



Forecast and uncertainty analysis of extreme precipitation in China from ensemble of multiple climate models

Peng Deng¹ · Jianting Zhu²

Received: 27 August 2020 / Accepted: 14 May 2021 / Published online: 24 May 2021
© The Author(s), under exclusive licence to Springer-Verlag GmbH Austria, part of Springer Nature 2021

Abstract

Global climate change is expected to have a major impact on the hydrological cycle. Understanding potential changes in future extreme precipitation is important to the planning of industrial and agricultural water use, flood control, and ecological environment protection. In this paper, we study the statistical distribution of extreme precipitation based on historical observation and various global climate models (GCMs), and predict the expected change and the associated uncertainty. The empirical frequency, generalized extreme value (GEV) distribution, and L-moment estimator algorithms are used to establish the statistical distribution relationships and the multi-model ensemble predictions are established by the Bayesian model averaging (BMA) method. This ensemble forecast takes advantage of multi-model synthesis, which is an effective measure to reduce the uncertainty of model selection in extreme precipitation forecasting. We have analyzed the relationships among extreme precipitation, return period, and precipitation durations for 6 representative cities in China. More significantly, the approach allows for establishing the uncertainty of extreme precipitation predictions. The empirical frequency from the historical data is all within the 90% confidence interval of the BMA ensemble. For the future predictions, the extreme precipitation intensities of various durations tend to become larger compared to the historic results. The extreme precipitation under the RCP8.5 scenario is greater than that under the RCP2.6 scenario. The developed approach not only effectively gives the extreme precipitation predictions, but also can be used to any other extreme hydrological events in future climate.

1 Introduction

Global climate change has had a major impact on the hydrological cycle in the past few decades, leading to large-scale fluctuations in the water resources system. Most studies suggest that global precipitation will increase as the global average surface temperature increases in the long term (Shao et al. 2016; Awange et al. 2019; Wang et al. 2019). Studying the potential change of extreme precipitation has great significance to planning industrial and agricultural water use, controlling floods, and protecting ecological environment (Du and Park 2019; Fang et al. 2019; Novoa et al. 2019).

The frequency distribution of extreme precipitation is an effective tool for assessing flood risks and meeting other

hydrological design objectives (Song et al. 2019; Zhao et al. 2019a, b). However, relationships based solely on historical data may not reflect future hydrological conditions (Yang et al. 2019), and new approaches are needed to incorporate expected changes and uncertainties into assessment, planning, and design. The Fifth Assessment Report of the Intergovernmental Panel on Climate Change (IPCC) pointed out that global warming is likely to further accelerate during the twenty-first century (IPCC 2013). Data from global climate models (GCMs) can provide a reference for future climate scenarios when addressing future climate impacts. Many studies have assessed the uncertainty of climate change impact analysis due to climate model selection (e.g., Qi et al. 2017; Graham et al. 2007; Minville et al. 2008). Chen et al. (2011) studied the hydrological impacts by combining the results from various GCMs, initial conditions, downscaling techniques, and hydrological model structures, and pointed out that the choice of GCM is consistently a major contributor to uncertainty. Hawkins and Sutton (2009, 2011) quantified the sources of uncertainty in regional precipitation changes using multi-model ensemble, and indicated that the climate model uncertainty is generally the dominant source of

✉ Peng Deng
dp@nuist.edu.cn

¹ School of Hydrology and Water Resources, Nanjing University of Information Science & Technology, Nanjing 210044, China

² Department of Civil and Architectural Engineering, University of Wyoming, Laramie, WY 82071, USA

uncertainty for longer lead times. Smith and Chandler (2010) pointed out that the ability of climate models to simulate past climates is a good indicator of their ability to model future climates in mid-high latitudes. Yuan et al. (2018) coupled the climate models with the hydrologic model to evaluate the impact of climate change on future extreme flood changes. Chen et al. (2017) studied the impacts of weighting climate models for hydro-meteorological climate change, and showed that uncertainty due to hydrological modeling is significantly smaller than that related to the choice of a climate model. Maraun and Widmann (2018) simulated summer mean precipitation at two locations in Norway by GCMs, and corrected the residual bias in terms of the observed and simulated climate change signals.

The traditional practice is to rely on a single climate model to implement predictions for the future projection. The number of GCMs developed by many research groups around the world is increasing quickly in recent decades, and the results of simulated climate variables differ widely among GCMs. No single model can outperform other models under all conditions. Different climate models have their own characteristics in describing different aspects of meteorological processes. Wootten et al. (2017) presented a method to partition and quantify the uncertainty in climate model ensembles that is attributable to downscaling, and suggested that overconfidence could be a serious problem in studies that employ a single set of downscaled GCMs. Multi-model ensembles have been used in various forecasting applications such as economic and weather forecasting (Christensen and Lettenmaier 2007; Wang and Robertson 2011; Zhao et al. 2019a, b). These multi-model methods generally combine individual models according to different weights to obtain a set of prediction results. The simplest is that all models have equal weights, or some regression analysis algorithms can be used to determine the different weights. These multi-model averaging methods yield better results than single-model predictions with large errors. However, the reliability of this method is not satisfactory, and its uncertainty is not well described (Wilby and Harris 2006; Knutti 2010; Sanderson and Knutti 2012).

To overcome the problem, the Bayesian theory has been used in multi-model averaging. For a monodrome forecasting application, Krzysztofowicz and Herr (2001) proposed the Bayesian processor output, which was applied to revise the prior probability and obtain the cumulative probability or probability density prediction. Raftery et al. (2005) used the Bayesian model averaging (BMA) to integrate the forecast results of different sources, and estimate the probability density function and weight of different members, and apply the BMA to the probabilistic precipitation forecasting (Sloughter et al. 2007). The BMA method is similar to other multi-model methods in that it uses the concept of a weighted average of the individual predictions from competing models, but differs

in that the BMA can provide a more reliable predictive uncertainty accounting for both between-model variances and within-model variances. Duan et al. (2007) used the BMA scheme to develop skillful and reliable probabilistic hydrologic predictions from multiple competing predictions made by several hydrologic models. Zhu et al. (2013) proposed an integrated approach to explore potential changes in intensity-duration-frequency relationships of rainfall based on the BMA.

To make informed planning and management adjustment with climate change, it is essential to study the future extreme precipitation. The ensemble prediction can take the strength of each individual model, which has been proven as an effective measure to reduce the uncertainty of model selection in precipitation forecasting. In this study, the multi-model ensemble prediction is made to predict the changes of extreme precipitation in representative climate locations in China with integrated uncertainty analysis as part of the comprehensive investigation. The multiple climate models serve as a useful tool to analyze the statistical distribution of extreme precipitation, return periods, and precipitation duration for future climate change. The study not only effectively provides the regional extreme precipitation predictions, but the general approach can also be applied to other extreme hydrological events under climate change scenarios.

2 Material and methods

2.1 Data sources and study locations

The historical and future precipitation data in a daily time step are available from the fifth phase of the Coupled Model Intercomparison Project (CMIP5), which is evolving and new models are being added continuously. We use seven typical GCMs from various countries in the world available from CMIP5 in this study as shown in Table 1. While the proposed approach can be applied to analyze any desired number of climate models, we only choose the seven representative GCMs from various countries so each individual model result can be distinguished and the differences can be highlighted easily. Therefore, the approach can be potentially applied to the GCMs in CMIP5 or the upcoming CMIP6. The standard observation of meteorological station in China generally started in 1951, while the historical data of CMIP5 generally stop in 2005. Thus, the time span of 55 years from 1951 to 2005 for historical period is used. Since the historical observations from meteorological stations have a long record of 55 years, the future scenarios are also analyzed using data of 55 years which is from 2021 to 2075. In the future scenarios, we use two climate change scenarios, Representative Concentration Pathways (RCP) 2.6 and RCP8.5, which means

Table 1 General information of GCMs on the CMIP5 used in the study

Model abbreviation	Research institute	Resolution (Lon×Lat)	Country
CNRM-CM5	Centre National de Recherches Météorologiques	256×128	France
GFDL-ESM2M	Geophysical Fluid Dynamics Laboratory	144×90	USA
HadGEM2-ES	Met Office Hadley Centre	192×145	Britain
IPSL-CM5A-LR	Institut Pierre Simon Laplace	96×96	France
MIROC-ESM-CHEM	Atmosphere and Ocean Research Institute	128×64	Japan
MPI-ESM-LR	Max Planck Institute for Meteorology	192×96	Germany
NorESM1-M	Norwegian Climate Centre	144×96	Norway

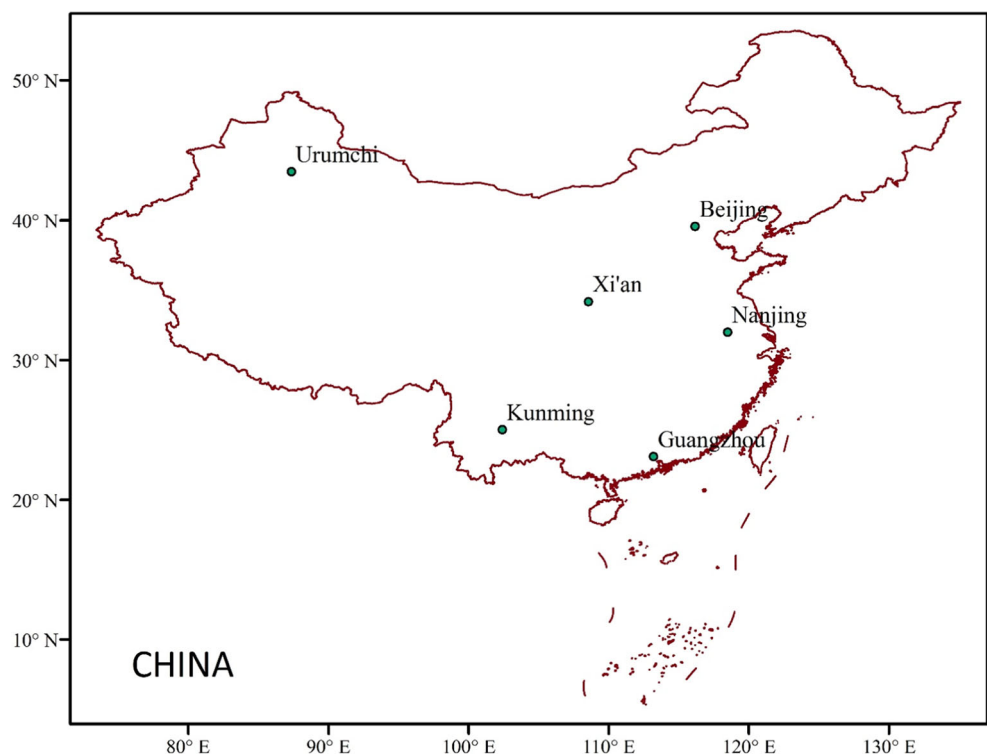
the radiated forcing achieved in 2100 is 2.6 W/m^2 and 8.5 W/m^2 , respectively.

Six locations (Beijing, Guangzhou, Kunming, Nanjing, Urumchi, Xi'an) from different representative climate regions in China are selected for this study as shown in Fig. 1. Beijing is the capital of China and has a typical warm temperate semi-humid continental monsoon climate. Urumqi is the most inland provincial capital of China farthest from the ocean and coastline, and it receives very low average annual precipitation of only 294 mm. Xi'an is a major central city in western China, and it has a semi-humid continental climate. Nanjing has a subtropical humid climate. Guangzhou has a maritime subtropical monsoon climate with abundant rainfall. Kunming is a northern and low latitude subtropical-plateau mountain monsoon climate with a mild climate all year around. The

daily historical precipitation data of 55-year span from 1951 to 2005 were obtained from the local meteorological stations.

2.2 Bias correction of GCM output

Bias correction is used for processing the raw output from GCMs and obtaining the point climate model data of each studied city from the climate grid where the city is located (Hay et al. 2000). As mentioned earlier, the GCM data of historical (1951~2005) and future (2021~2075) period and observed data of historical (1951~2005) period are obtained at a daily time step. For other x -day durations (2-day~5-day in this study), the x -day moving-window average from the available daily time step data can be calculated to generate the x -day duration precipitation intensity data. The annual

Fig. 1 The six study locations in different climate regions in China

maximum (AM) sampling is used to generate the extreme precipitation series. The AM allows the samples to be independently and identically distributed. For each studied city, the nearest neighbor is used to find the climate model grid where the study site is located and the GCM data of this grid are obtained. Then, the AM series of the GCM data in the historical period (H_{GCM}) and future period (F_{GCM}) can be obtained. The AM series of observed data in historical period of this study site is H_{obs} . Since the target in this study is extreme precipitation intensity, we use the correction factor of historical maxima to adjust the GCM data as

$$\begin{cases} \delta = \frac{\max(H_{obs})}{\max(H_{GCM})} \\ \widehat{H}_{GCM} = \delta \cdot H_{GCM} \\ \widehat{F}_{GCM} = \delta \cdot F_{GCM} \end{cases} \quad (1)$$

where $\max(\cdot)$ means the maximum of the data series, δ is the correction factor, and \widehat{H}_{GCM} and \widehat{F}_{GCM} are the corrected point GCM data for the historical and future periods, respectively. It ensures that the maximum values of the historical period between observed and climate model simulated are equal, which has a manner similar to quantile mapping (Wang and Chen 2014) based on the absolute maximum. The corrected outputs are time-synchronous with the GCMs.

2.3 Frequency analysis

2.3.1 Empirical frequency

The empirical frequency is often used as the basis for the frequency curve-fitting in hydrological variables. For a descending sorted AM series $X = \{x_1, x_2, \dots, x_n\}$, we use the mathematical expectation formula as the empirical probability,

$$P_m = \frac{m}{n+1} \quad (2)$$

where P_m is the empirical probability of the variable being equal to or greater than x_m , m is the rank order in the descending sorted sample, and n is the total sample size. The empirical frequency of historical observed data series (H_{obs}) is calculated in this study as the basis for assessing the agreement between the climate model results and observations.

2.3.2 Generalized extreme value (GEV) distribution

The generalized extreme value (GEV) distribution is a theoretical probability curve commonly used in the study of extreme value problems. The cumulative density function of GEV can be written as,

$$F(x) = \exp \left\{ - \left[1 - \frac{\kappa(x-\xi)}{\alpha} \right]^{1/\kappa} \right\} \quad (3)$$

where ξ is the location parameter, α is the scale parameter, and κ is the shape parameter.

The L-moments estimators are used to estimate the GEV parameters (Hosking and Wallis 1997). The L-moments estimators are more robust and unbiased than the traditional moment method. For the descending sorted series X , the first- to third-order L-moments can be calculated as follows,

$$\begin{cases} \lambda_1 = \beta_0 \\ \lambda_2 = 2\beta_1 - \beta_0 \\ \lambda_3 = 6\beta_2 - 6\beta_1 + \beta_0 \end{cases} \quad (4)$$

where $\beta_0 = \frac{1}{n} \sum_{i=1}^n x_i$, $\beta_1 = \frac{1}{n} \sum_{i=2}^n \frac{i-1}{n-1} x_i$, and $\beta_2 = \frac{1}{n} \sum_{i=3}^n \frac{(i-2)(i-1)}{(n-1)(n-2)} x_i$. Then, the parameters of GEV in terms of L-moments can be calculated (Zhu et al. 2013),

$$\begin{cases} c = \frac{2\lambda_2}{\lambda_3 + 3\lambda_2} - \frac{\ln(2)}{\ln(3)} \\ \kappa = \frac{7.8590c + 2.9554c^2}{\kappa\lambda_2} \\ \alpha = \frac{\Gamma(1 + \kappa)(1 - 2^{-\kappa})}{\alpha[\Gamma(1 + \kappa) - 1]} \\ \xi = \lambda_1 + \frac{\alpha[\Gamma(1 + \kappa) - 1]}{\kappa} \end{cases} \quad (5)$$

When the parameters are determined, the extreme value x of given return period (equal to $\frac{1}{1-F(x)}$) can be calculated based on Eq. 3. The GEV distributions are calculated for the historical observation data (H_{obs}) and each GCM data (\widehat{H}_{GCM} , \widehat{F}_{GCM}). The distribution results of the different GCMs are then integrated by the BMA algorithm presented below.

2.4 Ensemble forecast and uncertainty analysis

The BMA is a popular statistical scheme to infer a probabilistic prediction by several competing models. We use the BMA scheme for ensemble forecast and uncertainty analysis. Consider y to be the forecasted variable, series $D = \{y_1, y_2, \dots, y_T\}$ to be the observed data with data length T , and $M = \{M_1, M_2, \dots, M_K\}$ to be a set of K considered models, then the probability density function (PDF) of the BMA probabilistic prediction of y given the observed data D can be represented as,

$$p(y|D) = \sum_{k=1}^K p(M_k|D) \cdot p(y|M_k, D) \quad (6)$$

where $p(y|M_k, D)$ is the posterior probability of y given model prediction M_k and observed data set D , and $p(M_k|D)$ is posterior probability of model prediction M_k being the correct

prediction given the observed data D . This term reflects how well this particular ensemble member agrees with the observations. Then, the weight of model M_k can be expressed as $w_k = p(M_k|D)$. It should be noted that $\sum_{k=1}^K w_k = 1$.

The ensemble forecast of the variable y by BMA is expressed in the form of a PDF, which can be used to determine the probability prediction of the variable and its uncertainty. The posterior mean and variance of the BMA prediction can be expressed as,

$$E[y|D] = \sum_{k=1}^K p(M_k|D) \cdot E[p(y|M_k, D)] = \sum_{k=1}^K w_k \cdot f_k \tag{7}$$

$$Var[y|D] = \sum_{k=1}^K w_k \left(f_k - \sum_{i=1}^K w_i f_i \right)^2 + \sum_{k=1}^K w_k \sigma_k^2 \tag{8}$$

where σ_k^2 is the variance associated with model prediction f_k with respect to observation D . In essence, the expected BMA prediction is the average of individual predictions weighted by the likelihood that an individual model is correct given the observations. The BMA variance is essentially an uncertainty measure of the BMA prediction. It contains two components: the between-model variance and the within-model variance, as shown in the first and second terms of Eq. 8, respectively.

With proper estimate of w_k , σ_k , and conditional probability distribution $p(y|M_k, D)$, we can generate probabilistic predictions based on Eq. 6. Before presenting the BMA algorithm, it is assumed that the conditional probability distribution $p(y|M_k, D)$ is Gaussian for computational convenience. The GEV distribution describes the probability (or return period) of an extreme value, and the Gaussian distribution indicates the likelihood of the extreme value in its sample. For $k = 1, 2, \dots, K$, and $t = 1, 2, \dots, T$, and denoting $\theta = \{w_k, \sigma_k\}$, the log-likelihood function is used which can be approximated as

$$l(\theta) = \log \left[\sum_{k=1}^K w_k \cdot p_k(y|f_k, D) \right] \\ = \log \left[\sum_{k=1}^K \left(w_k \cdot \sum_{t=1}^T g(y_t|f_{k,t}, \sigma_k) \right) \right] \tag{9}$$

where $g(y_t|f_{k,t}, \sigma_k)$ means the probability of y_t of the Gaussian distribution with mean $f_{k,t}$ and standard deviation σ_k . An iterative procedure can be used to obtain the solution of θ . We use the expectation maximization (EM) algorithm for this purpose (Raftery et al. 2005; Duan et al. 2007). The initial θ^{Itr} ($Itr = 0$), where the superscript “ Itr ” denotes iteration, can be calculated as

$$\begin{cases} w_k^{Itr} = \frac{1}{K} \\ \sigma_k^{Itr} = \sqrt{\frac{1}{KT} \sum_{t=1}^T \left[\sum_{k=1}^K (y_t - f_{k,t})^2 \right]} \end{cases} \tag{10}$$

Then, the initial likelihood $l(\theta^{Itr})$ ($Itr = 0$) can be calculated by Eq. 9. The expectation step is executed by estimating the latent variable

$$z_{k,t}^{Itr} = \frac{g(y_t|f_{k,t}, \sigma_k^{Itr-1})}{\sum_{k=1}^K g(y_t|f_{k,t}, \sigma_k^{Itr-1})} \tag{11}$$

Then, the maximization step is executed to compute the weight and update the variance as

$$\begin{cases} w_k^{Itr} = \frac{1}{T} \sum_{t=1}^T z_{k,t}^{Itr} \\ \sigma_k^{Itr} = \sqrt{\frac{\sum_{t=1}^T z_{k,t}^{Itr} \cdot (y_t - f_{k,t})^2}{\sum_{t=1}^T z_{k,t}^{Itr}}} \end{cases} \tag{12}$$

Similarly, the likelihood $l(\theta^{Itr})$ is updated by Eq. 9. The iteration stops until $|l(\theta^{Itr}) - l(\theta^{Itr-1})|$ is less than or equal to a pre-specified tolerance level (0.001 in this study). By this iteration algorithm, the proper w_k , σ_k can be estimated. Then, we can calculate the mean BMA predictions by Eq. 7, and estimate the relative uncertainties with the aid of Eq. 8. For the 90% confidence, the upper and lower limits of the confidence interval can be calculated as $[E - z \cdot \sqrt{Var}, E + z \cdot \sqrt{Var}]$ with $z = 1.6449$. We use this BMA algorithm to integrate the GCM results for the prediction of extreme precipitations.

3 Results and discussion

3.1 Model weights and frequency analysis

To examine the effects of bias correction for GCM output, the comparison of GCMs pre- and post-bias correction with observations is shown in Fig. 2. It can be seen that the raw outputs of GCMs for extreme precipitation are often smaller than observations. After the bias correction, the maximum value of each model output is equal to the observed value, and also the corrected outputs are more similar to observations in its distribution. There are differences between the corrected output and the observation when using different GCMs. Therefore, it is necessary to analyze the specific distribution of the model output and estimate the weights of them. Compared to the raw output of GCMs, the GCM data after bias correction can in general reflect the extreme precipitation range in the long time span.

The historical observed and GCM data are used to develop the BMA predictions of extreme precipitation intensity. The GEV distribution of each GCM was calculated, and then their ensemble is calculated using the BMA scheme. First, we calculate the extreme precipitation of 1-day duration. Figure 3

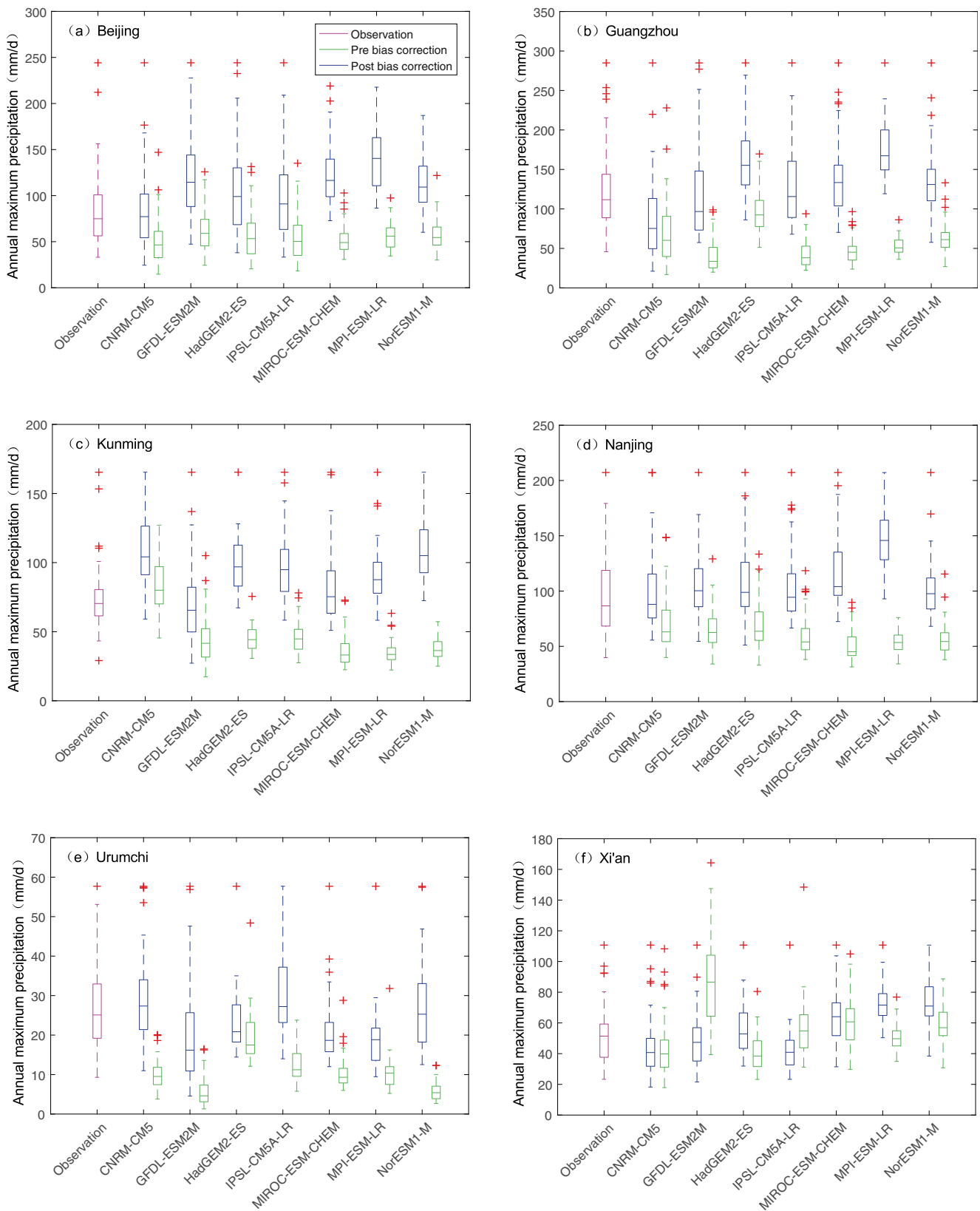


Fig. 2 Pre- and post-bias correction of GCM output compared to the observations at location (a–f)

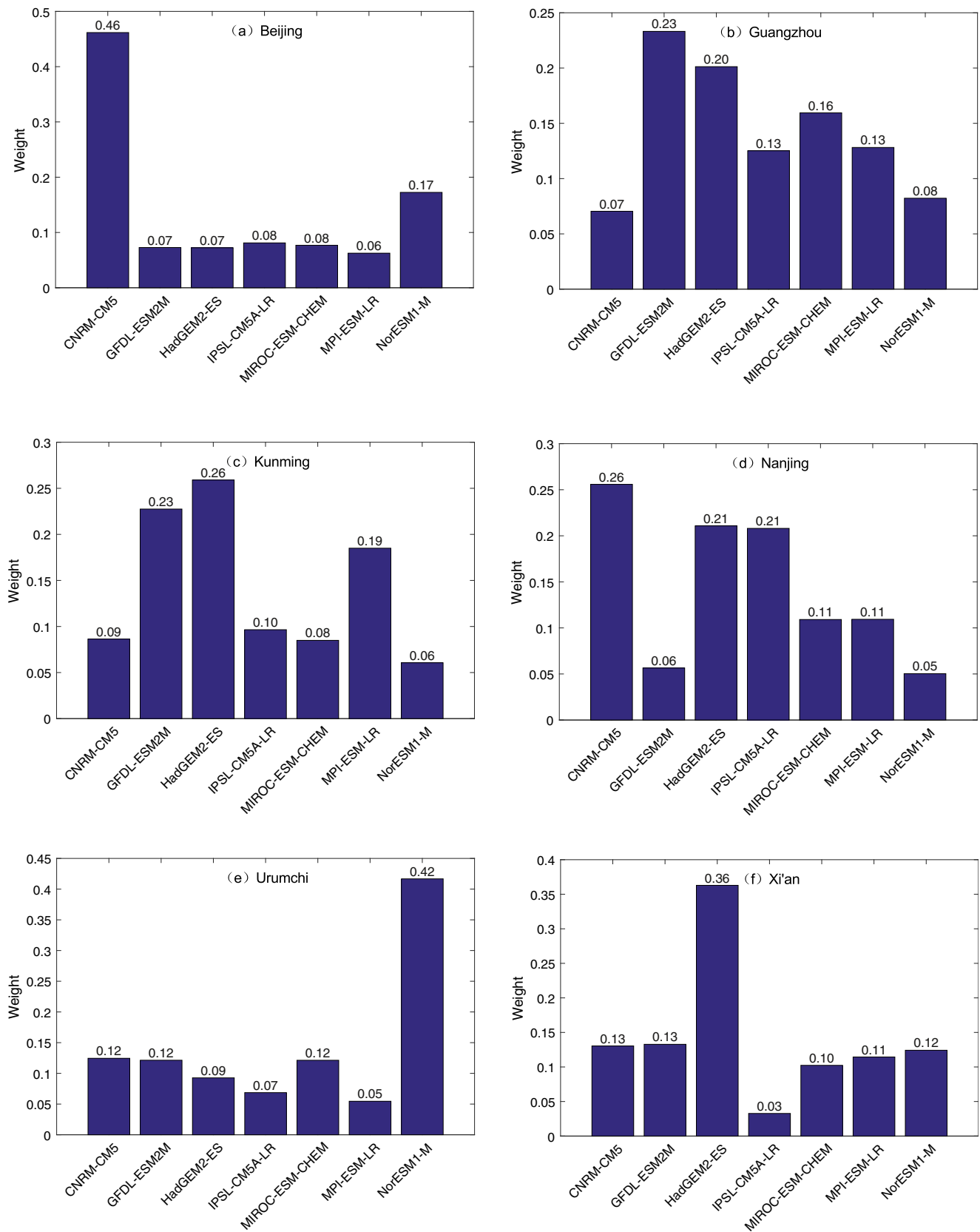


Fig. 3 Model weights of the seven GCMs after bias correction in relation to location (a–f)

shows the model weights of the seven GCMs after bias correction for the six climate locations. The performance of individual GCM varies depending on location, as reflected by the variations of weights among the GCMs. For Beijing, CNRM-CM5 receives the highest weight of 0.46, and NorESM1-M has the second highest weight of 0.17, while other four models have negligible weights of less than 0.1. For Guangzhou, GFDL-ESM2M and HadGEM2-ES have medium weights of 0.23 and 0.20, respectively, while the other models all have low weights of less than 0.2. For Kunming, GFDL-ESM2M, HadGEM2-ES, and MPI-ESM-LR have relatively higher weights, while the other four models have low weights of less than 0.1. For Nanjing, CNRM-CM5, HadGEM2-ES, and IPSL-CM5A-LR have medium weights of more than 0.2, which are higher than the other four models. For Urumchi, NorESM1-M receives a high weight of 0.42, while the other six models have low weights of less than 0.15. For Xi'an, HadGEM2-ES receives a high weight of 0.36, while the other models have low weights of less than 0.15. Generally, no particular single model is superior to other models for all locations, which also illustrates the significance and necessity of the BMA ensemble strategy.

Figure 4 shows the frequency analysis results of historical period. The x-axis of the figure uses the return period to represent the probability, because the reciprocal of the return period indicates the probability of the extreme event exceedance. The figure shows the empirical frequency (Eq. 2), GEV distribution (Eq. 3) of the observed historical AM series, and the BMA ensemble results from Eq. 7 and the 90% confidence interval based on the determined variance from Eq. 8. In general, the BMA predictions are consistent with the observations except for Beijing and Kunming. Because of the large singular values in the observations of the two stations, the BMA results over-estimate the observations. For Nanjing, the agreement between the BMA expectation and the observed GEV is the best, and the range of confidence interval is the smallest. For Beijing and Kunming, although there are a few outliers, both of these outliers are still within the 90% confidence interval. This is an advantage of the BMA probabilistic forecasting compared to the traditional single-valued curve prediction. For Guangzhou, Urumchi, and Xi'an, the BMA expectation curves also agree with the observed GEV curves well with the 90% confidence interval all embracing the observed historical extreme precipitation values. Therefore, the BMA schemes are reliable and useful for the ensemble predictions.

3.2 Future predictions and uncertainty assessment

After determining the BMA model based on the historical observed data, the future extreme precipitation can be predicted using the GCM data for the future period. Figure 5 shows the comparison of BMA ensemble results for future period with the historical distributions. We

consider the two future climate change scenarios of RCP2.6 and RCP8.5. We find that the average expected extreme precipitation by GCMs under the RCP8.5 scenario is higher than that of RCP2.6 for each location and any return period. For the RCP2.6 scenario, the extreme precipitation is also slightly higher than that of historical scenario with the exception of Beijing and Urumchi. The figures also show the 90% confidence interval for the predicted future extreme precipitation under the RCP2.6 and RCP8.5 scenarios. The confidence interval ranges under the RCP2.6 and RCP8.5 scenarios vary with the locations. For Beijing, the uncertainty of the RCP8.5 is smaller than that of the RCP2.6, while the opposite is true for Xi'an that the uncertainty of the RCP8.5 is significantly greater than that of RCP2.6. For Guangzhou and Kunming, the confidence interval of RCP2.6 is smaller than that of RCP8.5. For Nanjing and Urumchi, the confidence interval for the two scenarios is almost the same. The results indicate the necessity of the uncertainty analysis of BMA, since the uncertainty of climate model predictions is different for different locations and return periods.

3.3 Results for other precipitation durations

Next we analyze the extreme precipitation intensity of 2- to 5-day durations. The BMA approach is the same but used for the 2- to 5-day duration precipitation data, which are calculated using the moving-window average method. For each precipitation duration data, we separately perform the BMA. Figure 6 shows the weights of the GCMs after bias correction for the different precipitation durations. Comparing with Fig. 3, it can be found that there is no obvious pattern for the change of model weights under different precipitation durations. For Beijing, while CNRM-CM5 receives the highest weight for the 1-day duration, it has low weights for the 2- to 5-day durations. For Guangzhou, GFDL-ESM2M has high weight for the 1-day duration, but HadGEM2-ES has high weight for the 4-day and MPI-ESM-LR for the 5-day durations, respectively. For Kunming, HadGEM2-ES receives the highest weights for the 1-day and 4-day durations, but not for the other durations. For Nanjing, GFDL-ESM2M has a low weight for the 1-day duration, but it has a high weight for the 5-day duration, and MIROC-ESM-CHEM receives the highest weight for the 3-day duration. For Urumchi, NorESM1-M all receives the highest weights except for the 3-day duration. For Xi'an, HadGEM2-ES receives the highest weights for the 1-day and 2-day duration, but not for the other durations. In general, different models have different weights in various precipitation durations. No single model is superior for all the considered durations, indicating that the performance of climate models in

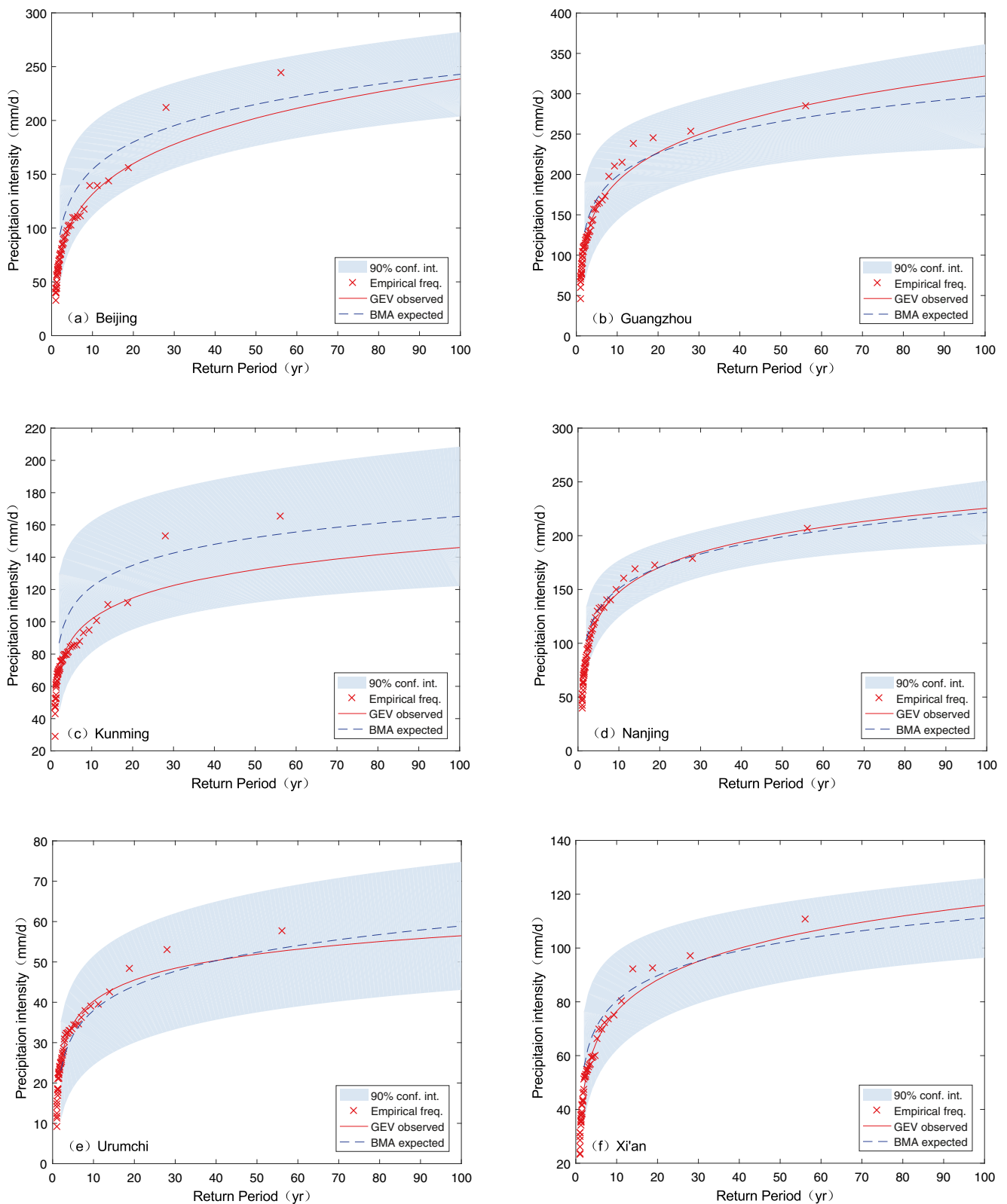


Fig. 4 The empirical frequency (freq.), GEV observed distribution, and BMA ensemble results of expected distribution and 90% confidence interval (conf. int.) for the historical period at location (a–f)

extreme precipitation prediction is also related to the durations.

Figure 7 shows the distribution of extreme precipitation over the historical period of different precipitation

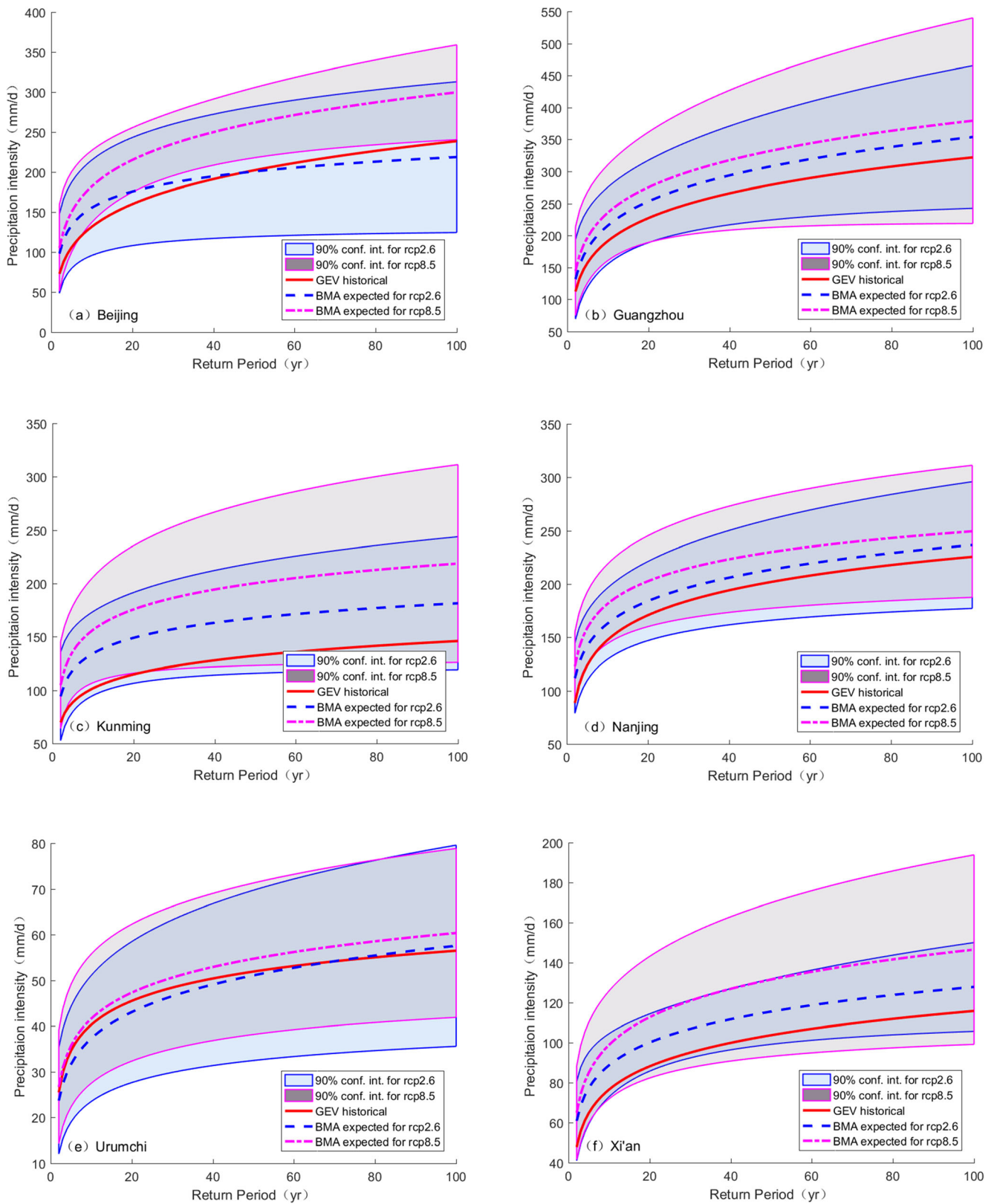


Fig. 5 The historical GEV distribution and BMA ensemble results of expected distribution and 90% confidence interval (conf. int.) for future period under the RCP2.6 and RCP8.5 scenarios at location (a–f)

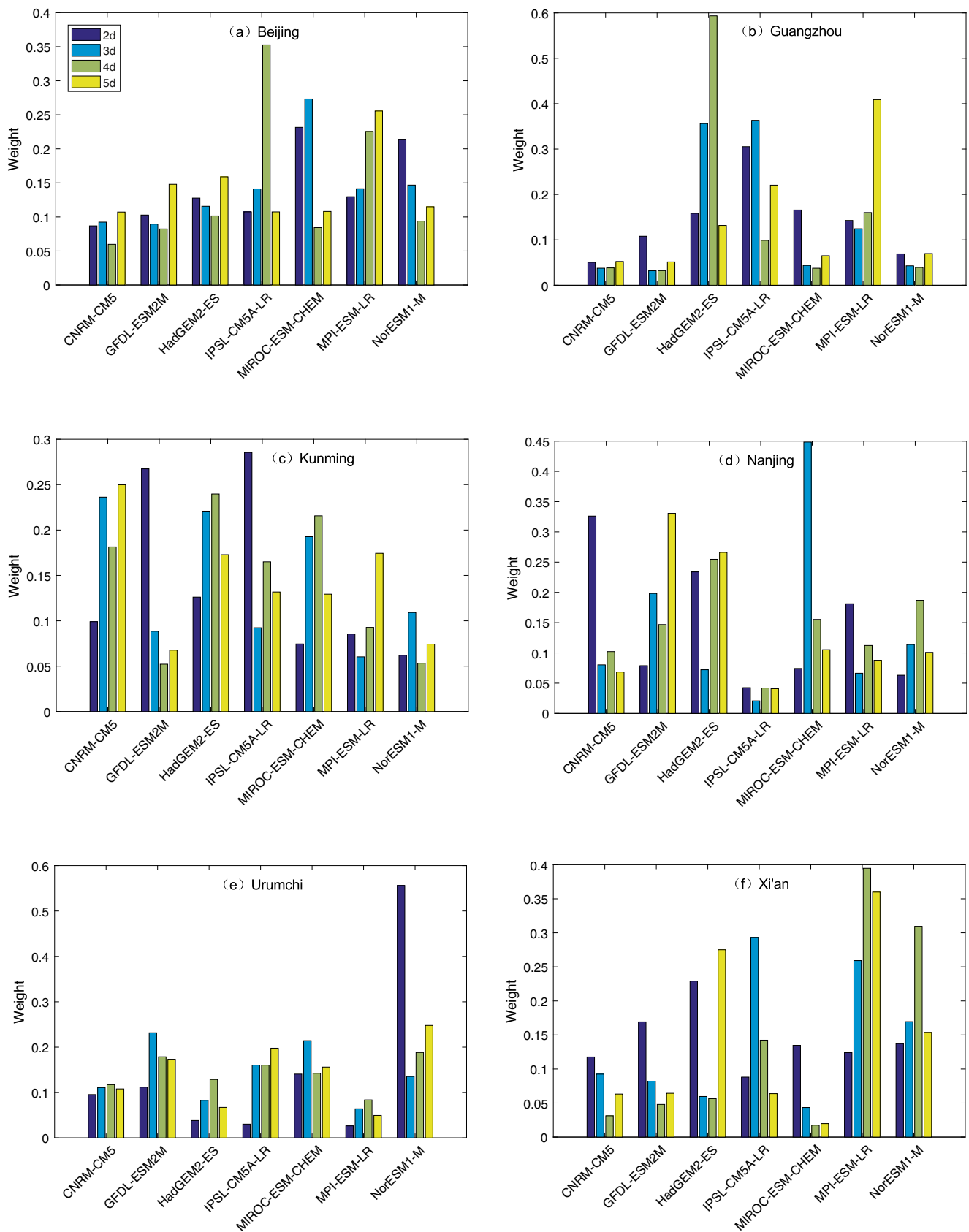


Fig. 6 Model weights for the 2- to 5-day (2d–5d) precipitation durations of the seven GCMs after bias correction in relation to location (a–f)

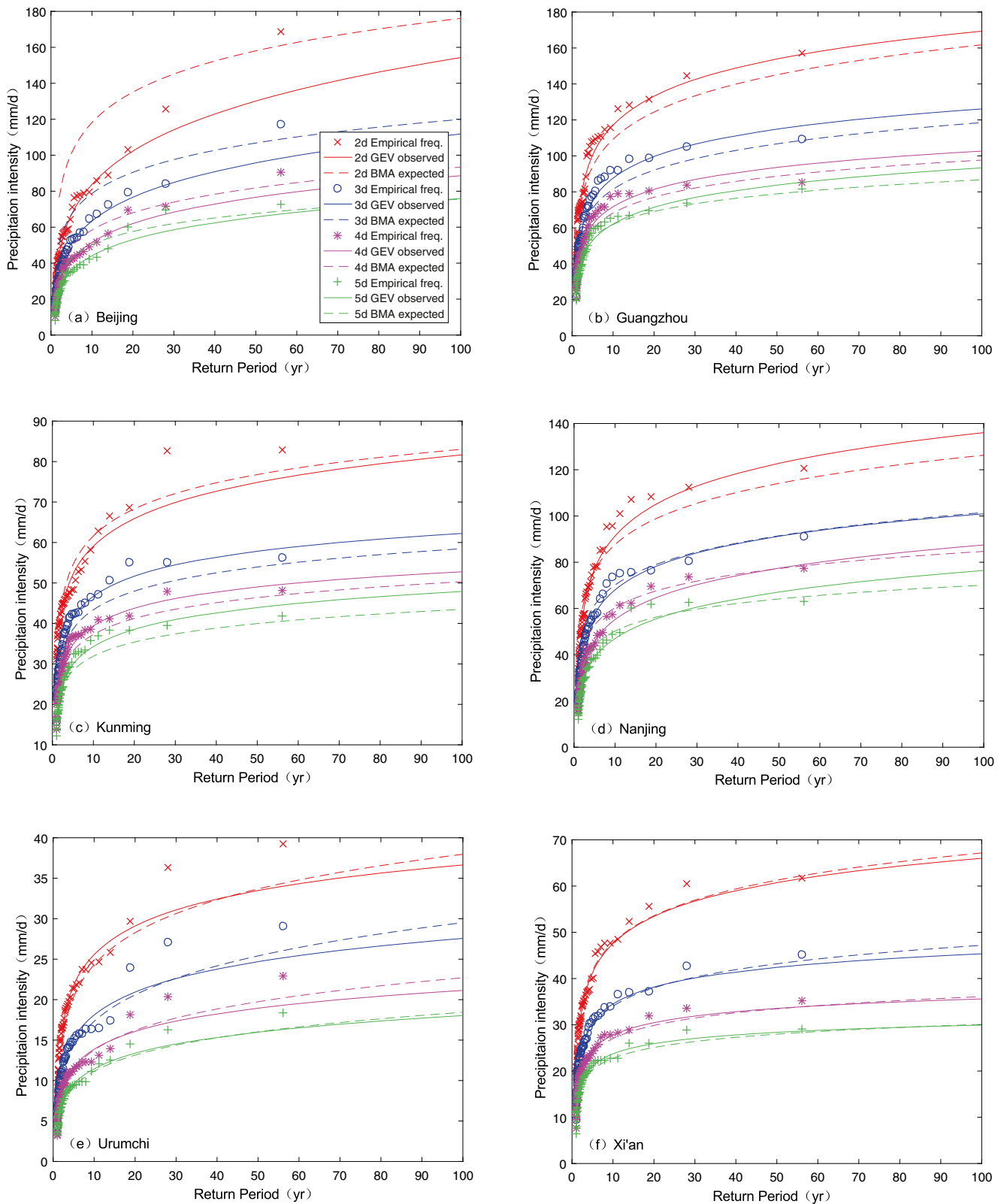


Fig. 7 The empirical frequency (freq.), GEV observed distribution, and BMA ensemble expected distribution of the 2- to 5-day (2d–5d) precipitation durations for the historical period at location (a–f)

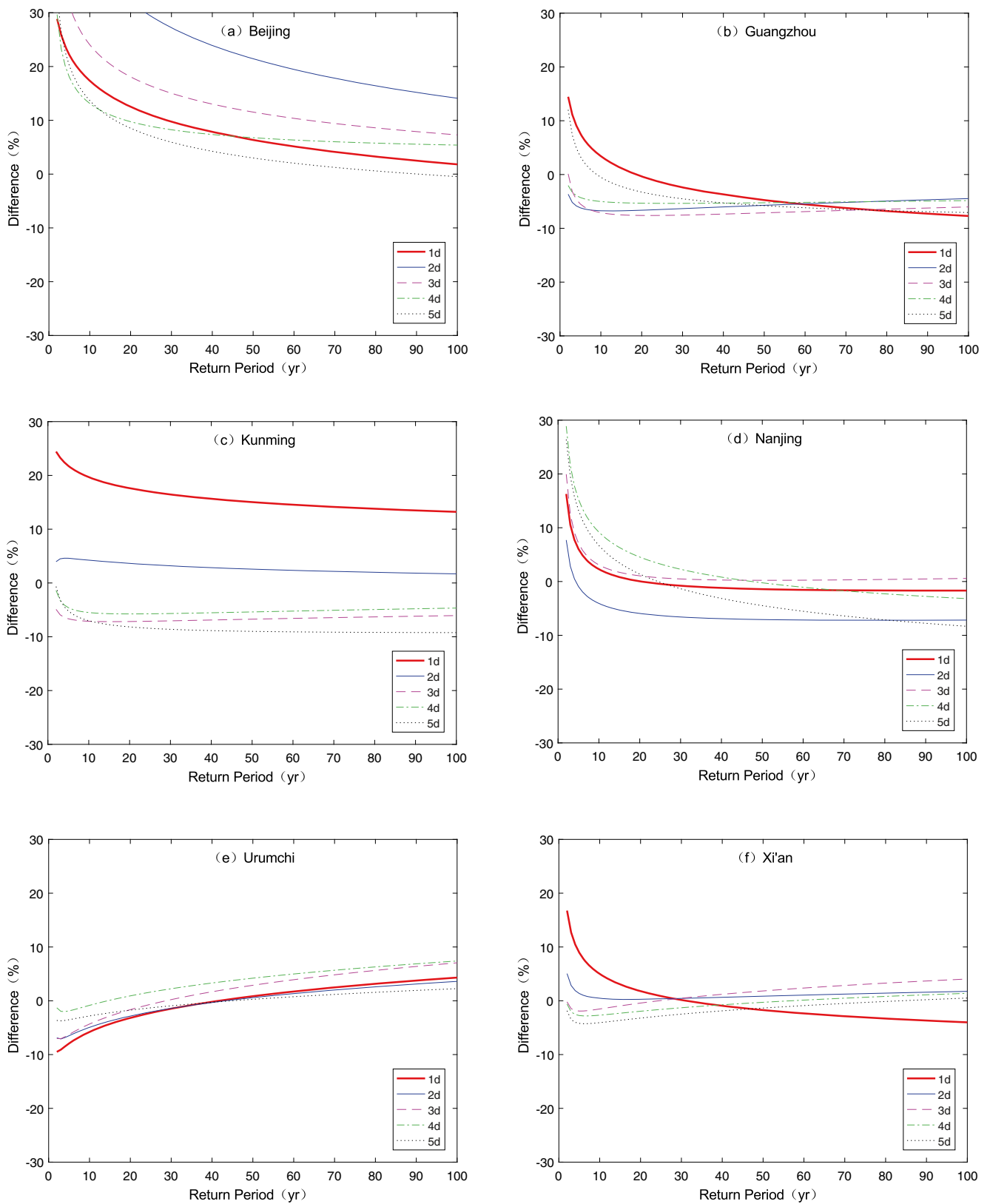


Fig. 8 Relative difference of precipitation intensity for various return periods and durations between the observed and BMA expected results for the historical period at location (a–f)

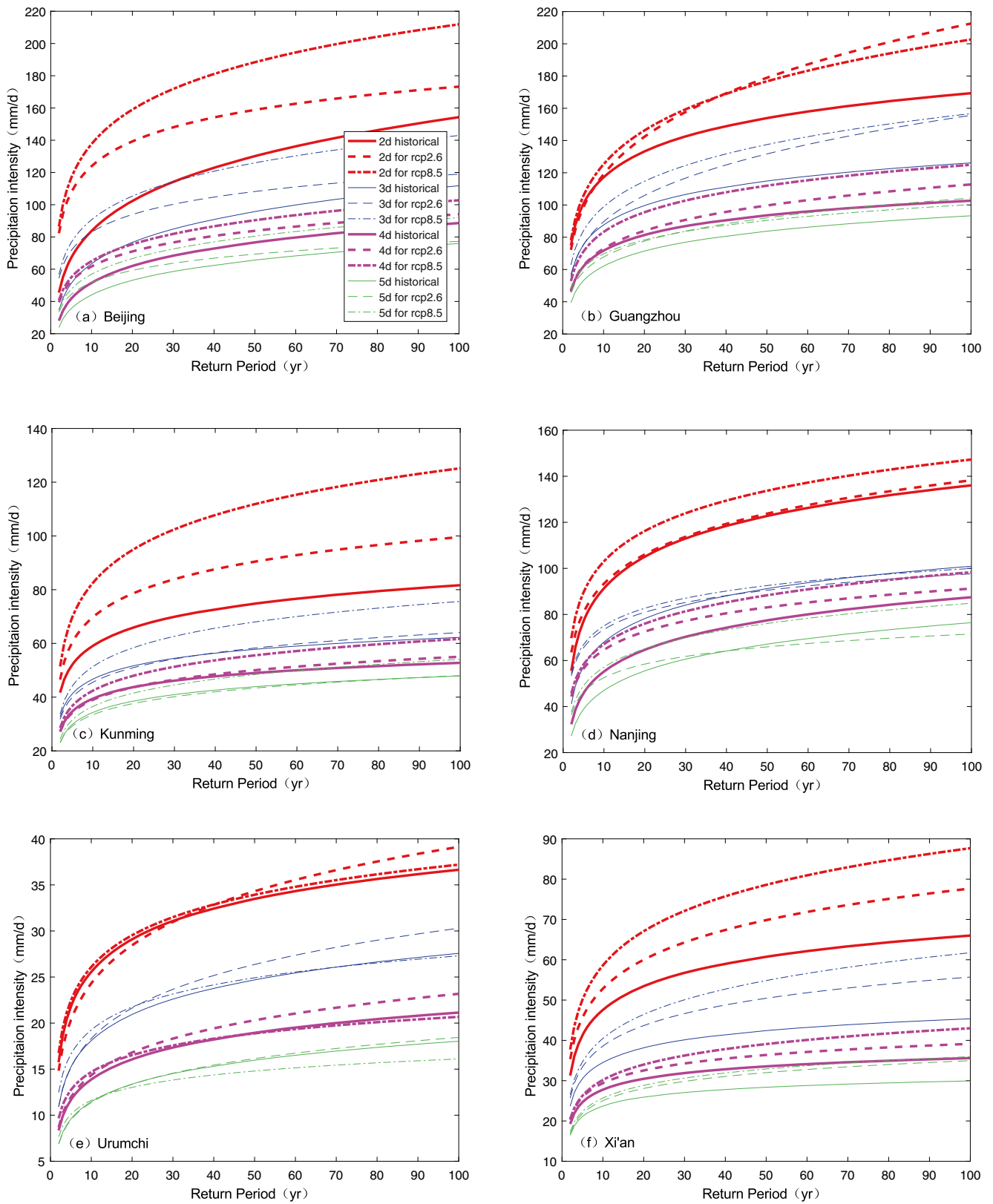


Fig. 9 The historical GEV distribution and BMA ensemble expected predictions for the future period under the RCP2.6 and RCP8.5 scenarios at location (a–f)

durations. The results in the figure include the empirical probability and the GEV distribution of the measured data, and the expected results from the BMA ensemble. It can be seen from the figure that the agreement between the BMA expected results and the measured probability distribution also varies with the duration. For Beijing, the 2-day duration results deviate more significantly than the 5-day duration results. For Guangzhou, the BMA expected results are all smaller than the observed results for the 2- to 5-day durations. For Kunming, the BMA results of 3- to 5-day durations are smaller than the observed results, but the 2-day results are slightly larger than observed ones. For Nanjing, the agreement of the 3-day duration results is the best, while the 2-day duration results deviate the most. For Urumchi and Xi'an, the BMA expected results agree well with the observed results for the 2- to 5-day durations. The variability of the BMA results for different durations in agreement with the observed results signifies the necessity of performing BMA for desirable durations separately for the extreme precipitation predictions.

Based on the results of Figs. 4 and 7, we can estimate the difference between the runs based on observed data and GCM data. Figure 8 shows the relative percent difference [$100 \times (I_{GCM} - I_{obs}) / I_{obs}$] where I_{GCM} is the precipitation intensity from the BMA ensemble while I_{obs} is the precipitation intensity from observed data, for various return periods and durations between the observed and BMA expected results for the historical period. In general, the relative difference in each case tends to be smaller as the return period becomes larger. For example, the difference is within the range between -9.23 and 14.11% for the 100-year return period. In particular, the differences are relatively large and positive in Beijing and 1-day duration in Kunming, which indicates over-estimation by the GCMs. In these two cities, a few large singular values of intensity from the historical data seen in Fig. 4 cause the BMA ensemble predictions to move toward large intensities to accommodate these large intensities and therefore to over-predict overall.

Based on the estimated BMA parameters for each duration calculated in the historical period, it is possible to predict the distribution of extreme precipitation for the same duration in the future climate. Figure 9 shows the distribution of extreme precipitation of different durations under the future RCP2.6 and RCP8.5 scenarios for each location. For Beijing, the extreme precipitation of RCP8.5 is higher than that of RCP2.6 and historical one for all durations considered. For Guangzhou, the results of RCP8.5 are also higher than that of RCP2.6 with the exception of the low return period of 2-day results. For Kunming and Nanjing, the results also show the higher extreme precipitation for the future scenarios, but the magnitude of increase tends to become smaller when the precipitation duration is longer. For Urumchi, the increase of future extreme precipitation is smaller. For Xi'an, the results

of RCP8.5 are higher than those of RCP2.6 and historical one. In general, the extreme precipitation under the future scenarios of both RCP2.6 and RCP8.5 is higher than the historical extreme precipitation, and the extreme precipitation of RCP8.5 is higher than that of RCP2.6.

3.4 Uncertainty analysis for other precipitation durations

In addition to the expected precipitation intensity from the BMA predictions, the confidence interval of the predictions is also determined for the other durations. The results of 90% confidence interval for the 100-year return period (exceedance probability of 1%) are shown in Fig. 10. It should be noted that the uncertainty ranges vary with return periods numerically. While the results in Fig. 10 are only for the 100-year return period, they also reflect the uncertainty of other return periods to a certain extent. For Beijing, the uncertainty range decreases with the decrease of precipitation intensity value under the RCP2.6 scenario, but the trend of this reduction is not obvious under the RCP8.5 scenario. For Kunming, Nanjing, and Xi'an, the uncertainty range of RCP8.5 results is usually larger than that of RCP2.6, but for Xi'an the uncertainty range of RCP8.5 is smaller than that of RCP2.6. There is no obvious pattern about the range of the uncertainty interval, which is related to the results of various climate models.

Similarly, the results of a 50-year return period (exceedance probability of 2%) with 90% confidence interval for the future period are shown in Fig. 11. It should be noted that the future expected results of the BMA ensemble for 2~100-year return periods can be found in Figs. 5 and 9. The results of two typical return periods 50 and 100 years that are shown with confidence interval are plotted in Figs. 10 and 11. Compared to the 100-year return period, the 50-year return period results demonstrate a similar trend but the intensity values and confidence interval ranges are both smaller. The future extreme precipitation predictions and confidence intervals of other specified durations and return periods by the BMA ensemble can be similarly determined.

In general, the BMA ensemble agrees well with the observations. However, caveats of the BMA ensemble should be noted. Sanderson et al. (2015a, b) pointed out the fact that models developed by different groups may be based on similar ideas and codes and therefore biases, and proposed a method to combine model results into single or multivariate distributions that are more robust to the inclusion of models with a large degree of interdependency. Knutti et al. (2017) argued that the growing number of models with different characteristics and considerable interdependence finally justifies abandoning strict model democracy, and provided guidance on when and how this can be achieved robustly. Massoud et al. (2019) tested a number of model weighting strategies that incorporate skill and interdependence for

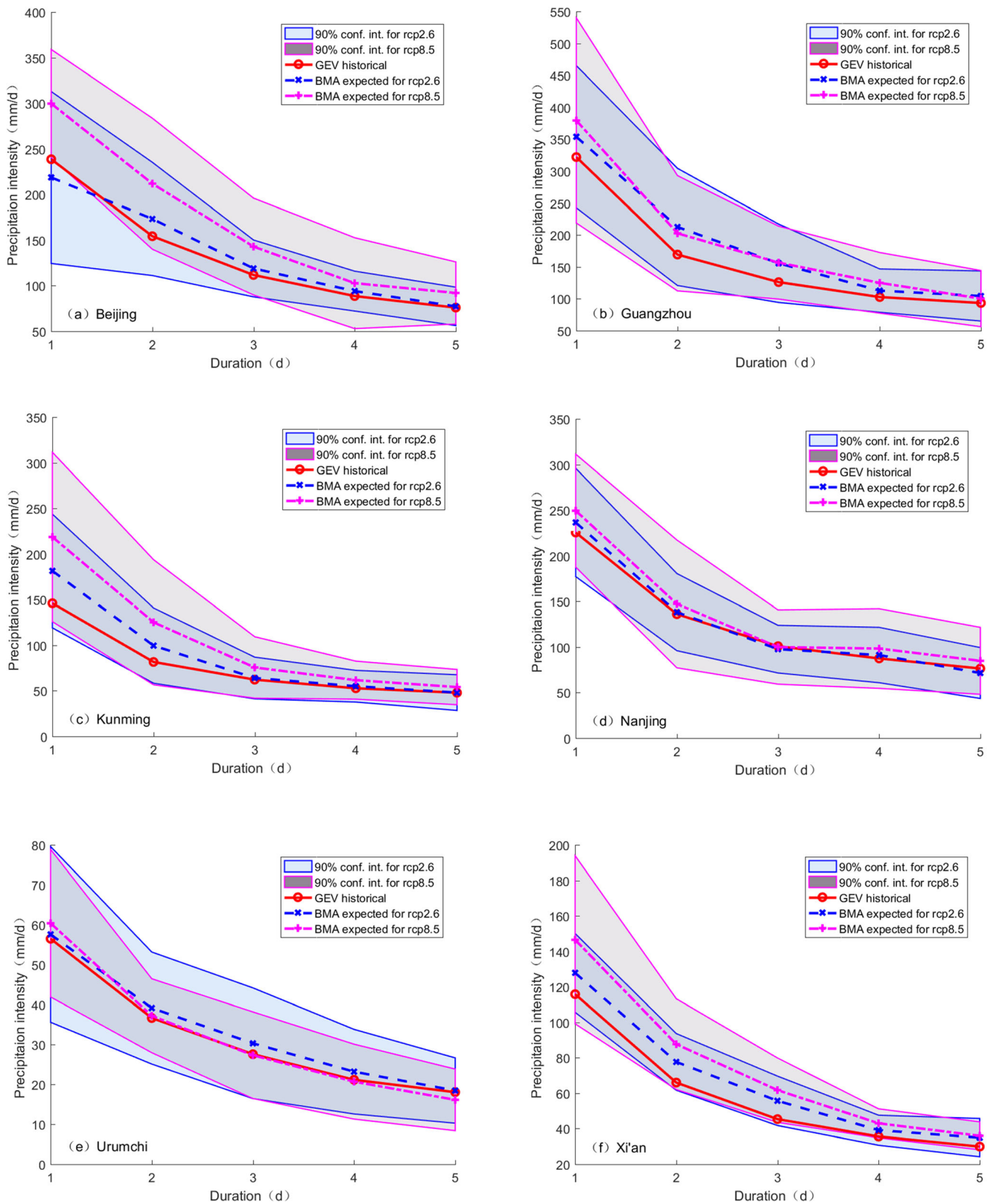


Fig. 10 The extreme precipitation and 90% confidence interval (conf. int.) of the historical and future scenarios under the RCP2.6 and RCP8.5 scenarios for the 100-year return period at location (a–f)

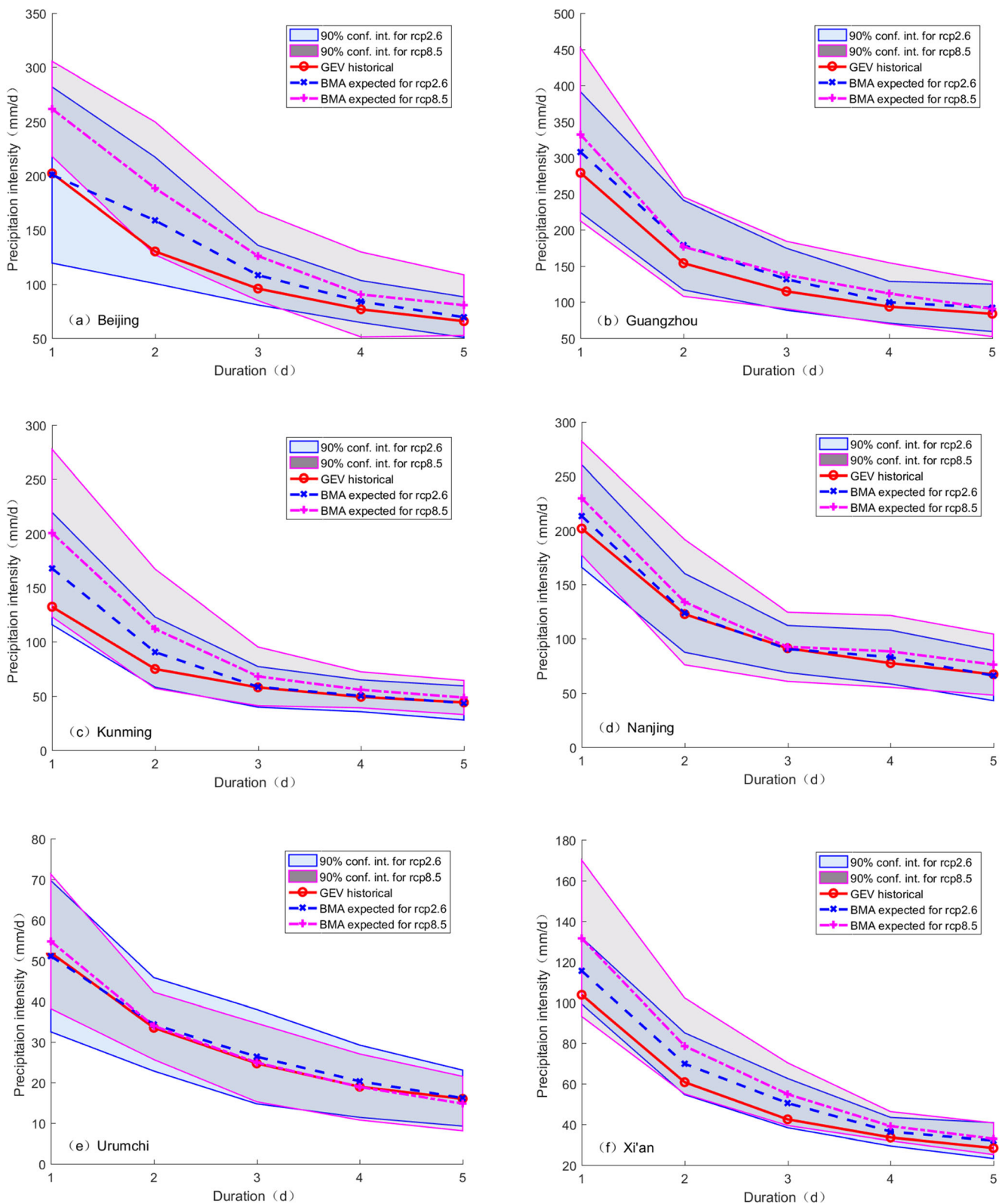


Fig. 11 The extreme precipitation and 90% confidence interval (conf. int.) of the historical and future scenarios under the RCP2.6 and RCP8.5 scenarios for the 50-year return period at location (a–f)

atmospheric rivers, and concluded that those weighting strategies produce future change estimates with lower

uncertainties than the ensemble mean approach, especially for the BMA method.

In this study, we used an approach of bias correction that gives the maximum precipitation the sole emphasis to correct climate model bias since the focus of this study is to predict extreme precipitation statistics. While the extreme statistics is about precipitation intensity in high percentiles not just the maximum intensity, it should be noted that the time span of climate models is 55 years. The approach could be justified given that the return period of interest is from 2 to 100 years. The generalized extreme value (GEV) distribution used in this study is to determine the precipitation intensities with the given return periods that could be longer or shorter than 55 years. Therefore, we would expect the intensities of long return periods could be under-estimated while those of short return periods over-estimated. The model weight of an individual model in the ensemble is strictly developed based on its likelihood. If some models in the ensemble are highly interdependent, they might be given similar weights and therefore the ensemble prediction may be highly skewed toward these models. One way to deal with the model interdependence issue might be to assign different priors for the individual models to account for the interdependence based on comprehensive analysis of these models before they are integrated into the BMA. To quantify the model interdependency, the among-model uncertainty and within-model uncertainty of the BMA ensemble could be separately analyzed. These considerations deserve further studies.

4 Concluding remarks

In this study, we present a BMA approach in ensemble prediction of extreme precipitation in locations of China based on a variety of CMIP5 climate models. The BMA is developed for the GEV distribution of historical measured data, which can predict the extreme precipitation in future climate scenarios and the uncertainty of the predicted results.

The BMA is implemented for prediction of extreme precipitation intensities for various durations and return periods. There is no obvious pattern in the weight of the GCMs. The expected results of BMA agree with the GEV distribution of historical extreme precipitation well, and the observed values are within the 90% confidence interval. In the future extreme precipitation forecast, the extreme precipitation intensity of each duration tends to become larger, and the extreme precipitation under the RCP8.5 scenario is greater than that under the RCP2.6 scenario. There is, however, no obvious relation of the uncertainty interval range under these two scenarios. This study develops an effective method for forecasting future extreme precipitation based on BMA and GCMs as well as quantifying the uncertainty of the extreme precipitation results so the future predictions are more comprehensive.

Author contribution Peng Deng designed and directed the project, wrote most of the article, proceeded the computation, and drew the figures; Jianting Zhu participated in writing and editing the manuscript.

Funding This work was partly supported by the Natural Science Foundation of Jiangsu Province of China (Grant No. BK20150922).

Availability of data and material Not applicable.

Code availability Not applicable.

Declarations

Ethics approval Not applicable.

Consent to participate Not applicable.

Consent for publication Not applicable.

Conflict of interest The authors declare no competing interests.

References

- Awange JL, Hu KX, Khaki M (2019) The newly merged satellite remotely sensed, gauge and reanalysis-based multi-source weighted-ensemble precipitation: evaluation over Australia and Africa (1981–2016). *Sci Total Environ* 670:448–465
- Chen J, Brissette FP, Poulin A et al (2011) Overall uncertainty study of the hydrological impacts of climate change for a Canadian watershed. *Water Resour Res* 47:W12509
- Chen J, Brissette FP, Lucas-Picher P, Caya D (2017) Impacts of weighting climate models for hydro-meteorological climate change studies. *J Hydrol* 549:534–546
- Christensen NS, Lettenmaier DP (2007) A multimodel ensemble approach to assessment of climate change impacts on the hydrology and water resources of the Colorado River Basin. *Hydrol Earth Syst Sci* 11:1417–1434
- Du JB, Park K (2019) Estuarine salinity recovery from an extreme precipitation event: hurricane Harvey in Galveston Bay. *Sci Total Environ* 670:1049–1059
- Duan Q, Ajami NK, Gao X, Sorooshian S (2007) Multi-model ensemble hydrologic prediction using Bayesian model averaging. *Adv Water Resour* 30(5):1371–1386
- Fang J, Yang W, Luan Y, du J, Lin A, Zhao L (2019) Evaluation of the TRMM 3B42 and GPM IMERG products for extreme precipitation analysis over China. *Atmos Res* 223:24–38
- Graham LP, Andreasson J, Carlsson B (2007) Assessing climate change impacts on hydrology from an ensemble of regional climate models, model scales and linking methods - a case study on the Lule River Basin. *Clim Chang* 81:293–307
- Hawkins E, Sutton R (2009) The potential to narrow uncertainty in regional climate predictions. *Bull Am Meteorol Soc* 90:1095–1107
- Hawkins E, Sutton R (2011) The potential to narrow uncertainty in projections of regional precipitation change. *Clim Dyn* 37:407–418
- Hay LE, Wilby RL, Leavesley GH (2000) A comparison of delta change and downscaled GCM scenarios for three mountainous basins in the United States. *J Am Water Resour Assoc* 36(2):387–397
- Hosking JM, Wallis JR (1997) *Regional frequency analysis: an approach based on L-moments*. Cambridge University Press, Cambridge
- IPCC (2013) In: Stocker TF, Qin D, Plattner GK et al (eds) *Climate change 2013: the physical science basis*. Contribution of Working

- Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change. Cambridge University Press, Cambridge
- Knutti R (2010) The end of model democracy? *Clim Chang* 102:395–404
- Knutti R, Sedlacek J, Sanderson BJ et al (2017) A climate model weighting scheme accounting for performance and interdependence. *Geophys Res Lett* 44:1909–1918
- Krzysztofowicz R, Herr HD (2001) Hydrologic uncertainty processor for probabilistic river stage forecasting: precipitation-dependent model. *J Hydrol* 249(1–4):46–68
- Maraun D, Widmann M (2018) Cross-validation of bias-corrected climate simulations is misleading. *Hydrol Earth Syst Sci* 22(9):4867–4873
- Massoud EC, Espinoza V, Guan B, Waliser DE (2019) Global climate model ensemble approaches for future projections of atmospheric rivers. *Earth's Future* 7:1136–1151
- Minville M, Brissette F, Leconte R (2008) Uncertainty of the impact of climate change on the hydrology of a Nordic watershed. *J Hydrol* 358:70–83
- Novoa V, Ahumada-Rudolph R, Rojas O (2019) Understanding agricultural water footprint variability to improve water management in Chile. *Sci Total Environ* 670:188–199
- Qi LY, Huang JC, Yan RH, Gao JF, Wang SG, Guo YY (2017) Modeling the effects of the streamflow changes of Xinjiang Basin in future climate scenarios on the hydrodynamic conditions in Lake Poyang, China. *Limnology* 18(2):175–194
- Raftery AE, Gneiting T, Balabdaoui F, Polakowski M (2005) Using Bayesian model averaging to calibrate forecast ensembles. *Mon Weather Rev* 113(5):1155–1174
- Sanderson BJ, Knutti R (2012) On the interpretation of constrained climate model ensembles. *Geophys Res Lett* 39:L16708
- Sanderson BJ, Knutti R, Caldwell P (2015a) Addressing interdependency in a multimodel ensemble by interpolation of model properties. *J Clim* 28:5150–5170
- Sanderson BJ, Knutti R, Caldwell P (2015b) A representative democracy to reduce interdependency in a multimodel ensemble. *J Clim* 28:5171–5194
- Shao J, Wang J, Lv SY, Bing J (2016) Spatial and temporal variability of seasonal precipitation in Poyang Lake basin and possible links with climate indices. *Hydrol Res* 47(S1):51–68
- Slughter JM, Raftery AE, Gneiting T, Fraley C (2007) Probabilistic quantitative precipitation forecasting using Bayesian model averaging. *Mon Weather Rev* 135(9):3209–3220
- Smith I, Chandler E (2010) Refining rainfall projections for the Murray Darling Basin of south-east Australia—the effect of sampling model results based on performance. *Clim Chang* 102(3–4):377–393
- Song XM, Zhang JY, Zou XJ, Zhang C, AghaKouchak A, Kong F (2019) Changes in precipitation extremes in the Beijing metropolitan area during 1960–2012. *Atmos Res* 222:134–153
- Wang L, Chen W (2014) Equiratio cumulative distribution function matching as an improvement to the equidistant approach in bias correction of precipitation. *Atmos Sci Lett* 15:1–6
- Wang QJ, Robertson DE (2011) Multisite probabilistic forecasting of seasonal flows for streams with zero value occurrences. *Water Resour Res* 47:W02546
- Wang R, Zhang JQ, Guo EL, Zhao C, Cao T (2019) Spatial and temporal variations of precipitation concentration and their relationships with large-scale atmospheric circulations across Northeast China. *Atmos Res* 222:62–73
- Wilby RL, Harris I (2006) A framework for assessing uncertainties in climate change impacts: low-flow scenarios for the River Thames, UK. *Water Resour Res* 42:W02419
- Wootten A, Terando A, Reich BJ, Boyles RP, Semazzi F (2017) Characterizing sources of uncertainty from global climate models and downscaling techniques. *J Appl Meteorol Climatol* 56:3245–3262
- Yang X, Yu X, Wang Y, Liu Y, Zhang M, Ren L, Yuan F, Jiang S (2019) Estimating the response of hydrological regimes to future projections of precipitation and temperature over the upper Yangtze River. *Atmos Res* 230:104627
- Yuan Z, Xu J, Wang Y (2018) Projection of future extreme precipitation and flood changes of the Jinsha river basin in China based on CMIP5 climate models. *Int J Environ Res Public Health* 15(11):2491
- Zhao HH, Pan XB, Wang ZW et al (2019a) What were the changing trends of the seasonal and annual aridity indexes in northwestern China during 1961–2015? *Atmos Res* 222:154–162
- Zhao T, Wang QJ, Schepen A, Griffiths M (2019b) Ensemble forecasting of monthly and seasonal reference crop evapotranspiration based on global climate model outputs. *Agric For Meteorol* 264:114–124
- Zhu J, Forsee W, Schumer R, Gautam M (2013) Future projections and uncertainty assessment of extreme rainfall intensity in the United States from an ensemble of climate models. *Clim Chang* 118:469–485

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