



Assessment of CMIP6 models performance in simulation precipitation and temperature over Iran and surrounding regions

Mohammad Javad Zareian · Hossein Dehban · Alireza Gohari · Ali Torabi Haghighi

Received: 1 March 2024 / Accepted: 28 June 2024 / Published online: 4 July 2024
© The Author(s), under exclusive licence to Springer Nature Switzerland AG 2024

Abstract This study investigates the performance of CMIP6 models in reproducing historical temperature and precipitation data for Iran and neighboring countries (Afghanistan, Pakistan, Turkmenistan, Azerbaijan, Armenia, Turkey, and Iraq) from 1980 to 2014. Reanalysis data from the ECMWF database (ERA5) for temperature and precipitation were utilized as a reference for the period 1980-2014. Additionally, ten Atmosphere-Ocean General Circulation Models (AOGCMs) from CMIP6 were employed to simulate temperature and precipitation data for the study region based on the IPCC Sixth Assessment Report databases. The Kling-Gupta Efficiency (KGE)

index was used to evaluate the accuracy of CMIP6 models in replicating daily temperature and precipitation. The results indicate that different CMIP6 models exhibit varying degrees of accuracy in simulating historical temperatures and precipitation, depending on the month and the country. For instance, the IPSL-CM6A-LR model demonstrated the best annual performance in estimating temperature in Azerbaijan (KGE = 0.5), while the HadGEM3-GC31-LL model showed the lowest annual performance in Pakistan (KGE = -1.4). Interestingly, the models were found to be more accurate in simulating temperatures during warm months compared to cold ones. Furthermore, the accuracy of different models in estimating annual precipitation varied significantly, ranging from -0.64 (MRI-EMS2-0 model in Afghanistan) to 0.05 (CMCC-ESM2 model in Armenia). Similar to temperature, the study found that models were generally more accurate in simulating precipitation during cold months compared to warm ones.

M. Zareian (✉)
Department of Water Resources Study and Research,
Water Research Institute (WRI), Tehran, Iran
e-mail: mj.zareian@gmail.com; m.zareian@wri.ac.ir

H. Dehban
Department of Water Resources Study and Research,
Water Research Institute (WRI), Tehran, Iran
e-mail: dehbanh@gmail.com; hdehban@wri.ac.ir

A. Gohari
Department of Water Science and Engineering, College
of Agriculture, Isfahan University of Technology, Isfahan,
Iran
e-mail: ar.gohari@cc.iut.ac.ir

A. Torabi Haghighi
Water, Energy, and Environmental Engineering Research
Unit, University of Oulu, P.O. Box 4300, FIN-90014 Oulu,
Finland
e-mail: Ali.TorabiHaghighi@oulu.fi

Keywords Climate change · CMIP6 · ERA5 · KGE · Iran

Introduction

In recent decades, climate change has emerged as a major concern due to its impact on the earth's climate. Defined as long-term alterations in greenhouse gas concentrations, it significantly influences

climatic processes (Goyal, 2004; Werndl, 2016). These changes have demonstrably affected Earth's temperature and precipitation patterns (Solomon et al., 2007; Calzadilla et al., 2013). Consequently, various related aspects of human life have experienced modifications, including extreme weather events (floods, droughts, heatwaves) (Mal et al., 2018; Luber and McGeehin, 2008; Ahmed et al., 2018), storm patterns (Michener et al. 1997; Brooks, 2013), and crop water requirements (Gohari et al., 2013; Li et al., 2022).

Climate change poses a significant threat to water resources by significantly impacting hydrological cycles (Allan et al., 2013). Projections for future climate change indicate continued and substantial impacts on water resource systems. Ignoring these consequences is no longer an option (Pachauri et al., 2014). Therefore, developing climate-resilient water resource systems becomes increasingly crucial (Charlton and Arnell, 2011; Gohari et al., 2017; Zareian, 2021).

The Intergovernmental Panel on Climate Change (IPCC) highlights the significant threat climate change poses to water resources, particularly in arid and semi-arid regions (IPCC, 2021). Their report details the anticipated impacts on these regions, including changes in temperature, precipitation, evaporation, soil moisture, and river flow. This challenge is evident in the case of Iran and its neighbors, where managing shared water resources amidst droughts and climate change has become increasingly complex. These complexities underscore the need to investigate how climate change specifically affects temperature and precipitation in these shared watersheds. Such insights are crucial for informing effective policy guidance for future water resource management in the region.

Atmosphere- Ocean General Circulation Models (AOGCMs) represent the primary tools for studying the Earth's future climate in the context of climate change. These models employ intricate three-dimensional networks to represent each component of the Earth's climate system. Notably, AOGCMs are constructed by coupling together distinct models representing atmospheric and oceanic processes (IPCC, 2021). Building upon this approach, the Intergovernmental Panel on Climate Change (IPCC) released its sixth assessment report (AR6) on climate change, highlighting the critical role of

CMIP6 models (Shukla et al., 2019). Consequently, CMIP6 models stand as one of the most prominent tools for analyzing and understanding future climate change scenarios.

Research suggests that combining AOGCMs through group implementation offers advantages over using individual models (Zareian et al., 2015). This approach has been explored in various studies, demonstrating potential improvements in future climate change forecasts. For instance, Chen et al. (2017) investigated weighting methods for AOGCMs in hydrological studies. They compared five methods applied to 28 CMIP5 models in Canada, finding that combining AOGCMs could enhance future climate predictions. Similarly, Ahmed et al. (2019) evaluated 20 CMIP5 models based on their ability to simulate precipitation between 1961 and 2005. They employed various criteria, including spatial efficiency and skill scores, and compared two methods for integrating precipitation data from the selected models: Simple Mean (SM) and Random Forest Regression (RF). Their findings indicated superior performance when combining models using the Random Forest Regression method.

Studies have explored various methods to improve the accuracy of climate change projections using AOGCMs. Gohari et al. (2015) investigated a probabilistic method based on statistical distribution fitting, demonstrating its ability to enhance the accuracy of CMIP3 models for temperature and precipitation changes in central Iran. Tanveer et al. (2016) employed the Reliability Ensemble Averaging (REA) method to combine 18 CMIP5 models, reducing uncertainty in climate predictions for the Han River basin in South Korea. They used the root mean square error (RMSE) to evaluate uncertainty across different models and found REA averaging significantly improved accuracy. Xiong et al. (2021) compared the ability of 24 CMIP5 models to simulate air temperature in central Asia. They used spatial and temporal analysis, trend analysis, and empirical orthogonal functions (EOF) to identify the most accurate models. CNRM-CM5, GFDL-CM3, and GISS-E2-H emerged as the most accurate for this specific region. Liu et al. (2022) evaluated 22 CMIP6 models in simulating extreme precipitation over Central Asia using Taylor diagrams and DISO (Distance between Indices of Simulation and Observation). Their findings revealed uncertainties in all models

for simulating extreme precipitation patterns in early spring and late winter.

Previous research has established the capability of CMIP models to capture the general spatial distribution of temperature and precipitation across Central Asia (Xiong et al., 2021; Yanmin & Haomin, 2013). Analyses using CMIP5 models have shown success in replicating trends, temporal evolution, and annual mean surface air temperatures (Yanmin & Haomin, 2013). However, limitations have also been identified. Studies highlight the inability of CMIP5 models to capture regional details of climate change impacts and accurately describe low-scale variations (Xiong et al., 2021).

Simulating precipitation poses a greater challenge for CMIP models. While some models show promise in replicating the annual cycle shape in specific regions, difficulties arise in capturing peak precipitation and accurately simulating Probability Density Functions (PDFs). Additionally, there is a tendency for overestimation, particularly in arid regions (Guo et al., 2021). Recent work on CMIP6 models suggests progress in simulating extreme precipitation events during specific seasons. However, substantial biases remain, particularly in early spring and late winter (Liu et al., 2022). These biases highlight the ongoing need for model refinement.

This study addresses several key gaps and inconsistencies in the current literature. First, we will employ daily data, providing a more granular understanding of model performance compared to previous studies which have primarily focused on annual or seasonal means (Xiong et al., 2021; Yanmin & Haomin, 2013). Second, our research concentrates on Iran and its neighboring countries, allowing for a more in-depth analysis of a climatically diverse region, compared to existing research that often encompasses broader regions within Central Asia. Finally, by focusing on CMIP6, our study will contribute to understanding the effectiveness of the latest generation of climate models in this specific region, whereas prior studies acknowledge advancements in CMIP6 models but lack a dedicated evaluation for this area using daily data (Guo et al., 2021; Liu et al., 2022).

By addressing these gaps, this study will provide valuable insights into the ability of CMIP6 models to simulate temperature and precipitation patterns in a region of significant geopolitical and climatic

importance (Iran and its neighboring countries). Improved understanding of model strengths and weaknesses will inform future climate projections and impact assessments for the region.

Materials and Methods

Study Area

This study focuses on Iran and seven neighboring countries (Iraq, Turkey, Azerbaijan, Armenia, Afghanistan, Turkmenistan and Pakistan) (Figure 1). Since the study area includes several shared watersheds, any change in temperature and precipitation in these countries due to climate change may result in new conflicts regarding the water resources management.

Extraction of CMIP6 outputs

The official IPCC database (<https://www.ipcc.ch>) was used to extract the outputs of 10 CMIP6 models. Table 1 summarizes the characteristics of the selected models. To evaluate the accuracy of the CMIP6 models, the historical period 1980 to 2014 was used as a baseline period. Selecting these specific models involved careful consideration of two key criteria. Firstly, advanced parameterization schemes for both the atmosphere and land surface were sought to accurately represent regional processes like monsoonal circulation patterns (He et al., 2023). Secondary, we looked for models with demonstrated skill in simulating temperature and precipitation over similar geographic regions or climatic zones based on historical performance evaluations (Yanmin & Haomin, 2013; Guo et al., 2021).

Extraction of temperature and precipitation data from ERA5

The European Centre for Medium-Range Weather Forecasts (ECMWF) provides a valuable source of reanalysis meteorological data. Renowned for its accuracy, ECMWF offers the highly regarded ERA5 dataset, which merges observational and numerical model data. This dataset offers high temporal (hourly to monthly) and spatial (0.1 degrees) resolution (Hersbach et al., 2020). Given the limited

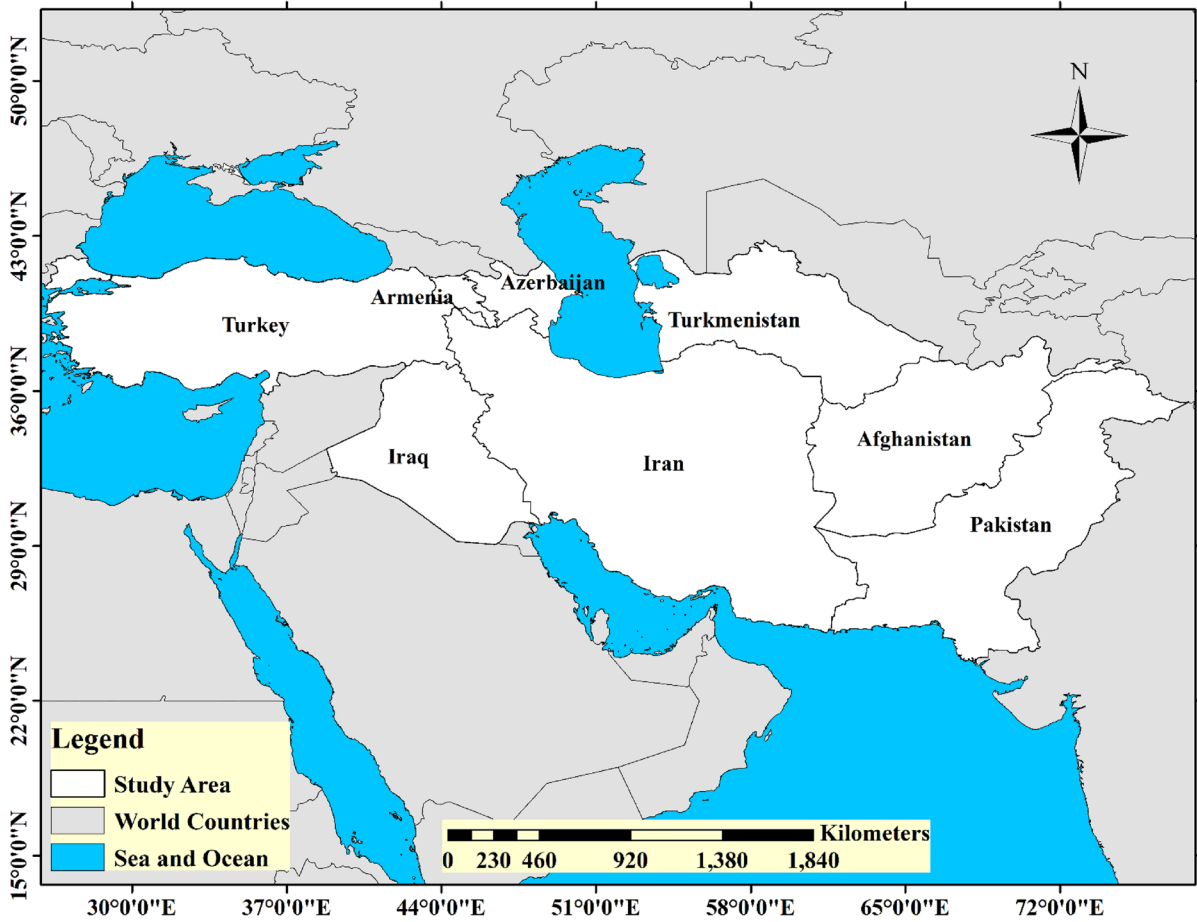


Figure 1 Location map of the study area

Table 1 Characteristics of selected CMIP6 models used in this study

Developer	Resolution	Model
Beijing Climate Center, China Meteorological Administration	1.12°×1.12°	BCC-CSM2-MR
Community Earth System Model Contributors	1.25°×0.94°	CESM2
Centre National de Recherches Météorologiques Scientifique (CNRM)	0.50°×0.50°	CNRM-CM6-1
Canadian Centre for Climate Modelling and Analysis-Canada	2.81°×2.81°	CanESM5
National Institute for Environmental Studies, The University of Tokyo	1.40°×1.40°	MIROC6
Meteorological Research Institute	1.12°×1.12°	MRI-EMS2-0
Institute Pierre-Simon Laplace	2.50°×1.27°	IPSL-CM6A-LR
Chinese Academy of Meteorological Sciences, China	1.12°×1.12°	CAMS-CSM1-0
Centro Euro-Mediterraneo sui Cambiamenti Climatici	1.25°×0.94°	CMCC-ESM2
NOAA-Geophysical Fluid Dynamics Laboratory	1.25°×1.00°	GFDL-ESM4
Meteorological Office Hadley Centre	1.88°×1.25°	HadGEM3-GC31-LL

density of meteorological stations in the study area, monthly temperature and precipitation data for the historical period 1980-2014 were retrieved from ERA5 on the ECMWF website (<https://www.ecmwf.int>) to serve as observed historical data. The selection of ERA5 dataset is motivated by two key considerations. Firstly, ERA5 has established credibility in representing climate variables like temperature and precipitation, particularly in regions with limited observational networks (Gomis-Cebolla et al., 2023). Studies have demonstrated strong agreement between ERA5 data and available in-situ observations in the study area (Song et al., 2022; Radmanesh et al., 2023). This established performance makes ERA5 a reliable alternative for evaluating climate model simulations. Secondly, the scarcity of high-quality, long-term observational meteorological data across Iran and neighboring countries presents a significant challenge.

Assessing the accuracy of CMIP6 models

The historical outputs of the chosen CMIP6 models can be employed to evaluate their accuracy in simulating future meteorological variables. This comparison allows us to assess how well each model replicates historical temperature and precipitation patterns by comparing its outputs against observed data from meteorological stations. In this study, the accuracy of the selected models was evaluated by comparing their historical outputs (1980-2014) with reanalyzed data from ERA5. The Kling-Gupta Combined Statistical Index (KGE) was used for this evaluation. The KGE index offers several advantages over simpler indices like Root Mean Squared Error (RMSE), Mean Squared Error (MSE), or the coefficient of determination (R^2). While RMSE, MSE, and R^2 are commonly used, they each focus on a single aspect of model performance. RMSE and MSE emphasize the magnitude of errors, and R^2 reflects the correlation between simulated and observed data. These indices fail to capture the multifaceted nature of model agreement, neglecting factors like mean bias and variability. Also, compared to dichotomous or probabilistic evaluations, the KGE index offers a single metric combining mean bias, variability agreement, and temporal correlation. This makes it easier to interpret and assess climate model accuracy in simulating precipitation

and temperature (Lamontagne et al., 2020). The KGE index can be calculated using Eqs. 1-4:

$$KGE = 1 - \sqrt{(r - 1)^2 + (\alpha - 1)^2 + (\beta - 1)^2} \tag{1}$$

$$\alpha = \frac{\sigma_s}{\sigma_o} \tag{2}$$

$$\beta = \frac{\mu_s}{\mu_o} \tag{3}$$

$$r = \frac{\sum(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{(\sum(x_i - \bar{x})^2)(\sum(y_i - \bar{y})^2)}} \tag{4}$$

where, x is the of historical precipitation or temperature that extracted from ERA5 dataset; y represents historical precipitation or temperature data that has been derived from the CMIP6 models; μ_s and μ_o are the averages of ERA5 and CMIP6 data, respectively; σ_s and σ_o are the standard deviations of the ERA5 and CMIP6 data, respectively; and r is the Pearson correlation coefficient (Knoben et al., 2019). KGE value ranges from $-\infty$ to 1, with values closer to 1 showing the strongest correlation between observational and simulated data (Patil and Stieglitz, 2015).

The complexity of calculations and the need for extensive pixel-based comparisons across a vast region necessitated the use of the R programming language for this analysis.

Results

The accuracy of CMIP6 models in simulating temperatures

Figure 2 depicts the KGE index, spatially distributed, which compares the accuracy of different CMIP6 models in estimating temperature across the study area. The BCC-CSM2-MR model shows fluctuating KGE values between 0 and 0.4 in most regions, except for eastern Afghanistan and western Turkey, where slightly higher accuracy is observed. Similar performance is seen with the CAMS-CSM1-0 model, except for small areas between Azerbaijan and Armenia, where its accuracy is lower. The CESM2 and CMCC-ESM2 models exhibit lower accuracy,

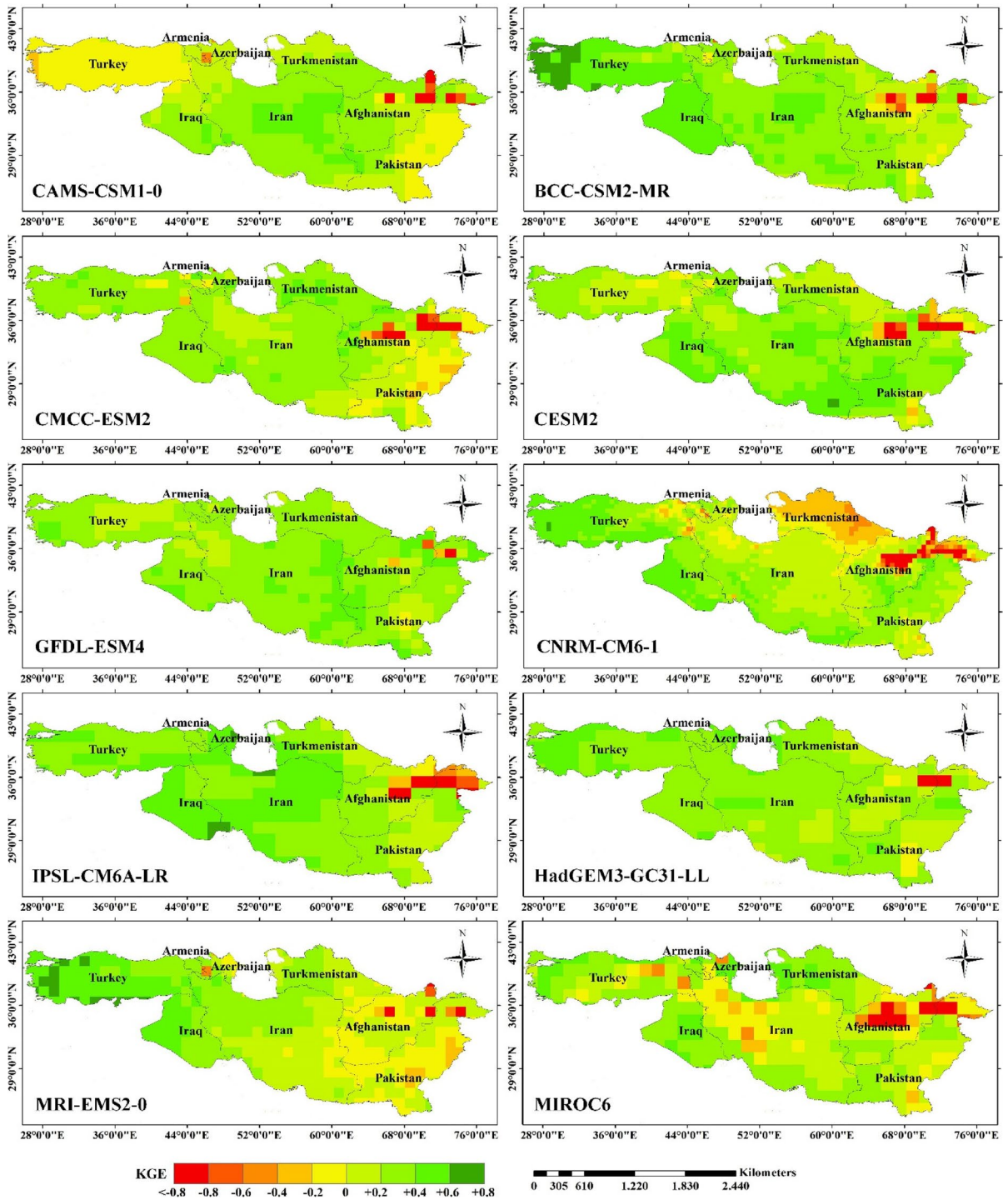


Figure 2 Spatial assessment of CMIP6 performance in temperature prediction using KGE index

particularly in central and eastern Afghanistan, as indicated by the KGE index. The KGE analysis revealed that the CNRM-CM6-1 model displayed lower performance compared to others, particularly in central and northeastern Afghanistan and Turkmenistan (Fig. 2). The GFDL-ESM4 model generally performed well across the study area, except for a small region along the northern Afghan-Pakistani border. HadGEM3-GC31-LL and IPSL-CM6A-LR models exhibited the highest error values in eastern Afghanistan and northern Pakistan. Additionally, MIROC6 and MRI-EMS2-0 models showed significant variations in error, particularly across Afghanistan, Pakistan, Iran, Azerbaijan, and Turkey (Fig. 2).

Figure 3 displays the average KGE values for historical temperature simulations across various countries. In January, the MRI-EMS2-0 model (KGE=0.25) exhibited the most accurate performance in Iraq, while the HadGEM3-GC31-LL model showed the weakest performance among the compared models for the same region. For February, MRI-EMS2-0 (KGE=0.31) remained the top performer in Iraq, followed by MIROC6 (KGE=-3.4) with the poorest performance. The comparison of KGE changes in March reveals that MIROC6 (KGE=0.34, Iraq) maintained its leading position, while the HadGEM3-GC31-LL model (KGE=-9.4, Armenia) demonstrated the weakest performance. Finally, in April, CMCC-ESM2 (KGE=0.28, Iraq) emerged as the best model for temperature prediction, while CNRM-CM6-1 (KGE=-8.8, Afghanistan) had the weakest performance (Fig. 3).

According to the results, the best and weakest CMIP6 models in May were HadGEM3-GC31-LL (KGE=0.3 in Turkmenistan) and MIROC6 (KGE=-1.5 in Afghanistan). There was highest accuracy (KGE=0.34) for the IPSL-CM6A-LR model in Azerbaijan in June, whereas the lowest accuracy (KGE=-0.81) was noted in Pakistan for the same model. In July, MRI-EMS2-0 model (KGE=-0.49) and CESM2 model (KGE=0.35) in Iraq have the best and weakest performance, respectively. In August IPSL-CM6A-LR model (KGE= 0.49 in Turkey) and MRI-EMS2-0 model (KGE=-0.48 in Pakistan), had the most accurate and weakest performances, respectively. Finally, the best and weakest performance for CMIP6 models were observed in September for IPSL-CM6A-LR model (KGE=0.41 in Armenia) and GFDL-ESM4 (KGE=-0.05 in Pakistan); in October for

IPSL-CM6A-LR model (KGE=0.29 in Armenia) and HadGEM3-GC31-LL model (KGE=-4.29, Pakistan); in November for CMCC-ESM2 model (KGE=0.15 in Iraq) and BCC-CSM2-MR model (KGE=-4.8 in Armenia) and in December for IPSL-CM6A-LR model (KGE=0.17 in Iraq) and CNRM-CM6-1 model (KGE=-3.2 in Iran) (Fig. 3).

KGE values, used to estimate historical annual temperatures, are presented in Figure 4 and Table 2. The IPSL-CM6A-LR model exhibited the best performance (KGE=0.51) in Azerbaijan, whereas the HadGEM3-GC31-LL model showed the lowest performance (KGE=-1.4) in Pakistan.

The standard deviation (SD) of the monthly KGE values was also calculated to compare the accuracy of CMIP6 models for estimating temperatures at monthly level (Figure 5). Winter months had the greatest SD, with January showing the peak value. In contrast, summer months (May-September) had the lowest SD. Armenia exhibited the highest (SD = 0.81) and lowest (SD = 0.07) monthly KGE variations for temperature estimation (Fig. 5). These findings suggest that CMIP6 models might exhibit larger variations in performance during colder months when simulating temperatures. Consequently, selecting the most suitable CMIP6 model for cold-month temperature simulations requires increased caution.

The lower performance of some models in central and eastern Afghanistan (e.g., CESM2, CMCC-ESM2) might be linked to factors like complex terrain or limited observational data in these regions. Investigating the specific characteristics of these models and the region's topography or data availability could shed light on this spatial discrepancy. Additionally, the standard deviation analysis (Figure 5) suggests higher performance variability during colder months. This could be attributed to factors like stronger atmospheric circulation patterns or less influence of local phenomena in winter compared to summer. Exploring the models' ability to capture these large-scale atmospheric dynamics in winter versus their performance in simulating smaller-scale summer processes would be a valuable addition. By delving deeper into the spatial and temporal variations, we can gain a more nuanced understanding of the strengths and weaknesses of each model. For instance, the consistent performance of IPSL-CM6A-LR in Azerbaijan (Figure 3) suggests its potential strength in simulating temperature dynamics for regions with specific geographical

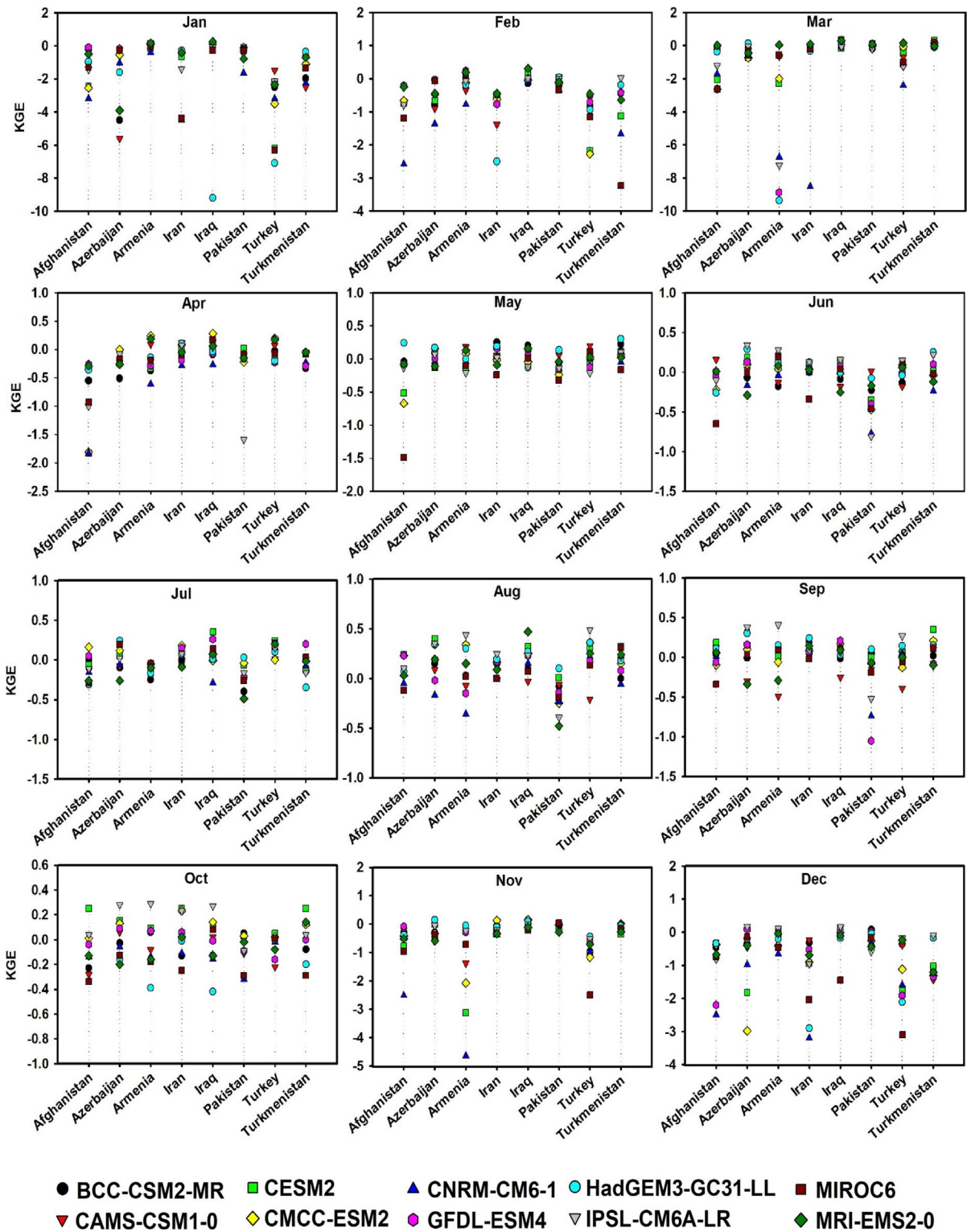


Figure 3 Average KGE index for estimating monthly temperature in different countries

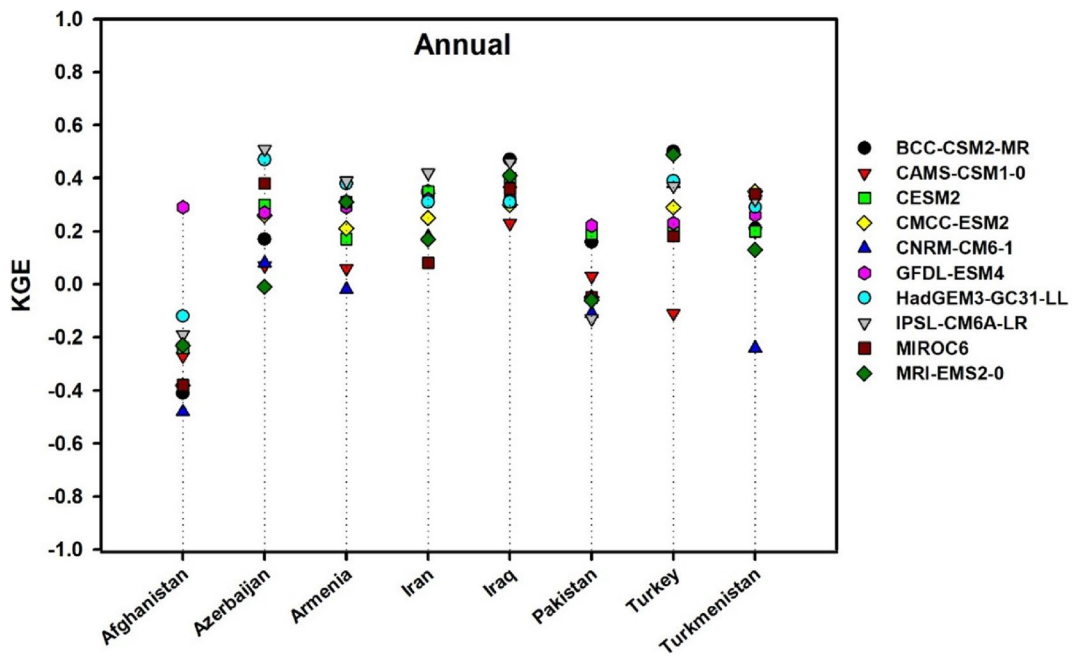


Figure 4 Average KGE index for estimating annual temperature in different countries

Table 2 Average KGE index for estimating annual temperature in different countries

Model	Afghanistan	Azerbaijan	Armenia	Iran	Iraq	Pakistan	Turkey	Turkmenistan
BCC-CSM2-MR	-0.41	0.17	0.31	0.35	0.47	0.16	0.5	0.21
CAMS-CSM1-0	-0.27	0.07	0.06	0.33	0.23	0.03	-0.11	0.2
CESM2	-0.24	0.3	0.17	0.35	0.4	0.19	0.2	0.2
CMCC-ESM2	-0.38	0.26	0.21	0.25	0.3	-0.05	0.29	0.35
CNRM-CM6-1	-0.48	0.08	-0.02	0.18	0.38	-0.11	0.38	-0.24
GFDL-ESM4	0.29	0.27	0.29	0.32	0.32	0.22	0.23	0.26
HadGEM3-GC31-LL	-0.12	0.47	0.38	0.31	0.31	-1.44	0.39	0.29
IPSL-CM6A-LR	-0.19	0.51	0.39	0.42	0.46	-0.13	0.37	0.32
MIROC6	-0.38	0.38	0.31	0.08	0.36	-0.05	0.18	0.34
MRI-EMS2-0	-0.23	-0.01	0.31	0.17	0.41	-0.06	0.49	0.13

characteristics (e.g., mountainous areas). Conversely, models like CNRM-CM6-1, consistently showing lower performance across the region (Figure 2), might require further investigation into their physical parameterizations or underlying processes.

The accuracy of CMIP6 models in simulating precipitation

Figure 6 shows the KGE index’s spatial variation for simulating precipitation across the study area.

Compared to temperature simulations, precipitation simulations exhibited greater spatial variability (larger differences in KGE values across the region). Notably, the BCC-CSM2-MR showed lower performance in the central and driest regions (Figure 6). Additionally, the CAMS-CSM1-0 model’s accuracy varied spatially. It’s worth mentioning that CAMS-CSM1-0, CESM2, and CMCC-ESM2 models showed higher accuracy in western regions (Iraq and Turkey) compared to other areas. The CNRM-CM6-1 model performed well in most regions, while the GFDL-ESM4

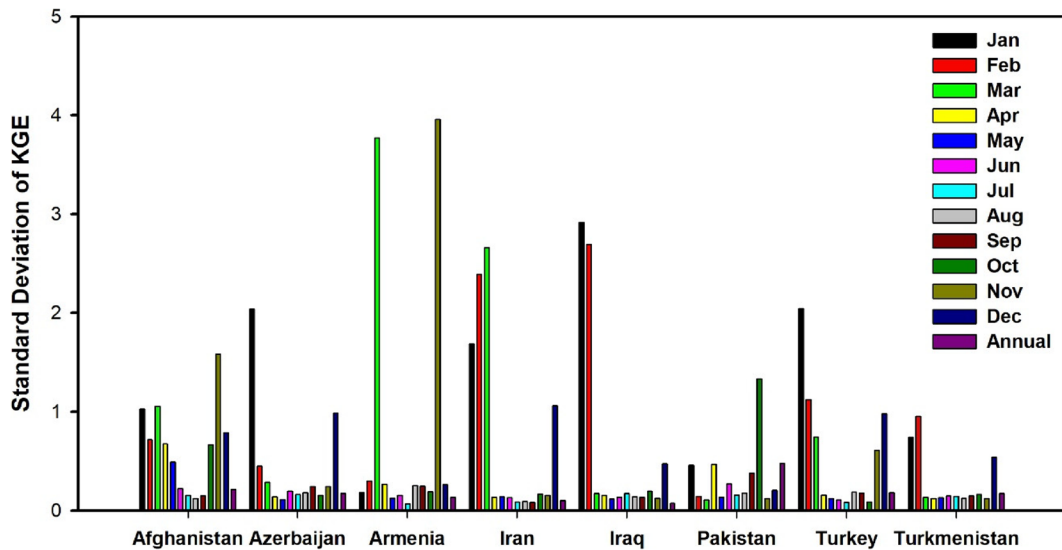


Figure 5 Standard deviation of the KGE index for monthly and annual temperature estimates by different CMIP6 models

model excelled in eastern areas (Afghanistan, Pakistan, and Turkmenistan). The HadGEM3-GC31-LL model showed good performance in central and western regions. The IPSL-CM6A-LR model had low accuracy in all areas except Turkmenistan. Additionally, MIROC6 and MRI-EMS2-0 models were inaccurate in most regions, except for Iraq (Fig. 6).

As shown in Figure 7, CMIP6 models exhibit variations in accuracy for predicting precipitation across different months. In January, the CAMS-CSM1-0 model ($KGE=0.19$) and the MRI-EMS2-0 model ($KGE=-0.63$) were identified as having the highest and lowest accuracy in Armenia and Pakistan, respectively. In February, HadGEM3-GC31-LL ($KGE=0.18$ in Armenia) and GFDL-ESM4 ($KGE=-0.5$ in Pakistan) were ranked as the most accurate and least accurate models, respectively. IPSL-CM6A-LR ($KGE=0.17$ in Armenia) and MIROC6 ($KGE=-0.79$ in Pakistan) displayed the most and least skillful performances in March, respectively. Finally, in April, GFDL-ES ($KGE=0.39$ in Armenia) and MIROC6 ($KGE=-1.7$ in Iran) showed the highest and lowest accuracy (Fig. 7).

Based on the findings, GFDL-ESM4 emerged as the most effective model in Armenia during May ($KGE=-0.01$), June ($KGE=0.12$), and July ($KGE=0.03$). However, MIROC6 ($KGE=-7.5$ in Iran), IPSL-CM6A-LR ($KGE=-8.2$ in Iraq) and BCC-CSM2-MR ($KGE=-9.4$ in Afghanistan)

exhibited the least proficient performance in May, June, and July, respectively. The most effective models identified in August, September, and October were HadGEM3-GC31-LL ($KGE=-0.13$ in Armenia), MIROC6 ($KGE=0.08$ in Armenia), and CESM2 ($KGE=0.09$ in Turkmenistan), respectively. Conversely, IPSL-CM6A-LR ($KGE=-9.8$ in Iran), BCC-CSM2-MR ($KGE=-9.8$ in Iraq) and CNRM-CM6-1 ($KGE=-8.7$ in Iran) were deemed the least proficient models in August, September and October, respectively (Fig. 7).

A comparison of the accuracy of different CMIP6 models in estimating precipitation indicated that MIROC6 ($KGE=0.13$ in Armenia) and CMCC-ESM2 ($KGE=0.12$ in Armenia) were the most accurate models in November and December, respectively. Furthermore, MRI-EMS2-0 ($KGE=-4$ in Armenia) and HadGEM3-GC31-LL ($KGE=-0.93$ in Armenia) were the weakest models during November and December, respectively. The results also indicate that MIROC6 ($KGE=0.13$ in Armenia) and CMCC-ESM2 ($KGE=0.12$ in Armenia) performed best in November and December, respectively. Conversely, MRI-EMS2-0 ($KGE=-4$ in Armenia) and HadGEM3-GC31-LL ($KGE=-0.93$ in Armenia) exhibited the lowest performance during these months (Fig. 7).

The KGE values for annual precipitation estimation are presented in Table 3 and Figure 8. The

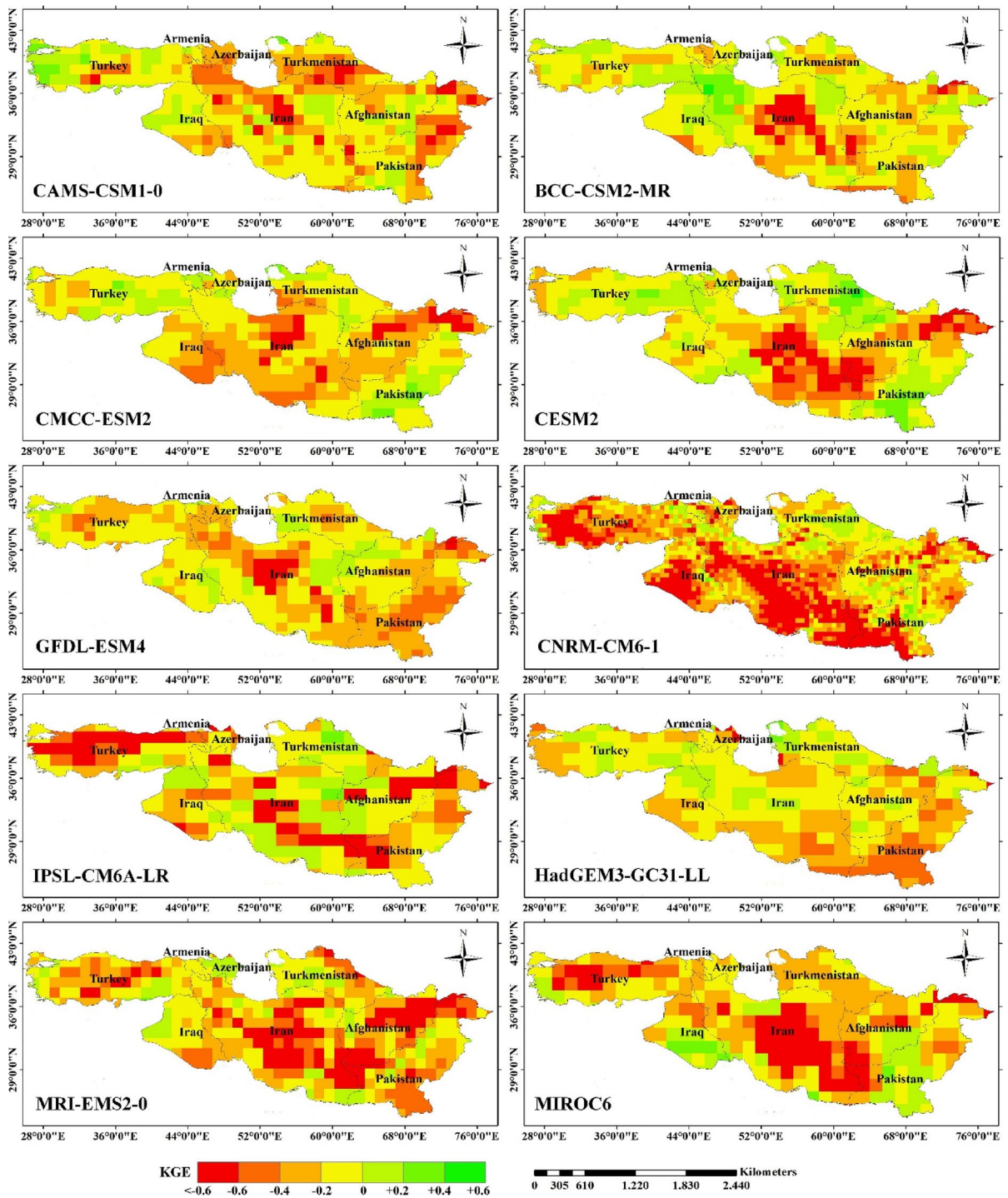


Figure 6 Spatial assessment of CMIP6 performance in precipitation prediction using KGE index

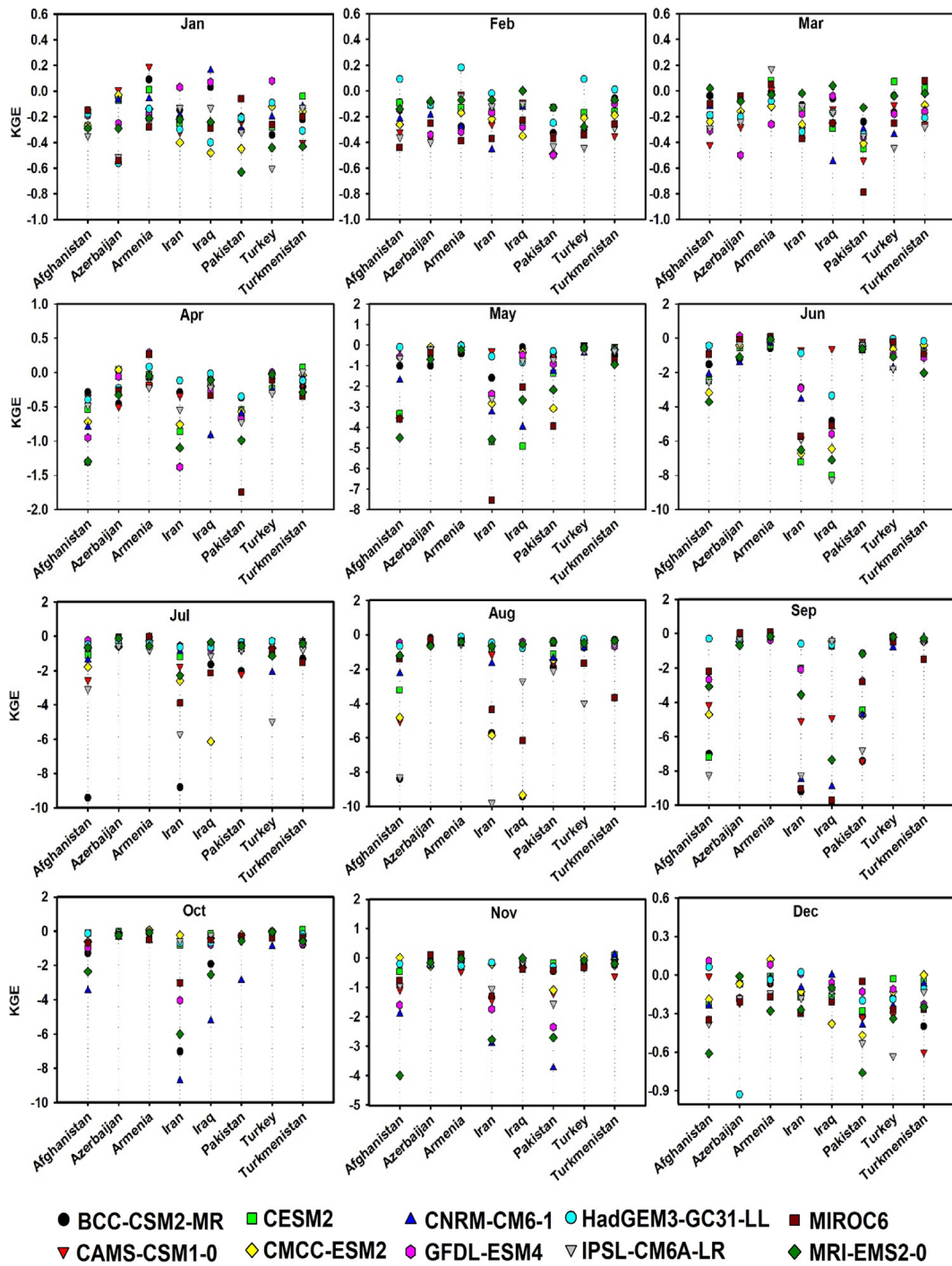


Figure 7 Average KGE index for estimating monthly precipitation in different countries

KGE index ranges from -0.64 (MRI-EMS2-0 in Afghanistan) to 0.05 (CMCC-ESM2 in Armenia). These results suggest a trend of better performance

Table 3 Average KGE index for estimating annual precipitation in different countries

Model	Afghanistan	Azerbaijan	Armenia	Iran	Iraq	Pakistan	Turkey	Turkmenistan
BCC-CSM2-MR	-0.2	0	-0.16	-0.25	-0.06	-0.16	-0.08	-0.07
CAMS-CSM1-0	-0.14	-0.25	-0.09	-0.27	-0.1	-0.29	-0.04	-0.28
CESM2	-0.29	-0.15	-0.03	-0.37	-0.09	-0.1	-0.05	0.04
CMCC-ESM2	-0.33	-0.15	0.05	-0.29	-0.28	-0.1	-0.05	-0.11
CNRM-CM6-1	-0.22	-0.33	-0.21	-0.59	-0.51	-0.4	-0.43	-0.09
GFDL-ESM4	-0.16	-0.19	-0.07	-0.26	-0.11	-0.35	-0.18	-0.15
HadGEM3-GC31-LL	-0.15	-0.16	-0.21	-0.18	-0.14	-0.34	-0.12	-0.06
IPSL-CM6A-LR	-0.31	-0.3	-0.42	-0.23	-0.14	-0.36	-0.57	-0.03
MIROC6	-0.34	-0.12	-0.25	-0.6	-0.05	-0.22	-0.29	-0.24
MRI-EMS2-0	-0.64	0.04	-0.14	-0.45	-0.17	-0.37	-0.26	-0.15

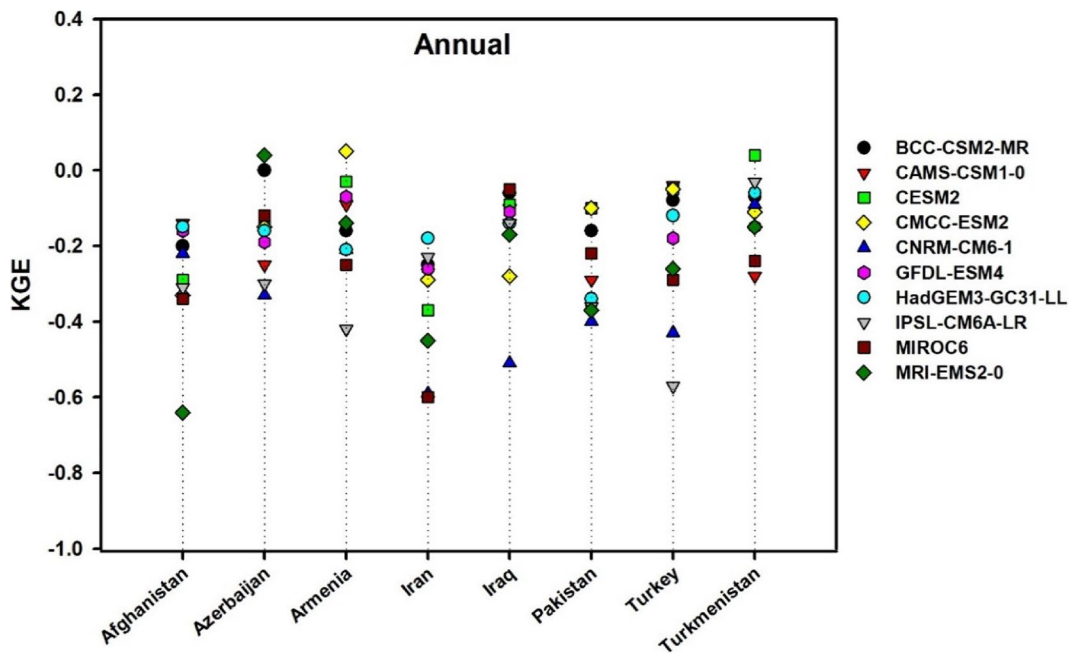


Figure 8 Average KGE index for estimating annual precipitation in different countries

by CMIP6 models in estimating precipitation during the winter months.

Figure 9 depicts the monthly variation of the KGE index for estimating precipitation across various countries. It reveals that model performance varied the most during warm months (May to October), peaking in Iraq during September (SD= 4.2). Conversely, Afghanistan exhibited the lowest variability in January (SD= 0.07). These findings suggest that CMIP6 models are more consistent in

predicting precipitation during rainy months compared to warm months.

The observed higher spatial variability in precipitation simulations compared to temperature (Figure 6) warrants further discussion. Potential reasons for this could include the complex interplay of factors like topography, convection, and large-scale atmospheric circulation that influence precipitation. Models might struggle to accurately represent these

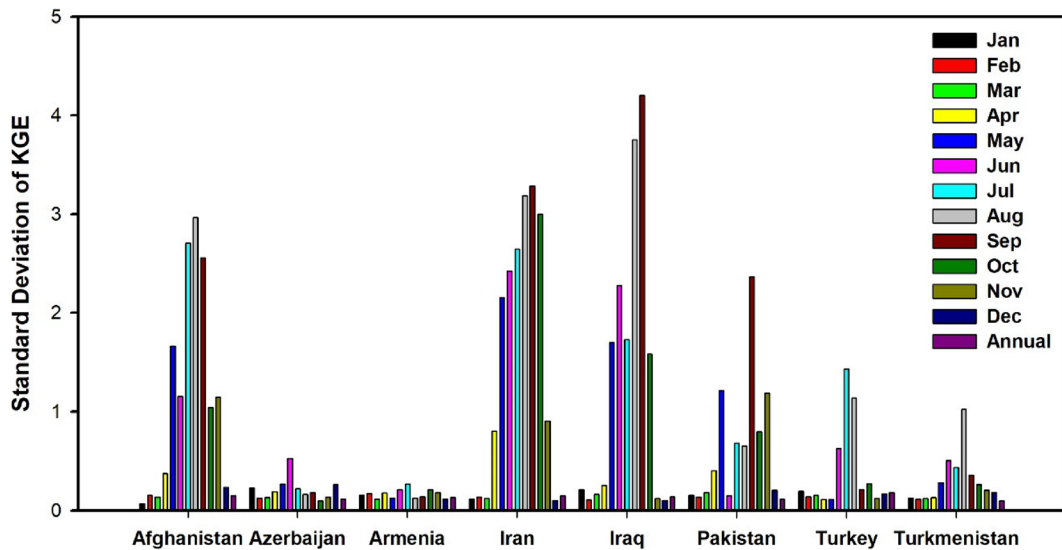


Figure 9 Standard deviation of the KGE index for monthly and annual precipitation estimates by different CMIP6 models

interactions, leading to greater spatial variability in performance compared to temperature simulations. Additionally, limited precipitation data, especially in mountainous regions, can hinder model calibration and validation, potentially contributing to the observed spatial variations in accuracy. Investigating the models' performance in areas with denser observational networks could provide insights into their intrinsic capabilities. Furthermore, the trend of better performance during winter months (Figure 8) could be related to the dominance of large-scale circulation patterns in winter precipitation compared to the more localized convective processes that influence summer precipitation. Analyzing the models' ability to capture these seasonal variations in precipitation drivers would be insightful. Understanding the spatial and temporal variations in precipitation simulations allows for a more targeted discussion of model strengths and weaknesses. For instance, the consistent performance of GFDL-ESM4 in eastern regions (Figure 6) suggests its potential strength in simulating precipitation patterns influenced by specific atmospheric dynamics in those areas. Conversely, models like BCC-CSM2-MR, consistently showing lower accuracy across the region (Figure 6), might require further investigation into their parameterizations for representing precipitation processes.

Discussion

This study evaluated the performance of various CMIP6 models in simulating temperature and precipitation across a region encompassing Afghanistan, Pakistan, Iran, Azerbaijan, and Turkey. The findings provide valuable insights into the strengths and weaknesses of these models for climate simulations in this specific region. Consistent with the findings of Babausmail et al. (2021) and Nguyen-Duy et al. (2023), the ensemble mean of multiple CMIP6 models often outperformed individual models in terms of capturing the overall trends. This emphasizes the potential benefits of utilizing multi-model ensembles for generating more robust climate projections, a practice advocated for by Sanderson et al. (2021) and Peng et al. (2023).

Our findings align with previous studies (e.g., Zhang et al., 2024; Babausmail et al., 2021; Chen et al., 2022; Jiang et al., 2020) in demonstrating that CMIP6 models exhibit varying skill in simulating temperature and precipitation. IPSL-CM6A-LR emerged as a relatively skillful model for precipitation in some areas (e.g., Azerbaijan), while models like BCC-CSM2-MR displayed poor performance (central arid regions). This highlights the importance of regional considerations when selecting appropriate models, as echoed by Xu et al. (2023) in their analysis of East Asian monsoon simulations.

Similar to Guo et al. (2021) and Ngoma et al. (2021) who reported wet biases in arid regions, our study observed overestimations of precipitation in some areas (e.g., Iran) by most CMIP6 models. This suggests potential limitations in these models' capability to accurately capture precipitation dynamics in arid and semi-arid climates, a challenge also recognized in studies by Kim et al. (2024) and Liu et al. (2022). Our results resonate with Iqbal et al. (2021) and Ngoma et al. (2021) who observed higher model discrepancies in simulating warm-season precipitation compared to colder months. This suggests that CMIP6 models require further development to improve their performance in capturing the complexities of seasonal precipitation patterns, as emphasized by Wang et al. (2023) in their recent contribution.

This study provides a detailed assessment of CMIP6 model performance in a specific region (Afghanistan, Pakistan, Iran, Azerbaijan, and Turkey) that has not been extensively explored in previous literature. This regional focus offers valuable insights into the suitability of these models for climate simulations in this area, complementing the works of Gao et al. (2022) on the Tibetan Plateau and He et al. (2023). By analyzing monthly variations in the KGE index, we demonstrate that model accuracy fluctuates throughout the year, with generally lower performance during warmer months for precipitation simulations. This information is crucial for selecting appropriate models for specific seasonal climate studies, as highlighted by Rivera (2023). The inclusion of standard deviation analysis for monthly KGE values provides additional insights into model consistency. We observed higher standard deviations during winter months for temperature estimations, suggesting greater variability in model performance during colder periods. This information can be helpful for researchers when interpreting model outputs and associated uncertainties, echoing the importance of uncertainty quantification stressed by Yazdandoost et al. (2021).

While this study offers valuable insights, certain limitations need to be acknowledged. The accuracy of the findings is contingent on the quality and completeness of observational data used for comparison, as noted by Chen et al. (2020) in their assessment of CMIP5 and CMIP6 data. The evaluation focused on historical simulations, and the performance of these models for future climate projections remains to be

assessed, as addressed by O'Neill et al. (2016) concerning the challenges of future climate projections. The coarse resolution of CMIP6 models might not adequately capture the influence of complex regional topography on climate, a limitation also recognized in studies by Giorgi & Raffaele (2022).

Future research directions include incorporating additional observational datasets to strengthen the evaluation of model performance, assessing how well these models project future climate changes in the study region, investigating the potential benefits of employing higher-resolution models for capturing regional climate dynamics, and identifying the underlying physical processes responsible for model biases to inform model development efforts. By addressing these limitations and pursuing further research avenues, we can gain a deeper understanding of CMIP6 model capabilities and limitations for climate simulations in the region of interest and beyond. This knowledge will be instrumental in generating more reliable climate projections and informing effective climate change adaptation strategies.

Conclusions

A crucial first step in assessing the impacts of climate change on temperature and precipitation is an initial evaluation of the accuracy of different AOGCM models. This is especially important in vulnerable regions with shared water sources, where changes in temperature and precipitation can have a substantial impact on water availability. This study found variations in the accuracy of CMIP6 models for estimating temperature and precipitation across Iran and its neighboring countries. During the historical period (1980-2014), these models performed relatively better at estimating temperature compared to precipitation. Additionally, the variability in temperature estimation was lower during warmer months compared to colder months. Interestingly, the opposite trend was observed for precipitation, with higher accuracy in colder months. However, despite the better performance in temperature estimation, the accuracy of CMIP6 models in predicting temperature remains insufficient in the some regions of the study area. Therefore, a thorough evaluation of CMIP6 models is critical before further analysis of climate change impacts on hydro-ecological aspects in these regions.

Competing interests

The authors declare no competing interests.

Author Contributions Mohamamd Javad Zareian, Hossein Dehban, Alireza Gohari and Ali Torabi Haghghi conceived the study, conducted the analysis, and examined and verified the results. All authors wrote and edited the manuscript.

Funding No funding was obtained for this study.

Data availability The datasets generated during the current study are available from the corresponding author on request.

Competing interests The authors declare no competing interests.

References

- Ahmed, K., Shahid, S., & Nawaz, N. (2018). Impacts of climate variability and change on seasonal drought characteristics of Pakistan. *Atmospheric research*, *214*, 364–374. <https://doi.org/10.1016/j.atmosres.2018.08.020>
- Ahmed, K., Sachindra, D. A., Shahid, S., Demirel, M. C., & Chung, E. S. (2019). Selection of multi-model ensemble of GCMs for the simulation of precipitation based on spatial assessment metrics. *Hydrology and Earth System Sciences*, *23*, 4803–4824. <https://doi.org/10.5194/hess-2018-585>
- Allan, C., Xia, J., & Pahl-Wostl, C. (2013). Climate change and water security: challenges for adaptive water management. *Current Opinion in Environmental Sustainability*, *5*(6), 625–632. <https://doi.org/10.1016/j.cosust.2013.09.004>
- Babaoumail, H., Hou, R., Ayugi, B., Ojara, M., Ngoma, H., Karim, R., Rajasekar, A., & Ongoma, V. (2021). Evaluation of the performance of CMIP6 models in reproducing rainfall patterns over North Africa. *Atmosphere*, *12*(4), 475. <https://doi.org/10.3390/atmos12040475>
- Brooks, H. E. (2013). Severe thunderstorms and climate change. *Atmospheric research*, *123*, 129–138. <https://doi.org/10.1016/j.atmosres.2012.04.002>
- Calzadilla, A., Rehdanz, K., Betts, R., Falloon, P., Wiltshire, A., & Tol, R. S. (2013). Climate change impacts on global agriculture. *Climatic change*, *120*, 357–374. <https://doi.org/10.1007/s10584-013-0822-4>
- Charlton, M. B., & Arnell, N. W. (2011). Adapting to climate change impacts on water resources in England—an assessment of draft water resources management plans. *Global Environmental Change*, *21*(1), 238–248. <https://doi.org/10.1016/j.gloenvcha.2010.07.012>
- Chen, J., Brissette, F. P., Lucas-Picher, P., & Caya, D. (2017). Impacts of weighting climate models for hydro-meteorological climate change studies. *Journal of Hydrology*, *549*, 534–546. <https://doi.org/10.1016/j.jhydrol.2017.04.025>
- Chen, H., Sun, J., Lin, W., & Xu, H. (2020). Comparison of CMIP6 and CMIP5 models in simulating climate extremes. *Sci. Bull*, *65*(17), 1415–1418. <https://doi.org/10.1016/j.scib.2020.05.015>
- Chen, R., Duan, K., Shang, W., Shi, P., Meng, Y., & Zhang, Z. (2022). Increase in seasonal precipitation over the Tibetan Plateau in the 21st century projected using CMIP6 models. *Atmospheric Research*, *277*, 106306. <https://doi.org/10.1016/j.atmosres.2022.106306>
- Gao, J., Du, J., Yang, C., Deqing, Z., Ma, P., & Zhuo, G. (2022). Evaluation and correction of climate simulations for the Tibetan Plateau using the CMIP6 models. *Atmosphere*, *13*(12), 1947. <https://doi.org/10.3390/atmos13121947>
- Giorgi, F., & Raffaele, F. (2022). On the dependency of GCM-based regional surface climate change projections on model biases, resolution and climate sensitivity. *Climate Dynamics*, *58*, 2843–2862. <https://doi.org/10.1007/s00382-021-06037-8>
- Gohari, A., Mirchi, A., & Madani, K. (2017). System dynamics evaluation of climate change adaptation strategies for water resources management in central Iran. *Water Resources Management*, *31*, 1413–1434. <https://doi.org/10.1007/s11269-017-1575-z>
- Gohari, A., Eslamian, S., Abedi-Koupaei, J., Massah Bavani, A., & Wang, D. (2013). A Survey of Exemplar Teachers' Perceptions, Use, And Access Of Computer-Based Games And Technology For Classroom Instruction. *Science of the Total Environment*, *442*. <https://doi.org/10.1016/j.scitotenv.2012.10.029>
- Gohari, A., Zareian, M. J., & Eslamian, S. (2015). A multi-model framework for climate change impact assessment. In W. L. Filho (Ed.), *Handbook of climate change adaptation* (pp. 17–35). Berlin Heidelberg: Springer. https://doi.org/10.1007/978-3-642-40455-9_91-1
- Gomis-Cebolla, J., Rattayova, V., Salazar-Galán, S., & Francés, F. (2023). Evaluation of ERA5 and ERA5-Land reanalysis precipitation datasets over Spain (1951–2020). *Atmospheric Research*, *284*, 106606. <https://doi.org/10.1016/j.atmosres.2023.106606>
- Goyal, R. K. (2004). Sensitivity of evapotranspiration to global warming: a case study of arid zone of Rajasthan (India). *Agricultural water management*, *69*(1), 1–11. <https://doi.org/10.1016/j.agwat.2004.03.014>
- Guo, H., Bao, A., Chen, T., Zheng, G., Wang, Y., Jiang, L., & De Maeyer, P. (2021). Assessment of CMIP6 in simulating precipitation over arid Central Asia. *Atmospheric Research*, *252*, 105451. <https://doi.org/10.1016/j.atmosres.2021.105451>
- He, L., Zhou, T., & Chen, X. (2023). South Asian summer rainfall from CMIP3 to CMIP6 models: biases and improvements. *Climate Dynamics*, *61*(3), 1049–1061. <https://doi.org/10.1007/s00382-022-06542-4>
- Hersbach, H., Bell, B., Berrisford, P., Hirahara, S., Horányi, A., Muñoz-Sabater, J., Nicolas, J., Peubey, C., Radu, R., Schepers, D., & Simmons, A. (2020). The ERA5 global reanalysis. *Quarterly Journal of the Royal Meteorological Society*, *146*, 1999–2049. <https://doi.org/10.1002/qj.3803>
- IPCC. (2021). Climate Change 2021: The Physical Science Basis. *Contribution of Working Group I to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change*. Cambridge University Press.

- Iqbal, Z., Shahid, S., Ahmed, K., Ismail, T., Ziarh, G. F., Chung, E. S., & Wang, X. (2021). Evaluation of CMIP6 GCM rainfall in mainland Southeast Asia. *Atmospheric Research*, 254, 105525. <https://doi.org/10.1016/j.atmosres.2021.105525>
- Jiang, J., Zhou, T., Chen, X., & Zhang, L. (2020). Future changes in precipitation over Central Asia based on CMIP6 projections. *Environmental Research Letters*, 15(5), 054009. <https://doi.org/10.1088/1748-9326/ab7d03>
- Kim, T., & Villarini, G. (2024). Projected changes in daily precipitation, temperature and wet-bulb temperature across Arizona using statistically downscaled CMIP6 climate models. *International Journal of Climatology*, 44, 1994–2010. <https://doi.org/10.1002/joc.8436>
- Knoben, W. J., Freer, J. E., & Woods, R. A. (2019). Inherent benchmark or not? Comparing Nash-Sutcliffe and Kling-Gupta efficiency scores. *Hydrology and Earth System Sciences*, 23(10), 4323–4331. <https://doi.org/10.1088/1748-9326/ab7d03>
- Lamontagne, J. R., Barber, C. A., & Vogel, R. M. (2020). Improved estimators of model performance efficiency for skewed hydrologic data. *Water Resources Research*, 56(9), e2020WR027101. <https://doi.org/10.1029/2020WR027101>
- Li, Y., Qin, Y., & Rong, P. (2022). Evolution of potential evapotranspiration and its sensitivity to climate change based on the Thornthwaite, Hargreaves, and Penman-Monteith equation in environmental sensitive areas of China. *Atmospheric Research*, 273, 106178. <https://doi.org/10.1016/j.atmosres.2022.106178>
- Liu, Z., Huang, J., Xiao, X., & Tong, X. (2022). The capability of CMIP6 models on seasonal precipitation extremes over Central Asia. *Atmospheric Research*, 278, 106364. <https://doi.org/10.1016/j.atmosres.2022.106364>
- Luber, G., & McGeehin, M. (2008). Climate change and extreme heat events. *American journal of preventive medicine*, 35(5), 429–435. <https://doi.org/10.1016/j.amepre.2008.08.021>
- Mal, S., Singh, R. B., Huggel, C., & Grover, A. (2018). Introducing Linkages Between Climate Change, Extreme Events, and Disaster Risk Reduction. In S. Mal, R. Singh, & C. Huggel (Eds.), *Climate Change, Extreme Events and Disaster Risk Reduction. Sustainable Development Goals Series* (Vol. 14, pp. 1–13). Cham: Springer.
- Michener, W. K., Blood, E. R., Bildstein, K. L., Brinson, M. M., & Gardner, L. R. (1997). Climate change, hurricanes and tropical storms, and rising sea level in coastal wetlands. *Ecological applications*, 7(3), 770–801. <https://doi.org/10.2307/2269434>
- Ngoma, H., Wen, W., Ayugi, B., Babaousmail, H., Karim, R., & Ongoma, V. (2021). Evaluation of precipitation simulations in CMIP6 models over Uganda. *International Journal of Climatology*, 41(9), 4743–4768. <https://doi.org/10.1002/joc.7098>
- Nguyen-Duy, T., Ngo-Duc, T., & Desmet, Q. (2023). Performance evaluation and ranking of CMIP6 global climate models over Vietnam. *Journal of Water and Climate Change*, 14(6), 1831–1846. <https://doi.org/10.2166/wcc.2023.454>
- O'Neill, B. C., Tebaldi, C., Van Vuuren, D. P., Eyring, V., Friedlingstein, P., Hurtt, G., Knutti, R., Kriegler, E., Lamarque, J. F., Lowe, J., Meehl, G. A., & Sanderson, B. M. (2016). The scenario model intercomparison project (ScenarioMIP) for CMIP6. *Geoscientific Model Development*, 9(9), 3461–3482. <https://doi.org/10.5194/gmd-9-3461-2016>
- Pachauri, R. K., Allen, M. R., Barros, V. R., Broome, J., Cramer, W., Christ, R., Church, J. A., Clarke, L., Dahe, Q., Dasgupta, P., & Dubash, N. K. (2014). Climate change 2014: synthesis report. *Contribution of Working Groups I, II and III to the fifth assessment report of the Intergovernmental Panel on Climate Change*. Geneva: IPCC.
- Patil, S. D., & Stieglitz, M. (2015). Comparing spatial and temporal transferability of hydrological model parameters. *Journal of Hydrology*, 525, 409–417. <https://doi.org/10.1016/j.jhydrol.2015.04.003>
- Peng, S., Wang, C., Li, Z., Mihara, K., Kuramochi, K., Toma, Y., & Hatano, R. (2023). Climate change multi-model projections in CMIP6 scenarios in Central Hokkaido Japan. *Scientific Reports*, 13(1), 230. <https://doi.org/10.1038/s41598-022-27357-7>
- Radmanesh, Y., Tabrizi, M. S., Etedali, H. R., Azizian, A., & Babazadeh, H. (2023). Comparative evaluation of the accuracy of re-analysed and gauge-based climatic data in Iran. *Journal of Earth System Science*, 132(4), 190. <https://doi.org/10.1007/s12040-023-02202-1>
- Rivera, P. (2023). Evaluation of historical simulations of CMIP6 models for temperature and precipitation in Guatemala. *Earth Systems and Environment*, 7(1), 43–65. <https://doi.org/10.1007/s41748-022-00333-x>
- Sanderson, B. M., Pendergrass, A., Koven, C. D., Brient, F., Booth, B. B., Fisher, R. A., & Knutti, R. (2021). On structural errors in emergent constraints. *Earth System Dynamics Discussions*, 2021, 1–30. <https://doi.org/10.5194/esd-12-899-2021>
- Shukla, P. R., Masson-Delmotte, V., Zhai, P., Pörtner, H. O., Roberts, D., Skea, J., Calvo, E., Priyadarshi, B., et al. (2019). *Climate Change and Land: An IPCC Special Report on climate change, desertification, land degradation, sustainable land management, food security, and greenhouse gas fluxes in terrestrial ecosystems*. Cambridge: Cambridge University Press. <https://doi.org/10.1017/9781009157988>
- Solomon, S., Qin, D., Manning, M., et al. (2007). Climate change 2007: the physical science basis. *Contribution of working group I to the fourth assessment report of the intergovernmental panel on climate change*. Cambridge: Cambridge University Press.
- Song, L., Xu, C., Long, Y., Lei, X., Suo, N., & Cao, L. (2022). Performance of seven gridded precipitation products over arid central Asia and subregions. *Remote Sensing*, 14(23), 6039. <https://doi.org/10.3390/rs14236039>
- Tanveer, M. E., Lee, M. H., & Bae, D. H. (2016). Uncertainty and reliability analysis of CMIP5 climate projections in South Korea using REA method. *Procedia engineering*, 154, 650–655. <https://doi.org/10.1016/j.proeng.2016.07.565>
- Wang, Y., Zhao, T., Hua, L., Guan, X., Xu, C., & Chen, Q. (2023). Influence of anthropogenic and natural forcings on future changes in precipitation projected by the CMIP6–DAMIP models. *International Journal of Climatology*, 43(8), 3892–3906. <https://doi.org/10.1002/joc.8064>

- Werndl, C. (2016). On defining climate and climate change. *The British Journal for the Philosophy of Science*, 67, 337–364. <https://doi.org/10.1093/bjps/axu048>
- Xiong, Y., Ta, Z., Gan, M., Yang, M., Chen, X., Yu, R., Disse, M., & Yu, Y. (2021). Evaluation of cmip5 climate models using historical surface air temperatures in central Asia. *Atmosphere*, 12, 308. <https://doi.org/10.3390/atmos12030308>
- Xu, X., Yun, X., Tang, Q., Cui, H., Wang, J., Zhang, L., & Chen, D. (2023). Projected seasonal changes in future rainfall erosivity over the Lancang-Mekong River basin under the CMIP6 scenarios. *Journal of Hydrology*, 620, 129444. <https://doi.org/10.1016/j.jhydrol.2023.129444>
- Yanmin, J., & Haomin, W. (2013). Simulation capabilities of 20 CMIP5 models for annual mean air temperatures in Central Asia. *Advances in Climate Change Research*, 9(2), 110. <https://doi.org/10.3969/j.issn.1673-1719.2013.02.005>
- Yazdandoost, F., Moradian, S., Izadi, A., & Aghakouchak, A. (2021). Evaluation of CMIP6 precipitation simulations across different climatic zones: Uncertainty and model intercomparison. *Atmospheric Research*, 250, 105369. <https://doi.org/10.1016/j.atmosres.2020.105369>
- Zareian, M. J. (2021). Optimal water allocation at different levels of climate change to minimize water shortage in arid regions (Case Study: Zayandeh-Rud River Basin, Iran). *Journal of Hydro-environment Research*, 35, 13–30. <https://doi.org/10.1016/j.jher.2021.01.004>
- Zareian, M. J., Eslamian, S., & Safavi, H. R. (2015). A modified regionalization weighting approach for climate change impact assessment at watershed scale. *Theoretical and Applied Climatology*, 122, 497–516. <https://doi.org/10.1007/s00704-014-1307-8>
- Zhang, L., Zhao, Y., Cheng, T. F., & Lu, M. (2024). Future changes in global atmospheric rivers projected by CMIP6 models. *Journal of Geophysical Research: Atmospheres*, 129(3), 2023JD039359. <https://doi.org/10.1029/2023JD039359>

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Springer Nature or its licensor (e.g. a society or other partner) holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.