Journal of
Water & Climate Change

IWA PUBLISHING

Journal of Water and Climate Change Vol 14 No 8, 2663 doi: 10.2166/wcc.2023.021

Evaluation and selection of CMIP6 GCMs for long-term hydrological projections based on spatial performance assessment metrics across South Korea

Nguyen Thi Huong[a], Yong-Tak Kim[b] and Hyun-Han Kwon[a,*]

[a] Department of Civil and Environmental Engineering, Sejong University, Seoul, South Korea
[b] Korea Institute of Civil Engineering and Building Technology, Goyang-si, South Korea
*Corresponding author. E-mail: hkwon@sejong.ac.kr

H-HK, 0000-0003-4465-2708

ABSTRACT

The selection of an appropriate subset of Global Climate Models (GCMs) from the Coupled Model Intercomparison Project 6 (CMIP6) for long-term hydrological simulation at the basin scale is a necessity. This study selected high-performing GCMs among 32 available CMIP6 GCMs on the basis of reproducing the observed precipitation over the main watershed in South Korea during the historical period. An integrated selection approach based on four spatial performance assessment metrics was proposed to better estimate changes in precipitation using the simulated GCMs precipitation. Results revealed that the spatial performance over different GCMs could provide an effective means of selecting GCMs for hydrological study over the major river watersheds in South Korea. The four top-ranked GCMs are FIO-ESM-2-0, CESM2-WACCM, CESM2, and CMCC-ESM2, and they overestimated precipitation in South Korea during the historical period, with a bias of 10–25%. However, this study confirmed that a formal bias correction for these GCMs is not recommended prior to model selection, and the ranking of GCMs under the bias correction could be problematic. The proposed approach in this study can be applied to numerous GCMs, climate variables, and other regions to select representative GCMs to reduce the uncertainties in terms of the spatial patterns observed.

Key words: CMIP6, comprehensive rating metric, GCMs, model selection, spatial performance indicators

HIGHLIGHTS

- The goodness-of-fit measures are proposed for the selection of climate models.
- A comprehensive rating metric approach is proposed to evaluate the overall ranking of the GCMs.
- Exploration of the spatial performance of different GCMs can effectively select GCMs.
- A formal bias correction for these GCMs is not recommended prior to model selection, and the ranking of GCMs under the bias correction could be problematic.

1. INTRODUCTION

The Coupled Model Intercomparison Project 6 (CMIP6) was used to support the Sixth Assessment Report (AR6) of the Intergovernmental Panel on Climate Change (IPCC). The AR6 report has a much clearer picture of recent rapid, widespread, and intensifying changes in the climate. The effects of climate change begin to appear in various regions in multiple ways (Nunes & Dias 2022). An increase in drought severity and frequency (Wetter *et al.* 2014) and flash floods (Marchi *et al.* 2016) have been observed in Europe. Changes in climate extremes have been explored around the world over the last three decades, including Africa (Mason *et al.* 2010) and Asia. The increasing trend of heatwaves in East Asia has also been often discussed (Im *et al.* 2019). Kwon & Lall (2016) identified a century-scale drought during 2012–2015 in California, and strong fire activity has also been reported in Australia (Canadell *et al.* 2021).

Climate change is largely related to spatiotemporal variability in precipitation (Lindsey & Dahlman 2020). In particular, global warming increase of 1.5–2 °C or more would lead to an increase in extreme rainfalls (Masson-Delmotte *et al.* 2021). This emphasizes the need for a spatiotemporal precipitation modeling framework to guide the water management policies at different spatial scales (e.g., regional, local, or basin scale) under climate change scenarios (Akhter *et al.* 2017).

The Global Climate Models (GCMs) for simulating future climates were released under a series of CMIPs in different phases, and they are considered the primary source for interdisciplinary scientific climate research. CMIP3 and CMIP5 comprise 25 and 40 GCMs, respectively, and CMIP6, the most recent experimental design, has 55 GCMs (Eyring *et al.* 2016). The GCM simulations are principally used to analyze and simulate atmospheric circulation and climate on a global scale (Pour *et al.* 2018). Various studies indicated that CMIP6 simulations generally have better performance than CMIP5 in terms of bias in precipitation at the annual and seasonal scales (Gusain *et al.* 2020), indicating the CMIP6 GCMs better reproduce historical precipitation extremes observed at the regional and global scale than the CMIP5 GCMs. However, the inherent uncertainty in climate change scenarios from GCM simulations and the necessity of adopting a suitable subset of GCMs in the development of climate change scenarios have been addressed. In particular, the uncertainty of climate change scenarios was caused by the selection of GCMs, emission scenarios, downscaling model approach, and hydrological model structures, where a parameter estimation (Mandal *et al.* 2019) indicated that it is essential to evaluate the model performance of each GCM in terms of the ability to reproduce key statistics of historical precipitation. Cook *et al.*(2017) discussed the importance of choosing a suitable GCM for engineering applications, demonstrating differences in the impact assessment at different spatiotemporal scales. Therefore, the selection of a set of suitable GCMs from the CMIP6 inventory could be an essential reliable assessment using long-term hydrological simulation.

To evaluate and prioritize the best-performing GCMs, previous studies have used various machine learning models and statistical metrics based on information entropy, probability distribution function, a Bayesian approach, correlation analysis, clustering approach, hierarchical analysis of variance (ANOVA) models, symmetrical uncertainty approach, and recursive feature elimination model. More specifically, Wu *et al.* (2018) used standard deviation (σ), deterministic coefficient (DC), correlation coefficient (CC), relative error (RE), and root mean square error (RMSE) to evaluate six CMIP5 GCMs for precipitation in the Huaihe River Basin, China. He *et al.* (2019) assessed nine CMIP5 GCMs using CC and RMSE for temperature in China. In general, studies have combined different indicators to evaluate the model performance from different perspectives. Parth Sarthi *et al.* (2016) considered combined indicators, such as the Taylor diagram, skill score (SS), CC, and RMSE, while Raju & Kumar (2014) combined several indicators, such as average absolute relative error (AARE), normalized root mean square error (NRMSE), absolute normalized mean bias error (ANMBE), and SS. These performance indicators have been widely used because of their advantages for systematic comparison of different GCMs (Guilyardi *et al.* 2009), demonstrating effective assessment of model performance (Gu *et al.* 2015). However, there have been few attempts to investigate the spatial representation of a large-scale climate in selecting GCMs (Koch *et al.* 2018). More importantly, the spatial pattern of climate variables (e.g., precipitation and temperature) should be captured by GCMs for the region that is affected by monsoons (Srinivasa Raju *et al.* 2017). Spatial climate patterns provide important information with regard to the overall spatial coverage of areas affected by drought and flood in response to climate extremes (Khan *et al.* 2018; Salman *et al.* 2018).

The full use of available GCMs could effectively reduce the uncertainty associated with climate models, leading to a computationally inefficient process of extensive data (Lutz *et al.* 2016). Therefore, the selection of appropriate GCMs can provide an effective way for future long-term hydrological simulations. Several studies have been performed on the selection of subsets of GCMs (Kim *et al.* 2021) and their combinations (multimodel ensembles, MMEs) for selective GCMs (Suh *et al.* 2016) in South Korea. Recently, a selection process of representative GCMs for East Asia and South Asia has been adopted (Kim *et al.* 2020) and for the areas affected by the East Asian monsoons, including South Korea (Kim *et al.* 2021). They selected several pairs of CMIP5 and CMIP6 GCMs using the selection process, and Kim *et al.* (2020) explored 29 CMIP5 and 25 CMIP6 participating GCMs that have the required climate variables for hydrological simulations.

The Korean peninsula is affected by East Asian monsoons with a distinct spatiotemporal pattern of precipitation (Kwon *et al.* 2008; Uranchimeg *et al.* 2020), and the associated climate variability under climate change is likely to increase the risks of the hydrological system (Lima *et al.* 2021). Under these circumstances, a comprehensive assessment of future climate informed by CMIP6 GCMs is required for a formal hydrological risk analysis. This study aims to explore the model performance of CMIP6 GCMs over South Korea in terms of the capability to reproduce spatial climate patterns using Goodness-of-Fit (GoF) measures such as the SPAtial EFficiency metric (SPAEF) (Demirel *et al.* 2018), Kling–Gupta efficiency (KGE) (Gupta *et al.* 2009), Cramer's V (Cramér 1999), and fractions skill score (FSS) (Roberts & Lean 2008).

The paper is organized as follows. We introduce a brief overview of climate model selection for hydrological simulations and the main objectives of the study in this current section. In the following section, the research area and hydrometeorological datasets are summarized. The methodology is described in Section 3. Results and discussions are provided in Section 4. Finally, conclusions and suggestions for future research directions are presented in Section 5.

2. STUDY AREA AND DATA

2.1. Study area

South Korea (35°50′N, 127°00′W) is mostly mountainous with wide coastal plains and lowlands located in the Southeastern and Western regions, as shown in Figure 1. Annual precipitation ranges from 1,000 mm to 1,800 mm, and South Korea has a relatively large seasonal variability of water resources. Approximately 60% of the annual precipitation is concentrated during the East Asian monsoon season (i.e., Changma) from late June to August, and the precipitation in the Autumn (from September to November) is usually less than 10% of the total annual precipitation.

There is a clear difference in temperature during the four distinct seasons: Spring (MAM, March–April–May) with an average temperature of approximately 12 °C, Summer (June–July–August) with maximum temperatures exceeding 33 °C leading to heatwaves, Autumn (SON, September–October–November) with an average temperature of approximately 20 °C, and Winter (DJF, December–January–February) because of influences from the Siberian high-pressure system with an average temperature of approximately 1.2 °C. Based on the aim of the study, five major river basins of South Korea (Yeongsan, Seomjin, Geum, Nakdong, and Han) with 56 weather stations, as shown in Figure 1, were primarily considered. The names of the weather stations are summarized in Table 1.

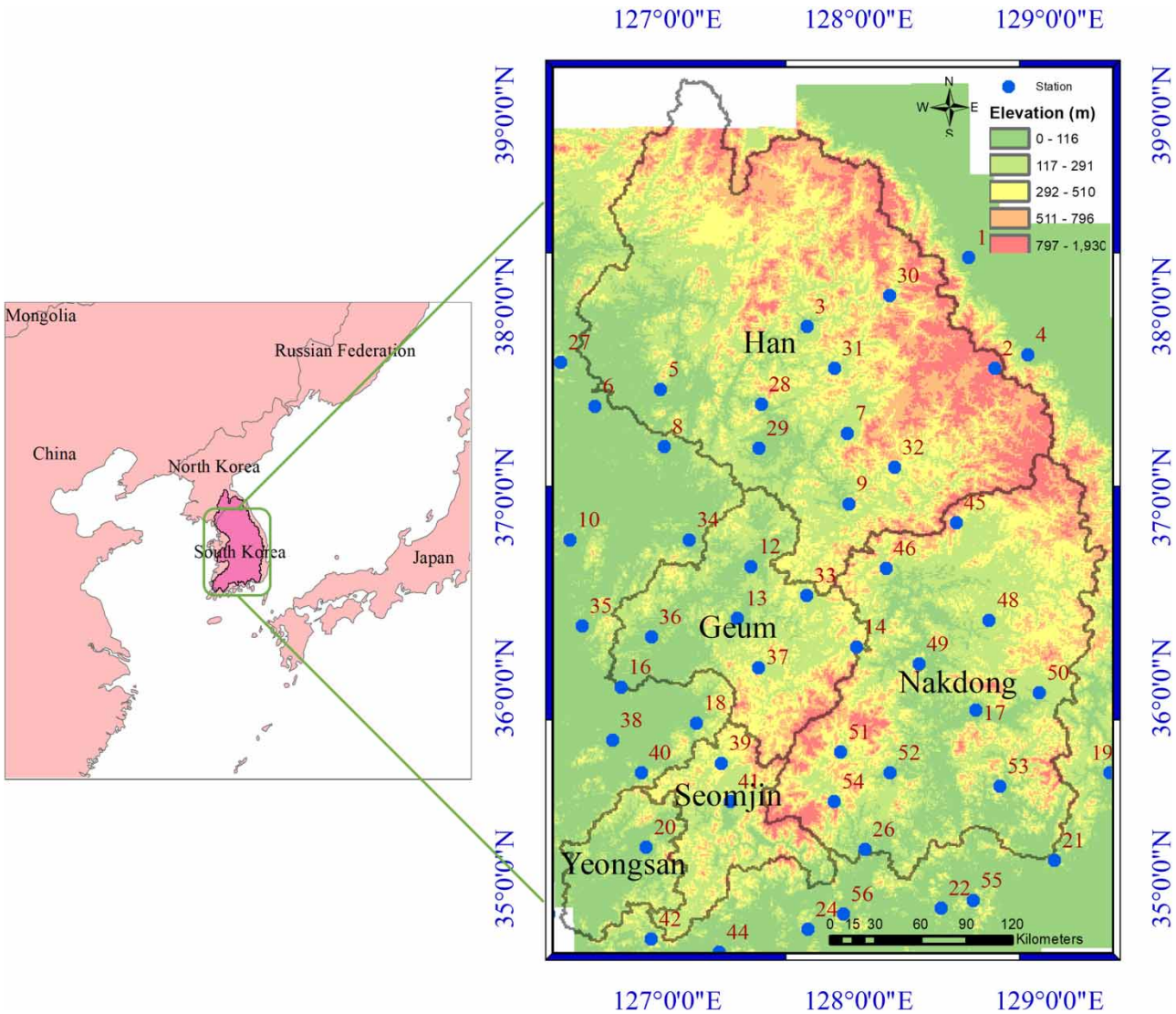

Figure 1 | Map showing South Korea and weather stations. The monthly precipitation data at 56 weather stations compiled by the Korea Meteorological Administration (KMA) are used to create spatial precipitation data using the inverse distance weighted (IDW) approach.

Table 1 | Weather stations involved in this study

No.	Code	Name	Lat. (N)	Lon. (E)	No.	Code	Name	Lat. (N)	Lon. (E)
1	90	Sokcho	128.5814	38.2648	29	203	Icheon	127.4842	37.264
2	100	Daeqwalyeong	128.7183	37.6772	30	211	Inje	128.1671	38.0599
3	101	Chuncheon	127.7357	37.9026	31	212	Hongcheon	127.8804	37.6836
4	105	Gangneung	128.891	37.7515	32	221	Jecheon	128.1943	37.1593
5	108	Seoul	126.9658	37.5714	33	226	Boeun	127.7341	36.4876
6	112	Incheon	126.6244	37.4776	34	232	Cheonan	127.1192	36.7767
7	114	Wonju	127.9466	37.3376	35	235	Boryeong	126.5574	36.3272
8	119	Suwon	126.9856	37.2728	36	236	Buyeo	126.9208	36.2724
9	127	Chungju	127.9527	36.9704	37	238	Geumsan	127.4817	36.1056
10	129	Seosan	126.4939	36.7766	38	243	Buan	126.7166	35.7295
11	130	Uljin	129.4128	36.9918	39	244	Imsil	127.2856	35.6123
12	131	Cheongju	127.4407	36.6392	40	245	Jeongeup	126.8661	35.5632
13	133	Daejeon	127.3721	36.372	41	247	Namwon	127.333	35.4054
14	135	Chupungryeong	127.9946	36.2202	42	260	Jangheung	126.9195	34.6887
15	138	Pohang	129.3796	36.0326	43	261	Haenam	126.569	34.5536
16	140	Gunsan	126.7613	36.0053	44	262	Goheung	127.2757	34.6182
17	143	Daegu	128.619	35.8852	45	272	Yeongju	128.517	36.8719
18	146	Jeonju	127.155	35.8215	46	273	Mungyeong	128.1488	36.6273
19	152	Ulsan	129.3203	35.5601	47	277	Yeongdeok	129.4094	36.5333
20	156	Gwangju	126.8916	35.1729	48	278	Uiseong	128.6886	36.3561
21	159	Busan	129.032	35.1047	49	279	Gumi	128.3205	36.1306
22	162	Tongyeong	128.4356	34.8455	50	281	Yeongcheon	128.9514	35.9774
23	165	Mokpo	126.3812	34.8169	51	284	Geochang	127.911	35.6712
24	168	Yeosu	127.7406	34.7393	52	285	Hapcheon	128.1699	35.565
25	170	Wan-do	126.7018	34.3959	53	288	Miryang	128.7441	35.4915
26	192	Jinju	128.04	35.1638	54	289	Sancheong	127.8791	35.413
27	201	Ganghwa	126.4463	37.7074	55	294	Geoje	128.6045	34.8882
28	202	Yangpyeong	127.4945	37.4886	56	295	Namhae	127.9264	34.8166

2.2. Weather station and CMIP6 datasets

The monthly precipitation data at 56 weather stations compiled by the Korea Meteorological Administration (KMA) were used to create spatial precipitation data using the inverse distance weighted (IDW) approach. The monthly simulated precipitation of 32 available CMIP6 GCMs for ensemble run *r1i1p1f1* was obtained from http://cmip-pcmdi.llnl.gov/cmip6/ for the historical period of 1973–2014 in terms of considering the validity and reliability of GCM outputs. Based on the Coordinate Reference System for Korea, 32 GCMs in the CMIP6, as summarized in Table 2, were adopted in this research. For an effective comparison with point rainfalls, this study used a bilinear interpolation method to map the point rainfalls into a high spatial resolution $0.125° \times 0.125°$ grid.

3. RESEARCH PROCESS AND METHODOLOGY

3.1. Research process

To facilitate the comparison, the observed precipitation across weather stations and simulated precipitation from the GCMs are first linearly interpolated to the same spatial grid of a 0.125° resolution, as shown in Figure 2. The historical period of 1973–2014 is selected for analysis, evaluation, and comparison between the monthly observed precipitation and simulated

Table 2 | Thirty-two CMIP6 GCMs adopted in this study

No.	Model	Institution	Country	Longitude (degree)	Latitude (degree)
1	ACCESS-CM2	CSIRO and ARCCSS	Australia	1.25	1.875
2	ACCESS-ESM1-5	CSIRO and ARCCSS	Australia	1.25	1.875
3	AWI-CM-1-1-MR	Alfred Wegener Institute Bremerhaven (AWI)	Germany	0.937	0.934
4	AWI-ESM-1-1-LR	Alfred Wegener Institute Bremerhaven (AWI)	Germany	1.875	1.86
5	BCC-CSM2-MR	Beijing Climate Center, Beijing (BCC)	China	1.125	1.121
6	BCC-ESM1	Beijing Climate Center, Beijing (BCC)	China	2.812	2.788
7	CAMS-CSM1-0	Chinese Academy of Meteorological Sciences	China	1	1
8	CanESM5	Canadian Centre for Climate Modelling and Analysis Environment and Climate Change Canada	Canada	2.812	2.789
9	CAS-ESM2-0	Chinese Academy of Sciences, Beijing	China	1.406	1.417
10	CESM2-FV2	National Center for Atmospheric Research	USA	1.9	2.5
11	CESM2-WACCM-FV2	National Center for Atmospheric Research	USA	1.9	2.5
12	CESM2-WACCM	National Center for Atmospheric Research	USA	0.9	1.25
13	CESM2	National Center for Atmospheric Research	USA	0.9	1.25
14	CMCC-CM2-HR4	Fondazione Centro Euro-Mediterraneo sui Cambiamenti Climatici	Italy	1	1
15	CMCC-CM2-SR5	Fondazione Centro Euro-Mediterraneo sui Cambiamenti Climatici	Italy	1	1
16	CMCC-ESM2	Fondazione Centro Euro-Mediterraneo sui Cambiamenti Climatici	Italy	1.25	0.942
17	E3SM-1-1-ECA	Lawrence Livermore National Laboratory (LLNL), Livermore	USA	1	1
18	E3SM-1-1	Lawrence Livermore National Laboratory (LLNL), Livermore	USA	1	1
19	EC-Earth3-AerChem	EC-Earth-Consortium	Europe	0.35	0.35
20	EC-Earth3-CC	EC-Earth-Consortium	Europe	0.35	0.35
21	EC-Earth3-Veg-LR	EC-Earth-Consortium	Europe	0.35	0.35
22	EC-Earth3-Veg	EC-Earth-Consortium	Europe	0.35	0.35
23	EC-Earth3	EC-Earth-Consortium	Europe	0.35	0.35
24	FGOALS-f3-L	Chinese Academy of Sciences, Beijing	China	1	1
25	FGOALS-g3	Chinese Academy of Sciences, Beijing	China	2	2
26	FIO-ESM-2-0	First Institute of Oceanography, Ministry of Natural Resources, Qingdao	China	0.9	1.25
27	GFDL-CM4	NOAA-GFDL	USA		
28	GFDL-ESM4	NOAA-GFDL	USA	1.25	1
29	GISS-E2-1-G-CC	Goddard Institute for Space Studies (NASA-GISS), New York	USA	1.25	1.25
30	GISS-E2-1-G	Goddard Institute for Space Studies (NASA-GISS), New York	USA	1.25	1.25
31	GISS-E2-1-H	Goddard Institute for Space Studies (NASA-GISS), New York	USA	1.25	1.25
32	GISS-E2-2-H	Goddard Institute for Space Studies (NASA-GISS), New York	USA	1.25	1.25

precipitation of 32 CMIP6 GCMs at all grid points using four spatial metrics, especially for mean annual precipitation (MAP) and mean monsoon precipitation (MMP): Cramer's *V*, SPAEF, KGE, and FSS. The GCMs are then ranked separately according to each indicator of the monthly precipitation reconstruction skill. Finally, the overall performance ranking of 32 GCMs is evaluated by comprehensive rating metrics (RM) to identify the best subset of GCMs. The performance of the selected GCMs

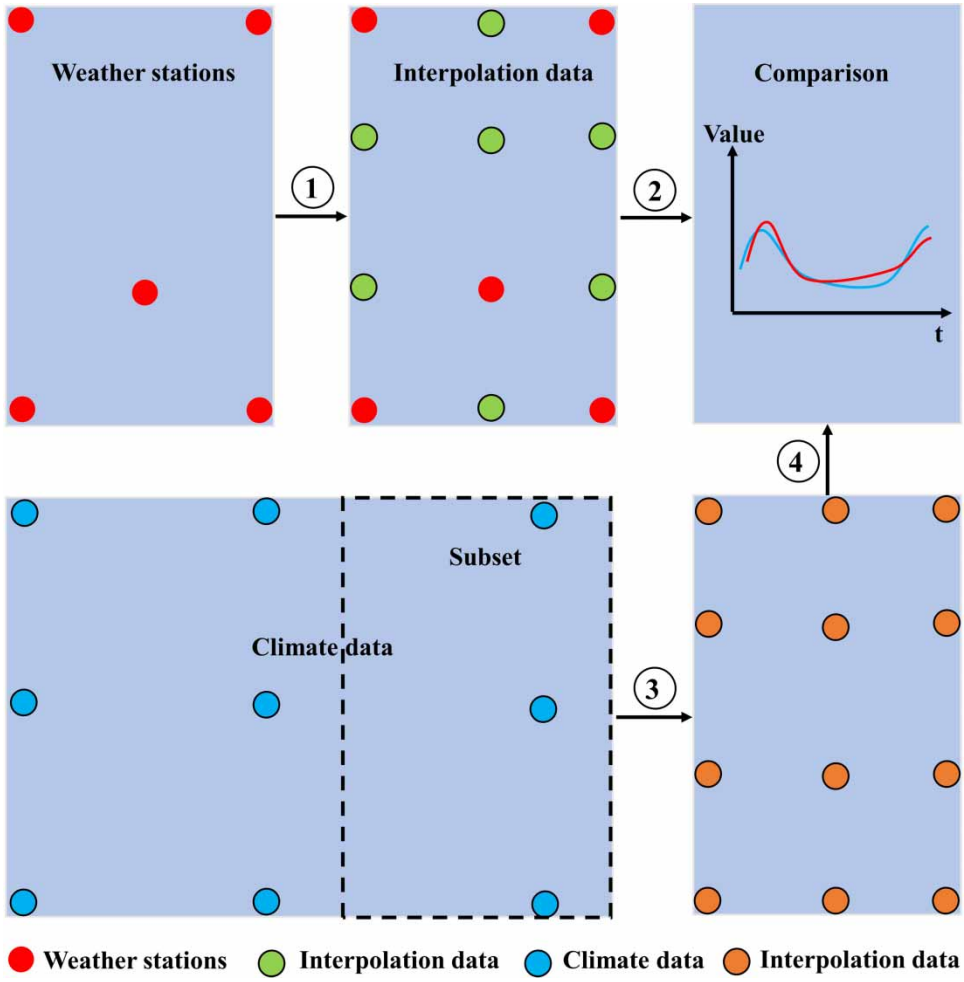

Figure 2 | Interpolation procedure of climate data from weather stations and CMIP6 GCMs. This study used a bilinear interpolation method to map the point rainfalls into a high spatial resolution 0.125° × 0.125° grid.

is also further evaluated by comparing the ability to reproduce the observed monsoon seasonal precipitation based on the biased precipitation obtained from the best-performing GCMs.

3.2. Spatial performance indicators

This study considered four indicators for the spatial representation of GCM models: Cramer's V, SPAEF, KGE, and FSS, to evaluate the association, variability, and error between observational data and GCMs output data.

3.2.1. Cramer's V

Cramer's V is a statistical test based on Pearson's χ^2 statistics to evaluate the strength of the relationship between two categorical variables, as written in Equation (1). Cramer's V ranges from 0 to 1 without any negative values, with 0 representing no relationship and 1 representing perfect association:

$$V = \sqrt{\frac{\chi^2 n^{-1}}{\min(c-1,\ r-1)}} \tag{1}$$

where χ^2 is the chi-square value, n is the number of data, and c and r are the numbers of columns and rows of the dataset. In our study, c is 2, representing the number of variables (i.e., observed and simulated precipitation), and $r = 511{,}056$ (the number of rows of data).

3.2.2. SPAtial EFficiency metric

SPAEF is a multiple-component spatial performance metric. SPAEF can be used to compare spatial patterns in two raster maps considering three distinct and complementary components of the (1) coefficient of variation (CV), (2) CC, and (3) histogram overlap, as written in Equation (2). The SPAEF value varies between $-\infty$ and 1, where a value closer to 1 refers to better agreement between observed and simulated data:

$$\mathrm{SPAEF} = 1 - \sqrt{(\alpha - 1)^2 + (\beta - 1)^2 + (\gamma - 1)^2} \tag{2}$$

where α represents the Pearson CC between the observed (obs) and GCMs simulated (sim) precipitation, as shown in Equation (3):

$$\alpha = \rho(\mathrm{obs},\ \mathrm{sim}) = \frac{\sum (x_i - \mu_G)(y_i - \mu_O)}{\sqrt{\sum (x_i - \mu_G)^2 \sum (y_i - \mu_O)^2}} \tag{3}$$

where x_i and y_i are the values of GCM-simulated precipitation and observation precipitations at ith, respectively. μ_G and μ_O are the mean values of GCM-simulated and observed precipitations, respectively. β is the CV that represents spatial variability, as written in Equation (4):

$$\beta = \mathrm{CV} = \frac{\sigma_G/\mu_G}{\sigma_O/\mu_O} \tag{4}$$

where σ_G and σ_O are the standard deviations of GCM-simulated and observed precipitations. γ represents the histogram overlap that is calculated by Equation (5):

$$\gamma = \frac{\sum_{j-1}^{b} \min(H_j,\ L_j)}{\sum_{j-1}^{b} H_j} \tag{5}$$

where H, L, and b represent histogram values of observations, histogram values of simulations, and the number of bins in a histogram, respectively.

3.2.3. Fractions skill score

FSS is a statistical measure of the spatial representation between the observed and simulated precipitation (Zhao & Zhang 2018), as written in Equations (6)–(8). FSS ranges between 0 and 1, and a value closer to 1 indicates a greater similarity in the spatial patterns over simulated and observed precipitations:

$$\mathrm{FSS} = 1 - \frac{\mathrm{MSE}_n}{\mathrm{MSE}_{n(\mathrm{wc})}} \tag{6}$$

$$\mathrm{MSE}_n = \frac{1}{N_{xy}} \sum_{i=1}^{N_x} \sum_{j=1}^{N_y} [\mathrm{ref}_{(n)ij} - \mathrm{scen}_{(n)ij}]^2 \tag{7}$$

$$\mathrm{MSE}_{n(\mathrm{wc})} = \frac{1}{N_{xy}} \left[\sum_{i=1}^{N_x} \sum_{j=1}^{N_y} \mathrm{ref}^2_{(n)ij} + \sum_{i=1}^{N_x} \sum_{j=1}^{N_y} \mathrm{scen}^2_{(n)ij} \right] \tag{8}$$

where MSE_n is the mean squared error between the observation and simulation fractions, defined as the proportion of the grid covered by precipitation compared with the full grid. Then, MSE is normalized with the worst case $\mathrm{MSE}_{n(\mathrm{wc})}$ representing zero agreement about the spatial patterns. N_x and N_y are the number of columns and rows in the observation or simulation map, respectively. The $\mathrm{ref}_{(n)}$ and $\mathrm{scen}_{(n)}$ are the observed and GCMs output fractions, respectively.

3.2.4. Kling–Gupta efficiency

The KGE is characterized by three main involved components: (1) the Pearson CC, (2) the ratio between the mean of observed and simulated precipitations, and (3) the ratio between the CV of the observed and simulated precipitations.

$$\mathrm{KGE} = 1 - \sqrt{(\alpha - 1)^2 + (\beta - 1)^2 + (\gamma - 1)^2} \tag{9}$$

α represents the spatial correlation coefficients based on the Pearson CC, and β indicates the relative variability based on the ratio of mean values of GCM-simulated precipitation (μ_G) to the mean of observed precipitation (μ_O):

$$\beta = \frac{\mu_G}{\mu_O} \tag{10}$$

γ represents the ratio between the standard deviation of GCM values (σ_G) to the standard deviation of observation values (σ_O), as written in Equation (11):

$$\gamma = \frac{\sigma_G}{\sigma_O} \tag{11}$$

The value of KGE varies between $-\infty$ and 1, and a value closer to 1 indicates better agreement between observations and simulations.

3.3. Comprehensive rating metric

After evaluating the performance of each GCM according to multiple metrics, the challenge is to assess the overall ranking because each GCM has a different rank for each evaluation index. To identify the desired subset of GCMs, this study used the information aggregation approach that combines the results to total ranking to find the top group of GCMs. A comprehensive rating metric (Chen *et al.* 2011; Jiang *et al.* 2015) of each GCM for four indices using Equation (12) to obtain the overall ranks of the GCMs:

$$\mathrm{RM} = 1 - \frac{1}{mn}\mathrm{rank}_i \tag{12}$$

where n and m represent the number of metrics and models, respectively. The rank_i indicates the rank of GCM based on each index ith. Equation (12) allows us to calculate the RM for the CMIP6 models with respect to Cramer's V, SPAEF, KGE, and FSS.

4. RESULTS AND DISCUSSION

4.1. Assessment of spatial pattern of MAP and MMP

Figure 3 represents the observed and CMIP6 GCM-simulated MAPs over the historical period 1973–2014, with the spatial correlation based on Pearson CC. The observed MAP in the study area is in the range of 1,000–1,800 mm. The higher MAP (1,600–1,800 mm) was observed in the eastern part of the Han River watershed. The average MAP (1,300–1,500 mm) was observed in the southern part of the Nakdong and Seomjin watersheds, while the relatively lower precipitation (1,000–1,100 mm) was observed in the central and northern parts of the Nakdong watershed.

The spatial pattern of the MAP from some of the GCMs (e.g., AWI-ESM-1-1-LR (0.4), E3SM-1-1-ECA (0.35), BCC-CSM2-MR (0.31), and CESM2-WACCM (0.28)), which is the model group with the highest correlation coefficients, appears to be similar to that of the observed, representing a better representation of the higher MAP in the lower Nakdong and Seomjin watersheds. Among all GCMs, AWI-ESM-1-1-LR showed the best performance in terms of representing the spatial distribution of MAP with a high CC of 0.4 that reproduces the precipitation in the Nakdong watershed. Alternatively, several GCMs appear to be a biased representation of the spatial distribution of MAP, highlighting the higher precipitation in the upper Nakdong watershed. Moreover, most GCMs fail to reproduce the spatial pattern of MAP in the eastern part of the Han River watershed, demonstrating underestimation or overestimation of the precipitation in the Han River and Geum River watersheds. The CESM2-WACCM tends to overestimate MAP in most Nakdong River watersheds with

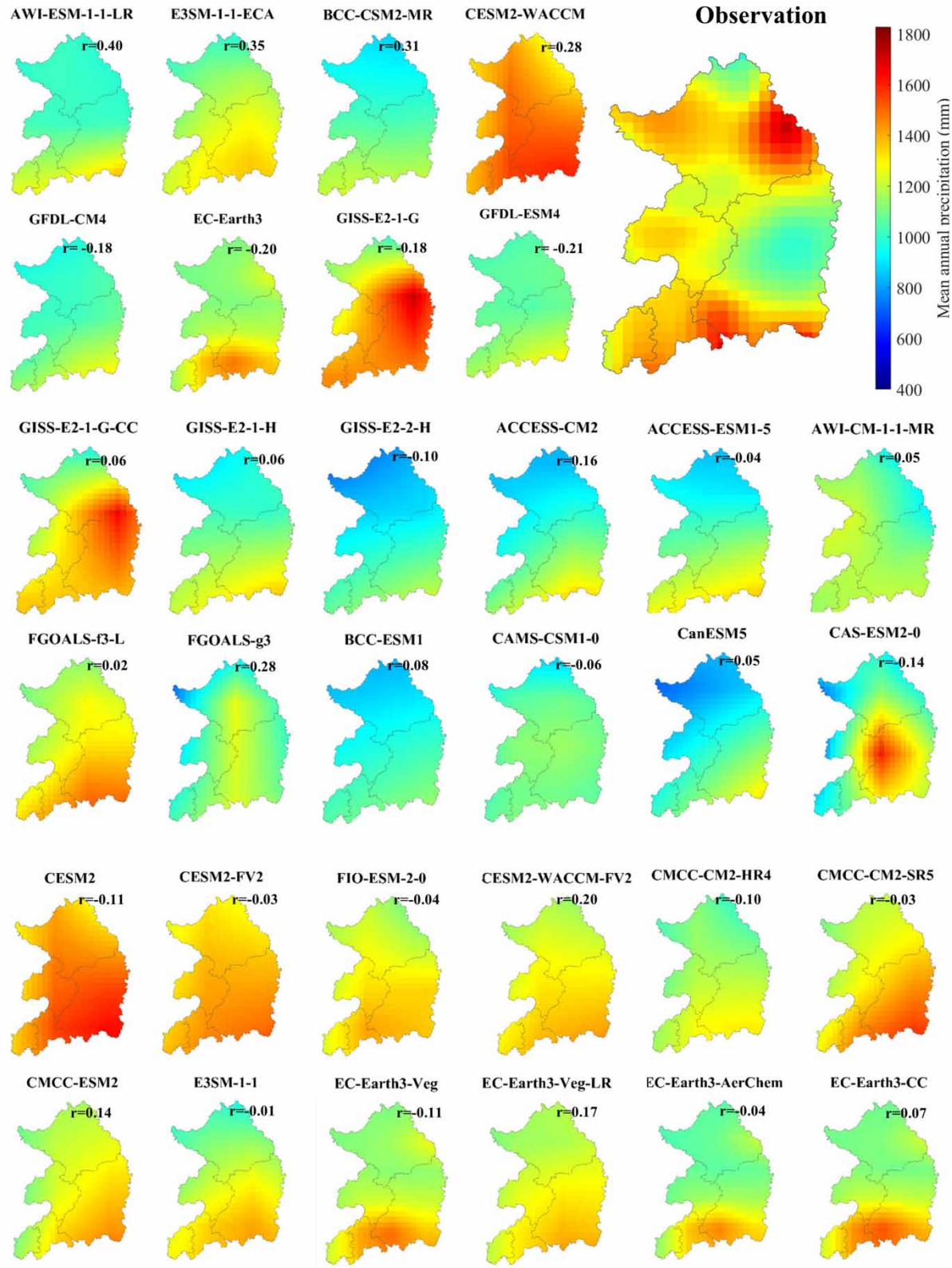

Figure 3 | Spatial distribution of mean annual precipitation (MAP) over the historical period 1973–2014 obtained from the observed and GCM-simulated precipitations. The spatial correlation coefficients (*r*) between the observed and simulated precipitations are illustrated for each GCM.

1,400–1,700mm, while the observed MAP is about 900–1,100 mm. Furthermore, the model group with the lower correlation coefficients (e.g., GFDL-CM4, GISS-E2–1-G, EC-Earth3, and GFDL-ESM4) showed a spatially biased representation of the MAP. In summary, the GCMs have somewhat different simulation capabilities representing the spatial distribution of the MAP for each watershed. With this perspective, exploration of spatial performance over different CMIP6 GCMs can provide an effective means of selecting GCMs for hydrological study over the major river watersheds in South Korea.

The majority of the total annual precipitation in South Korea is concentrated during the East Asian monsoon season (from June to the end of July). Assessing the distribution of MMP also plays a vital role in long-term hydrological forecasting studies. The observed and CMIP6 GCM-simulated MMPs over the historical period 1973–2014, with the spatial correlation based on Pearson CC, are illustrated in Figure 4. The relatively higher MMPs with a range of 500–600 mm are clearly observed in the Han River, Seomjin River, and lower Nakdong River watersheds, while the lower MMPs (350–400 mm) are distributed over the upper Nakdong watershed. As shown in Figure 4, the model group with higher correlation coefficients (BCC-CSM2-MR, FGOALS-g3, AWI-ESM-1-1-LR, and EC-Earth3-CC) still demonstrates a limited performance in representing MMP in terms of spatial distribution and rainfall amounts. Specifically, the EC-Earth3-CC model is able to describe the observed distribution of MMP quite well in the Seomjin and Nakdong watersheds but fails to reproduce the MMP for the rest of the areas. The AWI-ESM-1-1-LR model significantly underestimates MMP (i.e., 100–200 mm) compared with that of the observed (i.e., 350–600 mm). Meanwhile, some GCMs with lower correlation coefficients (FIO-ESM-2-0, CESM2, and CESM2-WACCM) showed better performance in terms of representing MMP with the range of 400–550 mm. Overall, the remaining models are able to reproduce the entire range of MMP depending on each area except for models such as AWI-ESM-1-1-LR, GISS-E2-1-H, GISS-E2-2-H, GISS-E2-1-G, CanESM5, and ACCESS-CM2 in all watersheds.

In general, some GCMs may perform better than others for specific regions, even if they do not have higher spatial correlation coefficients. The GCMs with the higher correlation coefficients may indicate overall good performance, but they may not accurately reproduce the observed distribution of MAP (or MMP) with the entire range of precipitation seen in the individual region. Similarly, the GCMs with the lower correlation coefficients may not necessarily indicate poor performance, especially if the model can reproduce important regional features and climate variability. Therefore, exploration of the spatial performance and combining multiple metrics over a large number of CMIP6 GCMs provide an effective means of selecting GCMs for hydrological study over the major river watersheds in South Korea.

4.2. Ranking of GCMs based on different indicators and comprehensive rating metric

The ranking of GCMs based on each spatial metric is first demonstrated. A comprehensive rating metric (RM) integrating four metrics is then considered to find the overall ranking of 32 GCMs and the subset of GCMs for both MAP and MMP. The overall ranking of 32 GCMs with RM values for the MAP is summarized in Table 3. They can be divided into three groups based on their RM values: Tier-1 (0.61–0.73) with 9 GCMs, Tier-2 (0.42–0.58) with 14 GCMs, and Tier-3 (0.19–0.37) with 9 GCMs. The range of ranking for each GCM was found to be quite variable over four indicators, leading to a narrower range of the RM value due to the mixed effect of different rankings associated with each indicator. The different rankings over four indicators could be due to inconsistent simulation results in various aspects of the spatial pattern seen in the observed climate system. For example, the GCM models in Tier-1 (e.g., ACCESS-ESM1-5, FGOALS-g3, AWI-ESM-1-1-LR, and ACCESS-CM2) noticeably underestimate MAPs over different regions for each model. Moreover, the models in Tier-1 showed a limited capability in simulating the extreme rainfalls in the Han River, Seomjin River, and Nakdong River watersheds. As illustrated, approximately 60% of the annual precipitation in South Korea occurred during the monsoon season from June to the end of July. Thus, the GCM selection process based on model performance during the monsoon season would be more informative than that of the annual.

The RM values of 32 GCMs based on the ranking of various spatial metrics for the spatial MMP distribution performance are summarized in Table 4. According to the KGE metrics, the FIO-ESM-2-0 model appears to be most relevant with the highest value (0.47), and then CESM2-WACCM, CESM2, and CMCC-ESM2 models are pertinent with a relatively higher KGE value ($>$0.43). It can be concluded that those three models show better performance in terms of the KGE value with a higher association of the observed precipitation. Alternatively, AWI-ESM-1-1-LR, GISS-E2-1-G-CC, GISS- E2-1-H, and GISS-E2-2-H models have fairly lower KGE values ($\leq$0.10). Moreover, these GCMs appear to be relatively lower SPAEF values, demonstrating poor representation of the spatial pattern of the historical precipitation. Higher SPAEF values were observed for FGOALS-g3, FIO-ESM-2-0, and BCC-CSM2-MR, with a value greater than 0.4, indicating a better representation of spatial patterns of precipitation over South Korea. The best performance in terms of the FSS metric was observed for the

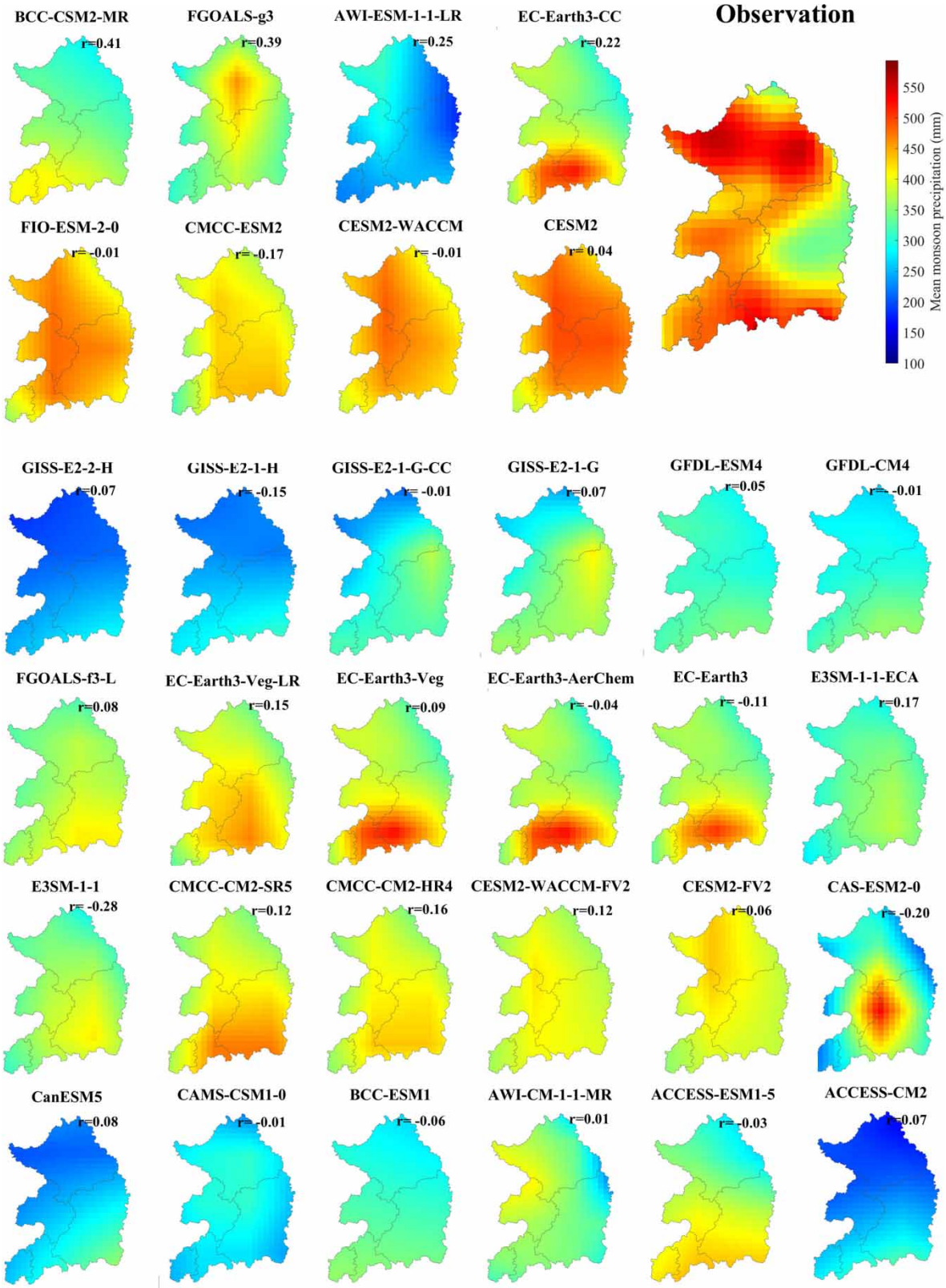

Figure 4 | Spatial distribution of mean monsoon precipitation (MMP) over the historical period 1973–2014 obtained from the observed and GCM-simulated precipitations. The spatial correlation coefficients (*r*) between the observed and simulated precipitations are illustrated for each GCM.

Table 3 | RM values of 32 GCMs based on the ranking of a variety of spatial metrics for the spatial MAP distribution performance

No.	Models	KGE	FSS	Cramer's V	SPAEF	RM	Overall Ranking
1	ACCESS-CM2	0.15 (5)	0.30 (19)	0.16 (5)	0.15 (2)	0.76	1
2	CESM2-WACCM	0.16 (2)	0.54 (0)	0.14 (21)	0.00 (14)	0.71	2
3	FGOALS-f3-L	0.08 (9)	0.46 (6)	0.15 (14)	0.03 (10)	0.70	3
4	CESM2-WACCM-FV2	0.15 (4)	0.51 (2)	0.14 (28)	0.07 (7)	0.68	4
5	E3SM-1-1-ECA	0.22 (1)	0.42 (10)	0.13 (29)	0.18 (1)	0.68	5
6	ACCESS-ESM1-5	0.06 (13)	0.33 (17)	0.17 (1)	0.03(11)	0.67	6
7	FGOALS-g3	0.14 (6)	0.22 (24)	0.15 (12)	0.13 (3)	0.65	7
8	AWI-ESM-1-1-LR	0.25 (0)	0.34 (16)	0.13 (30)	0.24 (0)	0.64	8
9	CMCC-ESM2	0.11 (8)	0.44 (9)	0.14 (23)	0.08 (6)	0.64	9
10	GISS-E2-1-G-CC	0.01 (19)	0.40 (12)	0.17 (3)	-0.03 (16)	0.61	10
11	EC-Earth3-Veg-LR	0.13 (7)	0.47 (4)	0.13 (31)	0.04 (9)	0.60	11
12	CMCC-CM2-SR5	0.07 (12)	0.47 (5)	0.15 (20)	−0.02 (15)	0.59	12
13	EC-Earth3-CC	0.08 (10)	0.45 (7)	0.14 (26)	0.02 (12)	0.57	13
14	BCC-CSM2-MR	0.16 (3)	0.07 (22)	0.14 (27)	0.09 (4)	0.56	14
15	GISS-E2-1-H	0.03 (16)	0.30 (18)	0.16 (4)	−0.04 (20)	0.55	15
16	CanESM5	0.05 (15)	0.19 (28)	0.15 (11)	0.09 (5)	0.55	16
17	CESM2	−0.07 (26)	0.52 (1)	0.16 (6)	−0.10 (25)	0.55	17
18	EC-Earth3-AerChem	0.04 (15)	0.29 (20)	0.15 (17)	0.05 (8)	0.53	18
19	E3SM-1-1	0.03 (17)	0.37 (14)	0.16 (9)	−0.06 (21)	0.52	19
20	EC-Earth3	0.01 (20)	0.36 (15)	0.15 (16)	0.01 (13)	0.50	20
21	AWI-CM-1-1-MR	0.07 (11)	0.27 (21)	0.15 (18)	−0.03 (18)	0.47	21
22	CESM2-FV2	−0.02 (22)	0.49 (3)	0.15 (19)	−0.11 (26)	0.45	22
23	GISS-E2-1-G	−0.11 (30)	0.44 (8)	0.17 (2)	−0.24 (31)	0.45	23
24	FIO-ESM-2-0	0.02 (18)	0.41 (11)	0.14 (24)	−0.08 (24)	0.40	24
25	CAS-ESM2-0	−0.11 (29)	0.21 (25)	0.18 (0)	−0.14 (28)	0.36	25
26	GFDL-CM4	−0.07 (27)	0.20 (27)	0.15 (10)	−0.04 (19)	0.35	26
27	GISS-E2-2-H	−0.04 (24)	0.11 (30)	0.16 (7)	−0.07 (22)	0.35	27
28	EC-Earth3-Veg	−0.02 (23)	0.39 (13)	0.14 (22)	−0.12 (27)	0.34	28
29	CAMS-CSM1-0	−0.09 (28)	0.14 (29)	0.16 (8)	−0.07 (23)	0.31	29
30	CMCC-CM2-HR4	−0.05 (25)	0.26 (23)	0.15 (15)	−0.15 (29)	0.28	30
31	BCC-ESM1	0.01 (21)	0.10 (31)	0.14 (25)	−0.03 (17)	0.27	31
32	GFDL-ESM4	−0.18 (31)	0.21 (26)	0.15 (13)	−0.18 (30)	0.22	32

Values in parentheses represent the ranking for evaluation metrics.

FIO-ESM-2-0 model (0.66). Similarly, CMCC-ESM2, EC-Earth3-Veg, and CESM2-WACCM models have comparably greater FSS values than the others, with a value greater than 0.60. A relatively lower FSS value was found for AWI-ESM-1-1-LR, GISS-E2-1- G-CC, GISS-E2-1-H, and GISS-E2-2-H models. Alternatively, conflicting results were identified for GISS-E2-1-H, GISS-E2-1-G-CC, and AWI-ESM-1-1-LR models in terms of Cramer's V metric, with a favorable result. However, there is a very narrow range of Cramer's V between 0.38 and 0.48 in our case, indicating that the metric is not sensitive to the GCMs considered in this study. Moreover, Cramer's V metric is only based on the strength of the relationship between the binary variables (0 or 1) with the limited number of grids, such that the sensitivity of the metric could be less than others.

The overall ranking of 32 GCMs is also summarized in Table 4. The ranking of GCMs based on each metric is readily available, and the overall ranking combining four indices is also required to differentiate the subset of GCMs. In this perspective, the study employed the RM metric to effectively define the overall ranking of 32 GCMs. The four top-ranked

Table 4 | RM values of 32 GCMs based on the ranking of a variety of spatial metrics for the spatial MMP distribution performance

No.	Models	KGE	FSS	Cramer's V	SPAEF	RM	Overall Ranking
1	FIO-ESM-2-0	0.47 (0)	0.66 (0)	0.41 (20)	0.45 (2)	0.84	1
2	CESM2-WACCM	0.44 (1)	0.63 (3)	0.41 (18)	0.33 (9)	0.76	2
3	CESM2	0.43 (3)	0.63 (5)	0.41 (15)	0.35 (8)	0.76	3
4	CMCC-ESM2	0.43 (2)	0.65 (1)	0.40 (27)	0.36 (6)	0.72	4
5	CMCC-CM2-HR4	0.36 (13)	0.62 (7)	0.41 (17)	0.35 (7)	0.66	5
6	EC-Earth3-CC	0.42 (4)	0.63 (6)	0.38 (31)	0.36 (5)	0.64	6
7	EC-Earth3-Veg	0.42 (5)	0.65 (2)	0.39 (29)	0.33 (10)	0.64	7
8	CESM2-WACCM-FV2	0.39 (8)	0.60 (12)	0.41 (14)	0.30 (14)	0.63	8
9	CMCC-CM2-SR5	0.41 (7)	0.63 (4)	0.40 (26)	0.31 (13)	0.61	9
10	EC-Earth3-AerChem	0.38 (11)	0.61 (8)	0.39 (28)	0.37 (3)	0.61	10
11	ACCESS-ESM1-5	0.30 (18)	0.54 (21)	0.44 (10)	0.37 (4)	0.59	11
12	CESM2-FV2	0.39 (9)	0.61 (9)	0.41 (16)	0.25 (18)	0.59	12
13	BCC-CSM2-MR	0.33 (16)	0.58 (17)	0.41 (21)	0.41 (2)	0.56	13
14	EC-Earth3-Veg-LR	0.41 (6)	0.60 (13)	0.40 (25)	0.32 (12)	0.56	14
15	FGOALS-g3	0.38 (12)	0.59 (15)	0.39 (30)	0.50 (0)	0.55	15
16	AWI-CM-1-1-MR	0.29 (19)	0.57 (19)	0.44 (9)	0.29 (15)	0.52	16
17	EC-Earth3	0.39 (10)	0.59 (16)	0.40 (24)	0.32 (11)	0.52	17
18	E3SM-1-1-ECA	0.34 (15)	0.61 (11)	0.40 (22)	0.25 (17)	0.49	18
19	E3SM-1-1	0.34 (14)	0.61 (10)	0.40 (23)	0.22 (20)	0.48	19
20	FGOALS-f3-L	0.32 (17)	0.57 (18)	0.43 (11)	0.21 (23)	0.46	20
21	BCC-ESM1	0.22 (22)	0.59 (14)	0.41 (19)	0.25 (16)	0.45	21
22	GFDL-ESM4	0.29 (20)	0.55 (20)	0.42 (13)	0.22 (19)	0.44	22
23	CAS-ESM2-0	0.20 (23)	0.53 (23)	0.44 (8)	0.21 (22)	0.41	23
24	ACCESS-CM2	0.17 (24)	0.47 (29)	0.46 (4)	0.22 (21)	0.39	24
25	GFDL-CM4	0.24 (21)	0.54 (22)	0.43 (12)	0.18 (24)	0.38	25
26	CAMS-CSM1-0	0.16 (25)	0.48 (26)	0.46 (6)	0.13 (26)	0.35	26
27	GISS-E2-1-G	0.14 (26)	0.49 (24)	0.47 (3)	-0.08 (30)	0.35	27
28	AWI-ESM-1-1-LR	0.08 (29)	0.47 (27)	0.47 (2)	-0.02 (27)	0.34	28
29	CanESM5	0.13 (27)	0.48 (25)	0.45 (7)	0.17 (25)	0.34	29
30	GISS-E2-1-G-CC	0.10 (28)	0.47 (28)	0.47 (1)	−0.04 (28)	0.34	30
31	GISS-E2-1-H	0.03 (30)	0.41 (31)	0.48 (0)	−0.08 (31)	0.28	31
32	GISS-E2-2-H	0.01 (31)	0.42 (30)	0.46 (5)	−0.08 (29)	0.26	32

Values in parentheses represent the ranking for evaluation metrics.

GCMs based on their overall ranks in reproducing spatial distribution of precipitation over South Korea were identified. The range of the RM values can be classified into three categories: Tier-1 (0.72–0.84): FIO-ESM-2-0, CESM2-WACCM, CESM2, and CMCC-ESM2 models; Tier-3 (0.26–0.39): AWI-ESM-1-1-LR, GISS-E2-1-G-CC, GISS-E2-1-H, GISS-E2-2-H models; and Tier-2 (0.41–0.66): the remaining GCMs. In general, one may expect that rankings of the GCM model can vary over different metrics due to the limited representation of the regional climate system. In this perspective, the use of a comprehensive rating metric could offer a systematic way to better characterize model performance over several GCMs. In this study, four top-ranked GCMs (i.e., FIO-ESM-2-0, CESM2-WACCM, CESM2, and CMCC-ESM2) can be selected as the best-performing GCMs for long-term hydrological projections based on spatial performance assessment metrics across South Korea.

4.3. Evaluation of the bias for the selected GCMs' subset

This section explored spatial patterns of the selected GCMs under the RM metric, as summarized in Table 4. Here, the preferred GCMs in reproducing MMP during the East Asian monsoon season (June–July) are demonstrated for the historical period of 1973–2014. The best-performing GCMs based on RM can better reproduce the spatial pattern of precipitation during monsoon season over the historical period compared with the GCMs in terms of representing the spatial distribution of MMP with a high CC (i.e., BCC-CSM2-MR, FGOALS-g3, AWI-ESM-1-1-LR, and EC-Earth3-CC), as shown in Figure 5. The best-performing GCMs tend to overestimate the MMP, whereas underestimation of the MMP in all watersheds is noticeable in GCMs with a high CC. Specifically, the selected GCMs based on the correlation coefficients for the MMP generally underestimate monsoon precipitation over large areas such as the Han River and Nakdong River watersheds.

Alternatively, the four top-ranked GCMs overestimate the precipitation in South Korea with a systematic bias of approximately 25% (100–150 mm) in the Nakdong River watershed, 10–15% (50–100 mm) in the Han River watershed, 15–25% (100–150 mm) in the Geum River, Yeongsan River, and Seomjin River watersheds. It can be inferred that selecting suitable climate change models based on the spatial performance metrics for the GCMs will yield appropriate results compared with the sole use of the spatial CC. The spatial patterns of the precipitation during monsoon season, simulated by the best-performing GCMs (i.e., ranked 1–4) and worst-performing GCMs (i.e., 29–32), were finally compared with the spatial patterns of observed precipitation, as seen in Figure 6. The best-performing GCMs more plausibly represent spatial patterns than those observed. Alternatively, the worst-performing GCMs showed larger differences than the spatial patterns observed in the precipitation during the historical period. Moreover, a significant underestimation of the precipitation is clearly identified over a large region.

In general, it is clearly evident that there is a substantial difference in rainfall amount between the observed precipitation and that simulated by the best-performing GCMs in the eastern part of the Han River watershed and the nearly entire Nakdong River watershed. This bias may be due to the misrepresentation of the regional climate system. Thus, a relevant bias correction approach for minimizing systematic bias in GCM precipitations is needed prior to their usage for regional impact assessment of climate change on water resources.

This study does not perform bias correction for the data during the model selection process. However, one may expect that the bias correction approach could offer a better option in selecting a set of suitable GCMs. Furthermore, there will be a change in the ranking pattern under the bias correction process. From this perspective, this study included the bias-corrected results for best- and worst-performing GCMs based on a quantile mapping (QM) approach, which is the most common bias

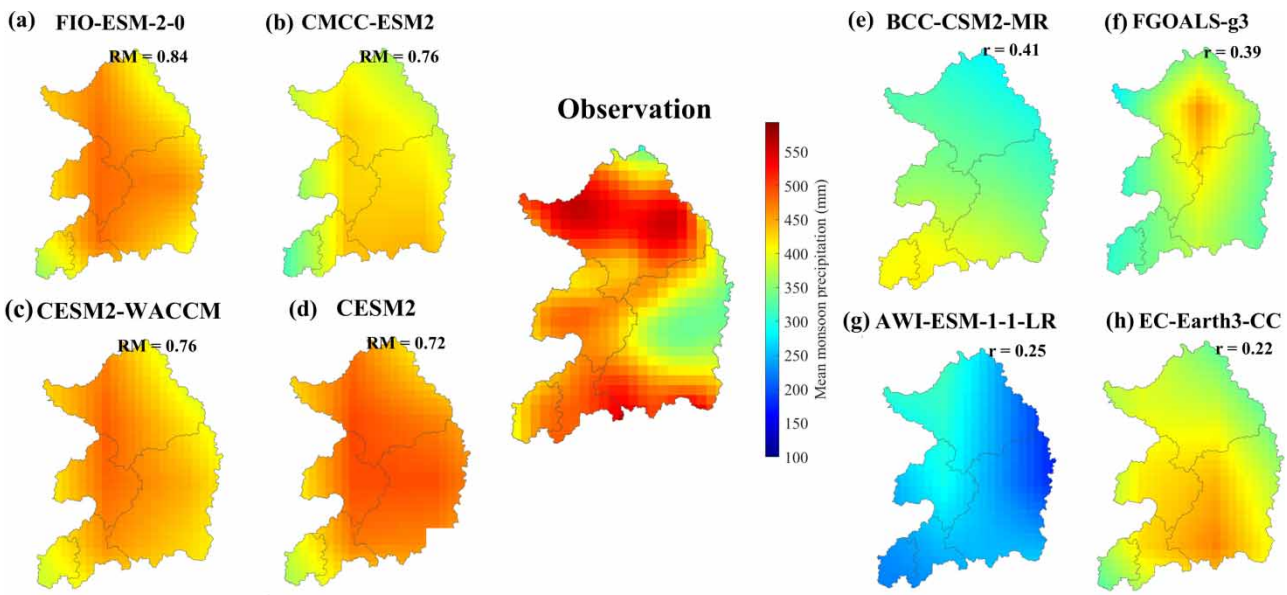

Figure 5 | Spatial patterns of the mean monsoon precipitation (MMP) over the historical period 1973–2014. The MMP distributions, obtained from four top-ranked GCMs based on the RM metric (a–d) and spatial correlation coefficient (e–h), are illustrated. Here, the RM values and spatial correlation coefficients (*r*) are given for each GCM.

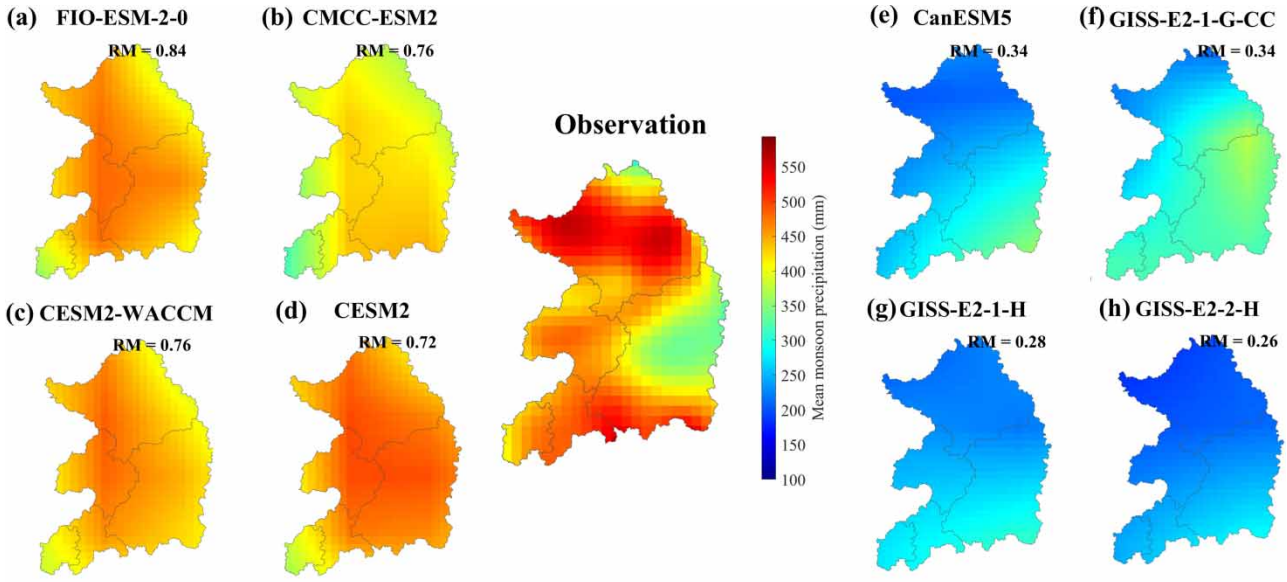

Figure 6 | Spatial patterns of the mean monsoon precipitation (MMP) over the historical period 1973–2014. The MMP distributions, obtained from the best-performing GCMs (a–d) and worst-performing GCMs (e–h) based on the RM metric, are presented. Here, the RM values are given for each GCM.

correction approach in the climate change study. The main idea of the QM-based bias correction for the simulated precipitation is to map its distribution to that of the observed. Thus, it should be noted that a marginal distribution of the bias-corrected precipitation is almost identical to the observed. As expected, the spatial distributions of bias-corrected MMPs for these GCMs are not discernable from those of the observed, regardless of the best- and worst-performing GCMs, as illustrated in Supplementary material, Figures S1 and S2. Under these circumstances, the bias correction for these GCMs is not recommended prior to model selection, and the ranking of GCMs under the bias correction process could be problematic.

5. CONCLUSION

The main objective of this study was to investigate the model performance of CMIP6 GCMs in South Korea in terms of capability to reproduce spatial climate patterns using GoF measures such as the SPAEF, KGE, Cramer's *V*, and FSS. Furthermore, this study proposed an integrated selection approach based on the four spatial performance assessment metrics to better demonstrate changes in precipitation in climate change studies. From this perspective, this study explored the high-performing GCMs among 32 available CMIP6 GCMs based on the GoF measures over the main watershed in South Korea for the historical period of 1973–2014.

It was found that a comprehensive rating metric (RM) approach was more effective to evaluate the overall ranking of the GCMs and then identify the desired GCMs due to the fact that the GCMs' performance differs from the spatial performance assessment metrics. Furthermore, the overall ranking of GCMs based on the KGE, FSS, and SPAEF metrics has relatively similar results, while inconsistent results based on the Cramer's *V* index are identified. In this perspective, the RM-based model selection process combining a variety of metrics could be more informative than a single metric by considering different aspects of model performance for hydrological study over the major river watersheds in South Korea. Finally, this study explored a suitable set of GCMs with RM values and further categorized the GCMs for both MAP and MMP. Among all GCMs, the co-selected GCMs based on RM values for representing the spatial distribution of both MAP and MMP are CESM2-WACCM and CMCC-ESM2.

According to the KGE, FSS, and SPAEF metrics, the FIO-ESM-2-0 model appears to be relevant, and the CESM2-WACCM, CESM2, and CMCC-ESM2 models have a better association with the observed precipitation. Furthermore, the best-performing GCMs showed more plausible spatial patterns than those observed. In contrast, the worst-performing GCMs showed more considerable differences in the spatial patterns observed in the precipitation during the historical period, and a significant underestimation of the precipitation was clearly marked over a large region. In this regard, this study explored the role of

the bias correction approach in selecting GCMs and further demonstrated there will be a change in the ranking pattern under the bias correction process. Here, a QM-based bias correction approach was applied to the simulated MMPs from GCMs. The results showed that the spatial distributions of bias-corrected MMPs for these GCMs were almost identical to those of the observed, irrespective of the GCM performance. Thus, this study confirmed that bias correction is not necessarily required and recommended prior to model selection.

In summary, the selection of an appropriate subset of CMIP6 GCMs may be related to various factors, such as the variables of interest, the number of GCMs, spatial and temporal resolutions, and evaluation metrics. This study used a limited number of evaluation metrics tailored for the assessment of several GCMs, so that the selected GCMs may be changed with the use of other metrics. Therefore, future work should focus on the use of more expanded spatiotemporal evaluation metrics and the sensitivity of these metrics to the variation in model selection, providing incremental information over the existing model selection process.

ACKNOWLEDGEMENTS

This work was supported by Korea Environment Industry & Technology Institute (KEITI) through Water Management Program for Drought, funded by Korea Ministry of Environment (MOE) (2022003610003).

DATA AVAILABILITY STATEMENT

All relevant data are available from an online repository or repositories https://esgf-node.llnl.gov/search/cmip6/.

CONFLICT OF INTEREST

The authors declare there is no conflict.

REFERENCES

Akhter, J., Das, L. & Deb, A. 2017 CMIP5 ensemble-based spatial rainfall projection over homogeneous zones of India. *Climate Dynamics* **49** (5), 1885–1916.

Canadell, J. G., Meyer, C. P., Cook, G. D., Dowdy, A., Briggs, P. R., Knauer, J., Pepler, A. & Haverd, V. 2021 Multi-decadal increase of forest burned area in Australia is linked to climate change. *Nature Communications* **12** (1), 111.

Chen, W., Jiang, Z. & Li, L. 2011 Probabilistic projections of climate change over China under the SRES A1B scenario using 28 AOGCMs. *Journal of Climate* **24** (17), 4741–4756.

Cook, L. M., Anderson, C. J. & Samaras, C. 2017 Framework for incorporating downscaled climate output into existing engineering methods: application to precipitation frequency curves. *Journal of Infrastructure Systems* **23** (4), 4017027.

Cramér, H. 1999 *Mathematical Methods of Statistics (PMS-9)*. Princeton University Press, Princeton, NJ.

Demirel, M. C., Mai, J., Mendiguren, G., Koch, J., Samaniego, L. & Stisen, S. 2018 Combining satellite data and appropriate objective functions for improved spatial pattern performance of a distributed hydrologic model. *Hydrology and Earth System Sciences* **22** (2), 1299–1315.

Eyring, V., Bony, S., Meehl, G. A., Senior, C. A., Stevens, B., Stouffer, R. J. & Taylor, K. E. 2016 Overview of the coupled model intercomparison project phase 6 (CMIP6) experimental design and organization. *Geoscientific Model Development* **9** (5), 1937–1958.

Gu, H., Yu, Z., Wang, J., Wang, G., Yang, T., Ju, Q., Yang, C., Xu, F. & Fan, C. 2015 Assessing CMIP5 general circulation model simulations of precipitation and temperature over China. *International Journal of Climatology* **35** (9), 2431–2440.

Guilyardi, E., Wittenberg, A., Fedorov, A., Collins, M., Wang, C., Capotondi, A., Van Oldenborgh, G. J. & Stockdale, T. 2009 Understanding El Niño in ocean–atmosphere general circulation models: progress and challenges. *Bulletin of the American Meteorological Society* **90** (3), 325–340.

Gupta, H. V., Kling, H., Yilmaz, K. K. & Martinez, G. F. 2009 Decomposition of the mean squared error and NSE performance criteria: implications for improving hydrological modelling. *Journal of Hydrology* **377** (1–2), 80–91.

Gusain, A., Ghosh, S. & Karmakar, S. 2020 Added value of CMIP6 over CMIP5 models in simulating Indian summer monsoon rainfall. *Atmospheric Research* **232**, 104680.

He, W.-p., Zhao, S.-s., Wu, Q., Jiang, Y.-d. & Wan, S. 2019 Simulating evaluation and projection of the climate zones over China by CMIP5 models. *Climate Dynamics* **52** (5), 2597–2612.

Im, E.-S., Thanh, N.-X., Kim, Y.-H. & Ahn, J.-B. 2019 2018 summer extreme temperatures in South Korea and their intensification under 3°C global warming. *Environmental Research Letters* **14** (9), 94020.

Jiang, Z., Li, W., Xu, J. & Li, L. 2015 Extreme precipitation indices over China in CMIP5 models. Part I: Model evaluation. *Journal of Climate* **28**, 8603–8619.

Khan, N., Shahid, S., Ahmed, K., Ismail, T., Nawaz, N. & Son, M. 2018 Performance assessment of general circulation model in simulating daily precipitation and temperature using multiple gridded datasets. *Water* **10**, 1793.

Kim, M.-K., Yu, D.-G., Oh, J.-S., Byun, Y.-H., Boo, K.-O., Chung, I.-U., Park, J.-S., Park, D.-S. R., Min, S.-K. & Sung, H. M. 2020 Performance evaluation of CMIP5 and CMIP6 models on heatwaves in Korea and associated teleconnection patterns. *Journal of Geophysical Research: Atmospheres* **125** (23), e2020JD032583.

Kim, Y.-T., Kwon, H.-H., Lima, C. & Sharma, A. 2021 A novel spatial downscaling approach for climate change assessment in regions with sparse ground data networks. *Geophysical Research Letters* **48** (22), e2021GL095729.

Koch, J., Demirel, M. C. & Stisen, S. 2018 The SPAtial EFficiency metric (SPAEF): multiple-component evaluation of spatial patterns for optimization of hydrological models. *Geoscientific Model Development* **11** (5), 1873–1886.

Kwon, H.-H. & Lall, U. 2016 A copula-based nonstationary frequency analysis for the 2012–2015 drought in California. *Water Resources Research* **52** (7), 5662–5675.

Kwon, H.-H., Khalil, A. F. & Siegfried, T. 2008 Analysis of extreme summer rainfall using climate teleconnections and typhoon characteristics in South Korea 1. *JAWRA Journal of the American Water Resources Association* **44**, 436–448.

Lima, C., Kwon, H.-H. & Kim, Y.-T. 2021 A Bayesian Kriging model applied for spatial downscaling of daily rainfall from GCMs. *Journal of Hydrology* **597**, 126095.

Lindsey, R. & Dahlman, L. 2020 Climate change: global temperature. https://www.climate.gov/news-features/understanding-climate/climate-change-global-temperature (Accessed 16 February 2023).

Lutz, A. F., ter Maat, H. W., Biemans, H., Shrestha, A. B., Wester, P. & Immerzeel, W. W. 2016 Selecting representative climate models for climate change impact studies: an advanced envelope-based selection approach. *International Journal of Climatology* **36**, 3988–4005.

Mandal, S., Arunkumar, R., Breach, P. A. & Simonovic, S. P. 2019 Reservoir operations under changing climate conditions: hydropower-production perspective. *Journal of Water Resources Planning and Management* **145**, 4019016.

Marchi, L., Cavalli, M., Amponsah, W., Borga, M. & Crema, S. 2016 Upper limits of flash flood stream power in Europe. *Geomorphology* **272**, 68–77.

Mason, J. B., Chotard, S., Bailes, A., Mebrahtu, S. & Hailey, P. 2010 Impact of drought and HIV on child nutrition in Eastern and Southern Africa. *Food and Nutrition Bulletin* **31**, S209–S218.

Masson-Delmotte, V., Zhai, P., Pirani, A., Connors, S. L., Péan, C., Berger, S., Caud, N., Chen, Y., Goldfarb, L., Gomis, M. I. & Huang, M. 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*. IPCC, Geneva, Switzerland.

Nunes, L. J. R. & Dias, M. F. 2022 Perception of climate change effects over time and the contribution of different areas of knowledge to its understanding and mitigation. *Climate* **10** (1), 7.

Parth Sarthi, P., Kumar, P. & Ghosh, S. 2016 Possible future rainfall over gangetic plains (GP), India, in multi-Model simulations of CMIP3 and CMIP5. *Theoretical and Applied Climatology* **124** (3), 691–701.

Pour, S. H., Shahid, S., Chung, E.-S. & Wang, X.-J. 2018 Model output statistics downscaling using support vector machine for the projection of spatial and temporal changes in rainfall of Bangladesh. *Atmospheric Research* **213**, 149–162.

Raju, K. S. & Kumar, D. N. 2014 Ranking of global climate models for India using multicriterion analysis. *Climate Research* **60** (2), 103–117.

Roberts, N. M. & Lean, H. W. 2008 Scale-selective verification of rainfall accumulations from high-resolution forecasts of convective events. *Monthly Weather Review* **136** (1), 78–97.

Salman, S. A., Shahid, S., Ismail, T., Ahmed, K. & Wang, X.-J. 2018 Selection of climate models for projection of spatiotemporal changes in temperature of Iraq with uncertainties. *Atmospheric Research* **213**, 509–522.

Srinivasa Raju, K., Sonali, P. & Kumar, D. N. 2017 Ranking of CMIP5-based global climate models for India using compromise programming. *Theoretical and Applied Climatology* **128** (3), 563–574.

Suh, M.-S., Oh, S.-G., Lee, Y.-S., Ahn, J.-B., Cha, D.-H., Lee, D.-K., Hong, S.-Y., Min, S.-K., Park, S.-C. & Kang, H.-S. 2016 Projections of high resolution climate changes for South Korea using multiple-regional climate models based on four RCP scenarios. Part 1: Surface air temperature. *Asia-Pacific Journal of Atmospheric Sciences* **52**, 151–169.

Uranchimeg, S., Kwon, H.-H., Kim, B. & Kim, T.-W. 2020 Changes in extreme rainfall and its implications for design rainfall using a Bayesian quantile regression approach. *Hydrology Research* **51** (4), 699–719.

Wetter, O., Pfister, C., Werner, J. P., Zorita, E., Wagner, S., Seneviratne, S. I., Herget, J., Grünewald, U., Luterbacher, J., Alcoforado, M.-J., Barriendos, M., Bieber, U., Brázdil, R., Burmeister, K. H., Camenisch, C., Contino, A., Dobrovolný, P., Glaser, R., Himmelsbach, I., Kiss, A., Kotyza, O., Labbé, T., Limanówka, D., Litzenburger, L., Nordl, Ø., Pribyl, K., Retsö, D., Riemann, D., Rohr, C., Siegfried, W., Söderberg, J. & Spring, J.-L. 2014 The Year-Long Unprecedented European Heat and Drought of 1540 – a Worst Case. *Climate Change* **125**, 349–363.

Wu, Y., Zhong, P.-a., Xu, B., Zhu, F. & Fu, J. 2018 Evaluation of global climate model on performances of precipitation simulation and prediction in the Huaihe River Basin. *Theoretical and Applied Climatology* **133** (1), 191–204.

Zhao, B. & Zhang, B. 2018 Assessing hourly precipitation forecast skill with the fractions skill score. *Journal of Meteorological Research* **32** (1), 135–145.

First received 14 January 2023; accepted in revised form 4 July 2023. Available online 14 July 2023