Water Supply

IWA PUBLISHING

Water Supply Vol 23 No 1, 396 doi: 10.2166/ws.2022.426

Flow prediction in the lower Yellow River based on CEEMDAN-BILSTM coupled model

Xianqi Zhang[a,b,c], Wenbao Qiao[a,*], Jiafeng Huang[a], Jingwen Shi[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, Henan Province, Zhengzhou 450046, China
*Corresponding author. E-mail: 18839783482@163.com

ABSTRACT

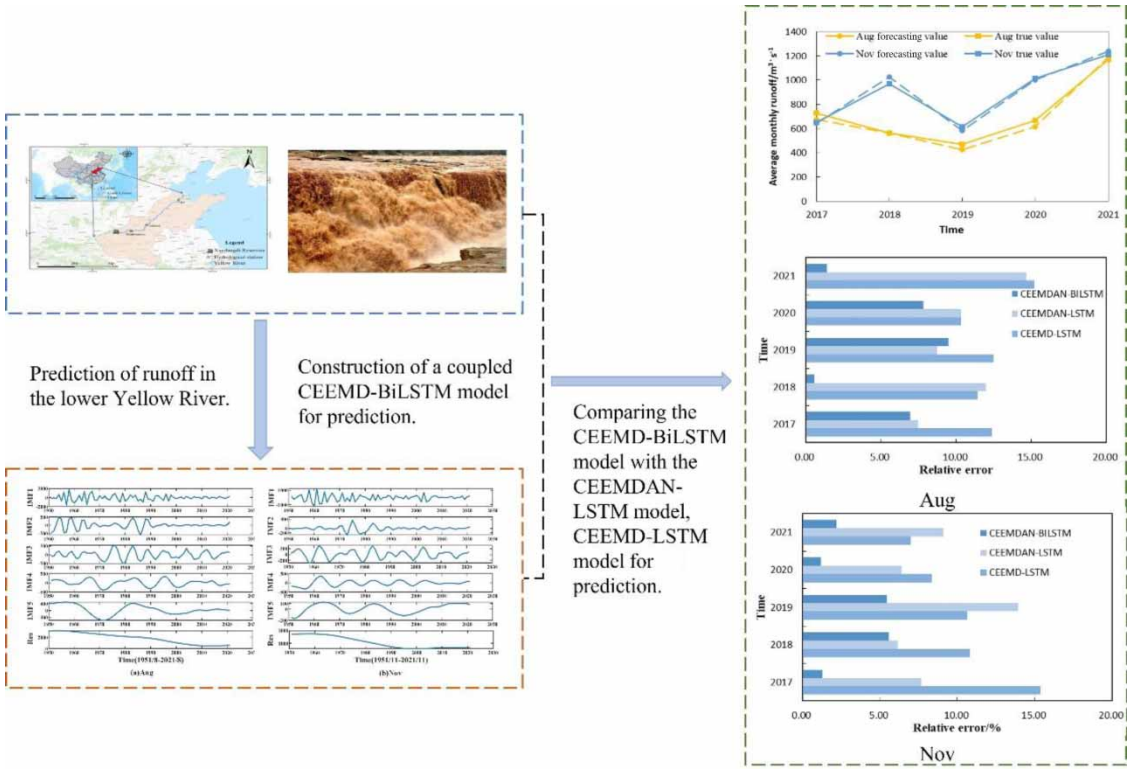
As one of the important hydrological elements of rivers, flow is of great significance to the development and utilization of water resources and the ecological environment. Based on the excellent nonlinear processing capability of CEEMDAN and the advantages of BILSTM in time-series data modeling, a coupled CEEMDAN-BILSTM model is constructed for flow prediction, and the i-month flows from 1951 to 2016 are used to predict the i-month flows from 2017 to 2021. The results show that the CEEMDAN-BILSTM coupled model predicts the trend more closely with the actual data variation, and the minimum relative error is 0.56 and maximum 9.48, which are maintained within 10%, and the deterministic coefficients are all greater than 0.9, so the prediction accuracy is high. The flow in month i of 5 years was picked up by monthly predictions for 66 consecutive years, which provides a new way of thinking about the prediction of river flow.

Key words: CEEMDAN-BILSTM, flow prediction, middle and lower reaches of the Yellow River, neural network

HIGHLIGHTS

- Predicting i-months in the 5 years of 2017–2021 by i-months in the 66 years of 1951–2016 can reduce the large flow volatility caused by the abundant and dry periods due to external conditions such as rainfall.
- CEEMDAN can effectively improve the modal mixing problem of EMD method, and the decomposition process of CEEMDAN is complete.

GRAPHICAL ABSTRACT



1. INTRODUCTION

Water is the most valuable natural resource in nature and the material basis for human life activities (Qian *et al.* 2009; Wang 2010). As an important part of water resources, rivers are involved in integrated development and utilization, scientific management, and optimal scheduling of water resources (Burger *et al.* 2007). Flow is one of the important hydrological variables of rivers, which is influenced by many uncertainties such as meteorological and human factors, and has strong randomness. Improving the accuracy of long-term river flow forecasting is important for scientific understanding of runoff processes, grasping the causes of changes and evolutionary characteristics, and optimizing water resource allocation and management (Tang 2008). Currently, the approaches to flow prediction can be broadly classified into two categories: hydrological models based on physical mechanisms and data-driven models (Ma *et al.* 2016). Hydrological model prediction methods based on physical mechanisms such as genesis analysis and hydrostatistical methods can only reflect linear time series or simple nonlinear time series (Yu *et al.* 2018; Zhao *et al.* 2009; Farajzadeh and Alizadeh 2018). Traditional hydrological models simulate flow prediction with poor accuracy. The data-driven model is usually based on a large amount of hydrological data accumulated by hydrological stations over the years, and establishes mathematical relationships to seek the optimal description of the forecast object and the forecast factor, such as artificial neural network models, regression analysis models, and other artificial intelligence models (Jiang *et al.* 2020; Li *et al.* 2021). Based on the good nonlinear mapping capability of neural network models of artificial intelligence and the unique advantages in big data processing and feature extraction (Zhao *et al.* 2009), such models do not need to consider the physical mechanisms of hydrological processes, are well-adapted to complex time series data, and have unique advantages in solving hydrological forecasts as well as multi-characteristic and multi-regular hydrological processes (Liu *et al.* 2021). In recent years, several common artificial intelligence models, such as Artificial Neural Networks (ANN), Random Forests, Recurrent Neural Networks (RNN), and Support Vector Machines (SVM), have been introduced in the field of hydrological forecasting and have achieved considerable prediction results (Burger *et al.* 2007; Nourani *et al.* 2011). However, it is easy to ignore the continuity and dependence implied by the sequence itself and treat it simply as a purely random sequence. It is still difficult for a single prediction model to effectively fit the entire hydrological process. Therefore, the combined idea of 'decomposition-prediction-reconstruction' provides a new way to solve this problem, which

is essentially a combination of signal decomposition methods and time series analysis methods (Wang *et al.* 2020; Xu *et al.* 2020). The overall performance of this class of methods is excellent and can effectively improve the predictive power of the model. Based on such models, a large number of studies have been conducted by domestic and foreign scholars. S. Zhang *et al.* (2020) and X.Q. Zhang *et al.* (2020) conducted a prediction study of the annual runoff of the lower Yellow River using a combined MEEMD-ARIMA model, which was found to be superior to the EEMD-ARIMA and CEEMD-ARIMA models. Komasi & Sharghi (2016) combined wavelet analysis with a support vector machine model for predictive analysis of rainfall-runoff processes in the Aghchai and Eel river basins. Park *et al.* (2014) used a real-time runoff prediction model with a hybrid ANN and Genetic Algorithm (GA) model and validated the feasibility and accuracy of the model using the Sumji watershed in Korea. Feng & Pan (2018) combined Long Short Term Memory (LSTM) with feedback (Back Propagation, BP) neural networks to successfully perform multiple successful runoff forecasts in the Ziwu River basin. Sen Zhang *et al.* (2020) and X.Q. Zhang *et al.* (2020) constructed a hybrid model based on the Multiple Population Genetic Algorithm (MPGA) with LSTM and used it for delayed regression prediction in the Jinjiang River basin. Li *et al.* (2021) used Empirical Mode Decomposition (EMD) to extract multi-scale runoff series features, and a hybrid differential Integrated Moving Average (ARIMA) autoregressive model with Generalized Regression Neural Network (GRNN) model to simulate the process of IMF components at different scales and effectively improve prediction accuracy.

The Complete Ensemble Empirical Mode Decomposition With Adaptive Noise (CEEMDAN) is a signal decomposition method proposed by Torres *et al.* (2011). This method is an improved algorithm of EMD (Napolitano *et al.* 2011) and EEMD (Wu & Huang 2009). By adding adaptive white noise in the decomposition stage several times, it can effectively improve the modal mixing problem of EMD method, and the decomposition process of CEEMDAN is complete. The method of signal decomposition based on the time scale characteristics of the data itself can accurately reconstruct the original signal and improve the decomposition efficiency. CEEMDAN was used to decompose the flow data, and the decomposed series had good smoothness. BILSTM neural network solves the problem that LSTM can only advance from one direction, and can be trained with sequence input from the forward and reverse layers. Due to the strong volatility of runoff, this paper combines the advantages of CEEMDAN and BILSTM and proposes a new coupled CEEMDAN-BILSTM model to predict the flow in the middle and lower reaches of the Yellow River. The water depth time series are smoothed using CEEMDAN decomposition, and the original series is decomposed into multiple components containing the information of the original series features, and then the individual components are input into the BILSTM model. In addition, to further improve the model accuracy and reduce the influence of rainfall variability in different months on the monthly average flow, a month-by-month prediction approach was adopted, i.e., a month i of 65 consecutive years from 1951 to 2016 was used to predict month i of 2017 to 2021. In this paper, we take August of the abundant water period and November of the dry water period as examples. The coupled model provides a new idea for river flow prediction.

2. MODELS AND METHODS

2.1. CEEMDAN model

CEEMDAN algorithm is developed based on the EMD algorithm, which not only integrates the advantages of EMD algorithm that can decompose any signal but also reduces the modal aliasing phenomenon to a certain extent. Meanwhile, the CEEMDAN algorithm overcomes the problems of incomplete decomposition, large reconstruction error, and residual exposure sound of the EEMD algorithm, and the cycle iteration takes only half the time of the EEMD algorithm (Chen *et al.* 2021; Jiang *et al.* 2021). The CEEMDAN algorithm decomposes the complex time series signal into a series of simple Intrinsic Mode Functions (IMF) and a residual term (Res).

Denote the eigenmode function obtained from the CEEMDAN decomposition as IMF_n, define the operator $E_c(\cdot)$ as the c order eigenmode component generated by the EMD algorithm, w_i is the white noise signal satisfying the N(0, 1) distribution, and $x(t)$ is the original signal, then the CEEMDAN method can be specifically described as:

(1) First-order eigenmode components obtained by decomposition with EMD algorithm

$$IMF_1(t) = \frac{1}{M}\sum_{i=1}^{m} IMF_1^i(t) \tag{1}$$

(2) Calculate the residuals after the first-order eigenmode decomposition

$$r_1(t) = x(t) - IMF_1(t) \tag{2}$$

(3) Conducting i experiments ($i - 1, \cdots, M$), The signal $r_1(t) + \varepsilon_1 E_1((w_i(t))$ is decomposed until the eigenmode component of the first-order EMD is obtained. and on this basis, continue to calculate the second-order eigenmode components

$$IMF_2(t) = \frac{1}{M}\sum_{i=1}^{m} E_1(r_1(t)) + \varepsilon_1 E_1((w_i(t))) \tag{3}$$

(4) For each of the remaining stages, i.e., $n = 2, \cdots, N$, calculate the signal of the first residual

$$r_2(t) = r_{n-1}(t) - IMF_n(t) \tag{4}$$

(5) Repeating the steps of Equation (3), the intrinsic mode component of order

$$IMF_{k+1}(t) = \frac{1}{N}\sum_{i=1}^{m} E_1(r_k(t)) + \varepsilon_k E_k((w_i(t))) \tag{5}$$

(6) Repeat Equations (4) and (5) until the residual signal satisfies the termination condition of the decomposition, and finally obtain a modal component. The final residual signal of the decomposition is:

$$R(t) = X(t) - \sum_{i=1}^{k} IMF_i(t) \tag{6}$$

The final raw signal can be decomposed as:

$$X(t) = R(t) - \sum_{i=1}^{k} IMF_i(t) \tag{7}$$

2.2. BI-LSTM neural network training

Directional Long Short-Term Memory (BILSTM) (Schuster & Paliwal 2002) is a combination of forwarding LSTM and backward LSTM (Ma *et al.* 2021). BILSTM neural network has a bi-directional property by constructing a pair of LSTM layers with opposite directions, i.e., adding a reverse LSTM layer to the original one. The two-way network enables recursive response training for both the states of the implicit layer at the beginning and the end of the sequence, which can further explore some connection between the current data and the data at the previous moments, as well as with the data at the future moments, and then deal with the reverse information more deeply to optimize the long-term dependence problem and improve the prediction accuracy of the model, whose structure is shown in Figure 1.

The BILSTM model analyzes the forward and reverse sequences through two independent hidden layers, and the final output prediction value in the BILSTM prediction model is determined by both the forward and reverse hidden layers. The output sum of the forward and reverse implicit layers and the output of the predicted values are calculated as follows.

The output of the forward implicit layer is

$$\vec{h}_t = \sigma(W_{x\vec{h}_t} x_t + W_{\vec{h}\vec{h}} \vec{h}_{t-1} + b_{\vec{h}}) \tag{8}$$

The output of the reverse implicit layer is

$$\overleftarrow{h}_t = \sigma\left(W_{x\overleftarrow{h}_t} x_t + W_{\overleftarrow{h}\overleftarrow{h}} \overleftarrow{h}_{t-1} + b_{\overleftarrow{h}}\right) \tag{9}$$

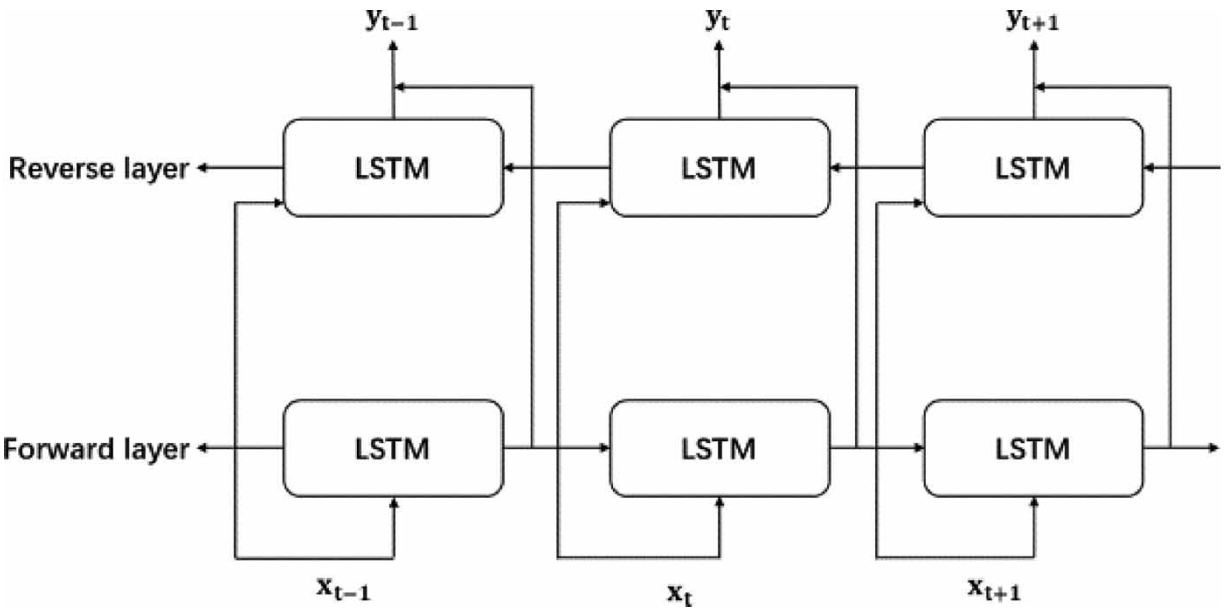


Figure 1 | BILSTM structure diagram.

In Equations (8) and (9), $\vec{h}_t$ is the positive implicit layer state output. $W_{x\vec{h}_t}$ is the output weight of the implicit layer of the forward propagation unit. $W_{\vec{h}\vec{h}}$ is the weight of the forward propagated previous moment state quantity to the current moment state quantity. $\vec{h}_{t-1}$ is the output value of the implied layer state at the previous moment of forwarding propagation. $b_{\vec{h}}$ is the biased term for forward propagation. Similarly, $\overleftarrow{h}_t$ is the reverse implied layer state output; $W_{x\overleftarrow{h}_t}$ is the backpropagation unit implied layer output weight; $W_{\overleftarrow{h}\overleftarrow{h}}$ is the weight of the backpropagated previous moment state quantity to the current moment state quantity; $\overleftarrow{h}_{t-1}$ is the backpropagated previous moment implied layer state output value; and $b_{\overleftarrow{h}}$ is the biased term of backpropagation.

2.3. CEEMDAN-BILSTM coupling model

The flow presents non-linearity and non-smoothness; considering the advantages of CEEMDAN in dealing with data volatility and the excellent performance of BILSTM in modeling time-series data, this paper tries to establish a coupled CEEMDAN-BILSTM model to improve the prediction accuracy, the structural model is shown in Figure 2, and the specific calculation process is shown as follows.

(1) CEEMDAN decomposition of the original data using MATLAB to obtain the different components IMF and the residual term.
(2) The data from 1951 to 2016 were used as training data for BILSTM and the data from 2017 to 2021 were used as test data.
(3) The decomposed IMF data are input into BILSTM, and the training degree of the BILSTM neural network on the training data is optimized by continuously tuning the grid parameters of the BILSTM neural network, to improve the prediction accuracy.
(4) Forecasts of IMF components and trend volumes for 1951–2016 by adjusted BILSTM.
(5) The IMF components and trend terms obtained from the forecast are reconstructed to obtain the forecast values of each component for 2017–2021.
(6) In turn, the process of (1)–(6) will be repeated to obtain the predicted values for different months. The specific flow chart is shown in Figure 2.

2.4. Statistical evaluation indicators

To more clearly reflect the error of the model prediction results and the prediction accuracy, four indicators, the Deterministic Coefficient (DC), the Mean Absolute Error (MAE), the Mean Absolute Percentage Error (MAPE), and the Root Mean

Square Error (RMSE), were used for the analysis, and the formulae were calculated as follows.

$$DC = 1 - \frac{\sum_{i=1}^{n} [y_c(i) - y_0(i)]^2}{\sum_{i=1}^{n} [y_0(i) - \bar{y}_0]^2}, i = 1, 2, \ldots, n \tag{10}$$

$$MAE = \frac{1}{n} \sum_{i=1}^{n} |y_c(i) - y_0(i)|, i = 1, 2, \ldots, n \tag{11}$$

$$MAPE = \sum_{i=1}^{n} \left| \frac{y_0(i) - y_c(i)}{y_0(i)} \right| \times \frac{100}{n}, i = 1, 2, \ldots, n \tag{12}$$

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^{n} [y_0(i) - y_c(i)]^2}, i = 1, 2, \ldots, n \tag{13}$$

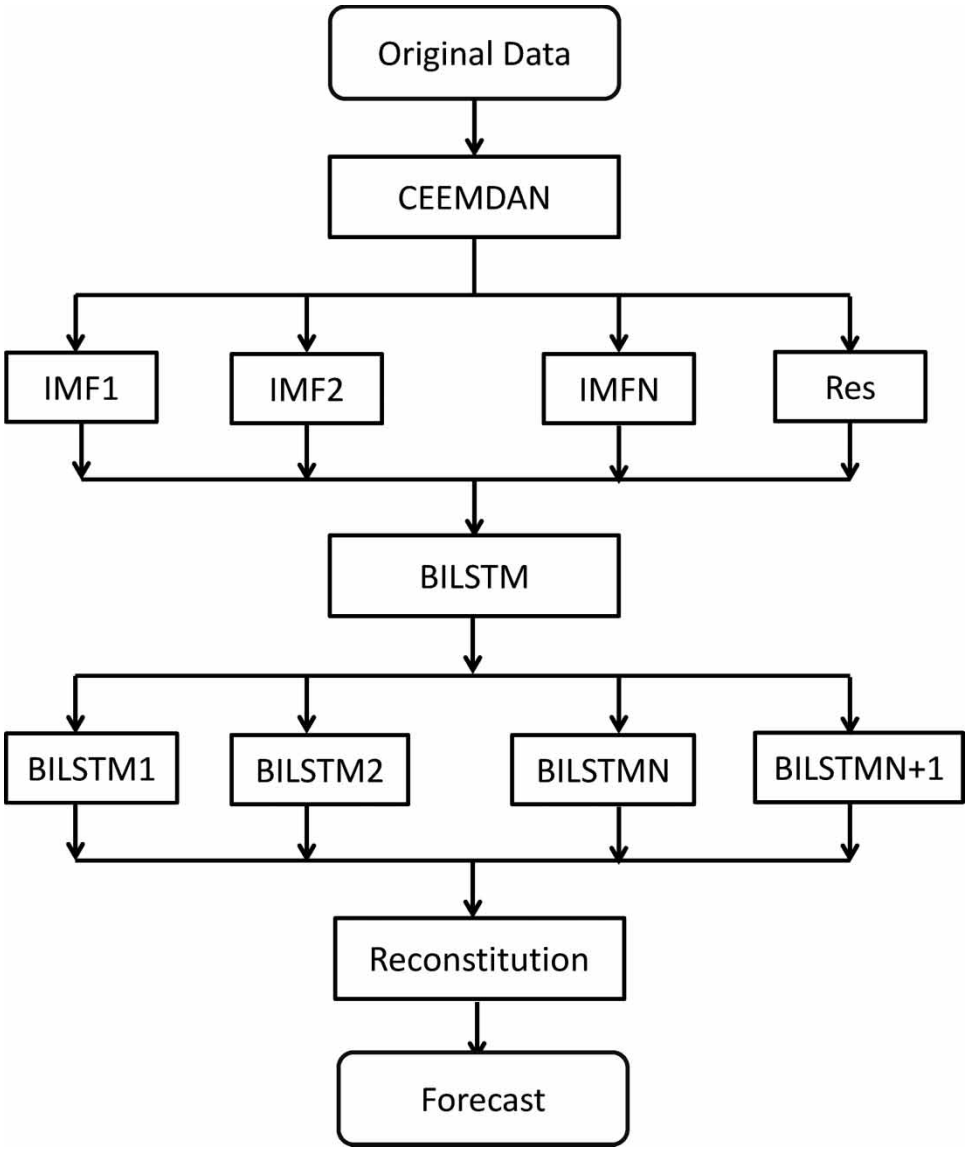


Figure 2 | Flow chart of CEEMDAN-BILSTM coupling.

where, $y_c(i)$ is the predicted value, $y_0(i)$ is the measured value, $\bar{y}_0$ is the mean value of the measured value, and n is the length of the series.

According to the Specification of Hydrological Information Forecasting (SL250–2000), the closer the DC value is to 1, the better the model prediction effect is. When DC $\geq$ 0.9, the prediction accuracy is grade A; when 0.7 $\leq$ DC $<$ 0.9, the prediction accuracy is grade B; when 0.5 $\leq$ DC $<$ 0.7, the prediction accuracy is grade C; when DC $<$ 0.5, the prediction result is not credible.

3. EXAMPLE APPLICATIONS

3.1. Data processing

GaoCun hydrological station is an important control station of the Yellow River flowing into Shandong, and the test section is located in the lower reaches of the Yellow River, at the end of the wandering section. The location is shown in Figure 3. Rainfall is an important factor affecting the flow of this section of the river, and the spatial and temporal distribution of rainfall in the region is not uniform, with rainfall mainly concentrated in summer and autumn. To improve the accuracy of the prediction and reduce the volatility caused by the large difference between the flow during the abundant and dry periods, the data of the ith month of the consecutive years 1951–2016 were used to predict the data of the ith month of 2017–2021. In this paper, the monthly average flow data from 1951 to 2021 at the GaoCun hydrological station in the middle and lower reaches of the Yellow River are used for the study. The data from this station are obtained from the official website of the Yellow River Commission, so the accuracy and reliability are high, and the time series is relatively long and representative, which can effectively reflect the flow variation characteristics of the middle and lower reaches of the Yellow River region. The data are divided into two parts: training data and validation data, where 1951–2016 data are the training numbers and 2017–2021 are the validation data. In this study, August in the abundant water period and November in the dry water period are used. The monthly average flow data are shown in Figure 4 below.

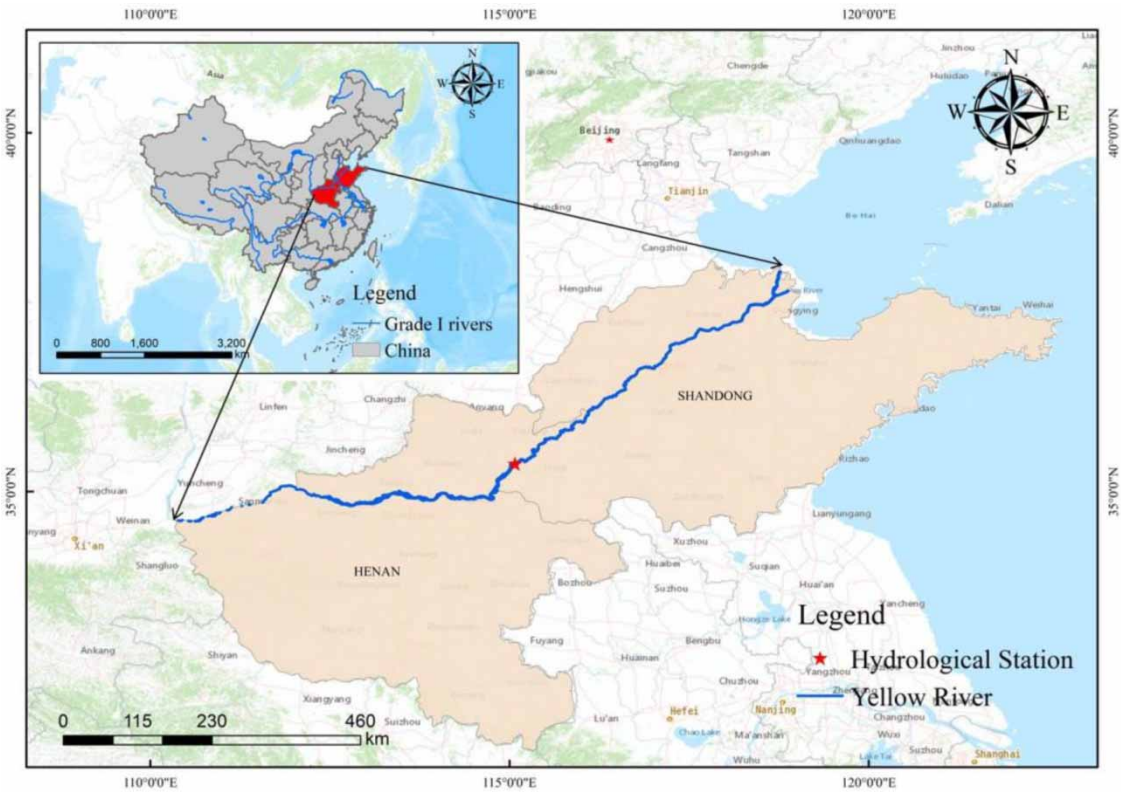


Figure 3 | Study area location.

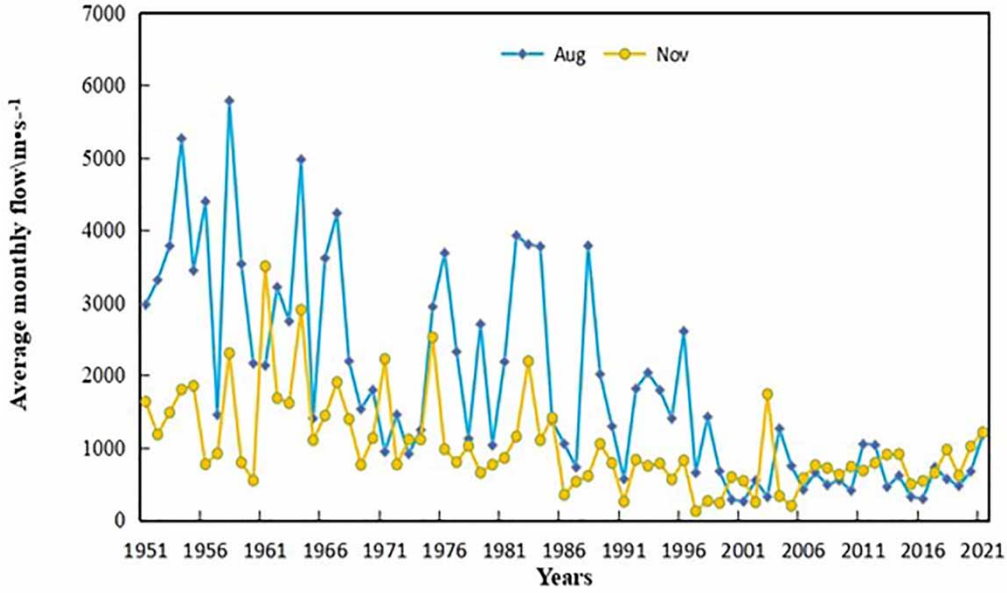


Figure 4 | Average monthly flow for August and November 1951–2021.

3.2. CEENDAN decomposition

The abundant water period in the lower Yellow River is usually July, August and September and the dry water period is January, February, November and December. In this study, samples were taken in August during the abundant water period and in November during the dry water period. The CEEMDAN decomposition model was constructed by using MATLAB to decompose the monthly flow data of Takamura hydrological station from 1951 to 2021 to obtain five IMF components and one residual Res. The decomposition results are shown in Figure 5 below.

In Figure 5 the IMF component shows the variation of the flow time series in terms of frequency, amplitude and wavelength. Among them, the IMF1 component has the highest frequency, the largest amplitude, and the shortest wavelength, and it can be seen from IMF1–IMF5 that the amplitude of the IMF component gradually decreases, the frequency gradually decreases and the wavelength gradually becomes longer as time passes. The volatility and non-stationarity of the series are greatly reduced after the flow data are decomposed by CEEMDAN (Zhang *et al.* 2018). It can be seen that the fluctuation range of each IMF component in August is higher than that in November. The remaining term represents the overall

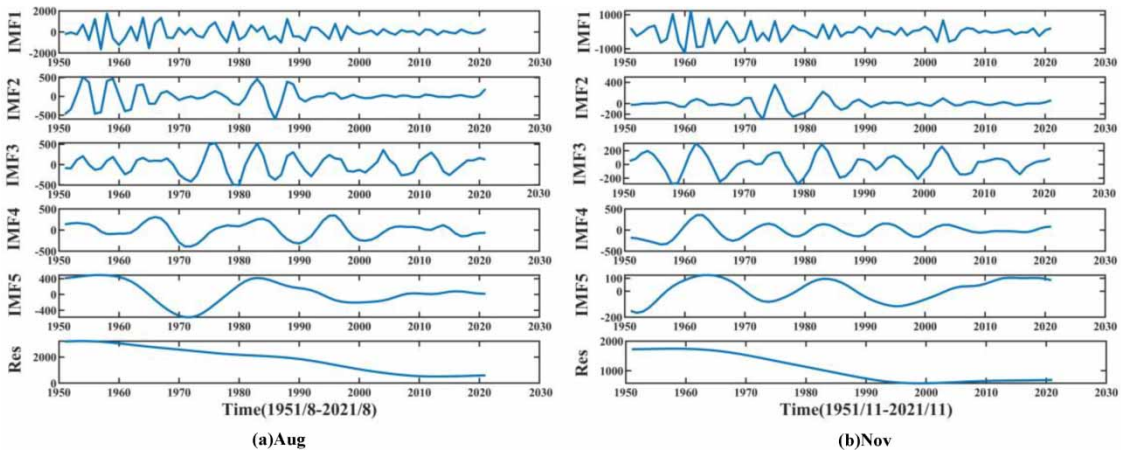


Figure 5 | Monthly average flow decomposition results for 1951–2019 at Takamura hydrological station.

trend of the monthly average flow at Takamura station from 1951 to 2021, with a decreasing trend in August and a decreasing and then increasing trend in November from 1951 to 2021.

3.3. BILSTM monthly average traffic forecast

To improve the prediction accuracy, the coupled CEEMDAN-BILSTM model of 'decomposition-prediction-reconstruction' is used, and the original data are divided into training samples and test samples. The years 1951–2016 are the training sample and the years 2017–2021 are the test sample. Through extensive experiments, the optimal number of hidden nodes for the derived network model is 37, the optimization algorithm Adam, the maximum number of iterations 500, and the gradient threshold 1.

The projections were made for each month's component and residual in turn, and then compared with the actual values, and the results are shown in Figure 6 below.

As can be seen from Table 1, the volatility errors of IMF1–3 are large, and the predicted values of IMF4–Res almost coincide with the actual data with small differences in values. The average relative errors of IMF1–3 in November are 19.92, 13.52, 11.99, respectively, indicating that the smoothness of the first three components is poor. The mean relative errors of IMF4–Res in August are 3.68, 6.35, 0.92, respectively; the mean relative errors of IMF4–Res in November are 7.97, 1.98, 0.62, respectively. The overall IMF1–Res error decreases sequentially, indicating that the component after CEEMDAN decomposition is smoother and the error is getting smaller. To test the prediction accuracy of the coupled CEEMDAN-BILSTM model more intuitively, each component of the prediction was reconstructed, and the results are shown in Figure 7 and Table 2.

From Figure 7 and Table 2, it can be seen that the coupled CEEMDAN-BILSTM model is used to predict the average monthly flow at the middle and lower reaches of the Yellow River at GaoCun hydrological station, and the trend of the prediction results is consistent with the actual trend, with a minimum error of 0.56 and a maximum error of 9.48, and the fluctuation range is within 10%, indicating that the prediction accuracy of this coupled model is high.

To further confirm the accuracy of the CEEMDAN-BILSTM prediction model, four indicators, namely, the coefficient of certainty (DC), the mean absolute error (MAE), the mean absolute percentage error (MAPE) and the root mean square error (RMSE), were selected for analysis, and the results are shown in Table 3 below. From Table 3, it can be seen that the

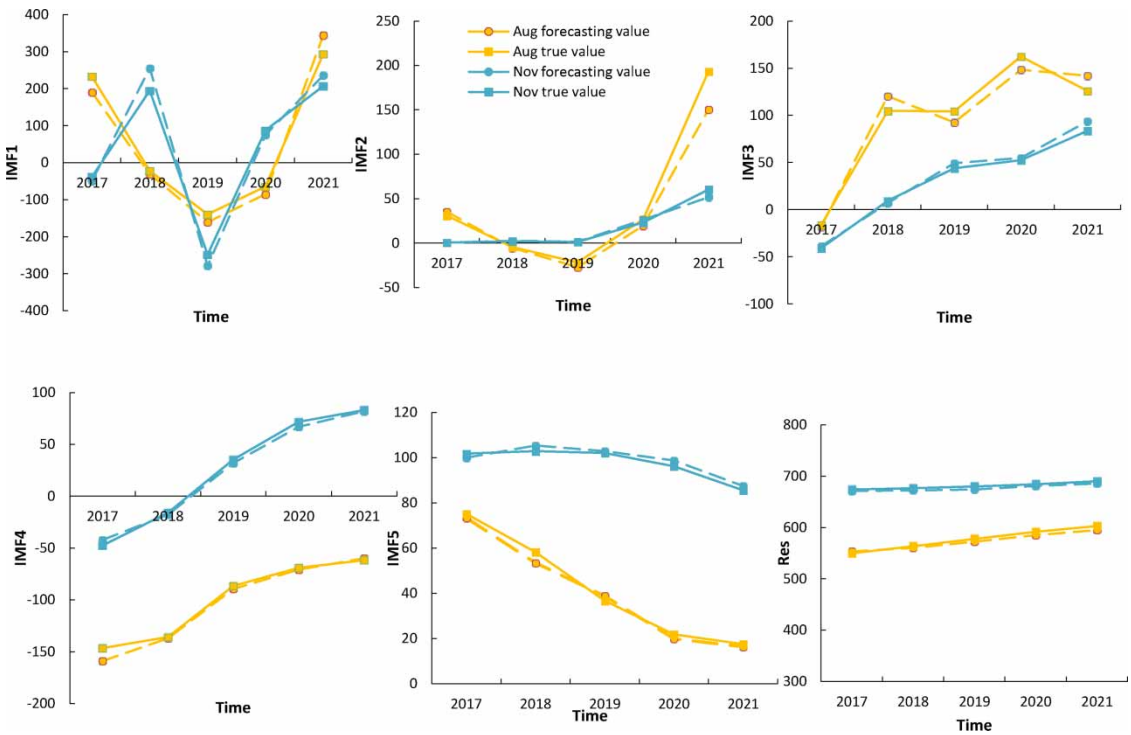


Figure 6 | Comparison of the components in August and November.

Table 1 | The relative error index of IMF

Month	Relative error	Maximum (%)	Minimum (%)	Average (%)
Aug	IMF1	38.42	14.92	24.17
	IMF2	27.01	14.18	22.03
	IMF3	14.95	8.64	11.92
	IMF4	8.44	0.84	3.68
	IMF5	8.73	2.36	6.35
	Res	1.30	0.55	0.92
Nov	IMF1	30.86	12.13	19.92
	IMF2	19.03	3.39	13.53
	IMF3	26.22	4.61	11.99
	IMF4	11.14	1.35	7.97
	IMF5	2.67	0.81	1.98
	Res	0.86	0.44	0.62

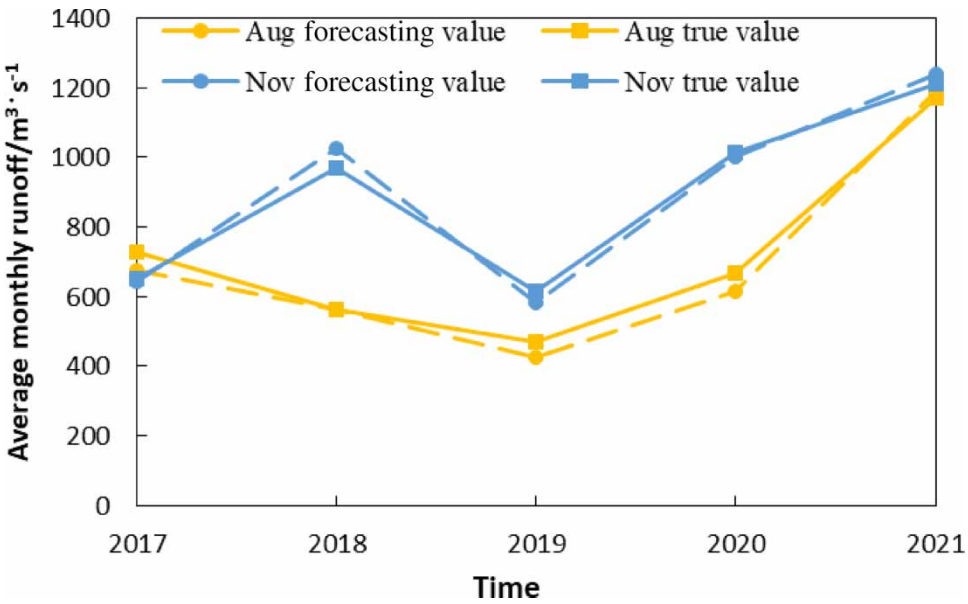


Figure 7 | Comparison of predicted results for August and November 2017–2021 at Takamura hydrological station.

Table 2 | The relative error index of 2017–2021 data

Month	Years	True value/$m^3\ s^{-1}$	Forecasting value/$m^3\ s^{-1}$	Relative error (%)
Aug	2017	726	675	6.95
	2018	564	561	0.56
	2019	471	426	9.48
	2020	669	617	7.84
	2021	1172	1188	1.42
Nov	2017	650	641	1.31
	2018	970	1025	5.60
	2019	614	581	5.44
	2020	1015	1003	1.19
	2021	1210	1237	2.21

Table 3 | Indicator evaluation table

Month	DC	MAE	MAPE	RMSE
Aug	0.92	26.82	5.25	87.07
Nov	0.97	22.18	3.15	70.73

deterministic coefficients of the forecast for August and November for five consecutive years from 2017 to 2021 are 0.92 and 0.97, respectively, which reach level A according to the Hydrological Intelligence Forecasting Code, and the three indicators of MAE, MAPE, and RMSE are also relatively small. It shows that the prediction results are more accurate and reasonable.

4. MODEL COMPARISON

To verify the accuracy and precision of the CEEMDAN-BILSTM coupled model prediction, two coupled models, CEEMD-LSTM and CEEMDAN-ElMAN, were selected for comparison in this study, and the results and error analysis of the three prediction models are shown in Figures 8 and 9 and Table 4 below.

As can be seen from Figure 8, under the comparison of the three different models, the CEEMDAN-BILSTM coupled model prediction fits the actual data the best, and the CEEMDAN-LSTM and CEEMD-LSTM models can also roughly predict the changing trend from 2017–2021, but the accuracy is poor. As can be seen from Figure 9, the prediction accuracy of the combined models is usually high, and the errors of the three combined models are maintained within 20%, with the relative error

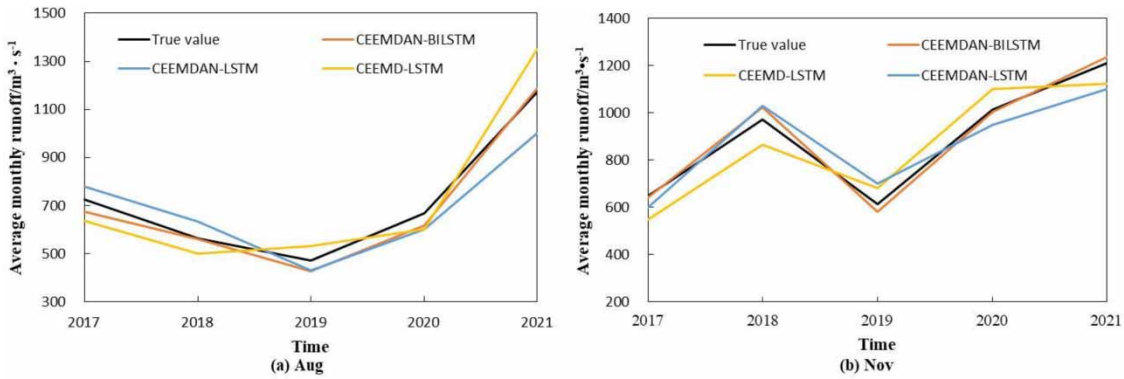


Figure 8 | Comparison of prediction effects of different models.

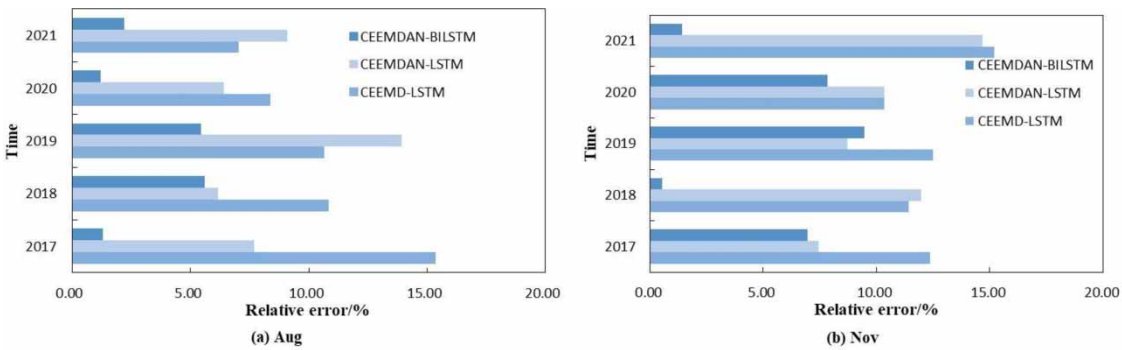


Figure 9 | Relative error plots of different models.

Table 4 | Comparison of indicators of different prediction models

Model	Month	DC	MAE	MAPE	RMSE
CEEMD-LSTM	Aug	0.80	64.41	7.37	228.49
	Nov	0.88	48.03	8.65	172.39
CEEMDAN-LSTM	Aug	0.90	66.72	8.26	194.51
	Nov	0.82	88.09	10.45	199.40
CEEMDAN-BILSTM	Aug	0.92	26.82	5.25	87.07
	Nov	0.97	22.18	3.15	70.73

of the CEEMDAN-BILSTM coupled model maintained within 10%. The error accuracy of the predictions of the other two models fluctuates widely, with a minimum error of 6.41% and a maximum error of 15.37%.

According to the four evaluation indexes selected, it can be seen from Table 4 that the deterministic coefficients DC of the three coupled models are all greater than 0.7, which reaches the hexagonal level in the regulation of the Hydrological Intelligence Forecasting Code, proving that the overall prediction accuracy of the coupled models is relatively accurate. As the model is optimized, the deterministic coefficient DC gets closer to 1, and MAE, MAPE, RMSE keep decreasing. All four metrics of the CEEMDAN-LSTM coupled model outperform the CEEMD-LSTM because CEEMDAN adds a limited number of times of adaptive white noise in each process, which reduces the reconstruction error of the signal with the same reduction in the average number of times. The prediction accuracy of the resulting model is improved. The CEEMDAN-BILSTM coupled model, based on LSTM, uses a bidirectional network so that the states of both the sequence start and end implicit layers can be trained for the recursive response, which can further explore some connection between the current data and the data of previous moments, as well as with the data of future moments, and further improve the prediction accuracy of the model. Through the comparison of three different combinations of prediction models, CEEMD-LSTM, CEEMDAN-LSTM, and CEEMDAN-BILSTM, it can be seen that the coupled CEEMDAN-BILSTM model has the best simulation effect in this flow prediction, and fits more closely with the actual data.

Although the overall prediction accuracy of the established CEEMDAN-BILSTM coupled model is high due to the limitation of the coupled CEEMDAN-BILSTM model, only a single sequence input and single sequence output are available. It is not possible to input multiple time series at the same time and output a single series. It is not possible to analyze the effect of rainfall, evaporation, or industrial and agricultural water use on the runoff volume. How to analyze runoff changes more comprehensively will be the next research focus of the model.

5. CONCLUSION

(1) CEEMDAN adds a limited number of times of adaptive white noise to each decomposition process, which also reduces the reconstruction error of the signal with the same reduction in the average number of times. BILSTM neural network solves the problem that LSTM can only advance in one direction and has the advantage of training the input sequence from the forward and reverse layers. The CEEMDAN-BILSTM coupled model was developed using August and November as samples to predict monthly flows for 2017–2021 using data from 1951 to 2016, with a minimum error of 0.56% and a maximum error of 9.48%, with an average error within 10%, and with determinants DC all greater than 0.9, indicating that the coupled model is feasible in predicting river flows.

(2) Using the data from the GaoCun hydrological station in the middle and lower reaches of the Yellow River as representative, we select August in the abundant period and November in the dry period as samples, and show that predicting i-months in the 5 years of 2017–2021 by i-months in the 66 years of 1951–2016 can reduce the large flow volatility caused by the abundant and dry periods due to external conditions such as rainfall. It makes the decomposition curve smoother, reduces the data volatility, and the prediction results closer to the actual flow.

(3) Through the comparative study of three different combinations of models, it can be seen that the flow prediction of CEEMDAN-BILSTM in the middle and lower reaches of the Yellow River is more accurate than the other two models. However, it is still a complicated process in the setting of parameters, which relies on the computer for optimal

tuning of parameters at a later stage, and the present model only considers the prediction of flow, while the performance in other aspects such as rainfall and climate needs to be further studied.

AUTHOR CONTRIBUTION

All authors contributed to the study conception and design. Writing and editing: Xianqi Zhang and Wenbao Qiao; chart editing: Jiafeng Huang; preliminary data collection: Jingwen Shi and Minghui Zhang. All authors read and approved the final manuscript.

DATA AVAILABILITY STATEMENT

All relevant data are included in the paper or its Supplementary Information.

REFERENCES

Burger, C. M., Kolditz, O., Fowler, H. J. & Blenkinsop, S. 2007 Future climate scenarios and rainfall-runoff modeling in the Upper Gallego catchment (Spain). *Environmental Pollution* **148**, 842–854.

Chen, S., Huo, X., Zhao, H. & Yao, Y. 2021 Axial unbalance identification of Gyrowheel rotor based on multi-position calibration and CEEMDAN-IIT denoising. *Measurement* **183**, 109852.

Farajzadeh, J. & Alizadeh, F. 2018 A hybrid linear–non-linear approach to predict the monthly rainfall over the Urmia Lake watershed using wavelet-SARIMAX-LSSVM conjugated model. *Journal of Hydroinformatics* **20**, 246–262.

Feng, J. & Pan, F. 2018 A LSTM BP combined hydrological forecasting method. *Computer and Modernization* **7**, 82–85 + 92 (in Chinese).

Jiang, S. C., Lu, J. Z., Chen, X. L. & Liu, Z. X. 2020 Runoff simulation of Fuhe River Basin in Poyang Lake based on LSTM network. *Journal of Central China Normal University (Natural Sciences)* **54** (1), 128–139 (in Chinese).

Jiang, L. L., Tan, H. C., Li, J. & Lei, J. L. 2021 Fault identification of spiral bevel gears based on CEEMDAN permutation with SVM. *Vibration, Testing and Diagnosis* **41** (1), 33–40.

Komasi, M. & Sharghi, S. 2016 Hybrid wavelet-support vector machine approach for modeling rainfall-runoff process. *Water Science & Technology* **73** (8), 1937–1953.

Li, F. X., Chen, F. L., Cai, W. J., He, C. L. & Long, A. H. 2021 Runoff multi-scale prediction based on EMD combination model. *Geoscience Frontiers* **28** (1), 428–437 (in Chinese).

Liu, H., Liu, Z., Jia, W. Q., Zhang, D. H. & Tan, J. R. 2021 Research status and challenges of residual service life prediction technology based on deep learning. *Computer Integrated Manufacturing System* **273** (1), 34–52 (in Chinese).

Ma, L. K., Qiu, Y. & Zhao, Y. 2016 Research on daily runoff prediction in small watersheds based on improved neural networks and support vector machines. *Journal of Water Resources and Water Engineering* **27** (5), 23–27.

Ma, D., Guo, Y. & Ma, S. 2021 Short-Term Subway Passenger Flow Prediction Based on GCN-BiLSTMIC. In *IOP Conference Series: Earth and Environmental Science*, Vol. 693(1). IOP Publishing, p. 012005.

Napolitano, G., Serinaldi, F. & See, L. 2011 Impact of EMD decomposition and random initialization of weights in ANN hindcasting of daily stream flow series: an empirical examination. *Journal of Hydrology* **406**, 199–214.

Nourani, V., Kisi, Ö. & Komasi, M. 2011 Two hybrid artificial intelligence approaches for modeling rainfall-runoff process. *Hydrology* **402**, 41–59.

Park, G., Moon, B. & Lee, H. 2014 The prediction of the stream runoff based on the hybrid model. *Journal of Korean Society Environmental Technology* **15** (1), 56–69.

Qian, J. L., Li, F. Q. & Wang, W. S. 2009 Time delay neural network applied to flood forecasting. *Journal of Hydropower* **28** (4), 18–21.

Schuster, M. & Paliwal, K. K. 2002 Bidirectional recurrent neural networks. *IEEE Transactions on Signal Processing* **45** (11), 2673–2681.

Tang, Y. C. 2008 *Modern Methods and Applications of Medium- and Long-Term Hydrological Forecasting*. China Water Conservancy and Hydropower Press, Beijing, pp. 159–186

Torres, M. E., Colominas, M. A. & Schlotthauer, G. 2011 A complete ensemble empirical mode decomposition with adaptive noise. In *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, Prague, Czech Republic, pp. 4144–4147.

Wang, C. 2010 *Research and Application of Improved Genetic Neural Network in Time Series Forecasting*. Sichuan Normal University, Chengdu.

Wang, D., Wei, J. H., Zhang, S. L. & Chu, H. B. 2020 Hydrological time series monthly runoff prediction based on CEEMD-BP model. *Journal of Beijing Normal University (Natural Science Edition)* **56** (3), 376–386.

Wu, Z. & Huang, N. E. 2009 Ensemble empirical mode decomposition: a noise-assisted data analysis. *Advances in Adaptive Data Analysis* **01** (01), 1–41.

Xu, B., Yang, F. G. & Li, Y. J. 2020 Application of two types of integrated learning algorithms in medium- and long-term runoff forecasting. *Hydropower* **46** (4), 21–24.34.

Yu, Y., Zhang, H. & Singh, V. P. 2018 Forward prediction of runoff data in data-scarce basins with an improved ensemble empirical mode decomposition (EEMD) model. *Water* **10**, 388.

Zhang, X. Q., Song, C. & Hu, D. K. 2018 Groundwater burial depth prediction model for irrigation areas based on EEMD and Elman network. *Water Conservation Irrigation* **12**, 86–91.

Zhang, S., Yan, Z. J., Xu, C. X. & Wang, H. R. 2020 Monthly runoff prediction model based on MPGA-LSTM and its application. *Hydropower Energy Science* **38** (5), 38–41 + 75 (in Chinese).

Zhang, X. Q., Tuo, W. & Song, C. 2020 Application of MEEMD-ARIMA combining model for annual runoff prediction in the Lower Yellow River. *Journal of Water and Climate Change* **11** (3), 865–876.

Zhao, K., Fu, H. Y., Li, W. & Chai, T. 2009 Research progress of watershed hydrological model. *Modern Agricultural Science and Technology* **23**, 267–270 (in Chinese).

First received 20 October 2022; accepted in revised form 6 December 2022. Available online 16 December 2022