Journal of Geoscience and Environment Protection, 2026, 14(8), 154-170
https://www.scirp.org/journal/gep
ISSN Online: 2327-4344
ISSN Print: 2327-4336

Evaluation of Simulated Summer Weather Patterns over the Yangtze River Basin Using Multiple Climate Models

Shengnan Wu[1*], Pei Wu[2]

[1]Chongqing Sub-Bureau of Southwest China Regional Air Traffic Management Bureau, CAAC, Chongqing, China
[2]Hunan Sub-Bureau of South China Regional Air Traffic Management Bureau, CAAC, Changsha, China
Email: *babywugood@qq.om

How to cite this paper: Wu, S. N., & Wu, P. (2026). Evaluation of Simulated Summer Weather Patterns over the Yangtze River Basin Using Multiple Climate Models. *Journal of Geoscience and Environment Protection, 14,* 154-170.
https://doi.org/10.4236/gep.2026.148008

Received: June 26, 2026
Accepted: August 28, 2026
Published: August 31, 2026

Abstract

Based on daily ERA5 reanalysis data from the European Centre for Medium-Range Weather Forecasts covering 2005-2014, this study adopts the Self-Organizing Map (SOM) method to evaluate and rank sea level pressure fields under historical forcing scenarios from 16 climate models of the Coupled Model Intercomparison Project Phase 6 (CMIP6). Optimal models are selected for multi-model ensemble, and the precipitation discrepancies between ensemble simulations and observations are further analyzed. The main conclusions are as follows: 1) Most CMIP6 climate models can reasonably reproduce the observed frequency characteristics of summer weather patterns, yet their simulation performance varies significantly across individual months. The top five models in the comprehensive ranking are MPI-ESM1-2-HR, IPSL-CM6A-MR1, IPSL-CM6A-LR, MPI-ESM1-2-HAM and FGOALS-f3-L. Multi-model ensemble effectively improves the overall simulation capacity for summer weather patterns, but more ensemble members do not guarantee better performance. An ensemble of the top three best-performing models (denoted MME3) can greatly boost model simulation skill. 2) Three ensemble schemes (MME3, MME5 and MME16) improve consistency with observations by reducing biases in simulating the frequency of noisy weather patterns, among which MME3 presents the smallest simulation errors. In terms of precipitation biases, MME3 can basically reproduce the spatial distribution of observed precipitation, with precipitation deviations ranging from 0 to 2 mm per day. Model performance varies widely across different weather patterns; overall, multi-model ensemble greatly optimizes the simulation of heavy rainfall magnitudes.

Keywords

CMIP6, Self-Organizing Map (SOM), Multi-Model Ensemble, Synoptic

Weather, Precipitation Bias

1. Introduction

Global climate system models (GCMs) serve as core tools for climate simulation and future climate projection, capable of reproducing large-scale atmospheric circulations well, yet they are generally limited by coarse spatial resolution and inconsistent simulation skills across various regions. Numerous researchers have carried out extensive model evaluations and climate projections for diverse regions using CMIP5 and CMIP6 multi-model datasets, laying a solid foundation for constructing regional climate assessment frameworks (Li et al., 2016; Dong et al., 2015; Seo et al., 2014; Sillmann et al., 2013; Cassano et al., 2006).

Synoptic-scale circulations dominate local climate conditions, and the intensity, duration and frequency of circulation patterns are key drivers of regional extreme climate events, making it vital to evaluate model performance in simulating daily synoptic weather patterns (Dayan et al., 2012; Paraschivescu et al., 2012; Raziei et al., 2012; Cohen et al., 2013). Conventional statistical methods such as principal component analysis (PCA) suffer from low accuracy and poor visualization when evaluating circulation characteristics (Reusch et al., 2005; Reusch et al., 2007; Liu et al., 2006). The Self-Organizing Map (SOM), originally proposed by Kohonen (1982), objectively clusters high-dimensional daily meteorological fields and extracts dominant circulation modes to overcome the drawbacks of traditional statistical tools. Hewitson and Crane (2002) first introduced SOM to meteorological research, and this technique has since been widely applied to assess model performance and diagnose regional climate features. Previous studies have proven that SOM can quantitatively capture simulation biases in circulation spatial structures and weather pattern frequencies.

The Yangtze River Basin is dominated by the East Asian summer monsoon with complicated circulation systems, leading to prominent uncertainties in climate simulation and projection. Evaluating circulation models can help select reliable models for regional downscaling and climate projection. Most existing synoptic classification and model evaluation studies for East Asia and the Yangtze River Basin are based on CMIP5 datasets. Wang et al. (2015) adopted the SOM method to assess the simulation of winter and summer synoptic patterns over East Asia under CMIP5 and produced model rankings. Compared with CMIP5, CMIP6 features improved physical parameterization schemes, higher spatial resolution and more accurate representation of climatic processes, making it the authoritative dataset for current climate change research. Nevertheless, few studies have systematically evaluated CMIP6's capacity to reproduce daily synoptic patterns over the Yangtze River Basin and decomposed precipitation biases using SOM (Yang et al., 2021; Luo et al., 2021; Song et al., 2024).

Against this background, this study uses CMIP6 historical run outputs from

2005 to 2014, adopts the mature evaluation framework proposed by Radić and Clarke (2011) together with SOM-based objective circulation classification, and takes the Yangtze River Basin as the research domain. We compare synoptic classification results from 16 CMIP6 models with ERA5 reanalysis data, screen optimal models according to their performance in simulating synoptic pattern frequencies, and further dissect precipitation biases and their underlying mechanisms via multi-model ensemble analysis, so as to provide theoretical support for climate simulation and future climate projection over the Yangtze River Basin.

2. Data

Daily ERA5 summer (June-August) daily reanalysis datasets for 2005-2014 from the European Centre for Medium-Range Weather Forecasts (ECMWF), including sea-level pressure (SLP) and surface precipitation, are used in this study. The dataset has a horizontal spatial resolution of 1.0˚ × 1.0˚, and these two variables are taken as observational references. As shown in **Table 1**, sixteen CMIP6 models are selected for evaluation, whose historical simulation outputs include the same variables as ERA5. Different CMIP6 models adopt either Gregorian or no-leap calendars. February 29 of leap years is excluded when extracting June-August daily data to ensure identical summer sample days across all models, which is essential for frequency-based evaluation. Prior to evaluation, all model outputs are interpolated to the same resolution as ERA5 via the bilinear interpolation method. The study domain is the Yangtze River Basin (25˚N - 35˚N, 108˚E - 125˚E).

Table 1. Overview of the 16 CMIP6 models.

Model Name	Institution	Resolution (lon × lat)
BCC-CSM2-HR	Beijing Climate Center, China Meteorological Administration, China	800 × 400
FGOALS-f3-L	Institute of Atmospheric Physics, Chinese Academy of Sciences, China	288 × 180
CanESM5	Canadian Centre for Climate Modelling and Analysis, Canada	128 × 64
CESM2-FV2	National Center for Atmospheric Research, United States	144 × 96
CNRM-CM6-1	Centre National de Recherches Météorologiques, CERFACS, France	256 × 128
HadGEM3-GC31-LL	Met Office Hadley Centre, United Kingdom	192 × 144
HadGEM3-GC31-MM	Met Office Hadley Centre, United Kingdom	288 × 216
INM-CM5-H	Institute for Numerical Mathematics, Russian Academy of Sciences, Russia	180 × 120
IPSL-CM6A-MR1	Institut Pierre-Simon Laplace, France	144 × 143
IPSL-CM6A-LR	Institut Pierre-Simon Laplace, France	144 × 143
MIROC6	Atmosphere and Ocean Research Institute, National Institute for Environmental Studies, Japan Agency for Marine-Earth Science and Technology, Japan	256 × 128
MIROC-ES2H	Atmosphere and Ocean Research Institute, National Institute for Environmental Studies, Japan Agency for Marine-Earth Science and Technology, Japan	256 × 128

Continued

MIROC-ES2H-NB	Atmosphere and Ocean Research Institute, National Institute for Environmental Studies, Japan Agency for Marine-Earth Science and Technology, Japan	256 × 128
MPI-ESM1-2-HAM	Max Planck Institute for Meteorology, Germany	192 × 96
MPI-ESM1-2-HR	Max Planck Institute for Meteorology, Germany	384 × 192
NorESM1-F	Norwegian Climate Centre, Norway	144 × 96

3. Methodology

3.1. SOM Synoptic Classification

Self-organizing map (SOM) is an unsupervised artificial neural network model first proposed by Kohonen (1982). This method repeatedly learns and trains meteorological data based on neural network principles to extract dominant circulation modes of synoptic systems. The detailed SOM classification procedure follows Wang et al. (2015). Following Wang et al. (2015) and Schuenemann and Cassano (2009), sea level pressure (SLP), which is closely related to near-surface meteorological variables, is adopted to represent atmospheric synoptic conditions in this study. The SLP field is processed into a spatial anomaly field by subtracting the regional daily mean value from the grid-point SLP values. A total of 920 daily summer SLP spatial anomaly fields during 2005-2014 are used as training samples and input into the two-dimensional SOM network. After iterative training, all daily samples are objectively classified into distinct synoptic weather types, and the specific training process is shown in **Figure 1**.

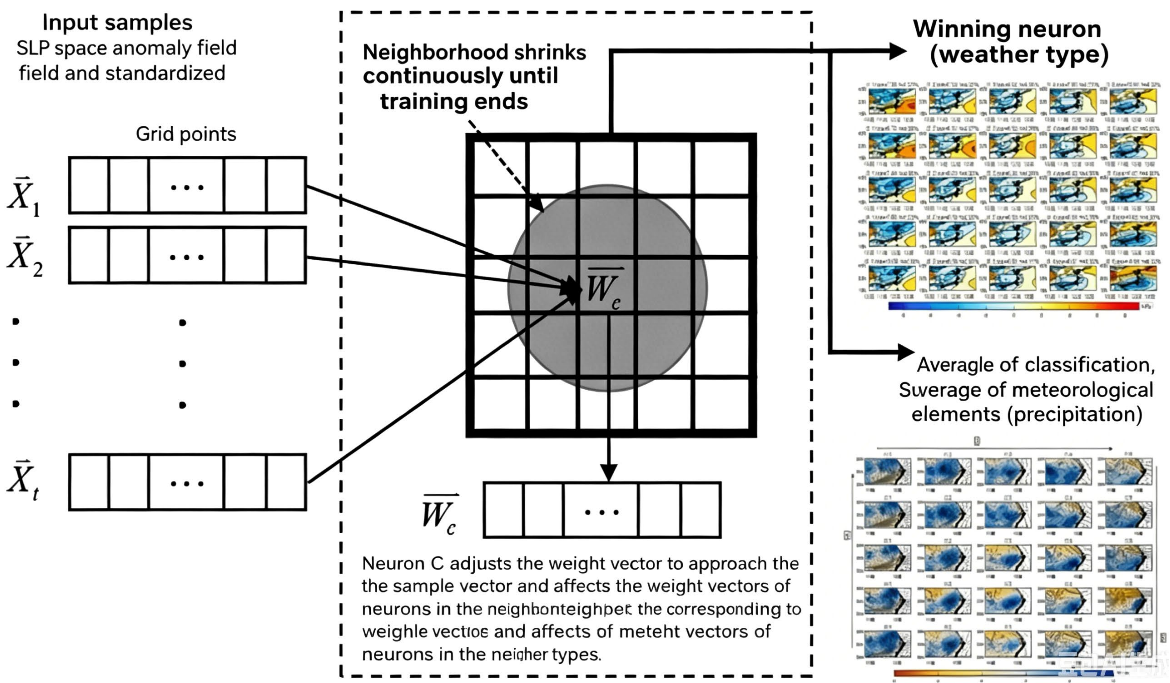

Figure 1. Diagram of the SOM's training process.

3.2. Determination of SOM Grid Size

Optimal SOM grid dimensions depend on the complexity of local synoptic regimes: too few categories fail to capture full synoptic variability and evolution, while excessive categories produce overly similar, non-representative patterns. Two core metrics quantify SOM classification quality: quantization error (QE) and topographic error (TE) (Kohonen, 1982). QE denotes the average Euclidean distance between samples within a cluster and the reference vector of their best-matching neuron, analogous to within-group variance in clustering. TE represents the percentage of samples whose first and second best-matching neurons are non-adjacent, quantifying inter-cluster dissimilarity. Lower QE and TE values indicate superior classification performance. Multiple grid dimension experiments are conducted (Figure 2), with QE and TE ranked ascendingly (smaller values correspond to higher ranks). The 5 × 5 grid yields the minimum summed rank of QE and TE, hence selected as the final SOM grid size.

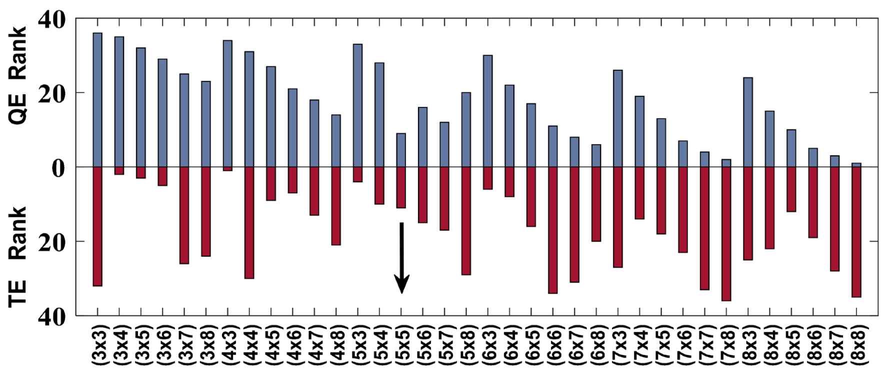

Figure 2. Ranked order of QE and TE values for alternative SOM grid dimensions (sorted from low to high).

3.3. Model Evaluation and Multi-Model Ensemble

SOM training assigns each daily SLP field sample to a single node on the two-dimensional map, enabling calculation of the occurrence frequency of each weather pattern (defined as the fraction of total samples, 920 summer days 2005-2014, belonging to a given pattern). A skillful climate model must accurately reproduce both the spatial morphology and occurrence frequency of observed synoptic patterns. The evaluation and ensemble workflow proceeds as follows:

- Daily ERA5 sea level pressure reanalysis data are used as observational reference. After spatial anomaly calculation, the data are fed into the SOM network to generate 25 benchmark synoptic patterns (5 × 5 layout), daily classification labels and corresponding occurrence frequencies for observed circulations.
- Daily sea level pressure outputs of the 16 CMIP6 models are preprocessed consistently with ERA5: leap-day February 29 is discarded, bilinear interpolation is performed to a uniform 1.0˚ × 1.0˚ resolution, and daily spatial anomalies are calculated by subtracting regional averages. Standardized circulation samples are projected onto the ERA5-trained SOM network to obtain daily synop-

tic classifications and simulated pattern frequencies, followed by frequency significance tests described by Schuenemann and Cassano (2009).

- Correlation coefficients of synoptic pattern frequencies between ERA5 and each model are computed for the whole summer, June, July and August. The average of the four coefficients serves as the comprehensive metric to rank the 16 models; higher values denote better performance in reproducing circulation frequencies.
- Three equally-weighted ensemble schemes are established based on rankings: MME3 (top 3 models), MME5 (top 5 models) and MME16 (all 16 models). Daily circulation and precipitation fields of each ensemble are derived by arithmetic averaging.
- Synoptic frequency biases and spatial precipitation discrepancies between the three ensembles and ERA5 are compared. Combined with pattern-wise precipitation maps and precipitation Q-Q plots, quantitative analysis is carried out on the improvements of regional and heavy rainfall simulation from different ensembles and well-performing individual models.

4. Results and Analysis

4.1. SOM-Derived Synoptic Weather Patterns

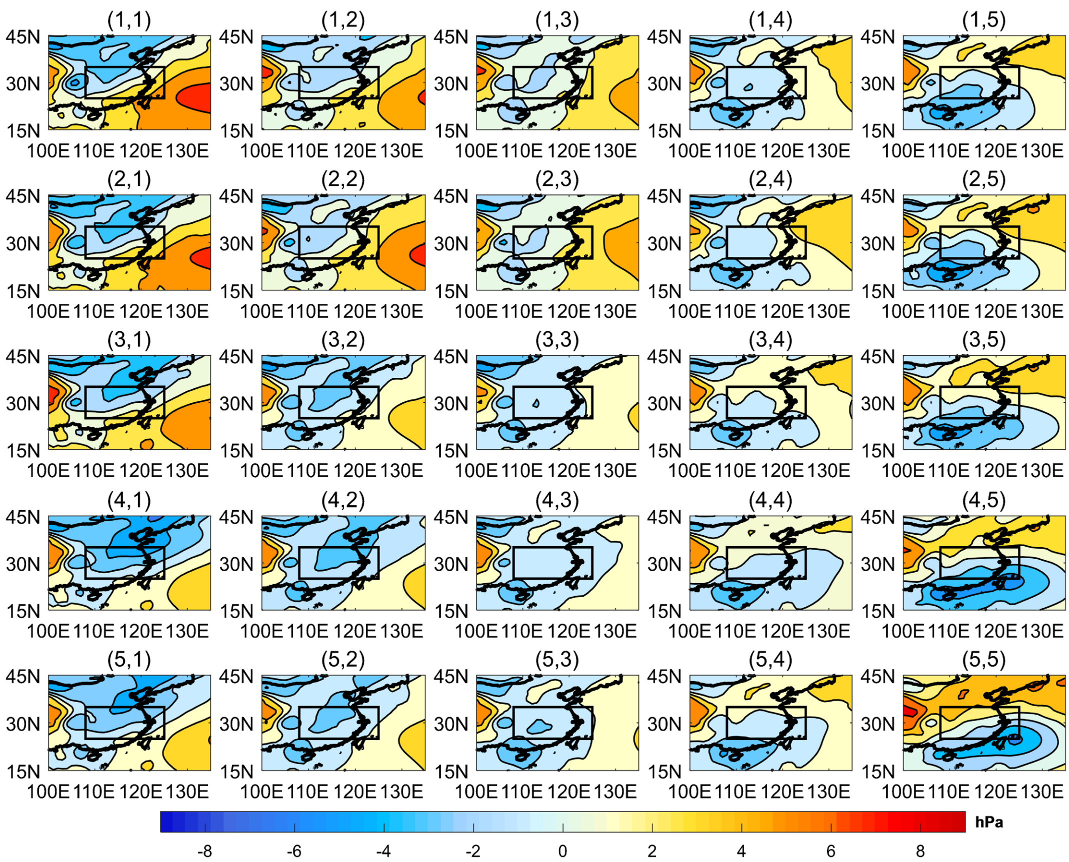

Figure 3. SLP spatial anomaly fields of SOM synoptic patterns derived from 2005-2014 ERA5 data (unit: hPa; rectangular box denotes study domain).

The 5 × 5 SOM grid generates 25 dominant SLP anomaly synoptic patterns over the Yangtze River Basin from 2005-2014 ERA5 data (**Figure 3**). Neighboring SOM nodes feature similar SLP anomaly distributions, while distant nodes exhibit stark contrasts, primarily driven by variations in the intensity and position of subtropical high-pressure systems over the ocean and low-pressure systems over land. Nodes (1,1), (2,1), (3,1), (1,2), (2,2) (top-left): Northwest low pressure, southeast subtropical high pattern. Nodes (5,5), (4,5), (3,5), (5,4), (4,4) (bottom-right): Northwest high pressure, southeast low-pressure pattern. Central nodes represent transitional circulation regimes.

4.2. Model Performance Evaluation

After SOM synoptic classification, the occurrence frequency of the 25 synoptic weather types can be further calculated. Radić and Clarke (2011) demonstrated that calculating the correlation coefficient between CMIP6 model outputs and ERA5 daily reanalysis data allows quantitative assessment of models' ability to reproduce the observed climatological features of synoptic patterns. A two-tailed Student's t-test for Pearson correlation coefficients was used to test the statistical significance of correlations between synoptic pattern frequencies derived from each model and ERA5. The paired sample size corresponds to the 25 synoptic weather types classified via SOM, resulting in a degree of freedom of df = N-2 = 23 and a significance threshold of alpha = 0.05; correlations with $p < 0.05$ are identified as statistically significant. The full t-statistics and p-values for each model are provided in **Table 2**.

Table 2. Two-tailed t-test results for Pearson correlation between synoptic pattern frequencies of individual CMIP6 models and ERA5.

Model	t-statistic	p-value	Significant ($\alpha = 0.05$)	Model	t-statistic	p-value	Significant ($\alpha = 0.05$)
MPI-ESM1-2-HR	7.741	2.51×10^{-7}	Yes	HadGEM3-GC31-LL	5.796	7.81×10^{-6}	Yes
IPSL-CM6A-MR1	6.395	1.87×10^{-6}	Yes	CESM2-FV2	6.144	3.26×10^{-6}	Yes
IPSL-CM6A-LR	6.492	1.53×10^{-6}	Yes	NorESM1-F	6.311	2.18×10^{-6}	Yes
MPI-ESM1-2-HAM	8.023	1.26×10^{-7}	Yes	BCC-CSM2-HR	9.327	6.63×10^{-9}	Yes
FGOALS-f3-L	5.078	4.22×10^{-5}	Yes	MIROC-ES2H-NB	4.605	1.14×10^{-4}	Yes
CanESM5	6.370	1.96×10^{-6}	Yes	MIROC-ES2H	3.604	1.47×10^{-3}	Yes
HadGEM3-GC31-MM	7.534	3.70×10^{-7}	Yes	MIROC6	4.473	1.50×10^{-4}	No
INM-CM5-H	8.621	3.14×10^{-8}	Yes	CNRM-CM6-1	−0.486	0.631	Yes

As illustrated in **Figure 4**, the summer occurrence frequency curves of synoptic patterns from ERA5 are broadly consistent with those from most CMIP6 models except CNRM-CM6-1, corresponding to statistically significant positive correlations. The correlation coefficient between CNRM-CM6-1 and ERA5 for SOM synoptic pattern occurrence frequencies is −0.101. Collectively, these results indicate that all 15 remaining CMIP6 climate models can reasonably capture the observed characteristics of synoptic pattern occurrence frequencies, while CNRM-CM6-1 exhibits poor simulation performance.

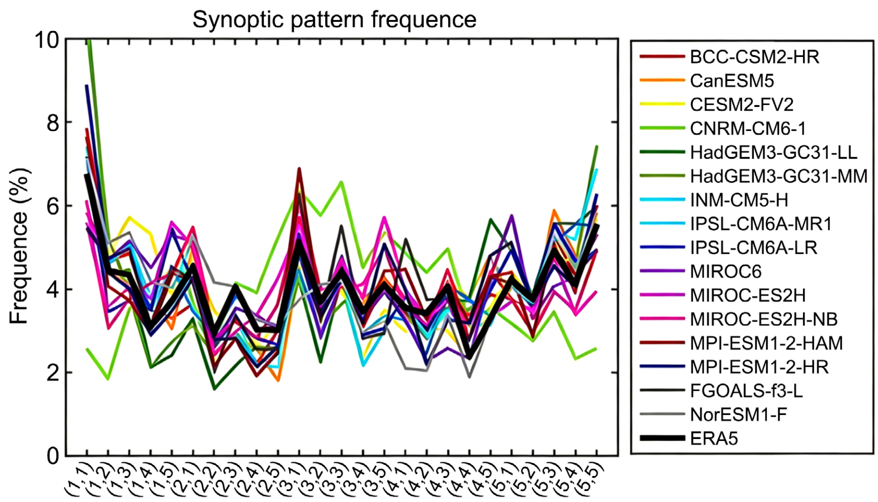

Figure 4. Time series of synoptic pattern occurrence frequencies for ERA5 and the 16 CMIP6 models (summer season).

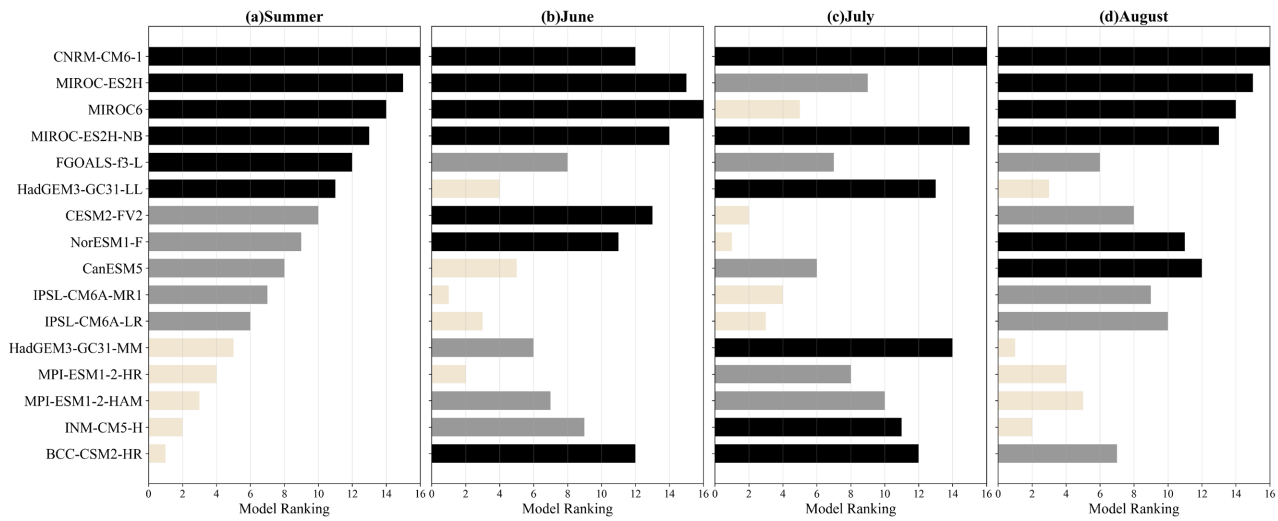

Figure 5. Correlation coefficients (sorted descending) between ERA5 and each CMIP6 model's synoptic pattern occurrence frequencies for (a) full summer, (b) June, (c) July and (d) August.

It is worth noting that the summer weather patterns over the Yangtze River Basin vary from month to month. Accordingly, this study calculates the correlation coefficients for the occurrence frequency of summer-mean weather patterns, and further discusses the correlation coefficients for individual months. Figure 5 ranks the 16 CMIP6 climate models according to their correlation coefficients, to evaluate their performance in simulating observed weather-pattern occurrence frequencies for summer as a whole and for each individual month. As shown in the figure, the rankings of the 16 CMIP6 models differ considerably between summer-mean and monthly conditions. For the full summer season, the top-five performing models are BCC-CSM2-HR, INM-CM5-H, MPI-ESM1-2-HAM, MPI-ESM1-2-HR, and HadGEM3-GC31-MM. For June, the five best-performing models for simulating weather-pattern occurrence frequency are IPSL-CM6A-MR1, MPI-ESM1-2-HR, IPSL-CM6A-LR, HadGEM3-GC31-LL, and CanESM5. For July, the top-five models are NorESM1-F, CESM2-FV2, IPSL-CM6A-LR, IPSL-CM6A-MR1, and MIROC6. For August, the five best-performing models are HadGEM3-GC31-MM, INM-CM5-H, HadGEM3-GC31-LL, MPI-ESM1-2-HR, and MPI-ESM1-2-HAM.

Table 3. Correlation coefficients and performance ranks between each CMIP6 model and ERA5 synoptic pattern occurrence frequencies.

Model	Correlations					Ranks				
	Summer	June	July	August	Average	Summer	June	July	August	Average
MPI-ESM1-2-HR	0.852	0.721	0.816	0.916	0.826	4	2	8	4	1
IPSL-CM6A-MR1	0.801	0.735	0.87	0.854	0.815	7	1	4	9	2
IPSL-CM6A-LR	0.805	0.698	0.873	0.831	0.802	6	3	3	10	3
MPI-ESM1-2-HAM	0.861	0.613	0.767	0.909	0.787	3	7	10	5	4
FGOALS-f3-L	0.735	0.597	0.856	0.902	0.773	12	8	7	6	5
CanESM5	0.8	0.657	0.862	0.755	0.768	8	5	6	12	6
HadGEM3-GC31-MM	0.845	0.626	0.644	0.952	0.767	5	6	14	1	7
INM-CM5-H	0.878	0.465	0.747	0.951	0.76	2	9	11	2	8
HadGEM3-GC31-LL	0.772	0.659	0.661	0.918	0.753	11	4	13	3	9
CESM2-FV2	0.787	0.355	0.893	0.858	0.723	10	13	2	8	10
NorESM1-F	0.796	0.387	0.928	0.78	0.723	9	11	1	11	11
BCC-CSM2-HR	0.894	0.406	0.698	0.891	0.722	1	10	12	7	12
MIROC-ES2H-NB	0.704	0.342	0.635	0.66	0.585	13	14	15	13	13
MIROC-ES2H	0.613	0.223	0.782	0.365	0.496	15	15	9	15	14
MIROC6	0.695	−0.047	0.865	0.44	0.488	14	16	5	14	15
CNRM-CM6-1	−0.101	0.379	0.114	0.297	0.172	16	12	16	16	16

Table 3 lists the correlation coefficients and corresponding ranks of SOM-derived weather-pattern occurrence frequencies between the 16 CMIP6 models and ERA5 reanalysis dataset. The BCC-CSM2-HR model exhibits the best performance in simulating summer-mean weather-pattern occurrence frequency, yet its skills degrade remarkably for individual months, leading to a low overall rank. Although INM-CM5-H reproduces August weather-pattern features reasonably well, it shows poor performance in June and July, which also results in a low comprehensive rank. HadGEM3-GC31-MM achieves the highest simulation skill for August weather-pattern frequency, but its performance deteriorates in June-July, yielding a low overall rank as well.

These results reveal that an individual climate model can show substantial discrepancies in simulating monthly weather-pattern occurrence frequencies across summer months. Such discrepancies are likely overlooked in previous model evaluations based on monthly-mean datasets, which highlights the necessity of model assessment using daily-scale outputs. Accordingly, the average of four correlation coefficients (summer-mean, June, July and August) is adopted to produce a comprehensive rank for selecting optimal models for multi-model ensemble. The top-five ranked models are MPI-ESM1-2-HR, IPSL-CM6A-MR1, IPSL-CM6A-LR, MPI-ESM1-2-HAM and FGOALS-f3-L. Multi-model ensembles are constructed by averaging the outputs of the top-3, top-5 and all 16 models, denoted as MME3, MME5 and MME16, respectively. Table 4 presents the correlation coefficients of weather-pattern occurrence frequencies between these ensemble products and ERA5. As shown in Table 4, MME3 and MME5 yield higher correlation coefficients than individual models and produce satisfactory ensemble performance, whereas MME16 shows low correlation coefficients and poor ensemble skills. This indicates that multi-model ensemble can improve the simulation of summer weather-pattern characteristics. Nevertheless, owing to large inter-monthly variations in individual model performance, the ensemble skill does not monotonically increase with the number of participating models. Selecting appropriate optimal models for ensemble is beneficial to enhance simulation capacity.

Table 4. Correlation coefficients of weather-pattern occurrence frequencies between multi-model ensembles (MME3, MME5, MME16) and ERA5 reanalysis dataset.

Model Ensembles	Correlations				
	Summer	June	July	August	Average
MME3	0.861	0.691	0.94	0.943	0.859
MME5	0.861	0.677	0.902	0.942	0.846
MME16	0.836	0.369	0.815	0.94	0.740

4.3. Biases in Simulated Synoptic Pattern Frequency and Precipitation

Figure 6 presents the occurrence frequencies of SOM synoptic weather types derived from ERA5 daily reanalysis data for summers 2005-2014, as well as the frequency biases between three multi-model ensembles (MME3, MME5, MME16) and ERA5. A one-sample two-tailed Student's t-test is conducted on the frequency bias of each synoptic weather type. The test samples consist of all 920 daily summer records from 2005 to 2014, with a degree of freedom of df = N-1 = 919. Only nodes with $p < 0.05$ are marked as statistically significant overestimated or underestimated frequencies in Figure 6.

As shown in Figure 6, the frequency biases between simulations from the three ensemble schemes and ERA5 observations are all constrained within ±2%, with consistent spatial bias patterns. The ensembles overestimate the occurrence of synoptic types located at the top-left and bottom-right of the SOM grid, while underestimating the central transitional weather types. Radić and Clarke (2011) (2011) noted that the central SOM nodes generally correspond to noisy synoptic regimes. Multi-model ensembles mitigate the oversimulation of these noisy weather types, thereby improving overall model performance in reproducing synoptic circulation characteristics. Nevertheless, the number of weather types with statistically significant frequency biases varies across ensemble configurations: MME16 exhibits significant deviations for 22 synoptic patterns, MME5 reduces this number to 20, and MME3 has the fewest (19). This result demonstrates that MME3 produces the smallest simulation errors for observed synoptic pattern frequencies, which is consistent with the correlation coefficient analysis presented in Section 4.2. Accordingly, subsequent precipitation bias analyses focus primarily on comparisons between MME3 outputs and ERA5 observations.

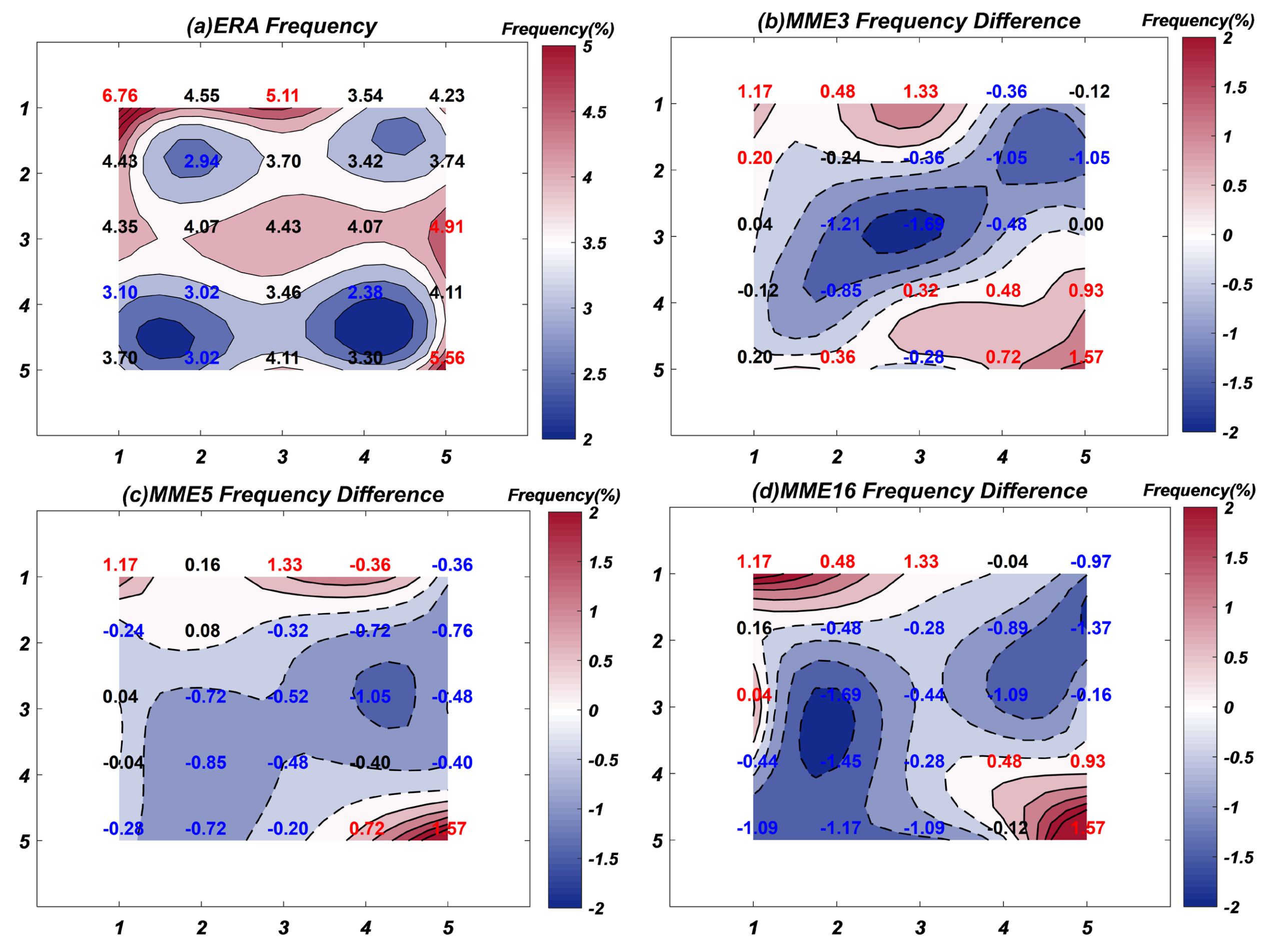

Figure 6. (a) ERA5 summer synoptic pattern occurrence frequencies (2005-2014); frequency differences between ERA5 and (b) MME3, (c) MME5, (d) MME16 (unit: %; red = significant overestimation, blue = significant underestimation).

Figure 7 compares observed precipitation against outputs from the three top individual models and MME3, alongside MME3 precipitation biases relative to ERA5. All three single models broadly reproduce observed precipitation spatial patterns, with MPI-ESM1-2-HR best capturing the two major precipitation maxima over the basin, followed by IPSL-CM6A-MR1 and IPSL-CM6A-LR. Averaging the three models into MME3 further aligns simulated precipitation with observations. MME3 precipitation is slightly lower than ERA5 across the domain, with absolute biases confined to 0 - 2 mm/day, demonstrating that ensemble outputs reasonably replicate observed precipitation spatial structure.Precipitation biases vary drastically across distinct synoptic patterns, illustrated via three representative SOM nodes (1,1), (3,3) and (5,5) in Figure 8:

- Pattern (1,1): Northern precipitation maxima, southern minima. All three single models overpredict northern rainfall and underpredict southern rainfall; ensemble averaging retains this spatial bias yet reduces its magnitude.
- Pattern (3,3): Disparate spatial bias structures among individual models produce mutual offsetting, drastically reducing overall precipitation bias in MME3 and delivering marked simulation improvements.
- Pattern (5,5): Southern precipitation maxima, northern minima. MME3 overestimates northern rainfall and underestimates southern rainfall, yielding minimal overall improvement for this regime.

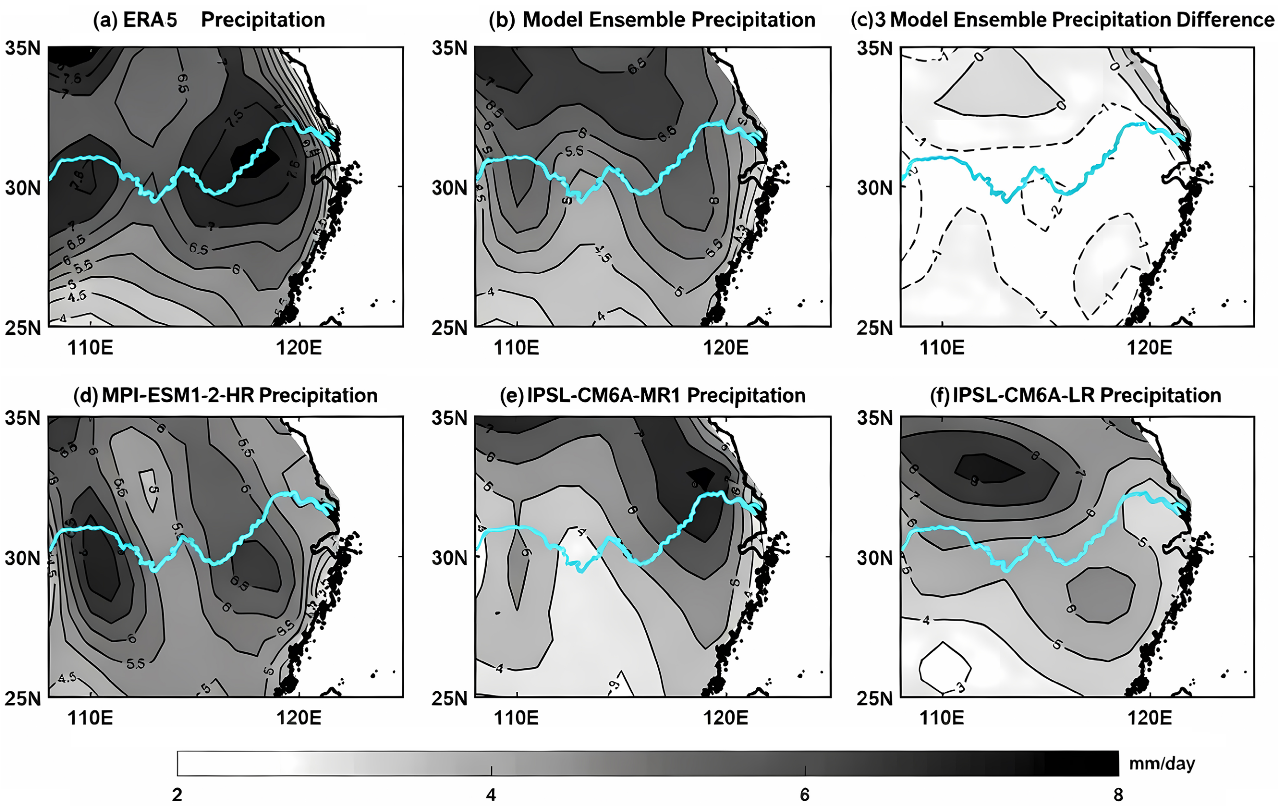

Figure 7. (a) ERA5 observed summer precipitation; (b) MME3 ensemble precipitation; (c) precipitation bias (MME3 minus ERA5); (d)-(f) precipitation from the three individual top models MPI-ESM1-2-HR, IPSL-CM6A-MR1 and IPSL-CM6A-LR (unit: mm/day).

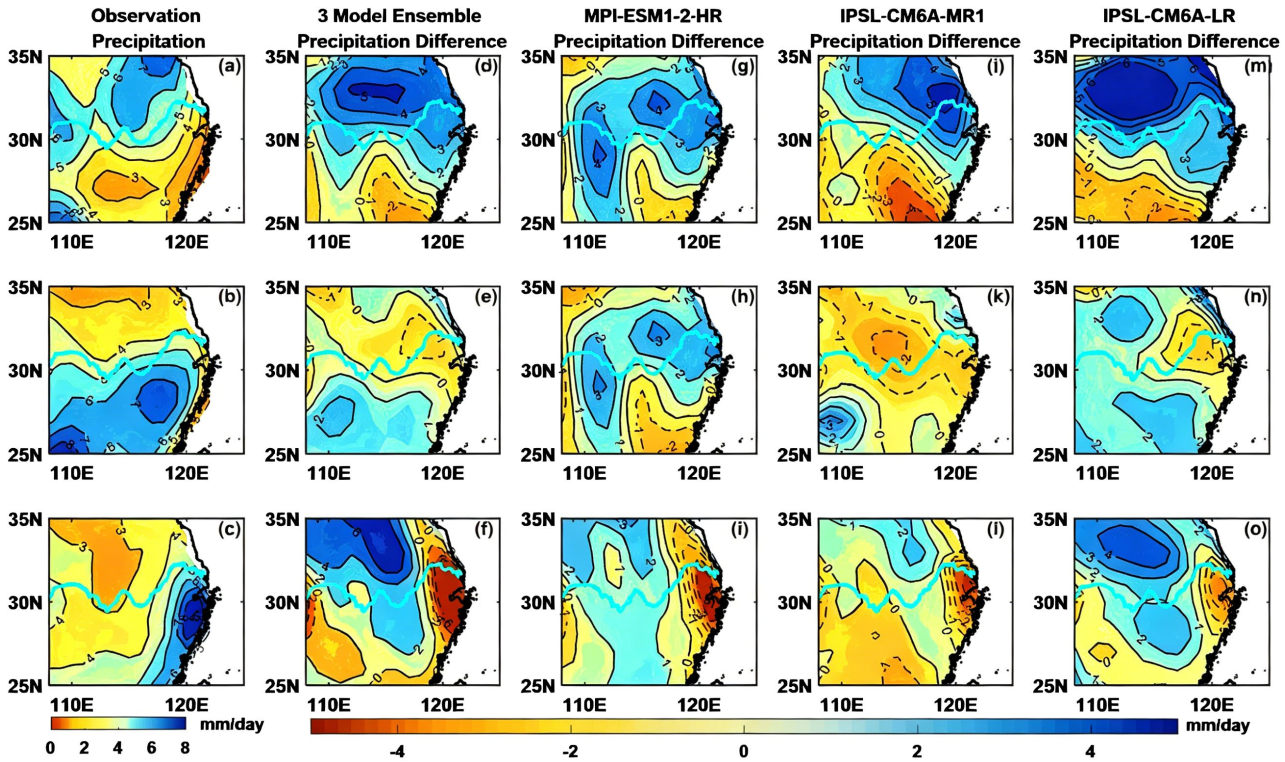

Figure 8. Mean daily ERA5 precipitation for representative synoptic patterns (a)-(c) (1,1), (3,3), (5,5); precipitation biases relative to ERA5 for (d)-(f) MME3, (g)-(i) MPI-ESM1-2-HR, (j)-(l) IPSL-CM6A-MR1, (m)-(o) IPSL-CM6A-LR (unit: mm/day).

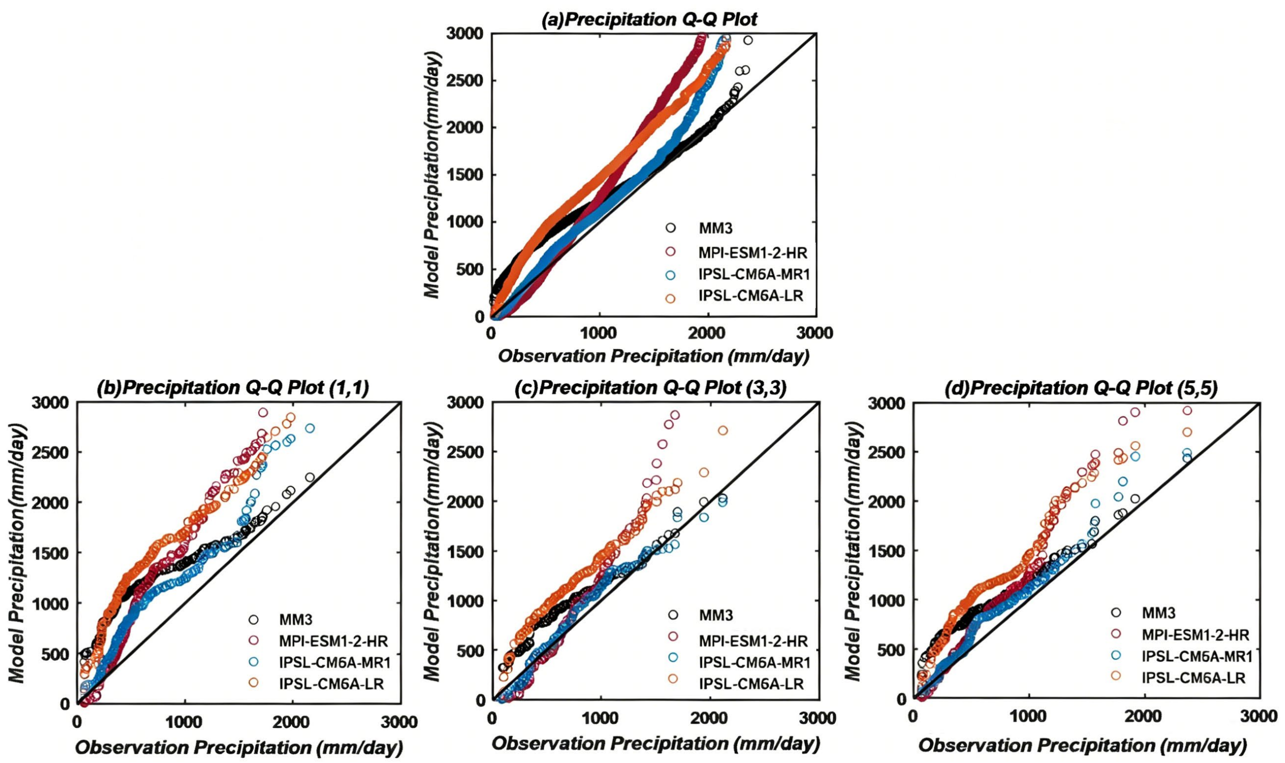

Figure 9. Q-Q plots of total regional precipitation (model vs ERA5) for (a) all synoptic pattern samples, (b) Pattern (1,1), (c) Pattern (3,3), (d) Pattern (5,5).

Quantile-quantile (Q-Q) plots of total regional precipitation for all samples and the three representative patterns (**Figure 9**) quantify ensemble performance across rainfall magnitudes. Points falling close to the 1:1 diagonal indicate accurate simulation:

- All synoptic samples (**Figure 9(a)**): MPI-ESM1-2-HR and IPSL-CM6A-MR1 perform well for light-to-moderate rainfall but diverge from observations above 2000 mm total precipitation; IPSL-CM6A-LR systematically overestimates rainfall across all magnitudes. MME3 shows minor biases for light rain yet substantially improves simulation of moderate to extreme heavy rainfall.
- Pattern (1,1) (**Figure 9(b)**): All single models overpredict rainfall at all magnitudes; MME3 progressively converges to observed heavy rainfall values with increasing precipitation intensity.
- Pattern (3,3) (**Figure 9(c)**): MPI-ESM1-2-HR and IPSL-CM6A-MR1 outperform MME3 due to large positive biases in IPSL-CM6A-LR.
- Pattern (5,5) (**Figure 9(d)**): Similar to Pattern (1,1), MME3 delivers clear improvements for heavy rainfall magnitudes.

Collectively, model performance varies across synoptic regimes and rainfall intensities; MME3 provides the most substantial improvements in simulating heavy rainfall volumes.

5. Conclusions

Taking daily ERA5 reanalysis data from the European Centre for Medium-Range Weather Forecasts as observational benchmark, this study employs the Self-Organizing Map (SOM) method to evaluate and rank 16 climate models from CMIP6. High-quality models are selected to construct multi-model ensembles, and the discrepancies between ensemble outputs and observations in terms of precipitation and occurrence frequency of each weather pattern are further analyzed. The main conclusions are listed as follows:

1) Most CMIP6 climate models can reasonably reproduce the observed occurrence frequency characteristics of summer weather patterns, while their skill in simulating weather pattern features varies markedly across individual months. A comprehensive ranking is calculated based on the average of four correlation coefficients corresponding to the full summer and June, July, August respectively. The top five models are MPI-ESM1-2-HR, IPSL-CM6A-MR1, IPSL-CM6A-LR, MPI-ESM1-2-HAM and FGOALS-f3-L. Multi-model ensemble can enhance the overall simulation performance of summer weather patterns. Nevertheless, given the large monthly disparities in each model's simulation skill, the performance of ensembles does not keep improving with the increase of model quantity. An ensemble named MME3, which consists of the three optimal models (MPI-ESM1-2-HR, IPSL-CM6A-MR1 and IPSL-CM6A-LR), can effectively promote model simulation capacity.

2) In terms of frequency biases between simulated and observed weather patterns, three multi-model ensembles (MME3, MME5 and MME16) achieve better

simulation consistency with observations by reducing biases in reproducing the frequency of noisy weather patterns, among which MME3 has the smallest errors in simulating the actual occurrence frequency of weather patterns. For precipitation simulation, MME3 can basically reproduce the spatial distribution of observed precipitation, with precipitation biases ranging from 0 to 2 mm per day. Model performance varies greatly in simulating precipitation distributions under different weather patterns, and multi-model ensemble greatly improves the simulation of heavy rainfall volume.

This paper classifies daily synoptic-scale circulations via the SOM method, assesses the performance of CMIP6 models in reproducing summer weather patterns over the Yangtze River Basin, and generates model rankings. The results provide references for selecting suitable models for regional downscaling and future climate projection in this basin. Although most models perform well in simulating the observed frequency of summer weather patterns, a single model shows inconsistent capability in reproducing precipitation distributions under different weather patterns. Such inconsistency may stem from biased representation of physical processes under corresponding weather regimes and requires further research. The attribution analysis of precipitation biases focuses on present-day climate conditions, which lays a foundation for our subsequent research on future climate scenarios.

Acknowledgements

We acknowledge the World Climate Research Programme Working Group on Coupled Modelling and all modeling institutions for providing CMIP6 model outputs, and the European Centre for Medium-Range Weather Forecasts (ECMWF) for supplying ERA5 atmospheric reanalysis datasets.

Author Contributions

Shengnan Wu: Conceptualization, Methodology, Formal analysis, Visualization, Writing-original draft, Writing-review & editing. Pei Wu: Data curation, Validation, Writing-review & editing.

Conflicts of Interest

The authors declare no conflicts of interest regarding the publication of this paper.

References

Cassano, J. J., Uotila, P., & Lynch, A. (2006). Changes in Synoptic Weather Patterns in the Polar Regions in the Twentieth and Twenty-First Centuries, Part 1: Arctic. *International Journal of Climatology, 26,* 1027-1049. https://doi.org/10.1002/joc.1306

Cohen, L., Dean, S., & Renwick, J. (2013). Synoptic Weather Types for the Ross Sea Region, Antarctica. *Journal of Climate, 26,* 636-649. https://doi.org/10.1175/jcli-d-11-00690.1

Dayan, U., Tubi, A., & Levy, I. (2012). On the Importance of Synoptic Classification Methods with Respect to Environmental Phenomena. *International Journal of Climatology, 32,* 681-694. https://doi.org/10.1002/joc.2297

Dong, S. Y., Xu, Y., Zhou, B. T., & Shi, Y. (2015). Assessment of Indices of Temperature Extremes Simulated by Multiple CMIP5 Models over China. *Advances in Atmospheric Sciences, 32,* 1077-1091. https://doi.org/10.1007/s00376-015-4152-5

Hewitson, B. C., & Crane, R. G. (2002). Self-Organizing Maps: Applications to Synoptic Climatology. *Climate Research, 22,* 13-26. https://doi.org/10.3354/cr022013

Kohonen, T. (1982). Self-Organized Formation of Topologically Correct Feature Maps. *Biological Cybernetics, 43,* 59-69. https://doi.org/10.1007/bf00337288

Li, W., Jiang, Z., Xu, J., & Li, L. (2016). Extreme Precipitation Indices over China in CMIP5 Models. Part II: Probabilistic Projection. *Journal of Climate, 29,* 8989-9004. https://doi.org/10.1175/jcli-d-16-0377.1

Liu, Y. G., Weisberg, R. H., & He, R. Y. (2006). Sea Surface Temperature Patterns on the West Florida Shelf Using Growing Hierarchical Self-Organizing Maps. *Journal of Atmospheric and Oceanic Technology, 23,* 325-338. https://doi.org/10.1175/jtech1848.1

Luo, N., Guo, Y., Chou, J., & Gao, Z. (2021). Added Value of CMIP6 Models over CMIP5 Models in Simulating the Climatological Precipitation Extremes in China. *International Journal of Climatology, 42,* 1148-1164. https://doi.org/10.1002/joc.7294

Paraschivescu, M., Rambu, N., & Stefan, S. (2012). Atmospheric Circulations Associated to the Interannual Variability of Cumulonimbus Cloud Frequency in the Southern Part of Romania. *International Journal of Climatology, 32,* 920-928. https://doi.org/10.1002/joc.2316

Radić, V., & Clarke, G. K. C. (2011). Evaluation of IPCC Models' Performance in Simulating Late-Twentieth-Century Climatologies and Weather Patterns over North America. *Journal of Climate, 24,* 5257-5274. https://doi.org/10.1175/jcli-d-11-00011.1

Raziei, T., Mofidi, A., Santos, J. A., & Bordi, I. (2012). Spatial Patterns and Regimes of Daily Precipitation in Iran in Relation to Large-Scale Atmospheric Circulation. *International Journal of Climatology, 32,* 1226-1237. https://doi.org/10.1002/joc.2347

Reusch, D. B., Alley, R. B., & Hewitson, B. C. (2005). Relative Performance of Self-Organizing Maps and Principal Component Analysis in Pattern Extraction from Synthetic Climatological Data. *Polar Geography, 29,* 188-212. https://doi.org/10.1080/789610199

Reusch, D. B., Alley, R. B., & Hewitson, B. C. (2007). North Atlantic Climate Variability from a Self-Organizing Map Perspective. *Journal of Geophysical Research: Atmospheres, 112,* D02104. https://doi.org/10.1029/2006jd007460

Schuenemann, K. C., & Cassano, J. J. (2009). Changes in Synoptic Weather Patterns and Greenland Precipitation in the 20th and 21st Centuries: 1. Evaluation of Late 20th Century Simulations from IPCC Models. *Journal of Geophysical Research: Atmospheres, 114,* D20113. https://doi.org/10.1029/2009jd011705

Seo, Y. W., Kim, H., Yun, K. S., Lee, J., Ha, K., & Moon, J. (2014). Future Change of Extreme Temperature Climate Indices over East Asia with Uncertainties Estimation in the CMIP5. *Asia-Pacific Journal of Atmospheric Sciences, 50,* 609-624. https://doi.org/10.1007/s13143-014-0050-5

Sillmann, J., Kharin, V. V., Zwiers, F. W., Zhang, X., & Bronaugh, D. (2013). Climate Extremes Indices in the CMIP5 Multimodel Ensemble: Part 2. Future Climate Projections. *Journal of Geophysical Research: Atmospheres, 118,* 2473-2493. https://doi.org/10.1002/jgrd.50188

Song, X., Xu, M., Kang, S., Wang, R., & Wu, H. (2024). Evaluation and Projection of Changes in Temperature and Precipitation over Northwest China Based on CMIP6 Models. *International Journal of Climatology, 44,* 5039-5056. https://doi.org/10.1002/joc.8622

Wang, Y., Jiang, Z., & Chen, W. (2015). Performance of CMIP5 Models in the Simulation of Climate Characteristics of Synoptic Patterns over East Asia. *Journal of Meteorological Research, 29,* 594-607. https://doi.org/10.1007/s13351-015-4129-4

Yang, X., Zhou, B., Xu, Y., & Han, Z. (2021). CMIP6 Evaluation and Projection of Temperature and Precipitation over China. *Advances in Atmospheric Sciences, 38,* 817-830. https://doi.org/10.1007/s00376-021-0351-4