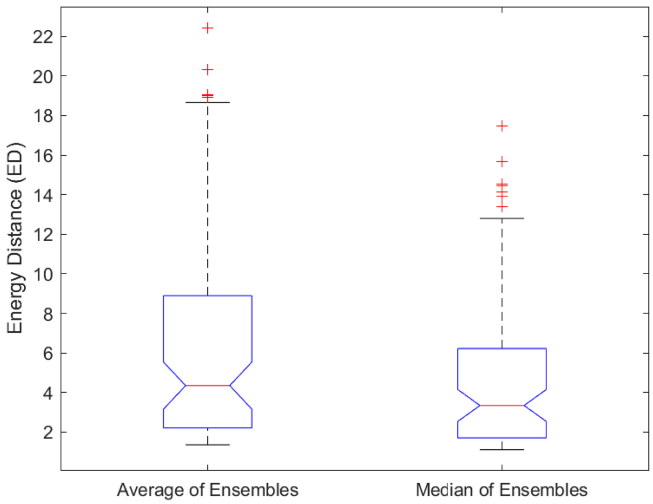

FIGURE 5 | Distributions of the energy distance (ED) between SEAS5 and observed precipitation values based on 87 stations. [Colour figure can be viewed at wileyonlinelibrary.com]

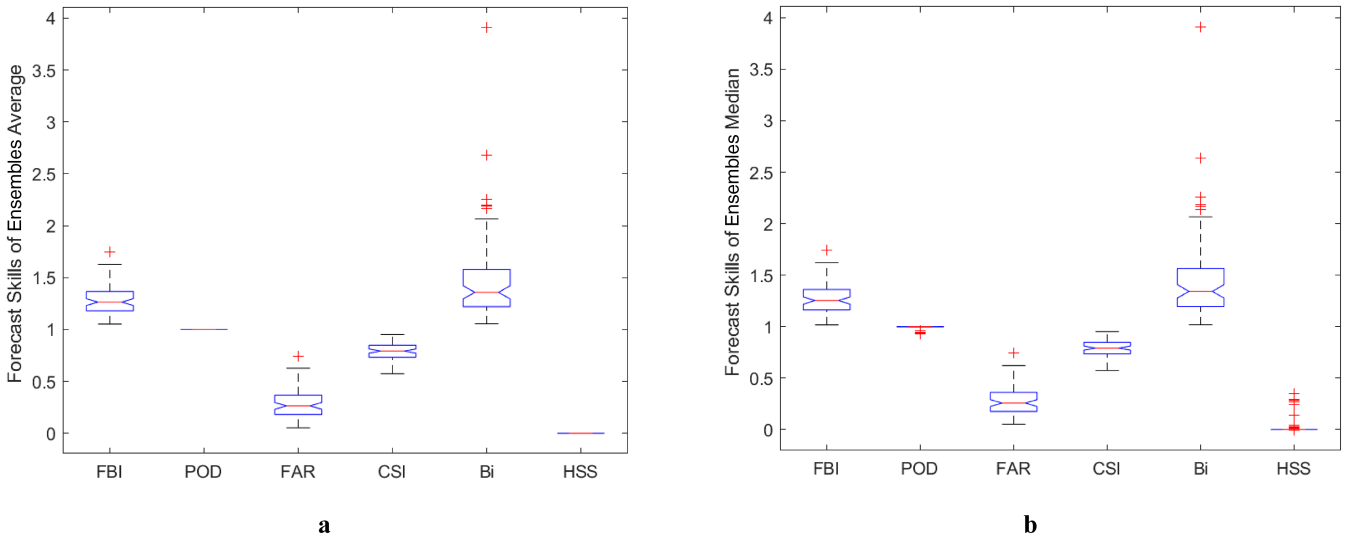

FIGURE 6 | Distributions of the forecast skill scores based on 87 stations (FBI, POD, FAR, CSI, Bi, HSS) for (a) average and (b) median of the ensemble members. [Colour figure can be viewed at wileyonlinelibrary.com]

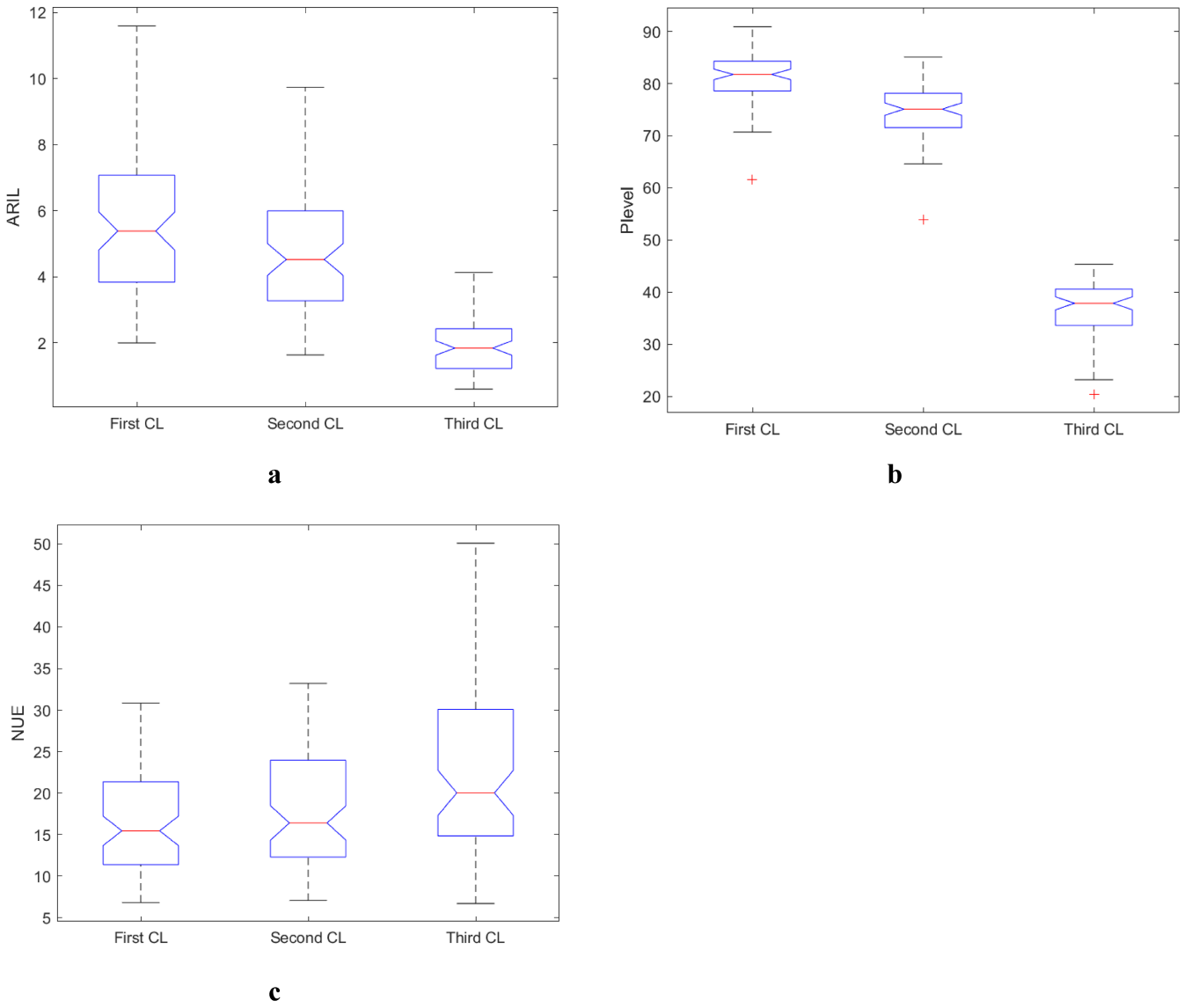

FIGURE 7 | Distributions of the three uncertainty metrics based on 87 stations (a): ARIL, (b) Plevel (%) and (c) NUE between the observed and various CIs of SEAS5 ensembles. [Colour figure can be viewed at wileyonlinelibrary.com]

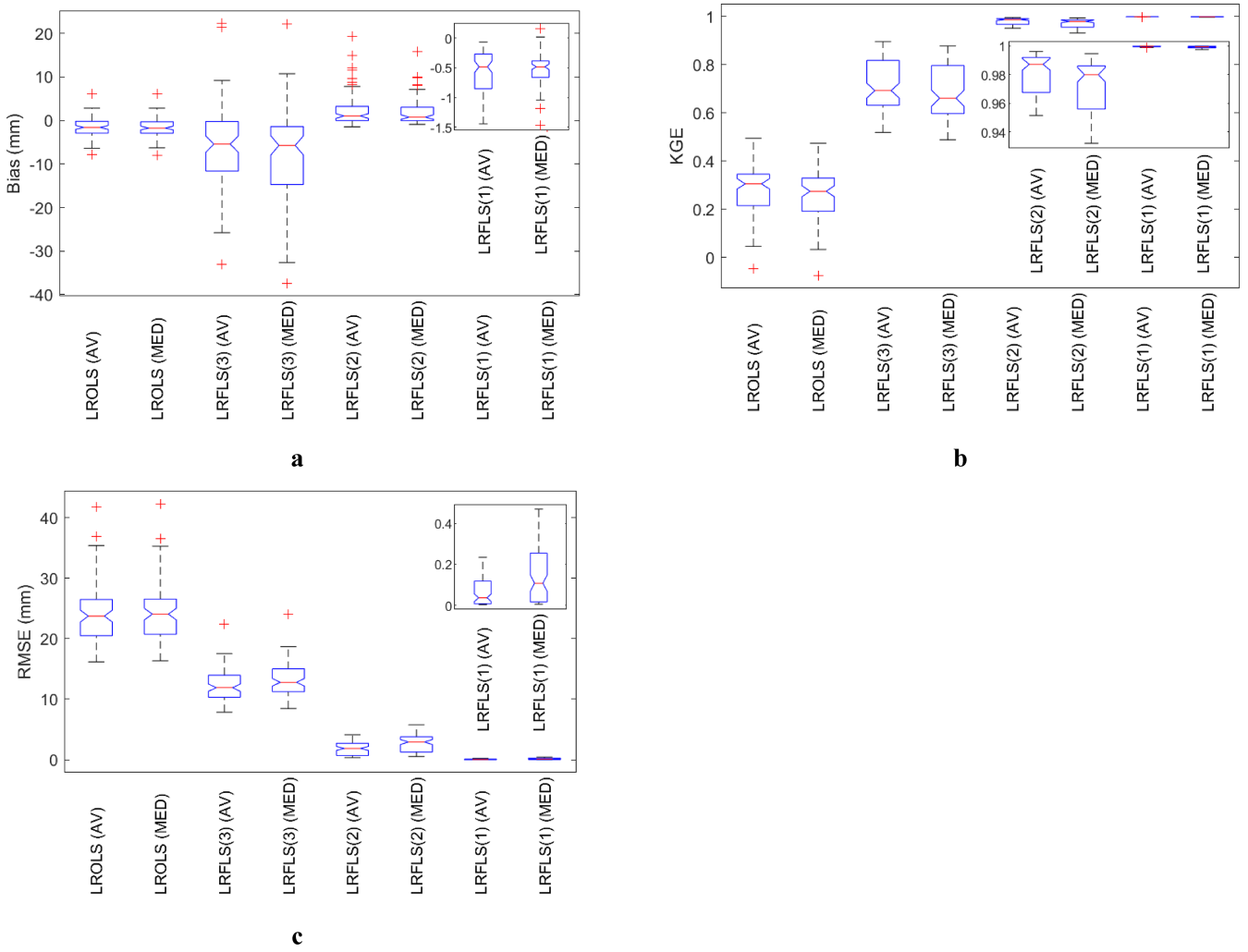

FIGURE 8 | Distributions of the statistical metrics based on 87 stations (a) bias (mm), (b), KGE and (c) RMSE (mm) between the observed and calibrated SEAS5 precipitation values using LROLS and different μ (1000, 10, 0.1) of LRFLS methods, respectively (model 1, model 2 and model 3 correspond to 0.1, 1 and 1000). [Colour figure can be viewed at wileyonlinelibrary.com]

97.5%], [5%, 95%] and [25%, 75%], representing the first, second and third CLs, respectively. A reduction in the width of the evaluated CIs corresponds to a decrease in the bounds and values of the ARIL, as illustrated in Figure 7a.

Furthermore, lowering the CL from 95% to 50% results in a decline in the median values of the Plevels' distributions, which range from 81% to 37%. This reduction in CL also enhances the efficiency metrics, with the median values of the NUE distributions increasing from 15.45 to 20.36. These metrics are valuable when analysing different CIs of the original SEAS5 ensemble members, as they provide insights into the efficiency of probabilistic forecasts.

5.2 | Statistical Evaluation of the Linear Methods

This section presents the results of the calibration and validation of the standard linear regression methods with dynamics and fixed linear parameters (FLS and OLS) for enhancing SEAS5 forecasts. As previously stated (Section 2.3.2), the variability of slope and offset values in the linear regression calibrated using the FLS approach is significantly influenced by the calibration scalar parameter (μ).

Figure 8 illustrates the distributions of the three statistical metrics (B, KGE and RMSE) for varying scaler parameters (0.1, 10 and 1000) alongside the linear regression calibrated via Ordinary Least Squares (LROLS) method. The results may be regarded as a sensitivity analysis of the scaler parameters (μ) on the calibration procedure. The boxplots of the models with median inputs are observed to be wider than those with average inputs. An increase in the scaler parameter value (from Models 1 to 3) results in a notable dissimilarity between the calibrated SEAS5 and observed precipitation values. As previously stated in Section 2.3.2, the scaler parameter determines the variability of the slope and offset values of the dynamic model. Furthermore, within the FLS framework, there is no established methodology for identifying the optimal parameter value. The next step will entail the augmentation of the SEAS5 with the calibrated offsets and slopes from the current stage for the scaler parameter 0.1, with the objective of demonstrating the effect of the most allowed variability of slope and offset values.

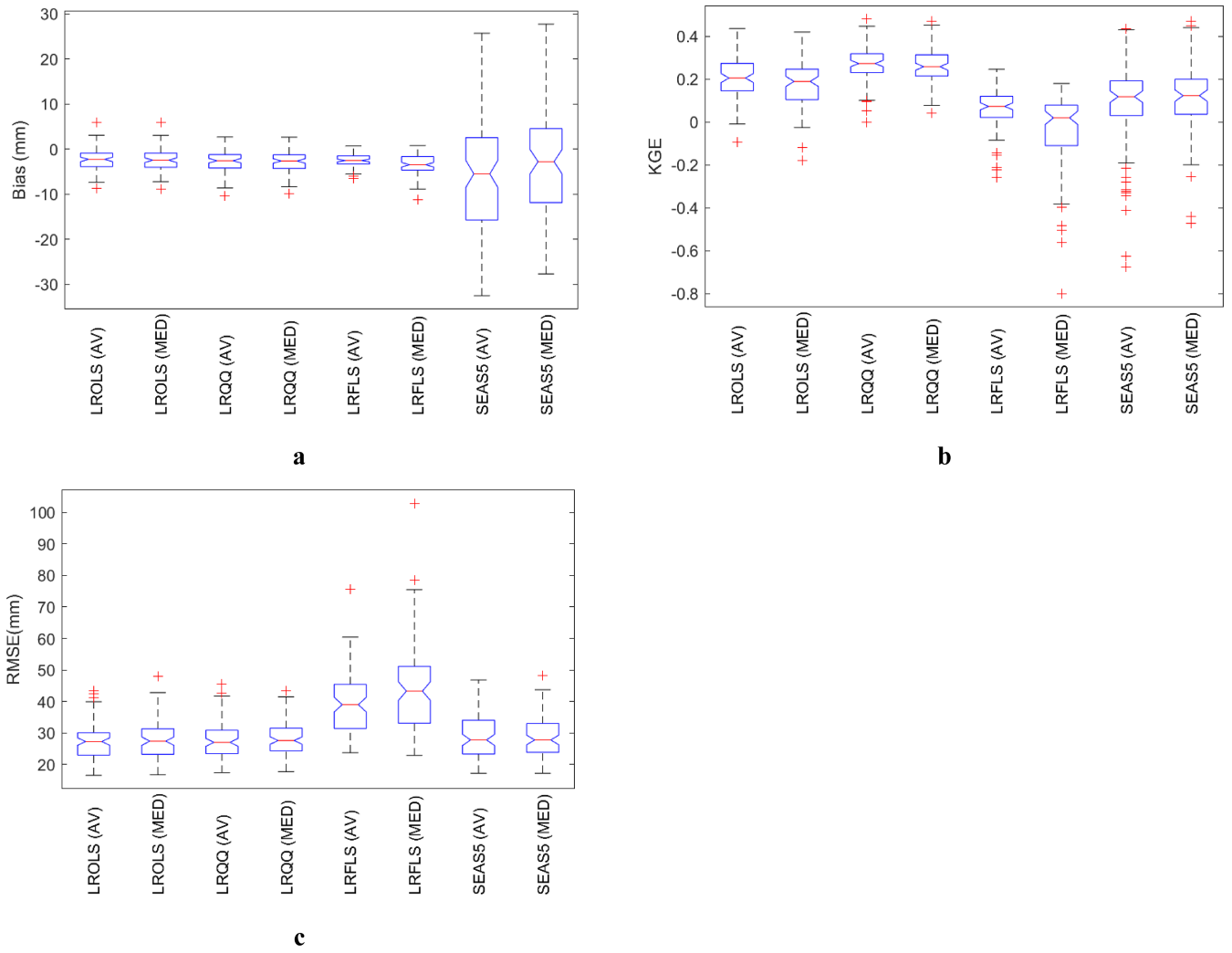

FIGURE 9 | Distributions of the statistical metrics based on 87 stations (a) bias (mm), (b) KGE and (c) RMSE (mm) between the observed and original and augmented SEAS5 at the first-coming monthly hindcasted values ($\mu = 0.1$ for LRFLS). [Colour figure can be viewed at wileyonlinelibrary.com]

5.3 | Augmenting SEAS5 Products via Linear Regression Methods

As illustrated in Figure 3, the slope and offset values estimated at each time step, incorporating all preceding observed precipitation data, are employed to enhance the SEAS5 hindcasted values. The updated slope (mL) and offset (a) values obtained from the selected linear regression methods (LROLS, linear regression with quantile-quantile (LRQQ) and linear regression with dynamic parameters using FLSs (LRFLS)) are employed in the recalculation of the linear regression parameters, a and mL, at a specific time (t). These parameters are then applied to augment the forthcoming hindcasted precipitation issued up to time t ($\widetilde{P}_{t+1}^{t-1}$ and $\widetilde{P}_{t+2}^{t}$).

In Figure 9, the distributions of the selected statistics (B, KGE and RMSE) are presented for the augmented average and median ensemble members using the regression methods including LROLS, LRQQ and LRFLS methods. These metrics assess the agreement between the first-coming hindcasted precipitation ($\widetilde{P}_{t+1}^{t-1}$) and observed values across the watershed's stations. Furthermore, the statistical assessment encompasses the unprocessed SEAS5 models for comparison. The LRFLS method employed a scaler parameter (μ) with a value of 0.1.

As illustrated in Figure 9a, all augmentation linear methods demonstrate a notable reduction in bias values in comparison to the bias observed in the raw SEAS5 values. The LRFLS method demonstrates the narrowest distribution of bias values among the methods under consideration. Figure 9b presents the KGE values between observed and augmented precipitation, as well as the SEAS5 methods. The LRQQ method demonstrates the highest KGE values and the narrowest distributions. The precipitation hindcast for the first month ahead, generated using the LROLS method and the original SEAS5 dataset, ranks second and third, respectively. The results of the LRFLS method are the least favourable. With regard to the RMSE statistics illustrated in Figure 9c, it can be observed that LRFLS (with a scaler value of 0.1) exhibits the highest RMSE values and the widest distributions for both the average and median values of the SEAS5 ensemble members. The distributions of LRQQ and LROLS are comparable, with SEAS5 results ranking third before LRFLS. It

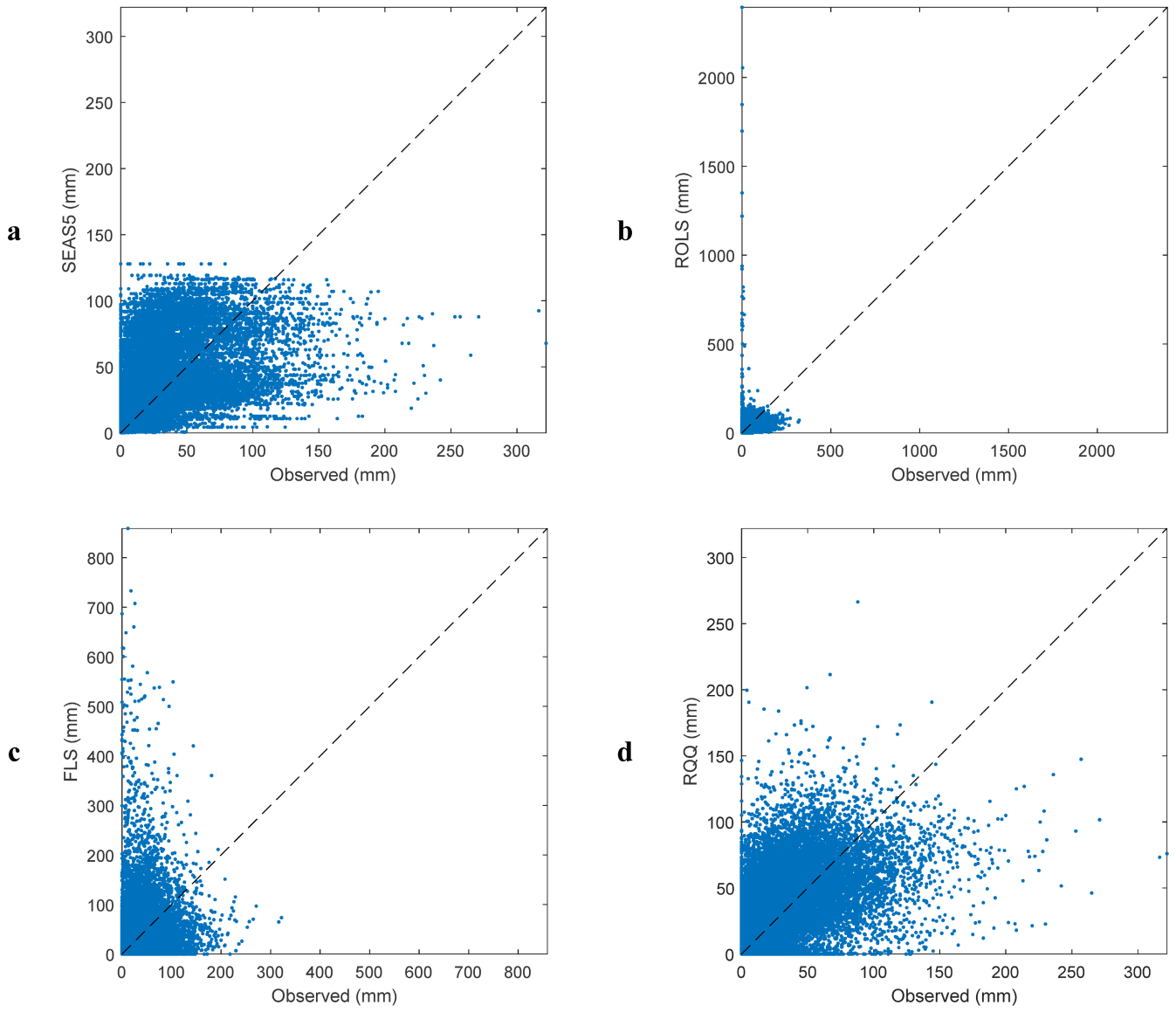

FIGURE 10 | Scatter plots of the raw and augmented SEAS5 at the first-coming monthly hindcasted values (a) raw SEAS5 and augmented with (b), LROLS, (c) LRFLS ($\mu = 0.1$) and (d) LRQQ methods for the first-coming lead time (average ensembles) based on 22,419 monthly samples. [Colour figure can be viewed at wileyonlinelibrary.com]

can thus be concluded that LRQQ and LROLS are the optimal models, respectively.

To evaluate the efficacy of the selected augmentation techniques, four scatter plots are presented in Figure 10. As the distribution around the first bisector line increases from the raw SEAS5 data to the augmented data obtained via the LRQQ method, the number of outlier augmented values decreases from the LROLS data to the LRQQ data. In light of the findings, it can be concluded that LRFLS (with a scaler value of 0.1) outperforms LROLS (or LRFLS with a scaler value of ∞).

In Figure 11, the distributions of the selected statistics related to the second-coming monthly hindcasted values are presented. As shown in Figure 11a, all calculated bias values for the augmented methods and their distributions are smaller than those of SEAS5 and its distributions. The boxplots of the KGE values in Figure 11b indicate that LRFLS exhibits the worst performance, while LRQQ performs the best. Excluding outliers, the KGE values for the LROLS method are comparable to the raw SEAS5 values, and LRQQ outperforms the raw SEAS5 values. Similarly, the RMSE values also reveal that LRFLS has the poorest performance, with the distribution of RMSE values for LRQQ being lower than that of the original hindcasted values.

Figure 12 illustrates the scatter plot of the augmented SEAS5 precipitation values (for the second-coming month) versus the observed precipitation. In Figure 12c, the LRFLS method shows the lowest performance due to its outlying values, while the LRQQ method demonstrates the best performance, as indicated by the distribution of scatter plot points around the bisector line.

Considering the selected statistical metrics, LRQQ outperforms LROLS in augmenting SEAS5-hindcasted values. Its statistical

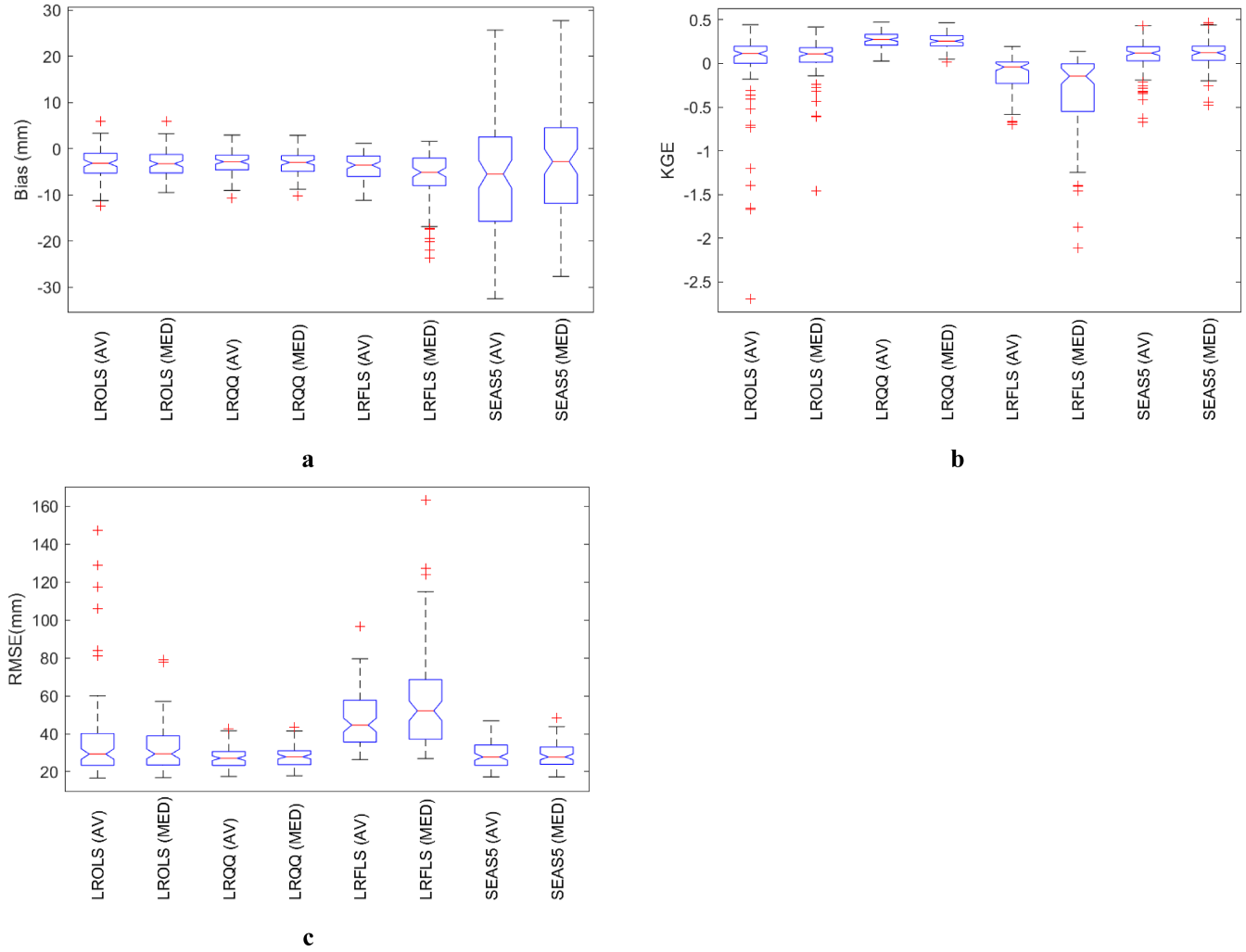

FIGURE 11 | Distributions of the statistical metrics based on 87 stations (a) bias (mm), (b) KGE and (c) RMSE (mm) between the observed, raw and augmented SEAS5 values at the second-coming monthly hindcasted values ($\mu = 0.1$ for LRFLS). [Colour figure can be viewed at wileyonlinelibrary.com]

efficiency is notable for enhancing the first-coming monthly precipitation and also improves the statistical efficiency of the second-coming monthly hindcasted precipitation. The FLS framework demonstrates significant competency in linear regression. It appears that LRFLS, with a small scaling parameter, has too much flexibility in capturing the mainstream behaviour of linear parameters during the augmentation process, potentially leading to overfitting.

To evaluate the performance of the proposed augmentation methods, eight time series, including observed values versus original SEAS5 (average and median of the ensembles) and their augmented hindcasted values, are depicted in Figure 13. These figures pertain to the Sezhab station (ID number: 31-001), located in the eastern part of the watershed. The average ensemble of SEAS5 exhibits more bias than its median counterparts. Considering the RMSE values (between raw or projected forecast time series and observations), all augmentation methods improve the behaviour of the hindcast precipitation time series. The LROLS method (considering both median and average signals) outperforms the LRFLS (with $\mu = 0.1$) and LRQQ methods, with LRQQ ranking second.

5.4 | Spatial Patterns of Augmentation Efficiencies

Figure 14 shows the spatial distribution of the bias values for the second hindcast for the median and mean of the ensemble members. From the figures (SEAS5 column), it can be seen that stations located in the central part of the ULW (near the lake) have the highest bias values, while those located in the west and southwest of the ULW have the lowest bias values. Most of the stations scattered in the western part of the ULW show underestimation and stations with overestimation conditions are located in the central and eastern parts of the ULW. In addition, it appears that stations at lower altitudes tend to have higher and overestimated values, while SEAS5 hindcast values in the highlands are lower and underestimated.

In contrast, the extended patterns shown in the figures show that stations with positive bias values are mainly concentrated to the north and northeast of the ULW. This pattern is generally observed when looking at the median of the ensembles across the catchment. Based on the figure, the southern and northern stations have negative and positive bias values using the LRFLS projection method with average and median datasets. Also, the

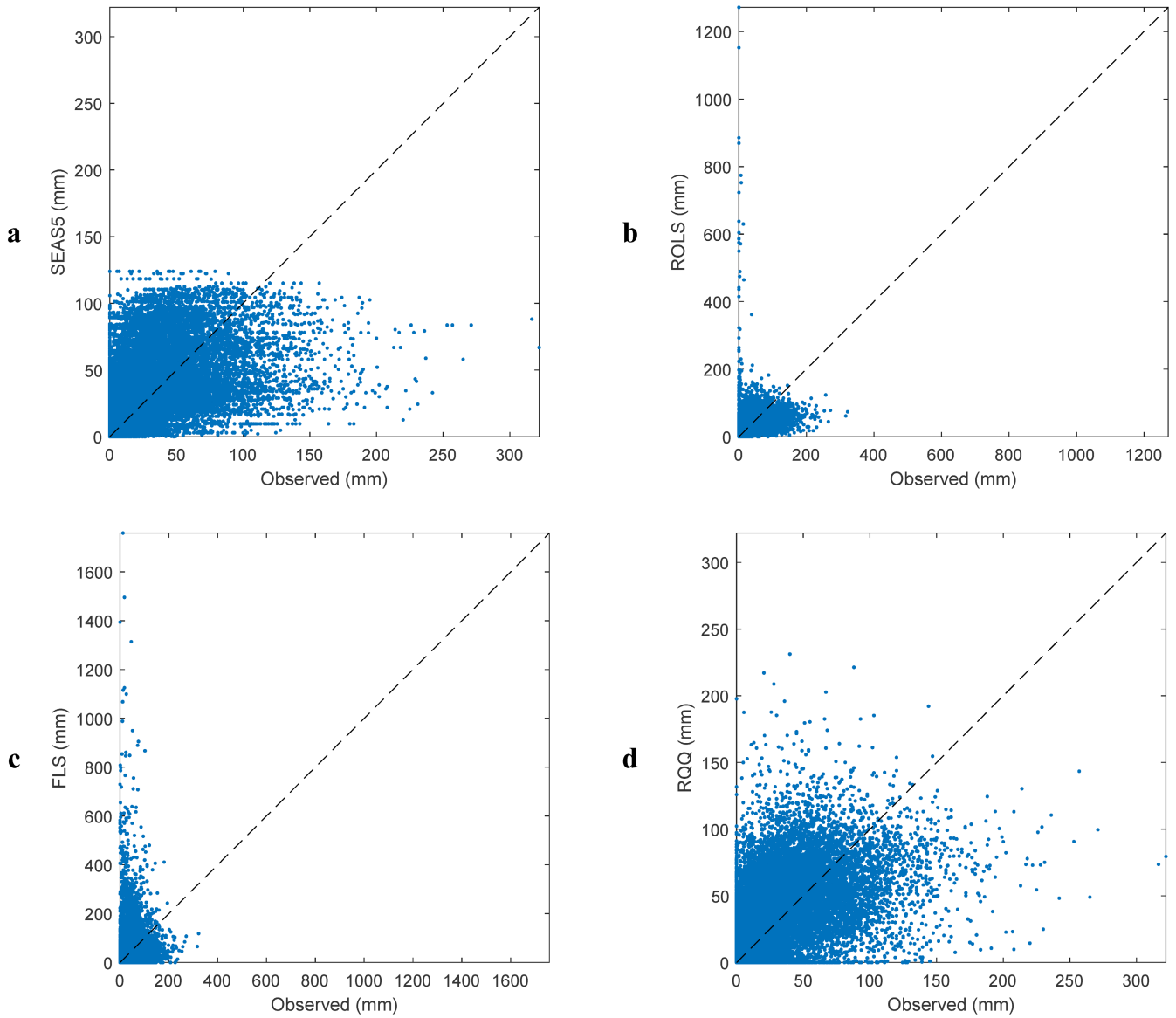

FIGURE 12 | Scatter plots of the original and augmented SEAS5 at the second-coming monthly hindcasted values (a) original SEAS5 and projected with (b), LROLS, (c) LRFLS ($\mu = 0.1$) and (d) LRQQ methods for the second-coming lead time (average ensembles) based on 22,419 monthly samples. [Colour figure can be viewed at wileyonlinelibrary.com]

bias values of the LRQQ projection method are semirandom patterns over the watershed.

6 | Conclusion

Researchers have emphasised the importance of calibrating bias values to effectively utilise SEAS5 hindcasted precipitation in hydroclimatological research and studies. In this article, the authors conducted a linear projection of the hindcast model using three regression methods including LROLS, LRFLS and LRQQ methods. These methods incorporated dynamic updates to obtain appropriate slope and offset values at each time step.

The ULW has been selected as a case study to assess the performance of the proposed linear projection methods. This is an endorheic watershed with semicomplex orographic conditions. The results presented in the previous sections indicate that these augmentation procedures, particularly LRQQ, significantly improve the statistical efficiency of the SEAS5 hindcast precipitation for both the first- and second-coming months.

However, while increasing the degrees of freedom for variability in the linear regression model parameters through the use of FLS improved certain statistical metrics, it did not necessarily translate into improved performance of the regression models in the augmentation mode. The findings suggest that overfitting may occur when the LRFLS method is allowed to adjust its model parameters at each time step. Interestingly, LROLS, which is a specific version of LRFLS, outperformed LRFLS with a lower scaling parameter in improving the accuracy of the hindcasted precipitation.

Among the various methods evaluated, LRQQ demonstrated the highest efficiency and played a considerable role in augmenting hindcasted precipitation. By employing the LRQQ method, the

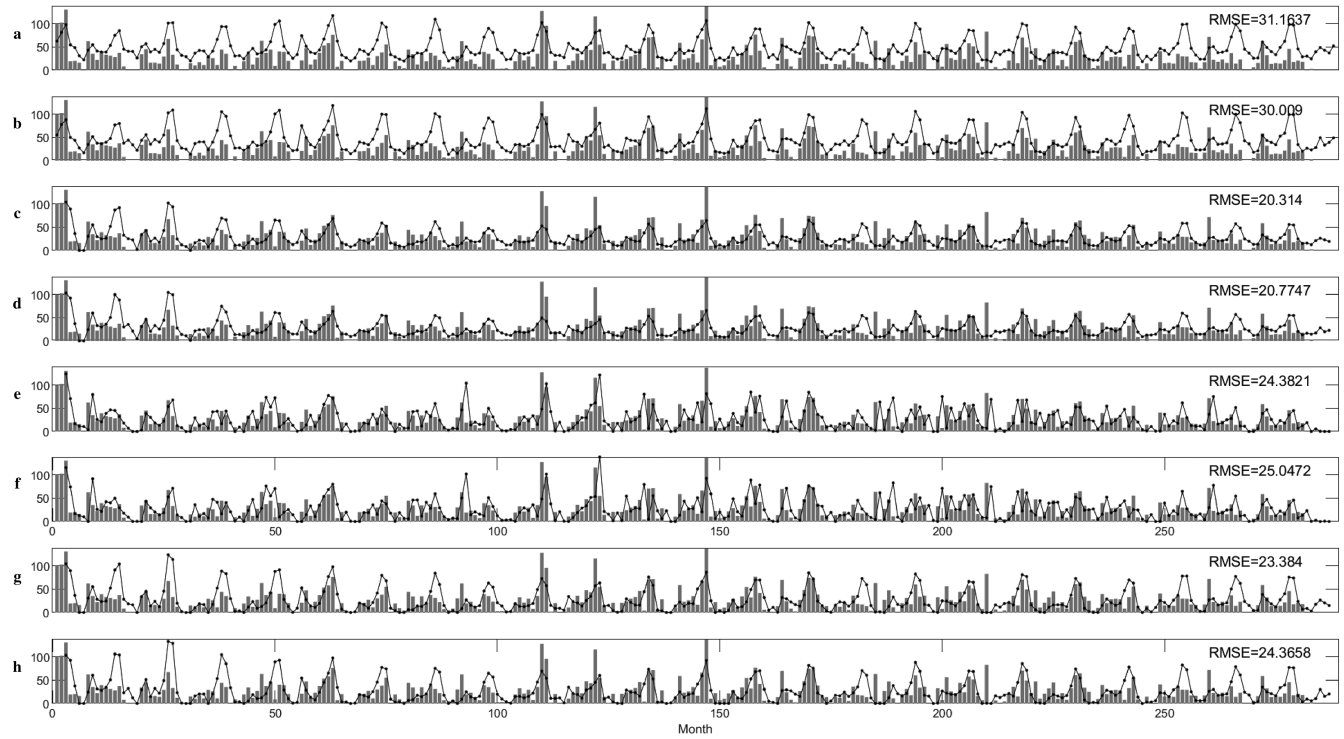

FIGURE 13 | Times series of the observed monthly precipitation (grey bars) at Sehzab station (with ID number: 31-001) versus first-coming monthly hindcasted SEAS5 values (a) average ensemble, (b) median ensemble, projected with (c) LROLS (average), (d) LROLS (median), (e) LRFLS ($\mu = 0.1$) (average), (f) LRFLS ($\mu = 0.1$) (median), (g) LRQQ (average) and (h) LRQQ (median) with their RMSE values.

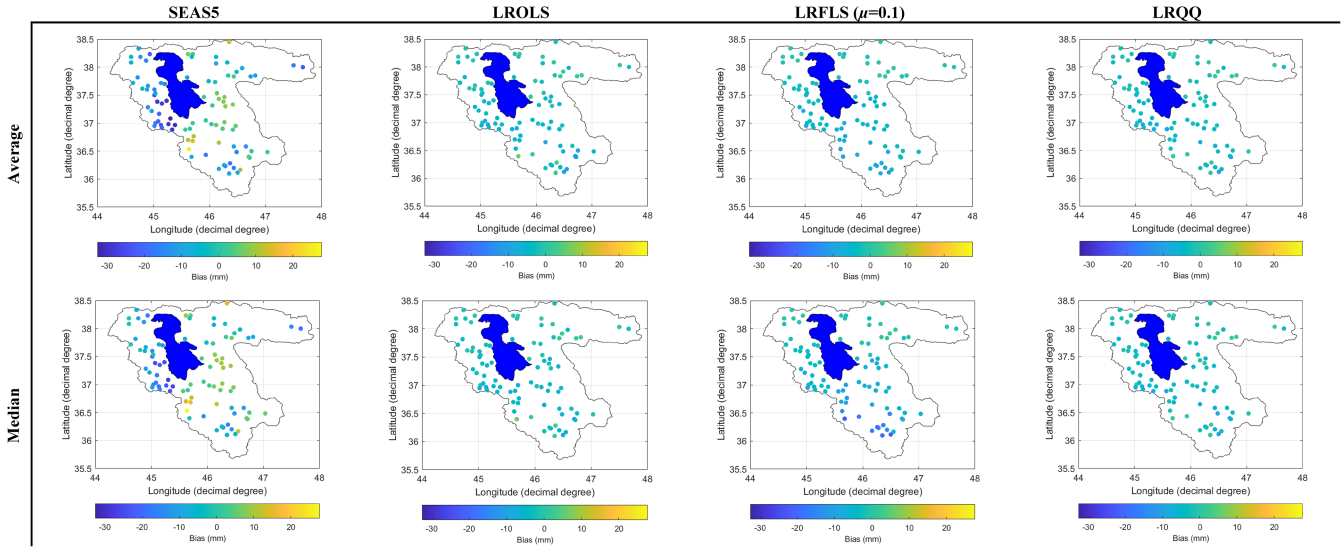

FIGURE 14 | Spatial distributions of bias values (mm) between monthly observations and raw SEAS5, augmented using LROLS, LRFLS ($\mu = 0.1$) and LRQQ methods for the second-coming hindcasted (average and median ensembles). [Colour figure can be viewed at wileyonlinelibrary.com]

proposed augmentation procedure effectively correlated the statistical distributions of ensemble SEAS5 precipitation with observed values, leading to improved forecast accuracy.

Additionally, both the average and median of the ensemble members were assessed in their raw and augmented states. While the statistical efficiencies, forecast skill scores and uncertainty metrics varied between the average and median ensembles, no distinct patterns emerged from the results. Furthermore, the effects of LROLS, LRQQ and LRFLS on the median and average of the ensemble members differed without revealing any notable or unique trends.

The augmentation approaches discussed are primarily based on linear methods. The proposed methodology is case-independent and applicable in any study area, and the performance of linear projection methods needs to be evaluated. Future research should explore distribution-free linear methods, such as generalised linear models (GLM) and support vector machines (SVMs), as well as nonlinear methods such as generalised regression neural

networks (GRNNs). Given the recursive and dynamic nature of these augmentation techniques, the adoption of parametric assimilation methods, such as the Kalman filter, is also recommended to enhance model performance and adaptability.

Author Contributions

Mohsen Nasseri: conceptualization, methodology, software, writing – original draft, writing – review and editing, visualization, formal analysis, investigation, resources, funding acquisition. **Gerrit Schoups:** conceptualization, methodology, writing – review and editing, project administration, supervision, validation. **Mercedeh Taheri:** data curation, software, investigation.

Acknowledgements

The first author was supported by Visitor's Travel Grant no. 040.11.731 from the Dutch Research Council (NWO).

Conflicts of Interest

The authors declare no conflicts of interest.

Data Availability Statement

The data that support the findings of this study are available from the corresponding author upon reasonable request.

References

Ahmadi A., and M. Nasseri. 2020. "Do Direct and Inverse Uncertainty Assessment Methods Present the Same Results?" *Journal of Hydroinformatics* 22, no. 4: 842–855. https://doi.org/10.2166/hydro.2020.190.

Ahmadi, A., M. Nasseri, and D. P. Solomatine. 2019. "Parametric Uncertainty Assessment of Hydrological Models: Coupling UNEEC-P and a Fuzzy General Regression Neural Network." *Hydrological Sciences Journal* 64, no. 9: 1080–1094. https://doi.org/10.1080/02626667.2019.1610565.

Alptekin, A., D. C. Broadstock, X. Chen, and D. Wang. 2018. "Time-Varying Parameter Energy Demand Functions: Bench-Marking State-Space Methods Against Rolling Regressions." *Energy Economics* 82: 26–41.

Berg, P., C. Donnelly, and D. Gustafsson. 2018. "Near-Real-Time Adjusted Reanalysis Forcing Data for Hydrology." *Hydrology and Earth System Sciences* 22, no. 2: 989–1000. https://doi.org/10.5194/hess-22-989-2018.

Bergman, D. L., L. Magnusson, J. Nilsson, and F. Vitart. 2019. "Seasonal Forecasting of Tropical Cyclone Landfall Using ECMWF's System 4." *Journal of Weather and Forecasting* 34: 1239–1255. https://doi.org/10.1175/WAF-D-18-0032.1.

Chevuturi, A., A. G. Turner, S. Johnson, et al. 2021. "Forecast Skill of the Indian Monsoon and Its Onset in the ECMWF Seasonal Forecasting System 5 (SEAS5)." *Climate Dynamics* 56: 2941–2957. https://doi.org/10.1007/s00382-020-05624-5.

Copernicus Climate Change Service, Climate Data Store. 2018. "Seasonal Forecast Monthly Statistics on Pressure Levels. Copernicus Climate Change Service (C3S) Climate Data Store (CDS)." https://doi.org/10.24381/cds.0b79e7c5.

Crochemore, L., M.-H. Ramos, F. Pappenberger, and C. Perrin. 2017. "Seasonal Streamflow Forecasting by Conditioning Climatology With Precipitation Indices." *Hydrology and Earth System Sciences* 21: 1573–1591. https://doi.org/10.5194/hess-21-1573-2017.

Dehghanipour, A. M., B. Zahabiyoun, G. Schoups, and H. Babazadeh. 2019. "A WEAP-MODFLOW Surface Water-Groundwater Model for the Irrigated Miyandoab Plain, Urmia Lake Basin, Iran: Multi-Objective Calibration and Quantification of Historical Drought Impacts." *Agricultural Water Management* 223: 105704. https://doi.org/10.1016/j.agwat.2019.105704.

Dorward, P., H. Osbahr, C. Sutcliffe, and R. Mbeche. 2019. "Supporting Climate Change Adaptation Using Historical Climate Analysis." *Climate and Development* 12: 469–480. https://doi.org/10.1080/17565529.2019.1642177.

Ehret, U., E. Zehe, V. Wulfmeyer, K. Warrach-Sagi, and J. Liebert. 2012. "Should We Apply Bias Correction to Global and Regional Climate Model Data?" *Hydrology and Earth System Sciences* 16: 3391–3404. https://doi.org/10.5194/hess-16-3391-2012.

Fang, G. H., J. Yang, Y. N. Chen, and C. Zammit. 2015. "Comparing Bias Correction Methods in Downscaling Meteorological Variables for a Hydrologic Impact Study in an Arid Area in China." *Hydrology and Earth System Sciences* 19: 2547–2559. https://doi.org/10.5194/hess-19-2547-2015.

Gbangou, T., F. Ludwig, E. van Slobbe, L. Hoang, and G. Kranjac-Berisavljevic. 2019. "Seasonal Variability and Predictability of Agro-Meteorological Indices: Tailoring Onset of Rainy Season Estimation to Meet Farmers' Needs in Ghana." *Climate Services* 14: 19–30. https://doi.org/10.1016/j.cliser.2019.04.002.

Gbangou, T., F. Ludwig, E. van Slobbe, L. Hoang, and G. Kranjac-Berisavljevic. 2020. "Rainfall and Dry Spell Occurrence in Ghana: Trends and Seasonal Predictions With a Dynamical and a Statistical Model." *Theoretical and Applied Climatology* 2020: 371–387. https://doi.org/10.1007/s00704-020-03212-5.

Ghajarnia, N., A. Liaghat, and P. D. Arasteh. 2015. "Comparison and Evaluation of High-Resolution Precipitation Estimation Products in Urmia Basin-Iran." *Atmospheric Research* 158: 50–65. https://doi.org/10.1016/j.atmosres.2015.02.010.

Ghazi, B., S. Dutt, and A. Torabi Haghighi. 2023. "Projection of Future Meteorological Droughts in Lake Urmia Basin, Iran." *Water* 15, no. 8: 1558. https://doi.org/10.3390/w15081558.

Golian, S., and C. Murphy. 2022. "Evaluating Bias-Correction Methods for Seasonal Dynamical Precipitation Forecasts." *Journal of Hydrometeorology* 23, no. 8: 1350–1363.

Grillakis, M., A. Koutroulis, and I. Tsanis. 2018. "Improving Seasonal Forecasts for Basin Scale Hydrological Applications." *Water* 10, no. 11: 1593. https://doi.org/10.3390/w10111593.

Gudmundsson, L., J. B. Bremnes, J. E. Haugen, and T. Engen-Skaugen. 2012. "Technical Note: Downscaling RCM Precipitation to the Station Scale Using Statistical Transformations—A Comparison of Methods." *Hydrology and Earth System Sciences* 16: 3383–3390. https://doi.org/10.5194/hess-16-3383-2012.

Habibi, M., I. Babaeian, and W. Schöner. 2021. "Changing Causes of Drought in the Urmia Lake Basin—Increasing Influence of Evaporation and Disappearing Snow Cover." *Water* 13, no. 22: 3273. https://doi.org/10.3390/w13223273.

Hamidi-Razi, H., M. Mazaheri, M. Carvajalino-Fernández, and J. Vali-Samani. 2019. "Investigating the Restoration of Lake Urmia Using a Numerical Modelling Approach." *Journal of Great Lakes Research* 45, no. 1: 87–97.

Hassanzadeh, E., A. Nazemi, J. Adamowski, T. H. Nguyen, and V. T. Van-Nguyen. 2019. "Quantile-Based Downscaling of Rainfall Extremes: Notes on Methodological Functionality, Associated Uncertainty and Application in Practice." *Advances in Water Resources* 131: 103371.

Hosseini-Moghari, S. M., S. Araghinejad, M. J. Tourian, K. Ebrahimi, and P. Döll. 2018. "Quantifying the Impacts of Human Water Use and Climate Variations on Recent Drying of Lake Urmia Basin: The Value of Different Sets of Spaceborne and In-Situ Data for Calibrating a

Hydrological Model." *Hydrology and Earth System Sciences* 24, no. 4: 1939–1956.

Johnson, S. J., T. N. Stockdale, L. Ferranti, et al. 2019. "SEAS5: The New ECMWF Seasonal Forecast System." *Geoscientific Model Development* 12: 1087–1117. https://doi.org/10.5194/gmd-12-1087-2019.

Kalaba, R., and L. Tesfatsion. 1988a. "The Flexible Least Squares Approach to Time-Varying Linear Regression." *Journal of Economic Dynamics and Control* 12, no. 1: 43–48.

Kalaba, R., and L. Tesfatsion. 1988b. "Exact Sequential Filtering, Smoothing, and Prediction for Nonlinear Systems." *Nonlinear Analysis* 12, no. 6: 599–615.

Kalaba, R., and L. Tesfatsion. 1989a. "Time-Varying Linear Regression via Flexible Least Squares." *Computers and Mathematics with Applications* 17, no. 8/9: 1215–1245.

Kalaba, R., and L. Tesfatsion. 1989b. "Sequential Nonlinear Estimation With Nonaugmented Priors." *Journal of Optimization Theory and Applications* 60, no. 3: 421–438.

Knoben, W. J. M., J. E. Freer, and R. A. Woods. 2019. "Technical Note: Inherent Benchmark or Not? Comparing Nash–Sutcliffe and Kling–Gupta Efficiency Scores." *Hydrology and Earth System Sciences* 23: 4323–4331. https://doi.org/10.5194/hess-23-4323-2019.

Kottek, M., J. Grieser, C. Beck, B. Rudolf, and F. Rubel. 2006. "World Map of the Köppen-Geiger Climate Classification Updated." *Meteorologische Zeitschrift* 15, no. 3: 259–263.

Lucatero, D., H. Madsen, J. C. Refsgaard, J. Kidmose, and K. H. Jensen. 2018. "On the Skill of Raw and Post-Processed Ensemble Seasonal Meteorological Forecasts in Denmark." *Hydrology and Earth System Sciences* 22: 6591–6609. https://doi.org/10.5194/hess-22-6591-2018.

Lucchetti, R., and F. Valentini. 2024. "Linear Models With Time-Varying Parameters: A Comparison of Different Approaches." *Computational Statistics* 39: 3523–3545. https://doi.org/10.1007/s00180-023-01452-3.

Manzanas, R., J. M. Gutiérrez, J. Bhend, et al. 2019. "Bias Adjustment and Ensemble Recalibration Methods for Seasonal Forecasting: A Comprehensive Intercomparison Using the C3S Dataset." *Climate Dynamics* 53: 1287–1305. https://doi.org/10.1007/s00382-019-04640-4.

Montana, G., K. Triantafyllopoulos, and T. Tsagaris. 2009. "Flexible Least Squares for Temporal Data Mining and Statistical Arbitrage." *Expert Systems with Applications* 36, no. 2: 2819–2830.

Mori, P., T. Schwitalla, M. B. Ware, K. Warrach-Sagi, and V. Wulfmeyer. 2021. "Downscaling of Seasonal Ensemble Forecasts to the Convection-Permitting Scale Over the Horn of Africa Using the WRF Model." *International Journal of Climatology* 41, no. Suppl. 1: E1791–E1811. https://doi.org/10.1002/joc.6809.

Nasseri, M., A. Ansari, and B. Zahraie. 2014. "Uncertainty Assessment of Hydrological Models With Fuzzy Extension Principle: Evaluation of a New Arithmetic Operator." *Water Resources Research* 50: 1095–1111. https://doi.org/10.1002/2012WR013382.

Nasseri, M., G. Schoups, and M. Taheri. 2022. "A Spatiotemporal Framework to Calibrate High-Resolution Global Monthly Precipitation Products: An Application to the Urmia Lake Watershed in Iran." *International Journal of Climatology* 42, no. 4: 2169–2194. https://doi.org/10.1002/joc.7358.

Nasseri, M., B. Zahraie, A. Ansari, and D. P. Solomatin. 2013. "Uncertainty Assessment of Monthly Water Balance Models Based on Incremental Modified Fuzzy Extension Principle Method." *Journal of Hydroinformatics* 15, no. 4: 1340–1360.

Ogutu, G. E. O., W. H. P. Franssen, I. Supit, P. Omondi, and R. W. A. Hutjes. 2016. "Skill of ECMWF System-4 Ensembles Seasonal Climate Forecast for East Africa." *International Journal of Climatology* 37, no. 5: 2734–2756. https://doi.org/10.1002/joc.4876.

Peel, M. C., B. L. Finlayson, and T. A. McMahon. 2007. "Updated World Map of the Köppen-Geiger Climate Classification." *Hydrology and Earth System Sciences Discussions* 4, no. 2: 439–473.

Rao, R. 1995. Multivariate Flexible Least Squares Analysis of Hydrological Time Series. *International Association of Hydrological Sciences, Publication* 231: 359–366.

Ratri, D. N., A. Weerts, R. Muharsyah, et al. 2023. "Calibration of ECMWF SEAS5 Based Streamflow Forecast in Seasonal Hydrological Forecasting for Citarum River Basin, West Java, Indonesia." *Journal of Hydrology: Regional Studies* 45: 10130. https://doi.org/10.1016/j.ejrh.2022.10130.

Ratri, D. N., K. Whan, and M. Schmeits. 2019. "A Comparative Verification of Raw and Bias-Corrected ECMWF Seasonal Ensembles Prediction Reforecasts in Java (Indonesia)." *Journal of Applied Meteorology and Climatology* 58: 1709–1723. https://doi.org/10.1175/JAMC-D-18-0210.1.

Ratri, D. N., K. Whan, and M. Schmeits. 2021. "Calibration of ECMWF Seasonal Ensemble Precipitation Reforecasts in Java (Indonesia) Using Bias-Corrected Precipitation and Climate Indices." *Weather and Forecasting* 36: 1375–1386. https://doi.org/10.1175/WAF-D-20-0124.1.

Rizzo, M. L., and G. J. Székely. 2016. "Energy Distance." *Wiley Interdisciplinary Reviews: Computational Statistics* 8: 27–38.

Roebber, P. J. 2009. "Visualizing Multiple Measures of Forecast Quality." *Weather and Forecasting* 24: 601–608. https://doi.org/10.1175/2008WAF2222159.1.

Sarojini, B. B., P. A. Stott, and E. Black. 2016. "Detection and Attribution of Human Influence on Regional Precipitation." *Nature Climate Change* 6: 669–675. https://doi.org/10.1038/nclimate2976.

Schepen, A., Q. J. Wang, and Y. Everingham. 2016. "Calibration, Bridging and Merging to Improve GCM Seasonal Temperature Forecasts in Australia." *Monthly Weather Review* 144: 2421–2441. https://doi.org/10.1175/MWR-D-15-0384.1.

Schick, S., O. Rössler, and R. Weingartner. 2019. "An Evaluation of Model Output Statistics for Subseasonal Streamflow Forecasting in European Catchments." *Journal of Hydrometeorology* 20, no. 7: 1399–1416. https://journals.ametsoc.org/view/journals/hydr/20/7/jhm-d-18-0195_1.xml.

Taheri, M., M. Emadzadeh, M. Gholizadeh, M. Tajrishi, M. Ahmadi, and M. Moradi. 2019. "Investigating the Temporal and Spatial Variations of Water Consumption in Urmia Lake River Basin Considering the Climate and Anthropogenic Effects on the Agriculture in the Basin." *Agricultural Water Management* 213: 782–791.

Tareghian, R., and P. Rasmussen. 2013. "Statistical Downscaling of Precipitation Using Quantile Regression." *Journal of Hydrology* 478: 122–135.

Wang, Q. J., Y. Shao, Y. Song, et al. 2019. "An Evaluation of ECMWF SEAS5 Seasonal Climate Forecasts for Australia Using a New Forecast Calibration Algorithm." *Environmental Modelling and Software* 122: 104550. https://doi.org/10.1016/j.envsoft.2019.104550.

Worku, G., E. Teferi, A. Bantider, and Y. Dile. 2020. "Statistical Bias Correction of Regional Climate Model Simulations for Climate Change Projection in the Jemma Sub-Basin, Upper Blue Nile Basin of Ethiopia." *Theoretical and Applied Climatology* 139: 1569–1588.

Xiong, L., M. I. N. Wan, X. Wei, and K. M. O'connor. 2009. "Indices for Assessing the Prediction Bounds of Hydrological Models and Application by Generalised Likelihood Uncertainty Estimation." *Hydrological Sciences Journal* 54, no. 5: 852–871.

Zeng, Q., Y. Wang, L. Chen, Z. Wang, H. Zhu, and B. Li. 2018. "Inter-Comparison and Evaluation of Remote Sensing Precipitation Products Over China From 2005 to 2013." *Remote Sensing* 10: 168.

Zhang, Q., Q. Tang, X. Liu, S.-M. Hosseini-Moghari, and P. Attarod. 2020. "Improving Princeton Forcing Dataset Over Iran Using the Delta-Ratio Method." *Water* 12: 630.