Hindawi Advances in Meteorology Volume 2020, Article ID 8680436, 17 pages https://doi.org/10.1155/2020/8680436



# Research Article

# **Downscaling and Projection of Multi-CMIP5 Precipitation Using Machine Learning Methods in the Upper Han River Basin**

Ren Xu, 1,2 Nengcheng Chen, 2,3 Yumin Chen, and Zeqiang Chen 2,3

Correspondence should be addressed to Zeqiang Chen; zeqiangchen@whu.edu.cn

Received 20 May 2019; Revised 5 January 2020; Accepted 30 January 2020; Published 9 March 2020

Academic Editor: Roberto Coscarelli

Copyright © 2020 Ren Xu et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Downscaling considerably alleviates the drawbacks of regional climate simulation by general circulation models (GCMs). However, little information is available regarding the downscaling using machine learning methods, specifically at hydrological basin scale. This study developed multiple machine learning (ML) downscaling models, based on a Bayesian model average (BMA), to downscale the precipitation simulation of 8 Coupled Model Intercomparison Project Phase 5 (CMIP5) models using model output statistics (MOS) for the years 1961-2005 in the upper Han River basin. A series of statistical metrics, including Pearson's correlation coefficient (PCC), root mean squared error (RMSE), and relative bias (Rbias), were used for evaluation and comparative analyses. Moreover, the BMA and the best ML downscaling model were used to downscale precipitation in the 21st century under Representative Concentration Pathway 4.5 (RCP4.5) and RCP8.5 scenarios. The results show the following: (1) The performance of the BMA ensemble simulation is clearly better than that of the individual models and the simple mean model ensemble (MME). The PCC reaches 0.74, and the RMSE is reduced by 28%-60% for all the GCMs and 33% compared to the MME. (2) The downscaled models greatly improved station simulation performance. Support vector machine for regression (SVR) was superior to multilayer perceptron (MLP) and random forest (RF). The downscaling results based on the BMA ensemble simulation and SVR models were regarded as the best performing overall (PCC, RMSE, and Rbias were 0.82, 35.07, mm and -5.45%, respectively). (3) Based on BMA and SVR models, the projected precipitations show a weak increasing trend on the whole under RCP4.5 and RCP8.5. Specifically, the average rainfall during the mid- (2040-2069) and late (2070-2099) 21st century increased by 3.23% and 1.02%, respectively, compared to the base year (1971-2000) under RCP4.5, while they increased by 4.25% and 8.30% under RCP8.5. Additionally, the magnitude of changes during winter and spring was higher than that during summer and autumn. Furthermore, future work is recommended to study the improvement of downscaling models and the effect of local climate.

#### 1. Introduction

General circulation models (GCMs) are widely utilized to study regional meteorological and hydrological responses under climatic changes. The low resolution of GCMs hinders their applications at a basin scale; thus, downscaling techniques are vital to obtain data at a local scale. Many studies have addressed this task based on various downscaling techniques in various regions of the globe [1–3]. There are two large groups for downscaling including dynamical

downscaling (DD), which uses regional climate models (RCMs) to downscale GCMs, and statistical downscaling (SD), which aims to achieve a statistical or empirical relationship between large-scale atmospheric variables (termed predictors) and regional variables (termed predictands). SD has the advantages of a low computational cost, relatively simple implementation, and reliable accuracy. Thus, SD techniques may be convenient if the study mainly concentrates on the downscaling and projection of precipitation at a basin scale.

<sup>&</sup>lt;sup>1</sup>School of Resource and Environmental Science, Wuhan University, 129 Luoyu Road, Wuhan 430079, China

<sup>&</sup>lt;sup>2</sup>State Key Laboratory of Information Engineering in Surveying, Mapping, and Remote Sensing, Wuhan University, 129 Luoyu Road, Wuhan 430079, China

<sup>&</sup>lt;sup>3</sup>Collaborative Innovation Center of Geospatial Technology, Wuhan University, Wuhan 430079, China

The SD methods are generally subdivided into two categories: perfect prognosis (PP) and model output statistics (MOS) [2]. For PP, the statistical relationship is established between a predictand and the observed largescale predictors, while for MOS, GCM-simulated predictors are exploited to establish the relationship with the observed predictand. Given the availability of a considerable database of historical forecasts and the ability to explicitly account for GCM-inherent bias, MOS has recently gained considerable popularity as an alternative to PP [4, 5]. It is possible for MOS to provide more information regarding the dynamic controls of precipitation or temperature, especially combining precipitation (temperature) data with circulation data as predictors [6]. MOS also has a high potential for projections of climate change [7, 8]. During recent years, MOS models based on regression have become popular for establishing the relationship between GCM-simulated variables and observed rainfall. There are successful studies using MOS downscaling based on regression techniques. For example, Jonathan et al. exploited maximum covariance analysis and principal component regression to downscale MOS in the extratropics and concluded that MOS clearly outperformed PP methods, which exploited observed largescale predictors [9].

However, the relationship between local precipitation and the spatial patterns of GCM hindcasts is often very complex. There have been several comparative studies that demonstrated the superiority of machine learning (ML) approaches compared to other SD approaches such as statistical downscaling models or conventional linear regression models [10–12]. ML approaches have proven to be efficient for modeling highly nonlinear relationships [7]. The artificial neural network (ANN) is competent to be applied for climate downscaling due to its ability to establish the nonlinear relationships between predictors and predictands [13]. Meanwhile, support vector machine for regression (SVR) models have also been successfully used to capture highly nonlinear relationships by applying kernel functions to map the low-dimensional input data to a high-dimensional feature space [10]. Furthermore, another ML method, random forest (RF), has been regarded as a competent and robust algorithm for representing complex relationships because it can implement different types of input variables and operates flexibly. This method was also successfully applied to precipitation downscaling [12]. However, the applicability of ML methods in MOS downscaling has not yet been widely explored. Sa'adi et al. compared two ML approaches, RF and SVR, to downscale monthly precipitation based on MOS. The results demonstrated that such strategies were capable of properly downscaling GCM precipitation, while the SVR showed a more accurate performance overall [7]. This result inspired the implementation of ML methods to downscale precipitation, based on MOS, in other regional basins.

Some studies have demonstrated that multimodel ensemble projections outperform the "best" single model over the long term, as they were effective to reduce the uncertainties [14, 15]. In addition, recent studies have suggested that the GCM is the largest quantified source of uncertainty

in the projected effects of climate change on hydrological processes. There are some proposed ensemble methods that have been proven to alleviate the GCM uncertainty, such as the simple model average (MME), Bayesian model average (BMA) [16, 17], and reliability ensemble average (REA) [18]. BMA is especially popular, as it assigns weights to each output model according to ensemble forecasts from different sources into a consensus probability density function (PDF). It is also popular to combine ensemble techniques with downscaling models for regional climate studies. For example, a simple statistical downscaling algorithm and superensemble methods were used to address forecast accuracy and regional predicted monsoon precipitation; it was demonstrated that such strategies outperformed each of the member models and their downscaling counterparts [19]. ANN and SVR were used for PP downscaling, and BMA was further exploited to obtain a multi-GCM ensemble. The corrected ensemble forecasts could represent the historical climate well [20].

However, there are no studies regarding the capacity of the projections for a combination of BMA-based ensembles and ML downscaling strategies based on MOS in the upper Han River. Thus, the major objectives of this study were to (1) develop MOS-based BMA\_ML ensemble downscaling models based on MLP, SVR, and RF methods in the upper Han River basin; (2) study the performance of the BMA method for ensemble model outputs; and (3) evaluate precipitation during the 21st century under RCP4.5 and RCP8.5 in upper Han River basin. Furthermore, based on the study results, a beneficial reference for future long-term management strategies of water resources, disaster mitigation, ecological layout, and other applications at a regional level is expected.

The remainder of the study is organized as follows: Section 2 describes the study area topography and the observational station data used. The applied methods are discussed in Section 3, including the data processing methods, assembling ensemble models, and downscaling. In Section 4, the results are discussed in detail. Lastly, the authors conclude and propose several prospects based on this study.

## 2. Study Area and Data

The Han River is the largest tributary of the Yangtze River, which covers 159 thousand km². In the region, the main terrain types are mountains and hills. The average observed annual rainfall is approximately 700~1100 mm and shows an obviously seasonal character. During the summer, heavy rainfall occurs due to the descending cold air from the south. Furthermore, flood and drought disasters in the basin are severe due to natural fluctuations and human intervention. In this study, the main region of the mountain and hill terrain of the upper Han River were selected, which contain 13 meteorological stations. Table 1 shows the position information of each station. The observed daily rainfall data for the years from 1961 to 2005 were utilized to calibrate and validate the downscaled models, and the data in 2006–2018 were used to validate the projected precipitation. These data

Station	Number	Longitude	Latitude	Elevation (m)
Taibai	1	107.19	34.02	1543.6
Liuba	2	106.56	33.38	1032.1
Hanzhong	3	107.02	33.04	509.5
Foping	4	107.59	33.31	827.2
Nanxian	5	109.58	33.52	742.2
Zhenan	6	109.09	33.26	693.7
Shangnan	7	110.54	33.32	523
Xishan	8	111.3	33.18	250.3
Shiquan	9	108.16	33.03	484.9
Ankang	10	109.02	32.43	290.8
Yunxi	11	110.25	33	249.1
Shiyan	12	110.47	32.39	286.5
Fangxian	13	110.45	32.03	426.9

were obtained from the China Meteorological Data Website (http://data.cma.cn/). Figure 1 shows the water system and topography of the Han River and station distributions.

For the CMIP5 data, the simulated historical and future precipitation under RCP4.5 and RCP8.5 from the GCMs were chosen for the regional simulation and projection. Eight GCMs were used, as they showed good performance in simulating precipitation in the study area. Table 2 provides a general description of these GCMs and Table 3 introduces RCP4.5 and RCP8.5. The GCM-simulated data were acquired from the Coupled Model Intercomparison Project website (http://cmip-pcmdi.llnl.gov/cmip5/). The resolution of these GCMs is not the same. To obtain consistent resolution of downscaling input, 8 GCMs are interpolated to 1° (longitude) × 1° (latitude), which is about 95 km × 111 km, based on bilinear interpolation. It is important to determine the appropriate climate domain for the reliable downscaling of climatic factors.

The considered domain size should be large enough to capture the rainfall mechanism without increasing the cost of CPU [21]. Thus, in this study, from the author's consideration that the domain should cover the area affected by climatic circulation patterns of precipitation in the study area, a spatial domain consisting of 171 grid points was selected, which covers the region between latitudes 99°E–117°E and longitudes 29°N–37°N (Figure 2). Then, the precipitation datasets of these 171 grid points were chosen to be the predictors.

# 3. Methods

This study aimed to develop an MOS-based BMA\_ML ensemble for downscaling models based on the MLP, SVR, and RF methods for the upper Han River basin. The main framework is shown in Figure 3.

The methodology of the present study encompasses the following steps: (i) predictor selection, (ii) GCM ensemble generation, (iii) downscaling using machine learning, and (iv) result evaluation.

3.1. Predictors Selection. This section mainly discusses the method for obtaining the selected predictors' data and

extracting the specific principal components (PCs) based on principal component analysis (PCA). The aim of PCA is to reduce the number of input variables without losing much information. PCA was first proposed by Hotelling in 1933 [22]. PCA is a multivariate statistical method that uses the central idea of dimensionality reduction to transform multiple indicators into several comprehensive indicators on the premise of reduced information loss. In statistics, PCA is an analysis for simplifying data. The main steps of PCA are as follows:

- (i) Computing the covariance matrix of sample data
- (ii) Calculating the eigenvalues and eigenvectors of this covariance matrix
- (iii) Calculating the cumulative contribution rate

The detailed equation can be seen in [22]. In this study, the data were standardized to eliminate the influence of single-sample data before PCA. The PCs of 171 GCM precipitation variables were calculated based on PCA, and the first few PCs were selected as predictors when the cumulative contribution rate was greater than 95% among all the PCs. The new variables express nearly the entirety of the raw variables and are regarded as predictors. These PCs are also used as the inputs of the ML downscaling models.

3.2. GCM Ensemble Generation. In this study, MME and BMA were used to generate a GCM ensemble. The MME only exploits the arithmetic mean of 8 precipitation models to obtain an ensemble simulation. As a statistical post-processing method, BMA combines multiple models. It has been recently applied to correct underdispersion in ensemble prediction and alleviate model uncertainty based on posterior probability [23]. The predictive PDF of the BMA for future variables is based on a weighted average, which centres on the individual downscaling forecasts [24]. In this study, the precipitation values from 8 CMIP5 models were used to obtain the ensemble output. Assuming that *y* is the target variable and that *D* is a given model's data, the combined forecast PDF of *y* is

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

where  $M = \{M_1, M_2, \dots, M_K\}$  are alternative CMIP5 models. K is equal to 8 in the current study.  $P(y | M_k, D)$  is the predictive PDF given for model  $M_k$  alone; k.  $P(M_k | D)$  denotes the posterior probability of model  $M_k$ , which is represented as

$$P(M_k \mid D) = \frac{P(D \mid M_k) P(M_k)}{\sum_{i=1}^{K} P(D \mid M_i) P(M_i)},$$
 (2)

where  $P(D \mid M_i)$  is the marginal likelihood, expressed as

$$P(D \mid M_i) = \int P(D \mid \theta_i, M_i) P(\theta_i \mid M_i) d\theta_i, \qquad (3)$$

where  $\theta_i$  is regarded as the parameter of model  $M_i$ , with  $P(\theta_i | M_i)$  denoting the prior probability for  $\theta_i$ .

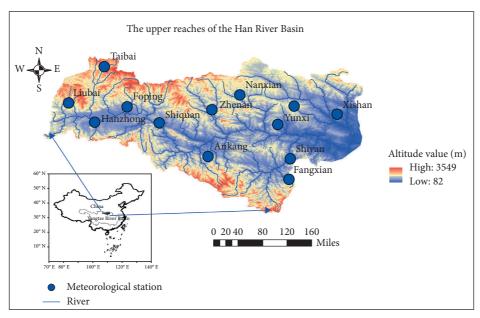


FIGURE 1: Sketch map of study area.

TABLE 2: Information of the 8 CMIP5 GCMs.

Model	Shorthand	Institution and country	Resolution (km)	
		Commonwealth Scientific and Industrial		
ACCESS1.0	AC10	Research Organization and Bureau of	192 * 145	
		Meteorology, Australia		
		Commonwealth Scientific and Industrial		
ACCESS1.3	AC13	Research Organization and Bureau of	192 * 145	
		Meteorology, Australia		
GFDL-ESM2G	GG	Geophysical Fluid Mechanics Laboratory, USA	144 * 90	
GFDL-ESM2M	GM	Geophysical Fluid Mechanics Laboratory, USA	144 * 90	
GISS-E2-H	GH	NASA Goddard Institute for Space Studies, USA	144 * 90	
GISS-E2-R	GR	NASA Goddard Institute for Space Studies, USA	144 * 90	
INMCM4	IN	UNM, Russia	180 * 120	
NorESM1-ME	Nor	NCC, Norway	144 * 96	

TABLE 3: Overview of RCP4.5 and RCP8.5.

RCP	Description
RCP4.5	Radiative forcing is stable at 4.5 W/m <sup>2</sup> (~650 ppm CO <sub>2</sub> eq) by 2100
RCP8.5	Radiative forcing increase to 8.5 W/m <sup>2</sup> (~1370 ppm CO <sub>2</sub> eq) after 2100

The respective weight for the 8 CMIP5 models was determined based on BMA. Then, the ensemble result was obtained based on these weights and models. Importantly, the posterior distribution of the multimodel was acquired based on a Markov chain Monte Carlo (MCMC) simulation algorithm [25]. The BMA procedure was run in R3.5.1, using the "BMS" package.

3.3. Machine Learning Downscaling. The ML methods applied in this study include MLP, SVR, and RF, which were used to downscale individual GCMs, MME, and BMA. The following subsections introduce these methods.

3.3.1. Multilayer Perceptron. As the classical type of neural network, MLP was used in this study [26]. The backpropagation (BP) algorithm was selected to train the MLP network [27]. The MLP consists of one input layer, one or more hidden layers, and one output layer. Figure 4 depicts a typical three-layer MLP network. In addition, equation (4) represents the output value of a three-layer MLP network [28]:

$$\widehat{y}_{k} = f_{o} \left[ \sum_{j=1}^{M} w_{kj} \cdot f_{h} \left( \sum_{i=1}^{N} w_{ji} x_{i} + w_{jo} \right) + w_{ko} \right], \tag{4}$$

where  $w_{ji}$  are the weights in the hidden layer connecting the *i*-th neuron in the input layer and *j*-th neuron in the hidden layer,  $w_{jo}$  is the bias for the *j*-th hidden neuron,  $f_h$  is the

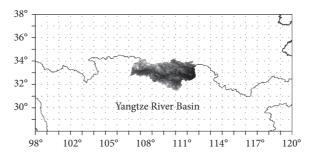


FIGURE 2: Grid points used from the GCM-simulated precipitation data.

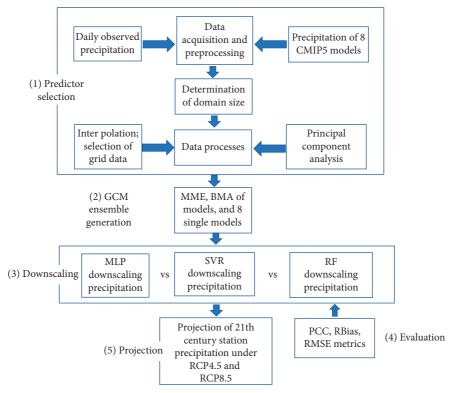


FIGURE 3: The framework of this study.

activation function of the hidden neuron,  $w_{ki}$  is the weight between the *j*-th neuron in the hidden layer and *k*-th neuron in the output layer,  $w_{ko}$  is the bias for *k*-th output neuron, and  $f_o$  is the activation function for the output neuron.

The main steps of BP algorithm are as follows:

- (i) Forward propagation: it includes the calculation of the sums of the input neurons, followed by the use of an activation function and the calculation of output neurons.
- (ii) Backpropagation: it includes the calculation of derivative of loss function, backpropagation of error signal through the network, and weight adjustment to minimize the overall error.
- (iii) Iteration of the above processes until overall error is satisfactorily small.

For the MLP network, the choice of hyperparameters is essential, such as the structure of the network, activation function, learning algorithm, and learning rate scheduling type and parameters [29]. In this study, the authors used Bayesian hyperparameter optimization to choose MLP hyperparameters. More details are provided in Section 4.2 and Section 1 of the supplementary materials.

3.3.2. Support Vector Machine. SVM is another machine learning method based on VC theory and the rule of structural risk minimization [30]. SVR can solve nonlinear regression problems by applying kernel functions to map the low-dimensional input data to a high-dimensional feature space. SVR methods have been successfully applied in precipitation downscaling based on MOS [8]. There are no documented MOS-based BMA\_SVR downscaling applications. Assuming a training sample  $D = \{(x_1, y_1), (x_2, y_2), \ldots, (x_n, y_n)\}$   $x_i \in \mathbb{R}^n$ ,  $y_i \in \mathbb{R}$ , a linear model is represented as

$$f(x) = w^T x + b, (5)$$

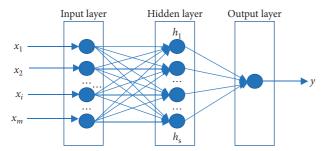


FIGURE 4: The construction of a typical three-layer MLP network.

where w and b denote the support vector and the constant, respectively. Generally, w should be as small as possible. T represents the transpose of matrix. The expectation is to yield an f(x) that is as close as possible to y. SVR assumes that it can tolerate the most deviations between f(x) and y. We define it as  $\varepsilon$ . Furthermore, a slack variable should be added to the model to overcome overfitting. Thus, an optimal hyperplane is selected based on

$$\min \frac{1}{2} \|w\|^2 + C \sum_{i=1}^{z} (\xi_i + \widehat{\xi}_i),$$

s.t. 
$$\begin{cases} y_i - f(x) \le \varepsilon + \hat{\xi}_i, \\ f(x) - y_i \le \varepsilon + \xi_i, \\ \xi_i, \hat{\xi}_i \ge 0, \quad i = 1, 2, 3 \dots, z. \end{cases}$$
 (6)

In equation (6),  $\xi_i$  and  $\xi_i$  denote slack variables, and C is the penalty factor, determining the tradeoff between the minimal loss and maximal margins. As an optimization problem, equation (6) can be solved based on Lagrange's dual formulas. The solutions to equation (6) can be represented as

$$f(x) = \sum_{1}^{z} (\alpha_{i} - \widehat{\alpha}_{i}) \langle x_{i}, x \rangle + b, \tag{7}$$

where  $\alpha_i$  and  $\hat{\alpha}_i$  are Lagrange multipliers. So far, equation (7) is the linear model. The kernel functions are exploited to achieve nonlinear regression.

$$f(x) = \sum_{1}^{z} (\alpha_{i} - \widehat{\alpha}_{i}) \text{Kernel}\langle x_{i}, x \rangle + b,$$
 (8)

where *K* denotes the exploited kernel function. Regarding SVR, the general kernel functions include linear, polynomial, and Gaussian kernels. There are also some hyperparameters in SVR such as *C*, kernel function, as previously mentioned above, and epsilon, which represents the tolerance of the termination criterion. The choice of these hyperparameters plays a great role in an SVR model. In this study, the authors used Bayesian hyperparameter optimization to choose SVR hyperparameters. More details are provided in Section 4.2 and Section 3 of the supplementary materials.

3.3.3. Random Forest. RF was proposed by Breiman [31] as a novel machine learning algorithm composed of multiple

classification and regression decision trees (CART) that may avoid overfitting and can regulate different types of input variables. For more details on CART analysis, refer to Breiman et al. [32]. RF can generate multiple independent trees and then make a final decision, which is determined by its characteristic of nonparametric statistical regression and randomness. Accordingly, the decision-making ability of the RF model depends on each CART. Each decision tree consists of a root node, subnode, and leaf node. Each subnode contains a judgement rule, and a leaf node corresponds to a judgement level. The RF model steps are approximately divided into the following:

- (i) First, in this study, a training set is extracted from the original dataset using a bootstrapping method. Each training set is approximately 2/3 of the original dataset. The other 1/3 of the dataset is described as out-of-bag (OOB). A total of *n* rounds of extraction are carried out such that *n* training sets are formed. These training sets are independent of each other, but the elements can be duplicated.
- (ii) Then, one training set is used to build a regression decision tree. Many decision trees form a forest. The principle adopted is the minimum mean square error.
- (iii) In the last step, based on the prediction results of every tree, the result of the forest is the average of the predicted values of all trees.

Using OOB, the RF can internally conduct cross validation. In this study, the authors exploited the OOB error  $(E_{OOB})$  to estimate the internal error:

$$E_{\text{OOB}} = \frac{1}{n} \sum_{i=1}^{n} \left[ \widetilde{Y}(X_i) - Y_i \right]^2. \tag{9}$$

Regarding RF, the number of trees in the forest and the maximum depth of trees are the main RF hyperparameters. The authors also used Bayesian hyperparameter optimization to choose RF hyperparameters. More details are provided in Sections 4.2 and 2 of the supplementary materials.

- 3.4. Evaluation of Downscaling Models. To evaluate the performance of the established model, the following statistical metrics were used for the verification period:
  - (i) Pearson's correlation coefficient (PCC) evaluates the linear correlation between modeled and observed precipitation series; a PCC equal to 0 denotes no correlation while 1 represents complete correlation.
  - (ii) Root mean squared error (RMSE) represents errors between modeled and observed precipitation series; the smaller the RMSE, the better the results.
  - (iii) Relative bias (Rbias) evaluates the relative deviation between simulated and observed data; the smaller the Rbias, the smaller the relative deviation.

Table 4 shows the detailed equations and description of these statistical metrics.

#### 4. Results and Discussion

4.1. Ensemble Downscaling Simulation of BMA and MME. First, the authors evaluated the precipitation simulation using individual GCMs. Based on the interpolated grid data of the CMIP5 models, the 16 nearest grid points to each station were used to simply downscale. An average value of the 16 nearest grid points to each station was used to represent the GCM precipitation simulation for each station. Tables 4–6 provide the monthly simulation performance for each station and the regional mean. Therein, the mean value of the whole region was acquired based on the arithmetic mean of the 13 stations. Regarding BMA, the selected period was from 1993 to 2005. Each model has a certain reliability to simulate real rainfall, though there are great differences, as shown in the tables. Clearly, ACCESS1.0, GFDL-ESM2M, and GISS-E2-R outperformed the other models in this region, as the PCC reaches 0.64-0.68, and the RMSE is approximately 57–82 mm. The good PCCs are paired with poor RMSEs and Rbias in some cases. Furthermore, the differences for each station are apparent. The simulation performance results of all the models for stations 5 and 13 are relatively poor. These results may be due to a local microclimate as the GCMs cannot consider regional simulation. This theme is worthy of further study.

As described in the Methods, BMA was applied for ensemble simulation of 8 CMIP5 models, and MME was compared to BMA. MME exploits the arithmetic mean of precipitation of 8 models to yield an ensemble simulation. Figure 5 shows the weighted value of the 8 models. Tables 5–7 also provide the performance results of the MME and BMA. The region mean presents the arithmetic mean precipitation of 13 stations. The MME provides an improvement compared to all single models. The PCC is improved from 0.50-0.68 to 0.76, and the RMSE is slightly less than that of ACCESS1.0 and IN for the region mean. Furthermore, the BMA results prove its superiority. The PCC, RMSE, and Rbias values between the BMA simulation and observed data are distinctly better than those of the single individual models. The PCC reaches 0.74, and the RMSE is reduced by 28%-60%. Compared with that of MME, the RMSE decreases from 60.86 mm to 41.03 mm and the Rbias is greatly reduced, although the PCC exhibits little change. For this region, the given weights of each model via BMA are shown in Figure 5. The performances of the 8 single models, MME, and BMA for precipitation simulation are shown through a Taylor diagram, as shown in Figure 6. The PCC, RMSE, and standard deviation are represented in the Taylor diagram. The results graphically conclude that the BMA has a greatly improved simulation performance in this study region. The satisfactory performance of multimodel ensemble simulation demonstrated their capacity to alleviate the structural differences in each model and the uncertainty that exists in individual models, as demonstrated by the identical conclusion of previous studies [33, 34].

From the results, the great performance of BMA proves that it is more reliable than the other approaches, as BMA reduced the errors. More effort is still needed to yield precipitation simulation at regional scales because it is difficult for the GCMs to represent regional climate situations [35]. Accordingly, there are large differences in precipitation simulation for each station, which is likely because the local climatic circulation of precipitation events can induce a greater interannual variability in rainfall [36]. In addition to BMA, there are other statistical ensemble methods that have been successfully applied, such as reliability ensemble averaging (REA) and the generalized linear model (GLM) [7, 18]. Recently, the GLM performed well with ensembles of downscaled GCM projections [7]; this finding could inspire further study, which would refer to the comparison of ensemble-downscaling and downscalingensemble strategies, applying the BMA and GLM methods. Huge uncertainties remain, requiring further efforts to reduce them.

4.2. Validation and Comparison of Three Machine Learning Downscaling Models. The authors conducted the process of hyperparameter choice of MLP, SVR, and RF based on Bayesian hyperparameter optimization. The Bayesian optimization maps hyperparameters to the scoring probability of the objective (e.g., loss of model performance) and can infer information on the unknown function [37]. The authors adopted the tree-structured Parzen estimator (TPE) algorithm under the framework of sequential model-based global optimization (SMBO) as better results are reported using it in several difficult learning problems [38]. In addition, a 10-fold cross-validation was used to promote the choice of optimal hyperparameters. More details regarding the hyperparameter optimization process and results can be found in Sections 1-3 of the supplementary materials. All ML downscaling models were built based on the optimal hyperparameters. Eventually, these models were evaluated based on PCC, RMSE, and Rbias, whose definitions are provided in Table 4.

The authors used Python's "hyperopt" package to implement Bayesian hyperparameter optimization for MLP and RF, and MATLAB's "fitrsvm" function for SVM. Furthermore, the authors used 10-fold cross-validation to identify the optimal parameters, where the data during the period of 1961-2005 were divided into 10 equal sized subdatasets. Nine out of 10 subdatasets were used for training, and the remaining one was used for model validation. Mean squared error (MSE) was used as the crossvalidation score (objective) for MLP and RF, while the loss function was used as the objective for SVR because it is a selfcontained function of "fitrsvm." Tables 8-10 provide the hyperparameter results of MLP, SVR, and RF for the region mean, respectively. The detailed description for each parameter and search range for optimization are provided in Sections 1-3 in the supplementary materials, as well as the optimization processes for each station.

Figure 7 shows the selected PC for each GCM, MME, and BMA. The selected PCs are the input of ML methods, while the observed precipitation of each station and region mean are the output. Table 11 shows a comparison of the three ML downscaling methods for the performance evaluation based on statistical metrics. Due to the length

Statistical metric	Equation	Description
PCC	$PCC_{x,y} = \sum_{i=1}^{n} (x_i - \overline{x}) (y_i - \overline{y}) / \sqrt{\sum_{i=1}^{n} (x_i - \overline{x})^2} \sqrt{\sum_{i=1}^{n} (y_i - \overline{y})^2}$	<i>n</i> is the sample size; $x_i$ , $y_i$ are individual samples; $\overline{x}$ , $\overline{y}$ are the arithmetic mean of $x$ and $y$ , respectively
RMSE	$RMSE = \sqrt{\sum_{i=1}^{n} (y_{Obs} - y_{pre})^2/n}$	$y_{\text{Obs}}$ denotes observed data; $y_{\text{pre}}$ is the prediction value; $n$ expresses the sample size
Rbias	Rbias = $\sum_{i=1}^{n} (y_{pre} - y_{Obs}) / \sum_{i=1}^{n} y_{Obs}$	Same as the description of RMSE equation

TABLE 4: Detailed equations and description of the statistical metrics.

Table 5: The PCC between the monthly observed precipitation and the simulation performance of individual GCMs, MME, and BMA simulation in the period of 1961–2005.

Station														
Model	1	2	3	4	5	6	7	8	9	10	11	12	13	Mean
AC1.0	0.63	0.62	0.58	0.61	0.60	0.60	0.58	0.54	0.62	0.57	0.57	0.59	0.41	0.68
AC1.3	0.59	0.59	0.54	0.58	0.51	0.54	0.49	0.51	0.53	0.54	0.51	0.59	0.42	0.62
GG	0.58	0.54	0.48	0.51	0.44	0.44	0.35	0.35	0.46	0.41	0.36	0.39	0.27	0.50
GH	0.63	0.64	0.56	0.65	0.56	0.59	0.46	0.49	0.59	0.54	0.47	0.54	0.41	0.67
GM	0.66	0.64	0.56	0.58	0.46	0.47	0.40	0.41	0.49	0.48	0.43	0.44	0.28	0.57
GR	0.62	0.63	0.55	0.61	0.53	0.56	0.45	0.48	0.55	0.49	0.45	0.51	0.43	0.64
IN	0.50	0.51	0.45	0.51	0.47	0.47	0.46	0.47	0.46	0.45	0.44	0.51	0.34	0.54
Nor	0.64	0.63	0.59	0.59	0.53	0.55	0.47	0.47	0.57	0.52	0.47	0.47	0.35	0.62
MME	0.75	0.74	0.67	0.73	0.67	0.68	0.61	0.62	0.67	0.65	0.61	0.66	0.48	0.76
BMA	0.74	0.73	0.66	0.72	0.66	0.64	0.61	0.61	0.67	0.63	0.60	0.66	0.47	0.74

Numbers 1-13 represent the meteorological stations. "Mean" denotes the region mean.

Table 6: The RMSE between the monthly observed precipitation and the simulation performance of individual GCMs, MME, and BMA simulation in the period of 1961–2005 (mm).

Station														
Model	1	2	3	4	5	6	7	8	9	10	11	12	13	Mean
AC1.0	56.75	65.83	71.83	69.35	59.78	66.33	65.95	70.09	69.73	74.46	65.39	71.28	83.17	57.18
AC1.3	59.74	69.38	78.32	73.22	65.98	72.10	74.83	75.12	77.83	82.21	72.70	77.47	87.97	63.70
GG	116.57	124.13	124.56	110.33	94.53	95.59	90.69	88.89	107.60	93.99	91.31	83.34	87.24	91.17
GM	145.31	162.49	152.88	111.79	66.73	70.68	74.20	73.84	97.37	70.90	68.69	67.22	77.92	80.56
GH	116.35	126.79	130.91	116.05	102.69	106.15	96.50	92.85	116.20	104.09	98.49	93.93	98.10	99.13
GR	132.94	164.76	161.73	124.68	68.67	74.66	71.59	70.15	105.95	75.78	68.33	66.24	71.69	82.67
IN	64.02	69.90	69.56	69.27	62.33	64.61	72.21	73.73	71.65	68.12	68.76	68.28	81.06	58.81
Nor	116.32	125.54	117.63	107.21	109.13	104.18	103.75	101.38	104.26	102.86	106.13	101.24	104.55	100.15
MME	77.90	88.49	90.94	74.32	60.91	64.17	63.95	63.21	74.05	67.39	63.00	60.98	70.24	60.86
BMA	44.26	53.58	56.09	55.15	41.80	50.23	56.95	58.91	57.82	51.24	51.42	46.24	55.79	41.03

Numbers 1-13 represent the meteorological stations. "Mean" denotes the region mean.

limitation, Table 11 only provides the regional mean, and the assessment results of individual stations are represented by Tables S4–S6 in the supplementary materials. No method has absolute superiority, as the performance of all individual GCMs, MME, and BMA is not superior to that of the other methods. Encouragingly, all the downscaled models show greatly improved station simulation performance compared with the raw GCM output. For the downscaling of individual GCMs, the marked improvements reveal that the PCC increased from 0.54–0.68 to 0.72–0.81, the RMSE changed from 57.18–100.15 to 36.66–41.97, and the Rbias values were clearly reduced. Moreover, there are relatively fewer improvements for MME and BMA. Downscaling of BMA especially showed that the RMSE decreased by approximately 5.4 mm. This

situation may be because the BMA simulation before machine learning training is relatively good, making it more difficult to significantly improve. From the final results, MME and BMA downscaling perform better than any of the single models.

Figures 8–10 show scatter plots between the down-scaled results and the observed precipitation in the region. The modeled data are closer to observed data if these scatters are more concentrated on the line. The SVR plots clearly show that the modeled data are closer to observed data than those of MLP and RF. According to Table 1, it is clear that there are relative differences in each down-scaling method. According to the performance assessed by PCC and RMSE, the SVR results outperform the others in most cases, especially for the GCM downscaling of GR,

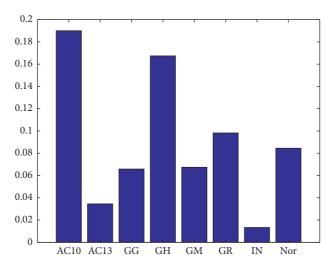


FIGURE 5: The weights of each CMIP5 model based on BMA.

Table 7: The Rbias between the monthly observed precipitation and the simulation performance of individual GCMs, MME, and BMA simulation in the period of 1961–2005 (%).

Station														
Model	1	2	3	4	5	6	7	8	9	10	11	12	13	Mean
AC1.0	34.03	33.31	42.76	30.24	49.45	46.20	28.88	23.62	34.40	56.97	42.00	50.45	46.71	39.63
AC1.3	30.17	33.33	40.93	25.25	40.93	39.33	25.37	21.99	28.56	54.66	36.84	50.45	47.93	36.40
GG	131.02	119.93	115.68	84.88	90.05	74.89	42.10	33.01	78.04	66.97	58.04	44.20	35.51	74.44
GM	128.46	129.44	125.40	81.10	53.77	50.09	27.68	23.07	66.20	46.00	34.26	38.91	40.09	65.04
GH	137.33	130.21	129.51	98.52	105.41	93.27	58.51	48.47	93.32	88.91	76.54	65.89	55.71	90.41
GR	124.50	140.22	140.15	95.72	57.81	58.18	25.26	15.75	77.82	54.96	35.66	40.22	37.57	69.77
IN	29.09	19.55	11.00	3.58	41.56	26.07	29.25	28.39	6.98	31.34	35.71	41.93	46.35	26.44
Nor	126.37	118.09	107.33	85.36	125.58	101.99	84.41	76.12	83.56	97.08	99.82	88.49	81.54	97.30
MME	92.62	90.51	89.10	63.08	70.57	61.25	40.18	33.80	58.61	62.11	52.36	52.57	48.93	62.43
BMA	-7.13	-12.14	-9.26	-6.23	-7.52	-7.66	2.92	-3.10	-1.24	-2.55	-0.17	-1.58	-0.65	-4.32

Numbers 1-13 represent the meteorological stations. "Mean" denotes the region mean.

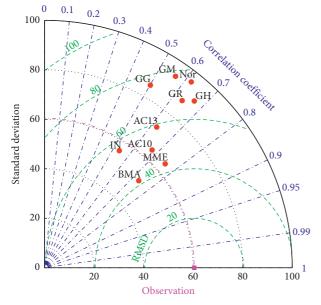


FIGURE 6: Taylor diagram for the 8 GCMs, MME, and BMA.

IN, and Nor. The difference in performance at specific stations (Tables S4–S6 in the supplementary materials) is maybe due to the differences in local climate and localized rainfall features. However, we did not explore this theme.

In this study, the BMA and SVR models were used for the downscaling future projection of precipitation. The overall outperformance of SVR based on the MOS method over RF has also been proposed in other regions, but there is no absolute superiority for every GCM and station [7]. This study also came to a similar conclusion. Although neural networks have been popularly applied for downscaling [39, 40], they were first tested on MOS. Compared with MLP and RF, the results of the MLP show a higher PCC and a lower RMSE. Those results demonstrated that there are relative uncertainties among the downscaling methods. There are some satisfactory studies using multimethod ensemble for modeling [41]. This has inspired future work using ensemble methods to reduce the uncertainties.

Generally, the modeling performance of the ML methods depends on their input and parameters [42]. For

T 0 0 1 1	(D : 1	The second of th
IADIE X. (Infimal recilife	ot Kawecian nynernarameter	ontimization for MILP for region mean
TABLE 6. Optimal results	of Dayesian hyperparameter	optimization for MLP for region mean.

	Activation	Alpha	HLZ	LR	Max_iter	Solver	Tol	Objective
AC10	tanh	1.024201	20	Adaptive	1500	adam	0.00319	1444
AC13	logistic	9.831136	12	Adaptive	125	lbfgs	0.008452	1605
GG	tanh	4.26283	16	Adaptive	1152	adam	0.002589	1501
GH	tanh	8.535319	14	Adaptive	1385	adam	0.000246	1511
GM	logistic	3.838534	14	Adaptive	413	adam	0.006199	1528
GR	tanh	9.454777	12	Adaptive	1521	adam	0.003523	1755
IN	logistic	7.885871	11	Invscaling	145	lbfgs	0.005522	1520
Nor	logistic	9.299373	10	Constant	1511	lbfgs	0.000955	1458
MME	tanh	3.534963	6	Adaptive	834	adam	0.000142	1340
BMA	logistic	9.045651	9	Constant	1360	lbfgs	0.007592	1278

HLZ, hidden\_layer\_sizes; LR, learning\_rate.

10

TABLE 9: Optimal results of Bayesian hyperparameter optimization of SVR for region mean.

,	BoxConstraint	KernelScale	Epsilon	KernelFunction	PolynomialOrder	Standardize	Objective
AC10	976.76	856.48	0.47992	Gaussian	NaN	False	7.3573
AC13	984.83	80.665	1.4186	Gaussian	NaN	False	7.3572
GG	43.612	89.569	0.5045	Gaussian	NaN	False	7.4439
GH	981.77	41.537	15.467	Gaussian	NaN	False	7.4516
GM	949.64	NaN	14.106	Polynomial (rbf)	1	True	7.3604
GR	637.68	61.876	0.071061	Gaussian	NaN	False	7.4265
IN	203.68	NaN	0.30124	Polynomial (rbf)	2	True	7.406
Nor	753.41	144.223	0.12571	Gaussian	NaN	True	7.4268
MME	12.972	NaN	8.73	Polynomial	2	True	7.2276
BMA	881.15	NaN	0.37308	Polynomial	2	True	7.2081

Table 10: Optimal results of Bayesian hyperparameter optimization for RF for region mean.

	max_depth	max_features	<i>n</i> _estimators	Objective
AC10	10	5	298	1562.
AC13	18	3	512	1604
GG	5	3	498	1581
GH	4	6	985	1551
GM	6	4	127	1630
GR	12	3	967	1657
IN	6	3	934	1549
Nor	13	2	451	1530
MME	4	2	622	1440
BMA	4	2	190	1401

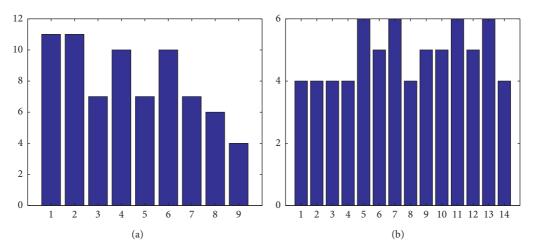


FIGURE 7: The size of selected PC for (a) each GCM and MME: 1–8 denote the GCM number, and 9 is MME; (b) BMA for each station: 1–14 denote station number.

Model	PCC			RMSE (mm)			Rbias (%)		
Model	MLP	SVR	RF	MLP	SVR	RF	MLP	SVR	RF
AC1.0	0.78	0.75	0.75	38.10	41.92	39.61	-6.49	-3.02	3.29
AC1.3	0.75	0.75	0.75	40.07	41.21	40.05	-3.61	-2.92	3.86
GG	0.76	0.75	0.74	38.91	44.5038.56	39.83	-3.18	1.90	-6.24
GM	0.76	0.77	0.76	38.94	38.80	39.42	-9.22	-3.58	-7.60
GH	0.75	0.76	0.74	39.76	39.27	40.42	9.48	-2.53	-5.80
GR	0.72	0.76	0.74	41.97	39.24	40.71	-15.93	-5.67	11.49
IN	0.76	0.79	0.76	39.00	37.80	39.34	2.15	-4.80	3.86
Nor	0.77	0.81	0.76	38.20	36.66	39.10	6.70	3.40	-4.88
MME	0.78	0.80	0.78	36.55	36.11	37.91	-6.88	-2.88	-6.16
BMA	0.80	0.82	0.79	35.58	35.07	37.43	6.13	-5.45	5.07

TABLE 11: The validation results of three methods for this region.

the input, it is clear that the more reliable the GCM, the better the performance of the downscaled results. For parameter set, though the authors used Bayesian hyperparameter optimization to choose the optimal hyperparameters of the ML method, there may be room for improving the optimization algorithm itself.

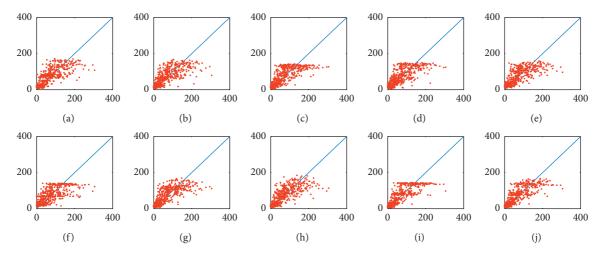
4.3. Projected Changes in Precipitation during the 21st Century. The projected precipitation outputs from 2006 to 2009 under RCP 4.5 and RCP 8.5 were assembled into an ensemble and downscaled through the SVR models using the established BMA, i.e., using the same GCM weights and selecting the same principal components. Firstly, we validated monthly precipitation projections against observational data for the 2006-2018 period. Table 12 shows the validation results. It can be seen that the projected precipitation under RCP4.5 is closer to the real data than RCP8.5, as the PCC is 0.78 and the RMSE is 43.57. In addition, the values of Rbias suggest that the real emission trend for the 2006-2018 period was somewhere between the RCP4.5 and RCP8.5 ones and was closer to the RCP4.5 one. As the change of climate is complex and uncertain, the changes of precipitation during the mid- (2040-2069) and late (2070-2099) 21st under RCP4.5 and RCP8.5 are worthy of study.

The monthly rainfall simulation was converted to annual and seasonal time series. Therein, the four seasons are spring (March-May), summer (June-August), autumn (September-November), and winter (December-February). Figure 11 shows the projected future annual and seasonal precipitation in this region. Considering 1971-2000 as the historical baseline, the average rainfall during the mid- and late 21st century increases by 3.23% and 1.02%, respectively, compared with the base year under RCP4.5; meanwhile, they increase by 4.25% and 8.30% under RCP8.5. From Figure 11, it can be seen that the rainfall during the 21st century has a weak increasing trend overall, and there is a slight downward fluctuation in the midterm. Under the RCP8.5 scenario, the increase in rainfall is more significant. For seasonal change, the mean value was obvious during winter and spring, while there was less change during summer and autumn. Figures 12 and 13 compare the statistics of the historical baseline and mid- and late 21st century time series based on quantilequantile plots under RCP4.5 and RCP8.5. It can be seen that most rain distributions are near the normal distributions. In each subfigure, three baselines that represent the corresponding normal distributions are shown. The interception and slope of these lines represent the mean and variance. Approximately, compared to the period of 1971–2000, the precipitation of the mid- and late 21st century under RCP4.5 and RCP8.5 clearly increases. The changes in magnitude during the spring and winter are great. This result may be due to the greater uncertainty in the downscaling models during the dry season or the bias in RCP4.5 and RCP8.5.

The trend of change throughout the 21st century in this region is consistent with that of previous studies, though the specific values are not the same [15, 44]. This difference is acceptable because the selected GCMs and downscaling strategies differ. The specific projection of each station is shown in Table S7 in the supplementary materials. The projection of precipitation change for each station is distinct. Significant precipitation seasonality exists, though this study did not take measures to remove it; thus, the evaluation of seasonal rainfall has more uncertainties. There are several studies that consider seasonality by separately training downscaling models based on each calendar season and month [43]. The authors considered the sufficient sample size for the training of the ML methods and thus used all monthly data, so far, for modeling. There are further plans to train monthly and seasonal models based on daily data, though much uncertainty exists in the daily rainfall. There are successful studies that have assessed extreme precipitation events based on daily downscaled precipitation [44]. Daily downscaled precipitation is also worth studying in a future work.

#### 5. Conclusion

Statistical downscaling of precipitation in complex-topography hydrological basins is of major importance, since the low resolution of GCMs hinders the effective study of climate change at this scale. Improving downscaling through machine learning techniques has been investigated in this study, through the use of three MOSbased machine learning methods (MLP, SVR, and RF). Prior to downscaling, the BMA model was developed into



12

FIGURE 8: Scatter plots between downscaled results and observed precipitation in the period of 1961–2005 for MLP. Horizontal axes show observed precipitation, and vertical axes show the MLP results. (a) AC10, (b) AC13, (c) GG, (d) GH, (e) GM, (f) GR, (g) IN, (h) Nor, (i) MME, (j) BMA.

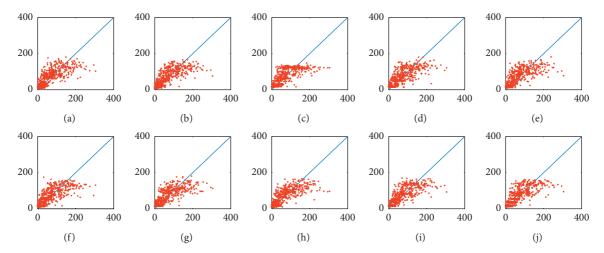


FIGURE 9: Scatter plots between downscaled results and observed precipitation in the period of 1961–2005 for RF. Horizontal axes show observed precipitation, and vertical axes show the RF results. (a) AC10, (b) AC13, (c) GG, (d) GH, (e) GM, (f) GR, (g) IN, (h) Nor, (i) MME, (j) BMA.

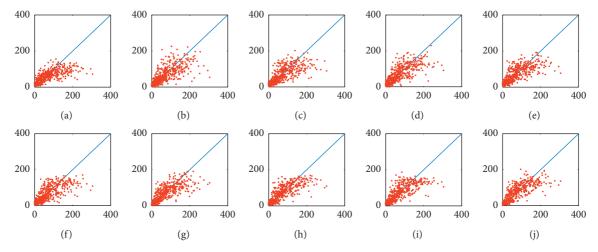


FIGURE 10: Scatter plots between downscaled results and observed precipitation in the period of 1961–2005 for SVR. Horizontal axes show observed precipitation, and vertical axes show the SVR results. (a) AC10, (b) AC13, (c) GG, (d) GH, (e) GM, (f) GR, (g) IN, (h) Nor, (i) MME, (j) BMA.

TABLE 12: The validation results of projected precipitation in 2006-2	TABLE	validation	12: The	results	of	projected	precipitation	in	2006-2018	3.
---	-------	------------	---------	---------	----	-----------	---------------	----	-----------	----

	PCC	RMSE	Rbias (%)
RCP4.5	0.78	43.57	-10.66
RCP8.5	0.74	48.86	14.28

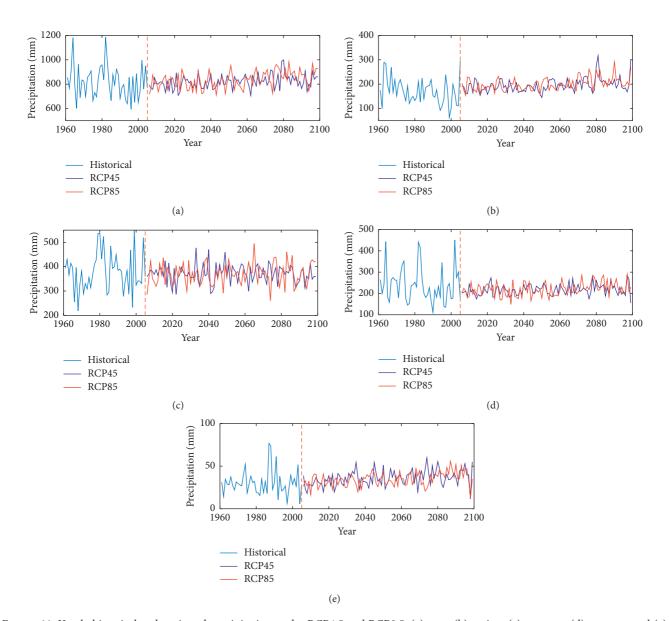


FIGURE 11: Yearly historical and projected precipitation under RCP4.5 and RCP8.5: (a) year, (b) spring, (c) summer, (d) autumn, and (e) winter, respectively. Horizontal axes represent the year, and vertical axes represent the rainfall (mm).

a multi-GCM ensemble, while MME was applied as a reference. A series of statistical metrics including PCC, RMSE, and Rbias were used to conduct an evaluation and comparative analyses. Then, BMA and SVR were applied to downscale the precipitation projection during the 21st century under RCP4.5 and RCP8.5. A validation of projected monthly precipitation in 2006–2018 was done for region mean, and the changes of precipitation during the mid- and late 21st century were analyzed. The conclusions are as follows:

(i) MME and BMA both demonstrated their superiority over the individual GCMs. Specifically, the PCC, RMSE, and Rbias between the BMA ensemble simulation and the observed data were clearly better than that of the single models. The PCC reached 0.74, and the RMSE was reduced by 28%–60% over all the GCMs. Compared with the MME, the RMSE decreased from 60.86 mm to 41.03 mm and the Rbias was greatly reduced, though the PCC showed little change.

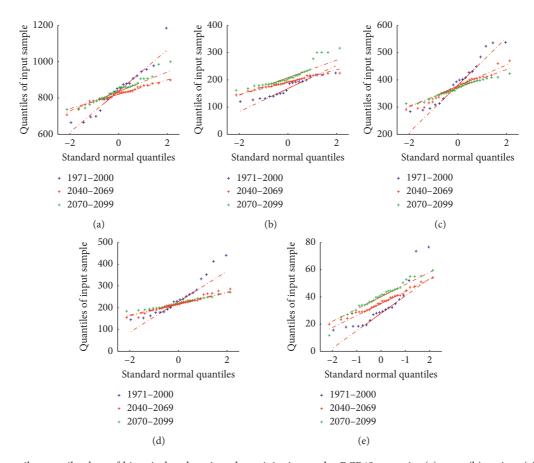
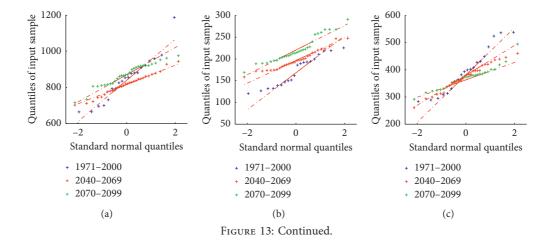
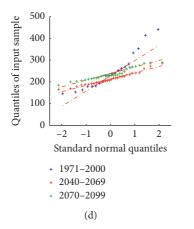


Figure 12: Quantile-quantile plots of historical and projected precipitation under RCP45 scenario: (a) year, (b) spring, (c) summer, (d) autumn, and (e) winter, respectively.





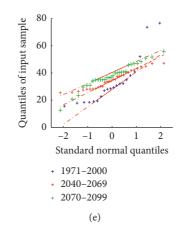


FIGURE 13: Quantile-quantile plots of historical and projected precipitation under RCP8.5 scenario: (a) year, (b) spring, (c) summer, (d) autumn, and (e) winter, respectively.

- (ii) All the downscaled models have greatly improved station simulation performance compared with that of the raw GCM output. The PCC increased from 0.54–0.68 to 0.72–0.81, and the RMSE changed from 57.18–100.15 mm to 36.66–41.97 mm. The SVR showed the best performance on the whole. The downscaling results based on the BMA ensemble simulation and the SVR models were regarded as the best performing (the PCC, RMSE, and Rbias were 0.82, 35.07 mm, and –5.45%, respectively)
- (iii) The projected rainfall under RCP4.5 is closer to the real data than RCP8.5 in 2006–2018. And the regional rainfall during the 21st century has a weak increasing trend on the whole, under RCP4.5 and RCP8.5, while there is a slight decreasing fluctuation in the midterm. Specifically, the average rainfall during the mid- and late 21st century increased by 3.23% and 1.02%, respectively, compared with the base year under RCP4.5; meanwhile, they increased by 4.25% and 8.30% under RCP8.5, and the magnitude of changes during winter and spring was higher than that during summer and autumn.

This study first developed MOS-based BMA\_ML ensemble downscaling models and provided preliminary downscaled projections of precipitation during the 21st century for the region, which can be referenced in studies of future climate, hydrological, economic change, etc. However, unresolved uncertainties remain. It may be advisable for researchers to study the improvement of downscaling models and the applicability for other regions for follow-up work, subsequently considering the effect of local climate and seasonality, based on multiple time scales.

#### **Nomenclature**

CMIP5: Coupled model intercomparison project phase 5

RCP: Representative concentration pathways

GCM: General circulation models RCM: Regional climate models SD: Statistical downscaling PP: Perfect prognosis SD. For PP, the statistical relationship is established between a predictand and the observed large-scale predictors

MOS: Model output statistics SD. For MOS, GCMsimulated predictors are exploited to establish the relationship with the observed predictand

ML: Machine learning
MLP: Multilayer perceptron

MLP: Multilayer perceptron SVR: Support vector regression

RF: Random forest
BMA: Bayesian model average
MME: Simple mean model ensemble
BP: Backpropagation algorithm

PCC: Pearson's correlation coefficient. PCC equal to 0 denotes no correlation while 1 represents total

correlation

RMSE: Root mean squared error. The smaller the RMSE,

the better the results

Rbias: Relative bias

PCA: Principal component analysis

AC10: ACCESS1.0 GCM AC13: ACCESS1.3 GCM GH: GISS-E2-H GCM GG: GFDL-ESM2G GCM GM: GFDL-ESM2M GCM GR: GISS-E2-R GCM IN: **INMCM4 GCM** Nor: NorESM1-ME.

# **Data Availability**

The observed daily rainfall data for the years from 1961 to 2005 were utilized to calibrate and validate the downscaled models, and they were obtained from the China Meteorological Data Website (http://data.cma.cn/). The GCM-simulated data were acquired from the Coupled Model Intercomparison Project website (http://cmip-pcmdi.llnl.gov/cmip5/).

# **Conflicts of Interest**

The authors declare no conflicts of interest.

# Acknowledgments

This work was supported by the National Key Research and Development Program of China (no. 2017YFB0503704) and the National Nature Science Foundation of China Program (nos. 41771422 and 41971351).

# **Supplementary Materials**

The supplementary materials mainly include the processes and details of Bayesian hyperparameter optimization for multilayer perceptron (MLP), support vector regression (SVR), and random forest (RF) for region mean. In addition, the supplementary materials include the evaluation of 21 NEX-GDDP models and MME model for specific stations in this region, and the changes of projected 21st century precipitation for each station. Each section has been cited in the manuscript. (Supplementary Materials)

## References

- E. P. Salathé Jr., "Downscaling simulations of future global climate with application to hydrologic modelling," *Interna*tional Journal of Climatology, vol. 25, no. 4, pp. 419–436, 2005.
- [2] D. Maraun, F. Wetterhall, A. M. Ireson et al., "Precipitation downscaling under climate change: recent developments to bridge the gap between dynamical models and the end user," *Reviews of Geophysics*, vol. 48, no. 3, Article ID RG3003, 2010.
- [3] S. H. Pour, S. B. Harun, and S. Shahid, "Genetic programming for the downscaling of extreme rainfall events on the East Coast of peninsular Malaysia," *Atmosphere*, vol. 5, no. 4, pp. 914–936, 2014.
- [4] J. M. Eden and M. Widmann, "Downscaling of GCM-simulated precipitation using model output statistics," *Journal of Climate*, vol. 27, no. 1, pp. 312–324, 2014.
- [5] M. Turco, P. Quintana-Seguí, M. C. Llasat, S. Herrera, and J. M. Gutiérrez, "Testing MOS precipitation downscaling for ENSEMBLES regional climate models over Spain," *Journal of Geophysical Research Atmospheres*, vol. 116, no. 18, 2011.
- [6] J. M. Eden, Development of a correction approach for future precipitation changes simulated by general circulation models, Ph.D. thesis, University of Birmingham, Birmingham, UK, 2011.
- [7] Z. Sa'adi, S. Shahid, E.-S. Chung, and T. b. Ismail, "Projection of spatial and temporal changes of rainfall in Sarawak of Borneo Island using statistical downscaling of CMIP5 models," *Atmospheric Research*, vol. 197, pp. 446–460, 2017.
- [8] S. H. Pour, S. Shahid, E.-S. Chung, and X.-J. Wang, "Model output statistics downscaling using support vector machine for the projection of spatial and temporal changes in rainfall of Bangladesh," *Atmospheric Research*, vol. 213, pp. 149–162, 2018
- [9] M. E. Jonathan and W. Martin, "Downscaling of GCM-simulated precipitation using model output statistics," *Journal of Climate*, vol. 27, no. 1, pp. 312–324, 2014.
- [10] L. Campozano, D. Tenelanda, E. Sanchez, E. Samaniego, and J. Feyen, "Comparison of statistical downscaling methods for monthly total precipitation: case study for the Paute River Basin in Southern Ecuador," Advances in Meteorology, vol. 2016, pp. 1–13, 2016.

- [11] D. Raje and P. P. Mujumdar, "A comparison of three methods for downscaling daily precipitation in the Punjab region," *Hydrological Processes*, vol. 25, no. 23, pp. 3575–3589, 2011.
- [12] G. Ibarra-Berastegui, J. Sáenz Aguirre, A. Ezcurra-Talegón, A. J. Elías-Saenz, and J. Díaz De Argandoña González, "Downscaling of surface moisture flux and precipitation in the Ebro Valley (Spain) using analogues and analogues followed by random forests and multiple linear regression," *Hydrology* and Earth System Sciences Discussions, vol. 15, no. 6, pp. 1895–1907, 2011.
- [13] A. Ochoa, L. Campozano, E. Sánchez, R. Gualán, and E. Samaniego, "Evaluation of downscaled estimates of monthly temperature and precipitation for a Southern Ecuador case study," *International Journal of Climatology*, vol. 36, no. 3, pp. 1244–1255, 2015.
- [14] J. M. Bravo, W. Collischonn, A. R. Da Paz, D. Allasia, and F. Domecq, "Impact of projected climate change on hydrologic regime of the upper Paraguay River Basin," *Climatic Change*, vol. 127, no. 1, pp. 27–41, 2014.
- [15] Z. Domecq, H. Gu, J. Wang, J. Xia, and B. Lu, "Effect of projected climate change on the hydrological regime of the Yangtze River Basin, China," Stochastic Environmental Research and Risk Assessment, vol. 32, no. 1, pp. 1–16, 2017.
- [16] Q. Duan and T. J. Phillips, "Bayesian estimation of local signal and noise in multimodel simulations of climate change," *Journal of Geophysical Research*, vol. 115, no. D18, Article ID D18123, 2010.
- [17] C. Miao, Q. Duan, Q. Sun, and J. Li, "Evaluation and application of Bayesian multi-model estimation in temperature simulations," *Progress in Physical Geography: Earth and Environment*, vol. 37, no. 6, pp. 727–744, 2013.
- [18] R. R. Torres and J. A. Marengo, "Uncertainty assessments of climate change projections over South America," *Theoretical* and Applied Climatology, vol. 112, no. 1-2, pp. 253–272, 2013.
- [19] X. Wang, T. Yang, X. Li, P. Shi, and X. Zhou, "Spatio-temporal changes of precipitation and temperature over the Pearl River Basin based on CMIP5 multi-model ensemble," *Stochastic Environmental Research & Risk Assessment*, vol. 31, no. 5, pp. 1–13, 2017.
- [20] U. Okkan and U. Kirdemir, "Downscaling of monthly precipitation using CMIP5 climate models operated under RCPs," *Meteorological Applications*, vol. 23, no. 3, pp. 514– 528, 2016.
- [21] M. R. Najafi, H. Moradkhani, and S. A. Wherry, "Statistical downscaling of precipitation using machine learning with optimal predictor selection," *Journal of Hydrologic Engineering*, vol. 16, no. 8, pp. 650–664, 2010.
- [22] H. Hotelling, "Analysis of a complex of statistical variables into principal components," *Journal of Educational Psychology*, vol. 24, no. 6, pp. 417–441, 1933.
- [23] W. Wang, Y. Ding, Q. Shao et al., "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, vol. 232, pp. 89–105, 2017.
- [24] H. Wu, X. Zhang, S. Liang, H. Yang, and G. Zhou, "Estimation of clear-sky land surface longwave radiation from MODIS data products by merging multiple models," *Journal of Geophysical Research: Atmospheres*, vol. 117, no. D22, 2012.
- [25] W. R. Gilks, S. Richardson, and D. Spiegelhalter, Markov Chain Monte Carlo in Practice, CRC Press, Boca Raton, FL, USA, 1995.

[26] M. Minsky and S. Papert, Perceptron: An Introduction to Computational Geometry, The MIT Press, Cambridge, MA, USA, 1969.

- [27] D. E. Rumelhart, G. E. Hinton, and R. J. Williams, Learning Internal Representations by Error Propagation. No. ICS-8506, California Univ San Diego La Jolla Inst for Cognitive Science, San Diego, CA, USA, 1985.
- [28] T.-W. Kim and J. B. Valdés, "Nonlinear model for drought forecasting based on a Conjunction of Wavelet transforms and neural networks," *Journal of Hydrologic Engineering*, vol. 8, no. 6, pp. 319–328, 2003.
- [29] Y. Bengio, "Practical recommendations for gradient-based training of deep architectures," in *Neural Networks: Tricks of* the Trade, pp. 437–478, Springer, Berlin, Heidelberg, Germany, 2012.
- [30] V. Vapnik, The Nature of Statistical Learning Theory; Information Science and Statistics, Springer-Verlag, New York, NY, USA, 2nd edition, 2000.
- [31] L. Breiman, "Random forests," *Machine Learning*, vol. 45, no. 1, pp. 5–32, 2001.
- [32] L. Breiman, J. Friedman, C. J. Stone, and R. A. Olshen, Classification and Regression Trees, CRC Press, Boca Raton, FL, USA, 1984.
- [33] T. Yang, Y. Tao, J. Li et al., "-criterion model ensemble of CMIP5 surface air temperature over China," *Theoretical and Applied Climatology*, vol. 132, no. 3-4, pp. 1057–1072, 2017.
- [34] H. Zhang and G. H. Huang, "Development of climate change projections for small watersheds using multi-model ensemble simulation and stochastic weather generation," *Climate Dynamics*, vol. 40, no. 3-4, pp. 805–821, 2013.
- [35] IPCC, Climate Change 2013: The Physical Science Basis, 2013, http://www.ipcc.ch/report/ar5/wg1/docs/WGIAR5\_SPM\_ brochure\_en.Pdf.
- [36] K. Venkataraman, S. Tummuri, A. Medina, and J. Perry, "21st century drought outlook for major climate divisions of Texas based on CMIP5 multimodel ensemble: implications for water resource management," *Journal of Hydrology*, vol. 534, pp. 300–316, 2016.
- [37] Y. Xia, C. Liu, Y. Li, and N. Liu, "A boosted decision tree approach using Bayesian hyper-parameter optimization for credit scoring," *Expert Systems with Applications*, vol. 78, pp. 225–241, 2017.
- [38] J. S. Liu, R. Bardenet, Y. Bengio, and B. Kégl, "Algorithms for hyper-parameter optimization," in *Proceedings of Advances in Neural Information Processing Systems*, pp. 2546–2554, Granada, Spain, December 2011.
- [39] T. Soares Dos Santos, D. Mendes, and R. Rodrigues Torres, "Artificial neural networks and multiple linear regression model using principal components to estimate rainfall over South America," *Nonlinear Processes in Geophysics*, vol. 23, no. 1, pp. 13–20, 2016.
- [40] S. Dorji, S. Herath, and B. K. Mishra, "Future climate of Colombo downscaled with SDSM-neural network," *Climate*, vol. 5, no. 1, 2017.
- [41] W. Zhang, G. A. Vecchi, G. Villarini, H. Murakami, R. Gudgel, and X. Yang, "Statistical-dynamical seasonal forecast of Western North Pacific and East Asia Landfalling Tropical Cyclones using the GFDL FLOR Coupled climate model," *Journal of Climate*, vol. 30, no. 6, pp. 2209–2232, 2017.
- [42] K. Whan and M. Schmeits, "Comparing area probability forecasts of (extreme) local precipitation using Parametric and machine learning statistical Postprocessing methods," *Monthly Weather Review*, vol. 146, no. 11, pp. 3651–3673, 2018.

[43] T. Raziei, "An analysis of daily and monthly precipitation seasonality and regimes in Iran and the associated changes in 1951–2014," *Theoretical and Applied Climatology*, vol. 134, no. 3-4, pp. 913–934, 2018.

[44] V. Moron, A. W. Robertson, M. N. Ward, and O. Ndiaye, "Weather types and rainfall over Senegal: part II: downscaling of GCM simulations," *Journal of Climate*, vol. 21, no. 2, pp. 288–307, 2008.