Water Resources Management (2023) 37:1539–1555
https://doi.org/10.1007/s11269-023-03442-y

Runoff Forecast Model Based on an EEMD-ANN and Meteorological Factors Using a Multicore Parallel Algorithm

Shengli Liao[1] · Huan Wang[1] · Benxi Liu[1] · Xiangyu Ma[1] · Binbin Zhou[2] · Huaying Su[3]

Received: 11 November 2022 / Accepted: 16 January 2023 / Published online: 25 January 2023

Abstract

Accurate long-term runoff forecasting is crucial for managing and allocating water resources. Due to the complexity and variability of natural runoff, the most difficult problems currently faced by long-term runoff forecasting are the difficulty of model construction, poor prediction accuracy, and time intensive forecasting processes. Therefore, this study proposes a hybrid long-term runoff forecasting framework that uses the antecedent inflow and specific meteorological factors as the inputs, is modeled by ensemble empirical mode decomposition (EEMD) coupled with an artificial neural network (ANN), and computed by a parallel algorithm. First, the framework can transform monthly inflow and meteorological series into stationary signals via EEMD to more comprehensively explore the relationships of the input factors through the ANN. Second, the selected meteorological factors that are closely related to inflow formation can be filtered out by the single correlation coefficient method, which contributes to reducing coupling between input factors, and increases the accuracy of the prediction models. Finally, a multicore parallel algorithm that is easily accessed everywhere and that fully utilizes multiple calculation resources while flexibly contending with various optimization requirements will improve forecasting efficiency. The Xiaowan Hydropower Station (XW) is selected as the study area, and the final results of the study show that (1) the addition of targeted meteorological factors does indeed greatly enhance the performance of the prediction models; (2) the five criteria for evaluating the prediction accuracy show that the EEMD-ANN model is far superior to the prediction performance from the ordinary ANN model when run under the same input conditions; and (3) the optimization time of the 32-core model can be reduced by as much as 25 times, which significantly saves time during the forecast process.

Keywords Long-term runoff forecast · Parallel algorithm · Ensemble empirical mode decomposition · ANN · Filtered meteorological factors

✉ Huan Wang
whkkll@163.com

Extended author information available on the last page of the article

1 Introduction

Due to the influence of climate change, typhoons and human activities (Chau et al. 2005), inflows have increased sharply, and the accuracy of long-term runoff forecasts has decreased, which has directly led to difficulty in hydropower station dispatch planning. Therefore, the surplus water from hydropower stations has increased greatly, resulting in a large amount of clean energy waste. Sichuan Province and Yunnan Province are regions with the most abundant hydropower resources in China and are also important power supply bases for China's "West–East Electricity Transmission Project", but they also face serious surplus water problems. Figure 1a shows that the abandoned energy of Sichuan and Yunnan Provinces is equivalent to the hydropower generation of Germany, and the excess energy accounted for a large proportion of hydropower generation from 2018 to 2020. This increasingly severe wastewater issue should be solved in time. Therefore, an effective long-term runoff prediction system is needed for hydropower stations to accurately predict the large amount of runoff caused by extreme weather events to help hydropower stations plan their operations in advance and reduce the loss of water and hydropower resources (Ma et al. 2015).

Due to the exponentially increasing complexity and randomness of runoff, the category and quantity of hydrological data required for runoff forecasting also increase accordingly (Yue et al. 2020). In the face of such a large amount of information, traditional runoff forecast methods, mathematical statistical methods, causal analysis methods, etc., may contain some shortcomings, such as the insufficient utilization of information, reduced learning ability, and large prediction error, which make it difficult to meet the actual hydrological data analysis needs. In contrast, machine learning algorithms with strong computing and analysis power are highly suitable for long-term runoff forecasting and can provide an effective tool for unknown prediction processes; for example, ANN (Shoaib et al. 2018), least squares support vector machines (LSSVMs) (Fahimi and El-Shafie 2014), and decision trees (Khader et al. 2013) are commonly used in long-term runoff forecasting. However, due to flaws in inherent mechanisms, these prediction models have difficulty dealing with nonstationary inflow series, and direct prediction often leads to increased error and low prediction accuracy. Thus, ensemble empirical mode decomposition (EEMD) can be used to stabilize runoff by decomposing the original runoff into multiple stationary components, and then an ANN model with a strong excavation capacity and high fault-tolerance ability can be applied to linearize runoff by predicting and reconstructing each component (Lin et al. 2006). Compared with other prediction models, EEMD-ANN can comprehensively disassemble and analyze the characteristics of the input runoff and has better prediction performance. In addition, it can also be combined with parallel algorithms to improve the efficiency of model construction and reduce the time consumed in the model optimization process. In recent years, EEMD-ANN has gradually become a popular research focus in the field of long-term runoff prediction because of their simple structure and good performance (Lu and Zhou 2014; Li et al. 2019).

At present, a perfect hydrological forecast system should not only ensure the accuracy of the prediction results but also ensure the efficiency of the prediction process. To guarantee the accuracy of hydrological forecasting, it must continuously model and adjust parameters according to the existing hydrological data until the optimal prediction model is fitted (Zhao et al. 2020). Although this can ensure the high accuracy of the system, the process of iterating the optimal solution will consume considerable time, which conflicts with the high efficiency of the forecast system. Thus, the multicore parallel algorithm can be used to accelerate the iterative process of the model and solve this contradiction perfectly (Liao et al. 2016; Niu et al. 2021), improve the accuracy of the forecast system, and ensure the high efficiency of the forecast system.

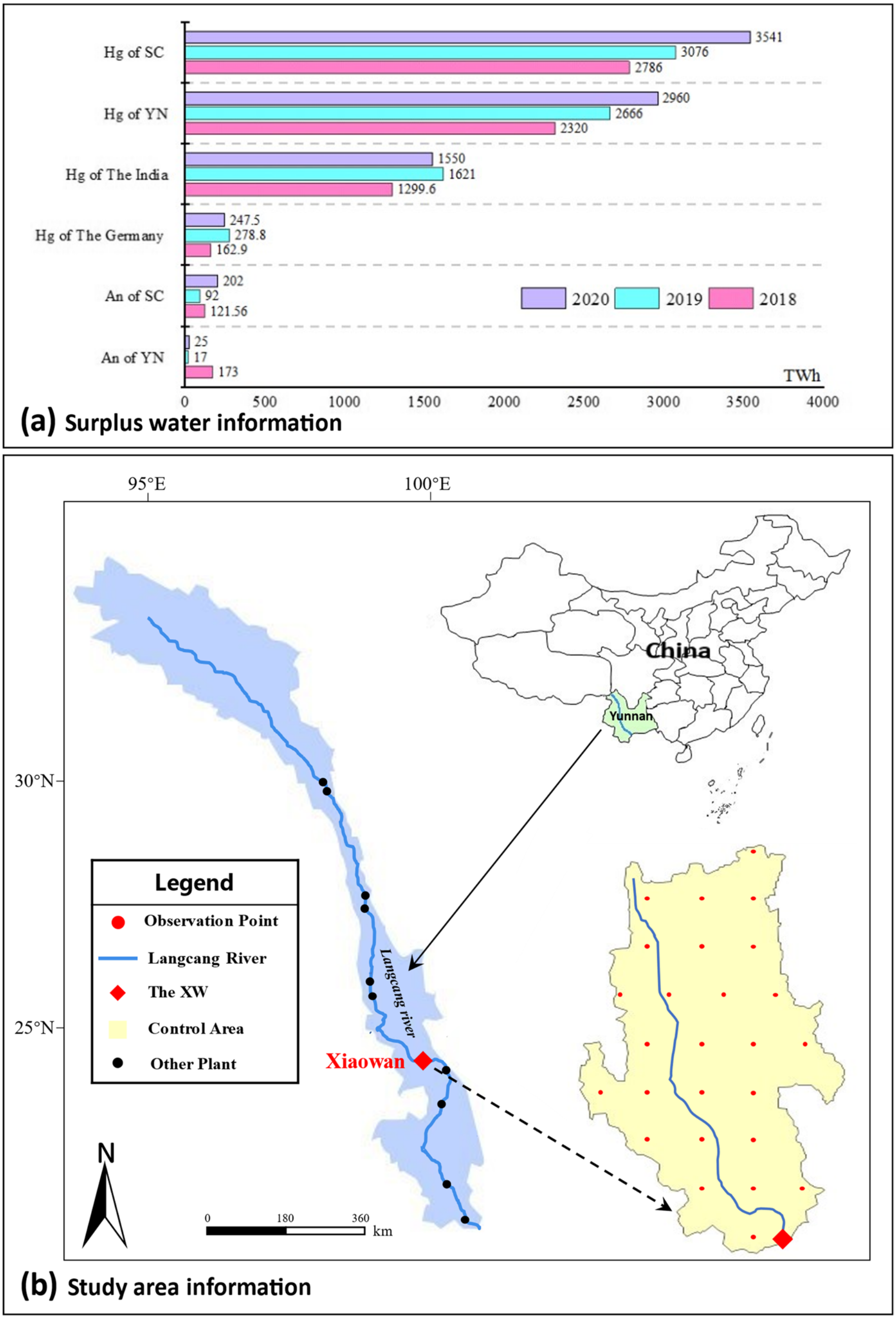

Fig. 1 **a** Surplus water information statistics (Hg: hydropower generation, An: abandoning energy, SC: Sichuan Province, YN: Yunnan Province); **b** Study area information: XW

Although a combination of the EEMD-ANN and parallel algorithm has made the forecast results more accurate, there is still a certain amount of room for improvement. At present, a relatively novel method is based on antecedent inflow and is coupled with specifically selected meteorological input factors to predict runoff (Wang et al. 2015). Some historical studies only used the antecedent water level as the input factor of the ANN that was trained by the particle swarm optimization (PSO) algorithm to predict the water level of the Manwan Hydropower Station (Chau 2006) or used the approximated meteorological factors as the input factors of the Long Short-Term Memory network (LSTM) model to predict rainfall runoff (Martin et al. 2020). While those studies have made several achievements, the impact of selected meteorological factors on the runoff formation process was not considered at all. Therefore, it has also been difficult for these methods to obtain accurate prediction results. It has been proven that the addition of the selected meteorological factors in the prediction system can improve the learning ability and prediction accuracy of the model (Li et al. 2017; Ravindran et al. 2021).

Subsequently, the rest of this paper is arranged as follows: Section 2 briefly introduces the basic theories, namely the ANN algorithm, parallel algorithm, EEMD algorithm and data preprocessing; Section 3 introduces the evaluation indicators and research process; and Section 4 provides the research results.

2 Study Area and Data

2.1 The Study Area and Data Utilized

The Xiaowan Hydropower Station (XW), which is a critical reservoir located in the middle and lower reaches of the Lancang River in Dali city, Western Yunnan Province, China, (longitude 100°04′48″E, latitude 34°42′36″N) is marked by the red square in Fig. 1b. Additionally, XW has the highest hyperbolic arch dam in the world and is the third largest active hydropower station in China. Therefore, it is very meaningful case study to adopt for this work. The Lancang River, which is also known as the Mekong River, is approximately 2000 km long, has a drainage area of 113300 km^2, originates from the Qinghai Tibet Plateau and is the largest international river in Southeast Asia, and the major source of its water comes from melting snow on the Tibetan Plateau (Liao et al. 2020).

The antecedent inflow of the XW from January 1979 to December 2019 and 25 meteorological data points were used as candidate input data. Table 1 shows the 25 candidate meteorological categories used for the analysis of the XW. In the control basin of the XW, there are 26 meteorological data monitoring points (small red dots in Fig. 1b), and the inverse distance weighted (IDW) method can be used to reduce the weight of the meteorological data at each monitoring point which was downloaded from the ERA5 website. The meteorological data used for the XW study area covers the time range from January 1979 to December 2019 (Sun et al. 2018).

2.2 Feature Selection and Feature Scaling

Because the input variables, including antecedent inflow and meteorological factors, are disordered and irregular, reasonable selection of input variables can reduce the dimensionality of the data to accelerate the learning efficiency of prediction models. The inverse distance-weighted method and the single-phase relation number method are used to process and filter input data from the reanalysis meteorological dataset. The correlation between lagged runoff and total precipitation (tp) can be demonstrated by the partial autocorrelation

Table 1 Description of candidate meteorological

No	Variable	Description
1	*v10*	10 m v-component of wind
2	*u10*	10 m u-component of wind
3	*d2m*	2 m dewpoint temperature
4	*t2m*	2 m temperature
8	*tp*	Total precipitation
6	*e*	Evaporation
7	*tcc*	Total cloud cover
8	*cdir*	Clear-sky direct solar radiation at surface
9	*sp*	Surface pressure
10	*ro*	Runoff
11	*smlt*	Snowmelt
12	*stl1*	Soil temperature level 1
13	*stl2*	Soil temperature level 2
14	*stl3*	Soil temperature level 3
15	*stl4*	Soil temperature level 4
16	*swvl1*	Volumetric soil water layer 1
17	*swvl2*	Volumetric soil water layer 2
18	*swvl3*	Volumetric soil water layer 3
19	*swvl4*	Volumetric soil water layer 4
20	*istl1*	Ice temperature layer 1
21	*istl2*	Ice temperature layer 2
22	*istl3*	Ice temperature layer 3
23	*istl4*	Ice temperature layer 4
24	*tcwv*	Total column water vapour
25	*tcw*	Total column water

function (PACF) and cross-correlation function (CCF), and the optimal lag order can be determined by 95% confidence intervals. In addition, when the correlation coefficient continues to change slowly and does not fall into the 95% confidence interval, it is necessary to determine the best lag order by the trial-and-error method and index of agreement (IA), which is an evaluation index that successively adds one or more lag sequences to the input from the first lag order (Lu and Zhou 2014; Li et al. 2019).

$$Y = \frac{\sum_{q=1}^{m}(Z_q/d_q^p)}{\sum_{q=1}^{m}(1/d_q^p)} \tag{1}$$

where Y is the meteorological data of the research point; Z_q is the meteorological value of each measure point q; d_q is the distance from each measure point q to the research point; m is the number of measure points; and P is the weight index, which is taken as 2 in this study (Liao et al. 2020).

$$\gamma = \frac{\sum_{j=1}^{J}(x_j - \overline{x})(y - \overline{y})}{\sqrt{\sum_{j=1}^{J}(x_j - \overline{x})^2 \sum_{j=1}^{J}(y_j - \overline{y})^2}} \tag{2}$$

$$\tau = \frac{\gamma\sqrt{J-2}}{\sqrt{1-\gamma^2}} \tag{3}$$

where x_j and y_j are the values of runoff and meteorological factors in sequence *j*, respectively; $\overline{x}$ and $\overline{y}$ are the multiyear averages of runoff and meteorological factors, respectively; γ and τ are the correlation coefficients between runoff and meteorological factors; and *J* is the number of samples (Zhang et al. 2021).

$$IA = 1 - \frac{\sum_{i=1}^{n}(\widehat{Q}_i - Q_i)^2}{\sum_{i=1}^{n}(|\widehat{Q}_i - \overline{Q}| + |Q_i - \overline{Q}|)^2} \tag{4}$$

where $\widehat{Q}_i$ and Q_i *re*present the inflow prediction and observed value at time *i*, respectively; $\overline{Q}$ represents the mean of the observed series; and *n* is the length of the runoff series (Liao et al. 2020).

To unify the input data and accelerate the learning efficiency of the model, the input signals can be normalized to convert them into interval [0,1]. The deviation standardization method (He et al. 2022) can be adopted to normalize the input data, as expressed below:

$$x_{scale} = \frac{x_b - x_{\min}}{x_{\max} - x_{\min}} \tag{5}$$

where x_{scale} is the scaled value; x_b is the original input value in sequence *b*; and x_{min} and x_{max} are the minimum and maximum values in the original data, respectively.

3 Materials and Methods

3.1 ANN Neural Network Model

The ANN prediction model adjusts the weight of the model through the propagation error E_{total} and learning rate η to gradually reduce the error and finally achieve the best prediction model.

The error propagation E_{total} of the ANN model is (Chau 2006; Wang et al. 2010):

$$E_{total} = \frac{1}{2}\sum_{u=1}^{p}(D_u - O_u)^2 \tag{6}$$

$$O_u = g\left(\sum_{l=0}^{L} y_l w_{lu}\right), y_l = f\left(\sum_{k=0}^{K} x_k w_{kl}\right) \tag{7}$$

The weights are adjusted via E_{total} and η:

$$\Delta w_{lu} = -\eta \frac{\partial E}{\partial w_{lu}} \quad 1 = 0, 1, \cdots, L; \; u = 1, 2, \cdots, p \tag{8}$$

$$\Delta w_{kl} = -\eta \frac{\partial E}{\partial w_{kl}} \quad \mathrm{k} = 0, 1, \cdots, K; \; 1 = 1, 2, \cdots, L \tag{9}$$

where E_{total} is the error in the propagation process; D_u is the actual value; O_u is the predicted value; x_k and y_l are the input value of the model and the output value of the hidden layer, respectively; *f(x) g(x)* are activation functions of the hidden layer and output layer, respectively; and w_{kl} and w_{lu} are the weights in each propagation and are updated with E_{total} and η.

3.2 Parallel Algorithm

The parallel algorithm is a method for solving problems with multiple processors. With advancements in computer science and technology, modern computers have been equipped with multicore systems (Niu et al. 2021). Therefore, large-scale data processing problems can be efficiently solved by a parallel algorithm that can now be easily accessed in modern computers (Ma et al. 2021). The steps for using the parallel algorithm are shown in Fig. 2a. The main steps of the parallel algorithm are as follows:

Step 1: Reading and analyzing the main task in parallel.
Step 2: Decomposing the main task into n subtasks according to the actual requirements and hardware environment.
Step 3: The processor is used to process and solve each subtask, and the output results of each subtask are obtained.
Step 4: Finally, the output results of each subtask are collated and combined to obtain the results of the main task.

3.3 EEMD Algorithm

EEMD is an improvement based on the algorithm empirical mode decomposition (EMD) algorithm, which can self-adaptively transform the nonstationary and nonlinear initial signal sequence *f(t)* into multiple stable components (Tan et al. 2018), including several intrinsic mode functions (IMFs) and one residue (Re). With the auxiliary addition of white noise, the decomposition results are continuously averaged (Wang et al. 2015), which effectively avoids the phenomenon of mode mixing in EMD and enhances the clarity and practicality of the signal (Wang et al. 2013). Finally, the corresponding decomposed signals (IMFs and Re) are integrated to obtain a stable initial signal.

3.4 Accuracy Evaluation Criteria of the Models

In this study, the correlation coefficient (CORR) (Zhang et al. 2021), the root mean square error (RMSE) (Zhang et al. 2021), the mean absolute error (MAE) (Callegari

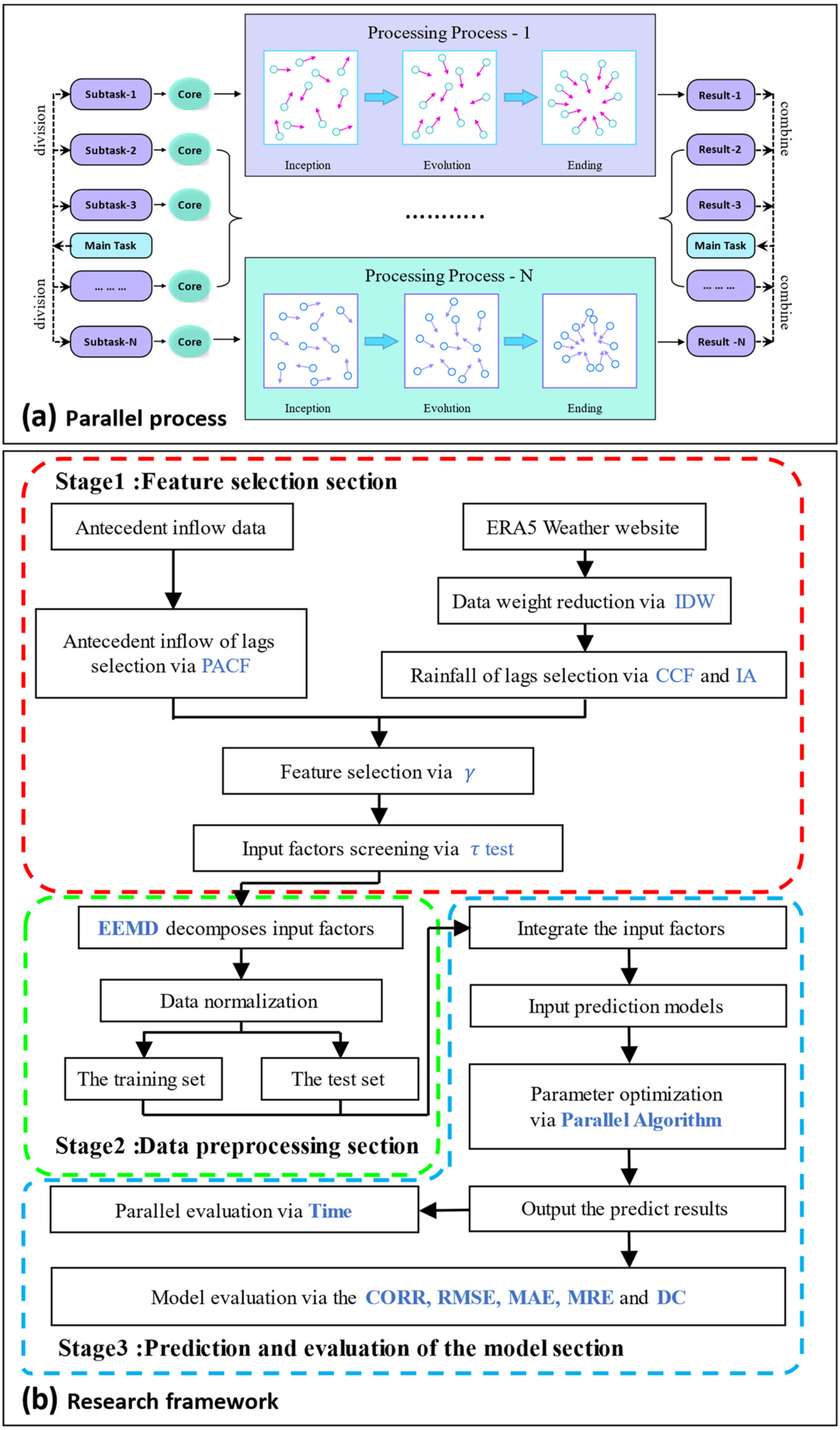

Fig. 2 **a** Parallel algorithm process; **b** Research framework for long-term runoff forecast

et al. 2015), the mean relative error (MRE) (Li et al. 2012), and the certainty coefficient (DC) (Lu and Zhou 2014) (Eqs. (10)-(14)) are used to evaluate the accuracy of the model results.

$$CORR = \frac{\sum_{i=1}^{n}(Q_i - \overline{Q})(\hat{Q}_i - \overline{\hat{Q}}_i)}{\sqrt{\sum_{i=1}^{n}(Q_i - \overline{Q})^2}\sqrt{\sum_{i=1}^{n}(\hat{Q}_i - \overline{\hat{Q}}_i)^2}} \tag{10}$$

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(\hat{Q}_i - Q_i)^2} \tag{11}$$

$$MAE = \frac{1}{n}\sum_{i=1}^{n}|\hat{Q}_i - Q_i| \tag{12}$$

$$MRE = \frac{1}{n}\sum_{i=1}^{n}\frac{|\hat{Q}_i - Q_i|}{Q_i} \tag{13}$$

$$DC = 1 - \frac{\sum_{i=1}^{n}(\hat{Q}_i - Q_i)^2}{\sum_{i=1}^{n}(Q_i - \overline{Q})} \tag{14}$$

where $\overline{\hat{Q}}_i$ represents the mean of the prediction series.

3.5 Overall Framework

To verify the effectiveness of the framework presented in this paper, the effects of different models and different input factors on the prediction accuracy are compared. The runoff and meteorological data are used as the input factors of the EEMD-ANN and ANN models, respectively, such that research models are constructed as follows:

1. Runoff is input into the EEMD-ANN prediction model (called the EEMD-ANN-Q model).
2. Runoff and meteorological factors are input into the EEMD-ANN prediction model (called the EEMD-ANN-Q-E model).
3. Runoff is input into the ANN prediction model (called the ANN-Q model).
4. Runoff and meteorological factors are input into the ANN prediction model (called the ANN-Q-E model).

Then, according to the length of the dataset, the input factors are divided into a training set and testing set proportional to the length and the parallel algorithm is used to increase the speed of the optimal parameter calibration process within the ANN model by applying different input factors. Finally, based on five evaluation criteria, the prediction models are compared and evaluated to obtain the optimal prediction model. The general framework of this process is shown in Fig. 2b.

4 Research Results and Discussion

4.1 Filtering of Input Factors

The antecedent inflow inputs: The lag order of the optimal antecedent inflow is determined by the PACF (Fig. 3a) of the XW and the corresponding 95% confidence interval. The figure shows that before the 12th order, the PACFs are basically outside the 95% confidence interval but after the 12th order, they are basically reduced to the 95% confidence interval, and the PACF is truncated at $p=12$. Therefore, the antecedent inflow with 12 lag orders $Q_{t\text{-}1}$:$Q_{t\text{-}12}$ (the lagging runoff Q from 1st to 12th order) is selected as the inflow input of the prediction models proposed by this study.

The total precipitation (tp) factor inputs: tp with different lag orders is filtered separately by the CCF and IA evaluation indices. As shown in Fig. 3a, the correlation between runoff and tp first slightly decreases and then slightly increases from lag 1 to 12 but never falls within the 95% confidence interval. Obviously, the optimal lag order of tp cannot be obtained only by the CCF. To obtain the accurate order, the trial-and-error method and IA index are adopted to further determine the optimal lag order: from 1 to the 12th order of tp lag are used as the input factors for testing the model, and then the correlation of each lagging input factor is evaluated by the IA index, and the most relevant input is selected (Liao et al. 2020). The experimental results of the IA index

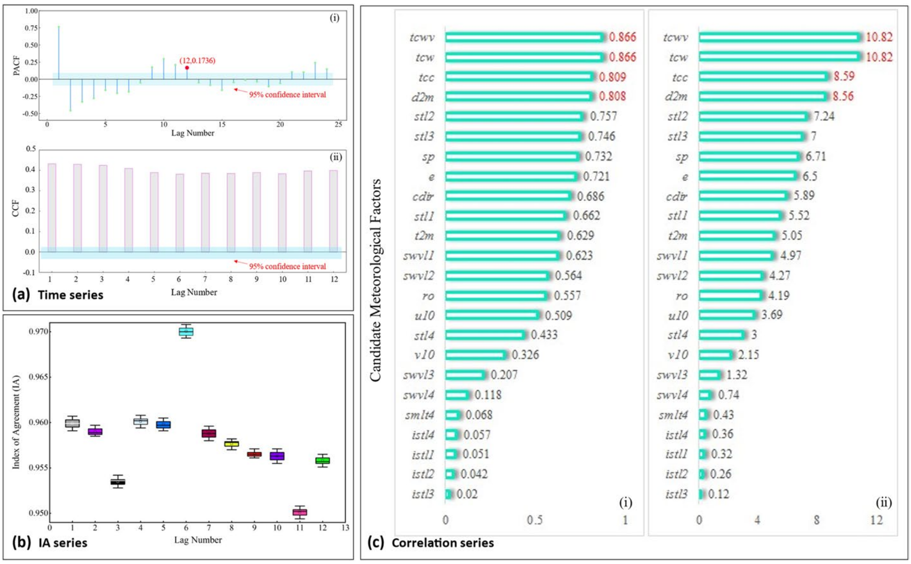

Fig. 3 **a** time series of XW (i) the PACF chart of the XW antecedent inflow, (ii) the CCF chart of the XW; **b** the IA values of the XW's tp lags; **c** single correlation coefficient of meteorological factor screening chart: (i) γ, (ii) τ

are shown in Fig. 3b. The IA value of input Factor 6 is the largest among the 12 input factors; therefore, the correlation between tp and runoff in input Factor 6 is the largest. Ultimately, it is determined that tp with 6 lag orders (R_{t-1}, R_{t-2}, R_{t-3}, R_{t-4}, R_{t-5}, and R_{t-6}) are selected as the total precipitation (tp) input for the prediction models proposed in this study.

The other meteorological factor inputs: Owing to the lag order of the other 24 candidate meteorological factors, excluding total precipitation (tp), they appear to have little lag effect on the runoff formation process, the correlation coefficient γ and the τ-test (confidence degree α as 0.05) can be directly used to identify the truly relevant meteorological factors to use as model input (Xie et al. 2016; Wang et al. 2022). As seen from the results in Fig. 3c, many of the candidate meteorological factors have a good single correlation with the antecedent inflow of the XW, and the independent coupling between meteorological factors is low. Finally, meteorological factors with $\tau > 8$ are selected in this study.

In summary, the five filtered meteorological factors selected for the prediction models proposed in this study are total precipitation (tp), total column water vapor (tcwv), total column water (tcw), total cloud cover (tcc), and 2 m dewpoint temperature (d2m), which is consistent with the input factors of the long-term runoff prediction framework proposed by this study.

4.2 Decomposition and Integration of Input Factors

Decomposition of input factors: The input factors in the study are composed of six signals; two of which are the antecedent inflow with 12-order lags, and tp with 6-order lags while the tcwv, tcw, tcc and d2m with 1-order lags comprise the remaining four. Then, the input factor is decomposed into a finite number of subsequences by the EEMD algorithm. The decomposition results are shown in Fig. 4a. As the data types and properties contained in each initial signal are different, the decomposition results are also varied. These consist of the antecedent inflow, tp, tcc and tcwv (which contain components of 7 IMFs), Re, and d2m and tcw (which contain the components of 6 IMFs and Re (Tan et al. 2018)).

Integration of input factors: The principle of the EEMD-ANN prediction model is to use EEMD to decompose the initial signal into several components, then use an ANN to predict each component, and finally add the predicted values of each component to obtain the final prediction results, which can improve the prediction accuracy by eliminating the "impurities" in the original signal. Due to the addition of the meteorological factors considered in the study, the diverse types of input factors lead to the diversity of the decomposition quantities. Thus, it is necessary to integrate the diverse decomposition components into the input factors of the model according to a certain law. Figure 4a shows that the IMF6 components of all input signals contain more than one signal period, while the IMF7 components of antecedent inflow, tp, tcc and tcwv do not have a complete period. Consequently, diverse components can be integrated according to this characteristic as the input of the ANN model, which is consistent with the method of modeling the long-term runoff prediction framework proposed in this work. The specific integration results are shown in Table 2.

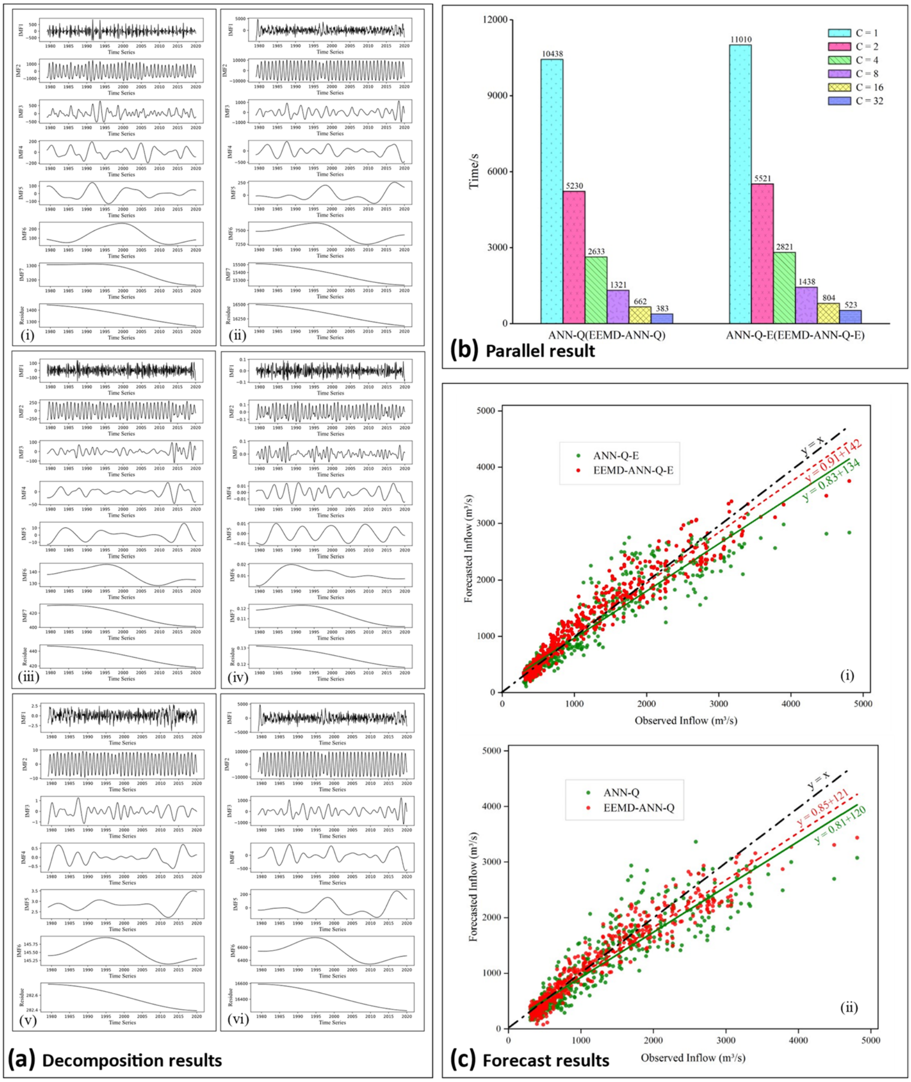

Fig. 4 **a** Decomposition result of input factors (i)runoff; (ii)tcwv; (iii)tcc; (iv)tp; (v)d2m; (vi)tcw; **b** Model parallel time comparison; **c** Forecast inflow of ANN and EEMD-ANN: (i) no adding refined meteorological factors; (ii) adding refined meteorological factors

4.3 Parallel Result Analysis

In this research, the EEMD decomposes the antecedent inflow signal into 8 components, which only affect the input factors of the ANN model and do not participate in the bottom prediction process of the ANN (Wang et al. 2015). Thus, when the number of nodes in the input layer (N_1), hidden layer (N_2) and output layer (N_3) is known, the parallel algorithm is used for other parameters, such as the learning rate (η) and training times (*epochs*), to optimize the model and determine the results. Additionally, the experimental results of the

Table 2 Integration results of input factors

Input signal	Components							
	IMF1	IMF2	IMF3	IMF4	IMF5	IMF6	IMF7	RE
antecedent inflow	√	√	√	√	√	√	√	√
tp	√	√	√	√	√	√	√	√
d2m	√	√	√	√	√	√	√	√
tcw	√	√	√	√	√	√	×	√
tcwv	√	√	√	√	√	√	×	√
tcc	√	√	√	√	√	√	√	√
Remark	Input-1	Input-2	Input-3	Input-4	Input-5	Input-6	Input-7	Input-8

fuzzy neural network (FNN) model were compared with the ANN as a reference control group.

Under the support of the parallel algorithm, the optimal model parameters of the prediction models are quickly obtained, and the results are shown in Table 3. Meanwhile, Fig. 4b intuitively shows the optimization time of the ANN models when parallel algorithms are run on processors with 2, 4, 8, 16 and 32 cores, optimization time of the ANN-Q (EEMD-ANN-Q) model ranges from 10438 s in single core to 383 s within 32 cores. Similarly, in terms of optimization time, the ANN-Q-E (EEMD-ANN-Q-E) model shows the same optimization process as ANN-Q (EEMD-ANN-Q). Moreover, due to the addition of the meteorological factors to the input data, the optimization time of the ANN-Q-E (EEMD-ANN-Q-E) model is always larger than that of the ANN-Q (EEMD-ANN-Q) model for the same number of cores. Thus, the acceleration effect of the parallel algorithm is perfectly optimized, which is consistent with the method of calculating the long-term runoff prediction framework proposed in this paper.

4.4 Comparison of Prediction Results

The prediction indicators of the models are shown in Table 4. According to the five accuracy evaluation indices, the ranking of the prediction performance of the models is as follows: EEMD-ANN-Q-E > EEMD-ANN-Q > ANN-Q-E > FNN-Q > ANN-Q. Because of the combination of fuzzy theory and neural networks, the prediction performance of the model FNN is better than that of the ordinary ANN model.

The correlation of the prediction results of the models is clearly shown in Fig. 4c (the red dot denotes the EEMD-ANN prediction inflow and the green dot denotes the ANN prediction inflow). We can see clearly from Fig. 4c that the red dots are densely clustered

Table 3 Optimal parameters of ANN models

Model	N_1	N_2	N_3	η	*epochs*
ANN-Q	12	10	1	0.02	1000
EEMD-ANN-Q	22, 22, 22, 22, 22, 22, 22	10	1	0.02	1000
ANN-Q-E	22	8	1	0.05	1000
EEMD-ANN-Q-E	22, 22, 22, 22, 22, 20, 22	8	1	0.05	1000

Table 4 Prediction performance of models

Model	Training					Testing				
	CORR	RMSE	MAE	MRE	DC	CORR	RMSE	MAE	MRE	DC
FNN-Q	0.90	384.0	247.42	19.90%	0.82	0.92	305.76	202.8	19.51%	0.82
ANN-Q	0.91	377.8	254.2	16.50%	0.83	0.92	359.3	248.4	21.30%	0.76
ANN-Q-E	0.93	352.7	225.4	16.80%	0.85	0.93	293.6	200.3	19.80%	0.84
EEMD-ANN-Q	0.95	272.7	189.9	15.70%	0.91	0.96	214.4	147.2	14.60%	0.91
EEMD-ANN-Q-E	0.96	252.74	182.3	16.50%	0.92	0.97	189.2	135.1	13.50%	0.93

around $y=x$, while the green dots are relatively scattered on both sides of $y=x$, which means that the prediction results of the EEMD-ANN model are better fitted with the antecedent inflow than the ordinary ANN when the input factors are the same, thus the prediction accuracy is higher. In addition, from the perspective of input factors, adding the selected meteorological factor in Fig. 4c(ii), the fitting slope of the predicted inflow of the model is greater than that without adding the selected meteorological factor (as shown in Fig. 4c(i)) under the same situation (EEMD-ANN-Q-E for 0.91; EEMD-ANN-Q for 0.85; ANN-Q-E for 0.83; ANN-Q for 0.81), which indicates that the prediction performance of the prediction models has been greatly improved by adding the selected meteorological factor.

In summary, the optimal prediction model for the XW inflow as determined by this study is EEMD-ANN-Q-E, which corresponds to the input factors for $Q_{t\text{-}1}$: $Q_{t\text{-}12}$, $R_{t\text{-}1}$, $R_{t\text{-}2}$, $R_{t\text{-}3}$, $R_{t\text{-}4}$, $R_{t\text{-}5}$, $R_{t\text{-}6}$, tcwv, tcw, tcc, and d2m.

5 Conclusion

In this study, the ANN was used to predict the future inflow of XW according to the antecedent inflow and meteorological factors. The parallel algorithm was adopted to improve the optimization efficiency of the ANN model. By adding the EEMD algorithm and selected meteorological factors for comparison, the following conclusions were drawn:

1. With the addition of the specifically chosen meteorological factors, the prediction performance of both the EEMD-ANN model and ANN model steadily improved. Thus, in long-term runoff forecasting, meteorological factors can effectively improve prediction accuracy and provide a guarantee for future hydrological system operation plans.
2. With the same input data, the EEMD-ANN prediction model has better performance than the ANN. Therefore, EEMD-ANN is more suitable for modeling long-term runoff forecasts than ANN.
3. The parallel algorithm can accelerate the model parameter calibration efficiency and save the training time of the model. In the parallel algorithm, the greater the number of cores, the higher the optimization efficiency, and the shorter the time.

For long-term runoff forecasting, the construction of a prediction model, prediction accuracy and model parameter calibration efficiency are very important. The framework of the long-term runoff forecast proposed in this study effectively solves these problems and has certain reliability and practicability for the operation of hydrological systems. However, the existing conclusions and methods presented in this paper cannot be applied to short-term prediction because the impact of weight on the short-term meteorological factors and the decomposition of daily runoff by the EEMD algorithm cannot be directly realized, and further detailed research is needed from the perspective of data processing and modeling, which is also our future research direction.

Authors' Contributions S. L. performed study design, data analysis and interpretation, and drafted the manuscript. H. W. and B.L. participated in the study design, data collection, algorithm and manuscript preparation. X M. participated in the design and coordination of experimental work. B.Z. and H.S. performed data collection and interpretation. All authors read and approved the final manuscript.

Funding This work was supported by the National Natural Science Foundation of China [grant number 51979023].

Data Availability Data, models, or code that support the findings of this study are available from the corresponding author upon reasonable request.

Declarations

Ethics Approval Not applicable.

Consent to Participate Not applicable.

Consent to Publish Not applicable.

Competing Interests The authors declare that they have no conflict of interest.

References

Callegari M, Mazzoli P, Pistocchi A (2015) Seasonal river discharge forecasting using support vector regression: a case study in the italian alps. Water 7(5):2494–2515. https://doi.org/10.3390/w7052494

Chau KW (2006) Particle swarm optimization training algorithm for ANNs in stage prediction of Shing Mun River. J Hydrol 329(3–4):363–367. https://doi.org/10.1016/j.jhydrol.2006.02.025

Chau KW, Wu CL, Li YS (2005) Comparison of several flood forecasting models in yangtze river. J Hydroelectr Eng 10(6):485–491. https://doi.org/10.1061/(ASCE)1084-0699(2005)10:6(485)

Fahimi F, El-Shafie AH (2014) Comment on “a hybrid model of self organizing maps and least square support vector machine for river flow forecasting” by Ismail et al. Hydrol Earth Syst Sci 18(7):2711–2714. https://doi.org/10.5194/hess-18-2711-2014

He S, Sang XF, Yin JX, Zheng Y, Chen HT (2022) Short-term runoff prediction optimization method based on bgru-bp and blstm-bp neural networks. Water Resour Manag. https://doi.org/10.1007/s11269-022-03401-z

Khader AI, Rosenberg DE, McKee M (2013) A decision tree model to estimate the value of information provided by a groundwater quality monitoring network. Hydrol Earth Syst Sci 17(5):1797–1807. https://doi.org/10.5194/hess-17-1797-2013

Li FF, Wang ZY, Zhao X, Xie E, Qiu J (2019) Decomposition-ANN methods for long-term discharge prediction based on fisher’s ordered clustering with MESA. Water Resour Manag 33(9):3095–3110. https://doi.org/10.1007/s11269-019-02295-8

Li KF, Ji CM, Zhang YK, Xie W, Zhang XX (2012) Study of mid and long-term runoff forecast based on back-propagation neural network. 188–191. https://doi.org/10.1109/ICICEE.2012.57

Li WT, Duan QY, Miao CY, Ye AZ, Gong W, Di ZH (2017) A review on statistical postprocessing methods for hydrometeorological ensemble forecasting. Wiley Interdiscip Rev Water 4(6). https://doi.org/10.1002/wat2.1246

Liao SL, Li G, Sun QY, Li ZF (2016) Real-time correction of antecedent precipitation for the Xinanjiang model using the genetic algorithm. J Hydroinform 18(5):803–815. https://doi.org/10.2166/hydro.2016.168

Liao SL, Liu ZW, Liu BX, Cheng CT, Jin XF, Zhao ZP (2020) Multistep-ahead daily inflow forecasting using the ERA-Interim reanalysis data set based on gradient-boosting regression trees. Hydrol Earth Syst Sci 24(5):2343–2363. https://doi.org/10.5194/hess-24-2343-2020

Lin JY, Cheng CT, Chau KW (2006) Using support vector machines for long-term discharge prediction. Hydrology Sci J 51(4):599–612. https://doi.org/10.1623/hysj.51.4.599

Lu D, Zhou HC (2014) Medium and long-term runoff forecasting based on mutual information and bp neural network. J China Hydrol 34(4):8–14

Ma F, Yuan X, Ye AZ (2015) Seasonal drought predictability and forecast skill over China. J Geophys Res Atmos 120(16):8264–8275. https://doi.org/10.1002/2015JD023185

Ma YF, Zhong PA, Xu B, Zhu FL, Li JY, Wang H, Lu QW (2021) Cloud-based multidimensional parallel dynamic programming algorithm for a cascade hydropower system. Water Resour Manag 35(9):2705–2721. https://doi.org/10.1007/s11269-021-02859-7

Martin G, Frederik K, Daniel K, Grey N, Jimmy L, Sepp H (2020) Rainfall–runoff prediction at multiple timescales with a single Long Short-Term Memory network. Hydrol Earth Syst Sci 25:2045–2062. https://doi.org/10.5194/hess-25-2045-2021

Niu WJ, Feng ZK, Feng BF, Xu YS, Min YW (2021) Parallel computing and swarm intelligence based artificial intelligence model for multi-step-ahead hydrological time series predictio. Sustain Cities Soc 66. https://doi.org/10.1016/j.scs.2020.102686

Ravindran SM, Bhaskaran SKM, Ambat SKN (2021) A deep neural network architecture to model reference evapotranspiration using a single input meteorological parameter. Environ Process 8:1567–1599. https://doi.org/10.1007/s40710-021-00543-x

Shoaib M, Shamseldin AY, Khan S, Khan MM, Khan ZM, Sultan T, Melville BW (2018) A comparative study of various hybrid wavelet feedforward neural network models for runoff forecasting. Water Resour Manag 32(1):83–103. https://doi.org/10.1007/s11269-017-1796-1

Sun QH, Miao CY, Duan QY, Ashouri H, Sorooshian S, Hsu K-L (2018) A review of global precipitation data sets: Data sources, estimation, and intercomparisons. Rev Geophys 56(1):79–107. https://doi.org/10.1002/2017RG000574

Tan QF, Lie XH, Wang X, Wang H, Wen X, Ji Y, Kang AQ (2018) An adaptive middle and long-term runoff forecast model using EEMD-ANN hybrid approach. J Hydrol 576:767–780. https://doi.org/10.1016/j.jhydrol.2018.01.015

Wang WC, Xu DM, Chau KW, Chen SY (2013) Improved annual rainfall-runoff forecasting using PSO–SVM model based on EEMD. J Hydroinf 15(4):1377–1390. https://doi.org/10.2166/hydro.2013.134

Wang WC, Chau KW, Qiu L, Chen YB (2015) Improving forecasting accuracy of medium and long-term runoff using artificial neural network based on EEMD decomposition. Environ Res 139:46–54. https://doi.org/10.1016/j.envres.2015.02.002

Wang YT, Liu J, Li R, Suo XY, Lu EH (2022) Medium and long-term precipitation prediction using wavelet decomposition-prediction-reconstruction model. Water Resour Manag 36(3):971–987. https://doi.org/10.1007/s11269-022-03063-x

Wang WC, Nie XT, Qiu L (2010) Support vector machine with particle swarm optimization for reservoir annual inflow forecasting. IEEE:184–188. https://doi.org/10.1109/AICI.2010.45

Xie XY, Tang HH, Wang JP, Li CH, Chen J, An W, Hu BX (2016) Runoff forecasting model in flood season based on weather factors. J Paper 33(10):12–14

Yue ZX, Ai P, Xiong CS, Song YH, Hong M, Yu JR (2020) Mid and long-term runoff forecasting based on improved deep belief networks model. J Hydroelectr Eng 39(10):33–46

Zhang X, Zhang XN, Wang GX, Wu W, Xu Y (2021) Medium and long-term forecast of the reservoir inflow in the upper yangtze river. Water Resour Protect 38(4):131–136

Zhao XY, Ping A, Ding BY, Chuan SX (2020) Ensemble approach for mid-long term runof forecasting using hybrid algorithms. J Ambient Intell Humaniz Comput. https://doi.org/10.1007/s12652-020-02345-9

Authors and Affiliations

Shengli Liao[1] · Huan Wang[1] · Benxi Liu[1] · Xiangyu Ma[1] · Binbin Zhou[2] · Huaying Su[3]

Shengli Liao
shengliliao@dlut.edu.cn

Benxi Liu
benxiliu@dlut.edu.cn

Xiangyu Ma
mxy0324@mail.dlut.edu.cn

Binbin Zhou
bbzhouchn@163.com

Huaying Su
suzz1234@126.com

1 Institute of Hydropower and Hydroinformatics, Dalian University of Technology, Dalian 116024, China

2 Yunnan Electric Power Dispatch Control Center of Yunnan Power Grid, Kunming 650011, China

3 Power Dispatching Control Center of Guizhou Power Grid, Guiyang 550000, China