



An evaluation of statistical and deep learning-based correction of monthly precipitation over the Yangtze River basin in China based on CMIP6 GCMs

An He¹ · Chao Wang¹  · Lei Xu² · Peng Wang³ · Wei Wang¹ · Nengcheng Chen^{1,2} · Zeqiang Chen²

Received: 21 July 2023 / Accepted: 30 April 2024
© The Author(s), under exclusive licence to Springer Nature B.V. 2024

Abstract

Precipitation, as one of the most indispensable and significant material resources for human survival and society development. Compared with observed data, numerical simulation precipitation products, exemplified by Coupled Model Intercomparison Project Phase 6 (CMIP6), provide a better way to understand past, present, and future climate changes, and have been widely used in hydrological, ecological, and environmental fields. However, the precipitation data of CMIP6 often contains a high degree of uncertainty and bias. This study corrected the historical monthly precipitation data over the Yangtze River Basin (YRB) based on four Global climate models (GCMs) from the CMIP6, by using one statistic method, principal component analysis (PCA), and two deep learning methods, deep neural network (DNN) method, and the long short-term memory (LSTM) based method. The revised monthly precipitation was analyzed temporally and spatially, and the results showed that the LSTM-based method is the most appropriate and gained better performance of correlation, root mean square error (RMSE) and relative bias(RBIAS). Then the precipitation of future scenes was also corrected by the LSTM-based method, the shortcoming of original CMIP6 data in overestimating rainfall has been significantly improved by reducing the peak of extreme rainfall. This result helps us to get a better understanding of the status and trends of rainfall data. Besides, this study not only takes advantage of the fact that LSTM is inherently suitable for long time-series data, but also takes into account spatial heterogeneity by dividing the study area into grids and the data in each grid were revised individually. Therefore, the method is flexible and can be widely used in other regions.

Keywords CMIP6 · YRB · LSTM · Precipitation · Correction method

Extended author information available on the last page of the article

1 Introduction

Global warming has been influencing regional climate change over the past century, and the effect has mainly been on the water cycle, resulting in significant extreme rainfall, leading to natural ecological and socioeconomic losses (Peng et al., 2022). This impact is progressively accumulated, leading to frequent extreme weather events (Jia et al., 2023; Molly E. Brown & Funk, 2008; Wang et al., 2018). Therefore, it is of great theoretical significance and application value to explore changes in historical precipitation patterns, and simulate and predict future precipitation events, especially extreme rainfall (Wang et al., 2023b). In particular, accurate precipitation data for future scenarios are essential for attribution analysis, disaster prevention and mitigation, and decision-making (Adhikari et al., 2015; Huang et al., 2023; Xu et al., 2021b).

There are estimated 30,000 meteorological stations globally to monitor various meteorological factors in real-time, according to the National Aeronautics and Space Administration (NASA). Under current climate change trends, extreme weather events will become more frequent, last longer, and be more widespread, such as heat waves, heavy precipitation (Gosai et al., 2009). However, meteorological stations can only collect historical and current meteorological data and cannot outline future climate change trends (Ehteram et al., 2023).

To assess the status of future climate by examining current and historical weather patterns, the World Climate Research Program (WCRP) launched the International Coupled Model Intercomparison Program (CMIP) in 1995 and has promoted the second to sixth CMIP in the past 25 years. The program is an integral part of the scientific assessment of the Intergovernmental Panel on Climate Change (IPCC), focusing on the evaluation of climate model simulation capabilities and future climate change scenarios (Eyring et al., 2016). Global climate models (GCMs) are core components of the CMIP and are essential for understanding the history and for predicting future climate change scenarios and impacts (Song et al., 2019). There are 23 CMIP6-endorsed model intercomparison projects organized and designed by different countries (Gusain et al., 2020). The GCM is an important tool for understanding the history and predicting potential future climate change. It is also the main way to predict future climate change and its impacts at this stage by studying the connection and evolution of the Earth's spheres through numerical simulation (Stevens, 2012). However, a series of studies have shown that there are large uncertainties and biases in CMIP6 climate simulation data due to scenarios, observations, model parameters, and various conditions in the process of climate simulation (Guo et al., 2011). Due to the geographical features such as altitude and topography between regions, the performance of simulated data varies at different regional scales (Masud et al., 2021). Unlike the temperature data output from the model, the precipitation data from CMIP6 are not directly simulated but are secondary diagnostic data of the simulation results, and their bias and uncertainty are larger (OFFICE, 2011).

The correction of meteorological data is more often done using the variational approach, whereby errors are estimated based on the field of analysis and thus the model data are corrected. There are also methods to establish statistical relationships to revise data based on the regressions of the historical observed data, such as the perfect prognostic method and the simulation model output statistic method (Klein, 1971). However, the disadvantages of these methods are the insufficient use of information from the original historical data, and they do not consider the nonlinear interactions between the internal and external errors of

the model and the spatial correlation characteristics of the error field (Danforth & Kalnay, 2008; Mohamadi et al., 2022). This is a correction to the results based on existing simulation model data, also known as state-dependent correction (Danforth et al., 2007). A common method of precipitation data revision is principal component analysis (PCA), which extracts the principal components of the observed and model data. It was the representative statistical method, and have been achieved good results in a number of studies. It uses a regression relationship between the two principal components. Each observed principal component can be obtained by regressing the principal components of its model field. The prediction projection can be obtained by projecting the data of the predicted component onto the spatial pattern of the model field (Qin et al., 2011a), and the predicted values of the observed projections can be obtained by using regression relationships. The revised data can be obtained after combining the predicted values with the spatial base, which is a type of revision based on a statistical method that does not consider the topographic and geomorphological factors of the study area (Huang et al., 2015).

With the continuous development of artificial intelligence (AI) technology, researchers have focused on machine learning algorithms (Zhiyuan et al., 2020). These algorithms have been widely applied to several research areas in earth system science, including meteorological data corrections (Huang et al., 2017; Xu et al., 2021a). The machine learning-based revision model can capture nonlinear variations of the model simulation results with the observed deviations (A. F. Bennett, 1996; Ridwan et al., 2021), and by building a deviation prediction model, more accurate model revision results can be obtained. Different algorithms are used to determine the dependencies between the input feature values and the system output variables based on this relationship (Kadow et al., 2020). Some studies have revised daily precipitation data of the Qinghai-Tibet region using a k-nearest neighbor algorithm by environmental factors such as topography and vegetation, meteorological factors, temperature, wind speed, humidity, and barometric pressure (Wang et al., 2016). The results demonstrated that the revised method outperforms the probability density function-matching-based method in the revision of the remote sensing products. A revised model combining ensemble empirical modal decomposition and base pair neural networks were also used to revise the Global Monthly Mean Sea Surface Temperature for CMIP6 for three future emission scenarios (SSP1–2.6, SSP2–4.5, and SSP5–8.5), with the addition of socio-economic scenarios to the CMIP6 data compared to CMIP5 (Huang & Ying, 2015). This allows for more possibilities for future model data to be more probable and closer to the contemporaneous observed data. AI methods can also reconstruct missing historical data from weather stations by using model data from the historical period, reduce uncertainty and bias in climate data, and reconstruct monthly climate data through image interpolation techniques (Dalcher & Kalnay, 2016).

These studies used statistical methods, bias correction, and machine learning methods to revise the meteorological data over the YRB based on CMIP5 and CMIP6 GCMs. These findings also demonstrated that the performance of different CMIP6 models were not the same over time and space (Rivera & Arnould, 2020). However, the applicability and differences of these methods are unknown, and there is no clear comparison of the revised effects of these methods. The majority of the current correction methods have improved in terms of correlation and root mean square error (RMSE), but do not provide accurate corrections in specific extreme rainfall cases and ignore the terrain and landscape characteristics (Zheng et al., 2009). Principal component analysis (PCA) and deep neural network (DNN) meth-

ods have been used in many studies to correct precipitation data. Long short-term memory (LSTM) is useful but rarely implemented for precipitation revisions (Kratzert et al., 2018). The advantages and disadvantages of these methods are also not clear. Therefore, in this study, PCA, DNN, and LSTM were applied to correct the precipitation data from the four CMIP6 models for the period 2002–2014 taking the YRB as a study area.

Therefore, this study took the Yangtze River Basin as the study region, and adopted statistical and deep learning-based methods to revise the precipitation data of the four CMIP6 models by $0.5^\circ \times 0.5^\circ$ grid. Three specific objectives of this study are: (1) to compare the performances of the different methods in the revision of the CMIP6 precipitation data and analyzing the reasons; (2) to analyze the revision effects in terms of spatial and seasonal aspects for future improvements; (3) to provide more accurate precipitation data in future scenarios to better support for water resources assessment and management.

2 Study area and methodology

2.1 Study area and data

The YRB was chosen as the study area, as shown in Fig. 1. It is a vast area of the Yangtze River's mainstream and tributaries, spanning the three major economic regions of eastern, central, and western China. It is the third largest river basin globally, covering 1.8 million km², and is rich in natural resources. It has a multistep topography, flowing through mountains, plateaus, basins, hills, and plains. The average annual precipitation in the region is 1067 mm (Zhou & Wang, 2007). The region is vast, with complex landscapes and a typical monsoon climate, with uneven spatial and temporal distributions of annual rainfall and rainstorms (Huang et al., 2022). The river source areas have less annual precipitation than the western and eastern edges of the Sichuan Basin, Jiangxi, and parts of Hunan and Hubei (Dai et al., 2008). Areas with annual precipitation exceeding 2000 mm are distributed in the mountainous areas and are smaller in extent. As shown in the Fig. 2, precipitation in YRB is unevenly distributed throughout the year, with winter precipitation being the lowest (Wang

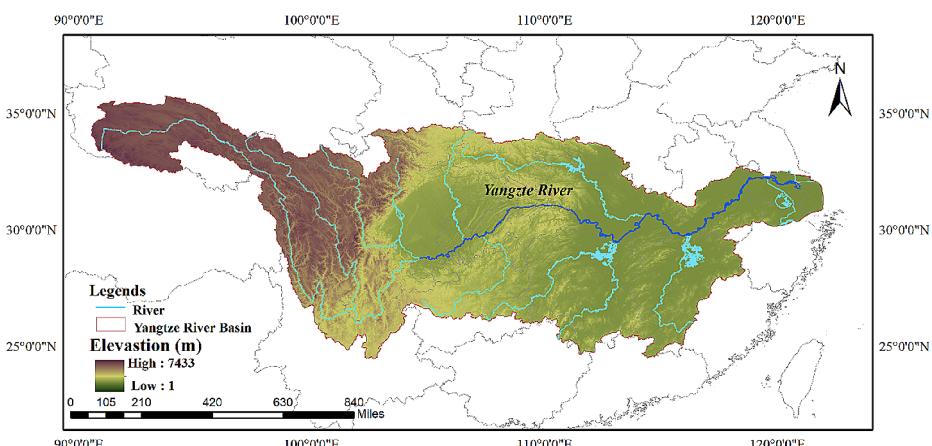


Fig. 1 Location of the YRB

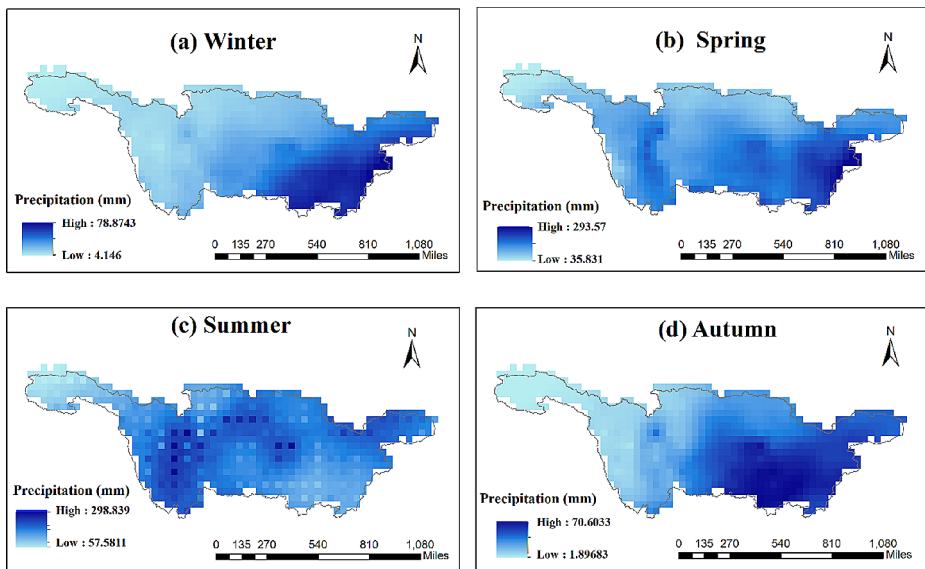


Fig. 2 Average precipitation in YRB in four seasons from 1962–2014

et al., 2023a). Spring precipitation increases monthly, with the highest monthly precipitation occurring in July and August, accounting for approximately 30% of the year, with higher precipitation in the upper reaches of the entire region and lower precipitation on the south coast in the middle and lower reaches (Tabari et al., 2012). Because of the vast territory of the YRB, it leads to a large deviation between simulated rainfall data and actual measured precipitation.

The observed monthly precipitation data used in this study were the national grid point analysis dataset with a resolution of $0.5^\circ \times 0.5^\circ$, published by the National Meteorological Information Center of China (<https://data.cma.cn>). The CMIP6 multi-model monthly precipitation data, including the historical climate simulation output data covering the period 1961–2014 and the simulation prediction data under three future climate scenarios (SSP1–2.6, SSP2–4.5, SSP5–8.5) during the period 2015–2100 in the CMIP6 archive (<https://esgf-node.llnl.gov/projects/CMIP6/>) was also used. To ensure data consistency, only the first realization (r1i1p1f1) was used from the four models. SSPs represent new emission scenarios driven by different socioeconomic models and socioeconomic changes in the areas of population, urban density, education, land use, and wealth (<https://www.ipcc.ch/report/ar6/wg1>). The three integrated scenarios were SSP126 (SSP1+RCP2.6), SSP245(SSP2+RCP4.5), and SSP585(SSP5+RCP8.5) (Gidden et al., 2019), which represent low, medium, and high emission scenarios. Representative Concentration Pathways (RCP) is a comprehensive set of concentration and emission scenarios that can be used as projections of future greenhouse gas emissions when concentrations of greenhouse gases, reactive gases, aerosols, and atmospheric constituents change in conjunction with prevailing demographics, socioeconomics, science and technology, energy consumption, and land use (Moss et al., 2010).

In this study, according to other literature, four models with good simulation results in the YRB were selected (Li et al., 2021), as shown below Table 1.

Table 1 List of the CMIP6 GCMs used in the study

Model	Affiliation (Country)	Marine mode resolution	Atmosphere mode resolution
BCC_CSM2	BCC(China)	$0.8^\circ \times 1.0^\circ$	$2.8^\circ \times 2.8^\circ$
GFDL-CM4	GFDL(USA)	$0.9^\circ \times 1.0^\circ$	$2.0^\circ \times 2.5^\circ$
IPSL-CM6A-LR	IPSL(France)	$1.2^\circ \times 2.0^\circ$	$1.9^\circ \times 3.8^\circ$
NorESM2-LM	NCC(Norway)	$0.5^\circ \times 1.1^\circ$	$1.9^\circ \times 2.5^\circ$

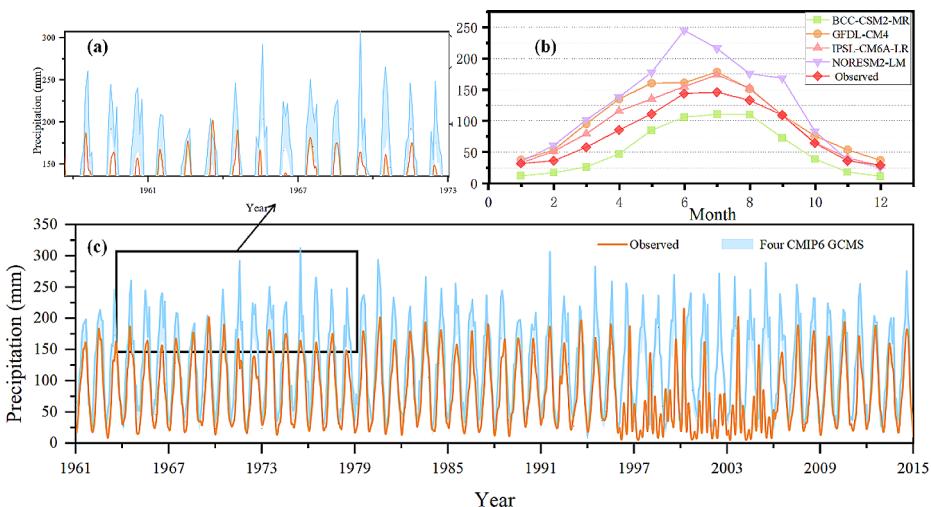
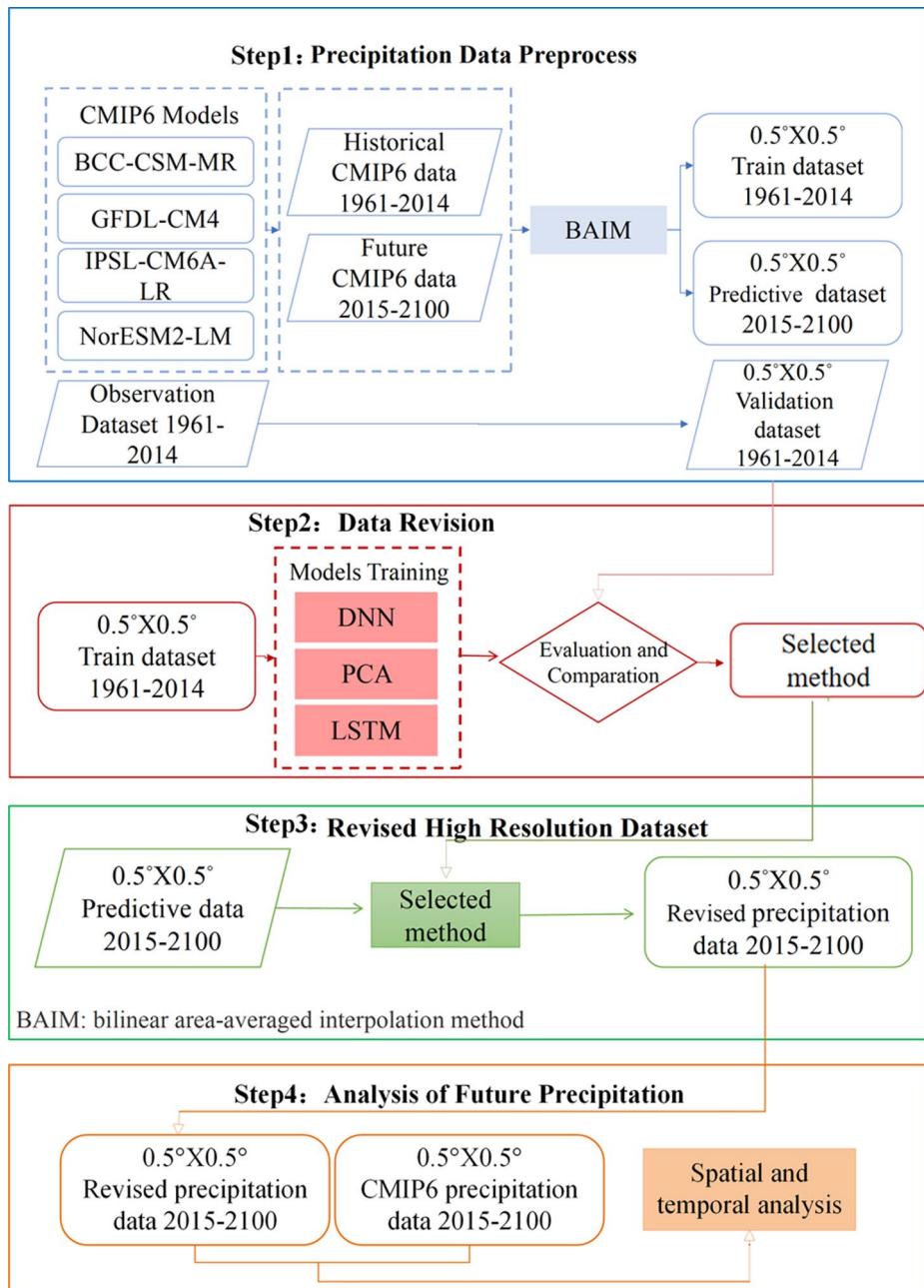


Fig. 3 Average monthly precipitation between four GCMs and observed (a) zoom-in picture (b) the average monthly precipitation for each month between GCMS and observed. (c) average monthly precipitation between GCMS and observed from 1961–2015

Figure 3-c is the comparison of the differences in time between the four models and the observed precipitation and multi-model ensemble (MME) of the four models, the overall precipitation trends of the model data and the observed data generally remain generally consistent but can differ in the extreme values. Figure 3-a is a zoomed-in plot of the difference between the observed and simulated data in terms of extremes, which clearly shows that the simulated data of several models obviously overestimated the peak value of precipitation during the rainy season. Figure 3-b shows the average monthly precipitation of the four GCMs and the observed data which indicate that GFDL-CM4, NorESM2-LM and IPSL-CM6A-LR overestimated the precipitation of YRB, especially the model Noresm2-LM.

2.2 Overall methodology

The flow chart of the study is outlined in Fig. 4. The precipitation data were interpolated onto a $0.5^\circ \times 0.5^\circ$ grid point using the Bilinear Area-averaged Interpolation Method (BAIM) before analysis for consistency (Asakawa & Kawanaka, 1991). The effect of bilinear interpolation on the quality of the data was also tested by comparing monthly mean precipitation, maxima and minima from the raw data and the output of bilinear interpolation. This approach has been demonstrated in existing studies (Yue et al., 2021). Three revision

**Fig. 4** Study flow chart

methods, PCA, DNN, and LSTM were used to revise the historical simulation of monthly precipitation data from 2002 to 2014. The precipitation from 1961 to 2100 was divided into three datasets, 1961–2001 was the training part, 2002–2014 was the validation part, and 2015–2100 was the test part. The performance of the three revised methods was evaluated and the best performing of the three was selected and used to correct the prediction monthly data from 2015 to 2100, under three SPPs in the YRB.

2.2.1 Principal components analysis method

The PCA method is a mathematical-statistical method that uses the regression relationship between two principal components by extracting the principal components of the observed data and model data. The PCA method is also called the empirical orthogonal function (EOF), as shown in Fig. 5 (Tian Mao, 2009). For this study, 1961–2001 was selected as the training segment, and 2002–2014 was the test segment. The precipitation data from the observed and CMIP6 data are decomposed separately by PCA, and then the observed field is subjected to multiple linear regression with the principal components of the modeled field to obtain the regression equation. In the revision, the model data are projected onto the existing eigenvectors to obtain the principal components, and then these principal components are substituted into the regression equation to obtain the revised precipitation data. This is a type of revision based on statistical methods and it has been used in many existing studies. (Qin et al., 2011b).

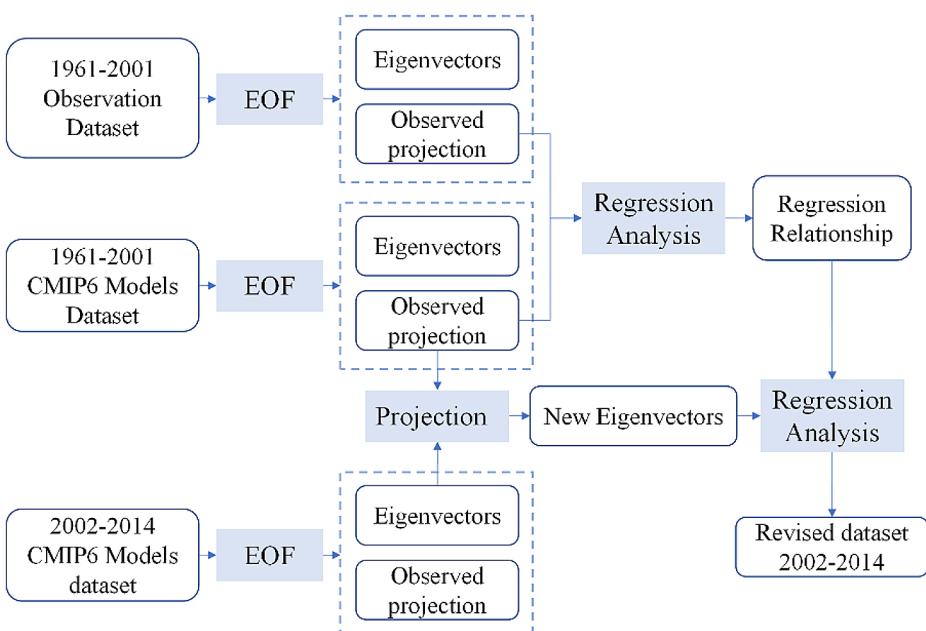


Fig. 5 PCA method flow chart

2.2.2 Deep neural network method

This study proposes a DNN-based revision method to model and revises each of the 686 grid points in the YRB. Neural network technology was launched in the 1950s and 1960s and was known as a perceptron, which included input, output, and hidden layers (Casarella, 2011). In this model, the input feature vector is sent to the output layer through the transformation of the hidden layer, and classification is performed at the output layer. DNN is a multilevel perceptual basis, and its internal neural network can be divided into three types: input, output, and hidden stages, based on which a forward neural network method is adopted. The forward transfer algorithm of the DNN uses several weighting coefficient matrices (W) and bias vector (b) to perform a series of linear operations and start-up operations on the input value vector (x). From the input layer, the operations are calculated one layer down to the output layer (Herrmann et al., 2019). In this study, the model used Relu as the active function, the input data was the historical precipitation of observed and GCMs.

2.2.3 Long short-term memory method

LSTM is a recurrent neural network model, a general term for a series of neural networks capable of processing sequential data, characterized by recurrent links between units, and mainly processes time-series data. LSTM is a specific form of RNN that has an additional pathway through the entire network, called unit states, in addition to the original output (h) of the RNN with only multiplication (Julia Gusak et al., 2022). The unique feature of LSTM is that it can control the amount of information that passes through the previous unit and the information that can be added to the current unit to pass to the next unit by adding various gating (Martin Sundermeyer et al., 2012). LSTM provides additional feature extraction functionality that can be used to complement the lack of a time-varying component in classical classifier algorithms (AlDahoul et al., 2023).

There are two states in the LSTM model, the cell and hidden states. Information can be added or deleted from the cell state, and each gate is a fully connected layer, with an activation function controlling the passage of information based on an activation value. The output of the forgotten gate and the functions of the other gates can be expressed as

$$f_t = \sigma(W_f \cdot [ht - 1, Xt] + bf) \quad (1)$$

$$c_t = f_t \times C_{t-1} + i_t \times \tilde{c}_{t-1} \quad (2)$$

$$i_t = \sigma(W_i \cdot [ht - 1, Xt] + bi) \quad (3)$$

$$\tilde{C}_t = \tanh(W_c \cdot [ht - 1, Xt] + bc) \quad (4)$$

$$O_t = \sigma(W_o \cdot [ht - 1, Xt] + bo) \quad (5)$$

$$h_t = O_t * \tanh(C_t) \quad (6)$$

f_t is a type of forgetting gate, which represents the characteristics of C_{t-1} , and then uses the sigmoid function to determine the data in the LSTM model, and then uses C_{t-1} and f_t to establish a new candidate value vector c_t then multiply the original state C_{t-1} and f_t to obtain the forgotten information. c_t is the unit state update value obtained by the input data X_t and hidden node $ht-1$ through the neural network, and tanh is generally used as a function of the unit state update. i_t is an input gate that is the same vector as f_t which is also computed with X_t and $ht-1$. Zero represents the one-unit product of two vectors. The output gate is used to update the battery cell, the sigmoid layer is used to determine which part of the current cell state is to be an output, and then tanh is used to process the cell state to obtain a value from -1 to 1, thereby obtaining an output value (Schmidhuber, 1997).

The LSTM model consists of three aspects: the first layer is data, the second is time series analysis, and the third is input-output analysis. Model simulations were performed based on the observed data from 1961 to 2001, the precipitation data of each grid in the model data, and the precipitation data from 2002 to 2014 as input. The precipitation data for each grid from 2002 to 2014 were outputted separately with the corrected precipitation data based on the relationship between the model data and observed data.

The model data and observed data from 1961 to 2001 were used for the simulation, fitting, parameter optimization, and calculation of historical simulated data predictions based on the resulting model. Within the model, the error term for each month of data was calculated by back-propagating the error term along the direction of the time series, and on the other side the error term was propagated up one layer and the gradient of each weight was calculated based on the corresponding error term. In the experiments in this study, when the learning number was set to 100, the learning rate to 0.1, and the loss function was the Euclidean distance, the RMSE was the lowest, and the revised data for 2002–2014 were obtained for each month from the predicted values of the previous five years.

2.3 Test criteria for the revised results

Various methods can be used to objectively assess the revised results of the meteorological data. In this study, three common tests, correlation, RMSE and relative bias(RBIAS) were selected to quantitatively assess the revised precipitation data. The value of the correlation ranges from 0 to 1, which mainly reflects the degree of correlation between the observed and model data. The closer the correlation is to 1, the more similar the two datasets are, and the better the simulation effect of the modeled data. Its calculation formula is as follows:

$$\gamma = \frac{\sum_{i=1}^n (X_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^n (X_i - \bar{x})^2} \sqrt{\sum_{i=1}^n (y_i - \bar{y})^2}} \quad (7)$$

The RMSE reflects the error between the observed and the model data; the smaller the value, the smaller the difference between the two datasets. The formula is as follows:

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (X_i - y_i)^2} \quad (8)$$

The RBIAS represents the bias as the percentage of the mean value. The formula is as follows:

$$RBIAS = \frac{\bar{X}_i - \bar{y}_i}{\bar{y}_i} \times 100\% \quad (9)$$

Where X_i and y_i represent CMIP6 simulation data and the observed precipitation, respectively, and \bar{x} and \bar{y} represent the long-term mean of the simulated and observed data, respectively.

3 Results

3.1 Evaluation of the corrected results by three methods

In this study, we used PCA, DNN and LSTM-based methods to revise the precipitation data of the four models in the YRB for the same period and compared the revised results. Using the PCA method to revise the principal component analysis, three contribution rates of 75, 85, and 90 were selected, and the different contribution rates mean that the numbers of the principal components selected, the higher the rate is, the more components are. The results showed that when the contribution rate is 85, the RMSE is the smallest.

After comparing the LSTM, PCA, and DNN methods, the LSTM-based method was found to be superior in terms of coefficient of correlation, RMSE and RBIAS, as shown in Table 2, where MEM represents the multi-model ensemble mean. MEM is a method that obtaining the optimal combination of several different numerical model forecasts according to certain statistical methods, in this study, the mean of four GCMs were selected (Krishnamurti et al., 1999). The DNN method has the lowest correlation and highest RMSE compared to the other methods. The LSTM performance was significantly better than PCA in terms of RMSE and RBIAS, and the revised results of the LSTM-based method were closer to the observed data than the data before modification.

Before the revision, as shown in Table 2, the correlation between the historical simulated precipitation and the observed data from 1961 to 2014 was approximately 0.45; the highest correlation was with the IPSL model, the lowest similarity was with the BCC-CSM2-MR model, the RMSE was above 80.89, the highest was with the NORESM2-LM model with an error of 109.32, and the lowest was with the IPSL-CM6A-LR model with an error of 83.67. In terms of the combined similarity and RMSE, the model closest to the observed data is the IPSL-CM6A-LR model. After calculating the average precipitation for each month, it was determined that the observed data and the model data were consistent in terms of precipitation trends, but the differences were larger in months with high precipitation during the rainy season. On a regional scale, Fig. 5 shows the average RMSE between the precipitation of the four CMIP6 models and the observed precipitation before and after correction. The RMSE of the northwestern part of the YRB is the smallest, which indicated that the simulation of precipitation is better in the northwestern side. The darker area represents the area with a larger RMSE, which indicated that the difference between the model data and observed data is large, the simulation is not effective, and the darker area is different for each model. The four pictures in the right side of Fig. 6 are the RMSE between four GCMs and

Table 2 The correlation and RMSE (mm) of four models after correction

Model	Before correct			PCA			LSTM			DNN		
	Cor	RMSE	RBIAS	Cor	RMSE	RBIAS	Cor	RMSE	RBIAS	Cor	RMSE	RBIAS
BCC-CSM2	0.38	80.89	43.70%	0.53	68	34.80%	0.6	64	24.40%	0.28	123	38.58%
GFDL-CM4	0.38	93.57	46.65%	0.48	76	37.60%	0.57	63	23.67%	0.23	134	37.45%
IPSL-CM6A-LR	0.42	83.67	28.23%	0.52	64	24.89%	0.56	62	21.22%	0.25	112	37.42%
NorESM2-LM	0.411	90.32	38.30%	0.47	80	32.67%	0.57	62	22.23%	0.32	108	34.48%
MEM	0.37	98.58	38.75%	0.47	75	31.45%	0.62	42	19.89%	0.27	145	32.50%

observed data after LSTM-based correction, which indicated that the RMSE of four models all decreased in the whole region. The lowest RMSE is in the northwest area, and the highest RMSE is in southeast area, after correction. Before the correction, all four models in the northwest region were optimal, whereas the worst regions differed, with the deviation in the southeast region being larger. After LSTM modification, the data characteristics of the four models became consistent.

3.2 Future changes in monthly precipitation in the YRB

3.2.1 Overall precipitation changes

Based on the results of previous historical data revisions, the best revisions were made to the MEM; therefore, the median precipitation data for three different scenarios using the four models were selected for analysis in this study. For all three scenarios, precipitation data increased slightly year-on-year, with little variation in monthly precipitation pattern compared to previous data. All three precipitation scenarios showed a fluctuating upward trend during 2015–2100, and there were some differences in the low emission scenario (SSP1–2.6), the medium emission scenario (SSP2–4.5), and the high emission scenario (SSP5–8.5). In all three scenarios, rainfall was likely to increase.

According to Fig. 7, shows the average monthly precipitation from 2015 to 2100 in three SSPs before and after LTSM-based correction. After correction, the precipitation all decreases in three SSPs, which shows that the original rainfall data from CMIP6 GCMs overestimated the precipitation in YRB. Besides, from the three figures, it shows that with higher SSP and RCP, precipitation gradually increased, which indicated that socio-economic patterns and carbon emission scenarios have an impact on precipitation. The higher the level of social development, the more significantly the precipitation in the rainy season will increase, raising the possibility of flooding. On the contrary, in autumn and winter, when rainfall is poor, the effect of different socio-economic patterns and carbon emission scenarios on rainfall is not as obvious. In general, from 2015 to 2100, the average precipitation over the YRB is projected to increase the most under SSP5–8.5, and the least under SSP1–2.6. Compared with the observed data, the precipitation corrected by LSTM has a good consistency, and the LSTM model can effectively reduce various uncertainty factors, thus providing a basis for future climate change analysis and evaluation.

3.2.2 Spatial and temporal variation in precipitation trends

Regarding temporal trends, Fig. 7 shows the average yearly precipitation under the three economic scenario models, which increase over time, all show an increasing trend, but the magnitude of the increase varies. The longer the time span, the more obvious the difference in precipitation data under the high scenario model, with little difference between the three scenarios during 2015–2035, and the difference between the three data becoming larger during 2075–2100. The higher and longer the duration of the three scenario emission patterns, the more the precipitation increases, and this characteristic change is consistent with the trend of the average monthly precipitation change of the three models. The study showed that the impact of precipitation data in the YRB was only apparent under long-term economic scenario model changes, while the impact of policy changes on precipitation under

Fig. 6 RMSE of four CMIP6 models in the YRB before and after LSTM-based correction

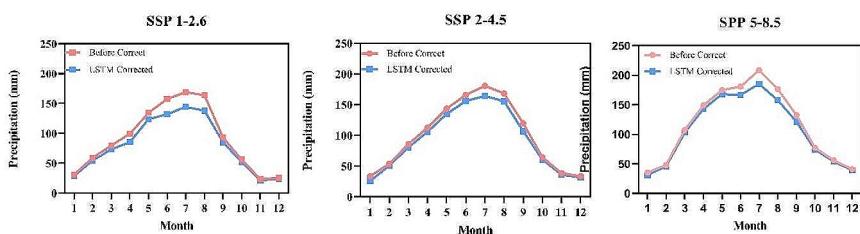
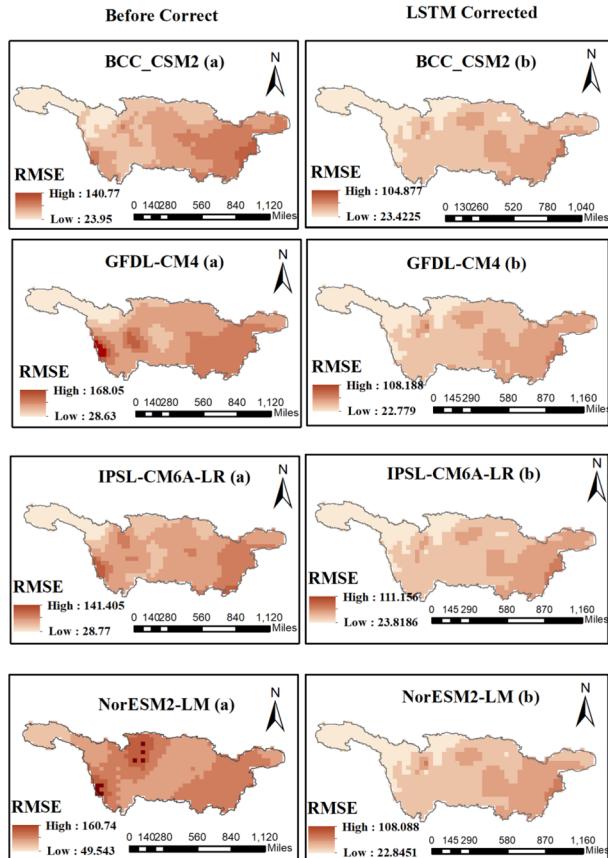


Fig. 7 Monthly average precipitation in the YRB from 2015–2100 in three SSPs

short-term scenario changes was mainly in months with more precipitation and can vary in extreme values (see Fig. 8).

Figure 9 shows the bias between average yearly precipitation of MEM after correction and historical observed data. Regionally, the northwest region of the YRB with less precipitation and the southeast region with more precipitation have differences in monthly precipitation data under the three scenario models from 2015 to 2100, and in general, each

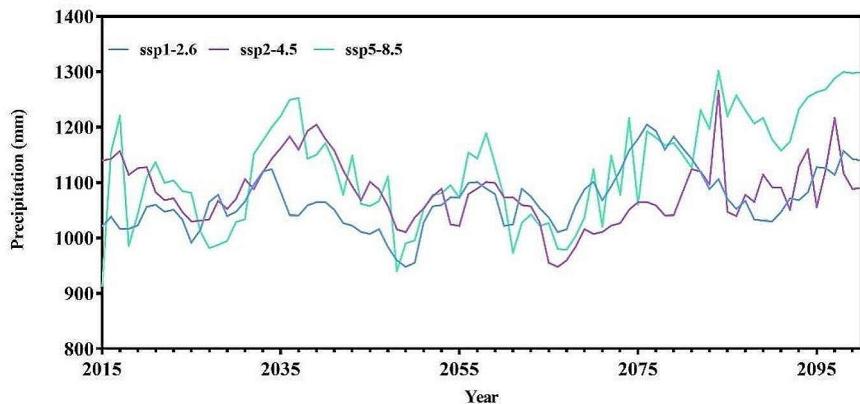


Fig. 8 Yearly average precipitation of MEM in YRB from 2015–2100 in three SSPs

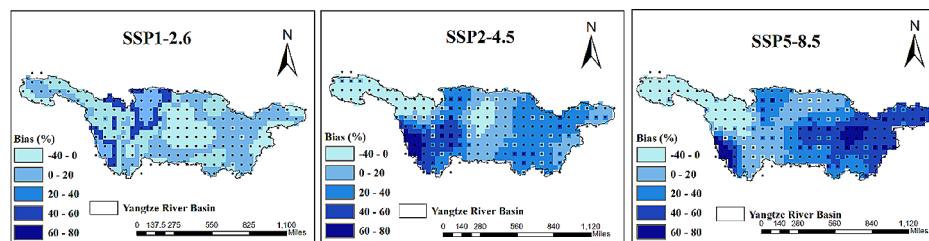


Fig. 9 Bias between average yearly precipitation of MEM and historical observed data in YRB under three SSPs

region has increased. Locally, the southeast region with more precipitation increases as the scenario model emission pattern increases, the higher the emission pattern, the more the increase, and the northwest region with less precipitation has a relatively less significant increase in SSP 1-2.6, and decreased in other two scenarios.

Comparing the precipitation data for the future phases of CMIP6 without the revision, the revised precipitation decreases in the extreme values, during the peak months of the rainy season. the CMIP6 model overestimates the precipitation in the rainy season on the YRB compared to the actual measured data, and the more the precipitation is overestimated as the discharge pattern increases. The revised data will better match the actual precipitation in the YRB and provide better support for future policy decisions.

4 Discussion

4.1 Accuracy analysis of the results

By comparing precipitation from CMIP6 GCMs with observed data confirms that there is an overall systematic overestimation of precipitation, especially in the cold season. Indeed,

both CMIP3 and CMIP5 simulations have been reported to consistently overestimate precipitation (Mueller & Seneviratne, 2014). The likelihood of this happening is due to the influence of the mid-latitude westerlies in winter, meanwhile, the complex circulation system in different regions of the YRB makes it difficult for the CMIP6 models to simulate precipitation precisely (Yao et al., 2013).

In terms of spatial distribution, YRB has the multi-step topography, where there are mountains, plateaus, hills and plains. The highland areas are all in the west, with little precipitation throughout the year and few seasonal changes, while the plain areas are concentrated in the southeast, with significant seasonal changes in rainfall. Before the revision, the regions with good simulation performance of the four models were all in the northeast, and the regions with poor simulation performance of different model data were all in the southeast, southwest, and central regions. It demonstrates that the CMIP6 model is effective in simulating regions with simple precipitation characteristics and little seasonal variation. After the LSTM-based revision, the performance of the four-model data tended to be consistent across the regions, with good results in the northwest and poor results in the southeast. In fact, the improvement of the correction is the most significant in the southeast region, while in the northwest region, where the simulation of CMIP6 GCMs is already effective, the revision didn't improve the results much or even made it worse. In terms of coefficient of correlation, R_{BIAS} and RMSE, the coefficient of correlation was increased, and the RMSE and R_{BIAS} were decreased, which indicated that there were significant improvements comparing with their pre-revision performance.

Figure 10 is the comparison of the monthly average precipitation of revised data with the observed data showed that the LSTM-based correction was relatively less effective on the days with extreme heavy rainfall during the rainy season, where the RMSE was twice the mean error. After correction, the performance of CMIP6 data has improved in summer. However, it is difficult to predict extreme rainfall events with historical data due to the episodic and lack of a certain pattern.

4.2 Comparison of statistics and deep learning revision methods

By comparing the correction of CMIP6 data by PCA, DNN-based and LSTM-based method, the LSTM-based method has the best correction effect. According to their principles, PCA

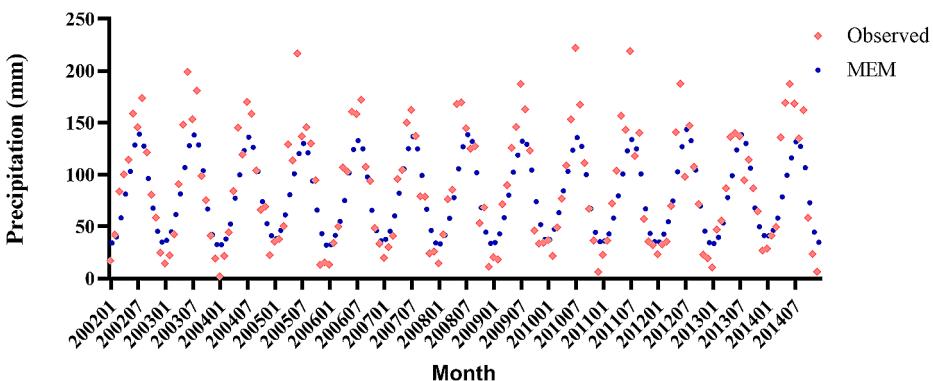


Fig. 10 Average monthly precipitation between observed and MEM from 2002–2015 after LSTM-based correction

method is a statistical method, based on the modeling of historical precipitation data, which is not so applicable to areas with complex topography and uneven precipitation distribution, the DNN model can not effectively handle large scale data with long time series. Comparatively, LSTM model can effectively capture the semantic association between long sequences, and mitigate the gradient disappearance or explosion phenomenon (Felix A. Gers, 2000), this has been demonstrated in many other studies (AlDahoul et al., 2021), so it can obtain the best revision results among these three methods. Table 3 lists the complexity, modelling principles, effectiveness in this study and interpretability of the three approaches.

In contrast to the other revision methods, there are studies that have used various statistical and bias correction methods to revise CMIP6 precipitation. One of the studies increased the correlation rate above 0.45, and another increased the R-value to 0.65 (Yang et al., 2019). The LSTM-based method can increase the correlation rate above 0.55, the R_{BIA}S is below 22%, and the RMSE is below 65. Most studies of CMIP6 precipitation data have used one method that simply revises the data and then performs an assessment within a region (Tong et al., 2015). These studies do not compare statistical methods with deep learning methods in a uniform study area. This study compared the statistical method PCA with DNN and LSTM method, and the results showed that the LSTM method was superior in both test metrics. Meanwhile, the LSTM-based method was modeled by dividing the grid points based on region, the spatial homogeneity was considered. So, the method can be used in the revisions of other complex regions.

4.3 Limitations and perspectives for future research

The results of this study showed the importance of the correction of CMIP6 precipitation data, and in addition the accuracy of the simulated data may be related to the complexity of the precipitation characteristics of the region itself. The accuracy of CMIP6 GCMs data is high in areas with simple precipitation characteristics, while the accuracy is low and the error is high in areas with complex and variable precipitation characteristics. The revised precipitation data based on the LSTM correction method still has room for improvement in terms of peak value, especially in June, July and September. The future precipitation correction method can build a more reasonable simulation model by region according to the precipitation characteristics of the region itself combined with the topography and geomorphology. By further subdividing the periods of extreme rainfall in seasons with high rainfall and improving the simulation accuracy of the model data, more accurate future precipitation simulation data will improve the understanding of future climate change trends, and thus more appropriate policies can be developed to improve the local ecological environment. By studying effective methods to correct model precipitation data, the CMIP6 future segment data can be corrected to obtain realistic precipitation data that are closer to future

Table 3 The complexity, principles, effectiveness, and interpretability of these three methods

Method	Complexity	Modelling principles	Effectiveness (this study)	Interpretability
PCA	Low	Linear transformation	Mediocre	Principal components are interpretable
DNN-based	Middle	Multi-layer perception	Worse	Lack of interpretability
LSTM-based	High	Improved recurrent neural network	Best	Lack of interpretability

scenarios. More realistic and refined future precipitation data can be useful for agriculture, carbon emissions, and government policy making.

5 Conclusion

This study evaluated the performance of four CMIP6 models in the YRB and compared statistical and deep learning-based correction methods, selected the LSTM-based method based on the coefficient of correlation, RMSE and R_{BIAS}, and then analyzed the LSTM-based method corrected precipitation from 2015 to 2100 in three SSPs. The results showed that there was an overall overestimate of original CMIP6 precipitation in the YRB, especially in the rainy season, due to the influence of the mid-latitude westerlies. The LSTM-based method improved the coefficient of correlation from 0.41 to 0.62, RMSE from 98 to 42, R_{BIAS} from 39.22 to 22.67%, and it also has a positive effect on the correction of extreme precipitation. Besides, this method improved the shortcoming of original CMIP6 data in overestimating rainfall. However, this method is more useful in the part of low RMSE and not that useful for the region which RMSE is already low. Regionally, the rainfall increases most in the southeast and less in the northwest over time, and the higher the SPP index the more the rainfall increases. In general, the LSTM-based method has advantages in modeling long time series data, it is also flexible by dividing the study area into grids which were revised individually. The corrected future precipitation can be helpful for water resources assessment and management, also provide the scientific evidence for future drought and flood control.

Acknowledgements This work was supported by the National Natural Science Foundation of China program (No. 41890822 and 41971351), the Key R&D Program of Hubei Province (No. 2022BAA048) and the Open Fund of National Engineering Research Center for Geographic Information System, China University of Geosciences, Wuhan 430074, China (Grant No. 2022KFJJ07). The numerical calculations in this paper have been done on the supercomputing system in the Supercomputing Center of Wuhan University.

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

Data availability The datasets generated during the current study are available from the corresponding author on reasonable request.

References

- Adhikari, U., Nejadhashemi, A. P., & Woznicki, S. A. (2015). Climate change and eastern Africa: A review of impact on major crops. *Food and Energy Security*, 4(2), 110–132. <https://doi.org/10.1002/fes3.61>.
- AlDahoul, N., Essam, Y., Kumar, P., Ahmed, A. N., Sherif, M., Sefelnasr, A., & Elshafie, A. (2021). Suspended sediment load prediction using long short-term memory neural network. *Scientific Reports*, 11(1), 7826. <https://doi.org/10.1038/s41598-021-87415-4>.
- AlDahoul, N., Momo, M. A., Chong, K. L., Ahmed, A. N., Huang, Y. F., Sherif, M., & El-Shafie, A. (2023). Streamflow classification by employing various machine learning models for peninsular Malaysia. *Scientific Reports*, 13(1), 14574. <https://doi.org/10.1038/s41598-023-41735-9>.
- Asakawa, E., & Kawanaka, T. (1991). *Seismic raytracing using linear traveltime interpolation* (p. 53). rd EAEG Meeting.
- Bennett, A. F., B. S. C., and, & Leslie, L. M. (1996). Generalized inversion of a global Numerical Weather Prediction Model. *Meteorology and Atmospheric Physics*, 60, 165–178.

- Casarella, J. M. (2011). *Frank Rosenblatt, Alan M. Turing, Connectionism, and Artificial Intelligence*. Seidenberg School of CSIS.
- Dai, Z., Du, J., Li, J., Li, W., & Chen, J. (2008). Runoff characteristics of the Changjiang River during 2006: Effect of extreme drought and the impounding of the Three Gorges Dam. *Geophysical Research Letters*, 35(7). <https://doi.org/10.1029/2008gl033456>.
- Dalcher, A., & Kalnay, E. (2016). Error growth and predictability in operational ECMWF forecasts. *Tellus A: Dynamic Meteorology and Oceanography*, 39(5), 474–491. <https://doi.org/10.3402/tellusa.v39i5.11774>.
- Danforth, C. M., & Kalnay, E. (2008). Impact of online empirical model correction on nonlinear error growth. *Geophysical Research Letters*, 35(24). <https://doi.org/10.1029/2008gl036239>.
- Danforth, C. M., Kalnay, E., & Miyoshi, T. (2007). Estimating and correcting global Weather Model Error. *Monthly Weather Review*, 135(2), 281–299. <https://doi.org/10.1175/mwr3289.1>.
- Ehteram, M., Ahmed, A. N., Khozani, S., Z., & El-Shafie, A. (2023). Convolutional neural network -support vector machine model-gaussian process regression: A New Machine Model for Predicting Monthly and Daily Rainfall. *Water Resources Management*, 37(9), 3631–3655. <https://doi.org/10.1007/s11269-023-03519-8>.
- Eyring, V., Bony, S., Meehl, G. A., Senior, C. A., Stevens, B., Stouffer, R. J., & Taylor, K. E. (2016). Overview of the coupled model Intercomparison Project Phase 6 (CMIP6) experimental design and organization. *Geoscientific Model Development*, 9(5), 1937–1958. <https://doi.org/10.5194/gmd-9-1937-2016>.
- Felix, A., Gers, J. S., & Cummins, F. (2000). Learning to forget: Continual prediction with LSTM. *Neural Computation*, 12, 2451–2471. <https://doi.org/10.1162/089976600300015015>.
- Gidden, M. J., Riahi, K., Smith, S. J., Fujimori, S., Luderer, G., Kriegler, E., & Takahashi, K. (2019). Global emissions pathways under different socioeconomic scenarios for use in CMIP6: A dataset of harmonized emissions trajectories through the end of the century. *Geoscientific Model Development*, 12(4), 1443–1475. <https://doi.org/10.5194/gmd-12-1443-2019>.
- Gosai, A., Salinger, J., & Dirks, K. (2009). Climate and respiratory disease in Auckland, New Zealand. *Australian and New Zealand Journal of Public Health*, 33(6), 521–526. <https://doi.org/10.1111/j.1753-6405.2009.00447.x>.
- Guo, J., Chen, H., Xu, C. Y., Guo, S., & Guo, J. (2011). Prediction of variability of precipitation in the Yangtze River Basin under the climate change conditions based on automated statistical downscaling. *Stochastic Environmental Research and Risk Assessment*, 26(2), 157–176. <https://doi.org/10.1007/s00477-011-0464-x>.
- Gusain, A., Ghosh, S., & Karmakar, S. (2020). Added value of CMIP6 over CMIP5 models in simulating Indian summer monsoon rainfall. *Atmospheric Research*, 232. <https://doi.org/10.1016/j.atmosres.2019.104680>.
- Herrmann, J., Beaumont, O., & L, E. D. (2019). Optimal checkpointing for heterogeneous chains: how to train deep neural networks with limited memory. *arXiv e-prints*.
- Stevens, S. B. a. B. (2012). Clouds, Circulation and Climate Sensitivity: How the interactions Between clouds, greenhouse gases and aerosols affect temperature and precipitation in a changing climate. *White Paper*.
- Huang, P., & Ying, J. (2015). A Multimodel Ensemble Pattern regression method to correct the Tropical Pacific SST change patterns under global warming. *Journal of Climate*, 28(12), 4706–4723. <https://doi.org/10.1175/jcli-d-14-00833.1>.
- Huang, J., Yu, H., Guan, X., Wang, G., & Guo, R. (2015). Accelerated dryland expansion under climate change. *Nature Climate Change*, 6(2), 166–171. <https://doi.org/10.1038/nclimate2837>.
- Huang, D., Yan, P., Zhu, J., Zhang, Y., Kuang, X., & Cheng, J. (2017). Uncertainty of global summer precipitation in the CMIP5 models: A comparison between high-resolution and low-resolution models. *Theoretical and Applied Climatology*, 132(1–2), 55–69. <https://doi.org/10.1007/s00704-017-2078-9>.
- Huang, S., Zhang, X., Yang, L., Chen, N., Nam, W. H., & Niyogi, D. (2022). Urbanization-induced drought modification: Example over the Yangtze River Basin, China. *Urban Climate*, 44. <https://doi.org/10.1016/j.uclim.2022.101231>.
- Huang, S., Gan, Y., Zhang, X., Chen, N., Wang, C., Gu, X., & Niyogi, D. (2023). Urbanization amplified asymmetrical changes of Rainfall and Exacerbated Drought: Analysis over five Urban agglomerations in the Yangtze River Basin, China. *Earth's Future*, 11(2). <https://doi.org/10.1029/2022ef003117>.
- Jia, M., Wang, Z., Mao, D., Ren, C., Song, K., Zhao, C., & Wang, Y. (2023). Mapping global distribution of mangrove forests at 10-m resolution. *Sci Bull (Beijing)*, 68(12), 1306–1316. <https://doi.org/10.1016/j.scib.2023.05.004>.
- Julia Gusak, D., Cherniuk, A., Shilova, A., Katrutsa, D., Bershatsky, & Zhao, X. (2022). Survey on Large Scale Neural Network Training.
- Kadow, C., Hall, D. M., & Ulbrich, U. (2020). Artificial intelligence reconstructs missing climate information. *Nature Geoscience*, 13(6), 408–413. <https://doi.org/10.1038/s41561-020-0582-5>.

- Klein, W. H. (1971). computer prediction precipitation *Journal of applied Meteorology*, 10, 903–915.
- Kratzert, F., Klotz, D., Brenner, C., Schulz, K., & Herrnegger, M. (2018). Rainfall-runoff modelling using long short-term memory (LSTM) networks. *Hydrology and Earth System Sciences*, 22(11), 6005–6022. <https://doi.org/10.5194/hess-22-6005-2018>.
- Krishnamurti, T. N., Kishtawal, C. M., LaRow, T. E., Bachiochi, D. R., Zhang, Z., Williford, C. E., & Surendran, S. (1999). Improved Weather and Seasonal Climate forecasts from Multimodel Superensemble. *Science*, 285(5433), 1548–1550. <https://doi.org/10.1126/science.285.5433.1548>.
- Li, Y., Yan, D., Peng, H., & Xiao, S. (2021). Evaluation of precipitation in CMIP6 over the Yangtze River Basin. *Atmospheric Research*, 253. <https://doi.org/10.1016/j.atmosres.2020.105406>.
- Masud, B., Cui, Q., Ammar, M. E., Bonsal, B. R., Islam, Z., & Faramarzi, M. (2021). Means and extremes: Evaluation of a CMIP6 Multi-model Ensemble in reproducing historical climate characteristics across Alberta, Canada. *Water*, 13(5). <https://doi.org/10.3390/w13050737>.
- Mohamadi, S., Sheikh Khozani, Z., Ehteram, M., Ahmed, A. N., & El-Shafie, A. (2022). Rainfall prediction using multiple inclusive models and large climate indices. *Environmental Science and Pollution Research International*, 29(56), 85312–85349. <https://doi.org/10.1007/s11356-022-21727-4>.
- Molly, E., Brown, & Funk, C. C. (2008). Food Security under Climate Change. *Science*, 319(NASA Publications), 580–581.
- Moss, R. H., Edmonds, J. A., Hibbard, K. A., Manning, M. R., Rose, S. K., van Vuuren, D. P., & Wilbanks, T. J. (2010). The next generation of scenarios for climate change research and assessment. *Nature*, 463(7282), 747–756. <https://doi.org/10.1038/nature08823>.
- Mueller, B., & Seneviratne, S. I. (2014). Systematic land climate and evapotranspiration biases in CMIP5 simulations. *Geophysical Research Letters*, 41(1), 128–134. <https://doi.org/10.1002/2013GL058055>.
- OFFICE, I. C. P. (2011). Decadal and bias correction for decadal climate predictions. *International CLIVAR Project Office*, 150.
- Peng, Q., Xie, S., Wang, D., Huang, R., Chen, G., Shu, Y., & Liu, W. (2022). Surface warming-induced global acceleration of upper ocean currents. *Science Advances*, 8. <https://doi.org/10.1126/sciadv.abj8394>.
- Qin, Z., Lin, Z., Chen, H., & Sun, Z. (2011a). The bias correction method based on the EOF/SVD for short term climate prediction and their application. *Acta Meteorologica Sinica*, 69(2), 289–296.
- Qin, Z., Lin, Z., Chen, H., & Sun, Z. (2011b). Short-term climate prediction error revision method based on EOF/SVD and its application. *Acta Meteorologica Sinica*, 69(2), 8.
- Ridwan, W. M., Sapitang, M., Aziz, A., Kushiar, K. F., Ahmed, A. N., & El-Shafie, A. (2021). Rainfall forecasting model using machine learning methods: Case study Terengganu, Malaysia. *Ain Shams Engineering Journal*, 12(2), 1651–1663. <https://doi.org/10.1016/j.asej.2020.09.011>.
- Rivera, J. A., & Arnould, G. (2020). Evaluation of the ability of CMIP6 models to simulate precipitation over Southwestern South America: Climatic features and long-term trends (1901–2014). *Atmospheric Research*, 241. <https://doi.org/10.1016/j.atmosres.2020.104953>.
- Schmidhuber, S. H., & a., J. (1997). Long short-term memory. *Neural Computation*, 9, 1735–1780. <https://doi.org/10.1162/neco.1997.9.8.1735>.
- Song, Z. Y., Bao, Y., & Qiao, F. L. (2019). Introduction of FIO-ESM v2.0 and its participation plan in CMIP6 experiments. *Climate Change Research*, 15(5), 558–565. <https://doi.org/10.12006/j.issn.1673-1719.2019.033>.
- Stevens, S. B., & a., B. (2012). Clouds, Circulation and Climate Sensitivity.
- Sundermeyer, M., Ralf Schluter, & Ney, H. (2012). LSTM Neural Networks for Language Modeling. *Interspeech*.
- Tabari, H., Aein, A., Talaee, P. H., & Some'e, B. S. (2012). Spatial distribution and temporal variation of reference evapotranspiration in arid and semi-arid regions of Iran. *Hydrological Processes*, 26(4), 500–512. <https://doi.org/10.1002/hyp.8146>.
- Tian Mao, W. W. (2009). *Xinan Yue,Lingfeng Sun,Biqiang Zhao, and Jianpeng Guo*. An empirical orthogonal function model of total electron content over China.
- Tong, L., Peng, X., & Cui, J. (2015). A study on the revision method of systematic errors in the GRAPES GFS model. *Journal of Chengdu University of Information Technolohy*, 30, 1671–1742. <https://doi.org/10.16836/j.cnki.jcuit.2015.03.011>.
- Wang, Y., Nan, Z., & Chen, H. (2016). Correction of CMORPH Daily Precipitation Data over the Qinghai-Tibetan Plateau with K-Nearest Neighbor Mode. *Remote Sensing Technology and Application*, 31(3), 607–616.
- Wang, C., Jia, M., Chen, N., & Wang, W. (2018). Long-Term Surface Water Dynamics Analysis Based on Landsat Imagery and the Google Earth Engine platform: A Case Study in the Middle Yangtze River Basin. *Remote Sensing*, 10(10). <https://doi.org/10.3390/rs10101635>.
- Wang, S., Zhang, X., Wang, C., & Chen, N. (2023a). Multivariable integrated risk assessment for Cyanobacterial blooms in eutrophic lakes and its spatiotemporal characteristics. *Water Research*, 228(Pt A), 119367. <https://doi.org/10.1016/j.watres.2022.119367>.

- Wang, S., Zhang, X., Wang, C., & Chen, N. (2023b). Temporal continuous monitoring of Cyanobacterial blooms in Lake Taihu at an hourly scale using machine learning. *Science of the Total Environment*, 857(Pt 2), 159480. <https://doi.org/10.1016/j.scitotenv.2022.159480>.
- Xu, L., Chen, N., Chen, Z., Zhang, C., & Yu, H. (2021a). Spatiotemporal forecasting in earth system science: Methods, uncertainties, predictability and future directions. *Earth-Science Reviews*, 222. <https://doi.org/10.1016/j.earscirev.2021.103828>.
- Xu, L., Chen, N., Zhang, X., Moradkhani, H., Zhang, C., & Hu, C. (2021b). In-situ and triple-collocation based evaluations of eight global root zone soil moisture products. *Remote Sensing of Environment*, 254. <https://doi.org/10.1016/j.rse.2020.112248>.
- Yang, Y., Dai, X., & Tong, H. (2019). CMIP5 Model Precipitation Bias-correction methods and projected China Precipitation for the next 30 years. *Climatic and Environmental Research*, 24(6), 769–784. <https://doi.org/10.3878/j.issn.1006-9585.2019.19021>.
- Yao, T., Masson-Delmotte, V., Gao, J., Yu, W., Yang, X., Risi, C., & Hou, S. (2013). A review of climatic controls on $\delta^{18}\text{O}$ in precipitation over the Tibetan Plateau: Observations and simulations. *Reviews of Geophysics*, 51(4), 525–548. <https://doi.org/10.1002/rog.20023>.
- Yue, Y., Yan, D., Yue, Q., Ji, G., & Wang, Z. (2021). Future changes in precipitation and temperature over the Yangtze River Basin in China based on CMIP6 GCMs. *Atmospheric Research*, 264. <https://doi.org/10.1016/j.atmosres.2021.105828>.
- Zheng, Z., Ren, H., & Huang, J. (2009). Analogue correction of errors based. *Acta Meteorologica Sinica*, 58(10), 7359–7367.
- Zhiyuan, K., Zhenya, S., & Changming, D. (2020). Study on the future projection of Global Sea Surface temperature over 21st Century using a biases correction model based on machine learning. *Climate Change Research Letters*, 09(04), 270–284. <https://doi.org/10.12677/ccrl.2020.94031>.
- Zhou, B., & Wang, H. (2007). Hadley circulation signal in summer precipitation in the Yangtze River basin. *Proceedings of the Annual Meeting of the Chinese Meteorological Society*, 313–327.

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

Springer Nature or its licensor (e.g. a society or other partner) holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.

Authors and Affiliations

An He¹ · Chao Wang¹  · Lei Xu² · Peng Wang³ · Wei Wang¹ · Nengcheng Chen^{1,2} · Zeqiang Chen²

 Chao Wang
c.wang@whu.edu.cn

 Peng Wang
i3dgis@outlook.com

¹ State Key Laboratory of Information Engineering in Surveying, Mapping and Remote Sensing, Hubei Luojia Laboratory, Wuhan University, Wuhan 430079, China

² National Engineering Research Center of Geographic Information System, China University of Geosciences, Wuhan 430074, China

³ College Information Science and Engineering, Wuchang Shouyi University, Wuhan 430064, China

Terms and Conditions

Springer Nature journal content, brought to you courtesy of Springer Nature Customer Service Center GmbH (“Springer Nature”).

Springer Nature supports a reasonable amount of sharing of research papers by authors, subscribers and authorised users (“Users”), for small-scale personal, non-commercial use provided that all copyright, trade and service marks and other proprietary notices are maintained. By accessing, sharing, receiving or otherwise using the Springer Nature journal content you agree to these terms of use (“Terms”). For these purposes, Springer Nature considers academic use (by researchers and students) to be non-commercial.

These Terms are supplementary and will apply in addition to any applicable website terms and conditions, a relevant site licence or a personal subscription. These Terms will prevail over any conflict or ambiguity with regards to the relevant terms, a site licence or a personal subscription (to the extent of the conflict or ambiguity only). For Creative Commons-licensed articles, the terms of the Creative Commons license used will apply.

We collect and use personal data to provide access to the Springer Nature journal content. We may also use these personal data internally within ResearchGate and Springer Nature and as agreed share it, in an anonymised way, for purposes of tracking, analysis and reporting. We will not otherwise disclose your personal data outside the ResearchGate or the Springer Nature group of companies unless we have your permission as detailed in the Privacy Policy.

While Users may use the Springer Nature journal content for small scale, personal non-commercial use, it is important to note that Users may not:

1. use such content for the purpose of providing other users with access on a regular or large scale basis or as a means to circumvent access control;
2. use such content where to do so would be considered a criminal or statutory offence in any jurisdiction, or gives rise to civil liability, or is otherwise unlawful;
3. falsely or misleadingly imply or suggest endorsement, approval , sponsorship, or association unless explicitly agreed to by Springer Nature in writing;
4. use bots or other automated methods to access the content or redirect messages
5. override any security feature or exclusionary protocol; or
6. share the content in order to create substitute for Springer Nature products or services or a systematic database of Springer Nature journal content.

In line with the restriction against commercial use, Springer Nature does not permit the creation of a product or service that creates revenue, royalties, rent or income from our content or its inclusion as part of a paid for service or for other commercial gain. Springer Nature journal content cannot be used for inter-library loans and librarians may not upload Springer Nature journal content on a large scale into their, or any other, institutional repository.

These terms of use are reviewed regularly and may be amended at any time. Springer Nature is not obligated to publish any information or content on this website and may remove it or features or functionality at our sole discretion, at any time with or without notice. Springer Nature may revoke this licence to you at any time and remove access to any copies of the Springer Nature journal content which have been saved.

To the fullest extent permitted by law, Springer Nature makes no warranties, representations or guarantees to Users, either express or implied with respect to the Springer nature journal content and all parties disclaim and waive any implied warranties or warranties imposed by law, including merchantability or fitness for any particular purpose.

Please note that these rights do not automatically extend to content, data or other material published by Springer Nature that may be licensed from third parties.

If you would like to use or distribute our Springer Nature journal content to a wider audience or on a regular basis or in any other manner not expressly permitted by these Terms, please contact Springer Nature at

onlineservice@springernature.com