

Using quantile mapping and random forest for bias-correction of high-resolution reanalysis precipitation data and CMIP6 climate projections over Iran

Maryam Raeesi¹ | Ali Asghar Zolfaghari¹  | Seyed Hasan Kaboli¹ | Mohammad Rahimi¹ | Joris de Vente² | Joris P. C. Eekhout²

¹Faculty of Desert Studies, Semnan University, Semnan, Iran

²Soil and Water Conservation Research Group, CEBAS-CSIC, Spanish Research Council, Murcia, Spain

Correspondence

Ali Asghar Zolfaghari, Faculty of Desert Studies, Semnan University, Semnan, Iran.
Email: azolfaghari@semnan.ac.ir

Abstract

Climate change is expected to cause important changes in precipitation patterns in Iran until the end of 21st century. This study aims at evaluating projections of climate change over Iran by using five climate model outputs (including ACCESS-ESM1-5, BCC-CSM2-MR, CanESM5, CMCC-ESM2 and MRI-ESM2-0) of the Coupled Model Intercomparison Project phase 6 (CMIP6), and performing bias-correction using a novel combination of quantile mapping (QM) and random forest (RF) between the years 2015 and 2100 under three shared socioeconomics pathways (SSP2-4.5, SSP3-7.0 and SSP5-8.5). First, bias-correction was performed on ERA5-Land reanalysis data as reference period (1990–2020) using the QM method, then the corrected ERA5-Land reanalysis data was considered as measured data. Based on the corrected ERA5-Land reanalysis data (1990–2020) and historical simulations (1990–2014), the future projections (2015–2100) were also bias-corrected utilizing the QM method. Next, the accuracy of the QM method was validated by comparing the corrected ERA5-Land reanalysis data with model outputs for overlapping years between 2015 and 2020. This comparison revealed persistent biases; hence, a combination of QM-RF method was applied to rectify future climate projections until the end of the 21st century. Based on the QM result, CMCC-ESM2 revealed the highest RMSE in both SSP2-4.5 and SSP3-7.0 amounting to 331.74 and 201.84 mm·year⁻¹, respectively. Particularly, the exclusive use of the QM method displayed substantial errors in projecting annual precipitation based on SSP5-8.5, notably in the case of ACCESS-ESM1-5 (RMSE = 431.39 mm·year⁻¹), while the RMSE reduced after using QM-RF method (197.75 mm·year⁻¹). Obviously, a significant enhancement in results was observed upon implementing the QM-RF combination method in CMCC-ESM2 under both SSP2-4.5 (RMSE = 139.30 mm·year⁻¹) and SSP3-7.0 (RMSE = 151.43 mm·year⁻¹) showcasing approximately reduction in RMSE values by 192.43 and 50.41 mm·year⁻¹, respectively. Although each bias-

corrected model output was evaluated individually, multi-model ensemble (MME) was also created to project the annual future precipitation pattern in Iran. By considering that combination of QM-RF method revealed the lower errors in correcting model outputs, we used the QM-RF technique to create the MME. Based on SSP2-4.5, the MME climate projections highlight imminent precipitation reductions ($>10\%$) across large regions of Iran, conversely projecting increases ranging from 10% to over 20% in southern areas under SSP3-7.0. Moreover, MME projected dramatic declines under SSP5-8.5, especially impacting central, eastern, and northwest Iran. Notably, the most pronounced possibly decline patterns are projected for arid regions (central plateau) and eastern areas under SSP2-4.5, SSP3-7.0 and SSP5-8.5.

KEY WORDS

annual precipitation projection, bias-correction methods, data mining, ERA5-Land reanalysis data, quantile mapping, random forest algorithm

1 | INTRODUCTION

Precipitation, as a key climatic component, has a central role in a wide-range of studies particularly regarding hydrological and soil erosion modelling, water balances and climate change projections (Talchabhadel et al., 2021). However, lack of sufficient reliable data has long been a potential barrier in generating a high-quality precipitation map for present and future time spans, especially under climate change. Reanalysis data can provide reasonable high spatiotemporal resolution and long-term coverage of precipitation variability for historic periods (Jiang et al., 2021). The European Centre for Medium-Range Weather Forecast (ECMWF) released the recent fifth generation of reanalysis dataset (Muñoz-Sabater et al., 2021) (ERA5-Land) that outperforms its predecessor (Amjad et al., 2020; Gleixner et al., 2020). Furthermore, the Coupled Model Intercomparison Project (CMIP6) provides climate change projections under new scenarios (Fan et al., 2022), including for Iran. Using global climate models (GCMs) outputs is often restricted because of relatively coarse resolution and systematic biases (Yang et al., 2019) that prevents models from representing regional climatic processes properly. Thus, considering the correction of model biases is an essential step (Beyer et al., 2020) for climate projection. Numerous bias-correction methods have been developed and evaluated in the latest years to minimize the model biases and their impacts on hydrological processes (Eekhout & de Vente, 2019). One commonly applied bias-correction method is quantile mapping (QM) concerned about erasing quantile-dependent biases. The probability transformation function in QM relies on the empirical distribution function (Zhu et al., 2022). Besides the QM

approach, other kinds of statistical approaches have been used including multiple linear regression (MLR) (Ahmed et al., 2015), principal component analysis (PCA) (Chu et al., 2008), support vector machines (SVM) (Chen et al., 2010), as well as random forest (RF) (Shi et al., 2015). RF is a relatively recent development in machine learning (ML) that has already had broad demand in various predictive tasks counting weather and climate forecasting (Li et al., 2019; Tong et al., 2019).

Several studies are found to address the impacts of climate change on precipitation and related hydrological processes. For example, Eekhout and de Vente (2019) assessed the application of three bias-correction methods, delta change (DC), scaled distribution mapping (SDM) and QM to project regional soil erosion using climate models. Results showed that individual climate models may forecast adverse changes to multi-model ensemble (MME) average. They further revealed that QM and SDM projected an increase in extreme precipitation, while DC method projected a decrease in most cases. Tang et al. (2021) conducted research on a spatial random forest downscaling LST method (SRFD). They showed that SRFD outperforms other methods in statistics due to the benefits from the supplement of the LST spatial feature.

In Asia, Buda et al. (2005) and Saddique et al. (2020) concluded an increase in extreme precipitation. By analysing dataset from 27 CMIP6 models, they examined the projected changes in temperature and precipitation over South Asian countries. They also found that the annual mean precipitation and temperature are projected to increase under the SSP1-2.6, SSP2-4.5 and SSP5-8.5 scenarios. Jiang et al. (2020) revealed their results on precipitation projection over central Asia. They used 15 models from CMIP6 and emphasized that annual mean

precipitation increases under SSP1-2.6, SSP2-4.5, SSP3-7.0 and SSP5-8.5 scenarios by end of the century. Piao et al. (2021) analysed the mean state and precipitation projection over the monsoon transitional zone (MTZ) in China. A significant increase was observed in annual total precipitation amount, and ensemble means of CMIP6 models presented a totally enhancement in obtaining the annual mean precipitation in study area. Kamruzzaman et al. (2023) used 18 GCMs to predict changes in future precipitation and air temperature across Bangladesh. They used simple quantile mapping (SQM) technique to remove the systematic biases in model outputs. By using MME of the 18 bias-corrected GCMs, they revealed MME outperformed individual models and significantly improved bias-corrected projections for climate variables. They also showed rainfall is expected a higher increase in distant future than in the near and mid-future.

In a review of the literature focused on analysing the impacts of climate change on precipitation in Iran, Doulabian et al. (2021) used six synoptic weather stations and monthly datasets for a 20-year period and outputs of 25 GCMs. They concluded mean precipitation will increase in wet seasons and decrease during dry seasons. Results indicated GCMs over/underestimated climatic variables on regional scales. Hong et al. (2021) used 24 MME to project the future changes of extreme rainfall over Iran. They indicated that rainfall will be increased notably in western and southwestern parts of Iran. Amini et al. (2022) conducted research on removing biases from raw datasets using QM method and assessed the performance of global rainfall forecasting systems. Results demonstrated in humid, semi-humid, Mediterranean and arid climate zones, for example, southwest, northwest and northeast parts of Iran, the predicted precipitation is highly correlated with ground-based observation. Even though, in semi-arid and extra-arid areas, the correlation coefficient between predicted and observed datasets was calculated very low.

The first step in addressing the impacts of climate change is to determine the climate change projection by the year 2100. Based on the literature, few researches conducted in Iran with the main reflection of precipitation projection using CMIP6 models either individually or as MME to understand the relative trend of annual precipitation. However, to date, no study has applied the RF methodology to enhance bias-correction analysis and project annual precipitation under shared socioeconomics pathways (SSP) scenarios in the region.

Therefore, this paper aims at (1) assessing the effectiveness of integrating QM and RF methodologies in correcting biases within the five CMIP6 model outputs, (2) evaluating the spatial patterns and uncertainty of projected annual mean precipitation under SSP scenarios

using both the five CMIP6 outputs and MME and (3) analysing the variability in precipitation projections in model outputs to provide the potential future climate trends in the country.

This study contributes to the advancement of knowledge on climate change projections in Iran, offering valuable insights for policymakers, researchers and stakeholders engaged in climate adaptation and mitigation strategies.

2 | DATA AND METHODOLOGY

2.1 | Study area

The study area, Iran, is located in western Asia, in an arid and semi-arid part of the world. Iran extends over 24°N and 40°N latitude, and 44°E and 64°E longitude, which has a total area of 1,640,195 km², and approximately a population of 81 million inhabitants (Figure 1). Iran has diverse climates; moderate and wet on the coast of the Caspian Sea, continental and arid in the plateau, cold in high mountains, dry and hot in deserts of the southern coast and southeast (Doulabian et al., 2021). Iran exhibits significantly lower mean precipitation levels compared to the global average. While regions like the Zagros mountainous ranges and the northern slopes of Alborz experience annual rainfall exceeding 1000 mm, a substantial portion encompassing over two-thirds of Iran's landmass sustains a mean yearly precipitation of less than 300 mm (Alijani et al., 2008). The geographical positioning of Iran plays a crucial role in shaping the overall distribution of precipitation. Therefore, annual rainfall ranges from lower than 100 mm in central areas to greater than 1200 mm across the Caspian Sea coast in the north of Iran (Sadeghi et al., 2017). Furthermore, the Alborz and Zagros mountain ranges contribute to contrasting spatial and temporal patterns in precipitation across northern and western regions of Iran, as discussed by Dinpashoh et al. (2004).

The distribution of available synoptic stations within the study area, Iran, is depicted in Figure 1a, and the black polygons represent the political boundaries of each province. Furthermore, the digital elevation model (DEM) is presented in Figure 1b, which ranges from -32 to 5379 m across the entire country.

2.2 | Data collection and preparation

Our study centred on three distinct categories of precipitation datasets: monthly precipitation records obtained from synoptic stations, monthly data sourced from the

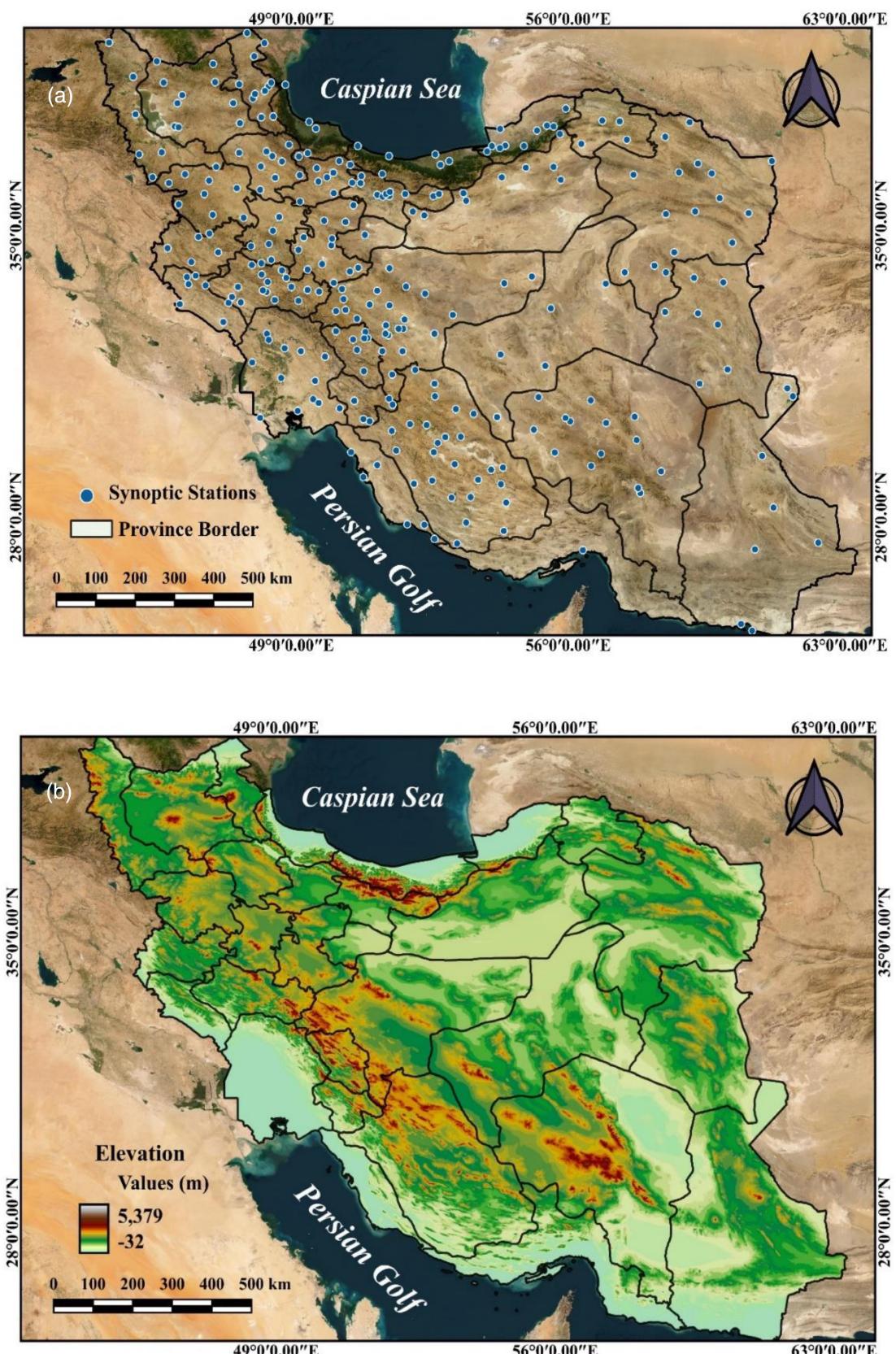


FIGURE 1 Geographical border of Iran and its provinces where distribution of available synoptic stations over the entire country (a), and topography is shown in the colour scale of background (b).

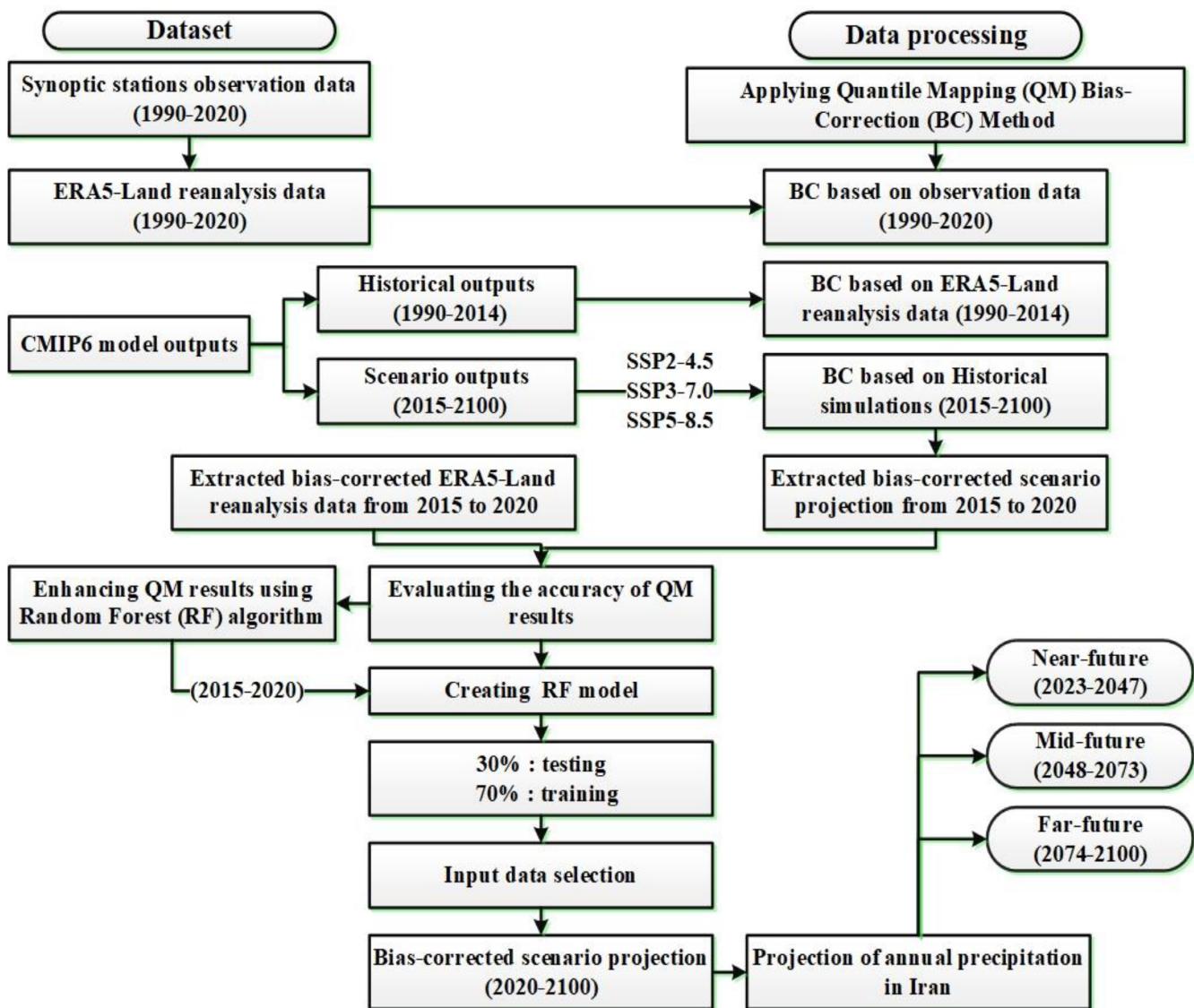


FIGURE 2 Conceptual methodology framework.

ERA5-Land grid-based reanalysis, and climate projections derived from five global circulation models (GCM) within the CMIP6 framework (Figure 2). These datasets were selected based on availability and prior literature reviews, all provided at a monthly resolution and examined across three distinct shared socioeconomic pathways emission scenarios (SSPs).

2.3 | Baseline data

Monthly precipitation data from 285 available synoptic stations were employed to calculate the annual average precipitation across Iran for the baseline period of 1990–2020, as illustrated in Figures 1 and 2, reflecting observational data. The absence of sufficient synoptic stations or recorded climatic data in certain regions has proven to be

a significant drawback or challenging issue in developing accurate precipitation maps. To address such challenges, gridded climatic datasets have emerged as effective solutions, with a particular recommendation to leverage reanalysis grid-based datasets to mitigate this disparity. The utilization of reanalysis precipitation products offers a significant advantage by alleviating the limitations associated with ground-based data, especially in areas lacking rain-gauge stations or adequate data coverage. The advantage of using reanalysis precipitation products is that they reduce the deficiency of ground-based data where neither rain-gauge stations nor sufficient data have been equipped.

In this study, monthly ERA5-Land reanalysis data was downloaded with approximately 9 km spatial resolution to examine the spatial and temporal precipitation patterns over the country. This product has been widely

TABLE 1 Summary of the five CMIP6 models used.

Row	CMIP6 name	Country	Horizontal resolution ^a	Nominal resolution ^b	Key references
1	ACCESS-ESM1-5	Australia	1.9° × 1.2°	100 km	Law et al. (2017)
2	BCC-CSM2-MR	China	1.1° × 1.1°	100 km	Wu et al. (2019)
3	CanESM5	Canada	2.8° × 2.8°	100 km	Swart et al. (2019)
4	CMCC-ESM2	Italy	1° × 1°	100 km	Lovato et al. (2022) and Peano et al. (2019)
5	MRI-ESM2-0	Japan	0.5° × 0.5°	100 km	Yukimoto et al. (2019)

^aLon. by lat. in degree.

^bAtmos, land, ocean.

accepted in global climate-related research (Zou et al., 2022). After that, the yearly ERA5-Land reanalysis data was generated in 5 km spatial resolution and used for next analysis.

2.4 | Historical simulations and scenario projections

In this study, we selected five GCMs (Table 1) from the CMIP6 projections to obtain the yearly precipitation which downloaded from the ESGF official website (<https://esgf-node.llnl.gov/search/cmip6>). These CMIP6 models are selected based on the common accessibility of precipitation for both historical simulations (1990–2014) and projected scenarios (2015–2100). Since GCMs produce results on the global scale (coarser resolution; Table 1), they tend to over/underestimate climatic variables on both regional and global scales, failing to resolve the finer scale climate variability (Ahmadalipour et al., 2017). Model outcomes were obtained for different future scenarios based on three shared socioeconomic pathways (SSPs). The SSP2-4.5 scenario represents a medium-forcing scenario, SSP3-7.0 represents a medium to high-end of the range of future forcing pathways, and SSP5-8.5 represents a high-forcing scenario. Detailed descriptions of the SSPs are available in O'Neill et al. (2017).

2.5 | Data processing and analysis

2.5.1 | Bias quantile mapping approach

Quantile mapping (QM) is a statistical method used to adjust the distribution of a variable to better match an observed distribution. It is designed to correct the probability density function (PDF) of climate data projected by GCMs to coordinate with the PDF of observation data (Boé et al., 2007). The QM method was applied based on empirical cumulative probability function (ECDF)

described in Michelangeli et al. (2009), which indicates the CDF relation between the observed datasets and simulations to reduce the systematic biases and model noises in precipitation simulations (1990–2014) and projections (2015–2100) in this study.

Although QM can be used for bias-correction of spatial data, it may not be an accurate technique for bias-correction of data with substantial differences from the measured data. Nevertheless, it is suitable for correcting spatial data, such as reanalyzed precipitation data that they have a strong correlation with rain-gauge data. Nevertheless, in this study, owing to the absence of pronounced disparities between both the station and the ERA5-Land reanalysis data (1990–2020), the QM method was employed to correct biases within the reanalysis dataset. This approach aimed to align annual precipitation values as closely as feasible to those derived from the station data. Although in some recent studies, ERA5-Land reanalysis data have been directly utilized as observational data without employing bias-correction methods (Muñoz-Sabater et al., 2021; Zhao & He, 2022; Zou et al., 2022).

The ERA5-Land reanalysis data were corrected based on station data for each year utilizing the QM method. The applied equations (Equations (1)–(3)) to address systematic biases in spatial data by considering the probability of precipitation occurrence (Prob_{gp}) on individual grid pixels derived from ERA5-Land reanalysis data (1990–2020),

$$\text{Prob}_{gp} = \text{ECDF}_{\text{ERA5-Land}}(P_{\text{ERA5-Land},gp}), \quad (1)$$

$$\text{CF}_{gp} = \text{ECDF}_{\text{obs,station}}^{-1}(\text{Prob}_{gp}) - \text{ECDF}_{\text{ERA5-Land}}^{-1}(\text{Prob}_{gp}), \quad (2)$$

$$P_{gp,BC} = P_{\text{ERA5-Land},gp} + \text{CF}_{gp}. \quad (3)$$

The term $\text{ECDF}_{\text{ERA5-Land}}$ stands for the empirical cumulative distribution function (ECDF) of ERA5-Land reanalysis data at station points, computed annually to account for variations over time.

The transfer function relating the ECDF of observed/station precipitation to ERA5-Land reanalysis data can be acquired from Equation (2). Moreover, the correction factor (CF_{gp}) indicates the discrepancies between observed and ERA5-Land reanalysis precipitation data in given pixel determined by the $Prob_{gp}$. Subsequently, the CF_{gp} is applied to adjust the ERA5-Land reanalysis precipitation data ($P_{ERA5-Land, gp}$) at each pixel, yielding the bias-corrected ERA5-Land reanalysis precipitation data ($P_{gp, BC}$), as detailed in Equation (3). For next steps, the corrected ERA5-Land reanalysis data was considered as measured/baseline data (1990–2020) in the study area.

The QM method was used for bias-correction of CMIP6 models. The bias-correction in time series was done in each pixel. The probability (Prob) of future precipitation in a pixel (2015–2100) is determined by comparing the model's empirical cumulative distribution function (ECDF) from the baseline period with the model's projected perceptions for that pixel ($P_{mod,f}$) (Piani et al., 2010). The probability (Prob) arises due to the coarse resolution of the climate model outputs (future time spans) in aligning them with the empirical cumulative distribution function (ECDF) during the calibration period (Equation (4)),

$$Prob = ECDF_{mod, ref}(P_{mod,f}). \quad (4)$$

Next, Equation (5) calculates the correction factor (CF), which reflects the disparity between the inverse ECDF ($ECDF^{-1}$) of the observed and projected data at a specific probability level. This disparity signifies the necessary adjustment to align the projected distribution with the observed distribution, aiming to enhance the accuracy of the model (Eekhout & de Vente, 2019).

Eventually, the correction factor (CF) added up to the raw future climate model outputs to acquire the bias-corrected value (Equation (6)),

$$CF = ECDF_{obs, ref}^{-1}(Prob) - ECDF_{mod, ref}^{-1}(Prob), \quad (5)$$

$$P_{mod,f, BC} = P_{mod,f} + CF. \quad (6)$$

In this research, QM analysis was conducted at an annual scale using the “qmap” package within the R environment. Initially, the ERA5-Land reanalysis data for the period spanning from 1990 to 2020 were subjected to bias-correction using ground-based observational data. Subsequently, the identical methodology was applied to rectify the systematic biases present in future precipitation projections at each pixel (grid-to-grid) by comparing historical simulations from 1990 to 2014 with the bias-

corrected ERA5-Land reanalysis data. Next, the future projections spanning the years 2015–2100 were subjected to bias-correction utilizing the QM method. A comparative analysis was then conducted between the results obtained from ERA5-Land reanalysis data and the bias-corrected future projections for six shared years between 2015 and 2020. Despite the presence of a notable biases observed during the comparison of results, the random forest (RF) algorithm was employed to mitigate and reduce these discrepancies. Further details regarding the RF approach will be elaborated in the subsequent sections.

2.5.2 | Random forest algorithm

Random forest (RF) is a machine learning algorithm (Shiru et al., 2019) integrating the output of numerous decision trees to extend a single prediction. Building multiple independent trees to form a climate change projection ensemble and the prediction is set to be the most popular class in ensemble predictions. RF algorithm utilizes both bagging and feature randomness to build up an uncorrelated forest of decision trees (Breiman, 2001).

RF method acquire nonlinear relationships that was spatially applied on both the annual bias-corrected ERA5-Land reanalysis from 2015 to 2020 and climate model projections for the period 2015–2020 based on SSP2-4.5, SSP3-7.0 and SSP5-8.5. RF model was created by two datasets including bias-corrected future projection (for each scenario) as X variable or model input, and the corrected ERA5-Land reanalysis data for the same periods of time as Y variable or model output from 2015 to 2020. The created RF models were performed to correct the remaining biases for the future projections in each scenario to the end of the 21st century. In current modelling experiments, randomly, 70% of all data was used to train RF model and the remaining 30% was then applied for testing the model. The data were randomly selected to build the RF model, using Classification and Regression Training (caret) package in R, which covered the entire range of precipitation amounts in Iran. Finally, QM outputs of the climate model projections for 2020–2100 were utilized as input for the RF model. This provided the final bias-corrected process for future climate projections.

2.5.3 | Multi-model ensemble analysis

Managing substantial uncertainty in GCM projections is a complex challenge. Notwithstanding numerous

TABLE 2 Evaluating model's performance by calculating statistical indices.

Statistical indices	Equation	Unit	Reference
Pearson correlation coefficient (CORR)	$r = \frac{\sum_{i=1}^n (p_i - \bar{p})(m_i - \bar{m})}{\sqrt{\sum_{i=1}^n (m_i - \bar{m})^2} \sqrt{\sum_{i=1}^n (p_i - \bar{p})^2}}$	-	Jeferson de Medeiros et al. (2022)
Root-mean-square error (RMSE)	$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (m_i - p_i)^2}$	mm	Liemohn et al. (2021)
Bias	$Bias = \frac{1}{n} \sum_{i=1}^n (m_i - p_i)$	mm	Zolfaghari et al. (2016)

Note: m_i is the observed value, \bar{m} is the mean of observed data, p_i is the predicted value, \bar{p} is the mean of simulated and n refers to the total observation number.

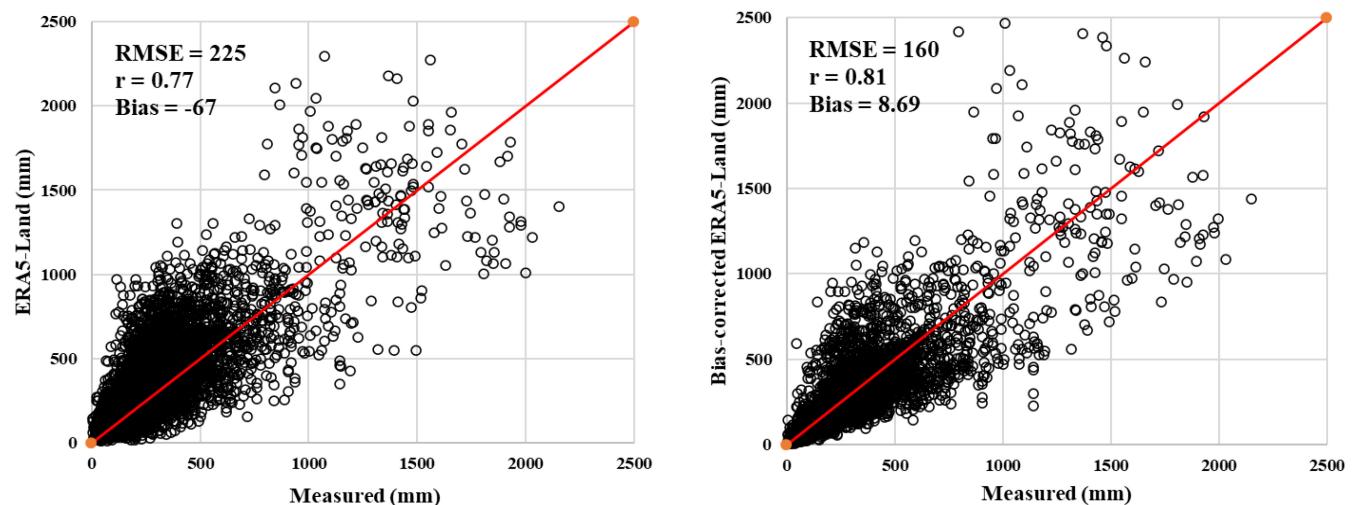


FIGURE 3 Distribution of measured/station's data versus ERA5-Land reanalysis precipitation data before (left) and after (right) BC using QM method.

methods have proposed to consider the climate projection's uncertainties. Multi-model ensemble is the most widely recommended solution to enhance projection accuracy and reliability in this context (Fowler & Kilsby, 2007; Heo et al., 2019; Kamruzzaman et al., 2020a, 2020b). For achieving this purpose, the results of bias-corrected QM-RF outputs were used for creating MME based on the projection annual mean precipitation under SSP2-4.5, SSP3-7.0 and SSP5-8.5 scenarios, separately. We considered whether using MME might minimize the uncertainties of each model in representation of precipitation patterns in Iran.

2.6 | Performance evaluation criteria

In the present study, statistical metrics such as root-mean-square error (RMSE), mean error (Bias) and correlation coefficient (r) were used to evaluate the performance of the bias-correction methods (Table 2). The Bias indicates whether there is a systematic error in the

method and the RMSE shows the accuracy of the bias-correction methods. RMSE, r and Bias equations are presented in Table 2.

3 | RESULTS

3.1 | ERA5-Land reanalysis assessment using QM method

The ERA5-Land data underwent bias-correction (BC) utilizing the QM method. Figure 3 displays scatter-plots showcasing the comparison between the observed data (x-axis) and the ERA5-Land reanalysis data before and after applying QM method on an annual scale. The right graph illustrates the level of agreement between the ERA5-Land data and the ground-based observations subsequent to the BC using QM method. The initial raw ERA5-Land data exhibited a significant bias, overestimating annual precipitation by -67 mm on average annually. However, the calculated bias in the corrected data was

determined to be 8.69, signifying a better match to the ground-based data. Although the RMSE value was determined to be $225 \text{ mm}\cdot\text{year}^{-1}$ before BC, it decreased to 160 mm annually using the QM method. Moreover, the correlation coefficient (r) between the ground-based and the corrected ERA5-Land data was calculated to be 0.81, indicating the corrected ERA5-Land data, exhibits a higher level of consistency with the spatial distribution of annual precipitation, as demonstrated in Figure 3.

Figure 4 illustrates the cumulative distribution function (CDF) of annual precipitation gauged directly from ground stations data (the blue line) and the raw ERA5-Land data (the orange line) in a given year. The findings reveal that the cumulative distribution function (CDF) curve of ERA5-Land data is lower than that of the measured data, suggesting an overestimation of annual precipitation in the study area by ERA5-Land. Following bias-correction, the CDF curve of ERA5-Land data aligns more closely with the CDF curve of the measured data (the black dash line), indicating a reduction in errors in the ERA5-Land data.

If we consider that the annual precipitation amount in the ERA5-Land data is 500 mm and its calculated cumulative frequency is 0.65, then within the same interval, the annual precipitation values corresponding to 0.65 in the cumulative frequency of observations are around 450 mm (Figure 4). It is clear that the measured data

covers annual precipitation values ranging from 100 to approximately 1250 mm. The original ERA5-Land reanalysis data shows a similar pattern, with values extending up to around 1450 mm, whereas the bias-corrected (BC) reanalysis data presents a distribution reaching up to about 1150 mm (the black dash line). The result provides evidence for the effectiveness of the bias-correction process by QM in enhancing the agreement between the ERA5-Land reanalysis data and the observed measurements of annual precipitation.

3.2 | CMIP6 models

3.2.1 | Assessing the raw historical simulations

The statistical analysis for raw simulations (before bias-correction) is provided in Table 3, derived from the five CMIP6 simulations for each individual model. The RMSE was computed as 231.08 for the BCC-CSM2-MR, and 313.42 for the MRI-ESM2-0 simulations, denoting the minimum and maximum errors, respectively. Specifically, the BCC-CSM2-MR simulation exhibited lower RMSE and Bias values, indicative of improved alignment between the simulated and observed data. It implies superior accuracy before bias-correction and the lowest deviation between its simulations and the observed data compared to other models.

The three simulations, ACCESS-ESM1-5, BCC-CSM2-MR and MRI-ESM2-0, demonstrate an overestimation in annual precipitation compared to observed values from 1990 to 2014. Furthermore, before using the QM method (Table 3), CanESM5 and MRI-ESM2-0 exhibited the highest positive and negative bias values, respectively, indicating systematic biases in their simulation. In that case, the simulations of CanESM5 model tended to underestimate the annual precipitation to $118.78 \text{ mm}\cdot\text{year}^{-1}$, while the MRI-ESM2-0 is inclined to overestimate the precipitation approximately $183.49 \text{ mm}\cdot\text{year}^{-1}$. The correlation coefficients (r) ranges from 0.21 to 0.40, for BCC-CSM2-MR and MRI-ESM2-0 demonstrating the highest r , suggesting a moderate positive linear relationship between the observed and simulated values (Table 3).

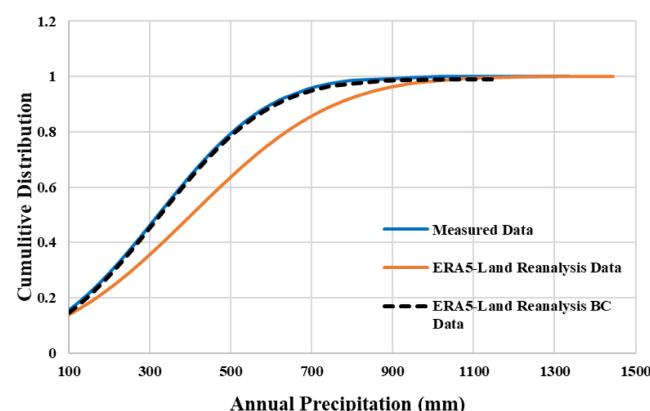


FIGURE 4 Cumulative distribution function (CDF) of annual measured precipitation data versus ERA5-Land and bias-corrected ERA5-Land data for a 1-year period.

TABLE 3 Statistical parameters calculated for raw simulations versus the ground data before BC (1990–2014).

Statistics/metrics	ACCESS-ESM1-5	BCC-CSM2-MR	CanESM5	CMCC-ESM2	MRI-ESM2-0
RMSE	244.01	231.08	282.24	263.46	313.42
Bias	-34.09	-21.79	118.78	99.74	-183.49
r	0.37	0.40	0.21	0.32	0.40

TABLE 4 Statistical parameters calculated the performance of QM method in bias-correcting five CMIP6 models (2015–2020).

Scenario	Models performance in QM method				
	ACCESS-ESM1-5	BCC-CSM2-MR	CanESM5	CMCC-ESM2	MRI-ESM2-0
SSP2-4.5					
RMSE	82.912	130.801	104.328	331.741	115.119
Bias	41.018	11.822	-3.837	-57.278	34.477
<i>r</i>	0.960	0.855	0.910	0.551	0.892
SSP3-7.0					
RMSE	136.092	122.897	99.250	201.849	110.261
Bias	35.981	56.164	3.562	-21.379	47.807
<i>r</i>	0.874	0.908	0.925	0.743	0.924
SSP5-8.5					
RMSE	431.396	227.832	106.502	161.143	145.846
Bias	11.646	25.696	-1.633	82.964	45.989
<i>r</i>	0.450	0.669	0.910	0.807	0.851

TABLE 5 Statistical parameters calculated the performance of QM-RF method in bias-correcting five CMIP6 models (2015–2020).

Scenario	Models performance in QM-RF method				
	ACCESS-ESM1-5	BCC-CSM2-MR	CanESM5	CMCC-ESM2	MRI-ESM2-0
SSP2-4.5					
RMSE	80.954	80.739	70.035	139.302	85.249
Bias	39.488	10.828	-3.671	-35.98	33.248
<i>r</i>	0.913	0.943	0.961	0.567	0.946
SSP3-7.0					
RMSE	79.016	107.279	63.759	151.429	85.561
Bias	34.543	55.920	2.838	-26.304	47.061
<i>r</i>	0.962	0.935	0.970	0.837	0.966
SSP5-8.5					
RMSE	197.756	143.487	66.133	137.487	107.493
Bias	12.321	26.625	-2.063	81.413	43.319
<i>r</i>	0.744	0.840	0.964	0.902	0.929

3.2.2 | Comparison the performance of QM and QM-RF methods

Certain statistical were calculated to evaluate the effectiveness of BC methods for five CMIP6 models, as depicted in Tables 4 and 5. The analysis was conducted using the corrected ERA5-Land data from 2015 to 2020 with the BC projected outputs spanning the same time frame under three SSPs. The calculations encompass RMSE, Bias and correlation coefficient (*r*) using QM in Table 4, and QM-RF in Table 5, respectively. Findings demonstrated that negative Bias values signify an overestimation in annual precipitation values, whereas positive

values indicate an underestimation of precipitation by the model outputs.

Within the SSP2-4.5 scenario, the QM method exhibited optimal performance in both ACCESS-ESM1-5 and CanESM5 models, with RMSE values of 82 and 104 mm, the correlation coefficients (*r*) of 0.96 and 0.91, respectively. Conversely, in the CMCC-ESM2 model, QM demonstrated the highest RMSE values of 331 and 201.849 mm under SSP2-4.5 and SSP3-7.0, separately. This suggests either the potential inefficacy of QM method for correcting the precipitation biases, or it may reflect limitations within the CMCC-ESM2 model itself.

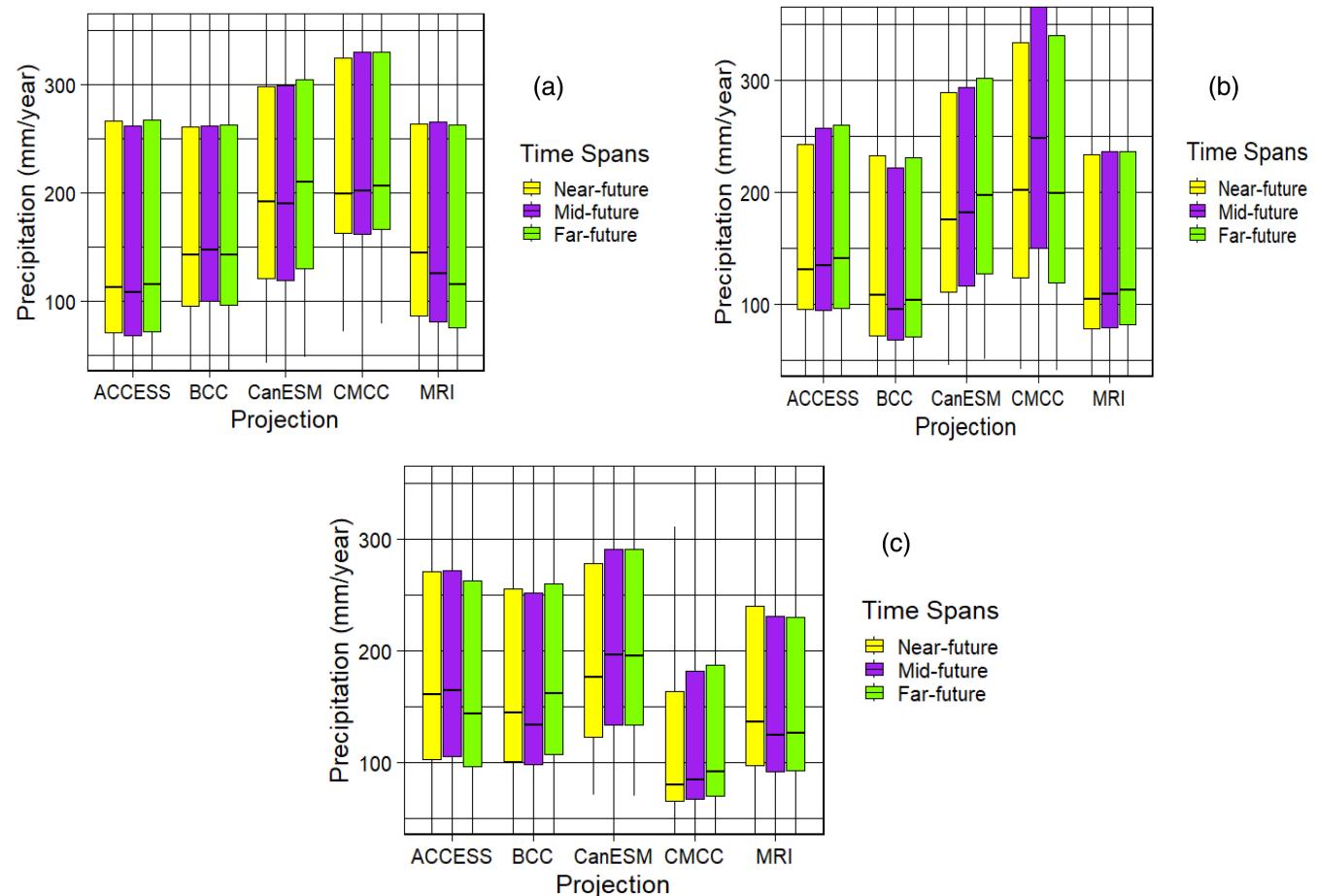


FIGURE 5 A series of boxplots for five CMIP6 models averaged in annual scale during three different future periods; near-future (2023–2047), mid-future (2048–2073) and far-future (2074–2100) under SSP2-4.5 (a), SSP3-7.0 (b) and SSP5-8.5 (c).

In the most extreme climate scenario (SSP5-8.5), QM effectively corrected biases in the CanESM5 model, yielding minimum errors and the highest correlation compared to the other models ($\text{RMSE} = 106 \text{ mm}$, $r = 0.91$) (refer to Table 4 for further details).

The statistical measurements of five CMIP6 models have been presented in Table 5, utilizing the QM-RF algorithm. Among the models, CanESM5 counted up the lowest error and CMCC-ESM2 had the highest RMSE values under SSP2-4.5. The remaining models, including ACCESS-ESM1-5, BCC-CSM2-MR and MRI-ESM2-0, displayed comparable performance with RMSE values falling within the range of 80–85 mm. In SSP3-7.0, although CanESM5 demonstrated superior performance in annual precipitation projection ($\text{RMSE} = 63.75$, $\text{Bias} = 2.83$ and $r = 0.97$), CMCC-ESM2 exhibited the highest error with the RMSE of 151.42 under the same scenario. Also, CanESM5 consistently outperformed other models with an RMSE of 66.13 under the SSP5-8.5.

The Bias values in both QM and QM-RF presented negative results for CanESM5 and CMCC-ESM2 under SSP2-4.5, signifying that both models may overestimate the future precipitation trend. Although both QM and QM-RF methods considered to minimize the systematic biases of model outputs, the combination of QM-RF presented a better performance in BC results. For example, the average RMSE of five CMIP6 models decreased by approximately 60, 40 and 84 mm in the scenarios of SSP2-4.5, SSP3-7.0 and SSP5-8.5, respectively, through the utilization of the bias-correction QM-RF method.

The boxplots in Figure 5a–c showcase the frequency of annual precipitation projections from five CMIP6 models under the SSP2-4.5, SSP3-7.0 and SSP5-8.5 scenarios. Specifically, in SSP2-4.5, both CanESM5 and CMCC-ESM2 projected a median of around $200 \text{ mm}\cdot\text{year}^{-1}$ across the near-, mid- and far-future timelines, while the other models projected values below $200 \text{ mm}\cdot\text{year}^{-1}$. Furthermore, under SSP2-4.5, it is projected that 75% of the overall country will

TABLE 6 Statistical summary for the total annual precipitation in baseline (1990–2020) and future periods (2023–2100).

MME summary (2023–2100)												
Time spans	SSP2-4.5				SSP3-7.0				SSP5-8.5			
	1st Q	Median	Mean	3rd Q	1st Q	Median	Mean	3rd Q	1st Q	Median	Mean	3rd Q
Near-future	107.50	156.95	232.07	282.42	96.01	143.82	221.50	265.66	97.89	139.10	208.00	241.26
Mid-future	107.00	155.90	232.20	287.41	97.00	144.82	221.60	267.66	98.10	139.24	209.47	243.00
Far-future	109.50	156.50	235.50	285.77	100.20	151.16	229.40	273.44	99.80	144.88	212.87	244.80
Baseline summary (1990–2020)												
1st Q	Median		Mean				3rd Q					
96.2	161.00		240.20				295.56					

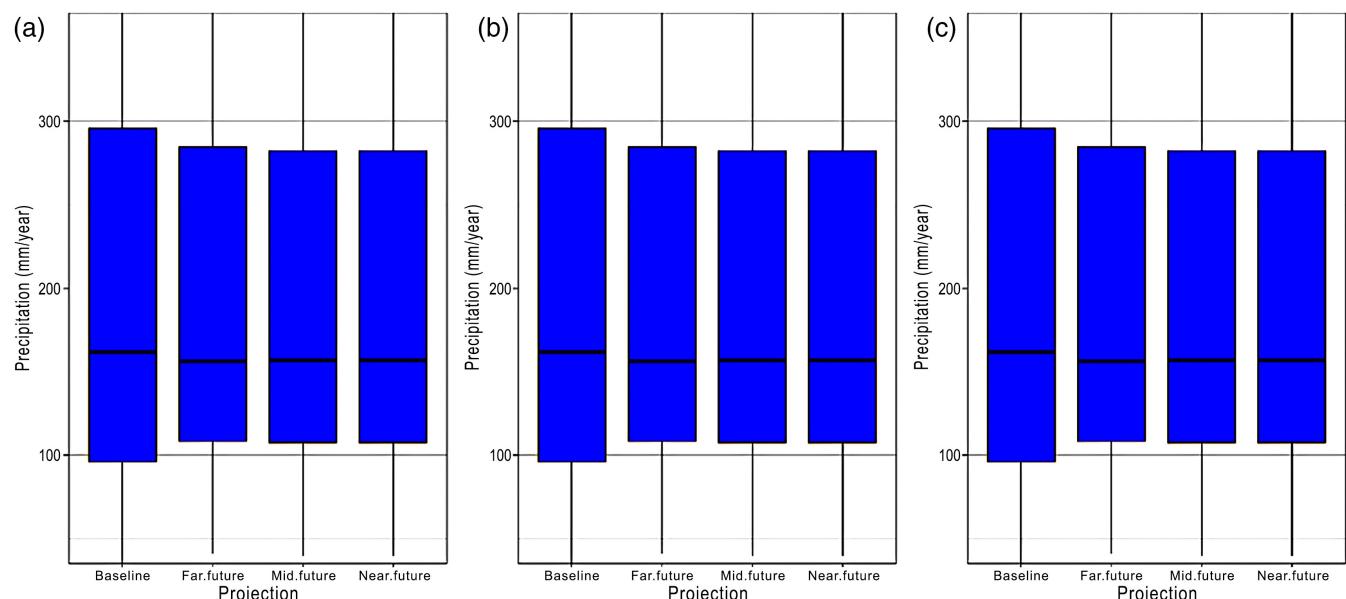


FIGURE 6 Boxplots visualization of annual mean precipitation in four time periods based on MME: baseline (1990–2020), near-future (2023–2047), mid-future (2048–2074) and far-future (2074–2100) under SSP2-4.5 (a), SSP3-7.0 (b) and SSP5-8.5 (c).

experience precipitation values below 260, 250, 300, 330 and 260 mm·year⁻¹ based on projection results of the ACCESS-ESM1-5, BCC-CSM2-MR, CanESM5, CMCC-ESM2 and MRI-ESM2-0, respectively (refer to Figures 5a and S1–S4 for detailed visual representation).

In both SSP2-4.5 (Figure 5a) and SSP3-7.0 (Figure 5b), the CMCC-ESM2 model revealed the highest amounts of precipitation across the three future intervals. However, in contrast to the findings for the CanESM5 model, CMCC-ESM2 indicated the lowest precipitation amounts in SSP5-8.5 (Figure 5c) implying a potential reduction in projected precipitation values for upcoming years. For further details, consult Figures S1–S4.

3.2.3 | Analysing MME outputs under three SSPs

The MME analysis revealed a decreasing trend in annual mean precipitation projections when compared to the baseline period (1990–2020), extending the end of 21st century (Table 6 and Figure 6). During the baseline period, the calculated mean and median precipitation values were 240.20, and 161 mm·year⁻¹, respectively; however, in all three SSP scenarios, lower mean and median precipitation values were observed for the near, mid- and far-future time spans (refer to Table 6 for detailed data).

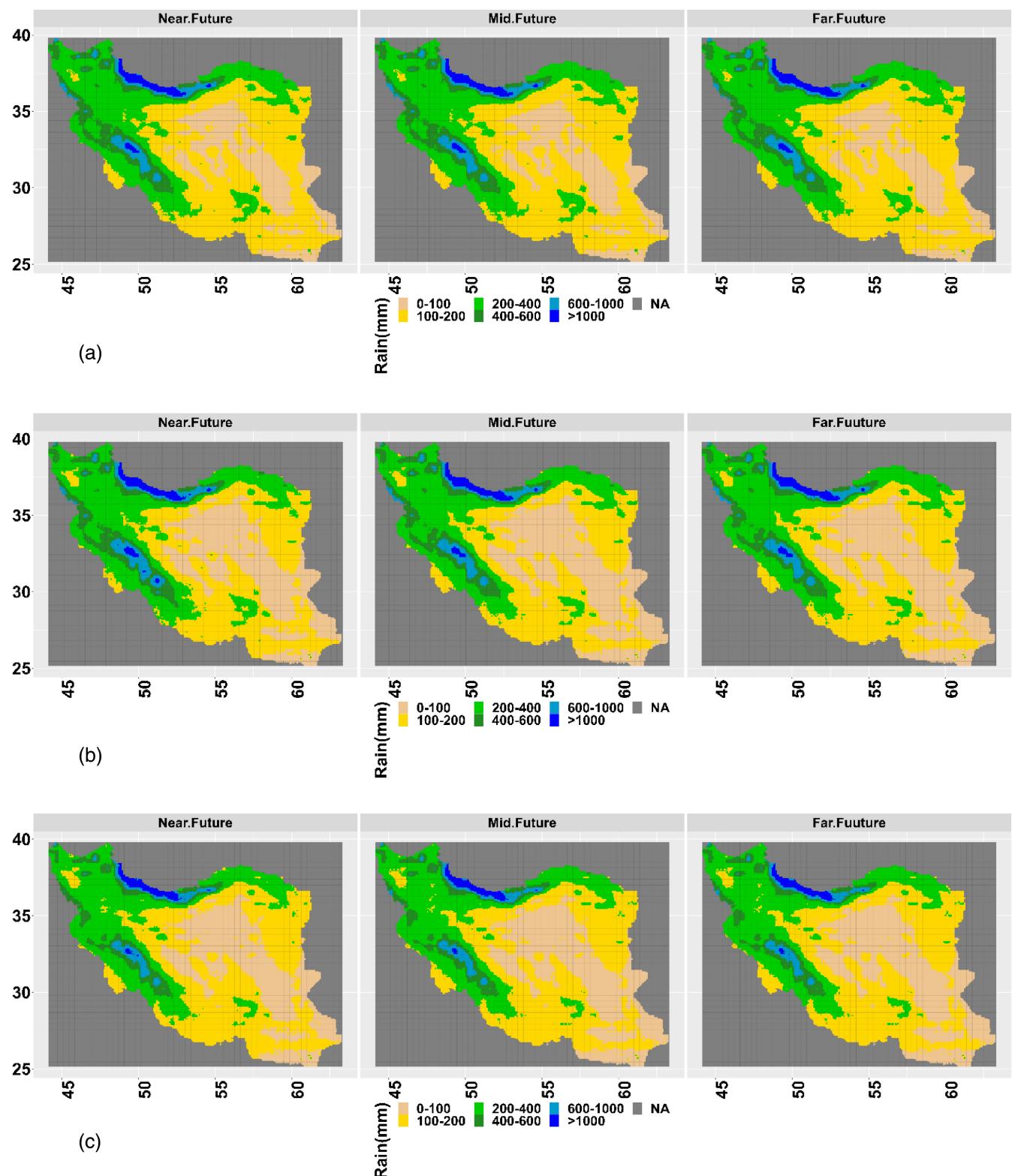


FIGURE 7 Classifying mean values of annual rainfall based on MME in three periods of time using QM-RF method; near-future (2023–2047), mid-future (2048–207) and far-future (2074–2100) under SSP2-4.5 (a), SSP3-7.0 (b) and SSP5-8.5 (c).

The reduction in mean precipitation values is more pronounced under the SSP5-8.5 scenario, decreasing from $232.07 \text{ mm}\cdot\text{year}^{-1}$ in SSP2.4.5 to $208 \text{ mm}\cdot\text{year}^{-1}$.

Specifically, the third quartile (Q_3) value during the baseline period was calculated to be $295.56 \text{ mm}\cdot\text{year}^{-1}$, indicating that only 25% of the annual mean precipitation

values exceeded this threshold. The model outputs quantified lower Q_3 compared to the baseline period, with the highest values recorded as $287.41 \text{ mm}\cdot\text{year}^{-1}$ under SSP2-4.5 in the mid-future and $273.44 \text{ mm}\cdot\text{year}^{-1}$ under SSP3-7.0 in the far-future time spans (Table 6 and Figure 5).

The distribution of the MME precipitation data demonstrates the degree of data clustering, facilitating comparisons across various SSP scenarios for the near-, mid- and far-future based on the baseline period spanning from 1990 to 2020. A visual representation of the mean annual precipitation data obviously follows a decreasing trend compares baseline period across each respective time interval (Figure 6).

The spatial distribution of projected annual precipitation was assessed under three distinct SSPs and three future periods using the MME approach over Iran (Figure 7). According to SSP2-4.5 (Figure 6a), the mean precipitation values predominantly range between 100 and $200 \text{ mm}\cdot\text{year}^{-1}$ in the central plateau, eastern and southern regions of Iran.

In the western (Zagros Mountain Range), northern and northwestern regions, the mean precipitation values vary from 200 to $600 \text{ mm}\cdot\text{year}^{-1}$ under SSP2-4.5, SSP3-7.0 and SSP5-8.5 scenarios. Notably, only a limited area along the Caspian Sea coastline, located in the north, and Zagros Mountain Range experiences annual mean precipitation exceeding $1000 \text{ mm}\cdot\text{year}^{-1}$. The MME analysis indicates a slightly decreasing trend in annual precipitation for the southeast and central plateau of Iran under SSP3-7.0 and SSP5-8.5. Moreover, the majority of Iran is projected to experience the annual precipitation lower than $100 \text{ mm}\cdot\text{year}^{-1}$ across the near-, mid- and far-future timeframes. It projected a reduction pattern in total precipitation which is sharper in SSP3-7.0 and SSP5-8.5, especially in central part, east and southeast of the study area.

A possible decline observed in mean annual precipitation in arid regions, for instance, the central plateau. Regarding SSP3-7.0 and SSP5-8.5, a reduction pattern is also evident in coastal areas of the Caspian Sea in the north and the western region comprising the Zagros Ranges. Based on MME analysis, each scenario projected a declining trend in mean annual precipitation across the entirety of Iran. For detailed comparisons between scenarios, refer to Figures S1–S3.

3.2.4 | The percentage of relative change for annual precipitation

In present study, the relative change (RC%) was computed for the MME to illustrate the changes in annual

precipitation compared to the baseline period. Negative and positive signs in the calculated RC values indicate decreasing and increasing patterns, respectively, in future projections (refer to Figure 8 for visual representation).

Under SSP2-4.5, a significant portion of Iran is projected to experience the RC falling within the -10 to 0 range in annual precipitation, depicting a decreasing trend (0 – 10%) until the end of 21st century. In the northern, northeastern, northwestern and central plateau regions, the annual precipitation is projected to decrease by 10% – 20% . Conversely, under SSP3-7.0, the southern areas are projected to witness an increase ranging between 10% and $>20\%$. The increasing trend is expected to exceed 20% in the central and southeastern parts of Iran. Under SSP5-8.5, the RC analysis presented notable variations, highlighting a decline in annual precipitation in various parts of the north, west, northwest and northeast regions, with reductions ranging from 10 to over 20% . Regarding GCMs, the extensive deviations are projected to emerge in SSP5-8.5 as it consists of the strongest forcing conditions (Mendoza Paz & Willems, 2022). Furthermore, an increase pattern exceeding 20% was observed in the central plateau, southeastern and southern regions according to each scenario (Figure 8). As these areas already experience annual precipitation below 150 mm , a 20% of increment may not either consider as a positive effect or improve environmental conditions significantly. Overall, the RC analysis indicates a declining trend in annual precipitation over Iran under SSP2-4.5, SSP3-7.0 and SSP5-8.5 scenarios (for further details, refer to Figures S5–S8).

4 | DISCUSSION

The future climate projections derived from the Coupled Model Intercomparison Project Phase 6 (CMIP6) serve as a crucial tool for climate research on both global and regional scales, considering that the performance of global climate models varies considerably. This study was conducted on five CMIP6 model outputs assessment and the application of BC methods to project annual precipitation patterns specifically over Iran. This evaluation has an advantage for improving the accuracy and reliability of climate projections in area under the study.

When projecting future precipitation, the stationarity assumption is critical and in climate modelling, it can be a more complex issue, because climate change can introduce nonstationary behaviour in meteorological variables like precipitation. In our study on projecting future precipitation, we have assumed weak or second-order stationarity in our time series data. The stationarity assumption between ground precipitation data and

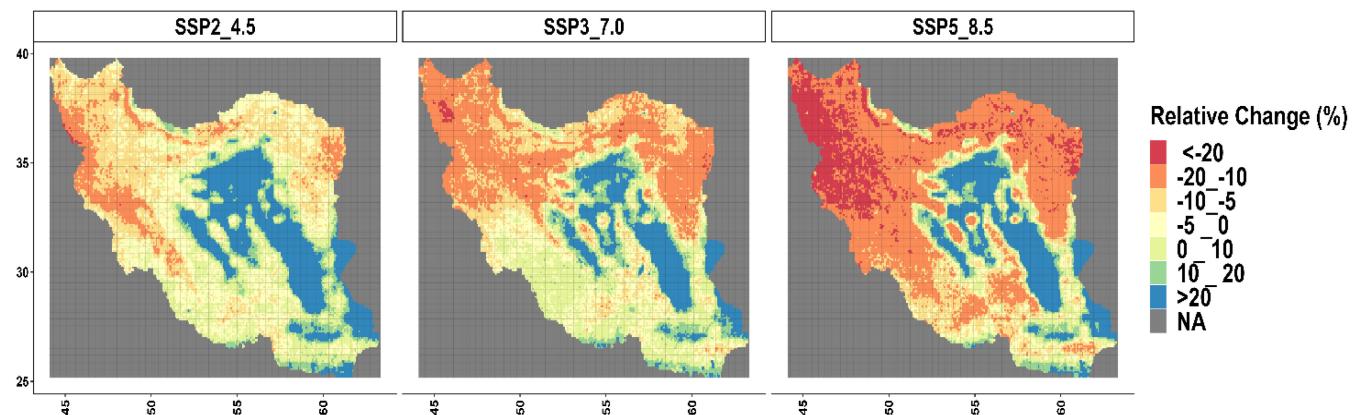


FIGURE 8 The relative changes (%) of MME mean annual precipitation under SSP2-4.5, SSP3-7.0 and SSP5-8.5.

ERA5-Land reanalysis data refers to the hypothesis that the statistical properties of precipitation data measured at ground stations remain constant over time and are comparable to the ERA5-Land reanalysis precipitation. This assumption implies that the distribution, mean, variance and other statistical characteristics of precipitation data observed at ground stations and those simulated by the reanalysis data are consistent and do not exhibit significant temporal variability (in the time series). This temporal variability could affect the accuracy of comparisons and/or analyses conducted between the ground and the ERA5-Land reanalysis datasets. Moreover, current investigation demonstrated that the statistical properties of precipitation in historical simulations analysis do not dramatically change over time and it implies that past trends can be used to project future precipitation. In other words, the stationarity assumption was primarily based on the analysis of historical precipitation simulations, which did not exhibit significant changes in mean and variance over the observational period. Thus, assuming stationarity might not always be valid. We employed the Pettitt test to verify the stationarity of our data, homogeneity or the presence of abrupt changes (point of change) in the annual precipitation time series (Nasrollahi et al., 2021), and the results supported the stationarity assumption.

Furthermore, for BC, it is essential to accurately establish the relationship between observed precipitation and its simulations. Although using QM method is a comprehensive used technique in adjusting precipitation projection, our findings demonstrate that integrating the QM method with the RF algorithm yielded more dependable results for BC of projected annual precipitation. Following the implementation of the QM method, the corrected climate projections exhibited a notable discrepancy when compared to the ERA5-Land reanalysis

precipitation data during the control period (1990–2020) in Iran. As a result, a combination of QM-RF approach was developed for future projections, by using QM bias-corrected outputs as inputs for creating the RF model. We acknowledge that climate change could introduce nonstationary behaviour in precipitation patterns, potentially affecting the accuracy of our projections. To mitigate the risks associated with nonstationarity, we compared our corrected projections with outputs from the bias-corrected ERA5-Land reanalysis data within an overlap of 6-year period (2015–2020) and test the accuracy of the applied bias-correction method (QM). The assessment findings indicated the QM-RF method was effective in reducing the systematic biases for precipitation projections in Iran. In essence, RF improved the BC procedure for the model outputs successfully.

Pang et al. (2017) reported that RF provided best results for downscaling temperature projections from GCMs. For instance, in SSP3-7.0, bias for CanESM5 reduces from 3.56 to 2.83, and RMSE had a dramatic decline of $35.49 \text{ mm}\cdot\text{year}^{-1}$ after using RF algorithm. According to the assessment, CMCC-ESM2 revealed a notable decrease in annual precipitation over Iran under SSP5-8.5. Conversely, CanESM5 demonstrated the most significant upward trend for the future period across various SSPs. Wang et al. (2023) also stated the majority of the CMIP6 models increase the total precipitation, which we encountered in projections before and after BC. Moreover, Almazroui et al. (2020) indicated that by the end of the 21st century, the CMIP6 simulations present much more precipitation over the South Asia. They pointed out the lowest annual mean precipitation occurs over several arid regions. According to the assessment, the outputs of BCC-CSM2-MR and MRI-ESM2-0 projected a decline in total precipitation in SSP5-8.5 significantly. It might be possible that some models perform

well in a specific climatic area with a determined precipitation pattern. In this case, ACCESS-ESM1-5 and BCC-CSM2-MR models indicated similar results to the control period in arid zones of Iran, suggesting less significant changes in precipitation patterns for areas. This is in agreement with previous findings in Iran centred on CMIP6 models (Yazdandoost et al., 2021). Although, BCC-CSM2-MR and MRI-ESM2-0 presented a decline in annual mean precipitation in coming years of the 21st century, CanESM5 and CMCC-ESM2 models projected an increase trend in central plateau, western and southern parts of Iran for near-, mid- and far-future spans under SSP3-7.0. Our results are in accordance with Farokhzadeh et al. (2022), highlighting the intensity of rainfall would increase in the central and the southern parts of Iran. Furthermore, Abdolalizadeh et al. (2023) applied NorESM2-MM in Urmia lake catchment, Iran, based on SSPs and QM method. They concluded the average rainfall in SSP5-8.5 will increase by 1.8% and 7.2% in the near- and far-future, respectively. According to Goodarzi et al. (2022), even though CanESM5 model projected an increase in rainfall between the years 2016 and 2100 under SSP5-8.5 in Gilan, Iran, the number of rainy days will drop dramatically. A study conducted on precipitation patterns also indicated that the future trends illustrated a considerable increase in precipitation between the years 2051 and 2080 under SSP1-2.6, SSP2-4.5 and SSP5-8.5 (Adib et al., 2022). According CMCC-ESM2 results, the highest projected values were obtained in both SSP2-4.5 and SSP3-7.0 ($>250 \text{ mm}\cdot\text{year}^{-1}$). While the lowest values were projected in SSP5-8.5 ($<200 \text{ mm}\cdot\text{year}^{-1}$).

Several studies have projected climate change impacts over Iran, pointed out that the maximum decrease in precipitation was projected for the end of the century (Babaeian et al., 2019). In addition, Yazdandoost et al. (2021) revealed a strong decrease in precipitation all over the rainiest zone in Iran. In general, the more/less spread in precipitation patterns over the central plateau of Iran, as an arid region, might be either associated with the topography or lack of ground data, especially, limited observations. Doulabian et al. (2021) revealed that the GCMs tend to over/underestimate climatic variables on regional and global scales, failing to deal with the micro-scale climate. In this case, considering the result of a single model may contains either a high level of uncertainty or variation in different climate conditions. Moreover, the relative errors vary substantially between locations (Mendoza Paz & Willems, 2022). In this study, MME revealed a total decline pattern in annual precipitation over Iran, in which covers the north, south, east and west regions under SSP2-4.5, SSP3-7.0 and SSP5-8.5. A number of study have confirmed that MEE projects a reduction in

precipitation trend as Seker and Gumus (2022) reported in the Mediterranean region. They created MME and downscaled outputs based on the artificial neural networks (ANN) statistical downscaling method and indicated a decrease in annual precipitation of 15% for SSP2-4.5 and 20% for SSP5-8.5. They further concluded MMEs are more reliable than individual GCMs in simulating historical temperature and precipitation. Zhang et al. (2022) investigated a decreasing trend in precipitation for Africa and Asia based on SSP5-8.5 by using CMIP6 ensemble. In this study, MME illustrated that annual precipitation has the greatest reduction based on SSP5-8.5 over the country, which is in parallel to a research carried out by Doulabian et al. (2021). They approved pessimistic scenario affirmed an extra decline in precipitation projection over Iran. To discuss the potential impact of the research assumption on our projections, climate change is expected to alter precipitation patterns based on SSP5-8.5 significantly, in which the obtained results might be less reliable. Conversely, the study area, Iran, presented limited signs of change in both historical simulations and under SSP2-4.5 projections, so that the stationarity assumption might be more defensible.

Although the MME projected a decline trend in overall precipitation, interestingly, it presented an increase in annual precipitation of 10%–20% in a confined area along the northern coastline under three SSP scenarios. This suggests a localized variation within the broader trend of declining precipitation. Our findings are in parallel with Hong et al. (2021), applied generalized extreme value distribution for 24 models and observations and concluded an increasing in precipitation relative change (%) are obvious in the northern parts of Iran. Even supposing ensemble projection has been progressed to express a powerful predictable signal, future studies should direct to minimize uncertainties in climate projections by performing high-resolution climate models (Alizadeh, 2022).

5 | CONCLUSION

This paper assessed the bias-corrected projections from five CMIP6 used for annual mean precipitation projection over Iran. Initially, QM was employed to rectify the systematic biases present in ERA5-Land reanalysis data and CMIP6 outputs. Subsequently, the combination of QM-RF method was utilized to enhance the results derived from QM. The findings indicated that QM exhibited substantial bias in projecting precipitation in Iran, a bias that was notably alleviated through the use of RF algorithm. Moreover, creating the MME outputs revealed a better understanding of precipitation projection and the

variability between individual climate models. Overall, results illustrated that Iran will encounter a more deficiency in annual precipitation over central and east regions under SSP3-7.0 and SSP5-8.5. Several key results can be acknowledged as follows: First, it recommended to apply QM in combination with RF to obtain better BC results in Iran, and it is recommended to use trend-preserving bias-correction methods like scaled distribution mapping (SDM) to minimize the systematic errors in model outputs (Eekhout & de Vente, 2019). Second, the future precipitation pattern highly depends on climate models; specifically, several GCMs/ESMs might either increasing or decreasing the precipitation projection over different SSPs, if they use individually. In the current study, the individual model exhibited variabilities and fluctuations in projecting the mean annual precipitation within the whole country, leading to uncertainties in future precipitation changes (refer to Figures S1–S5). The MME outcomes, on the other hand, displayed an expected trend in annual precipitation across all future time spans under the scenarios SSP2-4.5, SSP3-7.0 and SSP5-8.5 throughout Iran. It is evident that humid semi-arid regions are anticipated to experience a decline trend in total precipitation, whereas arid areas are projected to observe a slight increase in precipitation according to the MME analysis. These projected patterns align with the direction of observed alterations in the annual mean precipitation over Iran. Subsequent analysis and comparison following the application of the BC method can offer valuable insights into the capabilities and limitations of each model in projecting the annual precipitation. Moreover, superior understanding of precipitation projection requires scientific evidence to design further research and policy actions.

AUTHOR CONTRIBUTIONS

Maryam Raeesi: Conceptualization; investigation; writing – original draft; writing – review and editing; visualization; methodology; software; formal analysis; data curation. **Ali Asghar Zolfaghari:** Conceptualization; investigation; writing – original draft; writing – review and editing; validation; visualization; methodology; software; formal analysis; project administration; resources; supervision; data curation. **Seyed Hasan Kaboli:** Writing – review and editing; writing – original draft. **Mohammad Rahimi:** Writing – review and editing. **Joris de Vente:** Conceptualization; writing – review and editing; investigation; validation. **Joris P. C. Eekhout:** Conceptualization; investigation; writing – review and editing; validation.

CONFLICT OF INTEREST STATEMENT

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.

ORCID

Ali Asghar Zolfaghari  <https://orcid.org/0000-0001-7337-9849>

REFERENCES

- Abdalolizadeh, F., Khorshiddoust, M. & Jahanbakhsh, S. (2023) Projection and evaluation of the trend of temperature, precipitation and drought in Urmia lake catchment. *Hydrogeomorphology*, 10, 38–57.
- Adib, M.N.M., Harun, S. & Rowshon, M.K. (2022) Long-term rainfall projection based on CMIP6 scenarios for Kurau River basin of rice-growing irrigation scheme, Malaysia. *SN Applied Sciences*, 4, 70. Available from: <https://doi.org/10.1007/s42452-022-04952-x>
- Ahmadalipour, A., Rana, A., Moradkhani, H. & Sharma, A. (2017) Multi-criteria evaluation of CMIP5 GCMs for climate change impact analysis. *Theoretical and Applied Climatology*, 128, 71–87. Available from: <https://doi.org/10.1007/s00704-015-1695-4>
- Ahmed, K., Shahid, S. & Harun, S. (2015) Statistical downscaling of rainfall in an arid coastal region: a radial basis function neural network approach. *Applied Mechanics and Materials*, 735, 190–194. Available from: <https://doi.org/10.4028/www.scientific.net/AMM.735.190>
- Alijani, B., O'Brien, J. & Yarnal, B. (2008) Spatial analysis of precipitation intensity and concentration in Iran. *Theoretical and Applied Climatology*, 94, 107–124.
- Alizadeh, O. (2022) Advances and challenges in climate modeling. *Climatic Change*, 170, 18. Available from: <https://doi.org/10.1007/s10584-021-03298-4>
- Almazroui, M., Saeed, F., Saeed, S., Nazrul Islam, M., Ismail, M., Klutse, N.A.B. et al. (2020) Projected change in temperature and precipitation over Africa from CMIP6. *Earth Systems and Environment*, 4, 455–475. Available from: <https://doi.org/10.1007/s41748-020-00161-x>
- Amini, E., Zolfaghari, A., Kaboli, H. & Rahimi, M. (2022) Estimation of rainfall erosivity map in areas with limited number of rainfall station (case study: Semnan Province). *Iranian Journal of Soil and Water Research*, 53, 2027–2044. Available from: <https://doi.org/10.22059/ijswr.2022.343710.669279>
- Amjad, M., Yilmaz, M.T., Yucel, I. & Yilmaz, K.K. (2020) Performance evaluation of satellite-and model-based precipitation products over varying climate and complex topography. *Journal of Hydrology*, 584, 124707.
- Babaeian, I., Karimian, M., Modirian, R. & Mirzaei, E. (2019) Future climate change projection over Iran using CMIP5 data during 2020–2100. *Nivar*, 43, 62–71. Available from: <https://doi.org/10.30467/nivar.2019.142745.1103>
- Beyer, R., Krapp, M. & Manica, A. (2020) An empirical evaluation of bias correction methods for palaeoclimate simulations. *Climate of the Past*, 16, 1493–1508. Available from: <https://doi.org/10.5194/cp-16-1493-2020>
- Boé, J., Terray, L., Habets, F. & Martin, E. (2007) Statistical and dynamical downscaling of the Seine basin climate for

- hydro-meteorological studies. *International Journal of Climatology*, 27, 1643–1655. Available from: <https://doi.org/10.1002/joc.1602>
- Breiman, L. (2001) Random forests. *Machine Learning*, 45, 5–32. Available from: <https://doi.org/10.1023/A:1010933404324>
- Buda, S., Xiao, B., Zhu, D. & Tong, J. (2005) Trends in frequency of precipitation extremes in the Yangtze River basin, China: 1960–2003. *Hydrological Sciences Journal*, 50, 492.
- Chen, S.T., Yu, P.S. & Tang, Y.H. (2010) Statistical downscaling of daily precipitation using support vector machines and multivariate analysis. *Journal of Hydrology*, 385, 13–22. Available from: <https://doi.org/10.1016/j.jhydrol.2010.01.021>
- Chu, J.-L., Kang, H., Tam, C.-Y., Park, C.-K. & Chen, C.-T. (2008) Seasonal forecast for local precipitation over northern Taiwan using statistical downscaling. *Journal of Geophysical Research: Atmospheres*, 113, D12118. Available from: <https://doi.org/10.1029/2007JD009424>
- Dinpashoh, Y., Fakheri-Fard, A., Moghaddam, M., Jahanbakhsh, S. & Mirnia, M. (2004) Selection of variables for the purpose of regionalization of Iran's precipitation climate using multivariate methods. *Journal of Hydrology*, 297, 109–123.
- Doulabian, S., Golian, S., Toosi, A.S. & Murphy, C. (2021) Evaluating the effects of climate change on precipitation and temperature for Iran using RCP scenarios. *Journal of Water and Climate Change*, 12, 166–184. Available from: <https://doi.org/10.2166/wcc.2020.114>
- Ekhout, J.P.C. & de Vente, J. (2019) The implications of bias correction methods and climate model ensembles on soil erosion projections under climate change. *Earth Surface Processes and Landforms*, 44, 1137–1147. Available from: <https://doi.org/10.1002/esp.4563>
- Fan, X., Duan, Q., Shen, C., Wu, Y. & Xing, C. (2022) Evaluation of historical CMIP6 model simulations and future projections of temperature over the Pan-Third Pole region. *Environmental Science and Pollution Research*, 29, 26214–26229. Available from: <https://doi.org/10.1007/s11356-021-17474-7>
- Farokhzadeh, B., Bazrafshan, O., Singh, V.P., Choobeh, S. & Mohseni Saravi, M. (2022) Future rainfall erosivity over Iran based on CMIP5 climate models. *Water*, 14, 3861. Available from: <https://doi.org/10.3390/w14233861>
- Fowler, H.J. & Kilsby, C.G. (2007) Using regional climate model data to simulate historical and future river flows in northwest England. *Climatic Change*, 80, 337–367. Available from: <https://doi.org/10.1007/s10584-006-9117-3>
- Gleixner, S., Demissie, T. & Diro, G.T. (2020) Did ERA5 improve temperature and precipitation reanalysis over East Africa? *Atmosphere*, 11, 996. Available from: <https://doi.org/10.3390/atmos11090996>
- Goodarzi, M.R., Abedi, M.J. & Heydari Pour, M. (2022) Chapter 32 - Climate change and trend analysis of precipitation and temperature: a case study of Gilan, Iran. In: Zakwan, M., Wahid, A., Niazkar, M. & Chatterjee, U. (Eds.) *Current directions in water scarcity research*, vol. 7. Amsterdam, Netherlands. Elsevier, pp. 561–587. Available from: <https://doi.org/10.1016/B978-0-323-91910-4.00032-7>
- Heo, J.-H., Ahn, H., Shin, J.-Y., Kjeldsen, T.R. & Jeong, C. (2019) Probability distributions for a quantile mapping technique for a bias correction of precipitation data: a case study to precipitation data under climate change. *Water*, 11, 1475. Available from: <https://doi.org/10.3390/w11071475>
- Hong, J., Javan, K., Shin, Y. & Park, J.-S. (2021) Future projections and uncertainty assessment of precipitation extremes in Iran from the CMIP6 ensemble. *Atmosphere*, 12, 1052. Available from: <https://doi.org/10.3390/atmos12081052>
- Jeferson de Medeiros, F., Prestrelo de Oliveira, C. & Avila-Diaz, A. (2022) Evaluation of extreme precipitation climate indices and their projected changes for Brazil: from CMIP3 to CMIP6. *Weather and Climate Extremes*, 38, 100511. Available from: <https://doi.org/10.1016/j.wace.2022.100511>
- Jiang, J., Zhou, T., Chen, X. & Zhang, L. (2020) Future changes in precipitation over Central Asia based on CMIP6 projections. *Environmental Research Letters*, 15, 054009. Available from: <https://doi.org/10.1088/1748-9326/ab7d03>
- Jiang, Y., Yang, K., Shao, C., Zhou, X., Zhao, L., Chen, Y. et al. (2021) A downscaling approach for constructing high-resolution precipitation dataset over the Tibetan Plateau from ERA5 reanalysis. *Atmospheric Research*, 256, 105574. Available from: <https://doi.org/10.1016/j.atmosres.2021.105574>
- Kamruzzaman, M., Hwang, S., Choi, S.-K., Cho, J., Song, I., Jeong, H. et al. (2020a) Prediction of the effects of management practices on discharge and mineral nitrogen yield from paddy fields under future climate using APEX-paddy model. *Agricultural Water Management*, 241, 106345. Available from: <https://doi.org/10.1016/j.agwat.2020.106345>
- Kamruzzaman, M., Hwang, S., Choi, S.-K., Cho, J., Song, I., Song, J. et al. (2020b) Evaluating the impact of climate change on Paddy water balance using APEX-Paddy model. *Water*, 12, 1221. Available from: <https://doi.org/10.3390/w12030852>
- Kamruzzaman, M., Wahid, S., Shahid, S., Alam, E., Mainuddin, M., Islam, H.M.T. et al. (2023) Predicted changes in future precipitation and air temperature across Bangladesh using CMIP6 GCMs. *Heliyon*, 9, e16274. Available from: <https://doi.org/10.1016/j.heliyon.2023.e16274>
- Law, R.M., Ziehn, T., Matear, R.J., Lenton, A., Chamberlain, M.A., Stevens, L.E. et al. (2017) The carbon cycle in the Australian Community Climate And Earth System Simulator (ACCESS-ESM1)—part 1: model description and pre-industrial simulation. *Geoscientific Model Development*, 10, 2567–2590. Available from: <https://doi.org/10.5194/gmd-10-2567-2017>
- Li, H., Yu, C., Xia, J., Wang, Y., Zhu, J. & Zhang, P. (2019) A model output machine learning method for grid temperature forecasts in the Beijing area. *Advances in Atmospheric Sciences*, 36, 1156–1170. Available from: <https://doi.org/10.1007/s00376-019-9023-z>
- Liemohn, M.W., Shane, A.D., Azari, A.R., Petersen, A.K., Swiger, B.M. & Mukhopadhyay, A. (2021) RMSE is not enough: guidelines to robust data-model comparisons for magnetospheric physics. *Journal of Atmospheric and Solar-Terrestrial Physics*, 218, 105624. Available from: <https://doi.org/10.1016/j.jastp.2021.105624>
- Lovato, T., Peano, D., Butenschön, M., Materia, S., Iovino, D., Scoccimarro, E. et al. (2022) CMIP6 simulations with the CMCC earth system model (CMCC-ESM2). *Journal of Advances in Modeling Earth Systems*, 14, e2021MS002814. Available from: <https://doi.org/10.1029/2021MS002814>
- Mendoza Paz, S. & Willems, P. (2022) Uncovering the strengths and weaknesses of an ensemble of quantile mapping methods for downscaling precipitation change in southern Africa. *Journal of Hydrology: Regional Studies*, 41, 101104. Available from: <https://doi.org/10.1016/j.ejrh.2022.101104>

- Michelangeli, P.-A., Vrac, M. & Loukos, H. (2009) Probabilistic downscaling approaches: application to wind cumulative distribution functions. *Geophysical Research Letters*, 36, L11708. Available from: <https://doi.org/10.1029/2009GL038401>
- Muñoz-Sabater, J., Dutra, E., Agustí-Panareda, A., Albergel, C., Arduini, G., Balsamo, G. et al. (2021) ERA5-Land: a state-of-the-art global reanalysis dataset for land applications. *Earth System Science Data*, 13, 4349–4383. Available from: <https://doi.org/10.5194/essd-13-4349-2021>
- Nasrollahi, M., Zolfaghari, A.A. & Yazdani, M.R. (2021) Spatial and temporal properties of reference evapotranspiration and its related climatic parameters in the main agricultural regions of Iran. *Pure and Applied Geophysics*, 178, 4159–4179.
- O'Neill, B.C., Kriegler, E., Ebi, K.L., Kemp-Benedict, E., Riahi, K., Rothman, D.S. et al. (2017) The roads ahead: narratives for shared socioeconomic pathways describing world futures in the 21st century. *Global Environmental Change*, 42, 169–180. Available from: <https://doi.org/10.1016/j.gloenvcha.2015.01.004>
- Pang, B., Yue, J., Zhao, G. & Xu, Z. (2017) Statistical downscaling of temperature with the random forest model. *Advances in Meteorology*, 2017, 7265178. Available from: <https://doi.org/10.1155/2017/7265178>
- Peano, D., Lovato, T. & Materia, S. (2019) CMCC CMCC-ESM2-SR5 model output prepared for CMIP6 LS3MIP. <https://doi.org/10.22033/ESGF/CMIP6.1372>
- Piani, C., Weedon, G.P., Best, M., Gomes, S.M., Viterbo, P., Hagemann, S. et al. (2010) Statistical bias correction of global simulated daily precipitation and temperature for the application of hydrological models. *Journal of Hydrology*, 395, 199–215. Available from: <https://doi.org/10.1016/j.jhydrol.2010.10.024>
- Piao, J., Chen, W., Chen, S., Gong, H. & Wang, L. (2021) Mean states and future projections of precipitation over the monsoon transitional zone in China in CMIP5 and CMIP6 models. *Climatic Change*, 169, 35. Available from: <https://doi.org/10.1007/s10584-021-03286-8>
- Saddique, N., Khaliq, A. & Bernhofer, C. (2020) Trends in temperature and precipitation extremes in historical (1961–1990) and projected (2061–2090) periods in a data scarce mountain basin, northern Pakistan. *Stochastic Environmental Research and Risk Assessment*, 34, 1441–1455. Available from: <https://doi.org/10.1007/s00477-020-01829-6>
- Sadeghi, S.H., Zabihi, M., Vafakhah, M. & Hazbavi, Z. (2017) Spatiotemporal mapping of rainfall erosivity index for different return periods in Iran. *Natural Hazards*, 87, 35–56.
- Seker, M. & Gumus, V. (2022) Projection of temperature and precipitation in the Mediterranean region through multi-model ensemble from CMIP6. *Atmospheric Research*, 280, 106440. Available from: <https://doi.org/10.1016/j.atmosres.2022.106440>
- Shi, W., Schaller, N., MacLeod, D., Palmer, T.N. & Weisheimer, A. (2015) Impact of hindcast length on estimates of seasonal climate predictability. *Geophysical Research Letters*, 42, 1554–1559. Available from: <https://doi.org/10.1002/2014GL062829>
- Shiru, M.S., Shahid, S., Chung, E.-S., Alias, N. & Scherer, L. (2019) A MCDM-based framework for selection of general circulation models and projection of spatio-temporal rainfall changes: a case study of Nigeria. *Atmospheric Research*, 225, 1–16. Available from: <https://doi.org/10.1016/j.atmosres.2019.03.033>
- Swart, N., Cole, J., Kharin, V., Lazare, M., Scinocca, J., Gillett, N. et al. (2019) The Canadian Earth System Model version 5 (CanESM5.0.3). *Geoscientific Model Development*, 12, 4823–4873. Available from: <https://doi.org/10.5194/gmd-12-4823-2019>
- Talchhabhadel, R., Aryal, A., Kawaike, K., Yamanoi, K., Nakagawa, H., Bhatta, B. et al. (2021) Evaluation of precipitation elasticity using precipitation data from ground and satellite-based estimates and watershed modeling in Western Nepal. *Journal of Hydrology: Regional Studies*, 33, 100768. Available from: <https://doi.org/10.1016/j.ejrh.2020.100768>
- Tang, K., Zhu, H. & Ni, P. (2021) Spatial downscaling of land surface temperature over heterogeneous regions using random forest regression considering spatial features. *Remote Sensing*, 13, 3645. Available from: <https://doi.org/10.3390/rs13183645>
- Tong, X., Yan, Z., Xia, J. & Lou, X. (2019) Decisive atmospheric circulation indices for July–August precipitation in North China based on tree models. *Journal of Hydrometeorology*, 20, 1707–1720. Available from: <https://doi.org/10.1175/JHM-D-19-0045.1>
- Wang, G., He, Y., Zhang, B., Wang, X., Cheng, S., Xie, Y. et al. (2023) Historical evaluation and projection of precipitation phase changes in the cold season over the Tibetan Plateau based on CMIP6 multimodels. *Atmospheric Research*, 281, 106494. Available from: <https://doi.org/10.1016/j.atmosres.2022.106494>
- Wu, T., Lu, Y., Fang, Y., Xin, X., Li, L., Li, W. et al. (2019) The Beijing Climate Center Climate System Model (BCC-CSM): the main progress from CMIP5 to CMIP6. *Geoscientific Model Development*, 12, 1573–1600. Available from: <https://doi.org/10.5194/gmd-12-1573-2019>
- Yang, Y., Bai, L., Wang, B., Wu, J. & Fu, S. (2019) Reliability of the global climate models during 1961–1999 in arid and semiarid regions of China. *Science of the Total Environment*, 667, 271–286. Available from: <https://doi.org/10.1016/j.scitotenv.2019.02.188>
- Yazdandoost, F., Moradian, S., Izadi, A. & Aghakouchak, A. (2021) Evaluation of CMIP6 precipitation simulations across different climatic zones: uncertainty and model intercomparison. *Atmospheric Research*, 250, 105369. Available from: <https://doi.org/10.1016/j.atmosres.2020.105369>
- Yukimoto, S., Koshiro, T., Kawai, H., Oshima, N., Yoshida, K., Urakawa, S. et al. (2019) MRI MRI-ESM2.0 model output prepared for CMIP6 CMIP. <https://doi.org/10.22033/ESGF/CMIP6.621>
- Zhang, P., Lu, J. & Chen, X. (2022) Machine-learning ensembled CMIP6 projection reveals socio-economic pathways will aggravate global warming and precipitation extreme. *Hydrology and Earth System Sciences Discussions*, 1–40. <https://doi.org/10.5194/hess-2022-235>
- Zhao, P. & He, Z. (2022) A first evaluation of ERA5-Land reanalysis temperature product over the Chinese Qilian Mountains. *Frontiers in Earth Science*, 10, 907730. Available from: <https://doi.org/10.3389/feart.2022.907730>
- Zhu, L., Kang, W., Li, W., Luo, J.-J. & Zhu, Y. (2022) The optimal bias correction for daily extreme precipitation indices over the Yangtze-Huaihe River basin, insight from BCC-CSM1.1-m. *Atmospheric Research*, 271, 106101. Available from: <https://doi.org/10.1016/j.atmosres.2022.106101>
- Zolfaghari, A.A., Taghizadeh-Mehrjardi, R., Moshki, A.R., Malone, B.P., Weldeyohannes, A.O., Sarmadian, F. et al. (2016) Using the nonparametric k-nearest neighbor approach for predicting cation exchange capacity. *Geoderma*, 265, 111–119. Available from: <https://doi.org/10.1016/j.geoderma.2015.11.012>

Zou, J., Lu, N., Jiang, H., Qin, J., Yao, L., Xin, Y. et al. (2022) Performance of air temperature from ERA5-Land reanalysis in coastal urban agglomeration of Southeast China. *Science of the Total Environment*, 828, 154459. Available from: <https://doi.org/10.1016/j.scitotenv.2022.154459>

SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

How to cite this article: Raeesi, M., Zolfaghari, A. A., Kaboli, S. H., Rahimi, M., de Vente, J., & Eekhout, J. P. C. (2024). Using quantile mapping and random forest for bias-correction of high-resolution reanalysis precipitation data and CMIP6 climate projections over Iran. *International Journal of Climatology*, 44(12), 4495–4514. <https://doi.org/10.1002/joc.8593>