



# Performance of bias corrected monthly CMIP6 climate projections with different reference period data in Turkey

Sertac Oruc<sup>1</sup>

Received: 3 November 2021 / Accepted: 12 January 2022 / Published online: 3 February 2022  
© The Author(s) under exclusive licence to Institute of Geophysics, Polish Academy of Sciences & Polish Academy of Sciences 2022

## Abstract

Decisions that are based on the future climate data, and its consequences are significantly important for many sectors such as water, agriculture, built environment, however, the performance of model outputs have direct influence on the accuracy of these decisions. This study has focused on the performance of three bias correction methods, Delta, Quantile Mapping (QM) and Empirical Quantile Mapping (EQM) with two reference data sets (ERA and station-based observations) of precipitation for 5 single CMIP6 GCM models (ACCESS-CM2, CNRM-CM6-1-HR, GFDL-ESM4, MIROC6, MRI-ESM2-0) and ensemble mean approach over Turkey. Performance of model-bias correction method-reference data set combinations was assessed on monthly basis for every single station and regionally. It was shown that performance of GCM models mostly affected by the region and the reference data set. Bias correction methods were not detected as effective as the reference data set over the performance. Moreover, Delta method outperformed among the other bias correction techniques for the computation that used observation as reference data while the difference between bias correction methods was not significant for the ERA-based computations. Besides ensemble approach, MIROC6 and MRI-ESM2-0 models were selected as the best performing models over the region. In addition, selection of the reference data sets also found to be a dominant factor for the prediction accuracy, 65% of the consistent performance at the stations achieved by the ERA reference used bias correction approaches.

**Keywords** CMIP6 · Bias correction · GCM · ERA · Climate change

## Introduction

Climate change is a fact now both globally and regionally, and this is evident with intensification in the hydrological cycle, increasing temperature, changing trends among others. This has impacted many sectors including engineering design and agricultural management (Ombadi et al. 2018). Moreover, these hydrological variables such as temperature, precipitation, evaporation are time and space dependent (Taylan and Aydin, 2018), and this makes prediction or modeling a complex phenomenon. Making precise decisions for the future climate and its consequences are directly

related with the accuracy of model outputs. Global Climate Models (GCMs) that are the basic source of information for climate change and climate change induced impact assessment. GCMs provides information of climate change both for local and global scale, yet it is rarely preferred to directly use this information because of the large errors in GCM simulations relative to historical observations and the coarse spatial resolution while impact studies require generally finer scales (Ramirez-Villegas et al. 2013; Navarro-Racines et al. 2015).

The 6th phase of Coupled Model Intercomparison Project (CMIP); CMIP6, promises certain developments when compared to CMIP5, its predecessor. The reason behind these developments is the quantification of the natural or human induced radiative forcing, incorporation of the aerosol and land-use effects (Eyring et al. 2016; Stouffer et al. 2017; Liu et al. 2020; Wyser et al. 2020; Bağcaci et al. 2021). Although CMIP 6 data are relatively new, there is an increasing interest regarding the application of newest projections (Checa-Garcia et al. 2018; Fu et al. 2020; Luo et al. 2020; Lin and Chen 2020; Monerie et al. 2020; Ngoma et al. 2021). While

---

Communicated by Dr. Mohammad Valipour (ASSOCIATE EDITOR).

✉ Sertac Oruc  
sertac.oruc@ahievran.edu.tr

<sup>1</sup> Faculty of Engineering and Architecture, Department of Civil Engineering, Kırşehir Ahi Evran University, 40100 Kırşehir, Turkey

these studies revealed different results considering the performance between the CMIP6 and the CMIP5 results over their study regions, Bağcaci et al. (2021) indicated that CMIP6 products outperform CMIP5 in terms of accuracy of precipitation statistics particularly that emerges the need to update the results of impact studies with the most recent data for Turkey.

Future climatic conditions have significant contribution to the knowledge, but acquiring this information is sophisticated, and the most up-to-date instrument for examining the future climate is Global Circulation Models (GCMs). Climate change projections become more relevant and meaningful when their reference simulations are taken into account however, large uncertainties may exist in the model simulations, and focusing just on the changes may lead to the ignorance of the systematic errors (Nissan et al. 2019; Goldenson et al. 2021; Jose and Dwarakish 2022). Consequently the inclusion of reference period into the analyses is vital, especially if more than one model is employed, to stabilize the various models' predictions and applications such as bias correction or downscaling, which use historical observation to calibrate model findings, are used in this respect to generate better-fit climate projections (Chen et al. 2013; Stellingwerf et al. 2021; Hosseinzadehtalaei et al. 2021).

The multi-model ensemble approach is a simpler method; however, taking an average from the climate model often hides the model-related uncertainties of future projections and gives a false impression of increased confidence in projection. However, these modeling systems cannot determine the sub-grid-based processes (topography, land use, land–water boundary) in detail to obtain the required output for smaller scales such as regional or local (Kara et al. 2016; Kundzewicz et al. 2014, 2016). To address this problem, dynamical or statistical downscaling and/or bias correction techniques are needed to improve GCM-derived hydrometeorological outputs. The bias correction technique is mostly applied to remove the systematic biases inherited by the climate models. As climate models often are at a coarse resolution and can't represent the sub-grid processes, downscaling methods (dynamical or statistical) are employed. Considering the improving GCM output performance regionally or locally, statistical downscaling methods and their implementations related to hydrological impact studies have been subject of interest in many research (Sunyer et al. 2015). Among these methods, Delta, Quantile Mapping (QM) and Empirical Quantile Mapping (EQM) methods were selected to downscale 5 CMIP6 GCM models at the city central stations in Turkey which is situated mostly on the Anatolian plateau in Western Asia, and is bordered on the west by the Aegean Sea, on the north by the Black Sea, and on the south by the Mediterranean Sea.

The aim of this study is to investigate the effects of different bias correction techniques over the GCM outputs and to

explore the best model and bias correction techniques for the stations by using observation data sets over Turkey. The station data used for bias correction for a point-grid correction while ERA5 data is used for 'grid-grid' correction with GCM results, and both corrected GCM results were compared with observed point data.

This paper is organized as follows. “Materials and methods” section introduces the study area, station list, data source and reviews the methodology, in particular the statistical downscaling (Delta, the Quantile Mapping (QM), the Empirical Quantile Mapping (EQM)) methods and performance criteria. “Results and discussion” section reveals the results of bias correction methods and their performance, GCM performance, reference data set performance and their relationship over the study region and provides a discussion. “Conclusions and remarks” section provides final remarks.

## Materials and methods

### Study area

Turkey is situated mostly on the Anatolian plateau which is in Western Asia. The coordinates of the country are 36° and 42° N and 26° and 45° E. Turkey has borders with the Mediterranean Sea, the Aegean Sea and the Black Sea in the South, West and North, respectively. The total land surface area is 783,562 km<sup>2</sup>. In general, it has a mild Mediterranean climate, however, because of complex topography, mountain positions, sea effect, etc., different climate conditions arise. The coastal areas experience milder climates, on the other hand mountains like Taurus and North Anatolian lay parallel to the sea and block the sea effects to pass into the inland areas (Atalay et al. 2008; Seventh National Communication of Turkey Under The UNFCCC 2018). This induces limited precipitation and continental climate conditions for the inland area with cold winters and dry-hot summers (Amjad et al. 2020). The eastern coast of the Black Sea region has the highest amount of rainfall, with annual rainfall over 2200 mm, while Central Anatolia has an annual rainfall around 400 mm in Turkey.

### Datasets

Figure 1. shows the meteorological stations of the city centers that used in this study. The study stations were selected from the central city stations to obtain the maximum length of continuous data. The daily and monthly precipitation data at 81 meteorological stations for the 1961 and 2020 period are obtained from the Turkish State Meteorological Service (TSMS) however, because of the missing continuous record at some of the stations and historical projections of the GCMs end in the year of 2014, 1965–2014 period is

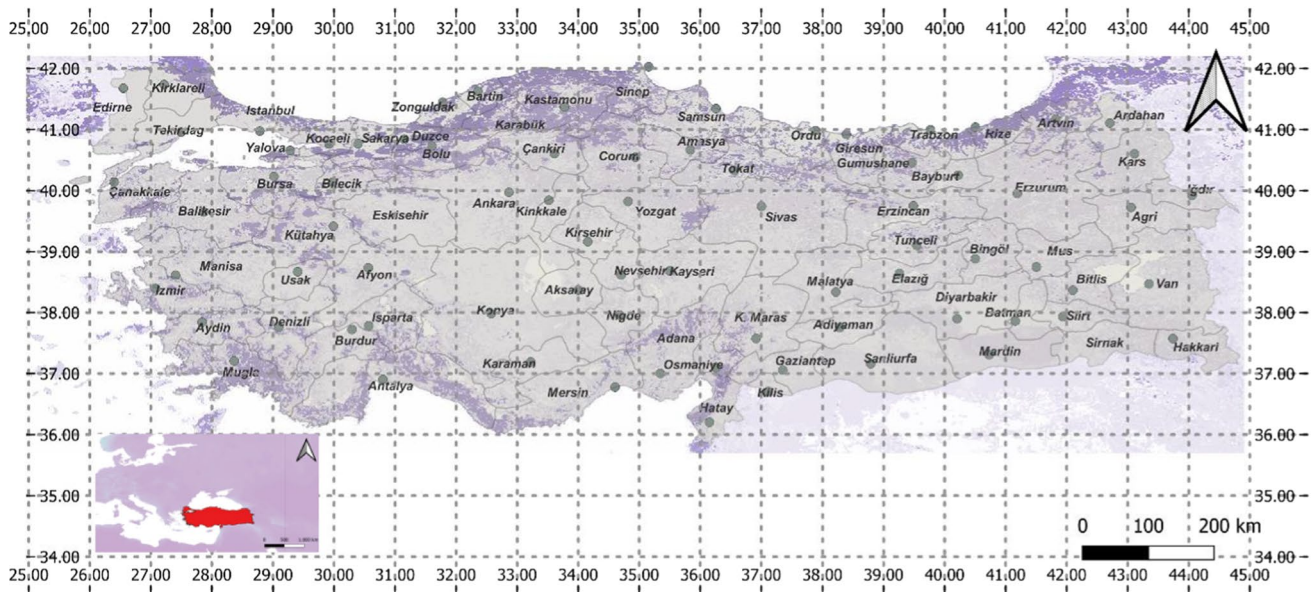


Fig. 1 Central observation stations

chosen for the analyses and Eskisehir, Karabuk and Sirnak stations were excluded because of the missing data or insufficient data length therefore 78 meteorological stations were used for the analyses. Furthermore, ERA5 monthly averaged precipitation data on single levels from 1979 to present was also used as reference data set for the bias correction process (Hersbach et al. 2020; CDS 2021; Muñoz-Sabater et al. 2021). ERA5 is ECMWF's fifth generation reanalysis of global climate and weather over the last 4 to 7 decades and it is currently available since 1950 (final release and timely updates after 1979) which replaced the ERA interim (CDS 2021).

Here we use monthly precipitation outputs of 5 CMIP6 models (Eyring et al. 2016; O'Neill et al. 2016) that are available in the Copernicus web site. The logic behind selecting these specific 5 GCMs is three of these models ranked in top 5, and all 5 models were ranked in top ten outperformed CMIP6 models for precipitation in Bağçaci et al. (2021) which considers 1045 meteorological observation stations over Turkey. The names of the model, corresponding

institution and horizontal resolution of the CMIP6 models are given in Table 1.

### Efficiency criteria

To assess the relative performance of the methods against observation data, combination of two metrics; Pearson's correlation coefficient ( $r$ ) and the index of agreement ( $d$ ) were computed on a monthly scale to explore the model performance, reference data sets and the bias correction methods.

First, the Pearson  $r$  coefficient was computed to identify the agreement between the modeled and the observed data. Moreover, the Index of Agreement ( $d$ ) (Willmott 1981) that varies between 0 and 1, was used to measure the degree of model prediction error. A value of 1 indicates an exact agreement (perfect fit) while a value of 0 means no match (no correlation).

The index of agreement “ $d$ ”, is the ratio of the mean square error and the potential error, calculated by Eq. (2) where  $O$  and  $P$  represent the observed and the predicted

Table 1 CMIP6 models used in the study

No	Model	Institution	Resolution
1	ACCESS-CM2	Commonwealth Scientific and Industrial Research Organization and Bureau of Meteorology, Australia	250 km
2	CNRM-CM6-1-HR	National Center for Meteorological Research, Météo-France and CNRS laboratory, France	250 km
3	GFDL-ESM4	NOAA/ Geophysical Fluid Dynamics Laboratory, USA	100 km
4	MIROC6	The University of Tokyo), National Institute for Environmental Studies, and Japan Agency for Marine-Earth Science and Technology, Japan	250 km
5	MRI-ESM2-0	Meteorological Research Institute, Tsukuba, Ibaraki 305–0052, Japan	100 km

values. More details of these well-known metrics can be found in Pereira et al. (2018), Willmott et al. (2012), Ullah et al. (2018).

$$r = \left( \frac{\sum_{i=1}^n (O_i - O) \times \sum_{i=1}^n (M_i - M)}{\sqrt{\sum_{i=1}^n (O_i - O)^2 \times \sum_{i=1}^n (M_i - M)^2}} \right) \quad (1)$$

$$d = 1 - \left( \frac{\sum_{i=1}^n (P_i - O_i)^2}{\sum_{i=1}^n (|P_i - O| + |O_i - O|)} \right) \quad (2)$$

### Statistical downscaling

There are many downscaling methods that can be encountered in the literature. In this study the simple Delta approach and the CDF matching based Quantile Mapping (QM) (Panofsky and Briar 1968) and Empirical Quantile Mapping (EQM) (Boe et al. 2007; Wetterhall et al. 2012) methods were used.

To summarize the process, observed data cumulative distribution function ( $CDF_{obs}$ ) and model data cumulative distribution function ( $CDF_{mod}$ ) were constructed by the QM method. Then a transfer function ( $h$ ) is derived from the two CDFs, and by applying, the transfer function model data are converted to unbiased by equaling the model cumulative frequency to the observed cumulative frequency, so its new distribution equals the distribution of the observation (Wuthiwongyothin et al. 2020). This transformation can in general be formulated as Eq. (3) where  $P_o$  and  $P_m$  denote observed and modeled precipitation, respectively, and  $F_m$  is the CDF of  $P_m$ ,  $F_o^{-1}$  is the inverse CDF (or quantile function) corresponding to  $P_o$  (Piani et al. 2010; Gudmundsson et al. 2012).

$$P_o = h(P_m) \quad (3)$$

$$P_o = F_o^{-1}(F_m(P_m)) \quad (4)$$

In addition, instead of assuming parametric distributions to solve Eq. (3), empirical quantiles are also a common approach to solve Eq. (3) by using the empirical CDF of observed and modeled values and for more details about this procedure refer to Wood et al. (2004), Boe et al. (2007), Wetterhall et al. (2012), Themeßl et al. (2011, 2012), Gudmundsson 2014, Gudmundsson et al. (2012).

Furthermore, the Delta method of SD approaches, which is relatively simple compared with the other methods was also used (Maraun et al. 2010; Themeßl et al. 2012; Salehnia et al. 2019; Beyer et al. 2020). Also called direct method; with modest data requirements relatively simple and easy to

use delta-approach is capable of incorporating the change signal of GCM to observed precipitation and obtain a new set of data (Hay et al. 2000; Wetterhall et al. 2012; Salehnia et al. 2019). In this study, Eq. (5) is used to calculate the downscaled precipitation data regarding the delta method.

$$P = P_m \times \left( \frac{P_o}{P_m} \right)_{\text{month}} \quad (5)$$

In this equation, where  $P$  is downscaled precipitation data,  $P_o$  indicates the mean observed precipitation,  $P_m$  is the mean precipitation data over the GCM historical run. Computations were performed by SD GCM tool (Agrimetsoft 2018) and R software (R Core Team 2021).

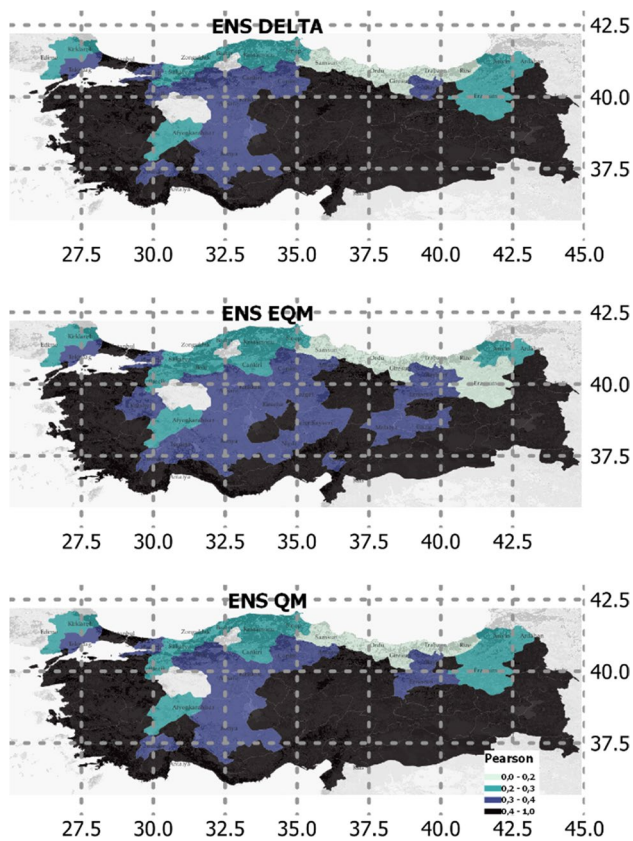
### Results and discussion

In this study, statistical indices of  $r$  and  $d$  were calculated for the bias corrected precipitation values based on observed and ERA precipitation reference data with three bias correction methods, namely, Delta, QM and EQM for the 78 meteorological stations in Turkey. The value of Pearson's correlation coefficient ( $r$ ) varies between 1 and -1, with +1, 0, and -1 indicating complete positive, uncorrelated, and completely negative correlations, respectively. Classification maps regarding the ( $r$ ) coefficient (0.0-0.2; 0.2-0.3; 0.3-0.4; and  $\geq 0.4$ ) were created in QGIS software for easy and accurate interpretation.

Moreover,  $d$  (index of agreement) values that are closer to 1.00 accepted as higher agreement of the bias corrected results. These analyses were conducted at monthly time scale for single (5 models) and ensemble model approach.

Figure 2 shows the Pearson coefficient distribution with ensemble mean for three methods for the study area. These results show that there are both overperformed and underperformed agreement between the station-observed precipitation data and the bias corrected precipitation data. The Eastern regions, together with Aegean and Mediterranean coasts of Turkey showed a better performance for the bias corrected monthly precipitation for all methods. However, Delta and QM methods exhibited a better performance compared with the EQM approach. Furthermore, all three methods with ensemble model approach showed a weaker performance for the Black Sea region, especially Middle and Eastern black sea regions showed no sign of considerable agreement. Besides, Central Anatolia exhibited a better performance compared with the Black Sea coasts and the Thrace part of the Marmara region. The  $r$  values of all regions were also calculated and given in Table 2. As shown in Table 2, the average  $r$  values ranged from 0.27 to 0.60 for Delta; from 0.25 to 0.52 for EQM and ranged from 0.26 to 0.59 for QM methods regionally and Delta 0.043-0.64; EQM





**Fig. 2** Pearson coefficient ( $r$ ) distribution with ensemble GCM mean with station-based reference data from 1965 to 2014 for three bias correction methods

**Table 2** Regional average  $r$  values for bias correction methods

Region	Delta	EQM	QM
Aegean Average	0.52	0.47	0.51
Black Sea Average	0.27	0.25	0.26
Central Anatolia Average	0.44	0.37	0.42
Eastern Anatolia Average	0.49	0.43	0.49
Marmara Average	0.38	0.34	0.37
Mediterranean Average	0.52	0.44	0.51
Southeastern Anatolia Average	0.60	0.52	0.59
All stations min–max range	0.043–0.64	0.055–0.58	0.04–0.63

0.055–0.58; QM 0.04–0.63 for 78 single stations. According to the ensemble analyses results, bias corrected precipitation data are consistent with the station-observed precipitation data for Eastern Anatolian, Aegean and Mediterranean coasts for Turkey.

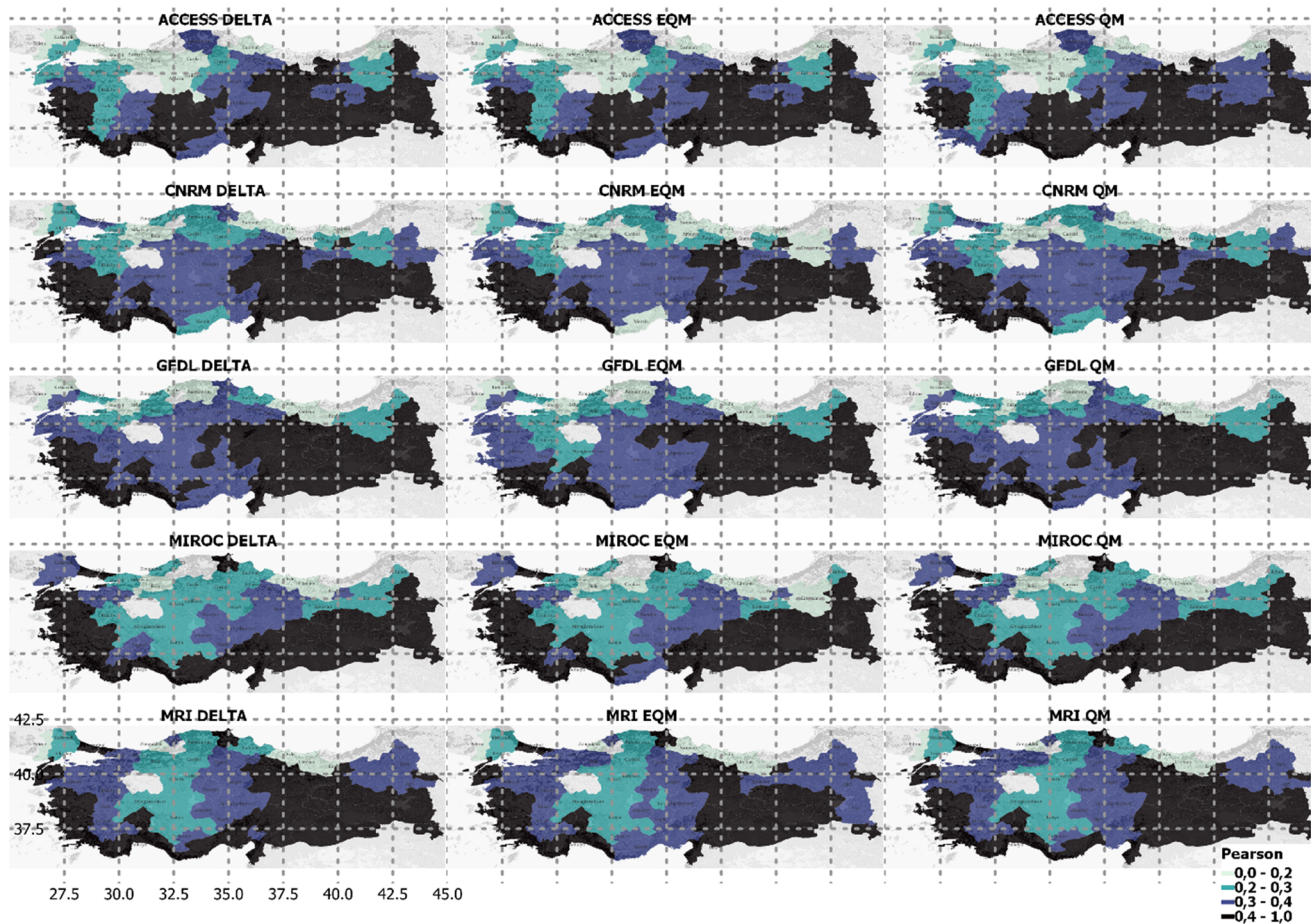
Furthermore, the analyses were conducted for every single model and bias correction method by using both observed station based and gridded data sets as reference period. In this approach, it was also possible to explore the model and method performance separately. In Fig. 3, the

result of the ERA-based reference analyses, and in Fig. 4, the results of observation-based reference analyses can be seen. The bias corrected data  $r$  values showed similar results for the single model evaluation with ERA reference data. On the other hand, the performance of ensemble model approach was generally better than the single model evaluation.

It can be seen from Table 3 that 50, 37 and 47 of 78 stations exhibited consistent results with the ensemble mean approach by using Delta, EQM and QM methods. Among these methods, Delta method exhibited a better performance while QM method also revealed similar results. On the other hand, EQM method did not perform well with the ensemble approach compared to others.

Further, single model performance also compared with the ensemble mean approach. Model, bias correction method and reference data set performance can be seen separately in Figs. 3 and 4 and Tables 4 and 5. Ensemble approach showed a significant better performance compared to single model approach for every reference data set (ERA and observation) and bias correction method (Delta, EQM, QM) and ensemble approach outperformed and revealed better correlation results. For instance, ACCESS model revealed better performance at the 33 stations for Era and 20 stations for observed reference data set for the Delta method. Furthermore, the model also exhibited better performance at the 33 and 32 stations for Era and 15 and 14 stations for observed reference data set for the EQM and QM methods, respectively. Besides single and ensemble model approach comparison, there is also a significant difference determined between two reference data sets, namely, Era and observation. For all bias correction methods, Era-based bias corrected data outperformed the bias corrected data that used observation data as reference for the ACCESS model. ERA5 gridded observation data can probably better represent the region while it is not possible to find a fully representation of the area with a single or few stations as Haerter et al. (2015) mentioned.

Considering the CNRM model, it is possible to observe the behavior of the previous model (ACCESS) in this one as well. Total number of satisfactory results among the 78 stations are two times of the observation data-based bias correction results for the Era-based results. When these results of two models were examined, it can be concluded that bias correction method is less effective over the results of Era-based approach while Delta method showed better performance for the observation-based bias correction results. GFDL model results showed the significant performance of Era compared to observation-based results again. Regarding this model, Delta approach can be pronounced as better than EQM and QM for the observation reference set-based results while Delta and QM revealed the same performance for the Era reference-based results.



**Fig. 3** Pearson coefficient ( $r$ ) distribution with ERA5 (reanalysis) reference data from 1980 to 2014 for three bias correction methods and five GCMs

MIROC and MRI results exhibited also similar results in terms of reference data set performance. These two models also revealed the highest total number of satisfactory results (164 and 163 in Table 4) for the study area. When reference data sets analyzed separately, MIROC outperformed with a total number of 112 for Era and MRI outperformed with a total number of 64 for observation-based bias correction results. MIROC model results showed that this model also outperformed for the Delta, EQM and QM bias correction methods with the Era-based reference data while observation-based reference data used bias correction methods exhibited better performance for the MRI model.

Performance of the models, bias correction methods and reference data sets were also investigated in a regional scale. IoA metric was used for this purpose.

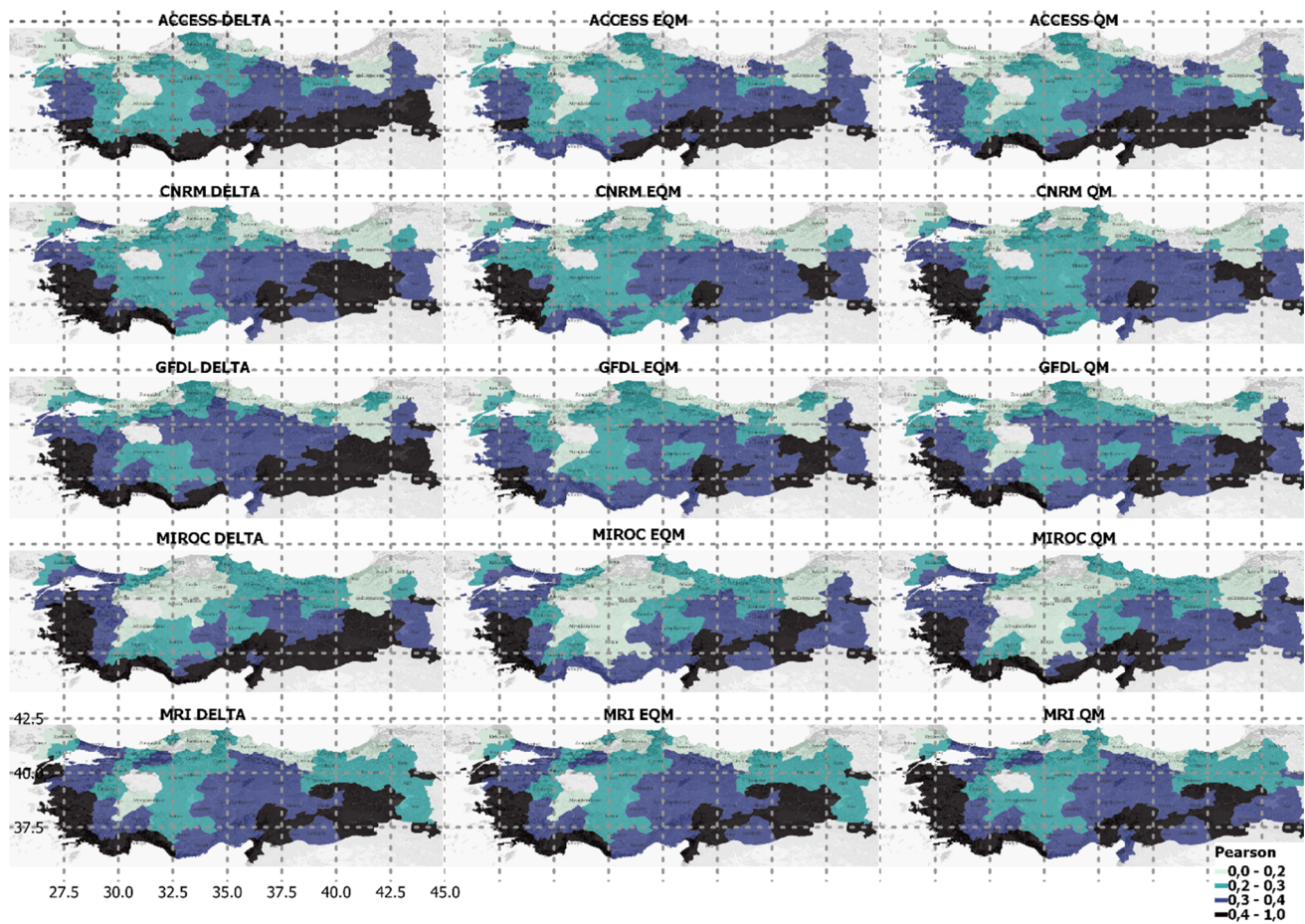
Index of agreement (IoA) values derived from ensemble model approach in Table 6, and the Era and observed precipitation data are shown in Tables 7 and 8. To obtain and reflect the regional results, average of stations in the region were computed.

Considering the ensemble mean approach, Southeastern, Mediterranean and Aegean Regions presented the highest model accuracy among the other regions for all bias correction methods. These regions were followed by Eastern and Central Anatolia. Furthermore, Quantile mapping bias correction method was found to be the outperformed method for all regions for the ensemble mean approach.

The IoA values revealed higher values for the Southeastern Anatolia and smaller values for the Black Sea regions among all the 7 geographical regions for all bias correction methods and reference data sets. Mediterranean, Aegean and Eastern Anatolian regions also showed satisfactory results. Among these models, ensemble mean approach again outperformed for most of the cases. Furthermore, Era-based references data used bias correction results also revealed better performance in comparison with their observation-based reference data used equivalents.

Besides, the spread of the index values for three bias correction methods can be seen from Figure 5. Black Sea region showed wider range of IoA values than the other regions while Southeastern region exhibited a fairly small





**Fig. 4** Pearson coefficient ( $r$ ) distribution with station-based (point) reference data from 1965 to 2014 for three bias correction methods and five GCMs

**Table 3** Distribution of stations based on the bias correction methods for the ensemble approach for  $(r) \geq 0.4$

Model	DELTA	EQM	QM
Ensemble Mean	50/78	37/78	47/78

variability. The behavior of the stations revealed similar variability for all the bias correction methods. The variability within the region is also important since the average can be affected from the outlier values in that region and cause overestimated or underestimated regional average. According to IoA analyses, it can be concluded that there is an acceptable agreement between observed and bias corrected data especially for the Southeastern, Mediterranean and Aegean Regions regardless of the reference data set or bias correction method.

Robust future results are essential to make concrete assessments for many sectors in a continuously changing climate. Methods like statistical downscaling or bias correction improve the GCM results and attempt to capture

**Table 4** Distribution of stations—bias correction methods vs. models for  $(r) \geq 0.4$

Model/Reference Data	DELTA	EQM	QM	Total
ACCESS	53	48	46	147
Reanalyses	33	33	32	98
Station	20	15	14	49
CNRM	52	40	43	135
Reanalyses	31	29	30	90
Station	21	11	13	45
GFDL	58	45	52	155
Reanalyses	36	31	36	103
Station	22	14	16	52
MIROC	60	52	52	164
Reanalyses	37	37	38	112
Station	23	15	14	52
MRI	60	51	52	163
Reanalyses	35	30	34	99
Station	25	21	18	64
Total	283	236	245	

**Table 5** Distribution of stations reference data sets vs. bias correction methods for  $(r) \geq 0.4$ 

Bias Correction Method/Model	Reanalyses	Station	Total
DELTA	172	111	283
ACCESS	33	20	53
CNRM	31	21	52
GFDL	36	22	58
MIROC	37	23	60
MRI	35	25	60
EQM	160	76	236
ACCESS	33	15	48
CNRM	29	11	40
GFDL	31	14	45
MIROC	37	15	52
MRI	30	21	51
QM	170	75	245
ACCESS	32	14	46
CNRM	30	13	43
GFDL	36	16	52
MIROC	38	14	52
MRI	34	18	52
Total	502	262	764

the regional or local climate characteristics to some extent. These methods, however, indicate different results for different conditions such as region, reference data sets or climatic

variable of interest e.g., precipitation or temperature. For instance, Tong et al. (2021) stated that seasonal variations and the variable that is chosen to be corrected can affect the bias correction quality and notices the probability that bias correction can change the simulated signal. This may be one of the reasons that model, and method performances show variations among the regions in this study since these regions differentiate in terms of annual and seasonal precipitation patterns. Bağçacı et al. (2021) mentioned the lack of accuracy of the macro models due to the mid or small-scale climatic regions. Furthermore, they revealed that model skills may depend on period or region which ultimately may cause projection performance to be model, space or time specific. Kundzewicz et al. (2016) also presented the potential reasons of inconsistencies between observations and projections which may increase the uncertainty for the regional and local scale studies.

Beyer et al. (2020) reported that the delta method as the promising bias correction method while quantile mapping is said to be a poor choice which they applied for a considerable large global dataset. Furthermore, Enayati et al. (2021) explored that quantile mapping methods cannot always overcome the systematic errors especially over the regions that have significantly diverse topographic variations. In this study, delta method also revealed promising results. Yet, here only three bias correction methods were used meanwhile method choice is a primary determinant over the climate change assessment as Gunavathi and Selvasidhu (2021) stated. Nevertheless, it can be suggested that

**Table 6** Regional average of IoA results for bias correction methods for ensemble approach Region

	Ensemble Delta	Ensemble EQM	Ensemble QM
Aegean	0.67	0.67	0.70
Black Sea	0.48	0.54	0.55
Central Anatolia	0.61	0.62	0.65
Eastern Anatolia	0.65	0.66	0.69
Marmara	0.58	0.60	0.62
Mediterranean	0.67	0.65	0.70
South Eastern Anatolia	0.73	0.70	0.75

**Table 7** Regional average of IoA results for bias correction methods with Reanalyses reference data

Region	ACCESS			CNRM			GFDL			MIROC			MRI		
	DELTA	EQM	QM	DELTA	EQM	QM	DELTA	EQM	QM	DELTA	EQM	QM	DELTA	EQM	QM
Aegean	0.61	0.61	0.61	0.65	0.65	0.64	0.64	0.64	0.63	0.66	0.67	0.66	0.65	0.65	0.64
Black Sea	0.46	0.47	0.47	0.48	0.49	0.49	0.49	0.49	0.49	0.46	0.48	0.47	0.52	0.53	0.52
Central Anatolia	0.60	0.61	0.60	0.61	0.61	0.61	0.63	0.63	0.62	0.58	0.58	0.58	0.59	0.60	0.58
Eastern Anatolia	0.66	0.66	0.67	0.64	0.64	0.63	0.68	0.68	0.67	0.65	0.65	0.65	0.66	0.66	0.66
Marmara	0.50	0.51	0.52	0.55	0.55	0.55	0.55	0.55	0.54	0.60	0.62	0.61	0.59	0.60	0.59
Mediterranean	0.68	0.67	0.66	0.64	0.64	0.63	0.66	0.66	0.65	0.67	0.68	0.67	0.66	0.66	0.66
South Eastern Anatolia	0.70	0.71	0.72	0.69	0.68	0.67	0.71	0.70	0.70	0.70	0.69	0.69	0.70	0.69	0.69



**Table 8** Regional average of IoA results for bias correction methods with Station reference data

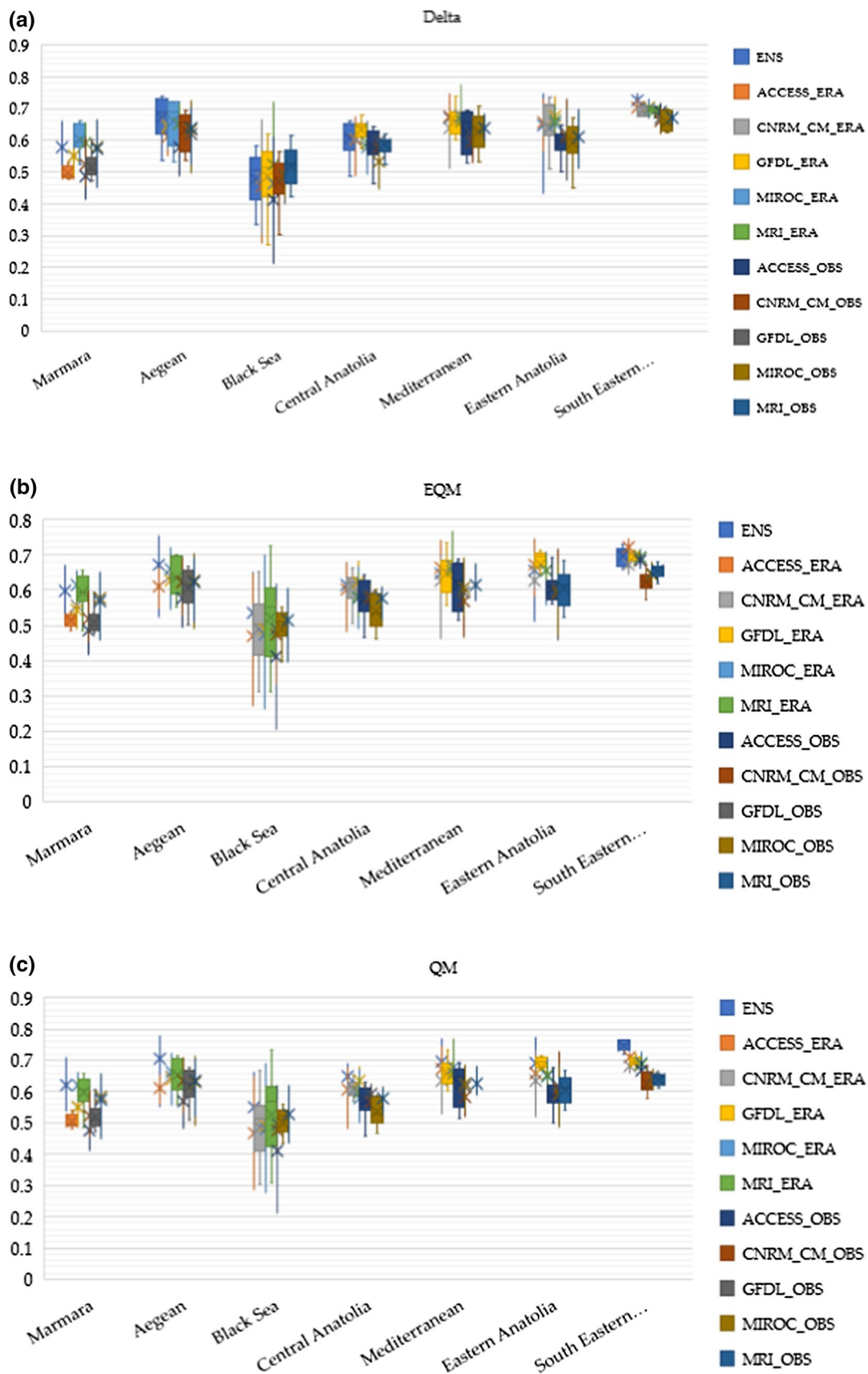
Region	ACCESS			CNRM			GFDL			MIROC			MRI		
	DELTA	EQM	QM	DELTA	EQM	QM	DELTA	EQM	QM	DELTA	EQM	QM	DELTA	EQM	QM
Aegean	0.58	0.57	0.58	0.63	0.63	0.62	0.62	0.62	0.61	0.64	0.63	0.63	0.64	0.64	0.62
Black Sea	0.41	0.41	0.41	0.47	0.48	0.47	0.49	0.50	0.49	0.49	0.50	0.49	0.52	0.53	0.52
Central Anatolia	0.58	0.57	0.57	0.58	0.57	0.58	0.60	0.59	0.59	0.53	0.54	0.54	0.59	0.58	0.58
Eastern Anatolia	0.60	0.59	0.60	0.61	0.61	0.60	0.62	0.62	0.61	0.59	0.59	0.59	0.61	0.61	0.61
Marmara	0.49	0.48	0.49	0.52	0.52	0.52	0.52	0.52	0.51	0.58	0.58	0.58	0.58	0.57	0.57
Mediterranean	0.63	0.62	0.62	0.60	0.59	0.57	0.62	0.62	0.59	0.63	0.62	0.60	0.64	0.62	0.62
South Eastern Anatolia	0.69	0.67	0.68	0.66	0.63	0.62	0.67	0.65	0.65	0.66	0.64	0.64	0.67	0.64	0.66

increasing the complexity of the methods does not always mean an improvement, based on the bias correction results in this study.

Feigenwinter et al. (2018) used three reference data set for bias correction which are stations, RCM grided and high-resolution grid with quantile mapping method, and they got satisfactory results for both. Yet, in this study it was found that using gridded data, ERA5, as a reference data set significantly improved the model adjustment results. Besides, there are studies that only use reanalysis/gridded data to validate the bias correction results such as Fauzi et al. (2020) or Tong et al. (2021). Era is assimilated by observations (stations) but ultimately a model, and the output values can be considered as the average value for the geography that grids cover based on the resolution. Nonetheless, Feigenwinter et al. (2018) also noticed that extremes or small-scale variability cannot be captured by the climate models and reveals that using quantile mapping might not appropriate for all climate impact studies. Cannon et al. (2015) claimed that although they are effective at removing historical biases relative to observations, quantile mapping can artificially corrupt future model-projected trends and proposed to ensure if the method is accurate for the purpose as Feigenwinter et al. (2018) also stated. This is particularly important if one method is selected for a large area that has various climatic and topographic features. In addition to these, Kundzewicz et al. (2016) stated that flood statistics are sensitive to the selected bias correction techniques and Casanueva et al. (2020) implied that bias correction methods can affect some certain magnitudes such as climate change signal or trends which can be induced by the observational reference or the inconsistent resolution. They also underlined the importance of the high-quality observational data sets especially for precipitation since calibration of models intensively rely on the reference data sets. There is still a complex mechanism behind the precipitation to entirely uncover and model. Particularly, nonstationary signal of the variable of interest must be sustained to obtain accurate prediction for the future since it is now accepted that there are indisputable nonstationary impacts.

Regardless of the bias correction methods, Black Sea region of Turkey results were not efficiently adjusted, which accommodate an irregular terrain with various land-use and land-cover types and, land-sea interaction dynamics (Amjad et al. 2020), while Southeastern region has demonstrated the highest accuracy. Sariş et al. (2010) found clear evidence of spatially varied seasonality over Turkey for the precipitation regimes. They also calculated a composite regime which presented Southeast Anatolia as the least diverse region while Black Sea and Mediterranean showed the highest number of different regimes over the regions. This can be one of the reasons of differentiated bias correction results among the regions over Turkey.

While it is very valuable and necessary to translate the GCM data into local scale for long-term risk assessment of establishing a scientific basis for the decision making, results must be representative of selected area and stations in that area. Statistical methods, on the other hand, assume statistical relationships and cannot provide a physically consistent results yet they are being preferred as a result of computational simplicity. Different studies investigated and tried to propose a solution for these capacity problems. Salehnia et al. (2019) compared statistical and dynamical downscaling methods and presented those dynamical downscaling method perform better than the statistical method for their study period. Tiwari et al. (2019) used a combination of statistical and dynamical methods and implied that better representation of orography, moisture and vertical pressure velocity in the regional climate model improved the results of downscaling methods, and LeRoy et al. (2021) used an integrated statistical dynamical downscaling method for high-resolution urban climate simulations. Song et al. (2021) interpolated the historical and projected data of the four nearest GCM grid points to the station's location and then bias corrected the interpolated data against the observed data using quantile mapping. Therefore, not only the selected statistical downscaling methods but it may be the statistical downscaling itself which cannot demonstrate a sufficient performance. In this study, city central stations were used, which may cause challenges such as heterogeneity, scale and



**Fig. 5** Regional Model evaluation by index of agreement for three methods **a** Delta, **b** EQM, **c** QM

local modifications of the temperature that affect the bias correction performance as LeRoy et al. (2021) indicated. Zhang et al. (2020) recommended the ensembles of models to overcome variations across climatic regions. Furthermore, increasing the number of stations in the GCM grid may also increase the bias correction performance as well since representing the reference period with a high number of station data may improve the spatial inconsistencies.

## Conclusion and remarks

This analysis has focused on the performance of three bias correction method with two reference data sets of precipitation (station-based observations and Era) for 5 single model and ensemble mean approach. Simulation performance of model-bias correction method-reference data set combination was assessed on monthly basis for every single station. Furthermore, regional analyses were added to station-based analyses to check and verify the results. It was shown that performance of models mostly affected by both region, reference data set. Bias correction methods were not detected as effective as the reference data set over the performance. While it is said to be a simple method, Delta method slightly outperformed among the other bias correction techniques for the computation that used observation as reference data. Considering the computations that was carried out with ensemble mean bias correction and single model-bias correction with ERA data, the difference between the bias correction methods was not significant. Besides ensemble approach, MIROC and MRI models were selected as the best performing models over the region. In addition, in this study selection of the reference data sets also found a dominant factor for the prediction accuracy. In this regard, an ensemble of ground-based station data can be used to capture the whole grid area that matches the corresponding GCM grid, and a single time series can be obtained to represent that grid to bias correct. Another alternative for this is using as much ground-based station data as possible for a specific grid and averaging the bias corrected results of these ground-based stations in the grid to decrease the uncertainty and bias.

One of the limitations of this study was that it has used 5 models which is not a high number. Better results may be obtained with different models. Furthermore, the choice of most appropriate bias correction technique may depend on the time scale of the data used (daily, monthly, etc.), model/models combination, regions that are subject of interest and the reference data set. Global climate models (GCMs), in general, are bias corrected to better match the observations. However, if a GCM does not accurately simulate the underlying physical processes, the rest of the analyse become critical. On the other hand, bias correction is also criticized in the literature. Regarding these issues, increase in model

resolution and ensemble prediction is another alternative. Furthermore, a long observational record is needed to build accurate statistical relationship which this relationship is assumed to be continue as same as history in future. Distorting the nonstationary signal is another drawback of the bias correction methods in literature. There are, however, developing methods that claim to preserving the trend, or the observed variability. Dynamical downscaling is an alternative for statistical downscaling methods, but the required computational loads and costs limits the application in widespread.

**Funding** This research received no external funding.

## Declarations

**Conflict of interest** The author declares no conflict of interest.

## References

- AgriMetSoft (2018). SD-GCM Tool [Computer software]. Available at: <https://agrimetsoft.com/SD-GCM.aspx>
- Amjad MR, Yilmaz MT, Yucel I, Yilmaz KK (2020) Performance evaluation of satellite- and model-based precipitation products over varying climate and complex topography. *J Hydrol* 584:124707
- Atalay I, Efe R, Soykan A (2008) Mediterranean ecosystems of Turkey: ecology of the taurus mountains.
- Bağçacı SÇ, Yucel I, Duzenli E, Yilmaz MT (2021) Intercomparison of the expected change in the temperature and the precipitation retrieved from CMIP6 and CMIP5 climate projections: A Mediterranean hot spot case, Turkey. *Atmospheric Research*.
- Beyer R, Krapp M, Manica A (2020) An empirical evaluation of bias correction methods for palaeoclimate simulations. *Clim past* 16:1493–1508. <https://doi.org/10.5194/cp-16-1493-2020>
- Boe J, Terray L, Habets F, Martin E (2007) Statistical and dynamical downscaling of the Seine basin climate for hydro-meteorological studies. *Int J Climatol* 27:1643–1655. <https://doi.org/10.1002/joc.1602>
- Cannon AJ, Sobie SR, Murdock TQ (2015) Bias correction of GCM precipitation by quantile mapping: how well do methods preserve changes in quantiles and extremes? *J Clim* 28(17):6938–6959
- Casanueva A, Herrera S, Iturbide M et al (2020) Testing bias adjustment methods for regional climate change applications under observational uncertainty and resolution mismatch. *Atmos Sci Lett* 21(21):e978. <https://doi.org/10.1002/asl.978>
- Chen J, Brissette FP, Chaumont D, Braun M (2013) Finding appropriate bias correction methods in downscaling precipitation for hydrologic impact studies over North America. *Water Resour Res* 49:4187–4205. <https://doi.org/10.1002/wrcr.20331>.
- Checa-Garcia R, Hegglin MI, Kinnison D, Plummer DA, Shine KP (2018) Historical tropospheric and stratospheric ozone radiative forcing using the CMIP6 database. *Geophys Res Lett* 45:3264–3273. <https://doi.org/10.1002/2017GL076770>
- Enayati M, Bozorg-Haddad O, Bazrafshan J, Hejabi S, Chu X (2021) Bias correction capabilities of quantile mapping methods for rainfall and temperature variables. *J Water Clim Change* 12(2):401–419. <https://doi.org/10.2166/wcc.2020.261>
- CDS (2021) ERA5 monthly averaged data on single levels from 1979 to present Copernicus CDS. <https://doi.org/10.24381/cds.f17050d7>

- Eyring V, Bony S, Meehl GA, Senior CA, Stevens B, Stouffer RJ, Taylor KE (2016) Overview of the coupled model intercomparison project phase 6 (CMIP6) experimental design and organization. *Geosci Model Dev* 9:1937–1958. <https://doi.org/10.5194/gmd-9-1937-2016>
- Fauzi FA, Kuswanto H, Atok RM (2020) Bias correction and statistical downscaling of earth system models using quantile delta mapping (QDM) and bias correction constructed analogues with quantile mapping reordering (BCCAQ)
- Feigenwinter I, Kotlarski S, Casanueva A, Fischer AM, Schwierz C, Liniger MA (2018) Exploring quantile mapping as a tool to produce user-tailored climate scenarios for Switzerland, Technical Report MeteoSwiss, 270, 44 pp
- Fu Y, Lin Z, Guo D (2020) Improvement of the simulation of the summer East Asian westerly jet from CMIP5 to CMIP6. *Atmosph Oceanic Sci Lett* 13:550–558
- Goldenson N, Thackeray CW, Hall AD, Swain DL, Berg N (2021) Using large ensembles to identify regions of systematic biases in moderate to heavy daily precipitation. *Geophys Res Lett* 48:e2020GL092026. <https://doi.org/10.1029/2020GL092026>
- Gudmundsson L (2014) Qmap: statistical transformations for post-processing climate model output
- Gudmundsson L, Bremnes JB, Haugen JE, Engen-Skaugen T (2012) Technical note downscaling RCM precipitation to the station scale using statistical transformations – a comparison of methods. *Hydrol Earth Syst Sci* 16(9):3383–3390. <https://doi.org/10.5194/hess-16-3383-2012>
- Gunavathi S, Selvasidhu R (2021) Assessment of various bias correction methods on precipitation of regional climate model and future projection, 07 April 2021, PREPRINT (Version 1) available at Research Square [<https://doi.org/10.21203/rs.3.rs-339080/v1>]
- Haerter JO, Eggert B, Moseley C, Piani C, Berg P (2015) Statistical precipitation bias correction of gridded model data using point measurements. *Geophys Res Lett* 42:1919–1929
- Hay LE, Wilby RL, Leavesley GH (2000) A comparison of delta change and downscaled GCM scenarios for three mountainous basins in the united states. *JAWRA J Am Water Resour Assoc* 36:387–397. <https://doi.org/10.1111/j.1752-1688.2000.tb04276.x>
- Hersbach H, Bell B, Berrisford P et al (2020) The ERA5 global reanalysis. *Q J R Meteorol Soc* 146:1999–2049. <https://doi.org/10.1002/qj.3803>
- Hosseinzadehtalaei P, Ishadi NK, Tabari H, Willems P (2021) Climate change impact assessment on pluvial flooding using a distribution-based bias correction of regional climate model simulations. *J Hydrol* 598:126239
- Jose DM, Dwarakish GS (2022) Ranking of downscaled CMIP5 and CMIP6 GCMs at a basin scale: case study of a tropical river basin on the South West coast of India. *Arab J Geosci* 15:120. <https://doi.org/10.1007/s12517-021-09289-0>
- Kara F, Yucel I, Akyurek Z (2016) Climate change impacts on extreme precipitation of water supply area in Istanbul: use of ensemble climate modelling and geo-statistical downscaling. *Hydrol Sci J* 61:2481–2495
- Kundzewicz ZW, Kanae S, Seneviratne SI, Handmer J, Nicholls N, Peduzzi P, Mechler R, Bouwer LM, Arnell N, Mach K, Muir-Wood RG, Robert B, Wolfgang K, Gerardo B, Yasushi H, Kiyoshi T, Boris S (2014) Flood risk and climate change: global and regional perspectives. *Hydrol Sci J* 59(1):1–28. <https://doi.org/10.1080/02626667.2013.857411>
- Kundzewicz ZW, Krysanova V, Dankers R, Hirabayashi Y, Kanae S, Hattermann FF, Huang S, Milly PC, Stoffel M, Driessen PP, Matczak P, Quevauviller P, Schellnhuber HJ (2016) Differences in flood hazard projections in Europe – their causes and consequences for decision making. *Hydrol Sci J* 62:1–14
- Le Roy B, Lemonsu A, Schoetter R (2021) A statistical–dynamical downscaling methodology for the urban heat island applied to the EURO-CORDEX ensemble. *Clim Dyn* 56:2487–2508. <https://doi.org/10.1007/s00382-020-05600-z>
- Lin W, Chen H (2020) Assessment of model performance of precipitation extremes over the mid-high latitude areas of Northern Hemisphere: from CMIP5 to CMIP6. *Atmosph Oceanic Sci Lett* 13:598–603
- Liu X, Li C, Zhao T, Han L (2020) Future changes of global potential evapotranspiration simulated from CMIP5 to CMIP6 models. *Atmosph Oceanic Sci Lett* 13:568–575
- Luo N, Guo Y, Gao Z, Chen K, Chou J (2020) Assessment of CMIP6 and CMIP5 model performance for extreme temperature in China. *Atmosph Oceanic Sci Lett* 13:589–597
- Maraun D, Wetterhall F, Ireson AM, Chandler RE, Kendon EJ, Widmann M, Brienen S, Rust HW, Sauter T, Themeßl M, Venema VKC, Chun KP, Goodess CM, Jones RG, Onof C, Vrac M, Thiele-Eich I (2010) Precipitation downscaling under climate 20 change: recent developments to bridge the gap between dynamical models and the end user. *Rev Geophys* 48:RG3003. <https://doi.org/10.1029/2009RG000314>
- Monerie PA, Wainwright CM, Sidibe M, Akinsanola AA (2020) Model uncertainties in climate change impacts on Sahel precipitation in ensembles of CMIP5 and CMIP6 simulations. *Clim Dyn* 55:1385–1401. <https://doi.org/10.1007/s00382-020-05332-0>
- Muñoz-Sabater J, Dutra E, Agustí-Panareda A, Albergel C, Arduini G, Balsamo G, Boussetta S, Choulga M, Harrigan S, Hersbach H, Martens B, Miralles DG, Piles M, Rodríguez-Fernández NJ, Zsoter E, Buontempo C, Thépaut J-N (2021) ERA5-Land: a state-of-the-art global reanalysis dataset for land applications. *Earth Syst. Sci. Data* 13:4349–4383. <https://doi.org/10.5194/essd-13-4349-2021>
- Navarro-Racines CE, Tarapues-Montenegro JE (2015) Bias-correction in the CCAFS-climate portal: a description of methodologies. Decision and policy analysis (DAPA) research area. International Center for Tropical Agriculture (CIAT), Cali, Colombia
- Ngoma H, Wen W, Ayugi B, Babauosmail H, Karim R, Ongoma V (2021) Evaluation of precipitation simulations in CMIP6 models over Uganda. *Int J Climatol* 41:4743–4768. <https://doi.org/10.1002/joc.7098>
- Nissan H, Goddard L, de Perez EC et al (2019) On the use and misuse of climate change projections in international development. *WIREs Clim Change* 10:e579. <https://doi.org/10.1002/wcc.579>
- Ombadi M, Nguyen P, Sorooshian S, Hsu K (2018) Developing intensity-duration-frequency (IDF) curves from satellite-based precipitation: methodology and evaluation. *Water Resour Res* 54:7752–7766
- O’Neill BC, Tebaldi C, Vuuren DP, Eyring V, Friedlingstein P, Hurtt GC, Knutti R, Kriegler E, Lamarque J, Lowe JA, Meehl GA, Moss RH, Riahi K, Sanderson BM (2016) The scenario model intercomparison project (ScenarioMIP) for CMIP6. *Geosci Model Develop* 9:3461–3482
- Panofsky HA, Brier GW (1968) Some applications of statistics to meteorology. The Pennsylvania State University, 224 pp
- Pereira HR, Meschiatti MC, Pires RC, Blain GC (2018) On the performance of three indices of agreement: an easy-to-use r-code for calculating the Willmott indices. *Bragantia* 77:394–403
- Piani C, Weedon G, Best M, Gomes S, Viterbo P, Hagemann S, Haerter J (2010) Statistical bias correction of global simulated daily precipitation and temperature for the application of hydrological models. *J Hydrol* 395:199–215. <https://doi.org/10.1016/j.jhydrol.2010.10.024>
- R Core Team (2021) R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL <https://www.R-project.org/>.
- Ramirez-Villegas J, Challinor AJ, Thornton PK, Jarvis A (2013) Implications of regional improvement in global climate models for agricultural impact research. *Environ Res Lett* 8:24018



- Salehnia N, Hosseini FS, Farid A, Kolsoumi S, Zarrin A, Hashemina M (2019) Comparing the performance of dynamical and statistical downscaling on historical run precipitation data over a semi-arid region. *Asia-Pac J Atmos Sci* 55:737–749
- Sariş F, Hannah DM, Eastwood WJ (2010) Spatial variability of precipitation regimes over Turkey. *Hydrol Sci J* 55(2):234–249
- SEVENTH NATIONAL COMMUNICATION OF TURKEY UNDER THE UNFCCC (2018) SEVENTH NATIONAL COMMUNICATION OF TURKEY Under the United Nations Framework Convention on Climate Change Republic of Turkey Ministry of Environment and Urbanization. Republic of Turkey Ministry of Environment and Urbanization General Directorate of Environmental Management Department of Climate Change
- Song YH, Nashwan MS, Chung E, Shahid S (2021) Advances in CMIP6 INM-CM5 over CMIP5 INM-CM4 for precipitation simulation in South Korea. *Atmosph Res* 247:105261
- Stellingwerf S, Riddle E, Hopson TM, Knievel JC, Brown B, Gebremichael M (2011) Optimizing precipitation forecasts for hydrological catchments in Ethiopia using statistical bias correction and multi-modeling. *Earth and Space Sci* 8:e2019EA000933. <https://doi.org/10.1029/2019EA000933>
- Stouffer RJ, Eyring V, Meehl GA, Bony S, Senior CA, Stevens B, Taylor KE (2017) CMIP5 scientific gaps and recommendations for CMIP6. *Bull Am Meteor Soc* 98:95–105
- Sunyer MA et al (2015) Inter-comparison of statistical downscaling methods for projection of extreme precipitation in Europe. *Hydrol Earth Syst Sci* 19(1827–1847):2015. <https://doi.org/10.5194/hess-19-1827-2015>
- Taylan D, Aydın T (2018) The trend analysis of lakes region precipitation data in Turkey. *Cumhuriyet Science Journal* 39(1):258–273. <https://doi.org/10.17776/cs.j.406271>
- Thiemeßl MJ, Gobiet A, Heinrich G (2012) Empirical-statistical downscaling and error correction of regional climate models and its impact on the climate change signal. *Clim Change* 112:449–468. <https://doi.org/10.1007/s10584-011-0224-4>
- Thiemeßl MJ, Gobiet A, Leuprecht A (2011) Empirical-statistical downscaling and error correction of daily precipitation from regional climate models. *Int J Climatol* 31:1530–1544. <https://doi.org/10.1002/joc.2168>
- Tiwari PR, Kar SC, Mohanty UC et al (2019) Comparison of statistical and dynamical downscaling methods for seasonal-scale winter precipitation predictions over north India. *Int J Climatol* 39:1504–1516. <https://doi.org/10.1002/joc.5897>
- Tong Y, Gao X, Han Z et al (2021) Bias correction of temperature and precipitation over China for RCM simulations using the QM and QDM methods. *Clim Dyn* 57:1425–1443. <https://doi.org/10.1007/s00382-020-05447-4>
- Ullah A, Salehnia N, Kolsoumi S, Ahmad A, Khaliq T (2018) Prediction of effective climate change indicators using statistical downscaling approach and impact assessment on pearl millet (*Pennisetum glaucum* L) yield through genetic algorithm in Punjab. *Pakistan Ecol Indic* 90:569–576. <https://doi.org/10.1016/j.ecoli.2018.03.053>
- Wetterhall F, Pappenberger F, He Y, Freer J, Cloke HL (2012) Conditioning model output statistics of regional climate model precipitation on circulation patterns. *Nonlin. Processes Geophys.* 19:623–633. <https://doi.org/10.5194/npg-19-623-2012>
- Willmott CJ, Robeson SM, Matsuura KA (2012) A refined index of model performance. *Int J Climatol* 32:2088–2094. <https://doi.org/10.1002/joc.2419>
- Willmott CJ (1981) On the validation of models. *Phys Geogr* 2:184–194
- Wood AW, Leung LR, Sridhar V, Lettenmaier DP (2004) Hydrologic implications of dynamical and statistical approaches to downscaling climate model outputs. *Clim Change* 62:189–216. <https://doi.org/10.1023/B:CLIM.0000013685.99609.9e>
- Wuthiwongyothin S, Mili S, Phadungkarnlert N (2020) A study of correcting climate model daily rainfall product using quantile mapping in upper Ping River Basin, Thailand. In: Trung VN, Xiping D, Thanh TT (eds) Springer. Singapore. [https://doi.org/10.1007/978-981-15-0291-0\\_166](https://doi.org/10.1007/978-981-15-0291-0_166)
- Wyser K, Kjellström E, Königk T, Martins H, Doescher R (2020) Warmer climate projections in EC-Earth3-Veg: the role of changes in the greenhouse gas concentrations from CMIP5 to CMIP6. *Environ Res Lett* 15:054020
- Zhang L, Xu Y, Meng C, Li X, Liu H, Wang C (2020) Comparison of statistical and dynamic downscaling techniques in generating high-resolution temperatures in China from CMIP5 GCMs. *J Appl Meteorol Climatol* 59(2):207–235