

Deterministic and probabilistic projections and their credibility in analyzing future precipitation variations in the Yellow River Basin, China

Zhouliang Sun^{a,b}, Yanli Liu^{b,c,d}, Jianyun Zhang^{a,b,c,d,*}, Hua Chen^a, Zhangkang Shu^b, Tiesheng Guan^{b,c,d}, Guoqing Wang^{b,c,d}, Junliang Jin^{b,c,d}, Zhenxin Bao^{b,c,d} and Cuishan Liu^{b,c,d}

^a State Key Laboratory of Water Resources and Hydropower Engineering Science, Wuhan University, Wuhan 430072, China

^b State Key Laboratory of Hydrology-Water Resources and Hydraulic Engineering, Nanjing Hydraulic Research Institute, Nanjing 210029, China

^c Yangtze Institute for Conservation and Development, Nanjing 210098, China

^d Research Center for Climate Change, Ministry of Water Resources, Nanjing 210029, China

*Corresponding author. E-mail: jyzhang@nhri.cn

 ZS, 0000-0002-0388-9621

ABSTRACT

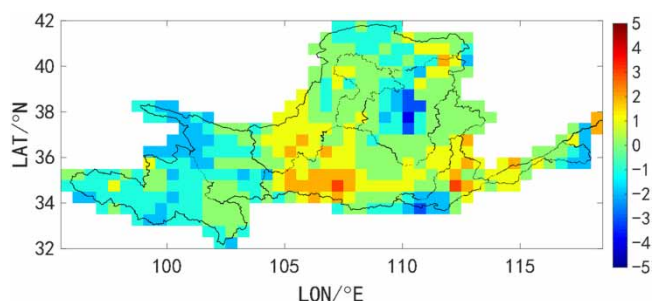
It remains a key challenge to obtain reliable future precipitation estimates and their reliability under different climate scenarios. In this study, the deterministic projection of future precipitation in the Yellow River Basin (YRB) was obtained within the Bayesian model averaging (BMA) framework. A probability estimation method based on the BMA weighting scheme was proposed to obtain the probabilistic projection of precipitation. We also analyzed the credibility of these two projections. The results showed that four indexes projected by the BMA method showed an increasing trend with a higher probability. The probabilities of increasing with varying degrees were more than those for decreasing for all the precipitation indexes. The credibility of the precipitation estimation under specific climate scenarios was testified by the lower E_D (the mean of long-term annual relative simulation deviation) and V_D (the variance of long-term annual relative simulation deviation). The estimation based on the BMA model is more trustworthy than any other model. For the four precipitation indicators, the accuracy between the calculated VR (Variation range, to describe the interval of variation of the indicators) with the greatest likelihood and the actual VR was 38.31–53.74%. In 81.93–94.70% of grids, the deviations were smaller than one level. Both the deterministic and probabilistic projections have high geographic distribution and variation trend consistency.

Key words: Bayesian model averaging, climate change, credibility assessment, precipitation indexes, probability estimation, Yellow River Basin

HIGHLIGHTS

- A probability estimation method was proposed for the probabilistic projection of future climate.
- The deterministic projection and probabilistic projection of future precipitation were consistent.
- The projection accuracy was positively correlated with historical occurrences of precipitation.
- The probability of increase is greater than that of decrease for the future precipitation of YRB.

GRAPHICAL ABSTRACT



This is an Open Access article distributed under the terms of the Creative Commons Attribution Licence (CC BY 4.0), which permits copying, adaptation and redistribution, provided the original work is properly cited (<http://creativecommons.org/licenses/by/4.0/>).

1. INTRODUCTION

The precipitation over the Yellow River Basin (YRB) has undergone continuous and drastic variation in the past decades (Liu *et al.* 2020), especially in terms of extremes and potential redistribution of water resources in time and space, threatening the lives of residents in the whole basin. Another concern lies in the future precipitation variation in the forthcoming decades and centuries under global environmental change. Studies have confirmed the incredible impact of climate change on people and the environment (Meshram *et al.* 2018, 2020; Almazroui *et al.* 2021a, 2021b), but it is still a significant challenge to reliably estimate future climate and its credibility under different climate scenarios and climate models.

The vulnerable natural system in the YRB is of interest to researchers worldwide because it reflects the evolution of water resources under climate change and human activity in the future (Wu *et al.* 2016; Wang *et al.* 2019b). Moreover, the precipitation in the YRB has changed in time and space for decades (Lu *et al.* 2014; Zhang *et al.* 2014, 2019; Iqbal *et al.* 2018; Zhao *et al.* 2019), arousing many concerns and prompting research on future precipitation change as well as its impact on hydrology and water resources. The GCMs provide scenario determination and forcing data, forming the basis of climate change projection. They are an original estimation of long-term precipitation without specific precipitation indexes, but most of the projected precipitation indexes derive from the GCMs data. Typical indexes include the mean over time and extreme value. The key to obtaining a reliable future precipitation projection is to solve uncertainty in multiple models and get valid information. Multi-model ensemble, including equal-weighted or unequal weighted averaging, is the most common and effective solution (Radermacher & Tomassini 2012; Chen *et al.* 2017; Guo *et al.* 2017). For the YRB previous studies mainly focused on the impact of future precipitation change, while there was little research on future precipitation change itself. Of the research that has been done in this area, Wang *et al.* (2017a) and Chen *et al.* (2019) investigated the impacts of climate change on water resources (indicated by runoff) by using the projected precipitation and temperature, applying the Variable Infiltration Capacity (VIC) model, in the whole YRB and the source region of the YRB. The impacting mechanisms usually come from comparing runoff in the baseline period (the model output) and runoff change caused by the precipitation change (the model input). As for the impacts of future precipitation change on hydrology and water resources, Yin *et al.* (2016) analyzed the future extreme precipitation with different return periods in the Huang-Huai-Hai region using five GCMs multi-scenarios. The result indicated that the climate models had errors in simulating extreme historical precipitation. Liu *et al.* (2017) studied the projected climatic year types (CYTs) throughout China between 2015 and 2099 from spatial patterns, temporal changes, and frequency, demonstrating better performance and superiority of multi-model ensembles than individual models in projecting changes in CYTs by offsetting systematic errors. Nevertheless, all future water resources projections depend on the future climate scenarios, and the reliability of conclusions is subject to the credibility of the GCMs. Thus, the main challenge is evaluating and generating more reliable precipitation from GCMs.

Multi-model ensembles are a practical solution for hydrological, meteorological, and climatological predictions or forecasts. The ensemble projection could be understood as a weighted model average from the member models. Raftery *et al.* (2005) applied Bayesian model averaging (BMA) to forecast the surface temperature and sea level pressure in the Pacific Northwest for better and sharper forecasting, confirming the excellent performance of BMA in the multi-model ensemble when the ensemble members are overdispersed. Duan *et al.* (2007) also used BMA to acquire more accurate hydrological predictions from some original competing predictions in terms of more satisfying daily root mean square error and daily absolute mean error, as well as a better description of total predictive uncertainty than the members. These results make it a typical case of ensemble hydrological forecasting using BMA. From research on climate change, the ensemble members mainly come from climate models, including Regional Climate Models (RCMs) and GCMs. Dars *et al.* (2017) proved that BMA performed better than individual models in multi-model ensemble prediction with a weighted average of four CMIP5 models to assess the impacts of climate change on future precipitation. The result will significantly help related managers, experts, and policymakers. A similar study on temperature (Min & Hense 2006; Miao *et al.* 2014) inferred that BMA enjoys more reliability and less uncertainty than other model averaging methods or individual models. The application of BMA in climate change has yielded encouraging results and has been widely recognized (Duan & Phillips 2010; Notaro *et al.* 2016). Irrespective of whether the ensemble members come from hydrological models or climate models, the consistent process in the ensemble is to capture the practical information from those member models for a more reliable future projection.

Researchers have focused on the uncertainty and credibility of climate change projections due to enormous uncertainty in climate models and downscaling data processing. The output of climate scenarios is also uncertain for the complicated

processes in the atmospheric system. It has been challenging to acquire desired results and assess the credibility of climate projection evaluations, especially without a practical evaluation system. The study (Fricker *et al.* 2013) pointed out validation of performance measures for the credibility of future climate projections. These issues about credibility aroused our keen interest in the credibility evaluation of future precipitation projection over water-scarce areas in the YRB. We aim to improve the projection level of climate elements by using appropriate indicators to measure the precipitation projection ability of GCM models.

In this paper, we proposed a probability precipitation estimation method based on the BMA method and the measurement of its credibility. For the future precipitation projection of the YRB, there would be two significant questions to resolve: (1) How to ensure good performance for BMA in terms of the precipitation indexes evaluation compared to the original climate models over the YRB and (2) How to ensure the credibility of the precipitation estimation based on BMA. To find the answer, we used the daily observed precipitation data during 1986–2005 and future precipitation projections selected from five GCMs during 2006–2099. Four precipitation indexes were adopted to describe the precipitation situation. Two indicators were used to assess the performance and credibility of precipitation ensemble projection conducted by the BMA method.

2. STUDY AREA AND DATA

2.1. The Yellow River Basin

With a length of 5,464 km and a drainage area of 795,044 km², the YRB (Figure 1) is the second-largest river in China. It originates in the Tibetan Plateau and runs through four geomorphic units – the Qinghai-Tibetan Plateau, the Inner Mongolia Plateau, the Loess Plateau, and the North China Plain. Supplying water resources to 100 million people, the YRB runs through an arid region, a semi-arid region, and a semi-humid region, before reaching the Bohai Sea (Wu *et al.* 2017; Wang *et al.* 2019a). These regions range in elevation between 1,000 and 4,000 m, have annual precipitation ranging from 123 to 1,021 mm, have annual evaporation varying from 700 to 1,800 mm, and have dramatic temperature fluctuations in time and space (Wang *et al.* 2019a). The long-time average annual value of precipitation, runoff (engaging a drastic decrease), and annual evapotranspiration were estimated to be about 460, 450, and 1,100 mm on average, respectively. Since the YRB is located in a temperate continental climate zone and is regulated by the East Asian monsoon, precipitation is the primary controlling source and restrictive factor of the runoff and water resources. This directly affects the water supply for 12% of China's population, 15% of China's arable land, and 8% of China's GDP approximately, even though the scarce runoff accounts for only 2% of China's total (Guan *et al.* 2016; Yin *et al.* 2017; Wang *et al.* 2019a). Therefore, the variation of precipitation plays a vital role in sustainable development and even in the daily life of the people of the YRB. Both in terms of time and space, the precipitation in the YRB is remarkably uneven, with June to October accounting for about 70% of annual total precipitation and the precipitation in the lower and middle reaches two times that of the upper reaches (Mu *et al.* 2013). Over the past 50 or 60 years, with the increase of air temperature, the precipitation of the YRB has fallen, and so too has runoff, exacerbating a profound contradiction between water demand and water supply and worsening the drought situation (Wu *et al.* 2017; Wang *et al.* 2018, 2019a).

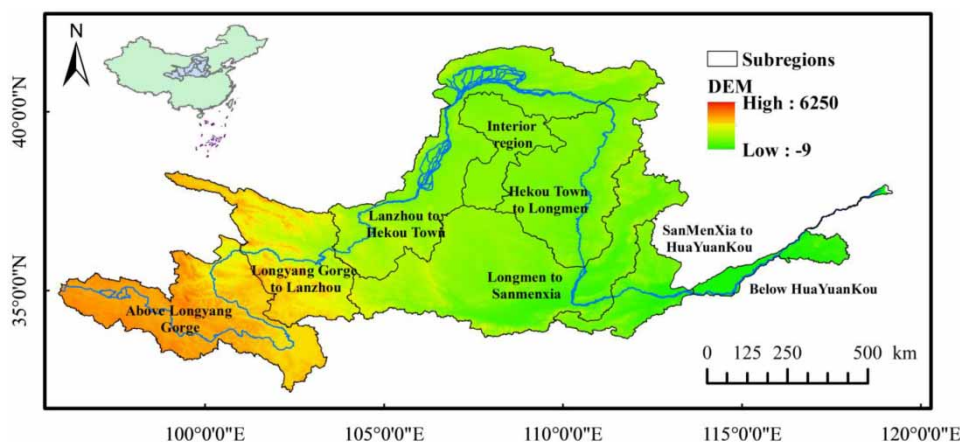


Figure 1 | Study area (the Yellow River Basin of China).

2.2. Data

The daily observed precipitation data during 1986–2015 came from the Meteorological Data Center of the China Meteorological Administration (CMA) (<http://data.cma.cn/>). It was gridded into the spatial resolution of $0.5^\circ \times 0.5^\circ$, the data of the YRB were extracted and put into use.

After a comprehensive multi-criteria assessment, Zhou & Han (2018) selected 5 models (Table 1) from 18 CMIP5 GCMs. Therefore, the total rank of these selected models is more satisfactory. In the model evaluation of Zhou & Han (2018), several statistics on precipitation and temperature were calculated. The statistics consisted of climatological mean, interannual variation, seasonal cycle, inter-annual modes, and probability distribution function. They were assessed by Mean Absolute Error, Correlation Coefficient, Root Mean Square Error, and S-value (for the degree of overlap between the simulated probability distribution and the observed probability distribution). The eighteen models were ranked by Rank Score (RS) performance, with these five models ranking better than other models. These models were corrected by the Quantile-Mapping method (Tong *et al.* 2017; Han *et al.* 2018).

Each model consists of daily simulated climate data during 1986–2005 and daily projected climate data during 2006–2099, including climatic elements of precipitation and temperature data under three emissions scenarios of RCP 2.6, RCP 4.5, and RCP 8.5. The detailed information of the selected models is shown in Table 1. The gridded climate data in the YRB were extracted as the observed precipitation data. Notably, for comparing the precipitation variation in a different time in the future, we chose two representative periods from the future period, the period of 2021–2050 (the near future scenario) and 2061–2090 (the distant future scenario). Given the homogeneity of different climatic elements and emissions scenarios, the climatic element of precipitation under emissions scenarios from RCP 4.5 was only analyzed in this study.

Considering the non-correspondence of the spatial resolution between the observed precipitation and the climate data, we unified the grids between observed data and climate data by matching the longitude and latitude of each grid, confirming that there are 415 coincident grids in both observation data and climate data.

3. METHODOLOGY

In this section, we state our chosen precipitation indexes. The research methods mainly included BMA and the probability estimation method. BMA was used to calculate the optimal weight of each model for the ensemble and to project each precipitation index deterministically (in Section 4.3). The probability estimation method calculated the probabilistic projections of future variations of each precipitation index based on the BMA weights (in Section 4.4).

3.1. Precipitation indexes

There are four precipitation indexes adopted in this study (Table 2). The *PY* indicates the total quantity of precipitation in the whole year over the YRB, while the *PS* indicates summer precipitation, which affects the water resources in the YRB. The R_{10} indicates the frequency of heavy rain that occurs over the YRB. It reflects the extreme attributes of precipitation, in other words, the frequency of rainstorm events. The *PI* indicates the concentration of precipitation. It shows the concentration of precipitation events in time in a year. The *PY* and the *PS* are calculated from the annual precipitation, while the R_{10} and the *PI* are daily. It is noted that the annual precipitation is the mean value of annual precipitation in all grids, and so is the daily precipitation.

Table 1 | The selected GCMs

No.	Model name	Temporal resolution	Spatial resolution
1	CNRM-CM5 (CN)	Daily	$0.25^\circ \times 0.25^\circ$
2	CSIRO-Mk3-6-0 (CS)	Daily	$0.25^\circ \times 0.25^\circ$
3	EC-EARTH (EC)	Daily	$0.25^\circ \times 0.25^\circ$
4	MIROC-ESM-CHEM (MI)	Daily	$0.25^\circ \times 0.25^\circ$
5	NorESM1-M (No)	Daily	$0.25^\circ \times 0.25^\circ$

Table 2 | The definition of precipitation indexes

Index name	Description	Reference	Unit
Annual mean precipitation (PY)	The mean value of annual precipitation	Libanda & Nkolola (2019)	mm
Summer precipitation (PS)	The mean value of summer precipitation	Qing <i>et al.</i> (2017)	mm
Heavy rain index (R_{10})	The mean number of days with daily precipitation is more than 10 mm.	Hanittinan <i>et al.</i> (2020)	day
Annual mean precipitation intensity (PI)	The mean ratio of annual precipitation and the number of days with daily precipitation greater than 1.0 mm.	Wu <i>et al.</i> (2019)	mm/day

3.2. Bayesian model averaging

BMA is a widely used probabilistic method for an ensemble of referable members (Hoeting *et al.* 1999), synthesizing the valuable information into a comprehensive and superior output from several comparable alternatives. BMA is typically applied to multi-model ensemble prediction of hydrometeorology, with temperature (Duan *et al.* 2007), streamflow (Yan *et al.* 2020), and precipitation (Liu & Xie 2014) among the typical cases. BMA can also integrate future projections from different climate models (Duan & Phillips 2010). In analyzing the changing mechanism of a variable under multiple factors, BMA can facilitate the desired results (Duan & Phillips 2010; Liu & Xie 2014). Wang *et al.* (2017b) applied the BMA ensemble method to obtain the multi-model ensembles projection of Kaifeng in the YRB from four GCMs under four emissions scenarios (RCP2.6, RCP4.5, RCP6.0, and RCP8.5). BMA was compared with the simple averaging method and individual models, and the merged projection could better capture the changing behaviors of climatic factors due to its lowest bias. These pieces of evidence show that BMA has advantages in dealing with multi-model issues. According to their simulation or reproduction competency to our object variables, the effect of BMA is usually regarded as assigning weights to each member for its extraordinary ability to reduce the uncertainty of different outputs. In this study, BMA is used for future precipitation projection. All the variables that appear below are directly related to the background of this study. In other words, the definition and description of variables are from the perspective of this study of precipitation projection based on multiple climate models.

If Pr is the predicted precipitation, $D = [Pr_1^o, Pr_2^o, \dots, Pr_T^o]$ is the observed precipitation data set used for BMA training with data length T , and $M = [M_1, M_2, \dots, M_K]$ is the ensemble of precipitation projections from K models, then, with the law of total probability, the BMA probabilistic projection of Pr can be expressed as:

$$p(Pr|D) = \sum_{k=1}^K p(M_k|D) \cdot p_k(Pr|M_k, D) \quad (1)$$

where $p(Pr|D)$ is the predicted probability of Pr given observed precipitation D , $p(M_k|D)$ is the posterior probability of a model projection M_k given observed precipitation D , and $p_k(Pr|M_k, D)$ is the posterior distribution of Pr given model projection M_k and observed precipitation D there is always $\sum_{k=1}^K p(M_k|D) = 1$.

Then, the posterior mean precipitation of the BMA projection is:

$$E(Pr|D) = \sum_{k=1}^K p(M_k|D) \cdot E[p_k(Pr|M_k, D)] = \sum_{k=1}^K w_k M_k \quad (2)$$

where $E(Pr|D)$ is the expected precipitation of BMA projection, w_k refers to $p(M_k|D)$, and we can obtain $\sum_{k=1}^K w_k = 1$, and w_k is the weight of the k th model as well as the right objective of the BMA method.

Considering the requirement of Gaussian distribution for $p_k(Pr|M_k, D)$, we transform both the observed precipitation data and model projection data using the Box-Cox transformation first, and all the data would be transformed with the same parameter.

Define $\theta = \{w_k, \sigma_k, k = 1, 2, \dots, K\}$ as the parameters to be solved, then the log-likelihood function is:

$$L(\theta) = \log \left(\sum_{k=1}^K w_k \cdot p_k(Pr|M_K, D) \right) = \log \left(\sum_{k=1}^K w_k \cdot g(Pr|M_K, \sigma_k^2) \right) \quad (3)$$

where σ_k^2 is the variance of model projection M_k concerning observation D .

Generally, w_k is optimized by the Expectation-Maximization (EM) algorithm with the same distribution requirement as BMA. Since the transformed precipitation data approximately obeys the Gaussian distribution, the algorithm can be implemented legitimately. In brief, the EM algorithm consists of the E step and M step. The best solution of w_k is generated through the alternation between these two steps. The following steps describe the procedure of the EM algorithm in detail.

Step 1: Initialization

$$\text{Iter} = 0, w_k^{\text{Iter}} = \frac{1}{K}, \sigma_k^{2(\text{Iter})} = \frac{1}{K} \frac{\sum_{k=1}^K \sum_{t=1}^T (Pr_t - M_{k,t})^2}{T} \quad (4)$$

where T is the total number of data points in the training period.

Step 2: Computing the initial likelihood

$$L(\theta)^{(0)} = \sum_{t=1}^T \log \left(\left(\sum_{k=1}^K w_k^{(0)} \cdot g(Pr_t^{\text{obs}} | M_{k,t}^t, \sigma_k^{2(0)}) \right) \right) \quad (5)$$

where $g()$ denotes the Gaussian distribution.

Step 3: Computing a latent variable known as the expectation step

$$\text{Iter} = \text{Iter} + 1, \hat{z}_{k,t}^{\text{Iter}} = \frac{g(Pr_t | M_{k,t}, \sigma_k^{2(\text{Iter}-1)})}{\sum_{k=1}^K g(Pr_t^{\text{obs}} | M_{k,t}, \sigma_k^{2(\text{Iter}-1)})} \quad (6)$$

where $z_{k,t}$ is a latent variable.

Step 4: Computing the weights known as the maximization step

$$w_k^{\text{Iter}} = \frac{1}{T} \sum_{t=1}^T \hat{z}_{k,t}^{\text{Iter}} \quad (7)$$

Step 5: Computing the projection variance

$$\sigma_K^{2(\text{Iter})} = \frac{\sum_{t=1}^T \hat{z}_{k,t}^{\text{Iter}} \cdot (Pr_t^{\text{obs}} - M_{k,t})^2}{\sum_{t=1}^T \hat{z}_{k,t}^{\text{Iter}}} \quad (8)$$

Step 6: Updating the likelihood using Equation (7)

Step 7: Checking convergence

If $L(\theta^{\text{Iter}}) - L(\theta^{\text{Iter}-1})$ is no more than a prespecified tolerance level, stop; otherwise, go back to Step 3 for a new optimization circle.

3.3. Probabilities in different variation ranges

Based on the BMA weights, the probability of variation refers to the probability that the variation of a precipitation index locates in a particular range. According to the future changes of each index, six variation ranges are selected for each

Table 3 | Variation ranges (VRs) of precipitation indexes (Inf. means infinity, –Inf. means infinitesimal)

Indexes name	VR1	VR2	VR3	VR4	VR5	VR6
PY (mm)	–Inf. to –50	–50 to –25	–25 to 0	0–25	25–50	50 to Inf.
PS (mm)	–Inf. to –50	–50 to –50	–25 to 0	0–25	25–50	50 to Inf.
R ₁₀ (day)	–Inf. to –4	–4 to –2	–2 to 0	0–2	2–4	4 to Inf.
PI (mm/day)	–Inf. to –1	–1 to –0.5	–0.5 to 0	0–0.5	0.5–1	1 to Inf.

index. The variation ranges are shown in Table 3. In this paper, the variation ranges (VRs) are defined as events. For example, if the variation of PY is between –50 and –25 mm, a VR2 event is considered to have occurred in PY (the event level is 2).

The BMA method is based on the total probability formula. The principle is to solve the possibility of various causes (if they are mutually exclusive) through the result that has already occurred or find out the most likely cause for this result. Theoretically, the result must be caused by one of those causes. The cause of the current result must be found among all those causes that have been known in advance. Therefore, there is a prerequisite: every possible cause has an objective explanation for the result. For the precipitation projection, the observed precipitation is the result, and the precipitation in each climate model is the cause. There is no evidence to suggest that precipitation in these climate models can perfectly (or partially) simulate observed precipitation. However, we proposed that climate models can explain the observed precipitation process under the current understanding. There were two reasons for this. First, these climate models were designed to simulate precipitation as closely as possible, and researchers had considered the physical mechanisms of precipitation. Second, the climate model used in this study had been strictly evaluated, and the models with the best performance of precipitation simulation had been screened from many climate models. We believed that this hypothesis was reasonable and necessary for these two reasons.

Given the principle of the BMA method (to solve the probabilities of each cause based on the result) and the reasonable assumption, we can use the obtained probabilities (the BMA weights) of each cause (the models) to predict the future variations in precipitation and their probabilities. In this study, five climate models were applied. The probability of variation means the proportion of the projected results within a specific range. This can be defined as:

$$P^{LVR \leq \Delta_i < UVR} = \sum_{LVR \leq \Delta_{ij} < UVR} w_j \quad (9)$$

where P is the probabilities of variation of i (i means one of the precipitation indexes), Δ_i is the variation of the i th index, Δ_{ij} is the variation of the i th index in the j th model, w_j is the BMA weight of the j th model, LVR and UVR is the lower bound and upper bound in the current variation ranges for index i (Table 3), respectively. For example, if the projected future variations of PY ($i = 1$) from the five models ($\Delta_{i,j}$ or $\Delta_{PY,j}$) are 37, –6, 18, 20, and –9 mm, and the BMA weights of the five models (w_j) are 0.14, 0.23, 0.38, 0.09, and 0.16. Then, the probabilities of different variations of PY in the future are:

$$\begin{aligned}
 P^{\Delta_{PY} < -50} &= \sum_{\Delta_{PY,j} < -50} w_j = 0 \\
 P^{-50 \leq \Delta_{PY} < -25} &= \sum_{-50 \leq \Delta_{PY,j} < -25} w_j = 0 \\
 P^{-25 \leq \Delta_{PY} < 0} &= \sum_{-25 \leq \Delta_{PY,j} < 0} w_j = \sum_{j=2,5} w_j = 0.39 \\
 P^{0 \leq \Delta_{PY} < 25} &= \sum_{0 \leq \Delta_{PY,j} < 25} w_j = \sum_{j=3,4} w_j = 0.47 \\
 P^{25 \leq \Delta_{PY} < 50} &= \sum_{25 \leq \Delta_{PY,j} < 50} w_j = \sum_{j=1} w_j = 0.14 \\
 P^{50 \leq \Delta_{PY}} &= \sum_{50 \leq \Delta_{PY}} w_j = 0
 \end{aligned} \quad (10)$$

3.4. Performance assessment of climate models

To better reflect the simulation and projection performance of all models and their ensembles precipitation in the whole YRB, the mean value of all grids in the basin was taken as the precipitation assessment index for the basin. The BMA weight of each model was calculated by taking the projected annual precipitation of each model as the input and the observed annual precipitation as the output. The superiority of the ensemble was evaluated by Root Mean Squared Error (*RMSE*) and total deviation (*TD*). The smaller the *RMSE* is, the better the model's performance in process simulation. In contrast, the smaller the *TD* is, the better the model's performance in quantitative simulation. The *RMSE* and *TD* are defined as:

$$RMSE = \sqrt{\frac{1}{m} \sum_{i=1}^m (M_i - H_i)^2} \quad (11)$$

$$TD = \frac{1}{m} \sum_{i=1}^m M_i - \frac{1}{m} \sum_{i=1}^m H_i \quad (12)$$

where M_i is the i th historical simulation of each model, including the BMA, H_i is the i th historical observation of each model, including the BMA, m is the total number of years.

Considering the primary purpose of this study is to explore the variation of precipitation indexes in the future based on the BMA model, it is necessary to investigate the simulation performance of the BMA model for precipitation indexes. Therefore, to accurately describe the simulation deviation of a model, the mean of the absolute value of simulation deviation (*MASD*) is applied, the *MASD* here is defined as:

$$MASD = \frac{1}{m} \sum_{i=1}^m |P_{si} - P_{oi}| \quad (13)$$

where the P_{si} is the simulated precipitation index of the i th year and the P_{oi} is the observed precipitation index of the i th year.

The historical simulation performance of the climate model determines whether a model is satisfactory. In other words, the credibility of precipitation in the future predicted by a climate model rests with its historical performance, and simulation deviation is the right indicator to explain the historical performance of a climate model. The simulation deviation is always fluctuant, so the mean indicates its long-term average level of deviation from observation, while the variance indicates the stability of the simulation deviation. The following formula is used to calculate the mean and variance of annual relative simulation deviation:

$$E_D = \frac{1}{m} \sum_{i=1}^m D_i \quad (14)$$

$$V_D = \frac{1}{m-1} \sum_{i=1}^m (D_i - E_D)^2 \quad (15)$$

$$D_i = \frac{|M_i - H_i|}{H_i} \quad (16)$$

where E_D is the mean long-term annual relative simulation deviation, V_D is the variance of long-term annual relative simulation deviation, and M_i is the i th historical simulated value of each model, H_i is the i th historical observed value of each model, D_i is the absolute value of relative simulation deviations, and i is the number of years.

4. RESULTS

4.1. BMA weight of each model and the superiority of ensemble

The BMA weight of each model was calculated, and the historical simulating performance was assessed by *RMSE* and *TD* simultaneously (Table 4). The CS model has the maximum weight while the CN model has the minimum. The weight of the other models stands between the two. From a process perspective, the CS model has the minimum *RMSE*, and the

Table 4 | BMA weights and historical precipitation simulating assessment of each model

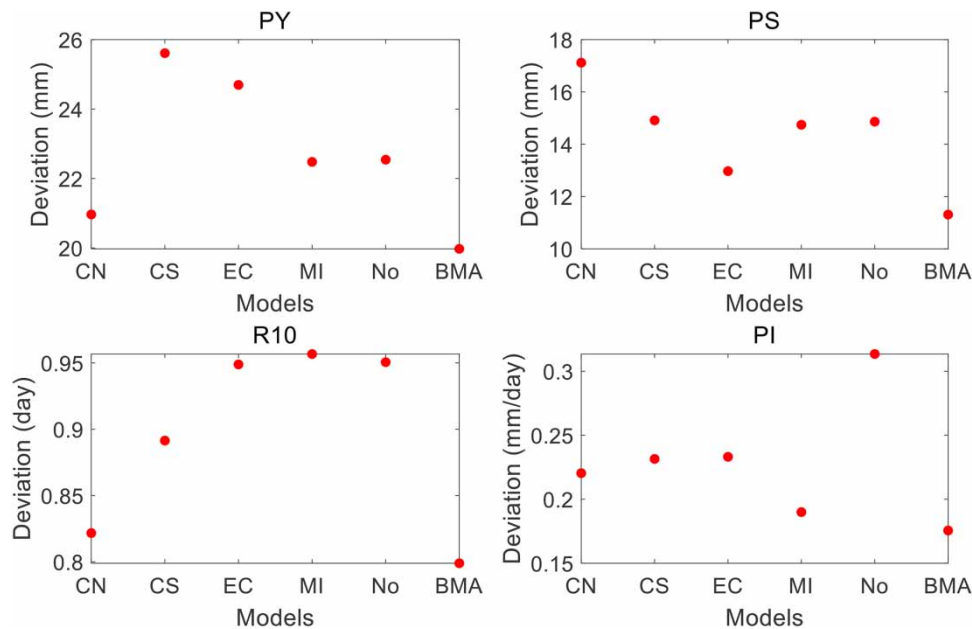
Model name	CN	CS	EC	MI	No	BMA
RMSE (mm)	65.9	63.9	69.5	85.2	83.2	51.7
TD (mm)	4.2	11.1	12.1	8.6	0.4	7.8
Weight	0.144	0.258	0.149	0.233	0.216	/

CN model and the EC model have a similar level to the CS model, while the MI model has the maximum *RMSE* and the No model is close to it. From a quantitative perspective, the No model has the minimum *TD*, the EC model has the maximum, and the CS model is close. The BMA model has the minimum *RMSE* among all the models (since the calculation of the BMA weights is based on process performance). It has better performance than most models (better than 3/5 of the models) from the quantitative perspective, though it is not the best.

The BMA model simulates the average state of historical precipitation indexes more precisely with the smaller *MASD* than any single model (Figure 2). Each model has different simulation performances in different precipitation indexes, but the BMA model consistently has the best performance in all four precipitation indexes with the minimum *MASD*. At the same time, every model among the five has its best or worst performance for different indexes. For instance, the CN model performs nicely for the *PY* and *R₁₀*, but not so for the *PS* and *PI*, and even worse for the *PS*. Thus, it is necessary to do a model ensemble to simulate and project the precipitation over the YRB.

Climate models are inherently uncertain, so is the BMA model derived from climate models. Therefore, credibility is a crucial reference to precipitation index estimation. The credibility of precipitation indexes estimation is quantitatively described by the E_D (mean long-term annual relative simulation deviation, Equation (14)) from the perspective of competence and the V_D (variance of long-term annual relative simulation deviation, Equation (15)) from the robustness perspective, respectively. The E_D and V_D on each grid in Figures 3 and 4 are based on the relative simulation deviation instead of the absolute simulation deviation. It is worth noting that, for the BMA model in Figures 3 and 4, the E_D and V_D are calculated from the annual time series of precipitation indexes ensembled by BMA weights. The blank grids in *R₁₀* are since the observed results of these grids are 0, which leads to abnormal calculation results.

There is a significant improvement in *PY*, *PS*, and *PI* estimation. Furthermore, from an E_D perspective (Figure 3), the simulation deviation of the BMA model is smaller than that of any other model. Finally, from a V_D point of view (Figure 4), the BMA model has a more stable simulation deviation than any other model.

**Figure 2** | The *MASD* of four precipitation indexes in the YRB by different models.

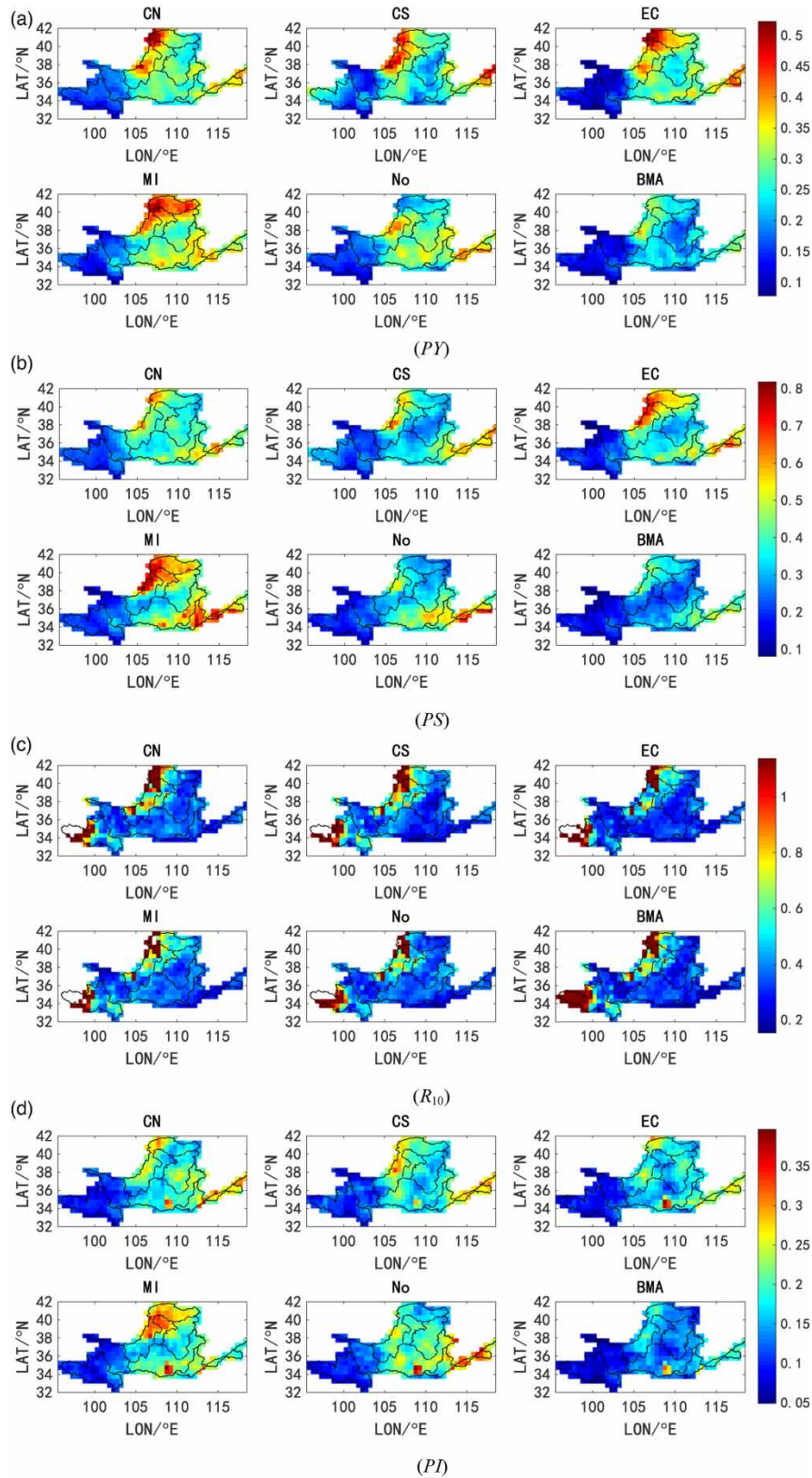


Figure 3 | Comparison of E_D (mean of long-term annual relative simulation deviation) between the BMA model and other models. The smaller the E_D in the BMA model, the smaller the deviation of the BMA model.

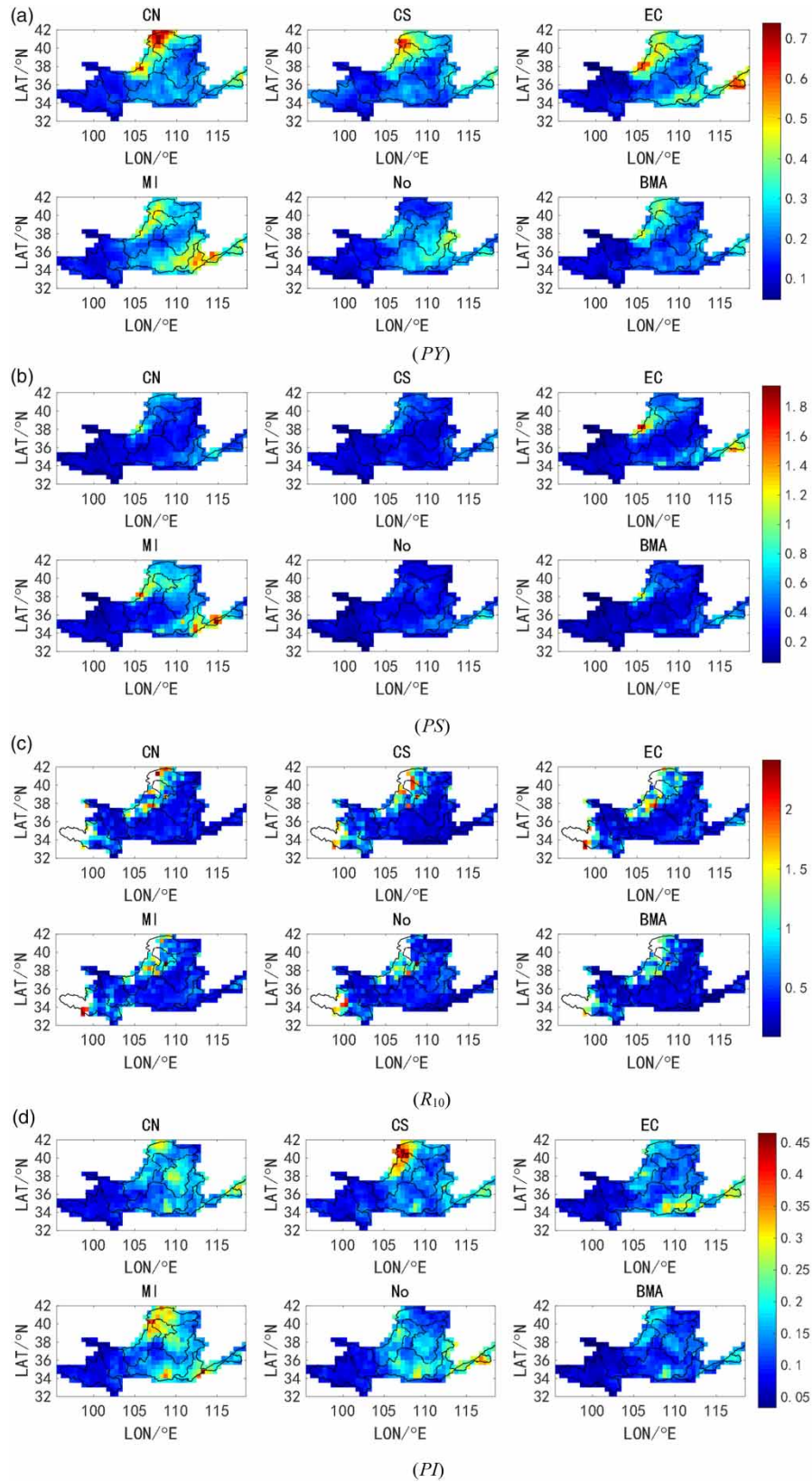


Figure 4 | Comparison of V_D (variance of long-term annual relative simulation deviation) between the BMA model and other models. The smaller the V_D in the BMA model, the more stable the deviation of the BMA model.

Concerning the *PY* (Figure 3(a)), the relatively large E_D in the north part of the CN, CS, EC, and MI models are as high as 0.5. However, that of the BMA model is no larger than 0.35. Thus, though the No model has less E_D than the other models, it is still higher than the BMA model. The same conclusion could be seen with the *PS* (Figure 3(b)) and *PI* (Figure 3(d)), as well as in the east of the YRB. In all models, the E_D in the west is smaller than in the east, coming in below 0.2 generally in the west and more significant than 0.3 in the east.

The R_{10} (Figure 3(c)) has a different result, and it seems that the E_D from the BMA model is not necessarily less than the other individual models. However, as can be seen from R_{10} , the result of the BMA model is not worse overall.

As with the V_D , the BMA model is not worse or better than any other model. Taking *PY* (Figure 4(a)) as a case, in the north YRB from the CN, CS, and EC model, and so with the north and southeast of the YRB from the MI model, the V_D of these regions is more significant than 0.5 and even up to 0.8. However, it is correspondingly less than 0.4 with the BMA model. In short, the performance of the BMA model is superior to that of other individual models on R_{10} (Figure 4(c)) and *PI* (Figure 4(d)). The *PS* (Figure 4(b)) is more acceptable than the EC and MI models while performing equivalent with the CN, CS, and No models.

In the projection of future precipitation, it is crucial to understand the credibility of future projection. The credibility here refers to whether the BMA model is more credible than any single model in the deterministic projection. In this study, the assessment indicators are set as E_D , the mean long-term annual relative simulation deviation, and V_D , the variance of long-term annual relative simulation deviation. As mentioned above, the simulation deviation is expressed as the absolute value of relative simulation deviation since it is interested in the deviation of both directions (i.e., positive and negative). The lower E_D means the smaller range of deviation, and the lower V_D , the more stable the deviation. As shown in Figures 2–4, compared with each model, the performance of reconstructing the historical precipitation situation of the BMA model is more satisfying in terms of lower *MASD*, E_D , and V_D . Thus, the BMA model has more significant advantages in predicting future climate under specific scenarios. Consequently, we get more rational ensemble weights and probability estimation of different changes in precipitation indexes. Based on the assumption of weight as a probability in the BMA model, the application of BMA improves the flexibility and the credibility of the projection.

Therefore, it is possible to conclude that the BMA model synthetically performs better than any single climate model in simulating the historical precipitation in the YRB due to valid information from each pattern. Thus, the precipitation in the future projected by the BMA model will be more rational than any single model.

4.2. Evaluation of the probability estimation method

We took 2006–2015 as the validation period to verify the probability estimation method used in this study. This period was the historical observation period and the projection period in the models, making it suitable to be the validation period. We compared the estimated variation with the observed variation and quantified the deviation of the method and its spatial distribution for each precipitation index. The detailed validation steps are as follows:

1. As above, we defined the variation ranges of precipitation index (VR) as the difference of annual mean value between the validation period (2006–2015) and the historical period (1986–2005).
2. According to the observed precipitation data, the variation of each precipitation index was calculated, and they were converted into VRs (observed VR, Figure 5) according to Table 3.
3. Similarly, for the projected precipitation data (GCMs), the estimated variations of each model (2006–2015) were calculated, respectively. Then, the probability of different variations of each precipitation index was calculated according to the probability estimation method. The VR with the highest probability (estimated VR) in each grid was selected as the estimated variation of each precipitation index.
4. The difference between the estimated VR and the observed VR was taken as the deviation of probability estimation. In each grid, the estimation deviation of each precipitation index was calculated separately. The proportion of deviations of different sizes in the whole basin was classified (the proportion of grids with various deviations in the total number of YBR grids). The statistical results of deviations are shown in Table 5.

We needed to learn about the observed precipitation variation during the validation period. There was no apparent spatial characteristic in the observed variation of each precipitation index shown in Figure 5. The VRs in the four precipitation indexes differed significantly in terms of area. The VR6 occurred in 35.1% of the grids in *PY*, while the VR1 (0%) and the

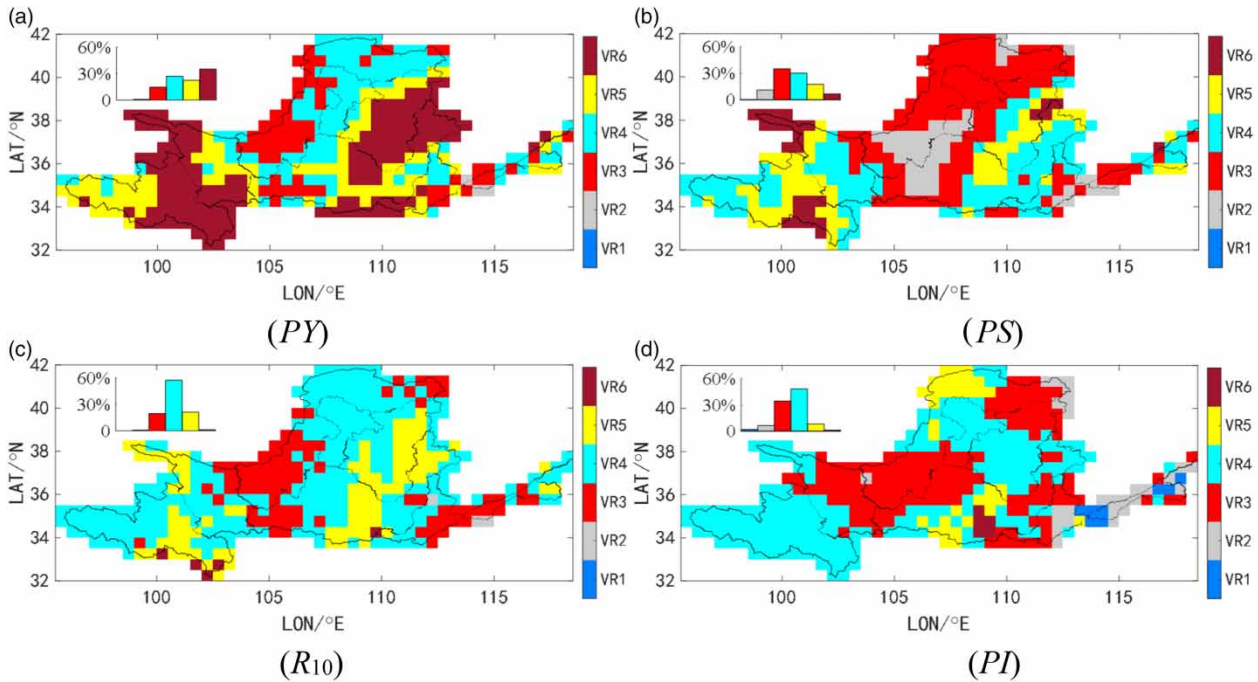


Figure 5 | Observed variation of four precipitation indexes in 2006–2015. This classified representation is based on every precipitation index’s variation range (VR) at each grid. For example, if the variation of *PY* in a grid is VR5, the average variation of *PY* is between 25 and 50 mm (see Table 3 for VRs). The proportions in the graph refer to the percentages of the number of grids occurring in each VR.

Table 5 | Proportions (%) of different deviations between observed VR and estimated VR

Deviation	-5	-4	-3	-2	-1	0	1	2	3	4	5
<i>PY</i>	0	0.24	1.21	10.60	25.31	38.31	18.31	5.54	0.48	0	0
<i>PS</i>	0	0	0	5.78	16.87	30.84	34.70	11.57	0.24	0	0
<i>R₁₀</i>	0	0	0	3.37	29.88	47.47	17.35	1.93	0	0	0
<i>PI</i>	0	0	0	3.37	13.74	53.74	18.31	6.51	2.89	1.20	0.24

The difference in the number of VR, if a grid was estimated as VR2 and observed as VR4, the deviation is -2. Negative deviations mean underestimating and positive deviations mean overestimating. A deviation of 0 indicates an accurate estimation.

VR2 (0.72%, only three grids) had no or almost none. In *PS*, the VR3 occurred in 34.9% of the grids, while the VR1 (0%) and the VR2 (6.5%) did not occur or very little. The VR4 occurred in 57.4% of grids in *R₁₀*, while the VR1 (0%), the VR2 (0.7%), and the VR6 (1.2%) did not or very little. The VR4 occurred in 48.0% of the grids in *PI*, while the VR1 (0%), the VR2 (6.5%), the VR5 (8.2%), and the VR6 (1.2%) had no or very few occurrences.

Similarly, we counted the estimated variation of the precipitation during the validation period, as shown in Figure 6. Compared with Figure 5, there was no apparent spatial relationship (in grids) between the observed results and the estimated results. Each VR (whether observed or estimated) had no apparent spatial distribution characteristics. In each precipitation index, the distribution of the estimated proportions was relatively close to the observed results for each VR, among which the *PY*, *R₁₀*, and *PI* were in good agreement. However, the estimated and observed results of *PS* were a bit different. The VR2 and the VR3 were underestimated by 8.9 and 20%, respectively (absolute value, not relative to the VR2 or the VR3, the same as below), and the VR4 was overestimated by 30.9%.

The deviation of the estimated result is a point of concern. Table 5 shows the proportion of grids with different estimation deviations in all grids of the whole basin. Among the four precipitation indexes of *PY*, *PS*, *R₁₀*, and *PI*, the grids with a deviation of 0 accounted for 38.31, 30.84, 47.47, and 53.74% of the whole basin, respectively, meaning that the VR with the highest estimated probability in these grids was consistent with observed VR. In other words, if the VR with the highest

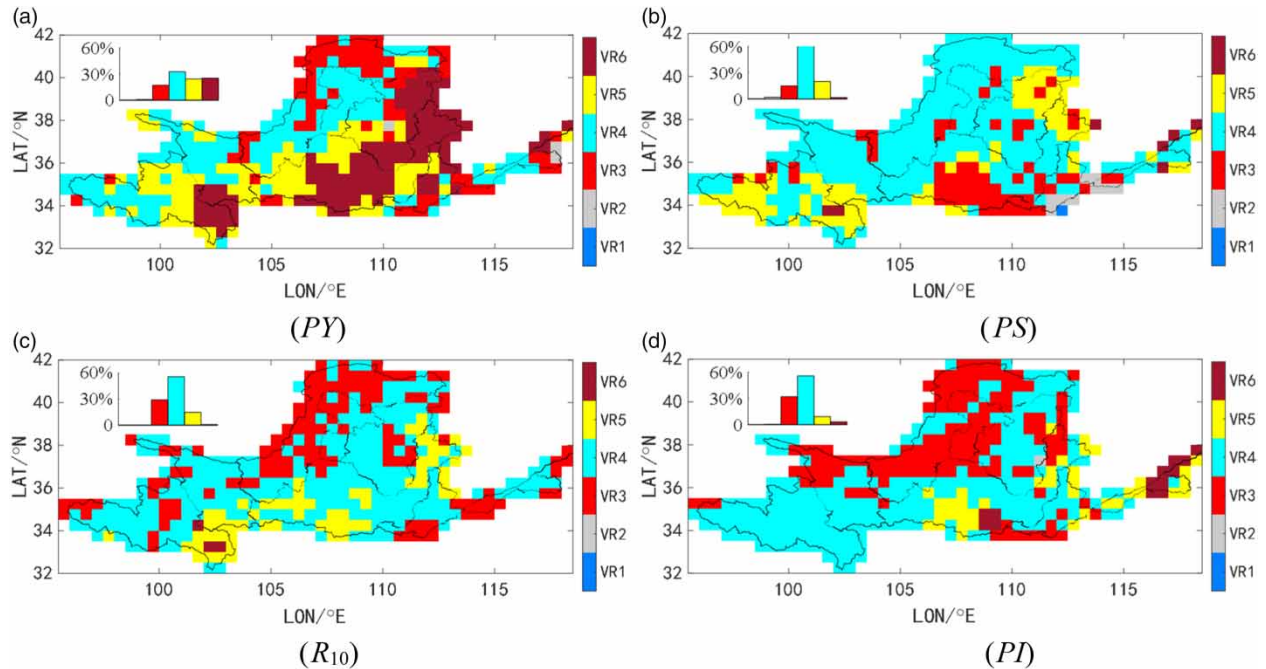


Figure 6 | Estimated average variation of four precipitation indexes in 2006–2015. As in Figure 5, the average variation is expressed as the classified VRs, but the VR at each grid has the highest estimated probability among six VRs.

estimated probability was used as the estimated result, the estimation accuracies of these four indexes were between 30.84 and 43.74%. The estimation deviations of 81.93, 82.41, 94.70, and 85.79% of the four precipitation indexes were between -1 and 1 . According to the current classification of precipitation indexes (Table 3), if the deviation of 1 level was taken as the allowable deviation, the probability estimation method could capture the actual occurrence of precipitation indexes in a significant probability (81.93–94.70%). The grids with an estimated deviation of -1 accounted for 13.74–29.88% of the whole basin, which meant that the VR with the highest estimated probability in these grids was underestimated by 1 level than the observed VR. The grids with an estimated deviation of 1 in the whole basin are between 17.35 and 34.70%. Therefore, the VR with the highest estimated probability in these grids was overestimated by 1 level than the observed VR. The proportions of grids with a deviation of 2 (overestimated by two levels) and -2 (underestimated by two levels) were between 1.93 and 11.57%. The total proportion of grids with overestimated or underestimated deviation in the four precipitation indexes of more than three levels was less than 4.33%.

Figure 7 shows the spatial distribution of the estimation deviation of each precipitation index and the main distribution areas of overestimation and underestimation. The overestimation of *PY* mainly appeared in the southern part of Subregion 3 and the western part of Subregion 5 and Subregion 7. The underestimation was mainly in the middle of Subregion 1, Subregion 2, and Subregion 6. The overestimation of *PS* was mainly in Subregion 3, Subregion 4, the west of Subregion 5, and the northwest of Subregion 6, while the underestimation was mainly in the southeast of Subregion 1, the northwest of Subregion 2, the east of Subregion 5, and the southeast of Subregion 6. The overestimation of R_{10} was mainly in the west of Subregion 5 and the east of Subregion 7. The underestimation was mainly in the middle of Subregion 1, the northwest of Subregion 2, the north of Subregion 3 and Subregion 6. The overestimation of *PI* was mainly in Subregion 7 and Subregion 8, the central and western parts of Subregion 5 and Subregion 3, and the northeast of Subregion 6, and the underestimation was mainly in Subregion 4 and the north of Subregion 3. Therefore, the projection of future precipitation variation in these regions should be focused on and treated with caution.

Analyzing the estimated composition of each observed VR (that is, which VR is estimated by the probability estimation method for each observed VR) helps us determine the deviations of the overestimation or underestimation in Figure 7. Figure 8 shows the composition of the estimation of each observed VR in different precipitation indexes. In the probability estimation method, the observed VR in a grid could be estimated as any VR1–VR6 (taking the VR with the highest estimated probability). For example, in *PY*, all grids where VR6 had occurred accounted for 35.1% of the whole basin (Figure 5(a)). In

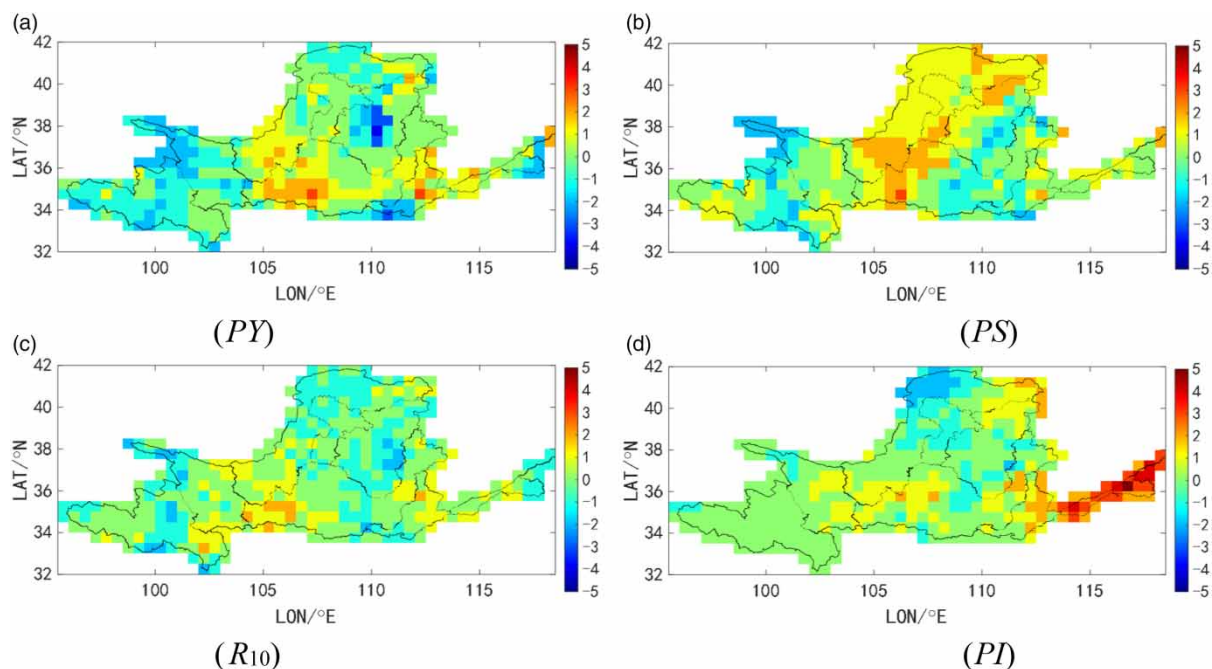


Figure 7 | Deviation between the observed VR and estimated the VR. The estimated VR has the highest probability among all VRs at each grid. The deviation is represented by the difference between the estimated VR number and the observed VR number. Where green means the deviation is 0 (accurate estimation), yellow means the deviation is 1 (overestimate by 1 level), and cyan means the deviation is 1 (underestimate by 1 level). Please refer to the online version of this paper to see this figure in colour: <http://dx.doi.org/10.2166/wcc.2022.359>.

the estimation results of these grids, the proportions of VR1–VR6 are 0, 0.7, 3.4, 21.2, 29.5, and 45.2%, respectively. The accuracy of the probability estimation method for VR6 in *PY* is 45.2%.

Figure 9 shows the accuracy of each VR in all precipitation indexes, corresponding to the occurrence proportion of each VR. *PY*, *PS*, and *R*₁₀ showed an apparent positive correlation between the estimation accuracy and the observed occurrence proportion. We also noticed that there were some differences in *PI*. This was due to the less observed occurrence proportion of VR1 (1.9%), VR2 (6.5%), and VR6 (1.2%), which led to large estimation deviations. There was a positive correlation between the estimation accuracy and the observed occurrence proportion of other VRs in *PI*.

We can explain why some areas in Figure 7 were overestimated or underestimated. For all the precipitation indexes in Table 5, the area of overestimation by 1 level accounted for the most significant proportion among the proportions of different degrees of overestimation, and the same is true in the underestimation. Since the proportion of overestimated or underestimated of each VR was based on the occurrence proportion of this VR, we would not consider those VRs with very few occurrences. For example, the occurrence proportion of VR6 in *PI* was 1.2% (five grids), although there were 20% underestimated by 1 level, and only one grid (5*20%) was underestimated. After statistical analysis, we decided not to consider the VRs that occurred in less than ten grids. In Table 6, both the underestimated and overestimated by 1 level came from the grids where the VR with a high proportion of occurrences. However, among VRs that occurred more often, higher-level VRs tended to be underestimated, while lower-level VRs tended to be overestimated. Therefore, in Figure 7, the proportions underestimated by 1 level and overestimated by 1 level mainly came from those VRs with higher occurrences.

In the probabilistic projection, credibility is mainly reflected in two aspects. The first is the deviation, the difference between the projection with the highest probability and the observation. Precipitation varies from place to place, leading to the difference in the ability of climate models to simulate precipitation in different places, which is the reason for the spatial heterogeneity of deviations. In Figure 7, the spatial distribution of the simulated deviations is a semi-quantitative measure of credibility. The larger the classified deviations, the lower the credibility is. For example, in the future projection of *PI* in Figure 7(d), the deviation of the source region (Subregion 1) is 0, it can be considered that the credibility of this region is relatively higher, and the deviation of the downstream region (Subregion 8) is between 2 and 5 (overestimated), it indicates that the credibility of this region is relatively lower.

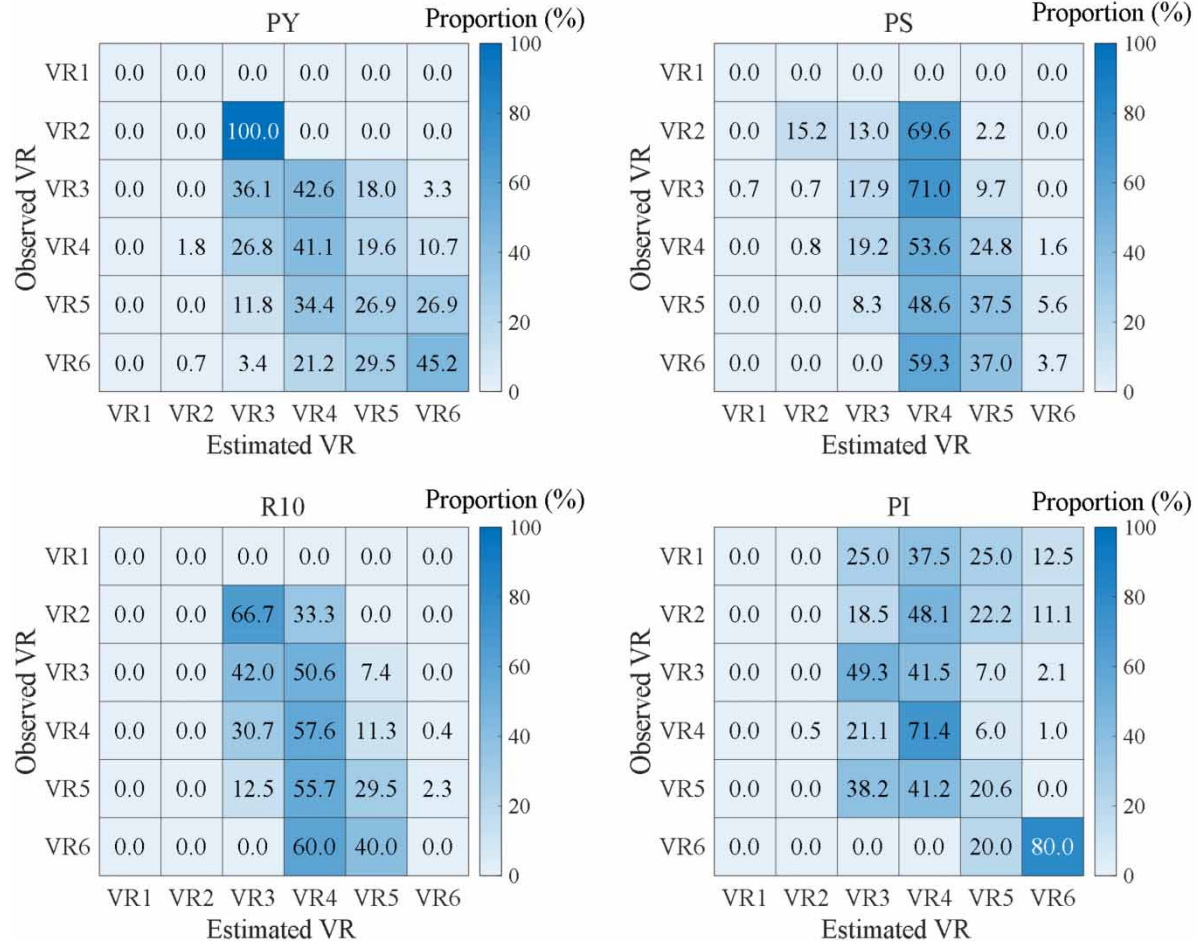


Figure 8 | Estimated composition of each observed VR in each precipitation index. The numbers represent the proportion of the number of grids estimated as different VRs in the observed VR. The numbers on the diagonal represent the estimation accuracy of the VR. In general, the sum of the numbers in a row of a VR is 100% in the subgraph of each precipitation index. If the sum is 0, the VR did not occur in reality.

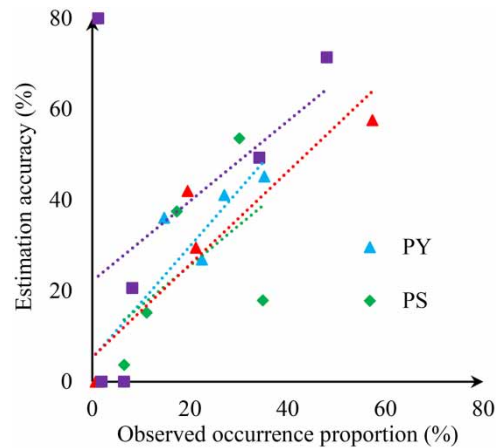


Figure 9 | Relationship between estimation accuracy and observed occurrence proportion for four precipitation indexes (in one precipitation index, the points come from all the VRs occurred).

Table 6 | VRs in main areas that were underestimated by 1 level and overestimated by 1 level in each precipitation index

Precipitation indexes	Underestimated by 1 level (the number of grids)	Overestimated by 1 level (the number of grids)
PY	VR4(30), VR5(32), VR6(43)	VR3(26), VR4(22), VR5(25)
PS	VR4(24), VR5(35), VR6(10)	VR3(103), VR4(31)
R ₁₀	VR4(73), VR5(49)	VR3(41), VR4(27)
PI	VR4(42), VR5(14)	VR3(59), VR4(12)

We also analyzed the accuracy of the probability projection method for different precipitation events (VRs). For each estimation, we can find the accuracy of the result in Figure 8, and we can also know the probability that the estimation might be for other VRs. The probability here is obtained by normalizing the numbers in Figure 8 by column, slightly different from the estimated composition discussed above. It is worth noting that the results in Figure 8 were mainly for the VRs that occurred more frequently in the past (as discussed above) because the more occurrences of VRs, the higher the accuracy of the projection would be (Figure 9).

4.3. Deterministic projection: the precipitation evolution in the future

The ensembled projected precipitation is calculated based on the BMA weights on all models. Figure 10 illustrates the estimation of the average variation for every precipitation index of the YRB in two 30-year periods (taking 2021–2050 as the near future scenario and 2061–2090 as the distant future scenario).

The *PY* in the YRB shows an increasing trend in both 2021–2050 (Figure 10(a-1)) and 2061–2090 (Figure 10(a-2)), and the annual precipitation variation in the northwest and the east (the downstream) of the YRB will be obviously and consistently more minor than the other region. The increase in annual precipitation will be greatest in the southwest of the YRB and up to more than 80 and 120 mm locally under the two future scenarios. There is a spatial pattern of low-high-low (L-H-L) for the distribution of *PY* variation from the northwest to the southeast of the YRB.

Compared with the *PY*, the *PS* (Figure 10(b-1) and 10(b-2)) in the YRB will vary. In terms of the overall trend, there will be a bit of a decreasing trend locally under both the future scenarios. The decrease in *PS* will reach 10 and 20 mm, respectively. Unlike the *PY*, the *PS* in the southwest and northeast tends to increase, but the southeast will be opposite, and both increases and decreases in 2061–2090 are more significant than that in 2021–2050. In particular, there is a spatial pattern of high-low-high (H-L-H) for the distribution of *PS* variation from the southwest to the northeast of the YRB.

The *R*₁₀ will have a trend of increasing in most areas of the YRB with an increment of 1–2 in 2021–2050 (Figure 10(c-1)) and 2–4 in 2061–2090 (Figure 10(c-2)). Dramatically, there will be a more significant increase of *R*₁₀ in the southwest of the YRB, where the *R*₁₀ will be up to 4 and 7 in 2021–2050 and 2061–2090, respectively. Thus, spatially, the distribution pattern is somewhat like the *PY*.

The *PI* of the YRB (Figure 10(d-1) and 10(d-2)) will have a slightly increasing trend in both periods. There is no recognizable spatial pattern of *PI* variation. Still, *PI* variation in the northwest of the YRB will be more stable wholly based on a consistent increase or decrease of less than 0.5.

With the future estimation of precipitation indexes in Figure 10 and the definition of these indexes in Table 2, the precipitation evolution in the future could be projected sequentially. The projection can be implemented from the four indexes. The annual precipitation is more concerning due to its dominant effect. In all areas of the YRB, the annual precipitation will increase. The increase from 2061 to 2090 is more significant than 2021–2050, while the annual increase in the southwestern region will be more significant than that in other regions. The summer precipitation is another focus for the impact on the flood problem. In the southwest and east regions, the increment will be dramatically higher than in both scenarios. Comparing the precipitation variation of *PY* and *PS*, the increment of annual precipitation mainly occurs in summer, which means the risk of flooding will increase in the future. The *R*₁₀, one of the most direct indicators of rainstorm disaster, will also increase on average about four to seven times per year in the southwest and two to four times in the midland. With the *PY*, *PS*, and *R*₁₀, the southwest of the YRB, namely the upstream, needs more attention with the most increment of these indexes. The situation will be more difficult in the distant future (2061–2090). By contrast, the *PI* will increase slightly, mainly in the south of the YRB, still increasing the whole YRB.

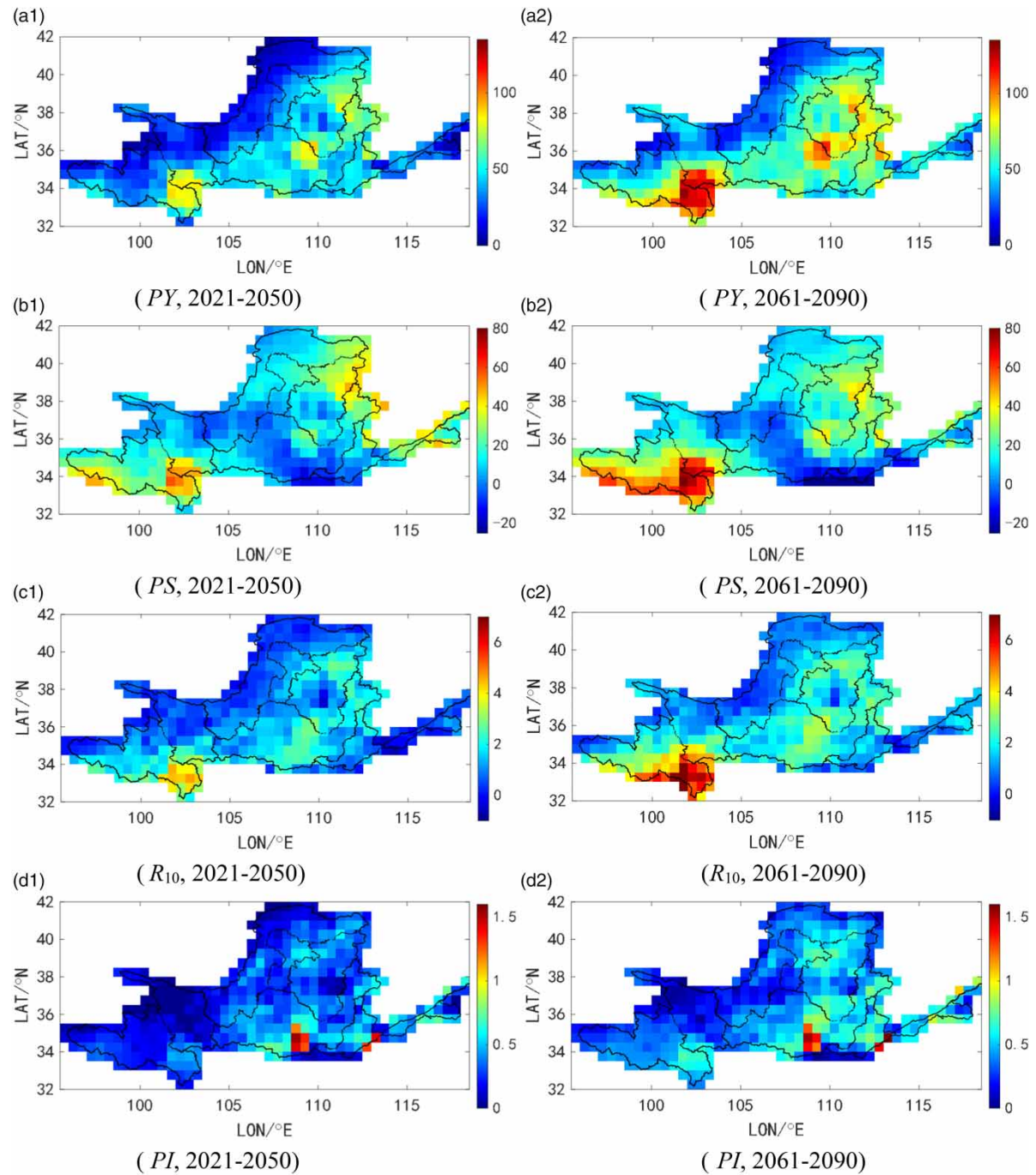


Figure 10 | Estimated average variation for four precipitation indexes of the YRB in the coming 30 years.

4.4. Probabilistic projection: probability estimation of future precipitation variation

As shown in Figure 11, the VR in every subgraph means the corresponding index has changed in a specific range. The meaning of VR corresponding to each indicator is shown in Table 3. For one grid, the sum of the probability values of all six VRs is 1, where the sum of the probabilities of VR1–VR3 is the probability of decreasing, and the sum of the probabilities of VR4–VR6 is the probability of increasing. Obviously, among all the indexes under both scenarios in Figure 11, the probability of increasing is greater than that of decreasing in most regions of the YRB.

In most regions of the YRB, the probability of *PY* increasing by more than 50 mm in both scenarios is much higher, and the probability in 2021–2050 (Figure 11(a-1)) is greater than that in 2061–2090 (Figure 11(a-2)), especially in the south of the

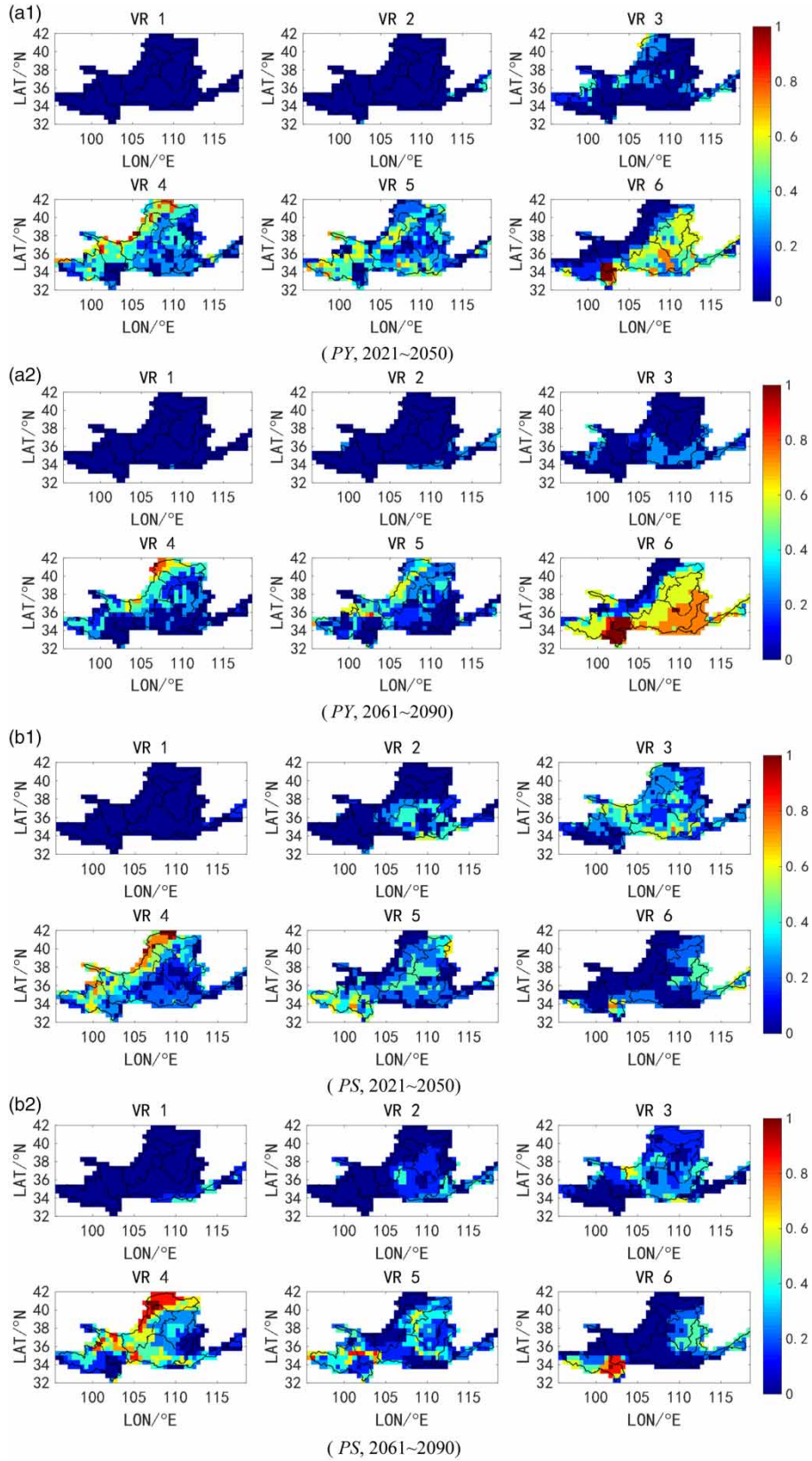


Figure 11 | Probability estimation of future variation in different ranges for precipitation indexes. Each VR represents a specific change in the precipitation index, VR1–VR3 represent the decreased change, and VR4–VR6 represent the increased change. The specific representations are shown in Table 3 (continued).

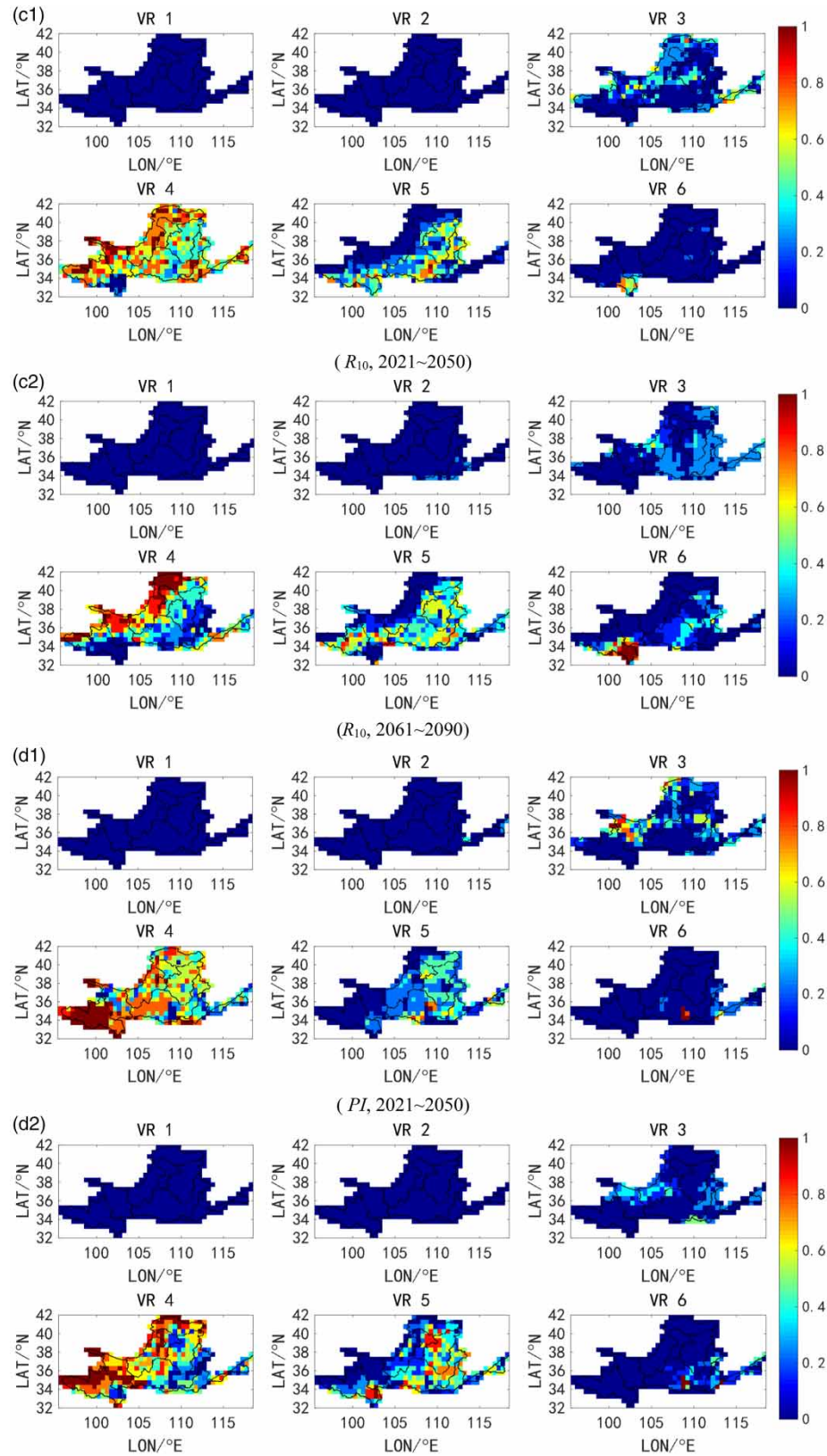


Figure 11 | Continued.

YRB. In the northwest of the YRB, it may increase 0–50 mm with a probability of 40–60%, whatever the scenario. The probability of *PY* decreasing by more than 25 mm in almost all regions (VR1 and VR2 in Figure 11(a-1) and 11(a-2)) is close to 0 under both scenarios.

PS has a more decentralized probability distribution than *PY* and is the largest among all indexes. By comparison with *PY*, the probabilities of *PS* increasing by 0–25 mm or decreasing by 0–25 mm will be proportionable under both scenarios. In particular, the difference in the spatial distribution pattern of probability between the two scenarios is slight. The probability of increasing by 0–25 mm in the north of the YRB is closer to 100% than 2021–2050 (Figure 11(b-1)). The same is true for the southwest part.

The probabilities of R_{10} changing concentrate more on a specific variation range than either *PY* or *PS*, overwhelmingly occurring to increase by 0–2 days and by 2–4 days. Compared with in 2021–2050 (Figure 11(c-1)), there are more significant and more recognizable probabilities of R_{10} increasing by 0–2 days in the northwest regions in 2061–2090 (Figure 11(c-2)). Like *PY*, the probability of R_{10} decreasing by more than two days in most regions of the YRB is 0.

The probability distribution in each VR of *PI* is similar to R_{10} . Most probability occurs in VR4 (*PI* increasing by 0–0.5 mm/day) and VR5 (*PI* increasing by 0.5–1 mm/day). Under both scenarios, the probability of *PI* increasing by 0–0.5 mm/day will be up to more than 0.8 in most regions in the west and north of the YRB, while the probabilities of *PI* increasing by 0.5–1 mm/day are between 20 and 40%. It mainly occurs in the east and south. The probabilities of other VRs are negligible.

5. DISCUSSION: THE CONSISTENCY BETWEEN THE DETERMINISTIC AND PROBABILISTIC PROJECTION

There is an apparent consistency between the deterministic projection in Section 4.3 and the probabilistic projection in Section 4.4, mainly manifested in spatial distribution and variation. In the deterministic projection in Figure 10(a-1), the increase of the *PY* of regions such as the southeast of Subregion 1 and Subregion 2 and Subregions 5–7 was significantly higher than that of other regions. In the probabilistic projection in Figure 11(a-1), the occurrences of VR6 (increased by more than 50 mm) in these regions were relatively high. However, as shown in Figure 10(a-1), the increases in the northwest of Subregions 1–2, Subregion 3, and Subregion 4 are mostly less than 50 mm. The variations of these regions are more likely to be between 0–25 mm or 25–50 mm, as shown in Figure 11(a-1).

We can also see that the future variations of the southeast of Subregions 1–2 in Figure 10(a-1) are between 50 and 100 mm, and the probability of increasing more than 50 mm in the future in these regions in Figure 11(a-1) is close to 1. Similarly, in the northeast of Subregion 5 and Subregion 6, the future variations of these regions are between 50 and 100 mm (Figure 10(a-1)), and the probabilistic projections indicate that these regions have the greatest probability of increasing by more than 50 mm in the future (Figure 11(a-1)). This shows consistency in the future variation. The projections of the two periods in the future have these consistencies in both aspects (Figures 10(a-2) and 11(a-2)).

In this study, we also found that the accuracy of the projection of a precipitation event (VR) is related to its occurrences (Figures 8 and 9). This leads us to think that this may be the best method to define precipitation events for each precipitation index to obtain more comprehensive and specific accuracy information of probability estimates. At present, we have considered four typical precipitation indexes. If we consider more different precipitation indexes, we can understand the ability of the probability estimation method to project other precipitation characteristics. This study is based on the data of five GCMs. If more models are applied, a higher estimation accuracy may be obtained. We will discuss the above issues in the subsequent work and constantly improve the probability estimation method of future precipitation.

6. CONCLUSION

This paper proposed an estimation method of future precipitation variation using the BMA method, constituting the deterministic and probabilistic projection of future precipitation. It was applied in the YRB. First, the spatial distribution of variations of four precipitation indexes (namely annual mean precipitation (*PY*), summer precipitation (*PS*), heavy rain index (R_{10}), and annual mean precipitation intensity (*PI*)) over the YRB was analyzed sequentially. Then, based on the precipitation projection and BMA weights of all selected models, the probability of variation in different ranges for all precipitation indexes at different times in the future was estimated spatially. Then, the credibility of the deterministic projection based on BMA was evaluated with E_D (mean of long-term annual relative simulation deviation) and V_D (variance of long-term annual relative simulation deviation). The deviation and accuracy evaluated the credibility of the probabilistic projection. The main conclusions are:

1. The BMA model performs well in reconstructing the precipitation conditions because of the lowest *MASD* (the mean of the absolute value of simulation deviation) in simulating the observed historical precipitation indexes. In addition, the relatively lower E_D (the mean of long-term annual relative simulation deviation) and V_D (the variance of long-term annual relative simulation deviation) in simulating precipitation situations under a given climate scenario give the BMA model more credibility than the individual models.
2. The probability estimation method proposed in this study can give sufficient probability information of different future precipitation variations. Among the four precipitation indexes, the accuracy between the estimated VR with the highest probability and the observed VR is 38.31–53.74%. The deviations in 81.93–94.70% of grids were no more than 1 level. Thus, the estimation accuracy of different precipitation events (VRs) varies greatly, and there was a positive correlation between the accuracy and the occurrence of events.
3. The precipitation situations indicated by four indexes of *PY*, *PS*, R_{10} , and *PI* have an increasing tendency in the whole YRB for a distinct higher probability of increasing with varying degrees rather than of decreasing, which suggests an increased risk of summer flooding in the future, compared to the near future. The increment of all the precipitation indexes will be even more significant in the distant future, as will the probability for a range of variation.
4. There is a good consistency between the deterministic projection and the probabilistic projection of the future precipitation in the YRB, which is reflected in the remarkable coincidence of spatial distribution and variation trend. With a better method to define precipitation events (classification method) and applying more and more reliable climate models, it is expected that more reliable precipitation projection can be obtained in the future.

ACKNOWLEDGEMENTS

This work was supported by grants from the National Key Research and Development Program of China (2018YFC1508104), the National Natural Science Foundation of China (52079079, 51779146, 51809174), the Natural Science Foundation of Jiangsu Province (BK20191129), and the Natural Science Foundation of Zhejiang Province (2021C03017). The authors are also grateful to the National Climate Center and China Meteorological Administration for providing data and the helpful comments by the editor and the reviewers. Han *et al.*'s early model evaluation work offered outstanding support for our research.

DECLARATION OF INTEREST STATEMENT

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

DATA AVAILABILITY STATEMENT

All relevant data are available from daily surface precipitation: <http://data.cma.cn/>. CMIP5: <https://esgf-node.llnl.gov/search/cmip5>.

REFERENCES

- Almazroui, M., Ashfaq, M., Islam, M. N., Rashid, I. U., Kamil, S., Abid, M. A., O'Brien, E., Ismail, M., Reboita, M. S., Sörensson, A. A., Arias, P. A., Alves, L. M., Tippett, M. K., Saeed, S., Haarsma, R., Doblas_Reyes, F. J., Saeed, F., Kucharski, F., Nadeem, I., Silva_Vidal, Y., Rivera, J. A., Ehsan, M. A., Martínez_Castro, D., Muñoz, Á. G., Ali, M. A., Coppola, E. & Sylla, M. B. 2021a *Assessment of CMIP6 performance and projected temperature and precipitation changes over South America*. *Earth Systems and Environment* **5**, 155–183. <https://doi.org/10.1007/s41748-021-00233-6>.
- Almazroui, M., Islam, M. N., Saeed, F., Saeed, S., Ismail, M., Ehsan, M. A., Diallo, I., O'Brien, E., Ashfaq, M., Martínez-Castro, D., Cavazos, T., Cerezo-Mota, R., Tippett, M. K., Gutowski, W. J. J., Alfaro, E. J., Hidalgo, H. G., Vichot-Llano, A., Campbell, J. D., Kamil, S., Rashid, I. U., Sylla, M. B., Stephenson, T., Taylor, M. & Barlow, M. 2021b *Projected changes in temperature and precipitation over the United States, Central America, and the Caribbean in CMIP6 GCMs*. *Earth Systems and Environment* **5**, 1–24. <https://doi.org/10.1007/s41748-021-00199-5>.
- Chen, H. P., Sun, J. Q. & Li, H. X. 2017 *Future changes in precipitation extremes over China using the NEX-GDDP high-resolution daily downscaled data-set*. *Atmospheric and Oceanic Science Letters* **10** (6), 403–410. <https://doi.org/10.1080/16742834.2017.1367625>.
- Chen, L., Chang, J. X., Wang, Y. M. & Zhu, Y. L. 2019 *Assessing runoff sensitivities to precipitation and temperature changes under global climate-change scenarios*. *Hydrology Research* **50** (1), 24–42. <https://doi.org/10.2166/nh.2018.192>.

- Dars, G. H., Najafi, M. R. & Qureshi, A. L. 2017 Assessing the impacts of climate change on future precipitation trends based on downscaled CMIP5 simulations data. *Mehran University Research Journal of Engineering and Technology* **36** (2), 385–394. <https://doi.org/10.22581/muet1982.1702.16>.
- Duan, Q. Y. & Phillips, T. J. 2010 Bayesian estimation of local signal and noise in multi-model simulations of climate change. *Journal of Geophysical Research: Atmospheres* **115** (18), 1–15. <https://doi.org/10.1029/2009JD013654>.
- Duan, Q. Y., Ajami, K. N., Gao, X. G. & Sorooshian, S. 2007 Multi-model ensemble hydrologic prediction using Bayesian model averaging. *Advances in Water Resources* **30** (5), 1371–1386. <https://doi.org/10.1016/j.advwatres.2006.11.014>.
- Fricker, T. E., Ferro, C. A. T. & Stephenson, D. B. 2013 Three recommendations for evaluating climate predictions. *Meteorological Applications* **20** (2), 246–255. <https://doi.org/10.1002/met.1409>.
- Guan, X. J., Liang, S. X. & Meng, Y. 2016 Evaluation of water resources comprehensive utilization efficiency in the Yellow River Basin. *Water Science and Technology: Water Supply* **16** (6), 1561–1570. <https://doi.org/10.2166/ws.2016.057>.
- Guo, J. H., Huang, G. H., Wang, X. Q., Li, Y. P. & Lin, Q. G. 2017 Investigating future precipitation changes over China through a high-resolution regional climate model ensemble. *Earth's Future* **5** (3), 1–19. <https://doi.org/10.1002/ef2.187>.
- Han, Z. Y., Tong, Y., Gao, X. J. & Xu, Y. 2018 Correction based on quantile mapping for temperature simulated by the RegCM4. *Climate Change Research* **14** (4), 331–340. (in Chinese) <https://doi.org/10.12006/j.issn.1673-1719.2017.156>.
- Hanittinan, P., Tachikawa, Y. & Ram-Indra, T. 2020 Projection of hydroclimate extreme indices over the Indochina region under climate change using a large single-model ensemble. *International Journal of Climatology* **40** (6), 2924–2952. <https://doi.org/10.1002/joc.6374>.
- Hoeting, J. A., Madigan, D., Raftery, A. E. & Volinsky, C. T. 1999 Bayesian averaging models. *Statistical Science* **14** (4), 382–417.
- Iqbal, M., Wen, J., Wang, S. P., Tian, H. & Adan, M. 2018 Variations of precipitation characteristics during the period 1960–2014 in the source region of the Yellow River, China. *Journal of Arid Land* **10** (3), 388–401. <https://doi.org/10.1007/s40333-018-0008-z>.
- Libanda, B. & Nkolola, N. B. 2019 Skill of CMIP5 models in simulating rainfall over Malawi. *Modeling Earth Systems and Environment* **5** (4), 1615–1626. <https://doi.org/10.1007/s40808-019-00611-0>.
- Liu, J. & Xie, Z. 2014 BMA probabilistic quantitative precipitation forecasting over the Huaihe Basin using TIGGE multi-model ensemble forecasts. *Monthly Weather Review* **142** (4), 1542–1555. <https://doi.org/10.1175/MWR-D-13-00031.1>.
- Liu, J., Du, H. B., Wu, Z. F., He, H. S., Wang, L. & Zong, S. W. 2017 Recent and future changes in the combination of annual temperature and precipitation throughout China. *International Journal of Climatology* **37** (2), 821–833. <https://doi.org/10.1002/joc.4742>.
- Liu, J., Jin, L. J., Huang, C. X. & Zhang, L. N. 2020 Variations of seasonal precipitation in the Yellow River Basin and its relationship to general circulation and SST. *Proceedings of the International Association of Hydrological Sciences* **383**, 5–12. <https://doi.org/10.5194/piahs-383-5-2020>.
- Lu, H. L., Jing, W. L., Zhao, J. C., Liu, X. J. & Huang, Z. 2014 Characteristics of the temporal variation in temperature and precipitation in China's Lower Yellow River Region. *Advances in Meteorology* **2014**. <https://doi.org/10.1155/2014/186823>.
- Meshram, S. G., Singh, S. K., Meshram, C., Ravinesh, C., Deo, R. C. & Ambade, B. 2018 Statistical evaluation of rainfall time series in concurrence with agriculture and water resources of Ken River basin, Central India (1901–2010). *Theoretical and Applied Climatology* **134**, 1231–1243. <https://doi.org/10.1007/s00704-017-2335-y>.
- Meshram, S. G., Kahya, E., Meshram, C., Ghorbani, M. A., Ambade, B. & Mirabbasi, R. 2020 Long-term temperature trend analysis associated with agriculture crops. *Theoretical and Applied Climatology* **140**, 1139–1159. <https://doi.org/10.1007/s00704-020-03137-z>.
- Miao, C. Y., Duan, Q. Y., Sun, Q. H., Huang, H., Kong, D. X. & Yang, T. T. 2014 Assessment of CMIP5 climate models and projected temperature changes over Northern Eurasia. *Environmental Research Letters* **9** (5), 055007. <https://doi.org/10.1088/1748-9326/9/5/055007>.
- Min, S. K. & Hense, A. 2006 A Bayesian approach to climate model evaluation and multi-model averaging with an application to global mean surface temperatures from IPCC AR4 coupled climate models. *Geophysical Research Letters* **33** (8), 1–5. <https://doi.org/10.1029/2006GL025779>.
- Mu, J. X., Liu, Q. C., Gabriel, H. F., Xu, D., Xu, J. D., Wu, C. L. & Ren, H. J. 2013 The impacts of climate change on water stress situations in the yellow river basin, China. *Irrigation and Drainage* **62** (5), 545–558. <https://doi.org/10.1002/ird.1784>.
- Notaro, V., Liuzzo, L. & Freni, G. 2016 A BMA analysis to assess the urbanization and climate change impact on urban watershed runoff. *Procedia Engineering* **154**, 868–876. <https://doi.org/10.1016/j.proeng.2016.07.461>.
- Qing, Y., Ma, Z. & Xu, B. 2017 Modulation of monthly precipitation patterns over East China by the Pacific Decadal Oscillation. *Climatic Change* **144**, 405–417. <https://doi.org/10.1007/s10584-016-1662-9>.
- Radermacher, C. & Tomassini, L. 2012 Thermodynamic causes for future trends in heavy precipitation over Europe based on an ensemble of regional climate model simulations. *Journal of Climate* **25** (21), 7669–7689. <https://doi.org/10.1175/JCLI-D-11-00304.1>.
- Raftery, A. E., Gneiting, T., Balabdaoui, F. & Polakowski, M. 2005 Using Bayesian model averaging to calibrate forecast ensembles. *Monthly Weather Review* **133** (5), 1155–1174. <https://doi.org/10.1175/MWR2906.1>.
- Tong, Y., Gao, X. J., Han, Z. Y. & Xu, Y. 2017 Bias correction of daily precipitation simulated by RegCM4 model over China. *Chinese Journal of Atmospheric Sciences* **41** (6), 1156–1166 (in Chinese). <https://doi.org/10.3878/j.issn.1006-9895.1704.16275>.
- Wang, G. Q., Zhang, J. Y., Jin, J. L., Weinberg, J., Bao, Z. X., Liu, C. S., Liu, Y. L., Yan, X. L., Song, X. M. & Zhai, R. 2017a Impacts of climate change on water resources in the Yellow River basin and identification of global adaptation strategies. *Mitigation and Adaptation Strategies for Global Change* **22** (1), 67–83. <https://doi.org/10.1007/s11027-015-9664-x>.

- Wang, W. G., Ding, Y. M., Shao, Q. X., Xu, J. Z., Jiao, X. Y., Luo, Y. F. & Yu, Z. B. 2017b Bayesian multi-model projection of irrigation requirement and water use efficiency in three typical rice plantation region of China based on CMIP5. *Agricultural and Forest Meteorology* **232**, 89–105. <https://doi.org/10.1016/j.agrformet.2016.08.008>.
- Wang, G. J., Pan, J., Shen, C. C., Li, S. J., Lu, J., Lou, D. & Hagan, D. F. T. 2018 Evaluation of evapotranspiration estimates in the Yellow River Basin against the water balance method. *Water* **10** (12), 1–18. <https://doi.org/10.3390/w10121884>.
- Wang, F., Wang, Z. M., Yang, H. B., Zhao, Y., Zhang, Z. Z., Li, Z. H. & Hussain, Z. 2019a Copula-based drought analysis using Standardized Precipitation Evapotranspiration Index: a case study in the Yellow river basin, China. *Water* **11** (6), 1–20. <https://doi.org/10.3390/w11061298>.
- Wang, Y. P., Zhao, W. W., Wang, S., Feng, X. M. & Liu, Y. X. 2019b Yellow River water rebalanced by human regulation. *Scientific Reports* **9** (1), 1–10. <https://doi.org/10.1038/s41598-019-46063-5>.
- Wu, X. S., Wang, Z. L., Zhou, X. W., Lai, C. G., Lina, W. X. & Chen, X. H. 2016 Observed changes in precipitation extremes across 11 basins in China during 1961–2013. *International Journal of Climatology* **36** (8), 2866–2885. <https://doi.org/10.1002/joc.4524>.
- Wu, H., Li, X. & Qian, H. 2017 Detection of anomalies and changes of rainfall in the Yellow River Basin, China, through two graphical methods. *Water* **10** (1). <https://doi.org/10.3390/w10010015>.
- Wu, S. Y., Wu, Y. J. & Wen, J. H. 2019 Future changes in precipitation characteristics in China. *International Journal of Climatology* **39** (8), 3558–3573. <https://doi.org/10.1002/joc.6038>.
- Yan, Z. Q., Zhou, Z. H., Liu, J. J., Han, Z. Y., Gao, G. & Jiang, X. T. 2020 Ensemble projection of runoff in a large-scale basin: modeling with a global BMA approach. *Water Resources Research* **56** (7), e2019WR026134. <https://doi.org/10.1029/2019WR026134>.
- Yin, J., Yan, D. H., Yang, Z. Y., Yuan, Z., Yuan, Y. & Zhang, C. 2016 Projection of extreme precipitation in the context of climate change in Huang-Huai-Hai region, China. *Journal of Earth System Science* **125** (2), 417–429. <https://doi.org/10.1007/s12040-016-0664-3>.
- Yin, Y. Y., Tang, Q. H., Liu, X. C. & Zhang, X. J. 2017 Water scarcity under various socio-economic pathways and its potential effects on food production in the Yellow River basin. *Hydrology and Earth System Sciences* **21** (2), 791–804. <https://doi.org/10.5194/hess-21-791-2017>.
- Zhang, Q., Peng, J. T., Singh, V. P., Li, J. F. & Chen, Y. D. 2014 Spatio-temporal variations of precipitation in arid and semi-arid regions of China: The Yellow River basin as a case study. *Global and Planetary Change* **114**, 38–49. <https://doi.org/10.1016/j.gloplacha.2014.01.005>.
- Zhang, Y., Xia, J. & She, D. X. 2019 Spatiotemporal variation and statistical characteristic of extreme precipitation in the middle reaches of the Yellow River Basin during 1960–2013. *Theoretical and Applied Climatology* **135** (1–2), 391–408. <https://doi.org/10.1007/s00704-018-2371-2>.
- Zhao, Y., Wang, M. Z., Li, J., Yang, X. J., Zhang, N. & Chen, H. 2019 Diurnal variations in summer precipitation over the Yellow River Basin. *Advances in Meteorology* **2019**, 1–10. <https://doi.org/10.1155/2019/2482656>.
- Zhou, W. C. & Han, Z. Y. 2018 Assessing CMIP5 climate simulations and objective selection of models over the Yellow River basin. *Journal of Meteorology and Environment* **34** (06), 42–55 (in Chinese). <https://doi.org/10.3969/j.issn.1673-503X.2018.06.005>.

First received 20 September 2021; accepted in revised form 21 February 2022. Available online 8 March 2022