Water Supply

IWA PUBLISHING

Water Supply Vol 23 No 9, 3597 doi: 10.2166/ws.2023.212

Precipitation prediction based on CEEMDAN–VMD–BILSTM combined quadratic decomposition model

Xianqi Zhang[a,b,c], Jingwen Shi[a,*], Haiyang Chen[a], Yimeng Xiao[a] and Minghui Zhang[a]

[a] Water Conservancy College, North China University of Water Resources and Electric Power, Zhengzhou 450046, China
[b] Collaborative Innovation Center of Water Resources Efficient Utilization and Protection Engineering, Zhengzhou 450046, China
[c] Technology Research Center of Water Conservancy and Marine Traffic Engineering, Zhengzhou, Henan Province 450046, China
*Corresponding author. E-mail: cicsjw@163.com

ABSTRACT

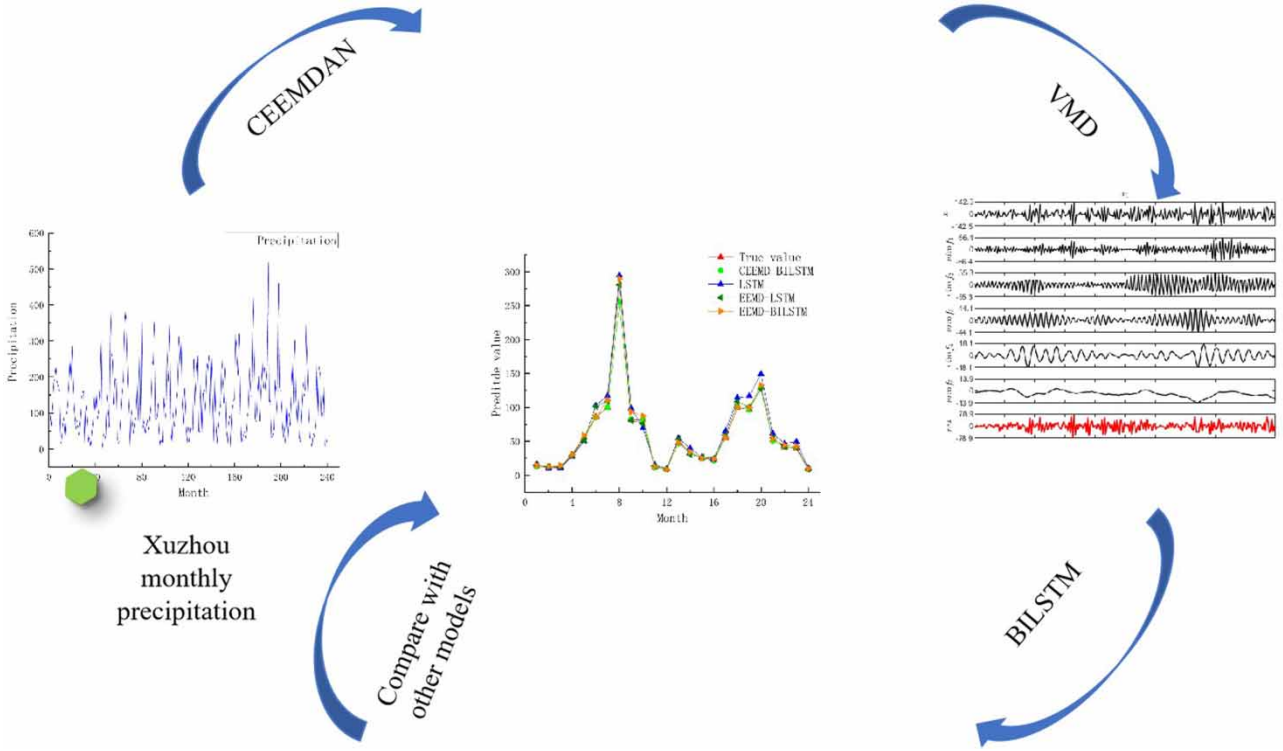
Accurate prediction of monthly precipitation is crucial for effective regional water resources management and utilization. However, precipitation series are influenced by multiple factors, exhibiting significant ambiguity, chance, and uncertainty. In this research, we propose a combined model that integrates adaptive noise-complete ensemble empirical mode decomposition (CEEMDAN), variational modal decomposition method (VMD), and bidirectional long- and short-term memory (BILSTM) to enhance precipitation prediction. We apply this model to forecast precipitation in Fuzhou City and compare its performance with existing models, including CEEMD–long and short-term memory (LSTM), CEEMD–BILSTM, and CEEMDAN–BILSTM. Our findings demonstrate that the combined CEEMDAN–VMD–BILSTM quadratic decomposition model yields more accurate predictions and captures the real variation in precipitation series with greater fidelity. The model achieves an average relative error of 1.69%, at a lower level, and an average absolute error of 1.32 m, with a Nash–Sutcliffe efficiency coefficient of 0.92. Overall, the proposed quadratic decomposition model exhibits excellent applicability, stability, and superior predictive capabilities in monthly precipitation forecasting.

Key words: BILSTM neural network, CEEMDAN and VMD, forecast accuracy, Fuzhou, precipitation forecasting

HIGHLIGHTS

- Based on CEEMDAN to effectively reduce the reconstruction error of time series, VMD to effectively reduce the non-smoothness of precipitation time series with high complexity and strong non-linearity, and bidirectional long and short-term memory (BILSTM) model to effectively learn the long-term dependence in time series.
- A combined model of adaptive noise-complete ensemble empirical modal decomposition (CEEMDAN), variational modal decomposition method (VMD) and bidirectional long and short-term memory (BILSTM) was constructed. which makes the prediction results more accurate and the coupled model can reflect the real changes of precipitation series in more reasonable details.

GRAPHICAL ABSTRACT



1. INTRODUCTION

The occurrence of water and drought extremes has become increasingly frequent due to global climate change, leading to a severe water resources situation (Mohammadi *et al.* 2021). Accurate long-term precipitation prediction serves as a critical indicator for efficient water resource utilization. Precipitation, a key source of recharge for regional water resources, significantly impacts various aspects of regional life and production. Precipitation anomalies often result in destructive floods, highlighting the importance of precise precipitation prediction. Monthly precipitation series are influenced by diverse factors, such as the atmosphere, region, and environment, exhibiting substantial ambiguity, contingency, and uncertainty. The study of monthly precipitation is a complex problem involving multiple levels and orders (Radhakrishnan & Dinesh 2006). Scholars worldwide have made substantial efforts to enhance prediction accuracy and optimize prediction models, yielding fruitful results. In recent years, the rapid development of artificial intelligence and modern statistical methods has propelled their application in precipitation prediction. Integrating these advanced techniques with traditional precipitation occurrence models is crucial to improve the reliability of water resource management decisions in scientific research and business (Jones & Pittock 2002).

Consequently, extensive research has been conducted by scholars worldwide, resulting in significant advancements in accurate precipitation prediction. Hua (2022) conducted a study analyzing the spatial and temporal characteristics of precipitation levels in Jilin Province. The study utilized daily data on precipitation and conventional meteorological elements from 2016 to 2020 for 50 cities. The parameters of the support vector machine were optimized using the gray wolf optimization algorithm and differential evolutionary algorithm. The study aimed to predict the occurrence and level of precipitation for selected stations. In a separate study, Wu *et al.* (2022) proposed a combined CNN-Attention-BP model that integrated the attention mechanism, convolutional neural network (CNN), and BP neural network. This model was developed through a comprehensive analysis of statistical prediction models for precipitation. The researchers empirically analyzed summer precipitation for different climate types at Changchun, Baicheng, and Yanji stations from 1961 to 2020. Shen *et al.* (2020) utilized the long-short-term memory (LSTM) network to predict summer precipitation in China for the years 2014 and 2015. They employed

the historical return data of the BCC-CSM seasonal climate prediction model system, along with monthly surface precipitation values provided by the National Meteorological Information Center. The researchers compared multiple methods and examined factors influencing the prediction results. In another study, Dong *et al.* (2020) compared the prediction accuracy of the LSTM network for monthly runoff using different feature inputs. They analyzed historical runoff data from Yingluo Gorge and Zamashk hydrological stations in the upper reaches of the Heihe River basin, as well as rainfall data from nearby meteorological stations. Han *et al.* (2022) identified that the LSTM-based precipitation prediction model is prone to overfitting and time lag. To address these issues, Hao-Ran (2022) proposed a depth-width prediction architecture that combines the advantages of width learning and depth learning. They established a CEEMD–LSTM–BLS single-factor monthly precipitation prediction model, considering the long-term memory function of LSTM and the noise elimination capability of complete ensemble empirical modal decomposition (CEEMD). The model was applied to analyze five representative stations in Hubei Province with diverse geographical and precipitation characteristics. In the field of data denoising, empirical modal decomposition (EMD) is a commonly used method (Norden *et al.* 1998). The ensemble empirical modal decomposition (EEMD) improves upon EMD by introducing Gaussian white noise and averaging the decomposition results to better reflect the original series' variation characteristics (Wu & Huang 2009). EEMD-based hybrid models are widely utilized in hydrological and meteorological forecasting. Complementary CEEMD (Xue *et al.* 2013) reduces signal reconstruction errors by adding white noise with zero mean and opposite amplitude to the original signal. Completely ensemble empirical modal decomposition with adaptive noise (CEEMDAN) addresses the slow decomposition of EEMD and the challenge of complete noise cancelation by decomposing the added white noise together (Torres *et al.* 2011). Chen *et al.* (2022) employed EMD, attention mechanism, and bidirectional long- and short-term memory (BILSTM) neural network in combination with an interpolation method for input data to enhance runoff prediction accuracy. Wei *et al.* (2017) applied the complete ensemble empirical modal decomposition (CEEMDAN) method to perform multi-scale analysis of flood precipitation in the Haihe River basin, identifying evolutionary patterns and selecting the best prediction model using methods such as nearest neighbor sampling regression model (NNBR), autoregressive model (AR), and neural network model. Zhang *et al.* (2021) achieved higher prediction accuracy for annual runoff using the CEEMDAN–ARMA model compared to a single ARIMA model. Kan *et al.* (2022) examined regional differences in climate change and circulation effects in Gansu Province using various methods, including least squares, complete ensemble empirical mode decomposition with adaptive noise (CEEMDAN), cross wavelet transform, wavelet coherence, and Hurst index analysis. Luo *et al.* (2022) developed a coupled model based on CEEMDAN and long- and short-term memory neural network (LSTM) to predict monthly precipitation in Zhengzhou City.

To explore new coupled models, we consider the strengths of different techniques. CEEMDAN effectively reduces the reconstruction error of time series, variational modal decomposition (VMD) addresses the non-smoothness of precipitation time series with high complexity and strong non-linearity, while the BILSTM model effectively captures long-term dependencies in time series. With these considerations, we construct a combined CEEMDAN–VMD–BILSTM prediction model for precipitation forecasting. This model follows a 'primary decomposition-quadratic decomposition-prediction-reconstruction' approach. The average relative error of the CEEMDAN–VMD–BILSTM combined prediction model is a low 1.69%, and the average absolute error of the CEEMDAN–VMD–BILSTM combined quadratic decomposition model is only 1.32 m. The Nash–Sutcliffe efficiency (NSE) coefficient reaches a high value of 0.92, indicating a superior prediction capability.

2. RESEARCH METHODS

2.1. CEEMDAN fundamentals

Empirical modal decomposition (EMD) is a signal decomposition technique developed by NASA. It is specifically designed to address the non-linearity and non-smoothness of signals, providing an adaptive spatio-temporal analysis method that ultimately achieves signal smoothing. EMD decomposes the signal into a sum of eigenmode functions (IMFs) and a trend term (Res), as shown in Equation (1). Each eigenmode function captures different time-scale features of the signal, while the trend term represents the overall signal trend. Building upon EMD and EEMD, Torres *et al.* (2011) proposed CEEMDAN, an enhanced signal decomposition method. CEEMDAN overcomes the modal mixing issue encountered in EMD by incorporating adaptive white noise multiple times during the decomposition process. This complete decomposition process of CEEMDAN ensures an accurate reconstruction of the original signal. Specifically, En(·) denotes the modal component of the nth stage

generated by the EMD algorithm, while the nth modal component generated by the CEEMDAN algorithm is denoted as IMFn. The algorithm is implemented as follows.

(1) The signal $x(t)$ to be decomposed is added to a Gaussian white noise sequence with N times the mean value of 0 to construct the sequence $x_i(t)$ to be decomposed for a total of N experiments ($i = 1, 2, \cdots, N$)

$$x_i(t) = x(t) + \varepsilon\delta_i(t) \tag{1}$$

where ε is the Gaussian white noise weighting coefficient; $\delta_i\ (t)$ is the ith added white noise sequence.

(2) The EMD algorithm is applied to decompose $x_i\ (t)$ to obtain the first modal component (IMF) and the first unique residual component $r_1\ (t)$.

$$\mathrm{IMF}_1(t) = \frac{1}{N}\sum_{i=1}^{N}\mathrm{IMF}_1^i(t) \tag{2}$$

$$r_1(t) = x(t) - \mathrm{IMF}_1(t) \tag{3}$$

(3) Adding noise to the residual components of the jth ($j = 2, 3, \cdots, N$) stage obtained after decomposition continues to apply EMD for decomposition.

$$\mathrm{IMF}_j(t) = \frac{1}{N}\sum_{i=1}^{N}E_1[r_{j-1}(t) + \varepsilon_{j-1}E_{j-1}(\delta_i(t))] \tag{4}$$

$$r_j(t) = r_{j-1}(t) - \mathrm{IMF}_j(t) \tag{5}$$

(4) Repeat (3) until the termination condition is satisfied. The termination criterion is that the number of residual signal extremum points does not exceed at most two. Finally, the original signal sequence is decomposed into N modal components and the residual term $R(t)$.

$$x(t) = \sum_{n=1}^{N}\mathrm{IMF}_n + R(t) \tag{6}$$

2.2. VMD principle

Compared with other time-frequency decomposition methods, VMD can effectively reduce the non-smoothness of reservoir water level time series with high complexity and strong non-linearity. The specific steps of VMD are as follows.

(1) Constructing variational problems
(2) Solving variational problems
(3) Find each mode component and center frequency

2.3. Bidirectional long- and short-term memory network

2.3.1. Long and short-term memory network

Conventional recurrent neural networks commonly employ a logistic nonlinear activation function for recurrent learning. However, due to the derivative values of the logistic function being bounded between 0 and 1, these networks encounter challenges when processing sequences with long time intervals. Specifically, when the time interval is substantial, the gradient tends to approach 0 or become excessively large, resulting in the issues of gradient disappearance or gradient explosion. Consequently, traditional recurrent neural networks face difficulties in effectively handling long-interval information sequences.

In 1997, Hochreiter & Schmidhuber introduced a variant of recurrent neural networks called the long-short-term memory network (LSTM) (Hochreiter & Schmidhuber 1997). This architecture incorporated a gating mechanism to address the issues of gradient explosion or disappearance encountered by traditional recurrent neural networks. The LSTM effectively manages the flow of information within each cell through the utilization of these gating mechanisms. Specifically, the LSTM consists of

three gates: the forget gate (denoted as f_t), the input gate (denoted as i_t), and the output gate (denoted as o_t). The internal structures of these gates are illustrated in Figure 1.

2.3.2. Bidirectional long and short-term memory network

While LSTM is effective for extracting sequential information in one direction, it may not fully capture the dependencies between the current state of a network and its preceding and subsequent states in the context of network security posture prediction. To enhance the prediction accuracy, a BILSTM is introduced. The BILSTM incorporates two LSTM layers that operate in both forward and reverse directions, allowing it to capture information from both past and future states. The structure of the BILSTM is depicted in Figure 2.

The output of the bidirectional long and short-term memory network (BILSTM) is determined through the collaboration of two LSTM layers. The forward LSTM layer performs computations in a forward direction, starting from the initial moment and progressing to the final moment. On the other hand, the reverse LSTM layer conducts computations in a reverse direction, starting from the final moment and moving back to the initial moment. Both layers follow the same computational process. Ultimately, the outputs of the forward and reverse layers at each moment are combined to yield the corresponding output for that moment (Chen *et al.* 2022).

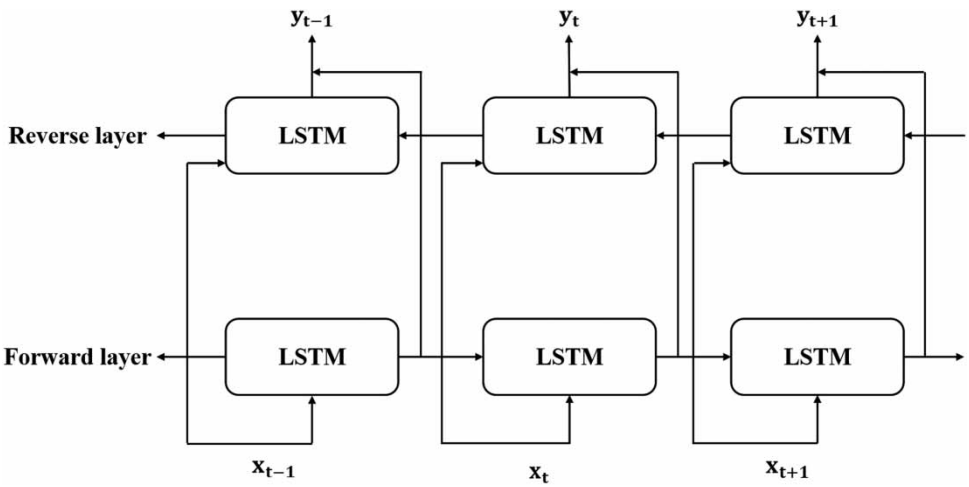


Figure 1 | LSTM basic structure.

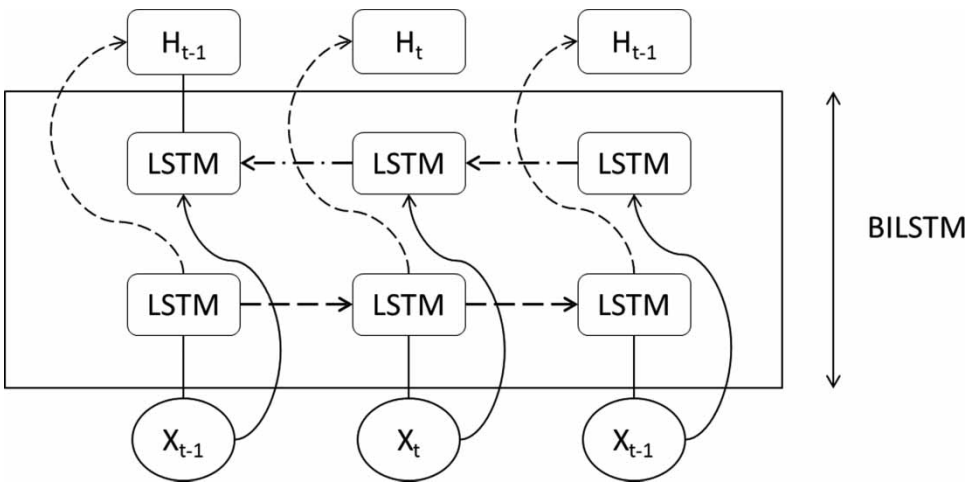


Figure 2 | Basic structure of BILSTM.

2.4. CEEMDAN–VMD–BILSTM combined model

The CEEMDAN–VMD–BILSTM combined prediction model follows a 'primary decomposition – secondary decomposition – prediction – reconstruction' framework. Initially, the original data undergo decomposition using CEEMDAN, yielding the intrinsic mode function (IMF) components and trend term of the time series. The highest frequency component IMF1 in these time series is decomposed by VMD twice, and the optimal number of VMD decompositions can be determined by the component center frequencies, and further extracting the component main information to obtain a number of variational modal components (VM). VMD not only simplifies the more complex high-frequency components into multiple subsequences but also ensures decomposition effectiveness with fewer components. After the decomposition stage, the BILSTM model is employed for predictions. Then, the predicted values are added up to form the predicted value of the new component IMF1, which further improves the prediction accuracy. Next, the training data and prediction data are segregated. The IMF components and trend terms of monthly precipitation from 2000 to 2018 serve as the training data for the BILSTM neural network, while the IMF components and trend terms from 2019 to 2020 are employed as the prediction data. The network parameters of the BILSTM neural network are fine-tuned to achieve the optimal training performance on the training data, thereby improving the prediction accuracy for the IMF components and trend terms of annual precipitation. Finally, the predicted data undergo reconstruction. The IMF components and trend terms obtained from the prediction phase are inverse normalized and reconstructed, resulting in the predicted values of monthly precipitation for 2019–2020. The overall technical route is illustrated in Figure 3.

2.5. Error analysis

To assess the prediction performance of the CEEMDAN–VMD–BILSTM model for precipitation, this study employs three widely recognized statistical indicators: root mean square error (RMSE), mean absolute error (MAE), and coefficient of determination (R^2). The calculations for these indicators are as follows (Luo *et al.* 2022):

$$\mathrm{RMSE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(P_i - P_i^*)^2} \tag{7}$$

$$\mathrm{MAE} = \frac{1}{n}\sum_{i=1}^{n}|P_i - P_i^*| \tag{8}$$

$$R^2 = 1 - \frac{\sum_{i=1}^{n}(P_i - P_i^*)^2}{\sum_{i=1}^{n}(P_i - \bar{P})^2} \tag{9}$$

3. CASE APPLICATION

3.1. Overview of the study area

Fuzhou, the capital city of Fujian Province, is situated in the eastern part of Fujian Province along the lower reaches of the Min River and the coastal area. Geographically, Fuzhou is positioned at the southeastern edge of the Eurasian continent, adjacent to the Pacific Ocean. It is situated at the mouth of the Min River on the southeastern coast of China and in the east-central part of Fujian Province, across the sea from Taiwan Province. The city's coordinates lie between latitude 25°15′ and 26°39′N and longitude 118°08′ and 120°31′E. Fuzhou is located to the east of the East China Sea, with Nanping and Sanming to the west, Ningde to the north, and Putian to the south. It resides in the Asia-Pacific economic circle on China's southeast golden coast. Fuzhou experiences a typical subtropical monsoon climate characterized by pleasant temperatures, warm and humid conditions, evergreen seasons, ample sunshine, abundant rainfall, minimal frost, and no snow. The city's basin topography makes it susceptible to flooding. Precipitation plays a vital role as a water source for Fuzhou, directly impacting the area's natural ecology. Figure 4 depicts the location map of Fuzhou City, while the monthly precipitation data collected from Fuzhou station between January 2000 and December 2020, sourced from the Fuzhou Water Resources Bulletin, consist of a total of 240 data points. The monthly precipitation data for Fuzhou City exhibit characteristics of randomness,

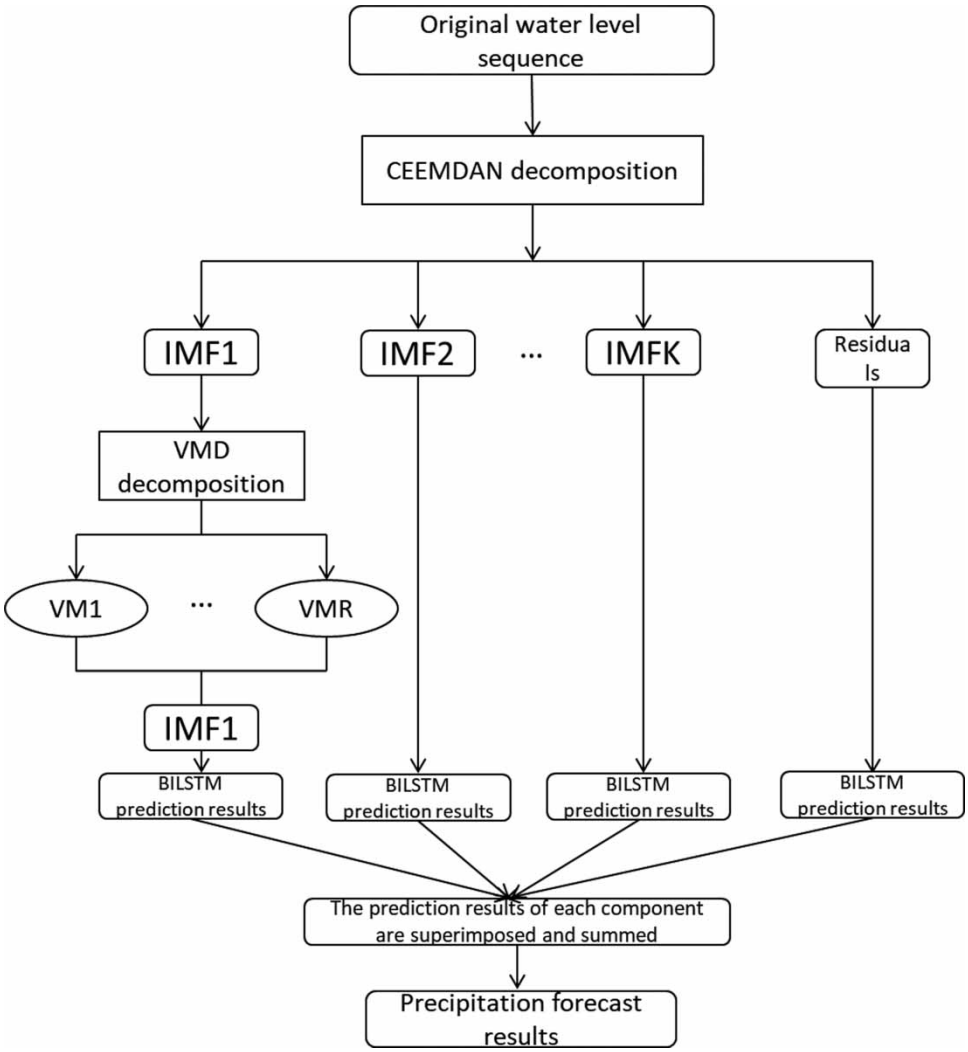


Figure 3 | Technology roadmap.

fluctuation, and non-linearity. Figure 5 illustrates the precipitation series of Fuzhou City, highlighting its pronounced non-stationarity and volatility.

3.2. Empirical modal decomposition

Using MATLAB software, we initially constructed the CEEMDAN decomposition model to decompose the monthly precipitation data of Fuzhou City spanning from 2000 to 2020. As the decomposition progressed, the amplitudes of IMF_2–IMF_7 and the trend term gradually diminished, the frequencies decreased, and the wavelengths became larger. Subsequently, we employed the MATLAB software to construct a VMD decomposition model, decomposing the IMF_1 component into five components and trend terms, as shown in Figure 6. Through the combined application of CEEMDAN and VMD, the monthly precipitation series underwent a reduction in volatility and improvement in non-stationarity.

3.3. Forecasting

The training samples consisted of the IMF components and trend terms extracted from the period 2000 to 2018, while the forecast samples comprised the IMF components and trend terms from 2019 to 2020. Using the CEEMDAN–VMD–BILSTM quadratic decomposition combination model, we made predictions for the monthly precipitation in Fuzhou City

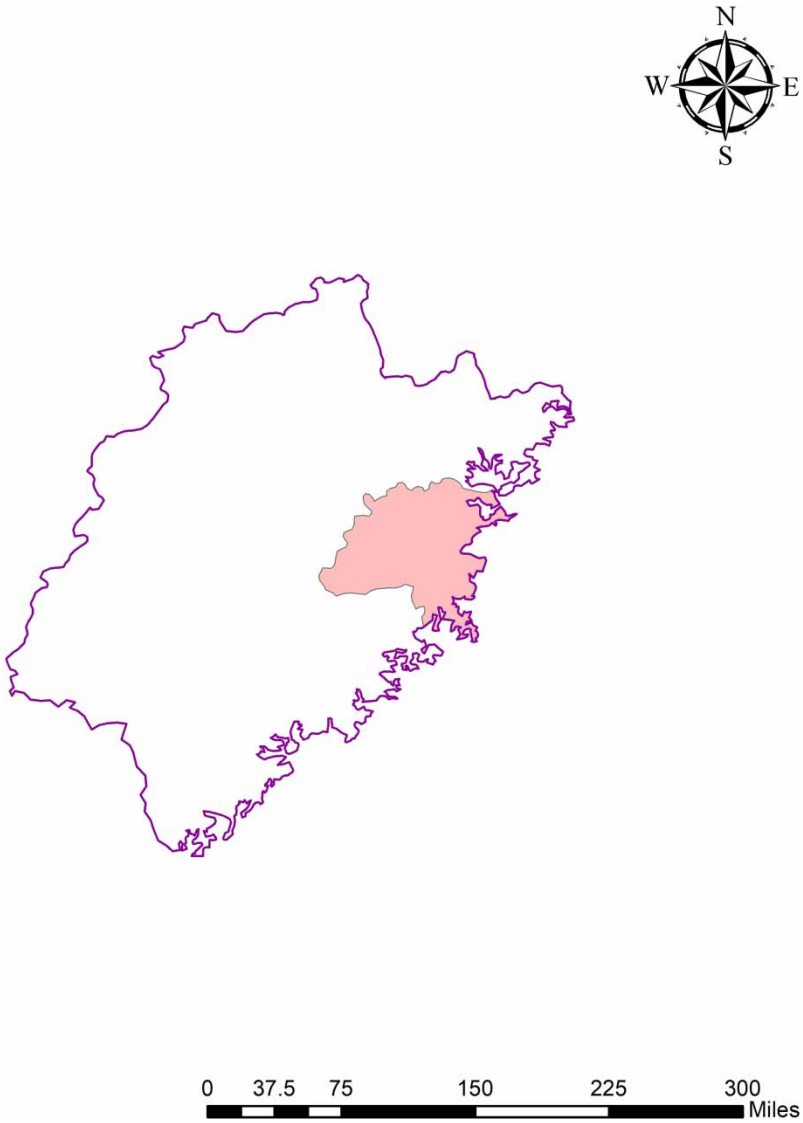


Figure 4 | Zhengzhou city location map.

for 2019–2020 and subsequently evaluated the model's effectiveness. Figure 7 illustrates a comparison between the predicted values and observed values for each component series.

Following the secondary decomposition, the precipitation time series in Fuzhou City exhibited enhanced smoothness and significantly reduced volatility. Moreover, the prediction errors for IMF_1 through IMF_7 progressively decreased, indicating an improved prediction effect with the secondary decomposition reconstruction.

The results in Table 1 show that IMF1 exhibits a small RMSE, MAE and a coefficient of determination close to 1, which can be attributed to its quadratic decomposition of the VMD. As we move from predicting IMF_1 to IMF_7, both the RMSE and MAE demonstrate a consistent downward trend. For IMF_2–IMF_7, there is a gradual reduction in the RMSE (from 0.7284 to 0.026316) and MAE (from 0.7488 to 0.25985). Additionally, the coefficient of determination progressively approaches 1, increasing from 0.89339 to 0.99222. Notably, IMF1 achieves relatively favorable levels of RMSE and MAE due to VMD's quadratic decomposition, resulting in improved prediction accuracy compared to the other components.

IMF_7, along with the residuals, exhibits a good fit. By referring to Figure 8, it becomes apparent that the predicted values of the residuals align well with the true values. As a result, after the preliminary CEEMDAN pre-processing, IMF1

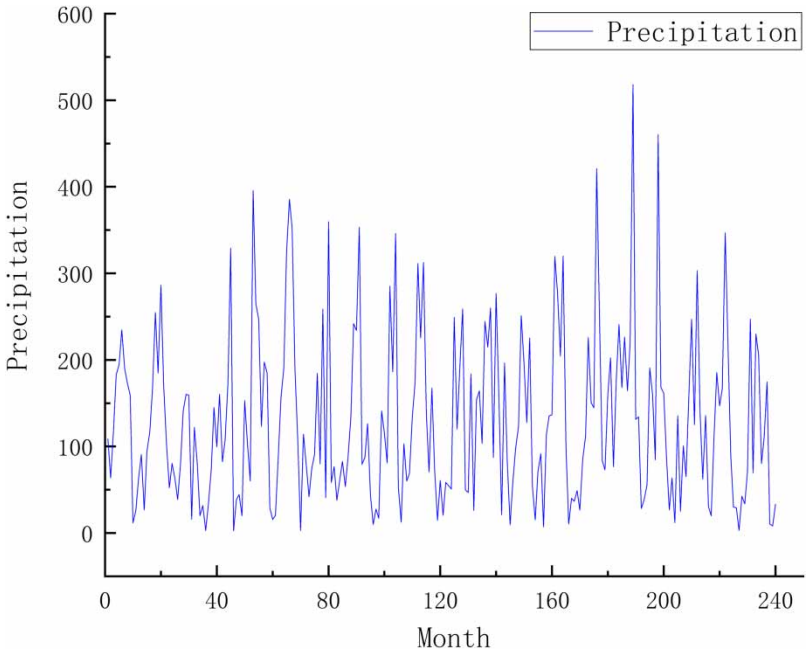


Figure 5 | Precipitation sequence in Zhengzhou City.

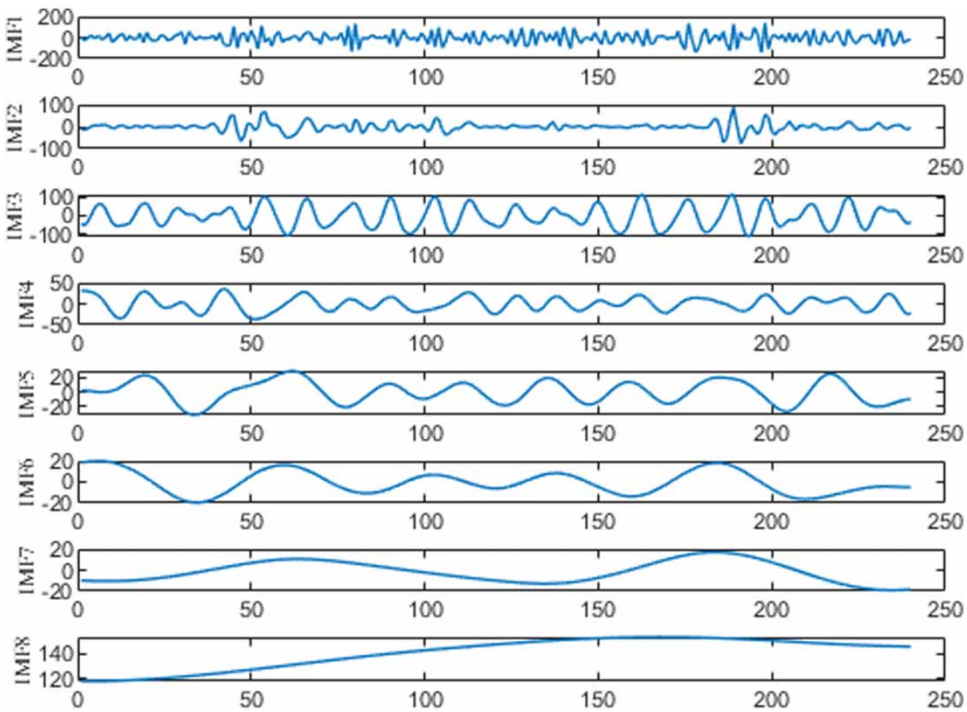


Figure 6 | CEEMDAN decomposition of monthly precipitation in Fuzhou.

undergoes further prediction following secondary decomposition using VMD. The employment of BILSTM for predicting IMF1–IMF7 and residuals yields a notable improvement. The prediction outcomes of IMF_1–IMF_7 and trend terms are then reconstructed from the data and compared against the original precipitation data of Fuzhou City, as demonstrated in Figure 9 and Table 2.

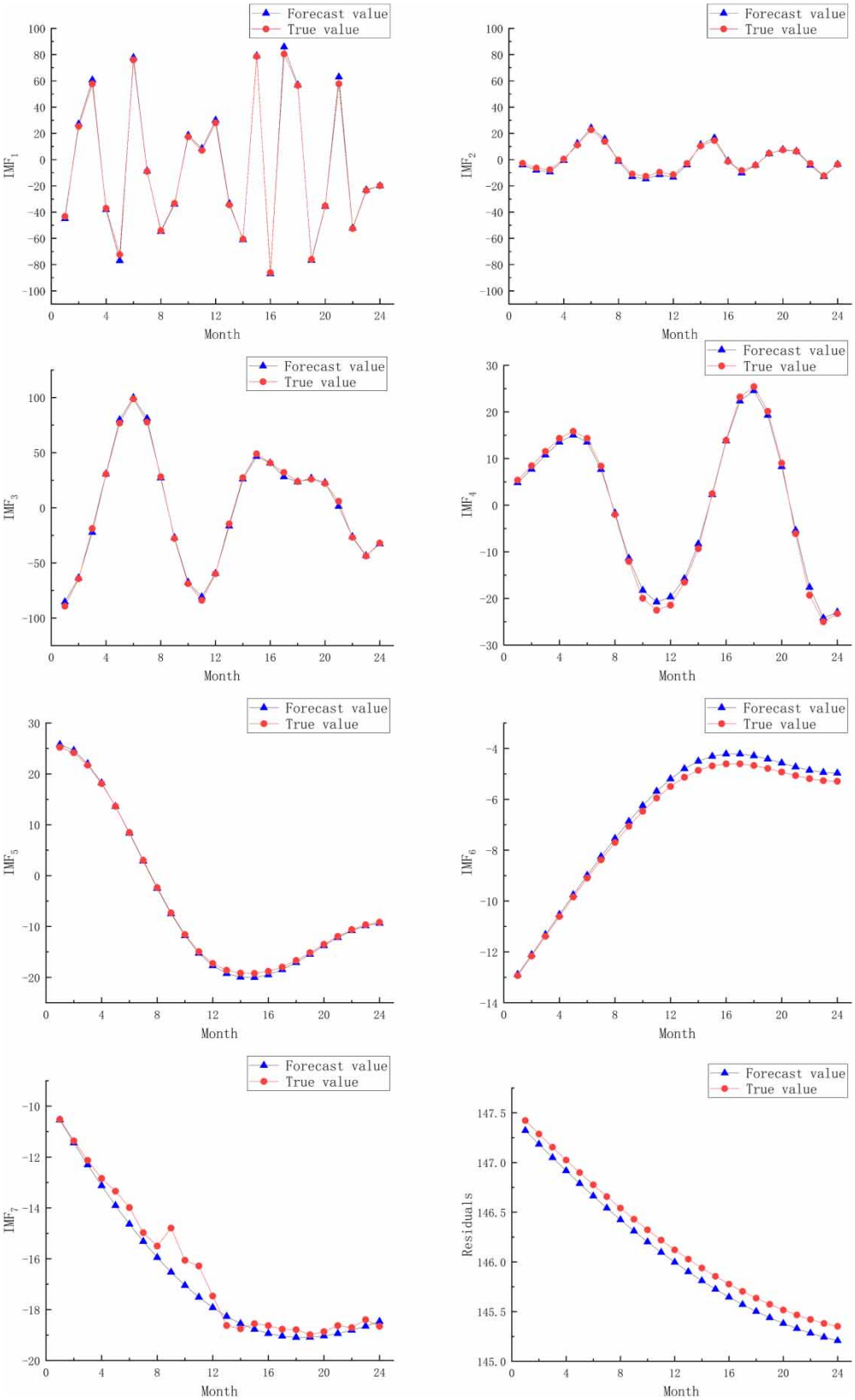


Figure 7 | Predicted versus observed values for each component series.

Table 1 | Errors and coefficients of IMF1–IMF7 and trend term forecasts

Sequence	RMSE	MAE	R^2
IMF_1	0.2108	0.2072	0.96365
IMF_2	0.7284	0.7488	0.89339
IMF_3	0.6693	0.5406	0.85274
IMF_4	0.4583	0.3943	0.93585
IMF_5	0.2333	0.21022	0.95999
IMF_6	0.16316	0.15985	0.97759
IMF_7	0.026316	0.025985	0.99222
Residuals	0.33433	0.25898	0.90367

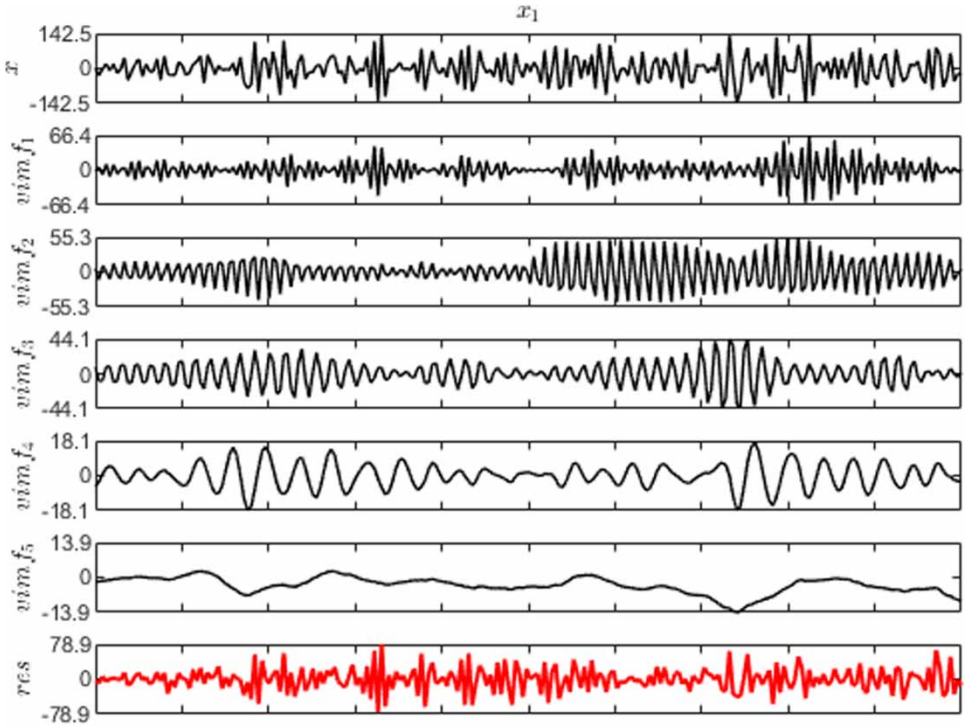


Figure 8 | VMD decomposition of IMF1 component of monthly precipitation in Fuzhou.

The results presented in Table 2 indicate that the combined CEEMDAN–VMD–BILSTM quadratic decomposition model achieves a low level of prediction error. The maximum relative error is 4.78%, the minimum is 0.00%, and the average relative error is 1.69%. These findings demonstrate that the model exhibits accurate predictions with a low relative error, a high success rate, and overall high prediction quality.

4. DISCUSSION

In order to assess the effectiveness of the CEEMDAN–VMD–BILSTM quadratic decomposition combination model, predictions were made using alternative models including the CEEMD–LSTM model, CEEMD–BILSTM model, and CEEMDAN–BILSTM model. The prediction results of the CEEMDAN–VMD–BILSTM quadratic decomposition combination model were then compared with those of the other models, as depicted in Table 3.

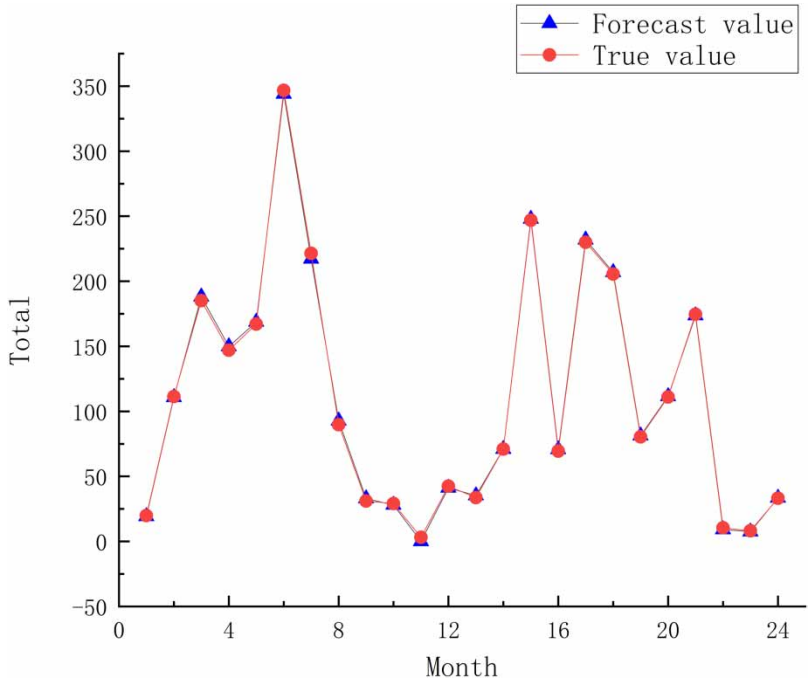


Figure 9 | CEEMDAN–VMD–BILSTM quadratic model prediction results compared with the original data.

As indicated in Table 3, the prediction accuracy of all three models falls within a reasonable range. The average relative errors of CEEMD–LSTM, CEEMD–BILSTM, CEEMDAN–BILSTM, and CEEMDAN–VMD–BILSTM were 15.89, 9.62, 5.10, and 1.69%, respectively. Among them, the combined CEEMDAN–VMD–BILSTM quadratic decomposition model demonstrates controlled prediction relative errors below 10%, with an average relative error smaller than that of the other models, resulting in superior prediction outcomes. The coupled models CEEMD–BILSTM and CEEMDAN–BILSTM also exhibit higher accuracy, with average relative errors below 10%, indicating a favorable performance at a lower level. Furthermore, the average relative error of the combined CEEMDAN–VMD–BILSTM quadratic decomposition model is lower than that of the coupled model CEEMDAN–BILSTM. This suggests that VMD quadratic decomposition contributes more effectively to the training of the BILSTM neural network. Moreover, the combined CEEMDAN–VMD–BILSTM quadratic decomposition model demonstrates significant advantages over other models in predicting monthly precipitation in Fuzhou. To gain a more visual understanding of the prediction effects of the four models, the prediction outcomes of the combined CEEMDAN–VMD–BILSTM quadratic decomposition model are compared to other models in Figure 10.

As observed in Table 3, the combined CEEMDAN–VMD–BILSTM quadratic decomposition model consistently outperforms the other three models in terms of average relative errors for precipitation prediction in Fuzhou City. Figure 9 further illustrates that the combined model exhibits superior precipitation prediction results compared to the other models. The combined CEEMDAN–VMD–BILSTM quadratic decomposition model demonstrates enhanced accuracy in predicting precipitation, while the coupled model effectively captures the realistic changes in the precipitation series with reasonable detail. Additionally, the trend and periodicity of the combined CEEMDAN–VMD–BILSTM quadratic decomposition model's prediction results align closely with the actual data.

Figure 9 and Table 3 reveal that the CEEMDAN–VMD–BILSTM quadratic decomposition model, followed by CEEMDAN–BILSTM, CEEMD–BILSTM, and CEEMD–LSTM models, yields the best prediction performance. The noise reduction prediction model significantly improves prediction accuracy compared to the single prediction model. Notably, the CEEMDAN–VMD–BILSTM prediction model demonstrates considerably higher accuracy than the other prediction models, closely resembling the actual measured precipitation variations. To further analyze the performance of different models, this study employs two metrics, namely MAE and NSE coefficient, to assess the accuracy of the models (Zhang

Table 2 | Relative error of the forecast month

Month	True value (mm)	Predicted value (mm)	Relative error (%)
217	19.93	19.27	3.31
218	111.44	110.78	0.59
219	185.15	188.11	1.60
220	147.02	149.82	1.90
221	167.02	168.70	1.01
222	346.73	343.92	0.81
223	221.46	217.18	1.93
224	89.69	92.79	3.46
225	31.02	31.02	0.00
226	29.16	28.09	3.67
227	3.10	3.04	1.73
228	42.53	41.16	3.22
229	33.64	35.25	4.78
230	71.03	70.96	0.10
231	246.97	247.94	0.39
232	69.39	70.76	1.98
233	229.95	231.87	0.84
234	205.60	206.91	0.64
235	80.32	81.35	1.27
236	111.08	111.46	0.35
237	174.64	173.66	0.56
238	10.45	10.15	2.81
239	8.20	8.43	2.89
240	33.20	33.47	0.81
Mean relative error (%)	1.69		

et al. 2017). The results are presented in Table 4.

$$\mathrm{MAE} = \frac{1}{n}\sum_{i=1}^{n}|P_i - P_i^*| \tag{10}$$

$$\mathrm{NSE} = 1 - \frac{\sum_{i=1}^{n}(\sigma_i - \sigma_0)^2}{\sum_{i=1}^{n}(\sigma_0 - \sigma)^2} \tag{11}$$

Here, σ_i is the predicted value at time i; σ_0 is the measured value at time i; σ is the mean value of the measured value.

Table 4 presents the comparative results of the model metrics. The MAE of CEEMD–LSTM, CEEMD–BILSTM, CEEMDAN–BILSTM, and CEEMDAN–VMD–BILSTM are 7.93, 6.04, 3.52, and 1.32 m. The NSE is 0.75, 0.78, 0.76, and 0.92, respectively, yielding the following analytical insights: as the single model is optimized towards the combined model, incorporating BILSTM treatment and more comprehensive VMD decomposition, the MAE index decreases, and the NSE approaches 1. This overall improvement indicates a higher preference for predictions. However, the single CEEMD–LSTM, CEEMD–BILSTM, and CEEMDAN–BILSTM models exhibit poor prediction performance, which can be attributed to the complexity, stochasticity, and ambiguity of precipitation sequences. The coupled CEEMDAN–VMD–BILSTM model stands out as optimal in terms of error reduction due to the combined use of CEEMDAN decomposition, VMD decomposition, and the BILSTM model, resulting in enhanced prediction performance.

Table 3 | Relative error of each model

		CEEMDAN–VMD–BILSTM		CEEMDAN–BILSTM		CEEMD–BILSTM		CEEMD–LSTM	
Month	True value (mm)	Predicted value (mm)	Relative error (%)	Predicted value (mm)	Relative error (%)	Predicted value (mm)	Relative error (%)	Predicted value (mm)	Relative error (%)
217	19.93	19.27	3.31	21.36	7.19	17.25	13.44	23.96	20.24
218	111.44	110.78	0.59	113.98	2.28	118.97	6.76	120.67	8.28
219	185.15	188.11	1.60	189.54	2.37	179.34	3.14	195.75	5.72
220	147.02	149.82	1.90	151.64	3.14	140.42	4.49	133.67	9.08
221	167.02	168.70	1.01	162.57	2.66	159.34	4.60	176.52	5.69
222	346.73	343.92	0.81	354.29	2.18	354.81	2.33	341.69	1.45
223	221.46	217.18	1.93	228.75	3.29	230.57	4.11	230.387	4.03
224	89.69	92.79	3.46	90.98	1.44	84.69	5.57	100.65	12.23
225	31.02	31.02	0.00	27.94	9.93	24.37	21.44	42.31	36.40
226	29.16	28.09	3.67	30.67	5.17	25.39	12.94	35.74	22.55
227	3.10	3.04	1.73	3.21	3.65	3.019	2.52	3.88	25.28
228	42.53	41.16	3.22	45.67	7.39	50.27	18.21	33.43	21.39
229	33.64	35.25	4.78	30.23	10.15	41.68	23.89	22.56	32.94
230	71.03	70.96	0.10	75.34	6.07	80.19	12.90	80.09	12.76
231	246.97	247.94	0.39	241.36	2.27	241.45	2.23	252.12	2.09
232	69.39	70.76	1.98	74.81	7.82	75.96	9.47	60.07	13.43
233	229.95	231.87	0.84	227.64	1.00	237.64	3.35	219.17	4.69
234	205.60	206.91	0.64	207.98	1.16	211.83	3.03	214.37	4.27
235	80.32	81.35	1.27	87.29	8.67	89.37	11.26	85.40	6.32
236	111.08	111.46	0.35	112.97	1.70	120.53	8.51	104.15	6.24
237	174.64	173.66	0.56	179.35	2.70	170.55	2.34	182.37	4.43
238	10.45	10.15	2.81	11.34	8.55	12.38	18.51	11.56	10.66
239	8.20	8.43	2.89	7.52	8.27	9.95	21.37	1.30	84.14
240	33.20	33.47	0.81	37.64	13.38	37.97	14.37	42.17	27.02
Mean relative error (%)		1.69		5.10		9.62		15.89	

In this study, we propose the CEEMDAN–VMD–BILSTM model for precipitation prediction in Fuzhou City and compare it with similar models used in previous studies (Özger *et al.* 2020; Serencam *et al.* 2022). The findings indicate that decomposition algorithms, wavelet noise reduction, and neural networks can be applied to other forecasting domains. Whether it is precipitation prediction, wind speed prediction, or drought prediction, these models demonstrate promising results. Our research builds upon CEEMDAN, effectively reducing time series reconstruction errors, VMD, effectively addressing non-smoothness in complex and nonlinear precipitation time series, and bidirectional long-short-term memory (BILSTM) model, effectively capturing long-term dependencies in time series. The constructed combined CEEMDAN–VMD–BILSTM quadratic decomposition model offers detailed and accurate predictions, reflecting the true variations in precipitation series with improved performance.

5. CONCLUSION

(1) The BILSTM model, with its incorporation of cell states and gate structure for information control, effectively captures long-term dependencies in time series. Complementing this, CEEMDAN reduces reconstruction errors in time series, while VMD mitigates non-smoothness in complex and nonlinear precipitation time series. By combining adaptive noise-CEEMDAN, variational modal decomposition (VMD), and bidirectional long-short-term memory (BILSTM), a

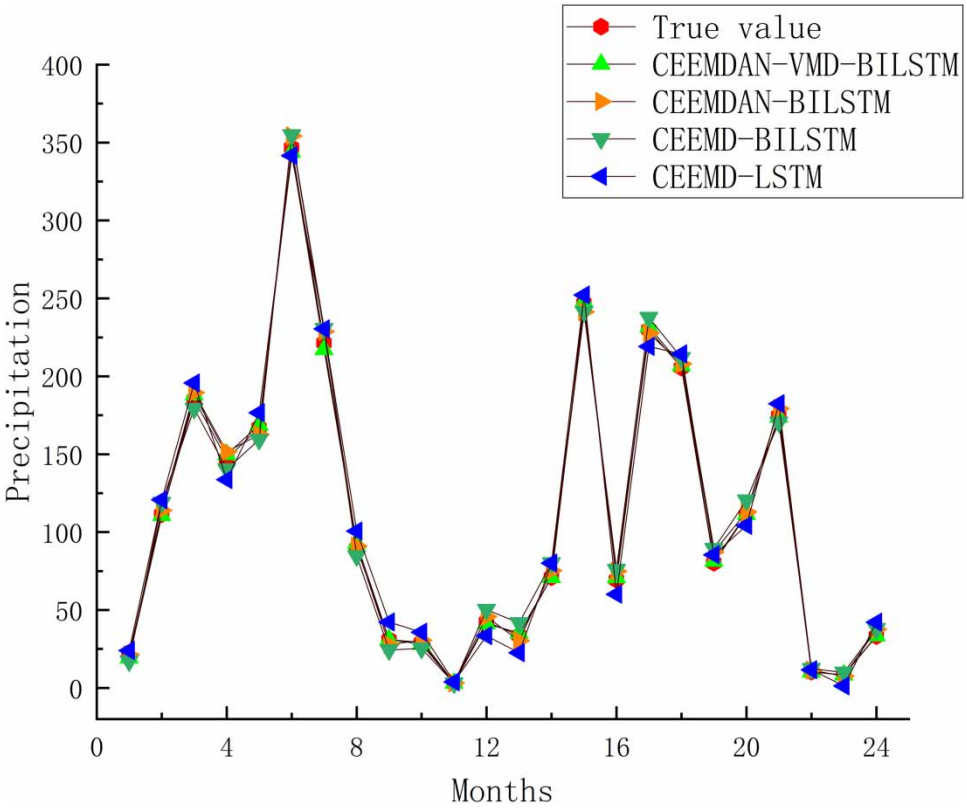


Figure 10 | Comparison of the prediction results of the CEEMDAN–VMD–BILSTM quadratic model with other models.

Table 4 | Comparison table of indicators of each model

	MAE (m)	NSE
CEEMD–LSTM	7.93	0.75
CEEMD–BILSTM	6.04	0.78
CEEMDAN–BILSTM	3.52	0.76
CEEMDAN–VMD–BILSTM	1.32	0.92

comprehensive model is constructed, yielding more accurate predictions and reflecting the true variations of precipitation series in greater detail. The model demonstrates a low average relative error of 1.69% and exhibits a higher level of prediction.

(2) A combined CEEMDAN–VMD–BILSTM quadratic decomposition model is developed and applied for urban monthly precipitation prediction. With an average relative error of 1.69% and a coefficient of determination close to 1, the model outperforms the CEEMD–LSTM, CEEMD–BILSTM, and CEEMDAN–BILSTM models in terms of prediction accuracy. Through the application of BILSTM processing and a more comprehensive VMD decomposition, the model achieves diminishing MAE index values and closer NSE values to 1. These findings confirm the feasibility of the combined CEEMDAN–VMD–BILSTM quadratic decomposition model for monthly precipitation prediction.

(3) It is important to acknowledge that the current model is data-based, and further enhancement is required to incorporate a deeper understanding of the underlying physical mechanisms. Future research endeavors should focus on strengthening the investigation of physical mechanisms within the model. This will contribute to a more comprehensive understanding and improve the model's performance in future predictions.

AVAILABILITY OF DATA AND MATERIALS

Data and materials are available from the corresponding author upon request.

AUTHOR CONTRIBUTION

All authors contributed to the study conception and design. X. Z. and J. S. wrote and edited the article. G. Z. edited the chart; Y. X. and H. C. collected the preliminary data. All authors read and approved the final manuscript.

FUNDING

This work was supported by the Key Scientific Research Project of Colleges and Universities in Henan Province (CN) [grant numbers 17A570004].

DATA AVAILABILITY STATEMENT

Data cannot be made publicly available; readers should contact the corresponding author for details.

CONFLICT OF INTEREST

The authors declare there is no conflict.

REFERENCES

Chen, L., Bi, X. & Zhou, X. 2022 Research on runoff prediction model based on EMD-ATT-BILSTM. *Modern Computer* **28** (1), 18–24.

Dong, G., Jia, D., Xue, H. & Lian, Y. 2020 Monthly runoff prediction in the Heihe River basin based on LSTM network. In *Proceedings of the 2020 Annual Conference of the Chinese Water Resources Society*, Vol. 3, October 18, 2020.

Han, Y., Guan, J., Cao, Y. & Luo, J. 2022 Application of LSTM-WBLS model in daily precipitation prediction. *Journal of Nanjing University of Information Engineering*, Nanjing, China.

Hochreiter, S. & Schmidhuber, J. 1997 Long Short-Term Memory. *Neural Computation* **9** (8), 1735–1780. http://dx.doi.org/10.1162/neco.1997.9.8.1735.

Hua, Y. 2022 *Research on Precipitation Prediction Based on GWO-DE-SVM Model*. Nanjing University of Information Engineering, Nanjing, China.

Jones, R. N. & Pittock, A. B. 2002 *Climate Change and Water Resources in an Arid Continent: Managing Uncertainty and Risk in Australia*, Springer, New York, NY, USA.

Kan, Y., Zhang, B. & Huang, H. 2022 Analysis of regional differences in climate change in Gansu over the past 58 years and the influence of circulation. *Highland Meteorology* **41** (5), 1291–1301.

Luo, S., Zhang, M., Nie, Y., Cao, R., Jia, X., Zhu, M. & Li, X. 2022 Monthly precipitation prediction in Zhengzhou city based on CEEMDAN-LSTM model. *Water Resources Planning and Design* **41** (5), 1291–1301.

Mohammadi, B., Moazenzadeh, R., Christian, K. & Duan, Z. 2021 Improving streamflow simulation by combining hydrological process-driven and artificial intelligence-based models. *Environmental Science and Pollution Research* **28** (46), 65752–65768.

Norden, E. H., Shen, Z., Long, S. R., Wu, M. C., Shih, H. H., Zheng, Q., Yen, N.-C., Tung, C. C. & Liu, H. H. 1998 The empirical mode decomposition and the Hilbert spectrum for nonlinear and non-stationary time series analysis. *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences* **454** (1971), 903–995.

Özger, M., Baakn Eyyup, E., Ekmekciolu, M. & Hacsüleyman, V. 2020 Comparison of wavelet and empirical mode decomposition hybrid models in drought prediction – ScienceDirect. *Computers and Electronics in Agriculture* **179**, 1–15.

Radhakrishnan, P. & Dinesh, S. 2006 An alternative approach to characterize time series data: case study on Malaysian rainfall data. *Chaos Solitons & Fractals* **27** (2), 511–518.

Serencam, U., Ekmekcioglu, O., Basakin, E. E. & Ozger, M. 2022 Determining the water level fluctuations of Lake Van through the integrated machine learning methods. *International Journal of Global Warming* **2022** (2), 27.

Shen, H., Luo, Y., Zhao, Z. & Wang, H. 2020 LSTM network-based prediction of summer precipitation in China. *Advances in Climate Change Research* **16** (03), 263–275.

Tan, H. R. 2022 *Depth-Width Learning Model and Its Application in Precipitation Prediction*. Nanjing University of Information Engineering, Nanjing, China.

Torres, M. E., Colominas, M. A. & Schlotthauer, G. 2011 A complete ensemble empirical mode decomposition with adaptive noise EEF international conference on acoustics. In *Speech and Signal Processing, 2013 (UCASSP) IFFE*. pp. 4144–4147.

Wei, L., Xu, S. & Zhang, Z. 2017 Research on multi-scale analysis and prediction of flood precipitation in Haihe River Basin. *Water Resources Planning and Design* **2017** (10), 60–63+86.

Wu, Z. & Huang, N. E. 2009 Ensemble empirical mode decomposition: a noise-assisted data analysis method. *Advances in Adaptive Data Analysis* **1** (1), 1–14.

Wu, X., Hua, Y., Guan, Y., Wang, W. & Liu, D. 2022 Research on precipitation occurrence prediction based on CNN-Attention-BP. *Journal of Nanjing University of Information Engineering: Natural Science Edition* **14** (2), 8.

Xue, C., Hou, W. & Zhao, J. 2013 The application of ensemble empirical mode decomposition method in multiscale analysis of region precipitation and its response to the climate change. *Acta Physica Sinica* **62** (10), 109203–109799.

Zhang, W., Qu, Z., Zhang, K., Mao, W., Ma, Y. & Fan, X. 2017 A combined model based on CEEMDAN and modified flower pollination algorithm for wind speed forecasting. *Energy Conversion and Management* **136**, 439–451.

Zhang, J., Xu, M., Zhang, X. & Xiao, H. 2021 Research on annual runoff prediction based on CEEMDAN-ARMA model. *People's Yellow River* **43** (1), 5.

First received 24 March 2023; accepted in revised form 31 July 2023. Available online 19 August 2023