



# Performance of the general circulation models in simulating temperature and precipitation over Iran

Mohammadsadegh Abbasian<sup>1</sup> · Sanaz Moghim<sup>2</sup> · Ahmad Abrishamchi<sup>2</sup>

Received: 6 November 2017 / Accepted: 7 March 2018 / Published online: 17 March 2018  
© Springer-Verlag GmbH Austria, part of Springer Nature 2018

## Abstract

General Circulation Models (GCMs) are advanced tools for impact assessment and climate change studies. Previous studies show that the performance of the GCMs in simulating climate variables varies significantly over different regions. This study intends to evaluate the performance of the Coupled Model Intercomparison Project phase 5 (CMIP5) GCMs in simulating temperature and precipitation over Iran. Simulations from 37 GCMs and observations from the Climatic Research Unit (CRU) were obtained for the period of 1901–2005. Six measures of performance including mean bias, root mean square error (RMSE), Nash-Sutcliffe efficiency (NSE), linear correlation coefficient ( $r$ ), Kolmogorov-Smirnov statistic (KS), Sen's slope estimator, and the Taylor diagram are used for the evaluation. GCMs are ranked based on each statistic at seasonal and annual time scales. Results show that most GCMs perform reasonably well in simulating the annual and seasonal temperature over Iran. The majority of the GCMs have a poor skill to simulate precipitation, particularly at seasonal scale. Based on the results, the best GCMs to represent temperature and precipitation simulations over Iran are the CMCC-CMS (Euro-Mediterranean Center on Climate Change) and the MRI-CGCM3 (Meteorological Research Institute), respectively. The results are valuable for climate and hydrometeorological studies and can help water resources planners and managers to choose the proper GCM based on their criteria.

## 1 Introduction

Climate change alters the long-term weather patterns and conditions. The Fifth Assessment Report (AR5) of the Intergovernmental Panel on Climate Change (IPCC) demonstrated that the surface temperature over the period of 2003–2012 has a warming of 0.78 °C (with a confidence interval of 0.72–0.85 °C) relative to the period of 1850–1900, and there is an upward trend in surface temperature of almost the entire globe, while there is a medium confidence in change of the

average precipitation over all lands after 1951 (IPCC 2013). As we know, there is an uncertainty in the future climate due to the non-stationarity of variables. Thus, the projected mean and the distribution of the variables are important to climatologists and water resources planners, particularly in semiarid and arid regions where climate change has greater impacts (IPCC 1996).

Currently, the General Circulation Models (GCMs) are the most promising tools for projecting future climate trends and variability (Kharin et al. 2007; Loukas et al. 2008). However, before assessing future climate changes based on the GCMs simulations, it is crucial to evaluate the performance of the GCMs in simulating climate variables (Dessai 2005; Reifen and Toumi 2009; Belda et al. 2015; Nasrollahi et al. 2015). Many studies aimed to assess the GCMs outputs relative to observations. Some researchers evaluated the GCMs at global scales. Gleckler et al. (2008) evaluated the performance of 22 CMIP3 GCMs in simulating several atmospheric variables relative to NCEP/NCAR reanalysis datasets. They ranked models for each variable using a measure of the relative error and showed that the performance of GCMs varies substantially. They also concluded that using just a single index of the model performance can be misleading. Their results showed that the relative ranking of models differs significantly for

---

✉ Sanaz Moghim  
moghim@sharif.edu

Mohammadsadegh Abbasian  
m.abbasian1@student.sharif.edu

Ahmad Abrishamchi  
abrisham@sharif.edu

<sup>1</sup> Department of Civil Engineering, Sharif University of Technology, Azadi Avenue, Tehran, Iran

<sup>2</sup> Department of Civil Engineering and UNESCO Chair in Water and Environment Management for Sustainable Cities, Sharif University of Technology, Azadi Avenue, Tehran, Iran

different variables; however, some GCMs (e.g., HadCM3) perform well in many respects. Hao et al. (2013) compared the joint occurrence of monthly continental precipitation and temperature extremes in 13 GCMs from the fifth phase of the Coupled Model Intercomparison Project (CMIP5) with the Climatic Research Unit (CRU) and the University of Delaware (UD) observations. The results showed that CMIP5 GCMs can reasonably represent the trends in the joint extremes. However, there are noticeable differences between the regional patterns and the magnitude of changes in the GCM simulations relative to the observations. Knutti et al. (2013) constructed a model family tree from CMIP3 and CMIP5 GCMs using a hierarchical clustering of pairwise distance matrix for temperature and precipitation. The results showed that most CMIP5 GCMs are very similar to their predecessors, while they are in close agreement with NCEP and Global Precipitation Climatology Center (GPCC) observations. Sillmann et al. (2013) evaluated CMIP5 GCMs in simulating 27 climate extreme indices and compared CMIP5 and CMIP3 GCMs. Their analyses showed that CMIP5 GCMs are able to represent climate extremes and their trends relative to four reanalysis datasets. They also indicated that CMIP5 GCMs have a better performance in representing the magnitude of the precipitation indices compared to CMIP3 GCMs. Furthermore, the results revealed that the variability of the CMIP5 GCMs for the temperature indices is smaller than those of the CMIP3 GCMs. Pascale et al. (2014) proposed two new indicators of the rainfall seasonality namely the relative entropy (RE) and the dimensionless seasonality index (DSI). They claimed that the RE and DSI allow objective metrics for model intercomparison and ranking. The results showed that the RE is overestimated over tropical Latin America and is underestimated in Western Africa, West Mexico, and East Asia in the CMIP5 simulations. In addition, no model can represent well the DSI spatial variability. Belda et al. (2015) evaluated 43 CMIP5 GCMs using the Köppen-Trewartha climate classifications. They suggested that the annual and monthly mean values of temperature and precipitation and their periodicity can be summarized in the climate classifications. The results showed that many models cannot capture the rainforest climate type (Ar), and almost half of the models underestimate the desert climate type (BW). Additionally, many models overestimate the boreal climate type (E). They also measured the similarity between the GCMs using a hierarchical cluster analysis. The analysis showed that the GCMs from the same modeling center are often grouped in the same cluster and therefore have quite similar results. McMahon et al. (2015) assessed the precipitation and temperature data from 23 CMIP3 GCMs. They compared the GCM estimates of the mean annual precipitation and temperature and the standard deviation of the annual precipitation and those of the observed CRU data using the Nash-Sutcliffe efficiency

(NSE), the coefficient of determination ( $R^2$ ), and six other statistics. They also assessed the ability of the GCMs to represent the Köppen-Geiger climate classification. They concluded that the HadCM3 is the best GCM among CMIP3 GCMs over globe. Nasrollahi et al. (2015) assessed the performance of 41 CMIP5 GCMs in representing the CRU observations of continental drought areas and their trends. They defined meteorological drought in terms of the Standardized Precipitation Index (SPI) and analyzed the trend of SPI using the Mann-Kendall test. The results showed that the envelope of the CMIP5 SPI time series encompasses the CRU SPI time series for drought regions during 1902–2005. However, most GCMs overestimate the index over the extreme drought regions. In addition, the trend in the CRU data can be represented by about three-fourths of the GCMs. They also evaluated the consistency of the precipitation distribution functions of the GCM simulations with those of the observations using the Kullback-Leibler divergent test. The test measures the distance between two distributions. The results indicated no significant difference between the distributions of the simulations and observations.

Many studies evaluated the GCMs at regional scales. Bonsal and Prowse (2006) assessed the ability of seven GCMs to simulate the mean values and the spatial variability of temperature and precipitation over four regions across Northern Canada. They calculated the differences between seasonal mean temperature and precipitation of the GCM simulations and those of the CRU data. They also plotted the Taylor diagram for annual precipitation and temperature. The results showed that the HadCM3 can best represent annual and seasonal temperature over all sub-regions, while there is an obvious regional difference between GCMs' precipitation and observed CRU one, and all models significantly overestimate this variable. Perkins et al. (2007) evaluated the performance of the CMIP3 GCMs in simulating precipitation, minimum temperature, and maximum temperature over Australia using Probability Density Functions (PDFs) of the variables. They claimed that it is better to evaluate the GCMs using their PDFs than their first or second moments. The results showed that although most GCMs can reasonably simulate the distribution of precipitation, they overestimate precipitation at low quantiles. Additionally, only 3 of the 14 GCMs can capture 80% of the observed PDFs of the precipitation. For minimum and maximum temperature, 10 of the 13 GCMs and 6 of the 10 GCMs can capture 80% of the observed PDFs, respectively. Johnson and Sharma (2009) assessed the performance of nine GCMs in simulating eight variables over Australia using a skill score. They concluded that pressure, temperature, and humidity are the variables with the highest skill score and precipitation is the variable with the lowest. Yin et al. (2012) evaluated the performance of 11 CMIP5 GCMs in simulating rainfall over tropical South America. The results showed that most models

underestimate both convective and large-scale precipitation during the dry season. Moreover, most models underestimate large-scale precipitation during the wet season. Chen and Frauenfeld (2014) evaluated the precipitation simulations of 20 CMIP5 GCMs over China and quantified CMIP5 improvements over CMIP3. They calculated the differences, the root mean square error (RMSE), the standard deviation of the error, and the linear correlation coefficient ( $r$ ) between the CRU and GPCC observed datasets and CMIP3 and CMIP5 outputs. In addition, they evaluated the performance of the GCMs in simulating the spatial distribution of precipitation over China using a skill score. They concluded that the CMIP5 models represent a significant improvement over the CMIP3 models. However, both CMIP5 and CMIP3 GCMs overestimate seasonal and annual precipitation. Miao et al. (2014) evaluated the performance of 24 CMIP5 GCMs in simulating intra-annual, annual, and decadal temperature over Northern Eurasia by the analysis of the bias, linear trend, and Taylor diagram. The analysis by the bias showed that although most GCMs give reasonably accurate simulations of the annual and the seasonal mean temperature, they mostly overestimate the annual mean temperature. Furthermore, most GCMs can approximate the trend of temperature. The Taylor diagram showed that the accuracy of the GCMs in simulating decadal temperature is better than those of the annual temperature. Sonali et al. (2016) assessed the performance of 46 GCMs from CMIP3 and CMIP5 in simulating monthly and seasonal maximum and minimum temperature over India. They compared the PDFs of the observations and the GCM simulations using a scale measure. In addition, they quantified the differences between the models and the observed datasets by the RMSE and  $r$ . Then, they defined a new scale measure that is an intersection of above three metrics to evaluate and rank the GCMs. This new metric demonstrated the superiority of the CMIP5 GCMs over the CMIP3 GCMs. They also compared the model trends with the observed trends using the Sen's slope estimator. Based on the studies, the performance of the GCMs varies significantly for different variables in different locations; therefore, it is necessary to evaluate the performance of the GCMs for the study region prior to using their outputs.

Iran is located in the arid and semi-arid belt of the northern hemisphere extended from 44° E to 64° E longitude and from 25° N to 40° N latitude. Many studies focused on the consequences of climate change over Iran using the GCMs' outputs, such as climate impacts on temperature and precipitation (e.g., Samadi et al. 2013; Abbasnia and Toros 2016; Nazemosadat et al. 2016), runoff (e.g., Zarghami et al. 2011; Samadi et al. 2012; Razmara et al. 2013; Shadkam et al. 2016), groundwater (e.g., Hashemi et al. 2015), drought (e.g., Sayari et al. 2012), flood (e.g., Khzaei et al. 2011), and crop yield (e.g., Abbaspour et al. 2009; Gohari et al. 2013). However, to the knowledge of the authors, there is no complete evaluation of

the GCMs' simulations over Iran, and hence the performance of different GCMs over this region is unclear. This paper presents the first comprehensive evaluation of the CMIP5 climate models in simulating temperature and precipitation over Iran. Temperature and precipitation are two predominant variables among climate and hydrological variables. Temperature plays a pivotal role in potential evapotranspiration, soil moisture, crop yield, water demand, and water quality (e.g., by its direct influence on the rate of reactions and microorganism growth). Precipitation affects runoff and groundwater recharge, and characterizes flood and different types of drought such as meteorological, agricultural, hydrological, and socioeconomic drought. The performance of the climate models is evaluated using performance criteria including the mean bias, RMSE, NSE,  $r$ , Kolmogorov-Smirnov statistic (KS), Sen's slope estimator, and Taylor diagram. Those criteria can determine the ability of the models to simulate the climate variables. Section 2 describes the CMIP5 and observational datasets. Section 3 explains the methodology and measures of the performance. Section 4 presents the results of the evaluation for temperature and precipitation simulations with discussion, followed by summary and concluding remarks in Section 5.

## 2 Datasets

### 2.1 General circulation models (GCMs)

GCMs are extremely complex models, which can represent physical processes with geochemical and biological (mostly carbon cycle) interactions in the atmosphere, ocean, land surface, and cryosphere using three-dimensional grids (IPCC 2007). In addition, they are the most advanced tools to predict the future climate under climate change scenarios. Since there are many climate models developed by different modeling groups around the world, the Coupled Model Intercomparison Project (CMIP) was established as a standard framework for the GCM evaluation, intercomparison, documentation, and access. CMIP5 is the latest phase of the CMIP that provides the simulations of more than 50 state-of-the-art GCMs from more than 20 modeling groups. The CMIP5 datasets include two types of the climate simulations: the long-term experiments (century time scales) and the near-term experiments (10–30 years). Both experiments are the outputs of the Atmosphere-Ocean General Circulation Models (AOGCMs) or the Earth System Models (ESMs). AOGCMs couple the Atmospheric GCMs (AGCMs) to the Oceanic GCMs (OGCMs). AGCMs aim to model the atmosphere using predefined sea surface temperatures, while OGCMs try to model the ocean with predefined fluxes from the atmosphere. AOGCMs intend to model the interaction between the atmosphere and ocean, and therefore, they do not need to predefine fluxes across the atmosphere-ocean

interface. ESMs are the AOGCMs that are coupled to biochemical components to calculate the predominant fluxes of carbon between the atmosphere, ocean, and terrestrial biosphere carbon reservoirs so that they can consider the effects of vegetation changes on climate (Hannah 2015). ESMs are capable of interactively determining the concentrations of constituents using their time-dependent emissions. For detailed information about the CMIP5 experiments, the reader is referred to Taylor et al. (2009, 2012). This study uses GCMs that have historical experiment (Table 1); thus monthly precipitation and temperature simulations of 37 GCMs for the period of 1901–2005 are obtained from the “historical experiment” of the CMIP5 datasets (<https://esgf-node.llnl.gov>). The historical experiment is a long-term experiment (1850–2005) used for evaluating the performance of the GCMs in the simulation of climate variables. Note that the original GCMs datasets are used.

## 2.2 Climatic Research Unit (CRU) datasets

CRU is an institution that provides a globally land-only monthly time series for commonly used surface climate variables including mean temperature and precipitation. CRU has been used in many fields including climate, hydrology, agriculture, ecology, biodiversity, and forestry. CRU uses the Climate Anomaly Method (CAM) to produce  $0.5^\circ \times 0.5^\circ$  gridded data from monthly observations of the meteorological stations. Each station should have more than 75% non-missing data in the base period (1961–1990) to calculate the average values of the baseline. The influence of each station is measured in terms of Correlation Decay Distances (CDDs). CDDs vary with time and between variables ranging from 1200 km for mean temperature to 450 km for precipitation. Each grid is filled with data interpolated from stations within the CDD range. Consequently, CRU aims to make the dataset as complete as possible using distant stations. For detailed information about CRU dataset, the reader is referred to Harris et al. (2013). This study used monthly CRU precipitation and temperature (CRU TS 3.24.01) for the period of 1901–2005 (<http://badc.nerc.ac.uk/data/cru/>).

## 3 Methodology

### 3.1 Data preparation

To evaluate the GCM simulations relative to the CRU observations, all GCM simulations and CRU data are re-gridded into a common  $3.75^\circ \times 3.75^\circ$  resolution by the bilinear interpolation, which is a widely used method (Wang and Chen 2013; Chen and Frauenfeld 2014; Belda et al. 2015; Aloysi et al. 2016).

The annual and seasonal time series of temperature in each grid are determined by averaging monthly temperature over each year and each season, respectively. Similarly, the annual and seasonal precipitation is calculated by aggregating monthly precipitation over each year and each season. Hence, each year has four seasonal values (winter: December to February, spring: March to May, summer: June to August, and fall: September to November).

### 3.2 Measures of the performance

In this study, six different measures of the performance are used to evaluate GCMs' ability to simulate temperature and precipitation over Iran. For each measure, GCMs are ranked such that the GCM with the best performance is designated as rank 1. Since all these metrics can be important to hydrologists and climatologists, they can choose suitable GCMs based on their criteria and priorities for a specific application.

#### 3.2.1 Mean Bias

Incomplete understanding of the climate system and simplified assumptions can cause biases in climate model simulations (Reichler and Kim 2008). Therefore, the mean bias is a critical indicator of the model performance. The mean bias is defined as follows:

$$\text{mean bias} = \bar{y}^m - \bar{y}^o \quad (1)$$

where  $\bar{y}^m$  and  $\bar{y}^o$  are the mean of the GCM simulations and observations, respectively. It is clear that the positive and negative values of the bias represent overestimation and underestimation of the GCMs, respectively. The GCMs are ranked based on the minimum absolute value of the mean bias, i.e., the GCM with the closest mean bias to zero has rank 1.

#### 3.2.2 Root mean square error

The root mean square error (RMSE), a very common metric providing the relative performance of the models in representing the exact observed values, is defined as follows:

$$\text{RMSE} = \sqrt{\frac{\sum_{i=1}^M \sum_{j=1}^T (y_{ij}^m - y_{ij}^o)^2}{N}} \quad (2)$$

where  $y_{ij}^m$  and  $y_{ij}^o$  are the GCMs simulations and observations at the  $j$ th time step in the  $i$ th grid, respectively,  $T$  is the total number of the time steps,  $M$  is the total number of the grid cells, and  $N$  is the total number of the data. The RMSE is a non-negative value and the GCM with the lowest RMSE has rank 1.

**Table 1** CMIP5 climate models evaluated in this study

Model name	Modeling center	Resolution (latitude × longitude)
ACCESS1.0	Commonwealth Scientific and Industrial Research Organization and Bureau of Meteorology, Australia	1.25° × 1.875°
ACCESS1.3		1.25° × 1.875°
CanCM4	National Center for Atmospheric Research, Canada	2.7906° × 2.8125°
CanESM2		2.7906° × 2.8125°
CMCC-CESM	Euro-Mediterranean Center on Climate Change, Italy	3.4431° × 3.75°
CMCC-CM		0.7484° × 0.75°
CMCC-CMS		3.7111° × 3.75°
CNRM-CM5	National Center for Meteorological Research/European Center for Research and Advanced Training in Scientific Computing, France	1.4008° × 1.40625°
CNRM-CM5-2		1.4008° × 1.40625°
FGOALS-s2.0	Atmospheric Sciences and Geophysical Fluid Dynamics/Institute of Atmospheric Physics, Chinese Academy of Sciences, China	1.6590° × 2.8125°
GFDL-CM2p1	NOAA Geophysical Fluid Dynamics Laboratory, USA	2.0225° × 2.5°
GFDL-CM3		2° × 2.5°
GFDL-ESM2G		2.0225° × 2°
GFDL-ESM2M		2.0225° × 2.5°
GISS-E2-H	NASA Goddard Institute for Space Studies, USA	2° × 2.5°
GISS-E2-H-CC		2° × 2.5°
GISS-E2-R		2° × 2.5°
GISS-E2-R-CC		2° × 2.5°
HadGEM2-AO	National Institute of Meteorological Research/Korea Meteorological Administration, Korea	1.25° × 1.875°
HadCM3	Met Office Hadley Center, UK	2.5° × 3.75°
HadGEM2-CC		1.25° × 1.25°
HadGEM2-ES		1.25° × 1.25°
INM-CM4	Institute for Numerical Mathematics, Russia	1.5° × 2°
IPSL-CM5A-LR	Institute Pierre-Simon Laplace, France	1.8947° × 3.75°
IPSL-CM5A-MR		1.2676° × 2°
IPSL-CM5B-LR		1.8947° × 3.75°
MIROC4h	Atmosphere and Ocean Research Institute (The University of Tokyo), National Institute for Environmental Studies, and Japan Agency for Marine-Earth Science and Technology, Japan	0.5616° × 0.5625°
MIROC5		1.4008° × 1.40625°
MIROC-ESM	Japan Agency for Marine-Earth Science and Technology, Atmosphere and Ocean Research Institute (The University of Tokyo), and National Institute for Environmental Studies, Japan	2.7906° × 2.8125°
MIROC-ESM-CHEM		2.7906° × 2.8125°
MPI-ESM-LR	Max Planck Institute for Meteorology, Germany	1.8653° × 1.875°
MPI-ESM-MR		1.8653° × 1.875°
MPI-ESM-P		1.8653° × 1.875°
MRI-CGCM3	Meteorological Research Institute, Japan	1.12148° × 1.125°
MRI-ESM1		1.12148° × 1.125°
NorESM1-M	Norwegian Climate Center, Norway	1.8947° × 2.5°
NorESM1-ME		1.8947° × 2.5°

### 3.2.3 Nash-Sutcliffe efficiency

The Nash-Sutcliffe efficiency (NSE) can evaluate the predictive performance of the models (Nash and Sutcliffe 1970). It reflects how well the plot of the model simulation versus observed data is close to the 1:1 line. The NSE is defined as follows:

$$\text{NSE} = 1 - \frac{\sum_{i=1}^M \sum_{j=1}^T (y_{ij}^m - y_{ij}^o)^2}{\sum_{i=1}^M \sum_{j=1}^T (y_{ij}^o - \bar{y}^o)^2} \quad (3)$$

The terms are defined in Eqs. (1) and (2). The NSE ranges from  $-\infty$  to 1, and negative values indicate that the variance of the model

errors exceeds the variance of the observations, and therefore, the predictive performance of the model is poor. In other words, a model with a negative NSE is weak in the prediction of the observed values. Note that the NSE is sensitive to extremes; hence, a negative NSE can indicate a poor performance of the model in simulation of the extreme values. The GCM with the highest NSE is designated as rank 1.

### 3.2.4 Linear correlation coefficient

The linear correlation coefficient ( $r$ ), the measure of the similarity between the model simulation and the observation, can be determined as follows:

$$r = \frac{\sum_{i=1}^M \sum_{j=1}^T (y_{ij}^m - \bar{y}^m) (y_{ij}^o - \bar{y}^o)}{\sqrt{\sum_{i=1}^M \sum_{j=1}^T (y_{ij}^m - \bar{y}^m)^2} \sqrt{\sum_{i=1}^M \sum_{j=1}^T (y_{ij}^o - \bar{y}^o)^2}} \quad (4)$$

All terms are defined previously. The value of  $r$  close to the unity indicates that the overall temporal change of the model simulations agrees with the observations. Hence, the larger the value of  $r$ , the lower the rank of the GCMs.

### 3.2.5 Kolmogorov-Smirnov statistic

The Kolmogorov-Smirnov test (KS, Kolmogorov 1933; Smirnov 1933) is a measure to determine the ability of the GCMs to represent the PDFs of the observed variables (CRU temperature and precipitation). The KS measures the maximum vertical distance between the Cumulative Distribution Functions (CDFs) of the model simulation and the observation. This statistic is sensitive to the median, variance, and the shape of the CDFs and can be expressed as follows:

$$KS = \max_y \{|F^m(y) - F^o(y)|\} \quad (5)$$

where  $F^m$  and  $F^o$  are the CDFs of the model simulations and observations, respectively. The GCM with the lowest value of KS has rank 1. For a better comparison of the GCM simulations and observations, the boxplots and the empirical CDFs (ECDFs) of the GCMs simulations and observations are provided (Figs. 3 and 7).

### 3.2.6 Sen's slope estimator

GCMs are expected to represent the long-term trends of the hydroclimatological variables. Hence, the Sen's slope estimator (Sen 1968) is used to assess the ability of the GCMs to represent the trend of the observations. The Sen's slope is a non-parametric method for estimating the magnitude of the trend in a time series, and it is robust against outliers. To calculate the Sen's slope for each GCM and the CRU data, first, the monthly data are averaged over all grids to obtain a single time series over Iran. Then, a set of linear slopes ( $S$ ) for the time series is determined as follows:

$$s_k = \frac{y_m - y_l}{m - l} \quad (6)$$

where  $s_k$  is the  $k$ th element of set  $S$ ,  $y_m$ , and  $y_l$  are the values of the variable at the  $m$ th and  $l$ th time steps, respectively,  $l$  and  $m$  are indices where  $1 \leq l < m \leq L$ , and  $L$  is the length of the time series. The Sen's slope is the median of  $S$ . To determine the significance of the trend, the non-parametric 95% confidence intervals of the Sen's slope are calculated by the method described by Sen (1968). The time series has a significant

upward trend if the lower and upper confidence intervals are both positive and has a significant downward trend if confidence intervals are negative. Absolute differences between the Sen's slopes of the GCMs simulations and the CRU time series are calculated and the GCM with the minimum absolute difference is labeled rank 1.

### 3.2.7 Taylor diagram

The Taylor diagram (Taylor 2001) can evaluate the performance of different models relative to observations. It can illustrate the statistical relationship between the model simulations and observations and represent three different statistics including the centered root mean squared error (CRMSE), the linear correlation coefficient, and the standard deviation. The means of the simulations and observations are removed before computing second-order statistics (Taylor 2001). The standard deviation of the observed times series is calculated as

$$sd^o = \sqrt{\frac{\sum_{i=1}^M \sum_{j=1}^T (y_{ij}^o - \bar{y}^o)^2}{N}} \quad (7)$$

and the standard deviation of the time series simulated by a GCM is given by

$$sd^m = \sqrt{\frac{\sum_{i=1}^M \sum_{j=1}^T (y_{ij}^m - \bar{y}^m)^2}{N}} \quad (8)$$

The correlation coefficient can be expressed as

$$r = \frac{\sum_{i=1}^M \sum_{j=1}^T (y_{ij}^m - \bar{y}^m) (y_{ij}^o - \bar{y}^o)}{N \cdot sd^o \cdot sd^m} \quad (9)$$

and the CRMSE as

$$RMSE = \sqrt{\frac{\sum_{i=1}^M \sum_{j=1}^T [(y_{ij}^m - \bar{y}^m) - (y_{ij}^o - \bar{y}^o)]^2}{N}} \quad (10)$$

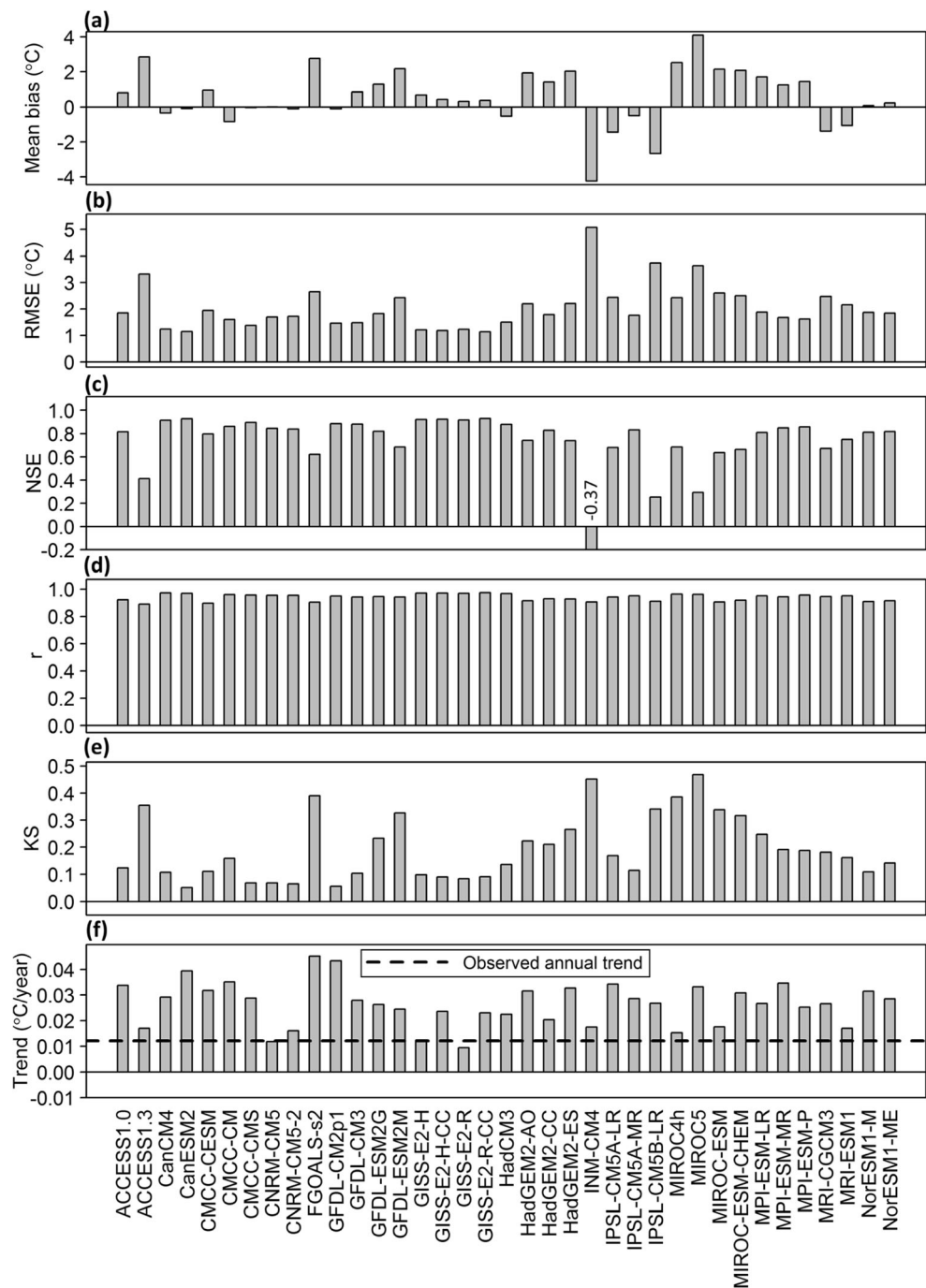
All terms in the above equations are defined in Eqs. (1), (2), and (3).

## 4 Results and discussion

The defined measures are used to evaluate the performance of the GCMs in simulating annual and seasonal temperature and precipitation as follows:

The measures of the performance (mean bias, RMSE, NSE,  $r$ , KS, and Sen's slope) are calculated for temperature. The results are shown in terms of the bar plot for annual temperature (Fig. 1) and seasonal temperature (Fig.

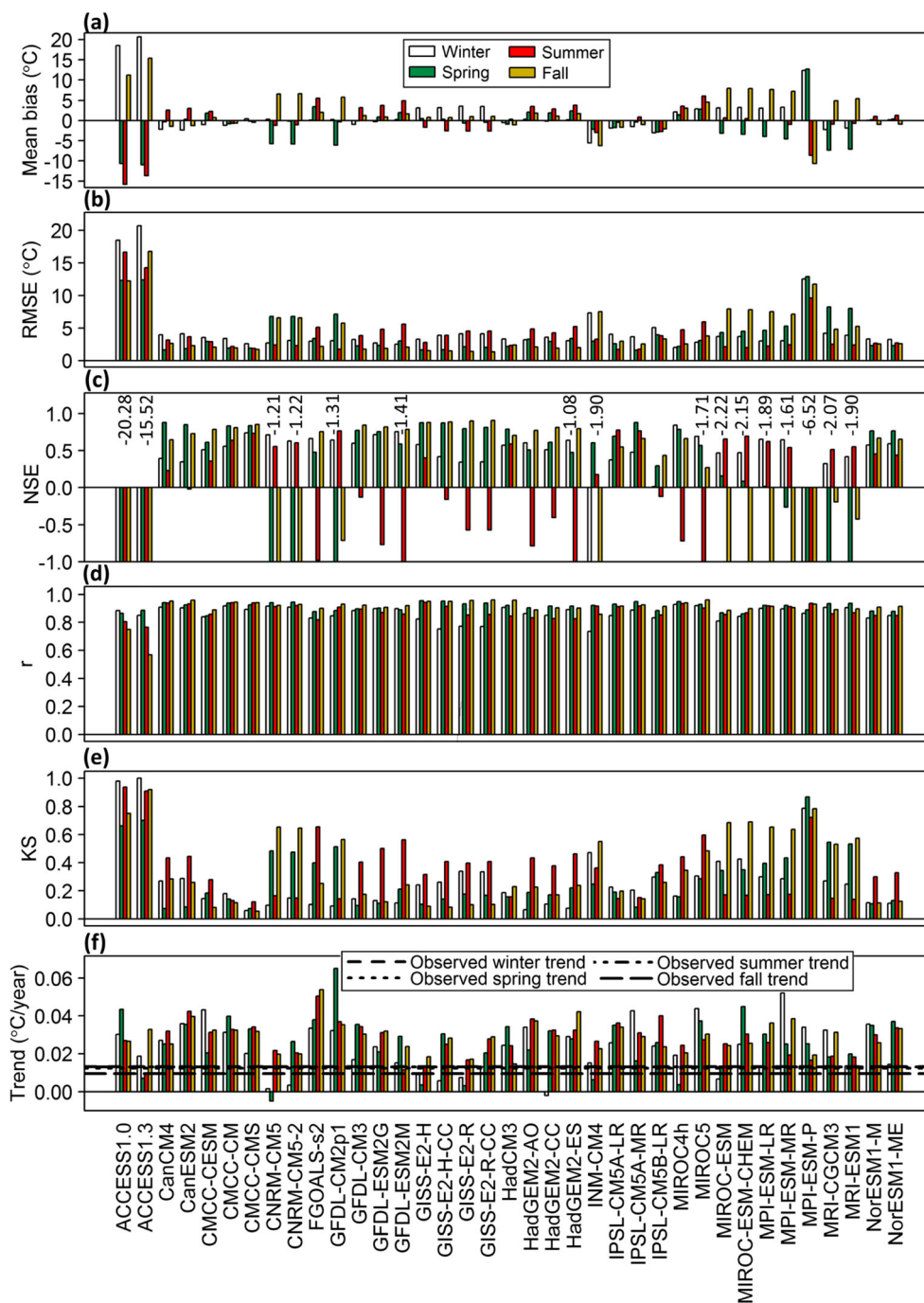
**Fig. 1** The measures of the performance for annual temperature. **a** Mean bias. **b** RMSE. **c** NSE. **d**  $r$ . **e** KS. **f** Sen's slope of the trend



2). The GCMs are ranked based on the values of the performance measures (Table 2). For the seasonal ranking, the sum of the ranks for all seasons is calculated for each GCM. Then, seasonal ranking is determined based on the sum of the ranks, such that the GCM with the minimum value of summation has rank 1. In other words, the best GCM at seasonal scale is the one with the overall best performance for all seasons. To have a more complete comparison between the GCMs' simulations and observations, the ECDFs and boxplots are illustrated (Fig. 3).

The mean bias of the majority of the GCMs for the annual temperature is less than 2 °C (Fig. 1a), which can indicate the high performance of the GCMs in simulating annual temperature. Figure 2a illustrates more detailed information about the GCMs' errors. Although some GCMs (e.g., the ACCESS1.0, CNRM-CM5, CNRM-CM5-2, and MPI-ESM-P) have relatively large positive and negative seasonal mean biases for temperature, they have small annual mean biases. The reason is that negative seasonal errors may cancel positive seasonal errors, and therefore, errors become relatively small for the

**Fig. 2** The measures of the performance for seasonal temperature **a** Mean bias. **b** RMSE. **c** NSE (for the GCMs that have at least one season with a NSE less than -1, the minimum value of seasonal NSE is written). **d**  $r$ . **e** KS. **f** Sen's slope of the trend



annual time scale. In addition, some GCMs (e.g., the GFDL-ESM2M, HadGEM2-AO, INM-CM4, IPSL-CM5B-LR, MIROC4h, and MIROC5) consistently overestimate or underestimate temperature in all seasons; therefore, they may have systematic errors in simulating temperature. As a result, assessing GCMs at both annual and intra-annual time scales may provide a better evaluation of the GCMs' performance. Based on the mean bias in the GCMs' rank at both annual and seasonal time scales (Table 2), the CMCC-CMS, NorESM1-M, and NorESM1-ME are defined as the best performing

GCMs. On the other hand, the ACCESS1.3, INM-CM4, and MIROC5 show a poor performance.

The average annual RMSE of the GCMs is 2.5 °C (Fig. 1b), i.e., the average variance of the GCMs' error in simulating annual temperature is 2.5 °C. The average seasonal RMSE of the GCMs (excluding the ACCESS1.0, ACCESS1.3, and MPI-ESM-P) is 3.5 °C (Fig. 2b). The average RMSEs can represent the overall errors of the GCMs in simulating temperature. The GISS-E2-H, CMCC-CMS, and GISS-E2-R-CC have the best and the ACCESS1.3, INM-CM4, and IPSL-



**Table 2** Rank of the GCMs for annual and seasonal temperature based on the mean bias, RMSE, NSE,  $r$ , KS, and the Sen's slope of the trend

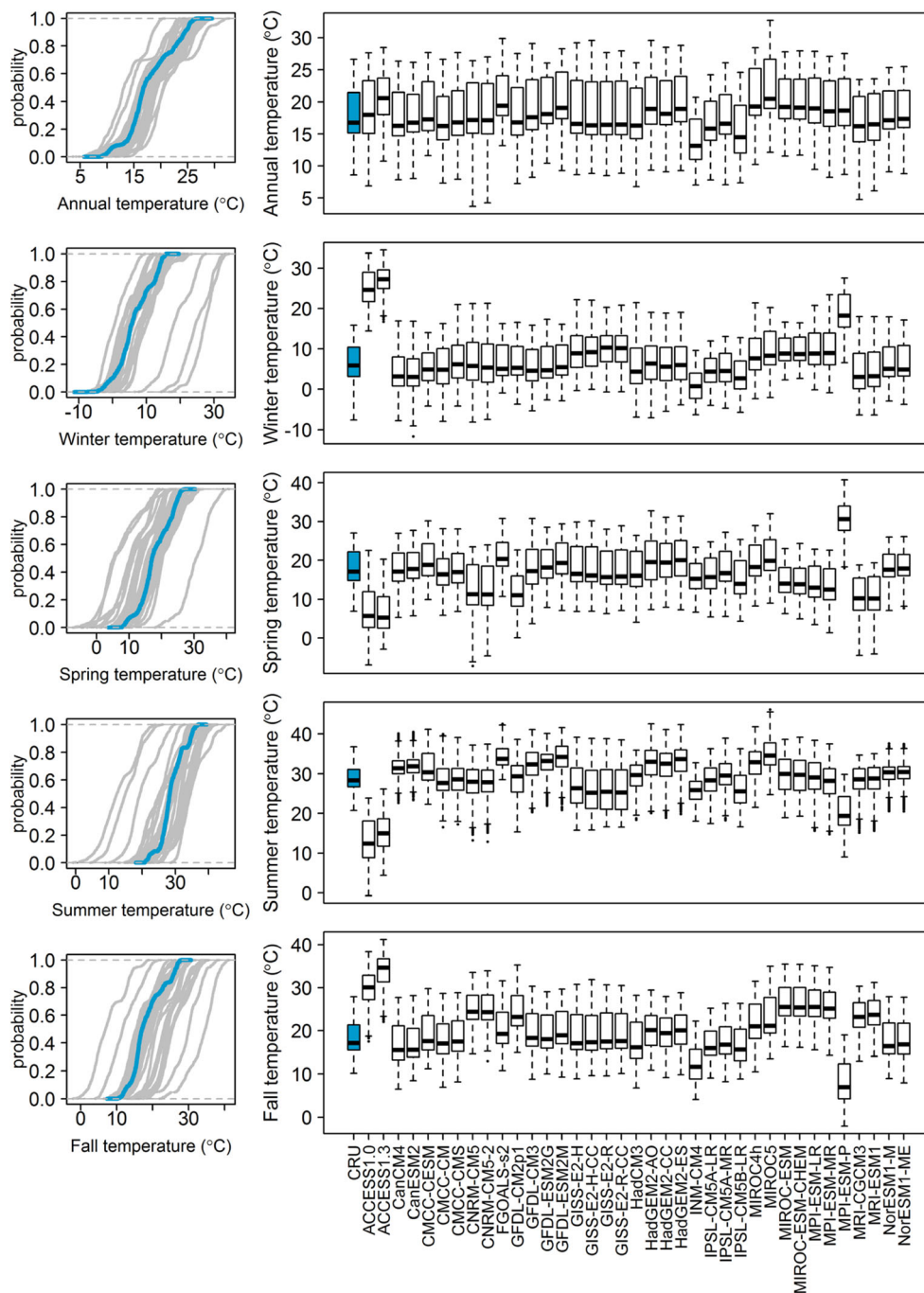
GCM	Measure of the performance											
	Mean bias		RMSE		NSE		$r$		KS		Sen's slope	
	Annual	Seasonal	Annual	Seasonal	Annual	Seasonal	Annual	Seasonal	Annual	Seasonal	Annual	Seasonal
ACCESS1.0	15	36	20	36	20	36	27	35	15	35	11	23
ACCESS1.3	35	37	34	37	34	37	37	34	33	37	31	24
CanCM4	9	14	6	17	6	17	2	1	11	17	5	8
CanESM2	4	15	2	18	2	18	6	6	1	23	25	18
CMCC-CESM	18	7	23	13	23	13	36	33	13	6	4	1
CMCC-CM	16	4	11	3	11	3	10	3	18	2	17	9
CMCC-CMS	2	1	7	1	7	1	11	9	4	1	7	6
CNRM-CM5	1	24	14	21	14	21	13	7	5	20	36	35
CNRM-CM5-2	5	20	15	26	15	26	14	4	3	24	33	32
FGOALS-s2	34	23	33	22	33	22	35	37	35	26	32	33
GFDL-CM2.1	6	16	8	20	8	20	18	24	2	15	27	12
GFDL-CM3	17	8	9	6	9	6	22	20	10	8	12	15
GFDL-ESM2G	21	9	18	7	18	7	19	22	26	12	19	3
GFDL-ESM2M	31	19	27	14	27	14	23	23	30	21	21	30
GISS-E2-H	14	12	4	2	4	2	3	8	9	7	35	36
GISS-E2-H-CC	11	10	3	8	3	8	4	11	7	13	22	26
GISS-E2-R	8	17	5	16	5	16	5	21	6	18	37	37
GISS-E2-R-CC	10	18	1	11	1	11	1	16	8	19	23	17
HadCM3	13	5	10	5	10	5	7	12	16	9	24	28
HadGEM2-AO	27	22	25	25	25	25	29	30	25	14	3	2
HadGEM2-CC	23	13	17	19	17	19	25	27	24	10	26	21
HadGEM2-ES	28	21	26	24	26	24	26	25	28	16	6	7
INM-CM4	37	34	37	33	37	33	34	26	36	31	29	29
IPSL-CM5A-LR	24	11	29	15	29	15	24	18	20	11	13	5
IPSL-CM5A-MR	12	6	16	4	16	4	15	10	14	3	9	4
IPSL-CM5B-LR	33	30	36	32	36	32	31	29	32	28	15	19
MIROC4h	30	29	32	30	32	30	33	36	31	32	28	27
MIROC5	29	31	31	29	31	29	28	31	29	33	1	25
MIROC-ESM	32	26	28	9	28	9	8	2	34	22	34	34
MIROC-ESM-CHEM	36	33	35	28	35	28	9	5	37	34	8	13
MPI-ESM-LR	26	27	22	23	22	23	17	13	27	30	16	20
MPI-ESM-MR	20	32	13	27	13	27	21	19	23	29	14	16
MPI-ESM-P	25	35	12	35	12	35	12	14	22	36	20	22
MRI-CGCM3	22	28	30	34	30	34	20	17	21	27	18	10
MRI-ESM1	19	25	24	31	24	31	16	15	19	25	30	31
NorESM1-M	3	2	21	10	21	10	32	32	12	4	2	14
NorESM1-ME	7	3	19	12	19	12	30	28	17	5	10	11

CM5B-LR have the poor performance to simulate temperature in annual and seasonal time scales (Table 2).

According to the annual NSE (Fig. 1c), the performance of the most GCMs for annual temperature is good, while the seasonal NSE (Fig. 2c) of the GCMs varies markedly. This difference can be well illustrated through the boxplots (Fig. 3). For example, for the ACCESS1.0, the annual NSE is 0.81,

which is a relatively high value, and the boxplot of its annual temperature is close to the boxplot of the CRU annual temperature. Conversely, the seasonal NSE values of this GCM are largely negative, and the boxplots of its seasonal simulations are entirely incompatible with those of the CRU ones for all seasons. Hence, better GCMs are those that have positive values for annual and seasonal NSEs. The best GCMs in terms

**Fig. 3** The ECDFs of the CRU observations (thick blue lines on the left plots), the ECDFs of the GCM simulations (thin gray lines on the left plots), the boxplots of the CRU observations (shaded boxes on the right plots), and the boxplots of the GCM simulations (blank boxes on the right plots) for annual (first row), winter (second row), spring (third row), summer (fourth row), and fall temperature (fifth row)



of the NSE for annual and seasonal time scales are the GISS-E2-H, CMCC-CMS, and GISS-E2-R-CC, while the ACCESS1.3, INM-CM4, and IPSL-CM5B-LR are the GCMs with the lowest performance (Table 2).

The correlation coefficient ( $r$ ) is relatively high for most GCMs at both annual and seasonal time scales (Figs. 1d and 2d). In other words, the GCMs' simulations can represent similar results to the observation for seasonal and annual temperature. Based on the annual and seasonal ranking of the GCMs (Table 2), the CanCM4, MIROC4h, and GISS-E2-H

have the best correlation with the observed temperature (the CMCC-CMS is among the top ten), while the FGOALS-s2, ACCESS1.3, and MIROC-ESM have the lowest correlation.

The KS shows that the simulated distribution of temperature varies significantly among the GCMs (Figs. 1e and 2e). This difference is also demonstrated by the boxplots, particularly for seasonal temperature (Fig. 3). Figure 3 illustrates that the CDFs of the CRU observation lie between those of the GCMs' simulations for annual and seasonal temperature. We can correct the biases of some GCMs' simulations (e.g., the

CMCC-CM, HadGEM2-CC, HadGEM2-ES, and MPI-ESM-LR) for annual temperature using the methods that can adjust only the mean of the simulations such as the additive linear scaling method. For some other GCMs, the variance scaling method can be proper to correct both the mean and variance of the simulations. The distribution of the temperature (observations and most simulations) is approximately normal with few extreme values (see Fig. 3); thus, most temperature simulations (but not those that have biases in their quantiles) can be corrected using simple bias correction methods. The CMCC-CMS, GISS-E2-H, and NorESM1-M are the best GCMs at the annual and seasonal scales, while the MIROC5, ACCESS1.3, and INM-CM4 show a poor performance in simulating temperature.

Trend, a long-term increase or decrease in values over time, is one of the most important properties of the time series in climate change analyses. The trend of the annual, winter, spring, summer, and fall temperature of the CRU dataset is 0.012, 0.013, 0.012, 0.013, and 0.009 °C per year, respectively (they are significant). There is also a significant upward trend in the annual and seasonal temperature of the most GCMs (Figs. 1f and 2f). None of the GCMs (excluding CNRM-CM5 and HadGEM2-CC for spring and winter, respectively) show a negative trend. The CMCC-CESM, HadGEM2-AO, CanCM4, and CMCC-CMS are the leading GCMs in the ranking for both annual and seasonal time scales, and the GISS-E2-R, GISS-E2-H, and CNRM-CM5 are at the end of the ranking.

The performance of the GCMs is further illustrated in Fig. 4 through the Taylor diagram for annual and seasonal temperature. Figure 4a illustrates that for the annual temperature, the CRMSE of the GCMs is less than 2 °C, the correlation between simulations and observed time series is larger than 0.9, and the standard deviations of the simulations are close to that of observations. Therefore, we can conclude that most GCMs can reasonably simulate annual temperature. Similarly, the Taylor diagrams show that most GCMs have a good performance in simulating spring and fall temperature (Fig. 4c, e). However, for winter and summer temperature (Fig. 4b, d), the two extreme seasons, most GCMs have a poor performance compared to spring and fall temperature. For winter temperature, many GCMs have correlations less than 0.9, and the standard deviations of the simulations are more spread out around the standard deviation of the observations. For summer temperature, most GCMs systematically overestimate the standard deviation. The CanCM4, CanESM2, CMCC-CM, CMCC-CMS, MIROC4h, and MIROC5 can be defined as the best GCMs in terms of the Taylor diagram. Note that the Taylor diagram assesses GCMs solely based on the three aforementioned statistics. Consequently, the final decision for the best GCM can be made by the users based on their priorities and performance criteria.

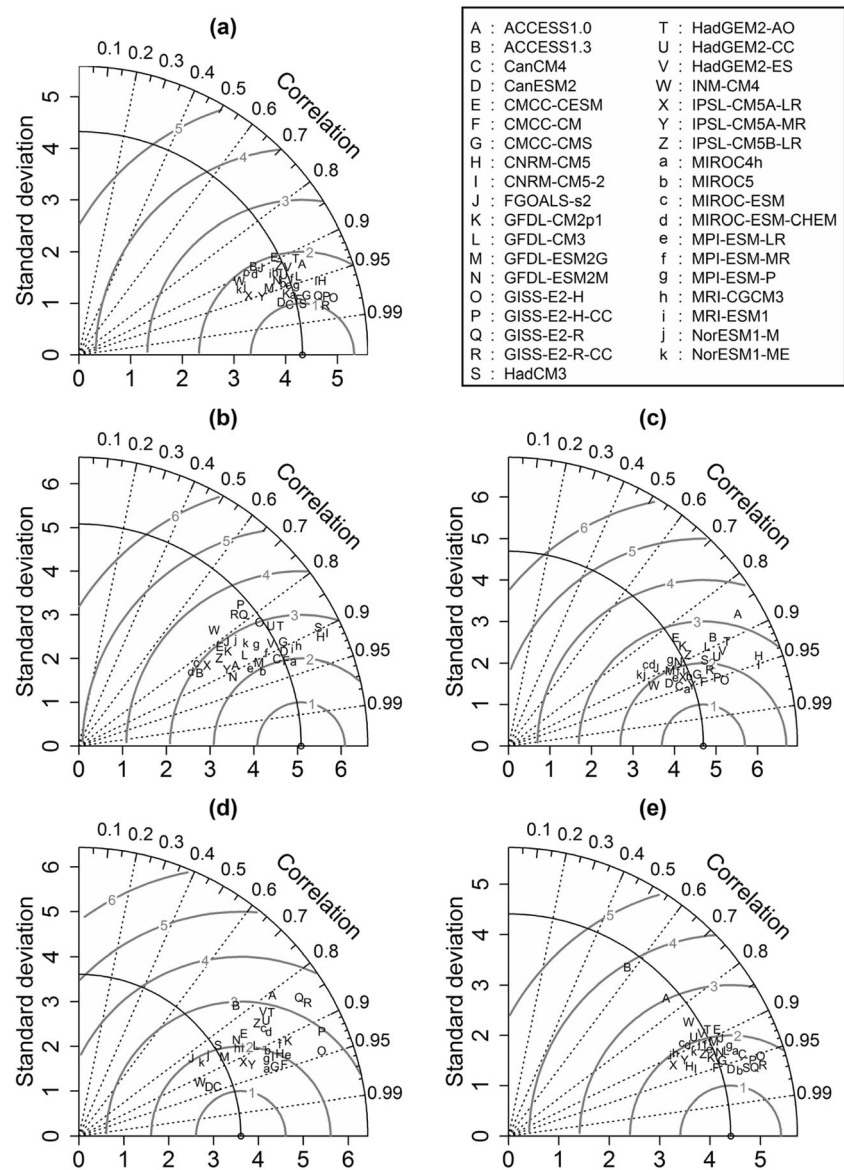
The measures of the performance for annual and seasonal precipitation are illustrated through bar plots in Figs. 5 and 6, respectively. Furthermore, the annual and seasonal ranking of the GCMs for each measure of the performance is presented in Table 3. To further investigate the performance of the GCMs in simulating annual and seasonal precipitation, the ECDFs and boxplots of the GCM simulations and the CRU observation are illustrated in Fig. 7. We can compare the mean, standard deviation, skewness, and extreme values of the GCMs' simulations and CRU observations using the ECDFs and boxplots.

Figure 5c shows that only 13 GCMs from 7 modeling groups have positive NSE for annual precipitation and, therefore, have a relatively good performance in simulating precipitation, while other GCMs have a poor performance over Iran, such that the mean annual precipitation of the CRU dataset is a better predictor than the GCMs' simulations. The result demonstrates that the performance of the GCMs in simulating temperature is much better than precipitation. The seasonal NSE shows that the performance of the GCMs is even more critical for the seasonal precipitation (Fig. 6c). The boxplots show that there is a large difference between the distribution of the simulations and observations, especially in the right tail of the distributions (Fig. 7). Precipitation is a skewed variable, but most GCMs overestimate or underestimate the skewness, and since the NSE is sensitive to extremes, it becomes negative for most GCMs. In terms of annual and seasonal NSE, The CMCC-CM, MRI-CGCM3, MRI-ESM1, and CMCC-CMS are the best performing GCMs, while ACCESS1.0, GISS-E2-R, and MIROC5 exhibit poor performance.

The average absolute mean bias of all GCMs for the annual precipitation is 37 mm, while the mean annual precipitation of the CRU dataset is 200 mm (Fig. 5a). Similar to the temperature simulations, some GCMs consistently overestimate or underestimate precipitation for all seasons, which can indicate a systematic bias in simulating physical processes of precipitation over Iran (Fig. 6a). For example, the IPSL-CM5A-LR underestimates temperature and precipitation, while the MIROC4h and MIROC5 overestimate those variables in all seasons. The GFDL-CM2p1, GFDL-ESM2G, NorESM1-M, and CMCC-CMS are the best GCMs based on the annual and seasonal mean bias ranking, and the ACCESS1.0, GISS-E2-R-CC, and MIROC4h have the largest biases (Table 3).

The average annual RMSE of all GCMs is 134 mm, which is a large value relative to the mean annual observed precipitation. Results reveal that the GCMs have relatively poor performance in representing annual precipitation values for Iran (Fig. 5b). According to Table 3, the CMCC-CM, MRI-CGCM3, MRI-ESM1, and CMCC-CMS are the best GCMs based on the annual and seasonal RMSE, and the ACCESS1.0, GISS-E2-R, and MIROC5 show a poor performance. Note that many GCMs have larger RMSEs for winter and spring precipitation (Fig. 6b) because the mean winter (87 mm) and the mean spring precipitation

**Fig. 4** The Taylor diagram of the GCMs simulations for **a** annual, **b** winter, **c** spring, **d** summer, and **e** fall temperature

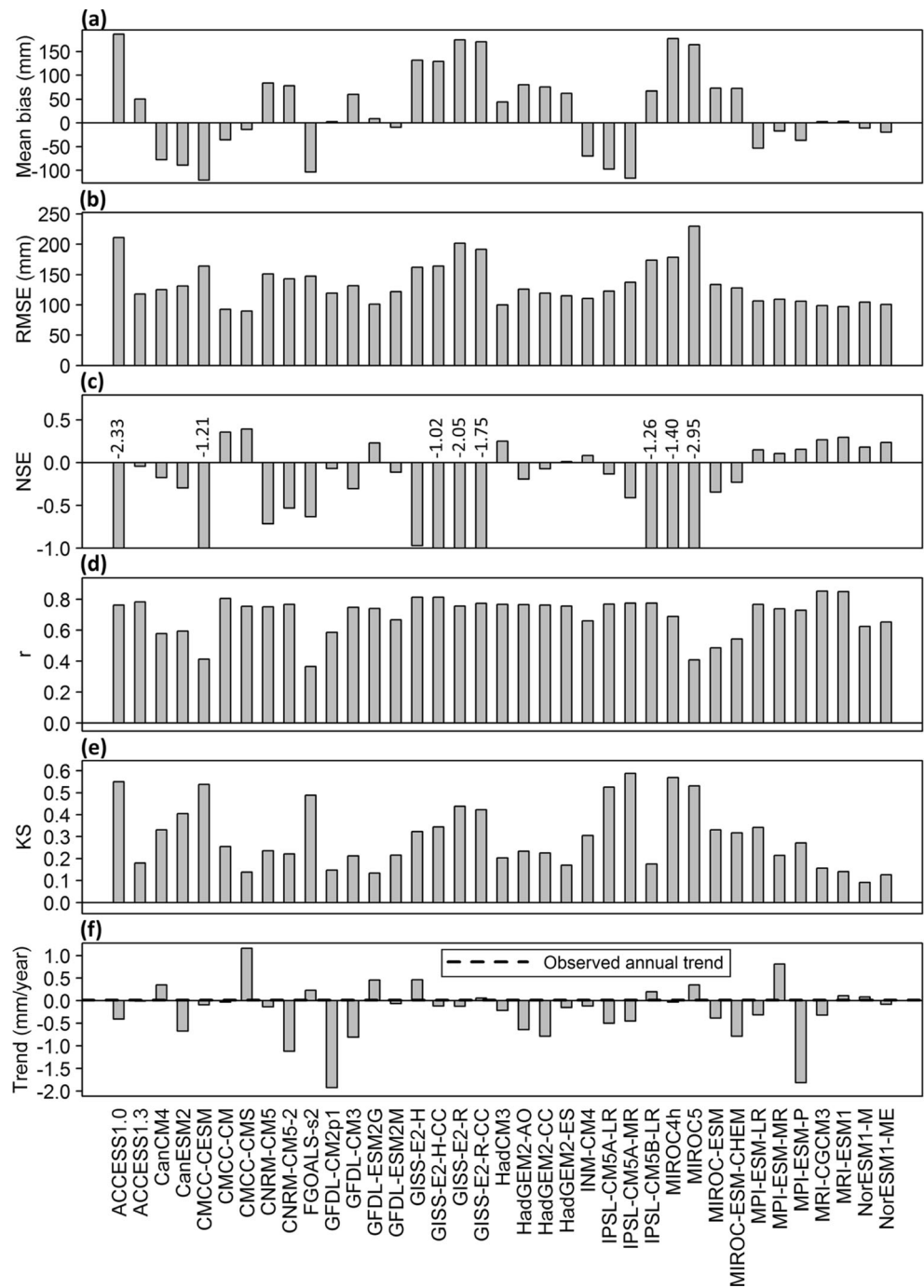


(78 mm) are larger than the mean summer (10 mm) and the mean fall precipitation (25 mm).

The average  $r$  is 0.70, 0.35, 0.54, 0.65, and 0.50 for the annual, winter, spring, summer, and fall GCMs' precipitation, respectively (Figs. 5d and 6d). The results can indicate that although most GCMs cannot properly represent observed precipitation values (according to the mean bias, RMSE, and NSE), they can approximately simulate the long-term time series (trend) of the observed annual and seasonal precipitation. The MRI-CGCM3, MRI-ESM1, and GISS-E2-H have the highest, and the FGOALS-s2, CMCC-CESM, and MIROC-ESM have the lowest correlations. The negative values of the NSE, relatively large values of the RMSE, and relatively low values of  $r$  indicate that the highly complex nonlinear processes of precipitation make it difficult to simulate.

The distribution of precipitation is important for various hydrological and climatological studies such as frequency analysis, forecasting, drought, and rainfall runoff. Figure 7 illustrates that the ECDF of the CRU precipitation is inside the envelope of the ECDFs of the GCMs for annual and seasonal precipitation. However, the boxplots of the GCMs simulations (Fig. 7) show that the GCMs cannot properly represent the distribution of the seasonal precipitation nor annual precipitation. The results confirm that the majority of the precipitation simulations need more complex bias correction methods than the temperature simulations. This conclusion is compatible with previous studies (e.g., Argüeso et al. 2013). The distribution mapping approach with the local intensity scaling method (for correcting the wet-day frequencies) can be useful for improving the GCM simulations. The annual and seasonal KS ranking of the GCMs (Table 3) shows that the annual and seasonal distributions of the CMCC-

**Fig. 5** The measures of the performance for annual precipitation. **a** Mean bias. **b** RMSE. **c** NSE. **d**  $r$ . **e** KS. **f** Sen's slope of the trend



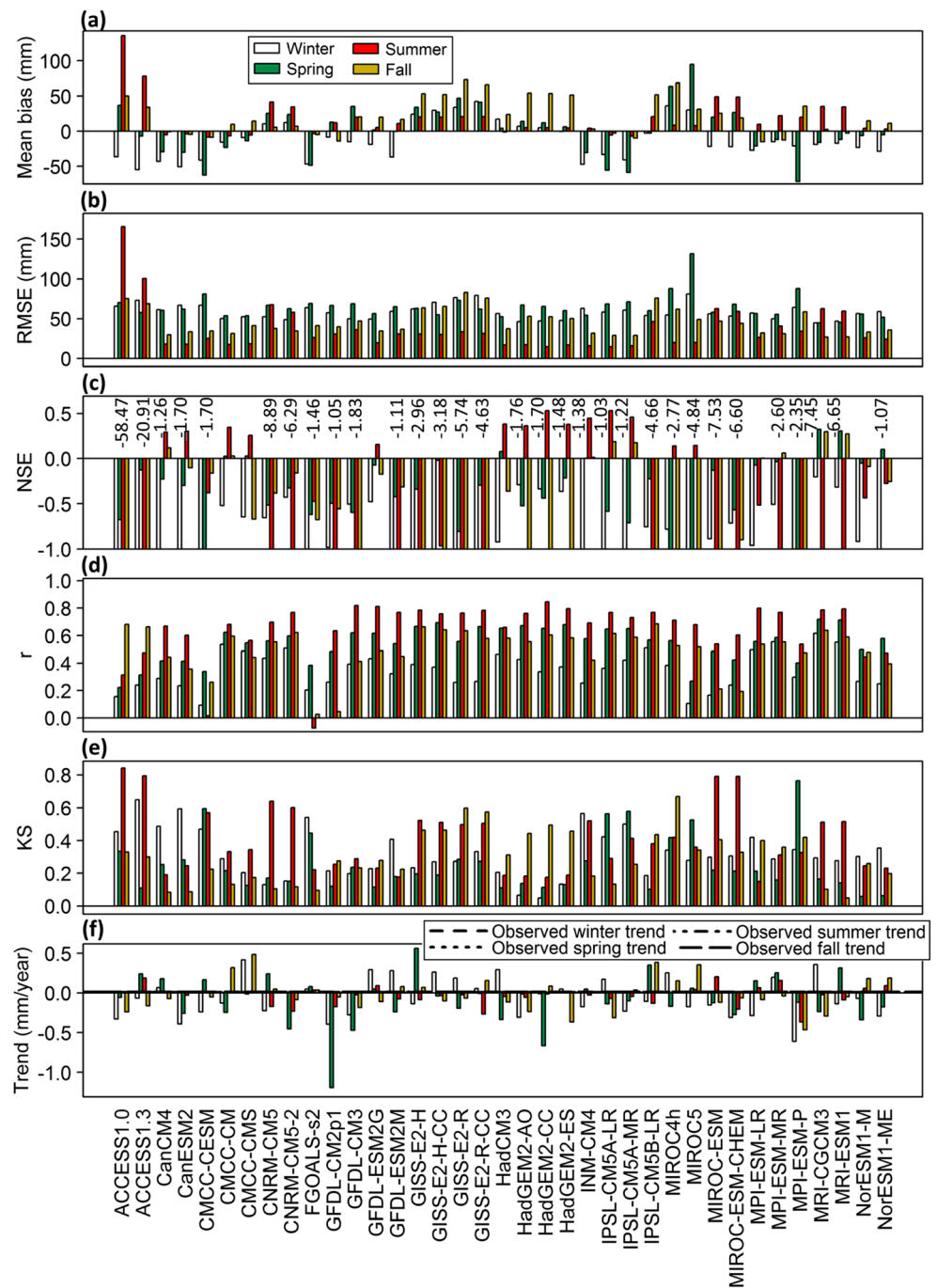
CMS, GFDL-ESM2G, and NorESM1-M are closest to those of the CRU dataset, and the distributions of the MIROC4h, ACCESS1.0, and IPSL-CM5A-MR have the largest difference from those of the observed precipitation.

The Sen's slope of the trend for the CRU precipitation is  $-0.35, 0.13, -0.27, -0.01,$  and  $-0.04$  mm/year for annual, winter, spring, summer, and fall precipitation, respectively. Note that all trends are insignificant. Similarly, most GCMs represent insignificant trends for annual and seasonal precipitation excluding the CanESM2 and MPI-ESM-P for winter, GFDL-CM2p1 and

MPI-ESM-P for spring, GISS-E2-R-CC and IPSL-CM5A-LR for summer, and CMCC-CM and CMCC-CMS for fall. The ACCESS1.0, MRI-CGCM3, and HadCM3 are at the top of the ranking for trend, while the MPI-ESM-P, CNRM-CM5-2, and MIROC5 are at the bottom.

To better evaluate the performance of the GCMs relative to the CRU observations, the Taylor diagrams for annual and seasonal precipitation are illustrated in Fig. 8. Precipitation simulations are more spread out on the Taylor diagram compared to the temperature simulations (Figs. 4 and 8). This can

**Fig. 6** The measures of the performance for seasonal precipitation. **a** Mean bias. **b** RMSE. **c** NSE (for the GCMs that have at least one season with a NSE less than -1, the minimum value of seasonal NSE is written). **d**  $r$ . **e** KS. **f** Sen's slope of the trend



confirm that the performance of the GCMs to simulate temperature is better than those of the precipitation. In addition, the diagram indicates that the GCMs from the same modeling group lie close together and behave similarly. The correlation of the simulations with observed annual precipitation varies between 0.4 and 0.7 (Fig. 8a), while this correlation is smaller for seasonal precipitation (Fig. 8b–e). According to the Taylor diagram, the CMCC-CM, CMCC-CMS, IPSL-CM5A-LR, IPSL-CM5A-MR, MRI-CGCM3, and MRI-ESM1 can be

identified as the best GCMs for annual and seasonal precipitation (Fig. 8).

### 5 Summary and concluding remarks

GCMs are the vital tool to simulate atmosphere-land-ocean circulations and to project climate change. Although many

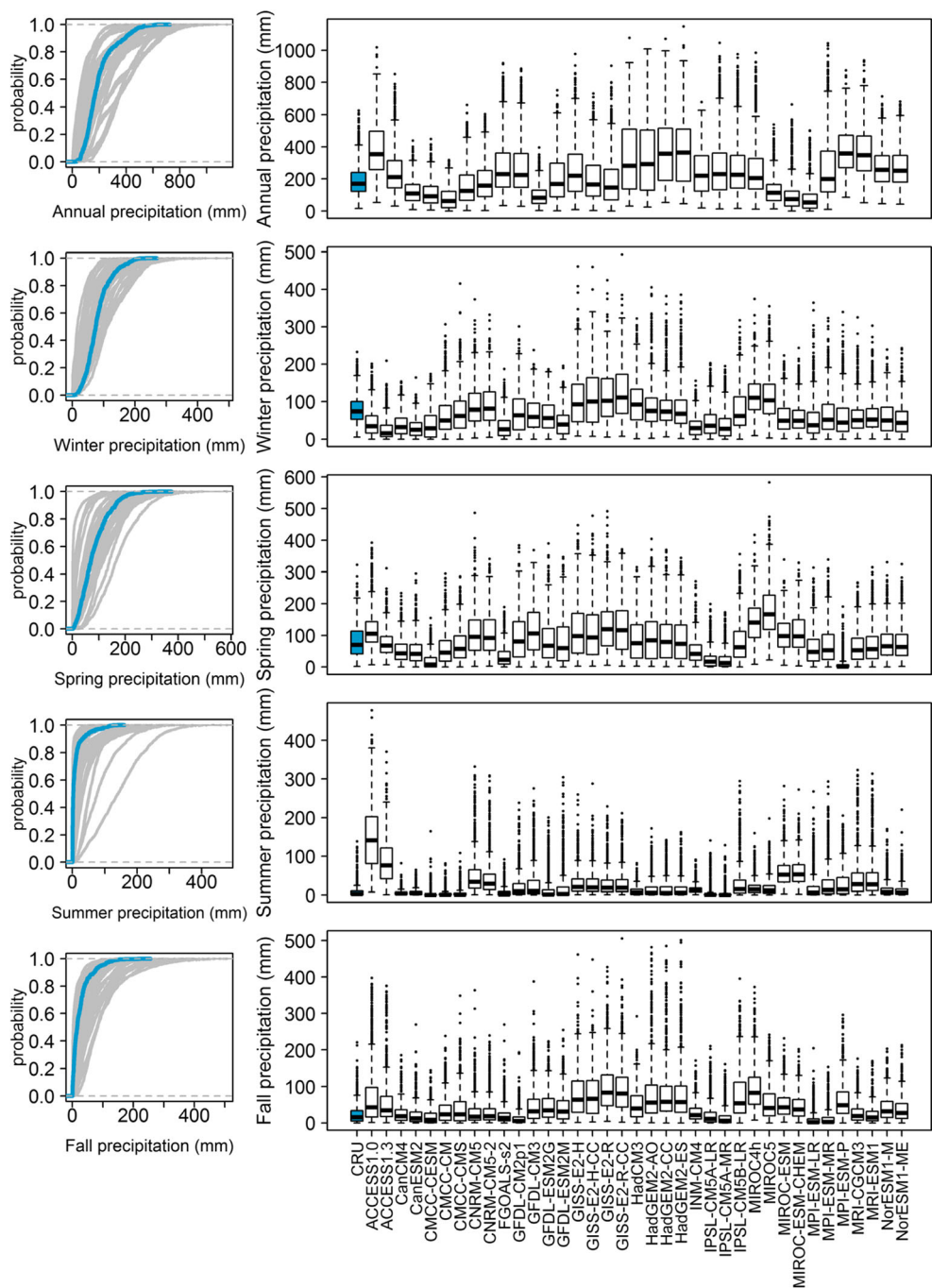
**Table 3** Rank of the GCMs for annual and seasonal precipitation based on the mean bias, RMSE, NSE,  $r$ , KS, and the Sen's slope of the trend

GCM	Measure of the performance											
	Mean bias		RMSE		NSE		$r$		KS		Sen's slope	
	Annual	Seasonal	Annual	Seasonal	Annual	Seasonal	Annual	Seasonal	Annual	Seasonal	Annual	Seasonal
ACCESS1.0	37	36	36	37	36	37	16	27	35	37	4	1
ACCESS1.3	13	33	14	34	14	34	6	25	10	27	20	27
CanCM4	23	16	19	14	19	14	32	26	24	16	29	32
CanESM2	26	19	22	19	22	19	30	34	28	22	18	5
CMCC-CESM	30	26	30	27	30	27	35	36	34	35	13	7
CMCC-CM	10	9	2	1	2	1	5	10	19	17	17	13
CMCC-CMS	7	5	1	7	1	7	19	23	4	3	36	14
CNRM-CM5	25	17	28	24	28	24	20	19	18	12	9	25
CNRM-CM5-2	22	15	26	18	26	18	13	7	15	11	31	34
FGOALS-s2	28	22	27	30	27	30	37	37	31	21	28	15
GFDL-CM2.1	1	7	15	20	15	20	31	31	6	6	37	11
GFDL-CM3	15	24	23	23	23	23	21	18	12	14	25	35
GFDL-ESM2G	4	6	7	4	7	4	22	12	3	4	32	17
GFDL-ESM2M	5	18	17	21	17	21	26	22	14	15	16	22
GISS-E2-H	32	31	29	31	29	31	3	3	23	26	33	8
GISS-E2-H-CC	31	29	31	25	31	25	4	8	27	24	12	31
GISS-E2-R	35	35	35	36	35	36	18	20	30	30	10	3
GISS-E2-R-CC	34	37	34	33	34	33	9	16	29	34	21	26
HadCM3	12	1	5	5	5	5	11	17	11	1	6	4
HadGEM2-AO	24	11	20	16	20	16	14	11	17	5	15	29
HadGEM2-CC	21	8	16	12	16	12	15	5	16	2	24	10
HadGEM2-ES	16	3	13	8	13	8	17	6	8	9	8	19
INM-CM4	18	20	12	6	12	6	27	24	21	29	11	9
IPSL-CM5A-LR	27	21	18	10	18	10	10	13	32	23	7	12
IPSL-CM5A-MR	29	25	25	17	25	17	7	14	37	32	5	18
IPSL-CM5B-LR	17	13	32	28	32	28	8	4	9	13	27	36
MIROC4h	20	27	24	22	24	22	34	35	25	31	2	24
MIROC5	19	28	21	26	21	26	33	33	22	28	22	20
MIROC-ESM	36	34	33	29	33	29	25	21	36	36	19	2
MIROC-ESM-CHEM	33	30	37	32	37	32	36	32	33	25	30	33
MPI-ESM-LR	14	23	10	15	10	15	12	15	26	20	3	28
MPI-ESM-MR	8	12	11	9	11	9	23	9	13	19	34	16
MPI-ESM-P	11	32	9	35	9	35	24	29	20	33	35	37
MRI-CGCM3	2	14	4	2	4	2	1	1	7	18	1	6
MRI-ESM1	3	10	3	3	3	3	2	2	5	10	26	23
NorESM1-M	6	4	8	11	8	11	29	28	1	8	23	21
NorESM1-ME	9	2	6	13	6	13	28	30	2	7	14	30

studies use GCMs for the climate change research and impact assessment, there is no credibility evaluation of the GCMs' simulations over Iran. This work is the first study that evaluated the performance of the GCMs' simulations (temperature and precipitation) including 37 CMIP5 GCMs (Table 1) with seven measures of the performance (mean bias, RMSE, NSE,  $r$ , KS, Sen's slope estimator, and Taylor diagram) over Iran.

The annual and seasonal analysis of the results showed that the GCMs are able to simulate annual and seasonal temperature better than precipitation over Iran. This can reveal the necessity of bias correction prior to using GCM outputs, in particular for the precipitation values. The results can help to understand the deficiencies of the GCMs in simulating temperature and

**Fig. 7** The ECDFs of the CRU observations (thick blue lines on the left plots), the ECDFs of the GCMs simulations (thin gray lines on the left plots), the boxplots of the CRU observations (shaded boxes on the right plots), and the boxplots of the GCMs simulations (blank boxes on the right plots) for annual (first row), winter (second row), spring (third row), summer (fourth row), and fall precipitation (fifth row)



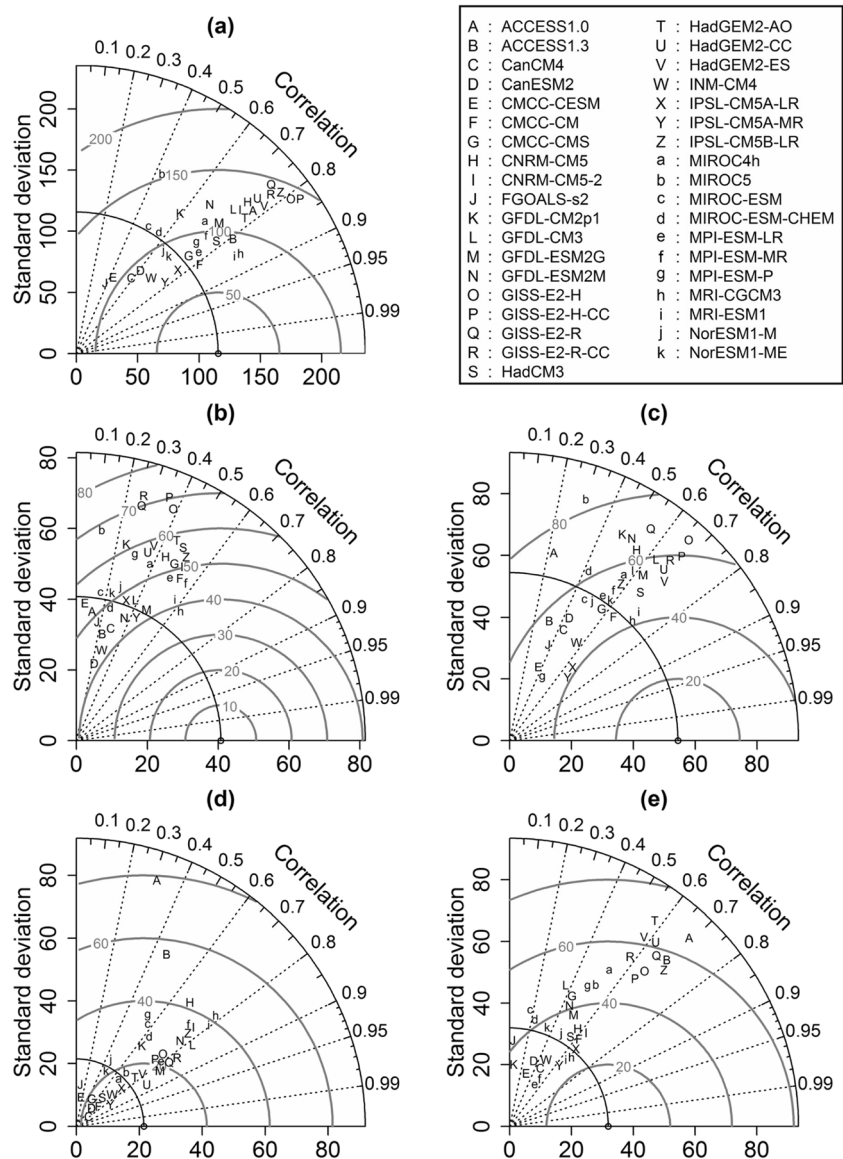
precipitation in order to find a proper bias correction method. The findings can be useful for the climate and process modeling groups who develop and modify the models.

In addition, results indicated that for temperature simulations, the CMCC-CMS has the best performance among all GCMs, while the ACCESS1.3 and INM-CM4 show a poor performance constantly. For precipitation, the MRI-CGCM3 is a top-ranking model based on the most performance measures, while MIROC5 and

ACCESS1.0 exhibit poor performance. Note that the CMCC-CMS, found as the best performing GCM for temperature, has the fourth rank in the mean bias, RMSE, and NSE, and the first rank in the KS for precipitation; therefore, we can consider this model as the best GCM in simulating both precipitation and temperature over Iran. Many authors believe that single models are unreliable (Weigel et al. 2008) and multi-model ensembles are superior to individual GCMs (Miao et al. 2014) since the ensemble can consider different



**Fig. 8** The Taylor diagram of the GCMs simulations for **a** annual, **b** winter, **c** spring, **d** summer, and **e** fall precipitation



information with possible uncertainties (Pincus et al. 2008). Hence, we suggest that researchers and users apply ensemble methods on the best performing GCMs based on our results.

The results can assist climatologists, hydrologists, and water resources managers to choose suitable GCMs for their related applications. Although this study ranked the GCMs for Iran, the results can be useful for other locations that have the same physical features (e.g., topography and climate) as Iran.

**Acknowledgements** The authors would like to thank the CMIP5 climate modeling groups (Table 1) and the Climatic Research (CRU) from the University of East Anglia for making their products publicly available. We recognize the UNESCO Chair in Water and Environment Management for Sustainable Cities, Sharif University of Technology for the help in this work. We are thankful to the reviewer whose comments and suggestions improved the paper.

## References

- Abbasnia M, Toros H (2016) Future changes in maximum temperature using the statistical downscaling model (SDSM) at selected stations of Iran. *Model Earth Syst Environ* 2:68. <https://doi.org/10.1007/s40808-016-0112-z>
- Abbaspour KC, Faramarzi M, Ghasemi SS, Yang H (2009) Assessing the impact of climate change on water resources in Iran. *Water Resour Res* 45:W10434. <https://doi.org/10.1029/2008wr007615>
- Aloysius NR, Sheffield J, Saiers JE, Li H, Wood EF (2016) Evaluation of historical and future simulations of precipitation and temperature in Central Africa from CMIP5 climate models. *J Geophys Res Atmos* 121:130–152. <https://doi.org/10.1002/2015jd023656>
- Argüeso D, Evans JP, Fita L (2013) Precipitation bias correction of very high resolution regional climate models. *Hydrol Earth Syst Sci* 17: 4379–4388. <https://doi.org/10.5194/hess-17-4379-2013>
- Belda M, Holtanová E, Halenka T, Kalvová J, Hlávka Z (2015) Evaluation of CMIP5 present climate simulations using the

- Köppen-Trewartha climate classification. *Clim Res* 64:201–212. <https://doi.org/10.3354/cr01316>
- Bonsal BR, Prowse TD (2006) Regional assessment of GCM-simulated current climate over Northern Canada. *Arctic* 59:15–128. <https://doi.org/10.14430/arctic335>
- Chen L, Frauenfeld OW (2014) A comprehensive evaluation of precipitation simulations over China based on CMIP5 multimodel ensemble projections. *J Geophys Res Atmos* 119:5767–5786. <https://doi.org/10.1002/2013jd021190>
- Dessai S (2005) Limited sensitivity analysis of regional climate change probabilities for the 21st century. *J Geophys Res Atmos* 110: D19108. <https://doi.org/10.1029/2005jd005919>
- Gleckler PJ, Taylor KE, Doutriaux C (2008) Performance metrics for climate models. *J Geophys Res Atmos* 113:D06104. <https://doi.org/10.1029/2007jd008972>
- Gohari A, Eslamian S, Abedi-Koupaei J, Massah Bavani A, Wang D, Madani K (2013) Climate change impacts on crop production in Iran's Zayandeh-Rud River Basin. *Sci Total Environ* 442:405–419. <https://doi.org/10.1016/j.scitotenv.2012.10.029>
- Hannah L (2015) *Climate change biology*. Elsevier, Amsterdam
- Hao Z, AghaKouchak A, Phillips TJ (2013) Changes in concurrent monthly precipitation and temperature extremes. *Environ Res Lett* 8:034014. <https://doi.org/10.1088/1748-9326/8/3/034014>
- Harris I, Jones P, Osborn T, Lister D (2013) Updated high-resolution grids of monthly climatic observations—the CRU TS3.10 Dataset. *Int J Climatol* 34:623–642. <https://doi.org/10.1002/joc.3711>
- Hashemi H, Uvo CB, Berndtsson R (2015) Coupled modeling approach to assess climate change impacts on groundwater recharge and adaptation in arid areas. *Hydrol Earth Syst Sci* 19:4165–4181. <https://doi.org/10.5194/hess-19-4165-2015>
- IPCC (1996) *Climate Change 1995: Impacts, adaptations, and mitigation of climate change: scientific-technical analyses*. Contribution of Working Group II to the second Assessment Report of the Intergovernmental Panel on Climate Change [Watson RT, Zinyowera MC, Moss RH (eds.)]. Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA
- IPCC (2007) *Climate models and their evaluation*. In: *Climate Change 2007: The Physical Science Basis*. Contribution of Working Group I to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change [Solomon S, Qin D, Manning M, Chen Z, Marquis M, Averyt KB, Tignor M, Miller HL (eds.)]. Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA
- IPCC (2013) *Summary for policymakers*. In: *Climate Change 2013: The Physical Sciences Basis*. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change [Stocker TF, Qin, Plattner GK, Tignor M, Allen SK, Boschung J, Nauels A, Xia Y, Bex V, Midgley BM (eds.)]. Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA
- Johnson F, Sharma A (2009) Measurement of GCM skill in predicting variables relevant for hydroclimatological assessments. *J Clim* 22: 4373–4382. <https://doi.org/10.1175/2009jcli2681.1>
- Kharin VV, Zwiers FW, Zhang X, Hegerl GC (2007) Changes in temperature and precipitation extremes in the IPCC Ensemble of Global Coupled Model Simulations. *J Clim* 20:1419–1444. <https://doi.org/10.1175/jcli4066.1>
- Khazaei MR, Zahabiyou B, Saghafian B (2011) Assessment of climate change impact on floods using weather generator and continuous rainfall-runoff model. *Int J Climatol* 32:1997–2006. <https://doi.org/10.1002/joc.2416>
- Knutti R, Masson D, Gettelman A (2013) Climate model genealogy: generation CMIP5 and how we got there. *Geophys Res Lett* 40: 1194–1199. <https://doi.org/10.1002/grl.50256>
- Kolmogorov AN (1933) Sulla determinazione empirica di una legge di distribuzione. *Giornale dell'Istituto Italiano degli Attuari* 4:83–91
- Loukas A, Vasiladias L, Tzabiras J (2008) Climate change effects on drought severity. *Adv Geosci* 17:23–29. <https://doi.org/10.5194/adgeo-17-23-2008>
- McMahon TA, Peel MC, Karoly DJ (2015) Assessment of precipitation and temperature data from CMIP3 global climate models for hydrologic simulation. *Hydrol Earth Syst Sci* 19:361–377. <https://doi.org/10.5194/hess-19-361-2015>
- Miao C, Duan Q, Sun Q, Huang Y, Kong D, Yang T, Ye A, di Z, Gong W (2014) Assessment of CMIP5 climate models and projected temperature changes over Northern Eurasia. *Environ Res Lett* 9:055007. <https://doi.org/10.1088/1748-9326/9/5/055007>
- Nash J, Sutcliffe J (1970) River flow forecasting through conceptual models part I—a discussion of principles. *J Hydrol* 10:282–290. [https://doi.org/10.1016/0022-1694\(70\)90255-6](https://doi.org/10.1016/0022-1694(70)90255-6)
- Nasrollahi N, AghaKouchak A, Cheng L, Damberg L, Phillips T, Miao C et al (2015) How well do CMIP5 climate simulations replicate historical trends and patterns of meteorological droughts? *Water Resour Res* 51:2847–2864. <https://doi.org/10.1002/2014wr016318>
- Nazemosadat MJ, Ravan V, Kahya E, Ghaedamini H (2016) Projection of temperature and precipitation in southern Iran using ECHAM5 simulations. *Iran J Sci Technol Trans A Sci* 40:39–49. <https://doi.org/10.1007/s40995-016-0009-8>
- Pascale S, Lucarini V, Feng X, Porporato A, Hasson SU (2014) Analysis of rainfall seasonality from observations and climate models. *Clim Dyn* 44:3281–3301. <https://doi.org/10.1007/s00382-014-2278-2>
- Perkins SE, Pitman AJ, Holbrook NJ, McAneney J (2007) Evaluation of the AR4 climate models' simulated daily maximum temperature, minimum temperature, and precipitation over Australia using probability density functions. *J Clim* 20:4356–4376. <https://doi.org/10.1175/jcli4253.1>
- Pincus R, Batstone CP, Hofmann RJP, Taylor KE, Glecker PJ (2008) Evaluating the present-day simulation of clouds, precipitation, and radiation in climate models. *J Geophys Res* 113:D14209. <https://doi.org/10.1029/2007JD00933>
- Razmara P, Massah Bavani AR, Motiee H, Torabi S, Lotfi S (2013) Investigating uncertainty of climate change effect on entering runoff to Urmia Lake Iran. *Hydrol Earth Syst Sci Discuss* 10:2183–2214. <https://doi.org/10.5194/hessd-10-2183-2013>
- Reichler T, Kim J (2008) Supplement to how well do coupled models simulate Today's climate? *Bull Am Meteorol Soc* 89:S1–S6. <https://doi.org/10.1175/bams-89-3-reichler>
- Reifen C, Toumi R (2009) Climate projections: past performance no guarantee of future skill? *Geophys Res Lett* 36:L13704. <https://doi.org/10.1029/2009gl038082>
- Samadi S, Carbone GJ, Mahdavi M, Sharifi F, Bihamta MR (2012) Statistical downscaling of climate data to estimate streamflow in a semi-arid catchment. *Hydrol Earth Syst Sci Discuss* 9:4869–4918. <https://doi.org/10.5194/hessd-9-4869-2012>
- Samadi S, Wilson CA, Moradkhani H (2013) Uncertainty analysis of statistical downscaling models using Hadley Centre Coupled Model. *Theor Appl Climatol* 114:673–690. <https://doi.org/10.1007/s00704-013-0844-x>
- Sayari N, Bannayan M, Alizadeh A, Farid A (2012) Using drought indices to assess climate change impacts on drought conditions in the northeast of Iran (case study: Kashaftood basin). *Meteorol Appl* 20: 115–127. <https://doi.org/10.1002/met.1347>
- Sen PK (1968) Estimates of the regression coefficient based on Kendall's tau. *J Am Stat Assoc* 63:1379–1389. <https://doi.org/10.1080/01621459.1968.10480934>
- Shadkam S, Ludwig F, Van Vliet MT, Pastor A, Kabat P (2016) Preserving the world second largest hypersaline lake under future irrigation and climate change. *Sci Total Environ* 559:317–325. <https://doi.org/10.1016/j.scitotenv.2016.03.190>
- Sillmann J, Kharin VV, Zhang X, Zwiers FW, Bronaugh D (2013) Climate extremes indices in the CMIP5 multimodel ensemble: part

1. Model evaluation in the present climate. *J Geophys Res Atmos* 118:1716–1733. <https://doi.org/10.1002/jgrd.50203>
- Smimov NV (1933) Estimate of deviation between empirical distribution functions in two independent sample (in Russian). *Bull Moscow Univ* 2:3–16
- Sonali P, Kumar DN, Nanjundiah RS (2016) Intercomparison of CMIP5 and CMIP3 simulations of the 20th century maximum and minimum temperatures over India and detection of climatic trends. *Theor Appl Climatol* 128:465–489. <https://doi.org/10.1007/s00704-015-1716-3>
- Taylor KE (2001) Summarizing multiple aspects of model performance in a single diagram. *J Geophys Res Atmos* 106:7183–7192. <https://doi.org/10.1029/2000jd900719>
- Taylor KE, Stouffer RJ, Meehl GA (2009) A summary of the CMIP5 experiment design. PCDMI Rep., 33 pp. [Available online at [http://cmip-pcmdi.llnl.gov/cmip5/docs/Taylor\\_CMIP5\\_design.pdf](http://cmip-pcmdi.llnl.gov/cmip5/docs/Taylor_CMIP5_design.pdf)]
- Taylor KE, Stouffer RJ, Meehl GA (2012) An overview of CMIP5 and the experiment design. *Bull Am Meteorol Soc* 93:485–498. <https://doi.org/10.1175/bams-d-11-00094.1>
- Wang L, Chen W (2013) A CMIP5 multimodel projection of future temperature, precipitation, and climatological drought in China. *Int J Climatol* 34:2059–2078. <https://doi.org/10.1002/joc.3822>
- Weigel AP, Liniger MA, Appenzeller C (2008) Can multi-model combination really enhance the prediction skill of probabilistic ensemble forecasts? *Q J R Meteorol Soc* 134:241–260. <https://doi.org/10.1002/qj.210>
- Yin L, Fu R, Shevliakova E, Dickinson RE (2012) How well can CMIP5 simulate precipitation and its controlling processes over tropical South America? *Clim Dyn* 41:3127–3143. <https://doi.org/10.1007/s00382-012-1582-y>
- Zarghami M, Abdi A, Babaeian I, Hassanzadeh Y, Kanani R (2011) Impacts of climate change on runoffs in East Azerbaijan, Iran. *Glob Planet Chang* 78:137–146. <https://doi.org/10.1016/j.gloplacha.2011.06.003>