



## RESEARCH ARTICLE

WILEY

# A stacked ensemble learning and non-homogeneous hidden Markov model for daily precipitation downscaling and projection

Qin Jiang<sup>1,2,3</sup>  | Francesco Cioffi<sup>1</sup> | Federico Rosario Conticello<sup>4</sup>  |  
Mario Giannini<sup>1</sup> | Vito Telesca<sup>5</sup> | Jun Wang<sup>2,3</sup>

<sup>1</sup>Dipartimento di Ingegneria Civile, Edile e Ambientale, "La Sapienza" University of Roma, Rome, Italy

<sup>2</sup>Key Laboratory of Geographic Information Science (Ministry of Education), East China Normal University, Shanghai, China

<sup>3</sup>School of Geographic Sciences, East China Normal University, Shanghai, China

<sup>4</sup>HydroScan, Leuven, Belgium

<sup>5</sup>Scuola di Ingegneria, Università della Basilicata, Potenza, Italy

## Correspondence

Francesco Cioffi, Dipartimento di Ingegneria Civile, Edile e Ambientale, "La Sapienza" University of Roma, Rome, Italy.  
Email: [francesco.cioffi@uniroma1.it](mailto:francesco.cioffi@uniroma1.it)

Jun Wang, Key Laboratory of Geographic Information Science (Ministry of Education), East China Normal University, Shanghai, China.  
Email: [jwang@geo.ecnu.edu.cn](mailto:jwang@geo.ecnu.edu.cn)

## Funding information

Director's Fund of Key Laboratory of Geographic Information Science (Ministry of Education), East China Normal University, Grant/Award Number: KLGIS2022C01; dual doctoral degree program in geography and environmental and hydraulics engineering between East China Normal University and "La Sapienza" University of Roma; National Natural Science Foundation of China, Grant/Award Number: 41971199; National Social Science Foundation of China, Grant/Award Number: 18ZDA105

## Abstract

Global circulation models (GCMs) are routinely used to project future climate conditions worldwide, such as temperature and precipitation. However, inputs with a finer resolution are required to drive impact-related models at local scales. The non-homogeneous hidden Markov model (NHMM) is a widely used algorithm for the precipitation statistical downscaling for GCMs. To improve the accuracy of the traditional NHMM in reproducing spatiotemporal precipitation features of specific geographic sites, especially extreme precipitation, we developed a new precipitation downscaling framework. This hierarchical model includes two levels: (1) establishing an ensemble learning model to predict the occurrence probabilities for different levels of daily precipitation aggregated at multiple sites and (2) constructing a NHMM downscaling scheme of daily amount at the scale of a single rain gauge using the outputs of ensemble learning model as predictors. As the results obtained for the case study in the central-eastern China (CEC), show that our downscaling model is highly efficient and performs better than the NHMM in simulating precipitation variability and extreme precipitation. Finally, our projections indicate that CEC may experience increased precipitation in the future. Compared with ~26 years (1990–2015), the extreme precipitation frequency and amount would significantly increase by 21.9%–48.1% and 12.3%–38.3%, respectively, by the late century (2075–2100) under the Shared Socioeconomic Pathway 585 climate scenario.

## KEYWORDS

downscaling, extreme precipitation, future projection, NHMM

## 1 | INTRODUCTION

The latest Intergovernmental Panel on Climate Change (IPCC) report has highlighted thought-provoking conclusions, such as the unprecedented warming of the climate due to human influence

(IPCC, 2021). It has been proven that water-related climate hazards and multiple risks to ecosystems and human settlements intensify with global warming (Sammen et al., 2022; Tabari, 2021), particularly in coastal cities at low elevations (Toimil et al., 2020; Voudoukas et al., 2018).

This is an open access article under the terms of the [Creative Commons Attribution-NonCommercial-NoDerivs](https://creativecommons.org/licenses/by-nc-nd/4.0/) License, which permits use and distribution in any medium, provided the original work is properly cited, the use is non-commercial and no modifications or adaptations are made.

© 2023 The Authors. *Hydrological Processes* published by John Wiley & Sons Ltd.

The global circulation model (GCM) can provide useful large-scale fields of climate variables for future projections, but its horizontal spatial resolution is too coarse to be used in water management at the basin or regional scales (Christensen & Kjellström, 2020; Wang et al., 2020). In addition, precipitation simulated by GCMs is strongly biased, with heavy precipitation being notably underestimated because of limitations in the modelling process (Huang et al., 2020; Kang et al., 2015). These inadequacies often yield non-negligible uncertainties in climate projections (Her et al., 2019; Norris et al., 2021; Yang et al., 2020). It is crucial to establish strategies to significantly reduce such bias, especially for extreme precipitation.

Statistical downscaling is a widely used approach to improve precipitation projections for GCMs. Wilby and Wigley (1997) reviewed commonly used downscaling techniques to solve the problem of an insufficient resolution of climate models compared with observed data. These techniques can be divided into four categories: regression, weather typing, stochastic weather generators and mesoscale modelling. In general, the local-scale relevant weather variables were considered as the predictands, whereas the large-scale atmospheric circulation information of GCMs were predictors. Several authors have examined the potential of the non-homogeneous hidden Markov model (NHMM), a good representation of the stochastic weather generator and a very powerful diagnostic tool for detecting the stochastic parameters between multiple sites precipitation and large-scale atmospheric circulation variables (Cioffi et al., 2016, 2017; Shahriar et al., 2021; Siabi et al., 2021). The basic assumptions of this statistical downscaling approach are as follows: (1) the daily precipitation occurrence at multiple sites depends on a finite number of unobserved weather states (hidden states) and (2) state transition probabilities depend on the former state and time-dependent atmospheric predictors (Bates et al., 2010; Charles et al., 2004; Hughes et al., 1999).

Relevant candidate predictors for the NHMM mainly include the geopotential height (GP), sea level pressure, wind fields, air temperature, humidity and their combinations (Ghamghami & Bazrafshan, 2021; Neykov et al., 2012; Zha et al., 2022). Given the high dimensionality of the original datasets, the inputs have to undergo a principal component analysis to find the most dominant behaviours of the large-scale atmospheric predictors (Ghamghami et al., 2019). For instance, Panda et al. (2022) only selected the first few components explaining 95% variance for each predictor. Guo et al. (2022) used a combination of the first 26 components of four predictors to run the NHMM downscaling. Liu et al. (2011) used two indicators, that is, the explained variances and partial correlation to select the predictors from 20 candidate variables for daily precipitation downscaling. Cioffi et al. (2016) adopted statistical criteria, that is, normalized log-likelihood to evaluate the accuracy of different combinations of potential predictors. Nonetheless, selecting the predictors of such models can be computationally intensive, especially when dealing with a large number of predictors or hidden states. This complexity can limit their applicability, particularly for operational use.

In addition, a multinomial logistic function is used to model the dependence of the hidden state transition matrix on such exogenous variables. Despite the reliable results of precipitation variabilities obtained by NHMM downscaling in several applications, the NHMM yields inaccurate simulations of mid-latitude precipitation in spring

and autumn as well as extreme precipitation (Cioffi et al., 2020). Given the spatio-temporal variability of predictors, the NHMM is limited in its ability to capture extreme precipitation behaviours. To address these limitations, we proposed a slight modification of the NHMM downscaling framework, employing a hierarchical modelling scheme.

At the first level, we used an ensemble learning model to construct intermediate features that are instrumental in classifying precipitation within homogeneous regions. This ensemble learning model combines the robustness of Random Forest (RF) and the efficiency of eXtreme Gradient Boosting (XGBoost). These intermediate features, produced by our tailored ensemble learning model, influence the NHMM at the second level of our hierarchy. Here, the NHMM extracts statistical precipitation characteristics for each rain gauge within the defined homogeneous regions. The efficacy of this integrated approach is rooted in the proven success of machine learning techniques in hydrological statistical downscaling. Machine learning-based approaches like RF, Support Vector Machine (SVM) and Gradient Boosting (GB) have demonstrated effectiveness in yielding accurate predictions, particularly for multi-class classification tasks, common in hydrological prediction (Ahmed et al., 2020; Zhou, 2021). Thus, our ensemble learning model, underpinned by the RF and XGBoost algorithms, extends these proven methods' capabilities, improving overall forecasting accuracy and reliability. The choice of an ensemble learning model is further justified by the work of Rahman et al. (2021). They developed a multi-class flood probabilities assessment model for Bangladesh using different meta-classifiers in a stacking method, which showed minimized overfitting, a common issue in machine learning applications for hydrology.

The decision to pair NHMM with an ensemble learning model in this research stems from the success of similar models in other fields. Bengio et al. (1992), for instance, enhanced the performance of HMM using Artificial Neural Networks (ANN). Similarly, Hassan et al. (2007) improved stock market forecasting by merging HMM with ANN and Genetic Algorithms. Trentin and Gori (2001) underscored the substantial improvements in automatic speech recognition achieved by hybrid systems combining models specializing in classification (e.g., RF, XGBoost, ANN) with those adept at recognizing temporal patterns (e.g., HMM, NHMM). Despite the proven efficacy of these individual components, the combination of NHMM with an ensemble learning model is a novel approach in hydrology. The motivation behind this new approach is to leverage the strengths of both methodologies: the NHMMs' ability to capture temporal dependencies in hydrological data and the ensemble learning model's proficiency in mitigating overfitting risk, while maximizing predictive accuracy. By leveraging these complementary strengths, our innovative approach marks a promising step forward in enhancing the precision and reliability of precipitation prediction in the field of hydrology. This paper will outline the methodology and validation of this novel framework, highlighting its potential contributions to the ongoing evolution of hydrological modelling and forecasting.

In this study, we considered ensemble learning to predict the occurrence probability of different levels of daily precipitation aggregated at multiple sites. The GP at 500 hPa and the vertically integrated water vapour transport (IVT) were selected as key factors affecting the precipitation occurrence (Conticello et al., 2018). Furthermore, the predicted probabilities were used as predictors to

model the transitional probabilities of hidden states in the NHMM. We determined if such a difference in predictors can improve the accuracy of the NHMM results. The aims of our study were as follows: (1) design an ensemble learning model for predicting the probability of daily precipitation events with different intensities; (2) improve the accuracy of the NHMM model concerning precipitation simulation, especially extreme events; and (3) project future precipitation and possible changes in extremes over central-eastern China (CEC) under different emission scenarios.

## 2 | STUDY AREA AND DATASETS

### 2.1 | Study area

Our study area is the CEC (25° N–35° N, 106° E–123° E), where precipitation is greatly affected by synoptic-scale circulation systems, such as the Meiyu front and typhoon cyclone. During mid-June and mid-July, a quasi-stationary front forms over the mid-lower reaches of the Yangtze River, which is related to the poleward moisture transport from the ocean derived by Western Pacific Subtropical High (WPSH) and low-level southwesterly jet (Ding et al., 2020). The Meiyu rain band in the lower troposphere can generally be observed at the northern boundary of the front and feeds precipitation system development and maintenance. Another important factor is the cyclone activity. It has been proven that the tropical cyclone circulation system, including frontal clouds and outward spiralling rain bands, could favour more intense precipitation over CEC (Tang et al., 2021).

### 2.2 | Datasets

Daily precipitation observations (1960–2015) were recorded at 228 gauges from the China Meteorological Data Service Centre (<http://data.cma.cn/>). One reanalysis product, the fifth generation atmospheric reanalysis (ERA5) with 0.25° spatial resolution was obtained from the Copernicus Climate Change Service Climate Data Store. One GCM of the Coupled Model Intercomparison Project Phase 6 (CMIP6), MPI-ESM1-2-HR (0.9° × 0.9°) was downloaded from the Max Planck Institute for Meteorology, Germany. To investigate the future precipitation projection, we used MPI-ESM1-2-HR to simulate both historical period precipitation (1960–2014) and that due to three climate scenarios with different shared socioeconomic pathways (SSP126, SSP245 and SSP585 from 2015 to 2100) (Meinshausen et al., 2020). Considering the historical simulation end by 2014 for GCMs, data from 2015 under the SSP585 scenario were employed as the historical to be consistent with the rain gauge period, while the future projection spans from 2016 to 2100.

For the domain of interest (10° S–70° N and 40° E–170° W), the used variables of ERA5 and MPI-ESM1-2-HR are geopotential height (GP) at 500 hPa, specific humidity and u/v wind fields from 250 to 1000 hPa. Their spatial resolution was resampled to 0.25° using the nearest neighbour interpolation method.

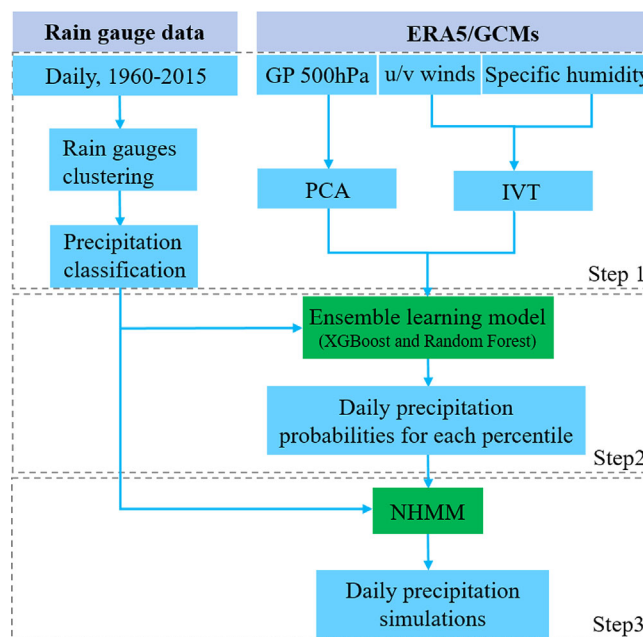


FIGURE 1 Flow chart of the proposed downscaling framework.

## 3 | METHODOLOGY

The overall structure of the stacked ensemble learning and non-homogeneous hidden Markov models ('Ensemble-NHMM') which is proposed in this study, consists of three steps: data preprocessing, ensemble learning model and NHMM (Figure 1). In the following subsection, the three steps are described in detail.

### 3.1 | Data preprocessing

Data preprocessing consists of four steps: (1) identification of different rain gauge clusters based on event synchronization of precipitation extremes; (2) for each cluster identifying different precipitation amount classifications; (3) choice of large-scale atmospheric predictors and (4) standardization and data reduction of predictors.

First, extreme precipitation refers to daily records exceeding the 95th threshold from 1960 to 2015 (only days with an amount above 0.1 mm are considered). For each rain gauge, the total amount and frequency of extreme precipitation per year were obtained. The synchrony degree matrix of extreme precipitation events among the rain gauges was calculated with the event synchronization method. Homogeneous regions for rain gauge classification were identified using a Complex Network approach (Newman, 2012) in conjunction with a pooling method known as the Louvain Method (Blondel et al., 2008; Conticello et al., 2020). The Complex Network approach effectively detects patterns and structures in the data, while the Louvain Method, celebrated for its efficiency in community detection within large networks, groups rain gauges into homogeneous regions based on similar rainfall patterns (Boers et al., 2019; Singhal et al., 2023). In this study, rain gauges were clustered into different clusters by

looking at the modularity values of the synchrony degree matrix. Within the same cluster, extreme precipitation events have a high degree of synchronization. For detailed information, please refer to Jiang et al. (2022).

Second, the daily precipitation was converted into categorical data. Let  $Y_t$  be a vector denoting the observed precipitation levels on the day  $t$ .

$$Y_t = \max(Pr_t \geq Pr_x^*) \quad (1)$$

where  $Pr_t$  is the original precipitation amount averaged by all rain gauges across a cluster,  $Pr_x^*$  indicates the precipitation thresholds of  $Pr_t$  at different percentiles and  $\max(\cdot)$  is a transform function that yields the maximum level on day  $t$  for  $Pr_t$ .

Furthermore, large-scale atmospheric circulation variables of the ERA5 or GCMs were utilized as predictors. It is recommended that predictors should be physically sensible in regional precipitation processes and reliable for future projections. Consequently, time sequences of selected variables were standardized using the z-score method, which is dimensionless. The principal component analysis (PCA) was applied to identify their leading behaviours for the domain of interest.

Here, GP at 500 hPa and the IVT (calculated from the specific humidity and u/v wind fields from 1000 to 250 hPa, Henny et al. (2023)) were utilized as predictors. We chose the two predictors based on the following considerations: IVT quantifies the intensity of moisture transport which affects a given specific geographic region and thus the probable intensity of precipitation. GP represents the atmospheric circulation features and in particular the fluid flow structure. Together, these two predictors describe the direction and intensity of humidity transport. Moreover, our previous study showed that the anomalous mid-level GP and strong water vapour transport from the adjacent seas could provide a favourable atmospheric circulation background for precipitation, especially extremes over CEC (Jiang et al., 2022).

We then extracted the standardized IVT values corresponding to the geographical location of the rain gauges as an independent variable ( $IVT\_set$ ). To better interpret the circulation variations, the first standardized 50 principal components of GP that explained more than 96% of the total variance were selected as the second predictor ( $GP\_pca$ ).

### 3.2 | Ensemble learning model

The task of the ensemble learning model is to learn and predict the probability of predefined mean daily precipitation levels of a rain gauge cluster based on two machine learning approaches: XGBoost and RF. RF, a tree-based machine learning method, is renowned for its high accuracy, robustness to overfitting and its capability to handle large datasets with high dimensionality (Breiman, 2001). XGBoost complements this with a regularized model formulation that controls overfitting, thereby enhancing overall performance (Chen & Guestrin, 2016). Relevant methodologies of XGBoost and RF are introduced in Supporting Information.

Figure 2 shows the workflow of the ensemble learning model. Its core concept is to use predictions from the lower level cast into the higher level for the meta classifier and minimize the errors. The inputs are represented by the atmospheric predictors  $I = \{(IVT\_set_{t,m}, GP\_pca_{t,p}, Y_t)\}$ ,  $t = 1, \dots, T$ ;  $m = 1, \dots, M$ ,  $p = 1, \dots, P$ , where  $T$  and  $M$  are the total number of days from 1960 to 2015 and the number of rain gauges in a cluster, respectively;  $P$  is the number of principal components of GP for the domain of interest. The descriptions of  $IVT\_set$ ,  $GP\_pca$  and  $Y_t$  were presented in Section 3.1.

At level 1, 90% of the samples were randomly separated as the training set to optimize the first XGBoost model. Its task is to obtain the preliminary probability prediction ( $Predict\_1$ ) of multi-level daily precipitation for each day. At level 2, an ensemble-based approach was adopted to combine the outputs of the basic learner (RF) with the meta-learner (second XGBoost) to achieve the best prediction through

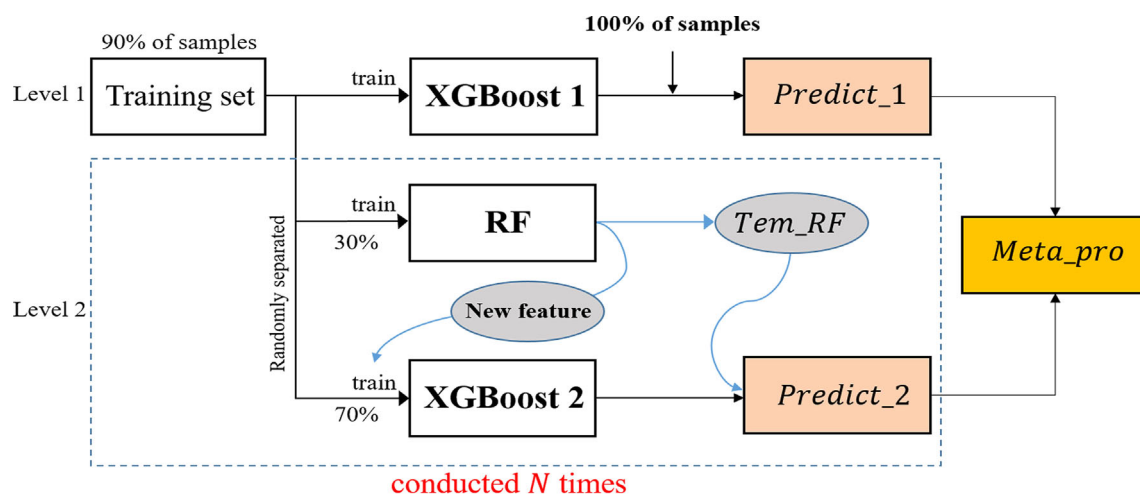


FIGURE 2 Flowchart of the ensemble learning model.



fast iteration. First, the training set was randomly divided into two subsets again by the ratio 3:7. The RF model was constructed on 30% of the training set and aimed to write out a new feature with the remaining. Subsequently, the second XGBoost model was trained using 70% of the training data, alongside the additional feature.

The prediction task at level 2 was conducted  $N$  times (in this experiment,  $N = 120$ ). For each round, all samples were applied to the RF model to obtain the additional features ( $Tem\_RF$ ) for each day. The  $Tem\_RF$  and the original feature data were then fed in to the second XGBoost model to make the 2nd prediction ( $Predict\_2$ ). Finally, the predictions of the basic learner (level 1) were stacked with those of meta classifiers (level 2) to output the final decision ( $Meta\_pro$ ), which is the average of the  $N + 1$  independent predictions.

The output  $Meta\_pro$  is a two-dimensional vector with  $c$  columns and  $T$  rows, containing the predicted probability of each data point belonging to each class of mean precipitation percentile as defined in Equation (1), where  $c$  is the total number of predefined daily precipitation levels. Therefore, possible weather states on day  $t$  can be quantified using the probability information contained in the  $c$  columns. The sum of the predicted probabilities of the  $c$  precipitation levels on day  $t$  is equal to 1.

The ensemble learning model was implemented in Rstudio using R version 4.1.3. Table 1 summarizes the specific parameter settings and explanations of the XGBoost and RF classifiers. First, the ratio of training samples was set to 0.8 to tune the hyperparameters, for example, tree depth (max.depth), learning rate (eta) and leaf node weight (min\_child\_weight) for XGBoost. The tuning task was to minimize the training logloss of multi-classes. In the RF, the out of bag score was referenced to select the number of trees (ntree) and candidates at each split (ntry). For the final predictions, we referred to the confusion matrix to evaluate the model's performance in simulating the probabilities of different precipitation levels.

In our downscaling framework, the simulated multi-level daily precipitation probabilities with homogeneous regions were used as the intermediate features to cast the NHMM downscaling scheme.

### 3.3 | Non-homogeneous hidden Markov model

The NHMM model is a stochastic dual process involving two variables: the daily rainfall amount of a multivariate set of rain gauges, and a finite and countable number of discrete hidden states (Hughes et al., 1999). Such variables and their mutual link are defined as follows:

1. Let  $R_t = (R_t^1, \dots, R_t^M)$  be a multivariate vector of the daily precipitation record from  $M$  rain gauges on day  $t$  belonging to a cluster, it is assumed that the observations  $R_t$  at time  $t$  are independent of all other variables in the model up to time  $t$ , and moreover they are depending on the hidden state  $S_t$  at time  $t$ , that is,  $P(R_t | R_{1:t-1}, S_{1:t}) = P(R_t | S_t)$ ;
2. Let  $S_t = (1, \dots, s)$  be the hidden state (or weather state) on day  $t$  whose probability transition from a hidden state to another one is modelled as a first-order Markovian process ( $S_t | S_{1:t-1}, X_{1:t}) = P(S_t | S_{t-1}, X_t)$  and it is conditioned by exogenous variables  $X_t$  which generally represent the influence of atmospheric predictors on the transition probability.

Subsequently, the log-likelihood of the data under the NHMM model can be formulated as

$$l = \log P(R|X) = \log \sum_S \left[ P(S_1 | X_1) \prod_{t=2}^T P(S_t | S_{t-1}, X_t) \right] \left[ \prod_{t=1}^T P(R_t | S_t) \right] \quad (2)$$

We used multinomial logistic regression to model multivariate hidden state transitions:

$$P(S_t = j | S_{t-1} = i, X_t = x) = \frac{\exp(\sigma_{ji} + \rho_j x^t)}{\sum_{h=1}^H \exp(\sigma_{jh} + \rho_h x^t)} \quad (3)$$

where  $H$  denotes the number of hidden states. An independent delta-gamma function was selected to model emission probabilities. The

**TABLE 1** Parameter settings in the ensemble learning model.

	Parameter	Explanation	Value
XGBoost	booster	Specify the type of booster	gbtree
	objective	Specify the learning task of the model	multi:softprob
	max.depth	Maximum depth of a tree	8 <sup>a</sup> , 11 <sup>b</sup>
	eta	Learning rate	0.1 <sup>a</sup> , 0.46 <sup>b</sup>
	nrounds	Number of rounds	1200 <sup>a</sup> , 60 <sup>b</sup>
	min_child_weight	Minimum weight for a child leaf node	3 <sup>a</sup> , 10 <sup>b</sup>
	subsample	Ratio of training samples	0.8
	eval_metric	Evaluation metrics for validation	mlogloss
RF	ntree	Number of trees	100
	ntry	Number of variables randomly used as candidates at each split	7

<sup>a</sup>XGBoost 1.

<sup>b</sup>XGBoost 2.

best NHMM fits the conditional probability of the observation sequence determined using the expectation–maximization technique (Moon, 1996). The most likely weather state sequence can then be generated from the NHMM using the Viterbi algorithm. Log-likelihood and Bayesian information criteria were selected to identify the most appropriate number of hidden states. In this study, a 5-fold cross-validation was used to train the NHMM.

Contrary to the conventional approach that relies on atmospheric circulation data to construct the NHMM downscaling model, we introduced a novel downscaling framework. This framework leverages the probabilities of various precipitation levels derived from ensemble learning model outputs as predictors. For each cluster, we carried out two distinct downscaling experiments:

- The Ensemble-NHMM downscaling model, which is anchored on daily precipitation probabilities.
- The standard NHMM that incorporates the original sequence of the primary principal components from standardized atmospheric predictors.

For our study, both models underwent 100 simulations.

### 3.4 | Evaluation indicators

Two indicators were selected to assess the performance of the Ensemble-NHMM and NHMM. The first is the coefficient of variation of the root-mean-squared error (CVRMSE), which quantifies the standard deviation of the differences between simulations and

observations. The other is the correlation coefficient (CC), which is used to estimate the linear relationship between simulations and observations.

$$\text{CVRMSE} = \frac{\sqrt{\frac{1}{n} \sum_{i=1}^n (sim_i - obs_i)^2}}{\frac{1}{n} \sum_{i=1}^n obs_i} \quad (4)$$

$$\text{CC} = \frac{\sum_{i=1}^n (sim_i - \bar{sim})(obs_i - \bar{obs})}{\sqrt{\sum_{i=1}^n (sim_i - \bar{sim})^2} \sqrt{\sum_{i=1}^n (obs_i - \bar{obs})^2}} \quad (5)$$

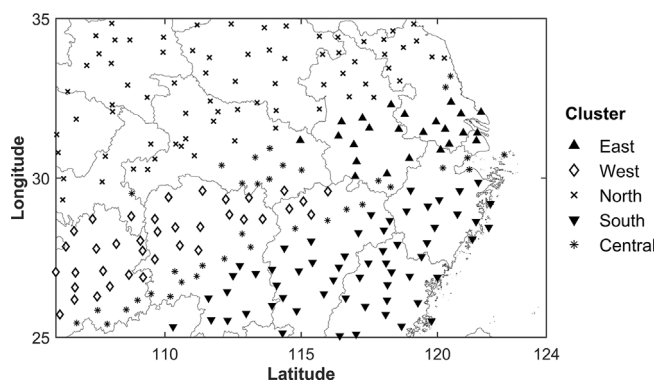
where *sim* and *obs* indicate the simulations and observations from the Ensemble-NHMM (or NHMM) and rain gauge, respectively, and *n* is the total number of samples.

## 4 | RESULTS

### 4.1 | Ensemble learning model performance in daily precipitation classification

#### 4.1.1 | Clustering of rain gauges

In our previous research (Jiang et al., 2022), rain gauges over the CEC were divided into four clusters for two time periods, 1960–1989 and 1990–2015, when considering abrupt changes in extreme precipitation (Figure S1). In this study, we categorized a new cluster named ‘Central’, (including the rainfall gauges whose communities differed between the two periods) as shown in Figure 3. The remaining clusters are labelled as ‘East’, ‘West’, ‘North’ and ‘South’. Subsequently, the average daily precipitation across each cluster was converted into categorical data. A number of model runs varying the number of classes and the precipitation percentile of each class were carried out. To prevent model overfitting, balance was maintained in the number of observations between the classes considering that, for each cluster, from 1960 to 2015, more than 25% of observations concerned no rain days. As explained below in Section 4.1.2, based on the accuracy rate in reproducing the observed time series of daily precipitation levels, five percentiles (i.e., no rain, 0.1 mm/day, 30th, 60th and 95th) were selected. Table 2 shows the intervals corresponding to different



**FIGURE 3** Spatial distributions of five rain gauge clusters.

**TABLE 2** Classifications of daily precipitation (unit: mm/day) for the five clusters.

	I	II	III	IV	V
East	[0, 0.1)	[0.1, 0.95)	[0.95, 4.19)	[4.19, 23.29)	≥23.29
West		[0.1, 0.86)	[0.86, 3.48)	[3.48, 19.58)	≥19.58
North		[0.1, 0.66)	[0.66, 2.54)	[2.54, 14.08)	≥14.08
South		[0.1, 1.23)	[1.23, 4.55)	[4.55, 21.16)	≥21.16
Central		[0.1, 1.09)	[1.09, 3.91)	[3.91, 17.62)	≥17.62

precipitation percentiles of the five clusters. The precipitation levels corresponding to these intervals were labelled as follows: no event (I), very light precipitation (II), light precipitation (III), moderate precipitation (IV) and heavy precipitation (V).

#### 4.1.2 | Performances of ensemble learning model

For each rain gauge cluster, the output of the ensemble learning model was a two-dimensional vector containing the predicted occurrence probability of the five predefined daily precipitation levels from 1960 to 2015. To investigate the performance of the ensemble learning model, we selected the simulated level with the highest occurrence probability as the eventually predicted level on a daily basis and compared it with the observed value [ $Y_t$ , in Equation (1)]. Table 3 shows the confusion matrix of the predicted and observed daily precipitation levels for the five clusters. The accuracy rate is the ratio of correct predictions for a precipitation level. Overall, the ensemble

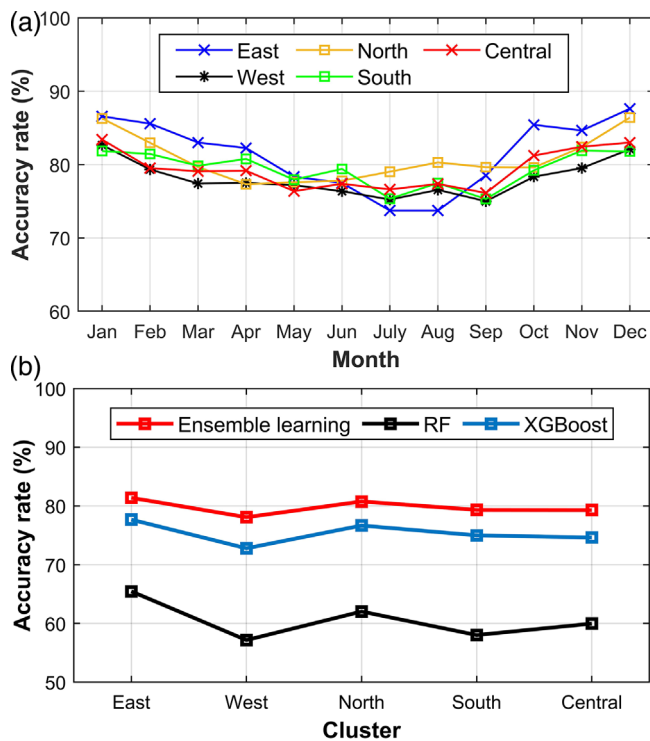
learning model shows a superior capability in classifying no rain days and in predicting moderate precipitation events for the five clusters (accuracy rates exceed 88% and 83%, respectively). A relatively lower accuracy rate was observed for very light and heavy precipitation levels; the confusion matrix indicates that some heavy precipitation events were misclassified as moderate. Furthermore, to evaluate the capability of the ensemble learning model in depicting precipitation seasonality, daily precipitation was classified into different months. As shown in Figure 4a, the ensemble learning model yields better performance in the non-rainy season (from October to March) for the five clusters, with an average classification accuracy rate of 82.14%; while the reported accuracy is 77.43% in the rainy season (from April to September). The lowest accuracy rate was observed in the East cluster from July to August.

To further understand the capability of the ensemble learning model, we separately examined the accuracy of using only the RF or XGBoost to predict the probabilities of multi-level daily precipitation. Figure 4b demonstrates the overall performance of the ensemble

**TABLE 3** Confusion matrix of the predicted and observed daily precipitation levels for the five clusters and related accuracy rate.

Cluster	Precipitation level	Actual	Predicted					Accuracy rate (%)
			I	II	III	IV	V	
East	Actual	I	<b>8922</b>	107	133	165	1	95.65
		II	908	<b>2008</b>	179	241	0	60.19
		III	634	127	<b>2094</b>	476	2	62.83
		IV	364	83	182	<b>3250</b>	8	83.61
		V	13	3	10	169	<b>361</b>	64.93
West	Actual	I	<b>5204</b>	273	180	201	0	88.84
		II	721	<b>2922</b>	325	408	0	66.77
		III	401	285	<b>2993</b>	693	2	68.43
		IV	204	163	347	<b>4374</b>	15	85.71
		V	13	12	16	217	<b>471</b>	64.61
North	Actual	I	<b>6384</b>	156	186	135	0	93.05
		II	899	<b>2616</b>	310	250	0	64.20
		III	397	195	<b>2841</b>	636	4	69.75
		IV	126	80	299	<b>4224</b>	23	88.89
		V	0	2	1	236	<b>440</b>	64.80
South	Actual	I	<b>4633</b>	357	126	96	1	88.87
		II	559	<b>3279</b>	400	327	3	71.78
		III	226	470	<b>3183</b>	685	4	69.68
		IV	100	213	362	<b>4635</b>	21	86.94
		V	1	4	6	262	<b>487</b>	64.08
Central	Actual	I	<b>4875</b>	316	165	108	0	89.22
		II	629	<b>3112</b>	407	349	0	69.20
		III	267	392	<b>3183</b>	646	2	70.89
		IV	115	188	376	<b>4539</b>	22	86.62
		V	1	4	11	230	<b>503</b>	67.16

Note: Bold represents the number of correct predictions of the ensemble learning model.



**FIGURE 4** Accuracy rate of the ensemble learning model in different months (a) and (b) the overall performance of the ensemble learning, RF, and XGBoost models for the five clusters.

learning, RF and XGBoost for the five clusters. The statistical results indicate that the ensemble learning model has better prediction skills than the RF and XGBoost. The accuracy rate of the ensemble learning model ranges from 78.10% to 81.38%, while it ranges from 57.16% to 65.47% for the RF and from 72.80% to 77.69% for the XGBoost, respectively. Merging base learners, even those with suboptimal performance, enhances the predictive strength of the multi-class probability model for daily precipitation.

## 4.2 | Comparison of Ensemble-NHMM and NHMM

Atmospheric circulation substantially affects the variability of regional precipitation and the magnitudes of extreme precipitation events under global warming (Zhang et al., 2021). In this section, we focus on evaluating the Ensemble-NHMM and NHMM downscaling models for depicting precipitation variability and extreme precipitation simulations. It is worth noting that the two models have distinctions in how they incorporate atmospheric predictors to influence the conditional hidden state transitional probability. The NHMM, as defined by Equation (3), directly uses the PCs of the atmospheric predictor within the framework of multinomial logistic regression. Conversely, the Ensemble-NHMM introduces an added layer of complexity by leveraging the probabilities of different precipitation levels, derived

from ensemble learning model outputs, as intermediate predictors that are influenced by atmospheric conditions.

### 4.2.1 | Precipitation variabilities

After evaluating the number of hidden states ranging from 2 to 10, the optimum number of hidden states for the two downscaling models was determined to be 6. For each rain gauge, 100 downscaled precipitation events were simulated by the Ensemble-NHMM and NHMM on a daily basis. Comparisons with observations were first conducted for monthly precipitation. Daily precipitation was accumulated on a monthly scale. Next, across the five clusters, we examined the CVRMSE indicator, which was calculated by comparing the monthly precipitation averaged by the rain gauges and the simulations of Ensemble-NHMM and NHMM. Table 4 presents the results calculated from the median of the 100 simulations. For CEC as a whole, the Ensemble-NHMM has a lower error (average CVRMSE of 0.26) concerning monthly precipitation simulations compared with the NHMM (average CVRMSE of 0.45). Furthermore, the temporal variations in the monthly mean precipitation for the five clusters agree more closely with the observations (Figure 5). This suggests that the Ensemble-NHMM can reproduce the seasonality of the precipitation very well over CEC when using the probabilities of different precipitation levels as predictors. The boxplots indicate that the NHMM made significant errors in estimating precipitation in May and June, especially for the West, South and Central clusters.

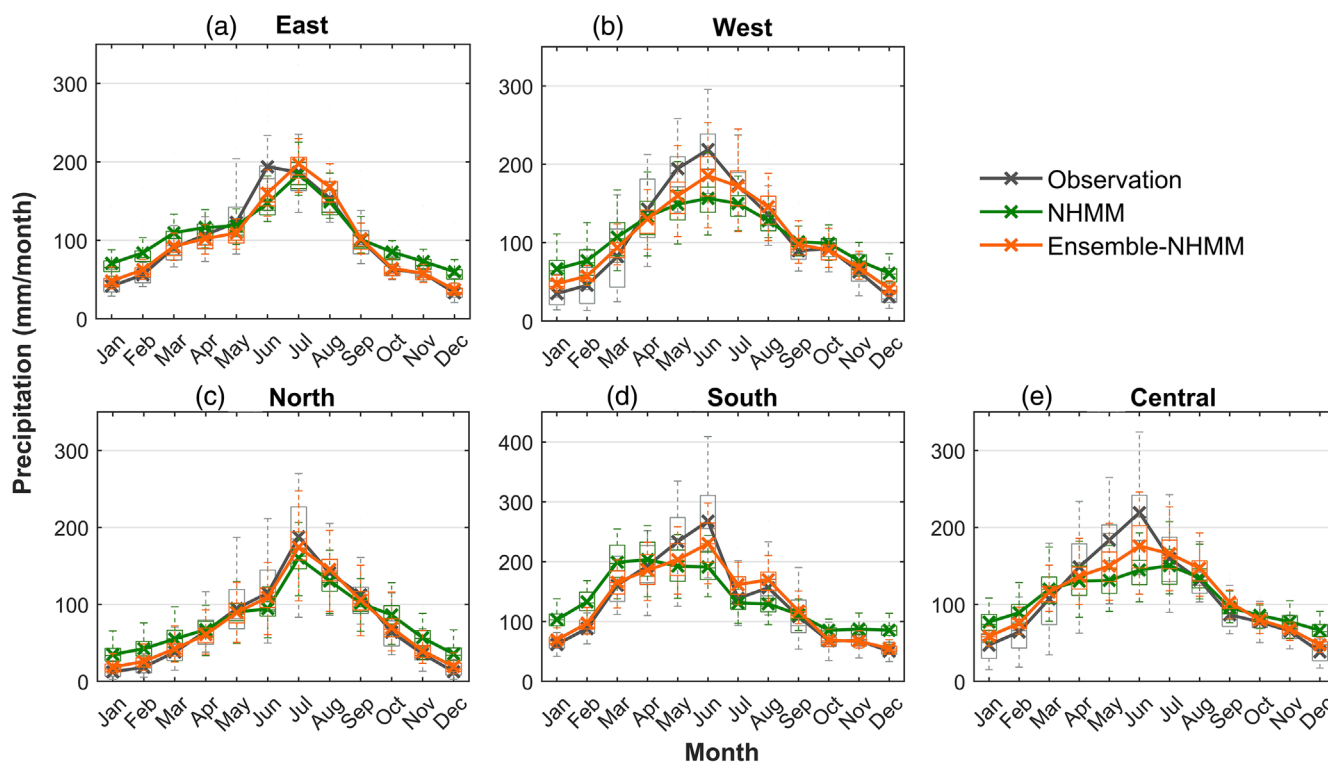
We compared the Ensemble-NHMM and NHMM for annual precipitation simulations accumulated from daily precipitation data. Table 5 summarizes the CVRMSE between the observations and simulations, considering different lengths of moving windows for the annual precipitation. Overall, the Ensemble-NHMM performs better than the NHMM, with a lower CVRMSE. As the CVRMSE tends to decrease with the increasing length of moving windows, time series with 15a weighted was chosen to depict the precipitation change trends more precisely. Figure 6 shows the observed and simulated mean annual precipitation under 15a moving windows for the five clusters, which were averaged by their respective rain gauges. We observed that the Ensemble-NHMM was able to capture the year-to-year and interannual variations in precipitation over CEC, whereas the NHMM tended to overestimate the annual precipitation for the five clusters and failed to capture the peak value in 1997 for the West cluster and in 2000 for the Central cluster.

### 4.2.2 | Extreme precipitation

The 95th percentile of the wet days (daily precipitation  $\geq 0.1$  mm/day) was used to define extreme precipitation for each rain gauge. For the simulations, extreme precipitation was directly calculated when considering the observation thresholds. Subsequently, we obtained the total frequency and amount exceeding the 95th percentile for each year from 1960 to 2015. Figure 7 compares the means of the

**TABLE 4** CVRMSE of the simulated monthly precipitation for the Ensemble-NHMM and NHMM.

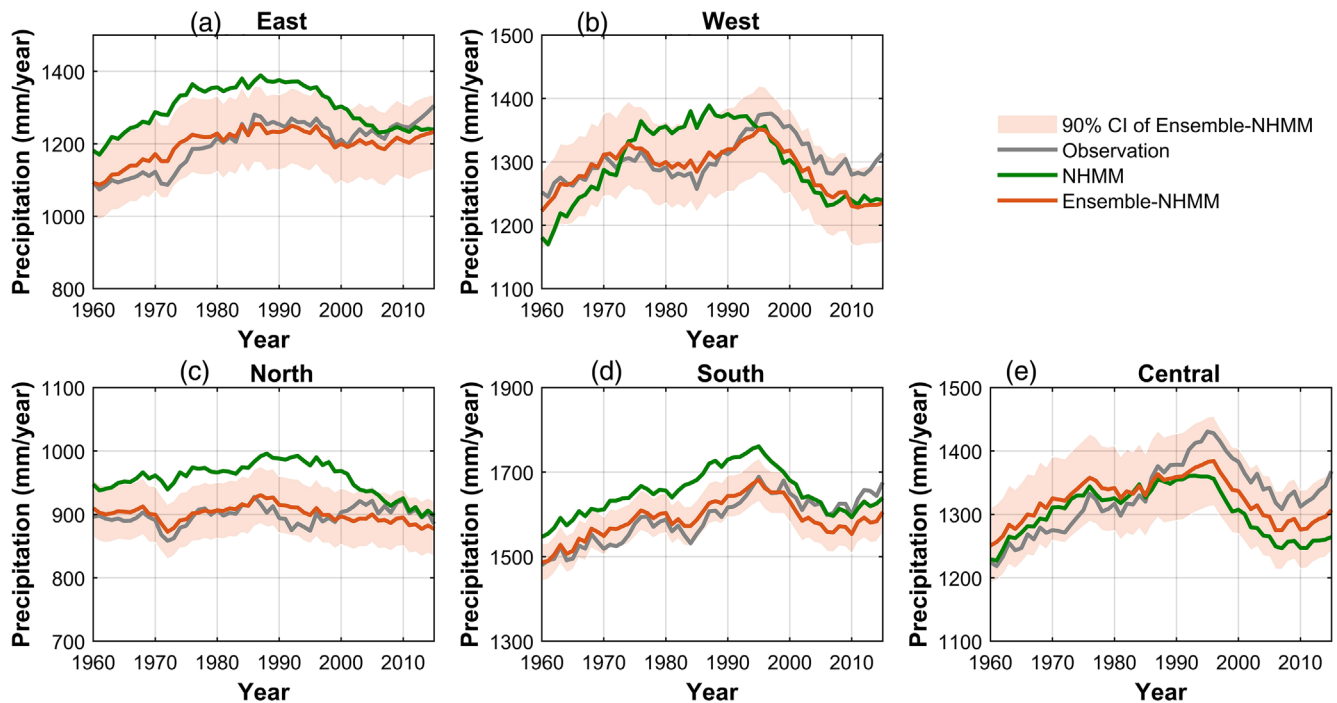
		Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec
Ensemble-NHMM	East	0.37	0.28	0.22	0.25	0.29	0.34	0.26	0.24	0.25	0.36	0.32	0.43
	West	0.35	0.27	0.18	0.23	0.28	0.24	0.25	0.22	0.22	0.27	0.27	0.36
	North	0.41	0.34	0.25	0.31	0.24	0.23	0.22	0.20	0.27	0.27	0.29	0.40
	South	0.22	0.21	0.16	0.17	0.22	0.24	0.23	0.17	0.22	0.31	0.29	0.37
	Central	0.26	0.20	0.18	0.20	0.27	0.27	0.22	0.22	0.22	0.22	0.26	0.32
NHMM	East	0.72	0.52	0.31	0.31	0.29	0.44	0.34	0.39	0.34	0.48	0.41	0.82
	West	0.77	0.63	0.35	0.27	0.35	0.39	0.38	0.34	0.37	0.35	0.38	0.83
	North	1.40	1.05	0.39	0.33	0.27	0.33	0.30	0.31	0.35	0.35	0.54	1.51
	South	0.64	0.52	0.29	0.22	0.29	0.37	0.31	0.34	0.30	0.44	0.40	0.71
	Central	0.59	0.41	0.26	0.27	0.38	0.41	0.33	0.34	0.28	0.32	0.33	0.68

**FIGURE 5** Boxplots of the observed (grey) and simulated (green: NHMM, orange: Ensemble-NHMM) mean monthly precipitation amount for the (a) East, (b) West, (c) North, (d) South and (e) Central clusters. Thick lines represent the average.**TABLE 5** CVRMSE of simulated annual precipitation trends under three periodic moving averages for the Ensemble-NHMM and NHMM.

	Ensemble-NHMM			NHMM		
	1a	5a	15a	1a	5a	15a
East	0.161	0.080	0.047	0.202	0.122	0.100
West	0.151	0.070	0.043	0.161	0.104	0.079
North	0.197	0.088	0.047	0.218	0.115	0.084
South	0.142	0.066	0.036	0.156	0.081	0.054
Central	0.162	0.076	0.047	0.171	0.083	0.057

total extreme precipitation frequency and amount over 56 years for the observations and simulations. The gauges are represented by dots. Referring to the 1:1 lines (dotted lines) and the evaluation results presented in these figures, we confirm the Ensemble-NHMM provides good approximations of the climatology of extremes across all gauges. The NHMM also performs well in reproducing extremes, but there is a relatively worse CC for extreme precipitation frequency. In addition, the Ensemble-NHMM is better at characterizing annual extreme precipitation than the NHMM. The Ensemble-NHMM captures the means of the total extreme precipitation frequency and amount compared to observations.





**FIGURE 6** Observed (grey) and simulated (green: NHMM, orange: Ensemble-NHMM) mean annual precipitation amount under the 15a moving window for the (a) East, (b) West, (c) North, (d) South and (e) Central clusters. Shadow represents the 90% confidence interval (CI) of Ensemble-NHMM.

### 4.3 | Future precipitation projection under different climate change scenarios

#### 4.3.1 | Analysis of the control run

In the above-mentioned analysis, we explored the Ensemble-NHMM in downscaling precipitation using the multi-level daily precipitation occurrence probabilities, which were predicted by the GP at 500 hPa and IVT from ERA5 data. We further analysed how the GCM model, specifically MPI-ESM1-2-HR, performs with the Ensemble-NHMM during the same historical period. All steps of the Ensemble-NHMM were repeated using the MPI-ESM1-2-HR data.

Because the Ensemble-NHMM is a data-driven model, changes in the data source could directly affect the recognition of the hidden states and their most probable sequence, which has a considerable influence on the simulated sequence of daily precipitation. We selected six hidden states and examined the goodness of fit of the hidden state sequence based on MPI-ESM1-2-HR with the results of ERA5. During the period 1960–2015, the consistency rate of hidden state occurrence was as high as 98% (98.62%, 98.56%, 99.69%, 98.01% and 96.63% of the East, West, North, South and Central clusters, respectively), indicating that MPI-ESM1-2-HR can capture the statistical behaviour of actual observations.

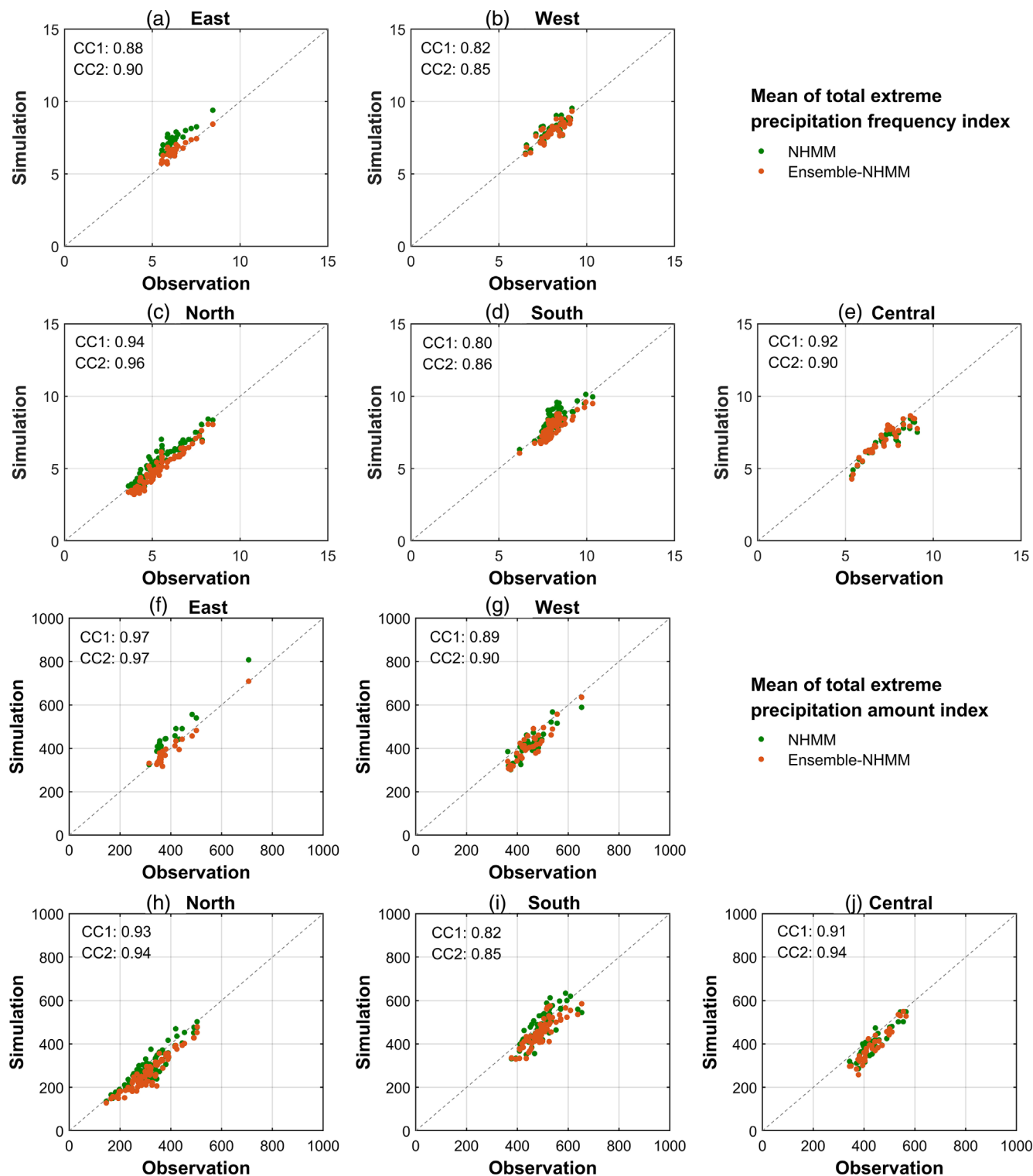
We then evaluated the accuracy of the downscaled precipitation based on MPI-ESM1-2-HR. Tables S1–S3 list the evaluation results of the mean monthly, annual and extreme precipitation simulations for MPI-ESM1-2-HR. Monthly CVRMSE statistics range between 0.22

and 0.59, with an average of 0.34. MPI-ESM1-2-HR performed better in the non-rainy months. The annual CVRMSEs of MPI-ESM1-2-HR were close to those of ERA5. Furthermore, when considering MPI-ESM1-2-HR in the estimations of the means of the total extreme precipitation frequency and amount, the CVRMSEs do not exceed 0.12 and 0.20, respectively. In summary, compared with the results produced by ERA5, the performance of MPI-ESM1-2-HR is satisfactory, and we can fully consider the downscaling model in scenario data.

#### 4.3.2 | Future precipitation projection

Before the projection of precipitation, we must state that there is a basic assumption, that is to say the future atmospheric circulation features of GCMs must respect the statistical consistency of ERA5 reanalysis ones in the historical period (Cioffi et al., 2016). Thus, we restructured the PCs in the period 2016–2100 for the three emission scenarios to be consistent with the mean and variance of those in the historical period, respectively. Subsequently, the restructured first 50 leading PCs of GP at 500 hPa and IVT were used to simulate the multi-level daily precipitation probabilities from 2016 to 2100 based on the ensemble learning model trained by the historical data of MPI-ESM1-2-HR. Thus, we can directly use the Ensemble-NHMM to perform the projections.

Figure 8 shows the mean annual precipitation projections under a 15a moving average window for the five clusters. Under the SSP245 and SSP585 emission scenarios, the annual precipitation over the CEC

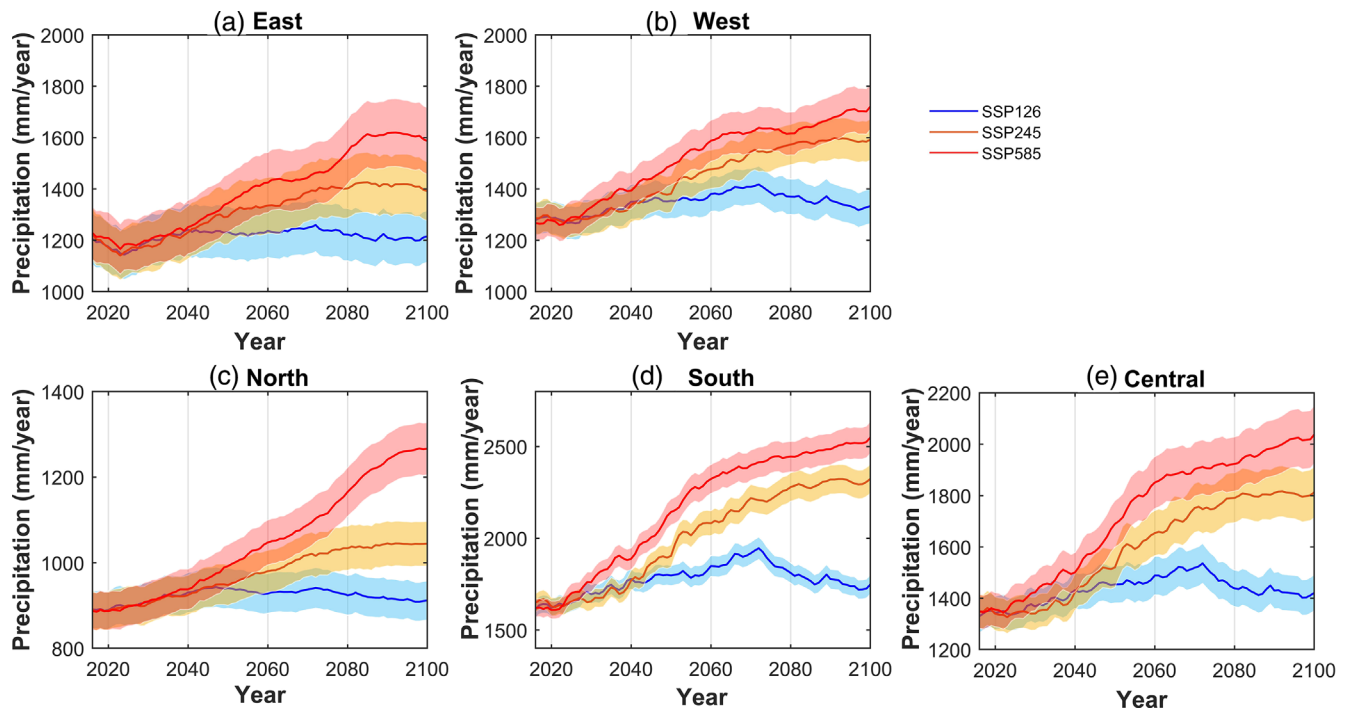


**FIGURE 7** Scatterplots of the means of total extreme precipitation frequency and amount (unit: mm/year) index for observations against the simulations for the (a, f) East, (b, g) West, (c, h) North, (d, i) South and (e, j) Central clusters. Numbers 1 and 2 represent the NHMM and Ensemble-NHMM, respectively.

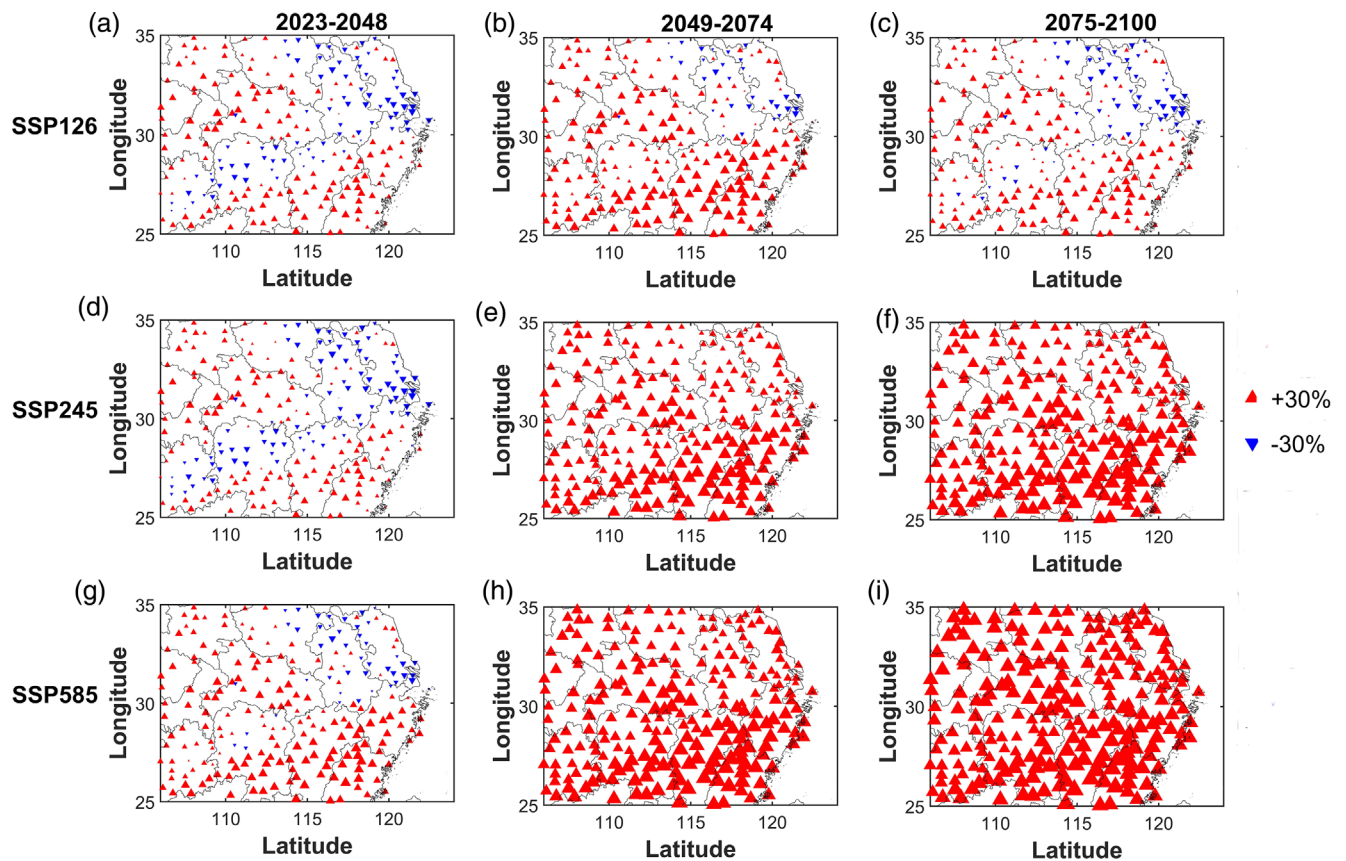
may significantly increase over the next 85 years. The largest boost is projected in 2060–2100 under the SSP585 scenario. While limiting anthropogenic greenhouse gas emissions at the lowest scenario, precipitation over the CEC shows a relatively slight increase from 2016

to 2070, followed by a decreasing trend in the last 30 years of the 21st century.

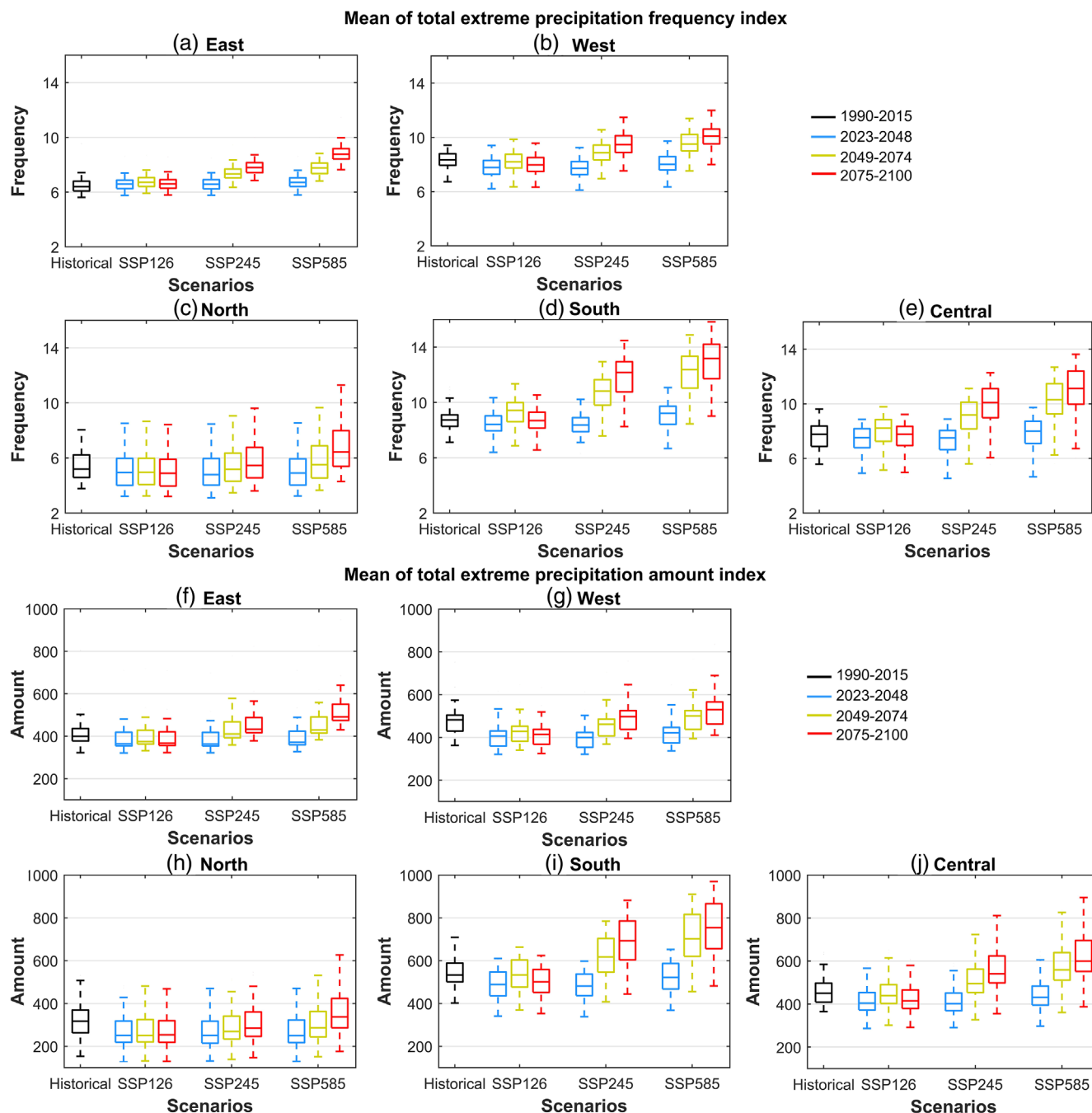
The future scenarios were divided into three sub-periods with a 26-year interval, that is, 2023–2048, 2049–2074 and 2075–2100,



**FIGURE 8** Annual precipitation projections under 15a moving average window for the (a) East, (b) West, (c) North, (d) South and (e) Central clusters. Shadows represent the 90% confidence intervals.



**FIGURE 9** Relative changes (%) of the projected mean annual precipitation amount in the early, mid and late 21st centuries under different emission scenarios compared with the historical period 1990–2015.



**FIGURE 10** Boxplots of the means of the total extreme precipitation frequency and amount (unit: mm/year) for the historical period (1990–2015) and the future projections for three sub periods (early, mid and late 21st centuries) under different warming scenarios. (a, f) East, (b, g) West, (c, h) North, (d, i) South and (e, j) Central clusters.

representing the early, mid and late 21st century, respectively. Projected changes in precipitation are presented relative to the historical 26-year period (1990–2015). Figure 9 shows the projected relative changes in the mean annual precipitation. Positive and negative values indicate more and less precipitation in the projection period than in the historical period, respectively. In the early 21st century, under the three emission scenarios, the mean annual precipitation amount indicates a slight drying tendency over the north-eastern CEC and sites in the centre (Figure 9a,d,g). The projected decrease remained over the

north-eastern CEC in the mid- and late century under the SSP126 scenario (Figure 9b,c). However, over the southern and western parts of CEC, a slightly wetter condition is projected in the early 21st century under the three scenarios. Furthermore, this wetter tendency is expected to be pronounced in the mid- and late periods, especially under the SSP585 scenario, with an increasing percentage beyond 30% (Figure 9f,i).

Figure 10 shows boxplots of the means of the total extreme precipitation frequency and amount index for historical and future

**TABLE 6** Relative changes (%) of the projected means of the total extreme precipitation frequency and amount in the early, mid and late 21st centuries under different emission scenarios compared with the historical period 1990–2015.

		2023–2048			2049–2074			2075–2100		
		SSP126	SSP245	SSP585	SSP126	SSP245	SSP585	SSP126	SSP245	SSP585
Frequency	East	1.78	1.75	3.61	4.24	<b>13.52</b>	<b>19.76</b>	2.17	<b>20.07</b>	<b>35.73</b>
	West	<b>-6.11</b>	<b>-7.07</b>	-2.65	-0.74	<b>6.67</b>	<b>15.33</b>	-3.54	<b>14.75</b>	<b>21.93</b>
	North	-7.75	-8.28	-7.68	-6.63	-1.08	4.84	-7.86	4.20	<b>23.22</b>
	South	-2.87	-3.95	4.04	<b>6.44</b>	<b>22.41</b>	<b>38.79</b>	-0.56	<b>35.70</b>	<b>48.07</b>
	Central	-4.26	-5.05	1.68	3.70	<b>16.87</b>	<b>32.31</b>	-1.81	<b>28.75</b>	<b>42.82</b>
Amount	East	-5.50	-5.58	-3.68	-3.13	5.44	11.28	-5.12	11.48	<b>26.21</b>
	West	<b>-13.65</b>	<b>-14.50</b>	<b>-10.41</b>	-8.64	-1.78	6.17	<b>-11.24</b>	5.68	<b>12.32</b>
	North	<b>-14.91</b>	<b>-15.42</b>	<b>-14.86</b>	<b>-13.88</b>	-8.75	-3.42	<b>-15.06</b>	-3.90	<b>16.61</b>
	South	<b>-9.69</b>	<b>-10.68</b>	-3.19	-0.89	<b>14.19</b>	<b>29.62</b>	<b>-7.55</b>	<b>26.68</b>	<b>38.33</b>
	Central	<b>-10.60</b>	-11.32	-4.95	-3.06	9.25	<b>23.85</b>	-8.31	<b>20.54</b>	<b>33.82</b>

Note: Bold fonts represent statistically significant changes at the 95% confidence level.

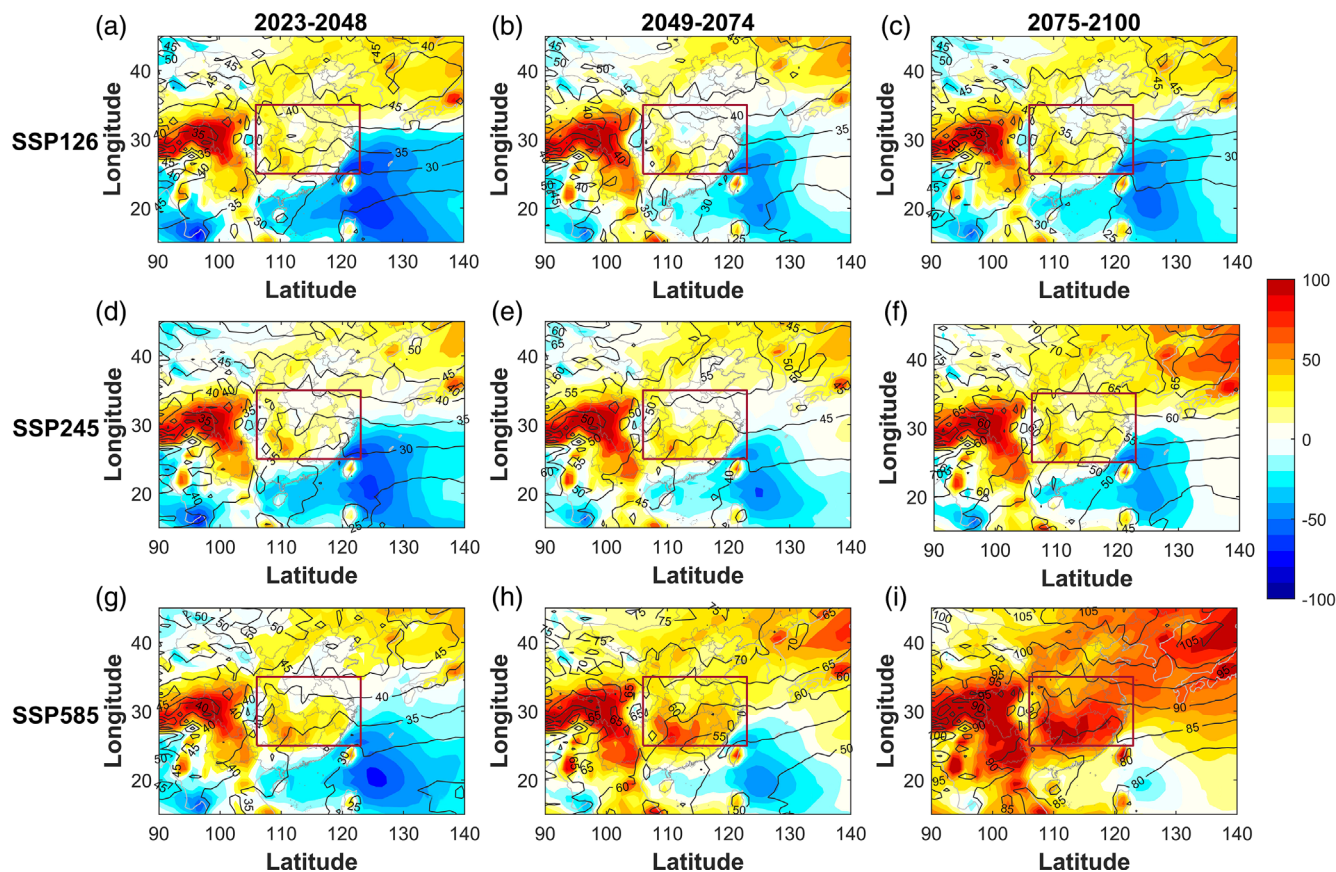
projections under different warming scenarios. The relative changes for the early, mid and late 21st centuries compared with the historical period of 1990–2015 are summarized in Table 6. Bold font indicates statistically significant changes at the 95% confidence level. In the early 21st century, the projected changes in the mean of the total extreme precipitation frequency show regional differences under the three scenarios. Except for a slight increase found for the east cluster, the remaining regions are likely to face less frequent extreme precipitation. Meanwhile, weakened extreme precipitation intensity is found over CEC. In the the North cluster, the negative changes are projected to decrease by -14.91%, -15.42% and -14.86% under the SSP126, SSP245 and SSP585 scenarios, respectively. By the mid-21st century, projected changes in extreme precipitation in most areas switch from negative (decreasing) to positive (increasing). Except for the North cluster, the mean of the total extreme precipitation frequency for the remaining clusters is projected to increase under the SSP245 and SSP585 scenarios. It produces the largest magnitude ranging from 15.33% to 38.79% in the worst scenario. For example, the projected mean extreme frequency of the South cluster increases to 11.89 times per year under the SSP585 scenario, which is higher than that of the historical period (8.76 per year). However, under the SSP126 scenario, the projected extreme amount may decrease by -0.89% to -13.88% for the five clusters. In the late century, the projected increases in precipitation extremes over CEC are expected to be larger than those in the early- and mid-periods. The means of the total extreme precipitation frequency and amount significantly increase by 21.93%–48.07% and 12.32%–38.33% respectively, in the scenario with the highest emission. The change rates of the North and West clusters are relatively smaller than those of the East, South and Central clusters. For the most severe cluster, the mean extreme precipitation amount can increase to 751.50 mm per year in the worst scenario. As expected, the projected extreme precipitation shows a continuous decrease over CEC under the SSP126 scenario.

Our results show that the annual and extreme precipitation over CEC have increased under global warming. Most regions will experience unprecedented precipitation and extremes in the late 21st century under the scenario with the highest emission. This increase is expected to be more severe over the southern CEC.

In general, changes in regional precipitation due to global warming can be decomposed into two contributors (Pfahl et al., 2017): (1) thermodynamic effects and (2) dynamic forcing. First, the enhanced atmospheric water-holding capacity resulting from increasing global temperature that approximately in accordance with the theoretical rate of the Clausius–Clapeyron equation, could increase the water vapour content and affect the magnitude of the precipitation (Gorman & Schneider, 2009). On the other hand, global warming leads to changes in atmospheric circulation characteristics relevant to specific geographic precipitation patterns (Mindlin et al., 2021; Vallis et al., 2015). These changes in the circulation system can further influence moisture transport (Liu et al., 2020). For example, Yang et al. (2022) showed that the strengthening of the WPSH under global warming could cause more moisture transport from the ocean to East China and induce heavy precipitation.

Figure 11 shows the anomaly fields of GP at 500 hPa and IVT during the rainy season (April–September) in the early, mid and late 21st centuries compared with the historical period 1990–2015. The figure shows that the GP at 500 hPa over CEC tends to increase due to global warming, but the contour lines of the anomaly fields denote an inhomogeneous feature in the horizontal direction during different sub-periods. Significant changes in GP at 500 hPa and IVT anomalies between the three scenarios can be detected only in the mid-21st century. Based on Tebaldi and Friedlingstein (2013)'s research, significant climate effects of different emission pathways cannot be detected very quickly; this process requires 30–45 years at the regional scale. In addition, ridge lines with enhanced pressure gradient located in the coastal areas would promote water vapour transport from the ocean to the south CEC, especially under the SSP585





**FIGURE 11** Under three emission scenarios, the anomalies fields of GP at 500 hPa (contours, unit: gpm) and IVT (shadow, unit:  $\text{kg m}^{-1} \text{s}^{-1}$ ) during the rainy season in the early, mid, and late 21st centuries compared with the historical period 1990–2015.

scenario. It is clear that in the mid- and late centuries, under different emission scenarios, the south of CEC would receive stronger water vapour transport than the north, which helps explain why the annual precipitation and precipitation extremes are expected to be more severe over the South cluster.

## 5 | DISCUSSION AND CONCLUSIONS

In this study, we proposed a new two-step precipitation hierarchical downscaling framework consisting of an ensemble learning model that combines XGBoost and RF to predict the probability of multi-level daily precipitation events for an ensemble of clustered rain gauges. Predicted probabilities were used as predictors for the NHMM to conduct precipitation downscaling at the scales of the individual rain gauges. The downscaling model, that is, Ensemble-NHMM, projects future trends in the daily precipitation amount in central-eastern China under different SSPs. The conclusions can be summarized as follows:

1. Our proposed ensemble learning model shows a superior capability in predicting the occurrence probabilities of multi-class daily precipitation. It obtains better prediction skills than the individual machine learning of RF and XGBoost.
2. Evaluation results show that the occurrence probability of different precipitation levels is a good predictor for the NHMM model. The Ensemble-NHMM performs well in simulating annual and extreme precipitation compared with the traditional NHMM, which directly uses the sequence of standardized IVTs and the first leading 50 PCs of GP at 500 hPa as atmospheric predictors. In previous studies using NHMM, many statistical indicators and processes were proposed to find the optimum predictors that are physically sensible in regional precipitation processes (Chen et al., 2018; Pineda & Willems, 2016), and this task often required intensive computational resources. Here we identified just two predictors, IVT and GP, whose reliability has been well demonstrated in representing the main features of atmospheric conditions affecting precipitation. Furthermore, the proposed Ensemble-NHMM is more suitable than NHMM in representing the possible nonlinearities in the link between daily precipitation and atmospheric circulation features, since such nonlinearity can be captured by the ensemble learning model. The multinomial logistic regression (Equation 3), is thus just used to represent the nearly linear link between multi-level daily precipitation probabilities of homogeneous regions and daily rainfall amount of single rain gauge belonging to the cluster.
3. We examined one of the CMIP6 models (MPI-ESM1-2-HR) for precipitation downscaling via the Ensemble-NHMM during 1960–

2015. Based on the statistical metrics, the downscaled results of MPI-ESM1-2-HR produce reliable trend simulations.
4. Based on the precipitation projection results of the MPI-ESM1-2-HR, central-eastern China may receive more precipitation in the period of 2016–2100, especially under the SSP245 and SSP585 emission scenarios. However, in the SSP126 scenario, the mean annual precipitation would decrease in the last 30 years of the 21st century. Furthermore, compared with ~26 years (1990–2015), a slight drying tendency is projected over the north-eastern CEC under the three emission scenarios. The extreme precipitation frequency and amount will increase throughout the century (2016–2100). In the late century, the extreme precipitation frequency and amount may significantly increase by 21.9%–48.1% and 12.3%–38.3%, respectively, in the case of the worst emission scenario.

Certainly, in our proposed downscaling framework, there is room for improvements. First of all, for the ensemble learning model, it shows a superior capability in classifying no rain or rain and capturing moderate precipitation events, but improvements need to be thought to increase the accuracy in detecting the regional heavy precipitation events. In addition, we assumed that the leading features of future atmospheric circulation remain the same, only changes in frequency and intensity. In the next study, the non-stationarity of precipitation will be considered in the future projections, attempting to introduce additional non-homogeneity into the emission distribution as suggested in Holsclaw et al. (2017).

## ACKNOWLEDGEMENTS

This research was funded by the National Social Science Foundation of China (Grant No. 18ZDA105), the National Natural Science Foundation of China (Grant No. 41971199), the Director's Fund of Key Laboratory of Geographic Information Science (Ministry of Education), East China Normal University (Grant No. KLGIS2022C01), and the dual doctoral degree program in geography and environmental and hydraulics engineering between East China Normal University and 'La Sapienza' University of Roma.

## DATA AVAILABILITY STATEMENT

Daily precipitation records are available at <http://data.cma.cn/>. ERA5 data are downloaded at <https://www.ecmwf.int/en/forecasts/datasets/reanalysis-datasets/era5>. GCM data are available at <https://esgf-node.llnl.gov/projects/cmip6/NHMM> toolbox is available at <http://www.sergeykirshner.com/software/mvnhmm>.

## ORCID

Qin Jiang  <https://orcid.org/0000-0001-6401-9067>

Federico Rosario Conticello  <https://orcid.org/0000-0003-0614-0656>

## REFERENCES

Ahmed, K., Sachindra, D., Shahid, S., Iqbal, Z., Nawaz, N., & Khan, N. (2020). Multi-model ensemble predictions of precipitation and

- temperature using machine learning algorithms. *Atmospheric Research*, 236, 104806.
- Bates, B. C., Chandler, R. E., Charles, S. P., & Campbell, E. P. (2010). Assessment of apparent nonstationarity in time series of annual inflow, daily precipitation, and atmospheric circulation indices: A case study from southwest Western Australia. *Water Resources Research*, 46(3), W00H02.
- Bengio, Y., Mori, R. D., Flammia, G., & Kompe, R. (1992). Global optimization of a neural network-hidden Markov model hybrid. *IEEE Transactions on Neural Networks*, 3(2), 252–259.
- Blondel, V. D., Guillaume, J.-L., Lambiotte, R., & Lefebvre, E. (2008). Louvain method: Finding communities in large networks. *Physics and Society*, P10008.
- Boers, N., Goswami, B., Rheinwalt, A., Bookhagen, B., Hoskins, B., & Kurths, J. (2019). Complex networks reveal global pattern of extreme-rainfall teleconnections. *Nature*, 566(7744), 373–377.
- Breiman, L. (2001). Random forests. *Machine Learning*, 45, 5–32.
- Charles, S. P., Bates, B. C., Smith, I. N., & Hughes, J. P. (2004). Statistical downscaling of daily precipitation from observed and modelled atmospheric fields. *Hydrological Processes*, 18(8), 1373–1394.
- Chen, J., Chen, H., & Guo, S. L. (2018). Multi-site precipitation downscaling using a stochastic weather generator. *Climate Dynamics*, 50(5), 1975–1992.
- Chen, T. Q., & Guestrin, C. (2016). XGBoost: A scalable tree boosting system. In *Proceedings of the 22nd ACM SIGKDD international conference on knowledge discovery and data mining* (pp. 785–794). Association for Computing Machinery.
- Christensen, O. B., & Kjellström, E. (2020). Partitioning uncertainty components of mean climate and climate change in a large ensemble of European regional climate model projections. *Climate Dynamics*, 54(9), 4293–4308.
- Cioffi, F., Conticello, F., & Lall, U. (2016). Projecting changes in Tanzania rainfall for the 21st century. *International Journal of Climatology*, 36(13), 4297–4314.
- Cioffi, F., Conticello, F., Lall, U., Marotta, L., & Telesca, V. (2017). Large scale climate and rainfall seasonality in a Mediterranean area: Insights from a non-homogeneous Markov model applied to the Agro-Pontino plain. *Hydrological Processes*, 31(3), 668–686.
- Cioffi, F., Conticello, F. R., & Lall, U. (2020). Stochastic scenarios for 21st century rainfall seasonality, daily frequency, and intensity in South Florida. *Journal of Water Resources Planning and Management*, 146(8), 04020058.
- Conticello, F. R., Cioffi, F., Lall, U., & Merz, B. (2020). Synchronization and delay between circulation patterns and high streamflow events in Germany. *Water Resources Research*, 56(4), e2019WR025598.
- Conticello, F. R., Cioffi, F., Merz, B., & Lall, U. (2018). An event synchronization method to link heavy rainfall events and large-scale atmospheric circulation features. *International Journal of Climatology*, 38(3), 1421–1437.
- Ding, Y. H., Liang, P., Liu, Y. J., & Zhang, Y. C. (2020). Multiscale variability of Meiyu and its prediction: A new review. *Journal of Geophysical Research: Atmospheres*, 125(7), e2019JD031496.
- Ghamghami, M., & Bazrafshan, J. (2021). Relationships between large-scale climate signals and winter precipitation amounts and patterns over Iran. *Journal of Hydrologic Engineering*, 26(3), 05021001.
- Ghamghami, M., Ghahreman, N., Olya, H., & Ghasdi, T. (2019). Comparison of three multi-site models in stochastic reconstruction of winter daily rainfall over Iran. *Modeling Earth Systems and Environment*, 5(4), 1319–1332.
- Gorman, P. A., & Schneider, T. (2009). The physical basis for increases in precipitation extremes in simulations of 21st-century climate change. *Proceedings of the National Academy of Sciences of the United States of America*, 106(35), 14773–14777.
- Guo, L. Y., Jiang, Z. H., Li, L., & Wang, H. J. (2022). Increase of future summer rainfall in the middle and lower reach of the Yangtze River basin projected with a nonhomogeneous hidden Markov model. *Geophysical Research Letters*, 49(7), e2021GL097325.

- Hassan, M. R., Nath, B., & Kirley, M. (2007). A fusion model of HMM, ANN and GA for stock market forecasting. *Expert Systems with Applications*, 33(1), 171–180.
- Henny, L., Thorncroft, C. D., & Bosart, L. F. (2023). Changes in seasonal large-scale extreme precipitation in the mid-Atlantic and Northeast United States, 1979–2019. *Journal of Climate*, 36(4), 1017–1042.
- Her, Y., Yoo, S.-H., Cho, J., Hwang, S., Jeong, J., & Seong, C. (2019). Uncertainty in hydrological analysis of climate change: Multi-parameter vs. multi-GCM ensemble predictions. *Scientific Reports*, 9(1), 4974.
- Holsclaw, T., Greene, A. M., Robertson, A. W., & Smyth, P. (2017). Bayesian nonhomogeneous Markov models via Pólya-Gamma data augmentation with applications to rainfall modeling. *The Annals of Applied Statistics*, 11(1), 393–426.
- Huang, X. Y., Swain, D. L., Walton, D. B., Stevenson, S., & Hall, A. D. (2020). Simulating and evaluating atmospheric river-induced precipitation extremes along the U.S. Pacific Coast: Case studies from 1980–2017. *Journal of Geophysical Research: Atmospheres*, 125(4), e2019JD031554.
- Hughes, J. P., Guttorp, P., & Charles, S. P. (1999). A non-homogeneous hidden Markov model for precipitation occurrence. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, 48(1), 15–30.
- IPCC. (2021). Climate change 2021: The physical science basis. In V. Masson-Delmotte, P. Zhan, A. Pirani, S. L. Connors, C. Péan, S. Berger, N. Caud, Y. Chen, L. Goldfarb, M. I. Gomis, M. Huang, K. Leitzell, E. Lonnoy, J. B. R. Matthews, T. K. Maycock, T. Waterfield, O. Yelekçi, R. Yu, & B. Zhou (Eds.), *Contribution of working group I to the sixth assessment report of the intergovernmental panel on climate change*. Cambridge University Press in press.
- Jiang, Q., Cioffi, F., Giannini, M., Wang, J., & Li, W. Y. (2022). Analysis of changes in large-scale circulation patterns driving extreme precipitation events over the central-eastern China. *International Journal of Climatology*, 43(1), 519–537.
- Kang, I.-S., Yang, Y.-M., & Tao, W.-K. (2015). GCMs with implicit and explicit representation of cloud microphysics for simulation of extreme precipitation frequency. *Climate Dynamics*, 45(1), 325–335.
- Liu, B. J., Tan, X. Z., Gan, T. Y., Chen, X. H., Lin, K. R., Lu, M. Q., & Liu, Z. Y. (2020). Global atmospheric moisture transport associated with precipitation extremes: Mechanisms and climate change impacts. *WIREs Water*, 7(2), e1412.
- Liu, Z. F., Xu, Z. X., Charles, S. P., Fu, G. B., & Liu, L. (2011). Evaluation of two statistical downscaling models for daily precipitation over an arid basin in China. *International Journal of Climatology*, 31(13), 2006–2020.
- Meinshausen, M., Nicholls, Z. R. J., Lewis, J., Gidden, M. J., Vogel, E., Freund, M., Beyerle, U., Gessner, C., Nauels, A., Bauer, N., Canadell, J. G., Daniel, J. S., John, A., Krummel, P. B., Luderer, G., Meinshausen, N., Montzka, S. A., Rayner, P. J., Reimann, S., ... Wang, R. H. J. (2020). The shared socio-economic pathway (SSP) greenhouse gas concentrations and their extensions to 2500. *Geoscientific Model Development*, 13(8), 3571–3605.
- Mindlin, J., Shepherd, T. G., Vera, C., & Osman, M. (2021). Combined effects of global warming and ozone depletion/recovery on southern hemisphere atmospheric circulation and regional precipitation. *Geophysical Research Letters*, 48(12), e2021GL092568.
- Moon, T. K. (1996). The expectation-maximization algorithm. *IEEE Signal Processing Magazine*, 13(6), 47–60.
- Newman, M. E. (2012). Communities, modules and large-scale structure in networks. *Nature Physics*, 8(1), 25–31.
- Neykov, N., Neytchev, P., Zucchini, W., & Hristov, H. (2012). Linking atmospheric circulation to daily precipitation patterns over the territory of Bulgaria. *Environmental and Ecological Statistics*, 19(2), 249–267.
- Norris, J., Hall, A., Chen, D., Thackeray, C. W., & Madakumbura, G. D. (2021). Assessing the representation of synoptic variability associated with California extreme precipitation in CMIP6 models. *Journal of Geophysical Research: Atmospheres*, 126(6), e2020JD033938.
- Panda, K. C., Singh, R. M., Thakural, L. N., & Sahoo, D. P. (2022). Representative grid location-multivariate adaptive regression spline (RGL-MARS) algorithm for downscaling dry and wet season rainfall. *Journal of Hydrology*, 605, 127381.
- Pfahl, S., O’Gorman, P. A., & Fischer, E. M. (2017). Understanding the regional pattern of projected future changes in extreme precipitation. *Nature Climate Change*, 7(6), 423–427.
- Pineda, L. E., & Willems, P. (2016). Multisite downscaling of seasonal predictions to daily rainfall characteristics over Pacific–Andean river basins in Ecuador and Peru using a nonhomogeneous hidden Markov model. *Journal of Hydrometeorology*, 17(2), 481–498.
- Rahman, M., Chen, N. S., Elbeltagi, A., Islam, M. M., Alam, M., Pourghasemi, H. R., Tao, W., Zhang, J., Shufeng, T., Faiz, H., Baig, M. A., & Dewan, A. (2021). Application of stacking hybrid machine learning algorithms in delineating multi-type flooding in Bangladesh. *Journal of Environmental Management*, 295, 113086.
- Sammen, S. S., Mohammed, T. A., Ghazali, A. H., Sidek, L. M., Shahid, S., Abba, S. I., Malik, A., & Al-Ansari, N. (2022). Assessment of climate change impact on probable maximum floods in a tropical catchment. *Theoretical and Applied Climatology*, 148(1), 15–31.
- Shahriar, S. A., Siddique, M. A. M., & Rahman, S. M. A. (2021). Climate change projection using statistical downscaling model over Chittagong Division, Bangladesh. *Meteorology and Atmospheric Physics*, 133(4), 1409–1427.
- Siabi, E. K., Kabobah, A. T., Akpoti, K., Anornu, G. K., Amo-Boateng, M., & Nyantakyi, E. K. (2021). Statistical downscaling of global circulation models to assess future climate changes in the Black Volta basin of Ghana. *Environmental Challenges*, 5, 100299.
- Singhal, A., Jaseem, M., & Jha, S. K. (2023). Spatial connections in extreme precipitation events obtained from NWP forecasts: A complex network approach. *Atmospheric Research*, 282, 106538.
- Tabari, H. (2021). Extreme value analysis dilemma for climate change impact assessment on global flood and extreme precipitation. *Journal of Hydrology*, 593, 125932.
- Tang, Y., Huang, A. N., Wu, P., Huang, D. Q., Xue, D. K., & Wu, Y. (2021). Drivers of summer extreme precipitation events over East China. *Geophysical Research Letters*, 48(11), e2021GL093670.
- Tebaldi, C., & Friedlingstein, P. (2013). Delayed detection of climate mitigation benefits due to climate inertia and variability. *Proceedings of the National Academy of Sciences of the United States of America*, 110(43), 17229–17234.
- Toimil, A., Losada, I. J., Nicholls, R. J., Dalrymple, R. A., & Stive, M. J. F. (2020). Addressing the challenges of climate change risks and adaptation in coastal areas: A review. *Coastal Engineering*, 156, 103611.
- Trentin, E., & Gori, M. (2001). A survey of hybrid ANN/HMM models for automatic speech recognition. *Neurocomputing*, 37(1), 91–126.
- Vallis, G. K., Zurita-Gotor, P., Cairns, C., & Kidston, J. (2015). Response of the large-scale structure of the atmosphere to global warming. *Quarterly Journal of the Royal Meteorological Society*, 141(690), 1479–1501.
- Vousdoukas, M. I., Mentaschi, L., Voukouvalas, E., Bianchi, A., Dottori, F., & Feyen, L. (2018). Climatic and socioeconomic controls of future coastal flood risk in Europe. *Nature Climate Change*, 8(9), 776–780.
- Wang, H.-M., Chen, J., Xu, C.-Y., Zhang, J. K., & Chen, H. (2020). A framework to quantify the uncertainty contribution of GCMs over multiple sources in hydrological impacts of climate change. *Earth’s Futures*, 8(8), e2020EF001602.
- Wilby, R. L., & Wigley, T. M. L. (1997). Downscaling general circulation model output: A review of methods and limitations. *Progress in Physical Geography: Earth and Environment*, 21(4), 530–548.
- Yang, K., Cai, W. J., Huang, G., Hu, K. M., Ng, B., & Wang, G. J. (2022). Increased variability of the western Pacific subtropical high under greenhouse warming. *Proceedings of the National Academy of Sciences*, 119(23), e2120335119.

- Yang, X. L., Yu, X. H., Wang, Y. Q., He, X. G., Pan, M., Zhang, M. R., Liu, Y., Ren, L. L., & Sheffield, J. (2020). The optimal multimodel ensemble of bias-corrected CMIP5 climate models over China. *Journal of Hydrometeorology*, 21(4), 845–863.
- Zha, X. N., Xiong, L. H., Chen, J., Kim, J.-S., Liu, D. D., & Wang, G. S. (2022). Estimating multisite precipitation by a stepwise NHMM-VAR model considering the spatiotemporal correlations of precipitation amounts. *Journal of Hydrology*, 612, 128065.
- Zhang, W. X., Furtado, K., Wu, P. L., Zhou, T. J., Chadwick, R., Marzin, C., Rostron, J., & Sexton, D. (2021). Increasing precipitation variability on daily-to-multiyear time scales in a warmer world. *Science Advances*, 7(31), eabf8021.
- Zhou, Z. H. (2021). Ensemble learning. In Z. H. Zhou (Ed.), *Machine learning* (pp. 181–210). Springer Singapore.

## SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

**How to cite this article:** Jiang, Q., Cioffi, F., Conticello, F. R., Giannini, M., Telesca, V., & Wang, J. (2023). A stacked ensemble learning and non-homogeneous hidden Markov model for daily precipitation downscaling and projection. *Hydrological Processes*, 37(9), e14992. <https://doi.org/10.1002/hyp.14992>