Water Supply

Water Supply Vol 23 No 5, 1760 doi: 10.2166/ws.2023.075

Model and application of annual river runoff prediction based on complementary set empirical mode decomposition combined with particle swarm optimization adaptive neuro-fuzzy system

Huifang Guo [a], Lihui Chen [b], Yuan Fang [a,*] and Shixia Zhang [a]

[a] Zhejiang Tongji Vocational College of Science and Technology, Hangzhou, Zhejiang, China
[b] Zhejiang Provincial Hydrological Management Center, Hangzhou, Zhejiang, China
*Corresponding author. E-mail: z20120190801@zjtongji.edu.cn

HG, 0000-0002-9909-3627; YF, 0009-0006-9518-2194

ABSTRACT

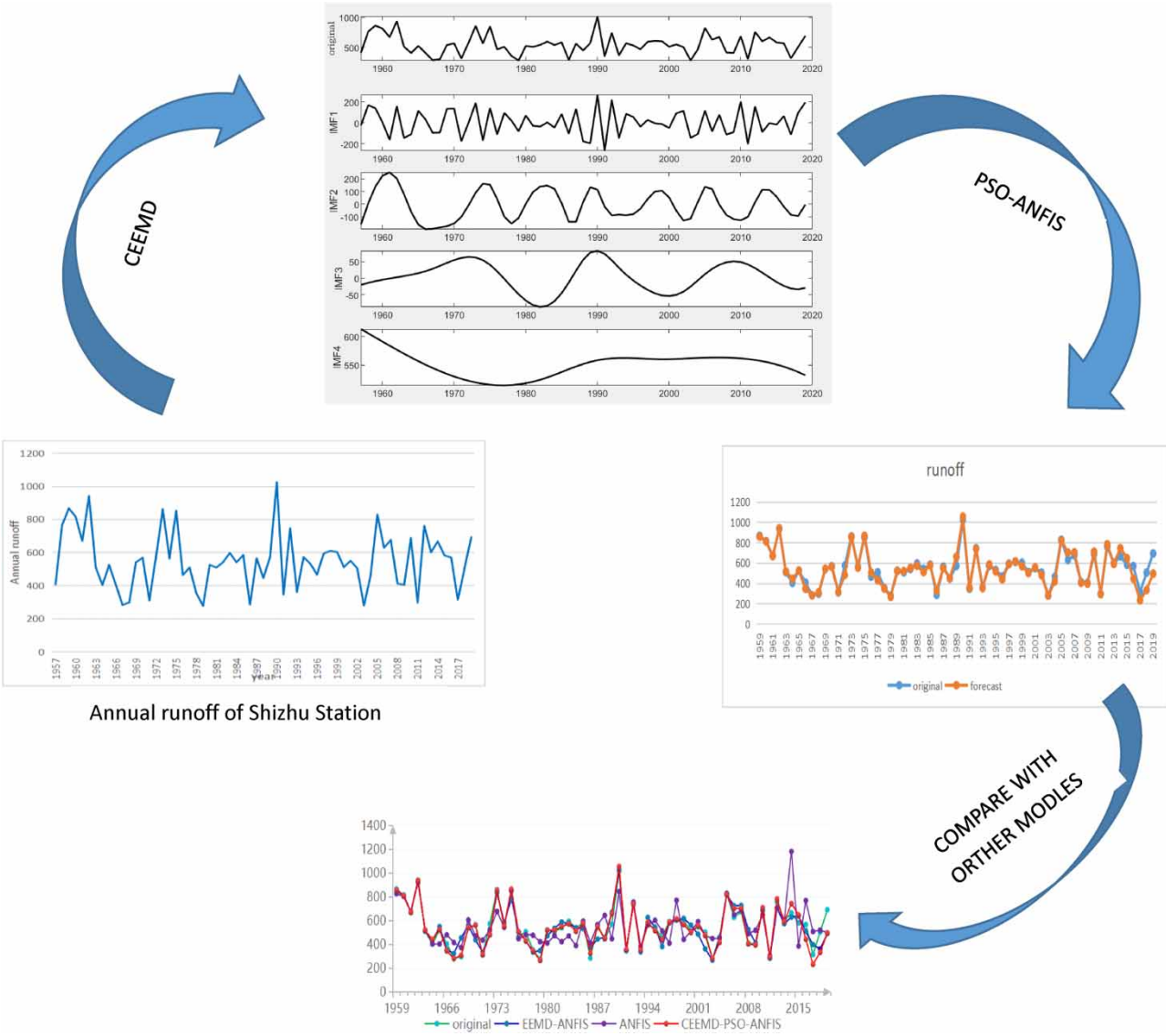
Runoff is affected by natural and nonnatural factors in the process of formation, and the runoff series is generally nonstationary time series. How to improve the accuracy of runoff prediction has always been a difficult problem for hydrologists. The key to solve this problem is to reduce the complexity of runoff series and improve the accuracy of runoff prediction model. Based on the aforementioned ideas, this article uses the complementary set empirical mode decomposition to decompose the runoff series into multiple intrinsic components that retain time–frequency information, thus reducing the complexity of the runoff series. The particle swarm optimization (PSO) adaptive neuro-fuzzy system is used to predict each intrinsic component to improve the accuracy of runoff prediction. After that, the trained intrinsic components of the model are reconstructed into the original runoff series. The example shows that the absolute relative error of the runoff forecasting model constructed in this article is 0.039, and the determination coefficient is 0.973. This model can be applied to the annual runoff series forecasting. Comparing the prediction results of this model with empirical mode decomposition algorithm-ANFIS model and ANFIS model, complementary set empirical mode decomposition algorithm-PSO-ANFIS model shows obvious advantages.

Key words: adaptive neuro-fuzzy system, annual runoff, complementary set empirical mode decomposition, forecast accuracy, runoff forecast

HIGHLIGHTS

- Complementary set empirical mode decomposition algorithm (CEEMD) is a good time series decomposition tool, which can reduce the nonstationary of time series.
- Particle swarm optimization (PSO)-ANFIS has good simulation and prediction results.
- CEEMD-PSO-ANFIS model can be applied to runoff series prediction.

GRAPHICAL ABSTRACT



1. INTRODUCTION

Runoff forecasting is an important research direction in the hydrological field. Runoff forecasting can not only provide basic reference for the use of water resources but also provide important data support for urban flood control scheduling. When the accuracy of runoff forecast is higher, the water resource management of the basin will be more reasonable, and the urban flood control and drainage scheduling will also be more scientific. However, in the process of runoff formation, it is not only affected by natural factors such as rainfall, temperature, and evaporation but also affected by many nonnatural factors such as human activities and urbanization process. The impact of these factors makes the runoff time series have nonstationary and nonlinear characteristics, and the impact of many factors also makes the runoff time series more complex. The complex runoff series, nonstationary, and nonlinear characteristics increase the difficulty of improving the accuracy of runoff forecast.

At present, the research in the field of runoff prediction at home and abroad can be roughly divided into two types (Hao *et al.* 2015): first, the quasi-physical model that uses mathematical methods to simulate the runoff process, such as the Xin'anjiang model; second, the formation process of runoff is simulated by the black box model. The black box model uses the computer's fast computing ability to establish the nonlinear relationship between the impact factors of runoff and runoff,

so as to simulate and predict. There are much research on the use of black box models at home and abroad, such as artificial neural networks, support vector machines, random forest methods, etc. For example, Yudong & Linsheng (1995) proposed that artificial neural network can be widely used in long-term runoff prediction. Xinmin *et al.* (1999) used conjugate gradient optimization and error back-propagation training algorithm to optimize the artificial neural network and applied it to the real-time prediction of river runoff. This method provides a way for the black box model of runoff prediction, that is, the runoff prediction model can be optimized by improving the model parameters. Jianyi & Chuntian (2006) applied the support vector machine model to the runoff forecast, and the results show that the accuracy of runoff forecast has been improved. Tongtie Gang *et al.* (2012) applied the stochastic forest method to the runoff forecast, and the forecast accuracy was improved. Han & Morrison (2022) used the deep learning method to improve the prediction effectiveness of the earlier hydrological model. The black box model is one of the widely used methods for runoff forecasting.

Due to the complexity and nonstationarity of runoff time series, it is difficult to improve the accuracy of runoff forecast. At present, the methods to improve the accuracy of runoff prediction include two key directions (Dongmei *et al.* 2022). The first is to reduce the complexity of the runoff series, that is, to decompose the complex runoff series into multiple simple time series. The second is to establish a black box model for runoff prediction with strong nonlinear approximation ability. The methods to reduce the complexity of runoff series include empirical mode decomposition (EMD), wavelet analysis, etc. It is to reduce the complexity of runoff series by decomposing the original nonstationary series into multiple series. For example, Shahabi *et al.* (2016) applied wavelet transform to decompose significant wave height (SWH) data, and the results show that this method can improve the prediction accuracy. Tian (2020) applied wavelet transform to decompose crop water demand data, extracted approximate signals, and decomposed signals of water demand, thus reducing the complexity of the model. Vousoughi (2023) applied wavelet analysis technology to decompose groundwater data, and the results showed that this method improved the accuracy of prediction. Zhang *et al.* (2017) applied the EMD model to extract the periodic and local characteristics of runoff. Ghasempour *et al.* (2021) used empirical mode decomposition algorithm (EEMD) and variational mode decomposition (VMD) to decompose the river flow, and the results show that the prediction accuracy of the model can be improved by using this method. After reducing the complexity of the model, the accuracy of runoff prediction can be effectively improved by using the model with strong nonlinear approximation ability. Ruiming (2019) applied wavelet transform and support vector machine to monthly runoff forecast, and the results showed that the method was effective. Rui *et al.* (2021) applied the linear network long short term memory (LN-LSTM)-particle swarm optimization (PSO) model to runoff forecasting, and the results showed that the accuracy of runoff forecasting was significantly improved. Xin *et al.* (2022) applied EMD and artificial neural network (ANN) to runoff prediction and obtain good results.

Based on the aforementioned idea of improving the accuracy of the model, this article introduces the complementary set EMD to decompose the runoff time series. In the process of runoff decomposition, this method can not only decompose the runoff series into simpler series but also give the correlation coefficient between the decomposed series and the original runoff series. The accuracy of runoff prediction can be increased by characterizing the intrinsic components of the original runoff series. Zhang *et al.* (2022a, 2022b) applied the CEEMD-bidirectional long short term memory (BISTM) model to monthly runoff forecast, and believed that this method could improve the accuracy of runoff forecast.

2. RESEARCH METHODS

2.1. Complementary ensemble empirical mode decomposition

Complementary set empirical mode decomposition algorithm (CEEMD), proposed by Yeh *et al.* (2010) in 2010, is improved on the basis of set EEMD. This method can effectively improve the decomposition efficiency by adding positive and negative white noise on the basis of traditional EMD. The traditional processing method of modal decomposition assumes that there is no noise in the original data, and the existence of noise is not considered in the process of decomposition, so modal aliasing will occur in the process of decomposition. In view of the aforementioned defects, EEMD solves the modal aliasing problem by adding uniformly distributed white noise several times to mask the noise of the original data. However, the white noise added by EEMD cannot be completely offset after the lumped average, and there is residual. To solve this problem, CEEMD adds positive and negative white noise in pairs. The positive and negative white noise will cancel each other when the set is averaged, thus improving the decomposition effect. The process of using CEEMD to decompose runoff time series is as follows:

(1) Add different positive and negative paired white noise $k(t)$ to the original runoff series $Y(t)$, the number of times of adding is k, the positive noise signal is $z_{k,a}(t)$, and the negative noise signal is $z_{k,b}(t)$.

(2) EMD decomposition of $z_{k,a}(t)$ and $z_{k,b}(t)$.

① The size and position of local maximum and minimum of $z_{k,a}(t)$ and $z_{k,b}(t)$ are analyzed.

② Apply the cubic spline difference curve to form the upper envelope lines $u_a(t)$ and $u_b(t)$ and the lower envelope lines $L_a(t)$ and $L_b(t)$ of the maximum and minimum values of $z_{k,a}(t)$ and $z_{k,b}(t)$.

③ Calculate the mean value of the upper envelope lines $u_a(t)$ and $u_b(t)$ of $z_{k,a}(t)$ and $z_{k,b}(t)$ and the lower envelope lines $L_a(t)$ and $L_b(t)$ at each time point $m_a(t) = [u_a(t) + L_a(t)]/2$ and $m_b(t) = [u_b(t) + L_b(t)]/2$. Calculate the difference $h_{1a}(t)$ between the mean $m_a(t)$ and $z_{k,a}(t)$ and the difference $h_{1b}(t)$ between the mean $m_b(t)$ and $z_{k,b}(t)$ to determine whether the requirements are met. If satisfied, record the current mean value $m_a(t)$, set it as the intrinsic modulus function $(\mathrm{IMF})_a$, and set the mean value $m_b(t)$ as the intrinsic modulus function IMF_b. If the assumed conditions are not met, replace the current data with the difference value and repeat the aforementioned steps until the assumed conditions are met.

④ When the set number of iterations n is reached:

$h_{1a,n}(t) = h_{1a,n-1}(t) - m_{1a,n-1}(t)$, $h_{1b,n}(t) = h_{1b,n-1}(t) - m_{1b,n-1}(t)$, if the set judgment conditions are met, the first intrinsic modulus of positive noise $\mathrm{IMF}_{1a}(t) = h_{1a,n}(t)$, and the first intrinsic modulus of negative noise $\mathrm{IMF}_{1b}(t) = h_{1b,n}(t)$.

⑤ Calculate residual sequence

$r_a(t) = z_{k,a}(t) - \mathrm{IMF}_{1a}(t)$, $r_b(t) = z_{k,b}(t) - \mathrm{IMF}_{1b}(t)$, judge whether it meets the EMD decomposition conditions. If it meets the conditions, the decomposition will be ended. If not, then another $h_{2,a}(t) = r_a(t)$, $h_{2,b}(t) = r_b(t)$, return to step ①.

⑥ Through the aforementioned steps, we can gradually calculate the positive noise sequence $z_{k,a}(t)$, the first to nth intrinsic modulus $\mathrm{IMF}_{ia}(t)$ and a residual $r_a(t)$, as well as the negative noise sequence $z_{k,b}(t)$, the first to nth intrinsic modulus $\mathrm{IMF}_{ib}(t)$ and a residual $r_b(t)$.

(3) Calculate the first to nth intrinsic modulus $\mathrm{IMF}_i(t)$ of the original runoff series $Y(t) = (\mathrm{IMF}_{ia}(t) + \mathrm{IMF}_{ib}(t))/2$, and the residual $r(t) = (r_a(t) + r_b(t))/2$.

(4) The sum of intrinsic modulus $\mathrm{IMF}_i(t)$ and residual $r(t)$ is the original data.

Due to the addition of white noise with positive and negative opposites, the two noises can cancel each other on average. This method not only improves the decomposition efficiency but also improves the shortcomings of large reconstruction error and poor decomposition completeness of EEMD. CEEMD has two important parameters during decomposition, including the number of noise additions and the standard deviation of noise additions. After consulting relevant literature (Zhang *et al.* 2022a, 2022b), the standard deviation is generally set as 0.10–0.2, and the number of noise additions is generally 50–100. The process of CEEMD decomposing runoff sequences is shown in Figure 1.

2.2. ANFIS model

Adaptive Network-based Fuzzy Inference System was proposed by J-S.R. Jang in the early 1990s. The model generally includes five layers: fuzziness, rule usage, normalized usage, Takagi-Sugeno-Kang (TSK) output layer, and total output layer (Daneshfaraz *et al.* 2022). The model is essentially a neural network based on fuzzy reasoning. It not only has the advantages of self-learning, adaptive, and nonlinear approximation of neural network (Talei *et al.* 2010) but also has fuzzy reasoning. This method has been proved to be effective in runoff forecasting (Nourani & Komasi 2013; Shabnam & Morteza 2022). It has two key parameters to learn: the former and the latter. Suppose a multi-input and single-output hierarchical fuzzy neural network, the output of the i nodes of its K layer is O_i^k, the expected output is R_i^L, and the actual output is O_i^L.

2.2.1. Learning of front part parameters

Assuming that the membership function used in the second layer is Gaussian type:

$$\mu_A(x) = \exp\left\{-\left(\frac{x - c_i}{a_i}\right)^2\right\} \tag{1}$$

where c_i and a_i are parameters related to nodes.

Assuming that there are training samples of P group, the training error function of $p(1 \le p \le P)$ group is defined as follows:

$$E_p = \sum_{m=1}^{\#k} (R_{m,p} - O_{m,p}^L)^2 \tag{2}$$

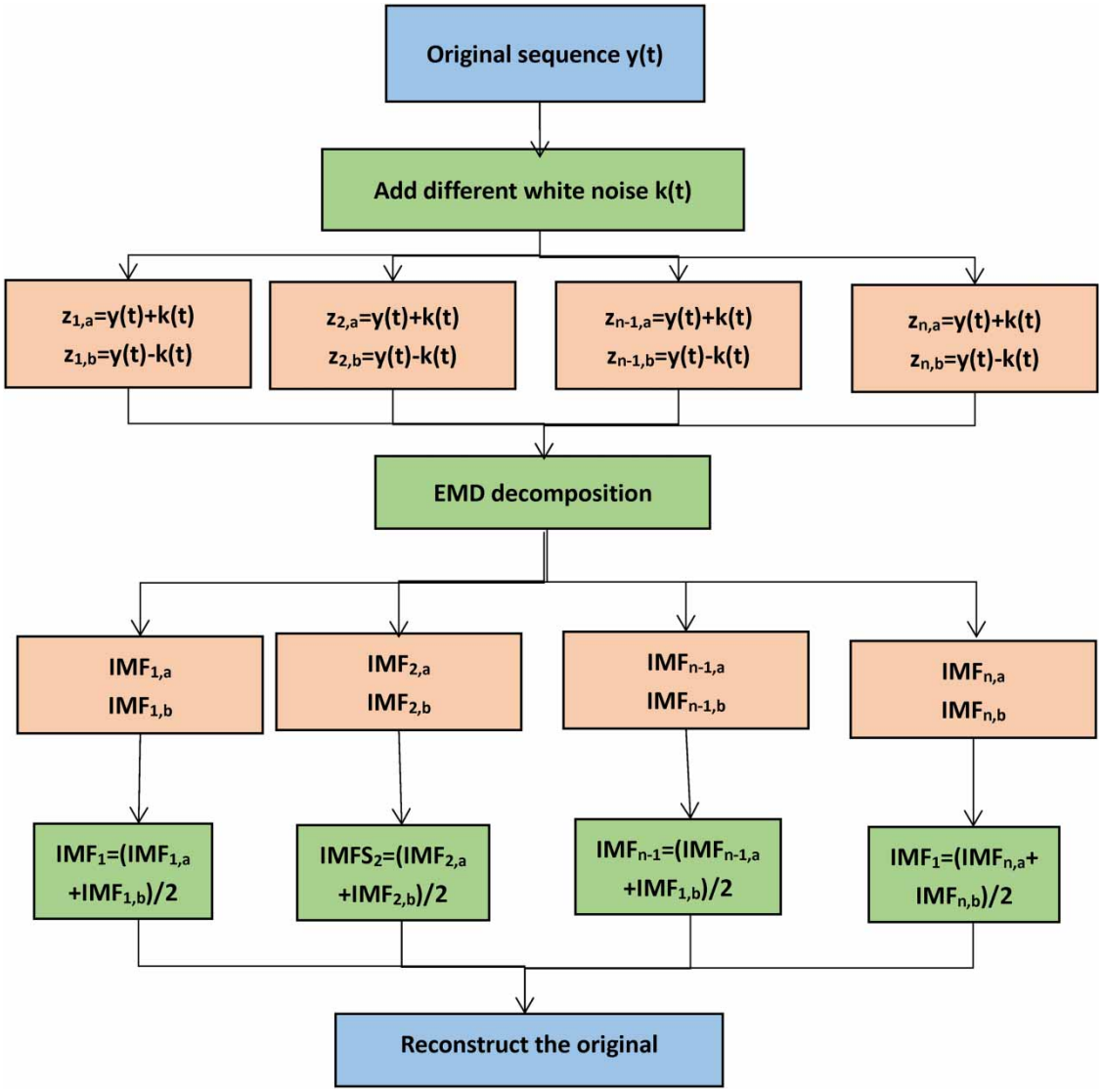


Figure 1 | Schematic diagram of CEEMD decomposition runoff sequence process.

The error function of the whole network is expressed as follows:

$$E = \sum_{p=1}^{P} E_p \tag{3}$$

The gradient descent method is used to improve the learning speed. First, the first derivative of the p group input vector to each output node O is calculated as follows:

$$\frac{\partial E_p}{\partial O_{m,p}^L} = -2(T_{i,p} - O_{i,p}^L) \tag{4}$$

The error change rate of (k, i) node is calculated as follows:

$$\frac{\partial E_p}{\partial O_{i,p}^k} = \sum \frac{\partial E_p}{\partial O_{m,p}^{k+1}} \frac{\partial O_{m,p}^{k+1}}{\partial O_{i,p}^k}, \quad 1 \le k \le L-1 \tag{5}$$

Convert it as follows:

$$\frac{\partial E_p}{\partial a} = \sum_{O^* \in S} \frac{\partial E_p}{\partial O^*} \frac{\partial O^*}{\partial a} \tag{6}$$

Among them, a is a parameter of the adaptive network, S is the set of nodes related to a output, and then the error index function of the whole network about a is calculated as follows:

$$\frac{\partial E}{\partial a} = \sum_{p=1}^{P} \frac{\partial E_p}{\partial a} \tag{7}$$

Taking a as an example, the parameter update formula is calculated as follows:

$$\begin{aligned} \Delta a &= -\eta \frac{\partial E}{\partial a} \\ a_{i+1} &= a_i + \eta \frac{\partial E}{\partial a} \end{aligned} \tag{8}$$

η is learning efficiency:

$$\eta = \frac{k}{\sqrt{\sum_a \left(\frac{\partial E}{\partial a}\right)^2}} \tag{9}$$

where k is the step size, and the speed of η can be changed by changing the speed of k.

2.2.2. Learning of subsequent parameters

The expected output R_i^L and actual output O_i^L have the following relationship with the corresponding expected input f_i:

$$O_1^L = \sum \overline{\omega_i} f_i = X^T W \tag{10}$$

where

$$\begin{aligned} X^T &= [\overline{w_1}x, \overline{w_1}y, \overline{w_1}, \overline{w_2}x, \overline{w_2}y, \overline{w_2}, \cdots \overline{w_n}x, \overline{w_n}y, \overline{w_n},] \\ W^T &= [p_1, q_1, r_1, p_2, q_2, r_2 \cdots p_n, q_n, r_n] \end{aligned} \tag{11}$$

$R = (h_1, h_2, \cdots h_n)$, and (x_i, y_i, h_i) is the i sample data pair.

According to the least square method, its performance index is defined as follows:

$$J = (2/1) \sum_{q=1}^{Q} \lambda^{Q-q} [R_1^L(q) - O_1^L(q)]^2 \tag{12}$$

Among them, λ is the forgetting factor, and the value is 0–1. The smaller the λ, the weaker the influence of historical data. Bring O_1^L into J:

$$J = (2/1) \sum_{q=1}^{Q} \lambda^{Q-q} [R_1^L(q) - X^T(q)W]^2 \tag{13}$$

From $\frac{\partial L}{\partial W} = 0$, we obtained

$$\begin{aligned} \varphi^T(q) &= [(\lambda^{Q-1})^{0.5} X(1), (\lambda^{Q-2})^{0.5} X(2), \cdots, (\lambda^{Q-q})^{0.5} X(q)] \\ \Gamma^T(q) &= [(\lambda^{Q-1})^{0.5} R_1^L, (\lambda^{Q-2})^{0.5} R_2^L, \cdots, (\lambda^{Q-q})^{0.5} R_q^L] \end{aligned} \tag{14}$$

The recursive least square method is used to solve the problem, and the corresponding estimation is calculated as follows:

$$W(q+1) = W(q) + S(q+1)X(q+1)[R_1^L(q+1) - X^T(q+1)W(q)] \tag{15}$$

where

$$S(q+1) = \frac{1}{\lambda}\frac{S(q) - S(q)X(q+1)X^T(q+1)S(q)}{\lambda + X^T(q+1)S(q)X(q+1)} \tag{16}$$

$S(q+1)$ becomes the covariance matrix. Usually, the initial piece is selected as $W(0) = 0$, $S(0) = \gamma I$, I is the unit matrix, γ is a large positive number, and a positive number of 103 can be taken. When $\gamma \to \infty$, the recursive least square method is close to the real value.

2.3. PSO optimized ANFIS model

The learning of the antecedent parameters and the consequent parameters is the most important content of the ANFIS model (Tao *et al.* 2021). In the calculation process of the ANFIS model, the antecedent parameters are calculated based on the gradient descent algorithm, which will lead to a large amount of calculation of the antecedent parameters and easy to fall into local minimization. This will slow down the calculation speed of the model and make it difficult to reach the best-fitting parameters. In order to improve this problem, the global optimization algorithm can be used.

PSO is a global optimization algorithm. Each particle represents a candidate solution in the solution space, sets the initial flight speed, lets the particles fly and search, and dynamically adjusts the flight speed during the search process. The i particle is expressed as $X_i = (X_{i1}, X_{i2,\ldots}X_{in})$. By substituting it into the fitness function, the fitness value of the particle at the position can be calculated. The size of the value represents the degree of the position. The best solution found by the particle is recorded as $P_i = (P_{i1}, P_{i2,\ldots}P_{in})$, which is called the individual extreme point pbest. The best position found by the group is recorded as $P_g = (P_{g1}, P_{g2,\ldots}P_{gn})$, which is called the global extreme point gbest. The velocity and position of any particle i in its d-dimensional space ($1 \le d \le D$) are updated as follows:

$$V_{id}(k+1) = w \times V_{id}(k) + c_1 \times \text{rand}() \times (P_{id}(k) - X_{id}(k)) + c_2 \times \text{rand}() \times (P_{gd}(k) - X_{id}(k)) \tag{17}$$

$$X_{id}(k+1) = X_{id}(k) + V_{id}(k+1) \tag{18}$$

where w is the inertia weight coefficient, c_1 and c_2 are the acceleration factors, rand() is the random number in the range [0,1], X_{id} is the position of particle I in the D-dimensional space, V_{id} is the velocity of particle i in the D-dimensional space, $V_{id} \in [-Vi\max_{i\max}, P_{id}$ are the individual extreme points, and P_{gd} is the global extreme points.

In order to prevent particles from exceeding the set variable range during flight, the maximum speed $V_{\max}$ of flight is generally set according to the variable range of the problem to be solved. If the current speed is $V > V_{\max}$ or $V < -V_{\max}$, take $V = V_{\max}$ or $V = -V_{\max}$.

The PSO-ANFIS model uses the global optimization ability of PSO to optimize the design of the network structure parameters of ANFIS. The purpose is to improve the accuracy of the parameter identification of ANFIS, thus improving the accuracy of the runoff forecast model. Its calculation process is as follows:

(1) Determine the scale, inertia weight and acceleration factor of PSO, and replace the antecedent parameters of ANFIS with initialized particles.
(2) Estimate the post-part parameters from the calculated pre-part parameters, calculate the model structure error at this time, and take the error as the fitness value of PSO.
(3) Compare the current fitness value with the best value to determine whether to replace the best value.
(4) Update the speed and position of PSO particles. Iteration is repeated until the optimal ANFIS parameter is found. The calculation process of PSO-ANFIS is shown in Figure 2.

2.4. CEEMD-PSO-ANFIS model

The design idea of the model is:

(1) By analyzing the frequency characteristics of the original runoff series, the frequency and frequency of adding noise to CEEMD are set.

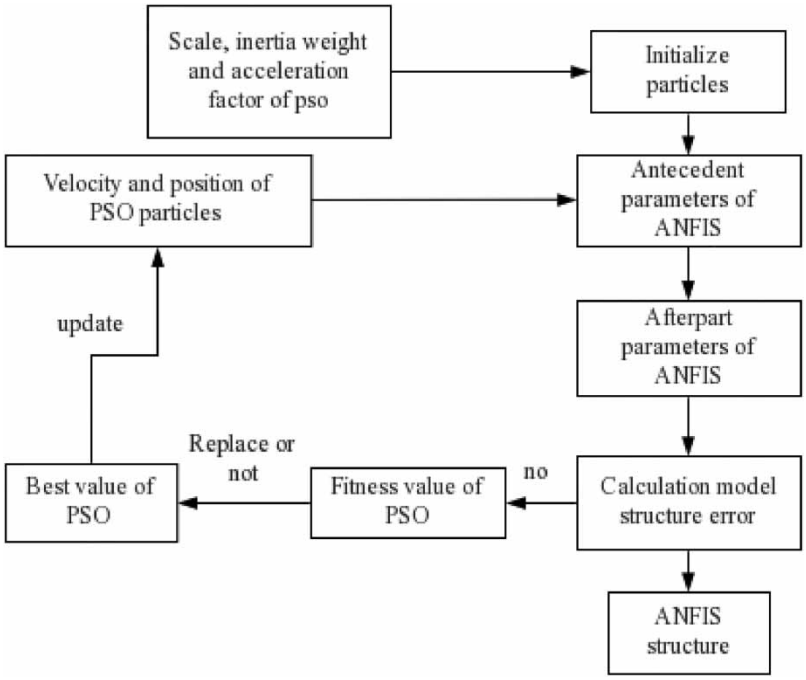


Figure 2 | PSO-ANFIS calculation process diagram.

(2) The runoff series is decomposed into several IMF components by CEEMD.
(3) The PSO-ANFIS model is used to predict each IMF component separately. According to the characteristics of each IMF component, the number of PSO particle swarm, inertia weight coefficient, acceleration factor, and membership function family number of ANFIS model are set.
(4) The simulation prediction results of each IMF component are reconstructed to obtain the simulation sequence results of the original runoff series.

2.5. Model evaluation criteria

In order to verify the results of the model, with reference to the relevant literature (Norouzi *et al.* 2021; Wu & Wu 2022), this article uses the average absolute value percentage error (M), root mean square error (R), average absolute pair error (M'), and determination coefficient (R^2) to evaluate the simulation and prediction results of the model. These parameters are calculated as follows:

$$M = \frac{1}{N}\sum_{i=1}^{n}\left|\frac{y_i - \widehat{y}_i}{y_i}\right| \times 100\%$$

$$R = \sqrt{\frac{1}{N}\sum_{i=1}^{n}(y_i - \widehat{y}_i)^2 \times 100\%}$$

$$M' = \frac{1}{N}\sum_{i=1}^{n}|y_i - \widehat{y}_i| \times 100\% \tag{19}$$

$$R^2 = 1 - \frac{\frac{1}{N}\sum_{i=1}^{n}(y_i - \widehat{y}_i)^2}{\frac{1}{N}\sum_{i=1}^{n}(y_i - \bar{y})^2}$$

where y_i is the original runoff value, $\widehat{y}_i$ is the predicted runoff value, N is the number of series, and $\bar{y}$ is the average value of runoff.

3. CASE APPLICATION

3.1. Overview of the study area

The Nanxi River basin is located on the eastern coast of Zhejiang Province, China. Its main stream has a total length of 142.2 km, a drainage area of 2,436 km^2, and an average gradient of 6.0‰. The main stream is a mountain stream channel above Shatou, 61 km above Xikou, with a river gradient of 10.46‰, and 48 km from Xikou to Shatou, with a river gradient of 1.12‰. The river below Shatou is affected by the tide of Oujiang River estuary, which is a tidal river section. The data are selected from Shizhu Hydrological Station, which is a national hydrological station. The station was set up by the Zhejiang Provincial Hydrological Bureau on 1 June 1955, with a catchment area of 1,273 km^2, to observe water level, water surface gradient, and precipitation. The data selected this time are the annual runoff data of Shizhu Station from 1957 to 2019, with a total of 63 data. From the figure, we can see that the annual runoff series of Shizhu Station on the Nanxi River is nonstationary.

3.2. CEEMD decomposition

We use MATLAB software to establish the CEEMD decomposition model of annual runoff at Shizhu Station of Nanxi River, where the variance of added noise is set to 0.1 and the frequency of added noise is 50. By referring to the optimal decomposition level of EEMD as four, the decomposition time of annual runoff CEEMD is set as four. The decomposition results are shown in Figure 3. From this figure, we can see that the annual runoff data of Shizhu Station in Nanxi River from 1957 to 2019 is decomposed into four intrinsic components. The variation process of the original runoff data is shown in Figure 4.

From Figure 5, we can see that the fluctuation of IMF1 is basically the same as that of the original sequence. With the increase in the number of decomposition layers, the fluctuation gradually becomes stable. It can be seen that CEEMD decomposition reduces the complexity of the runoff series and makes the runoff series gradually become stable, which is more conducive to the simulation and prediction of runoff. The correlation coefficient between IMF1–IMF4 and the original sequence is shown in Table 1.

From Table 1, we can see that the correlation between IMF1, IMF2, and the original sequence is the best, the correlation coefficient between IMF3 and the original sequence is general, and the correlation between IMF4 and the original sequence is weak.

3.3. Simulation and prediction of runoff series

The PSO-ANFIS model is used to simulate and predict the decomposed components. Use the data of T and T + 1 to predict the data of T + 2, in which the data of 1957–2014 is used as the training sequence, and the data of 2014–2019 is used as the prediction sequence. In the process of calculation, it is necessary to determine the population number, iteration number, individual acceleration, group acceleration, and speed change weight of the particle swarm according to the data of different components. In addition, it is necessary to determine the membership function and the cluster number of the ANFIS model. By observing the error change process and the change process of individual optimal value of particle swarm, it is determined that the number of IMF1, IMF2, IMF3, and IMF4 populations is 40, the number of iterations is 1500, the individual acceleration is 1, the group acceleration is 2, and the weight of velocity change is 0.5. ANFIS selects the Gaussian function as the membership function, the number of membership function clusters of IMF1 is 50, and the number of function clusters of IMF2, IM3, and IM4 is 10. The prediction results of IMF1–IMF4 and the original sequence are shown in Figure 6.

From Figure 6, we can see that the simulation and prediction effects of the trend item IMF4 are very good, which shows that the nonstationary series can be better predicted after being separated from the stationary series. In particular, we can see that the more the number of decomposition layers of the series, the better the simulation and prediction effect. This also verifies that the runoff series can improve the runoff prediction results after decomposition.

From Table 2, through statistics of the percentage error (M), root mean square error (R), average absolute pair error (M'), and determination coefficient (R^2) of the original runoff series, IMF1–IMF4, we can see that the calculation error is gradually decreasing. Among them, the simulation and prediction effect of IMF4 is the best. IMF1 has a strong volatility and stronger nonstationary. Its decision coefficient in the simulation stage is 0.967, but its decision coefficient in the prediction stage is relatively low, with a value of 0.191.

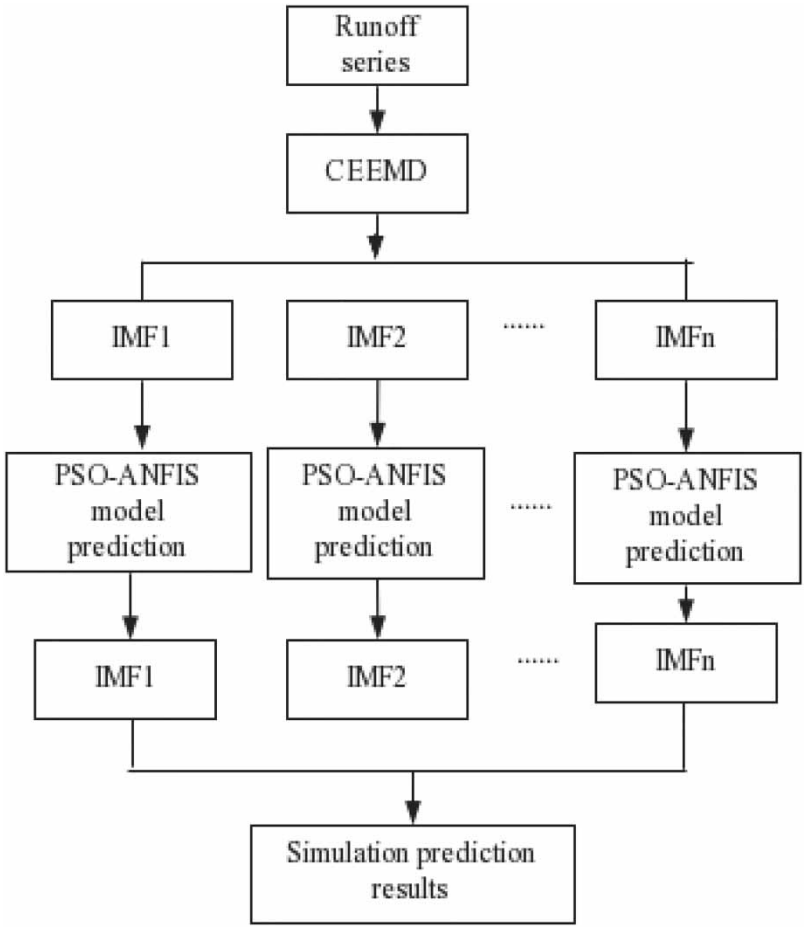


Figure 3 | CEEMD-PSO-ANFIS model design diagram.

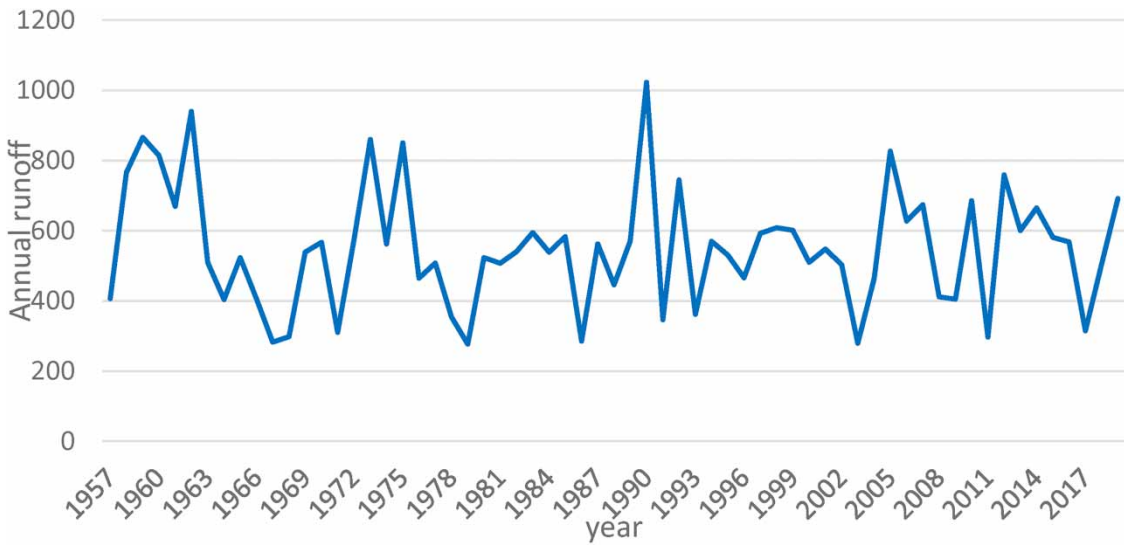


Figure 4 | Annual runoff of Shizhu Station on Nanxi River.

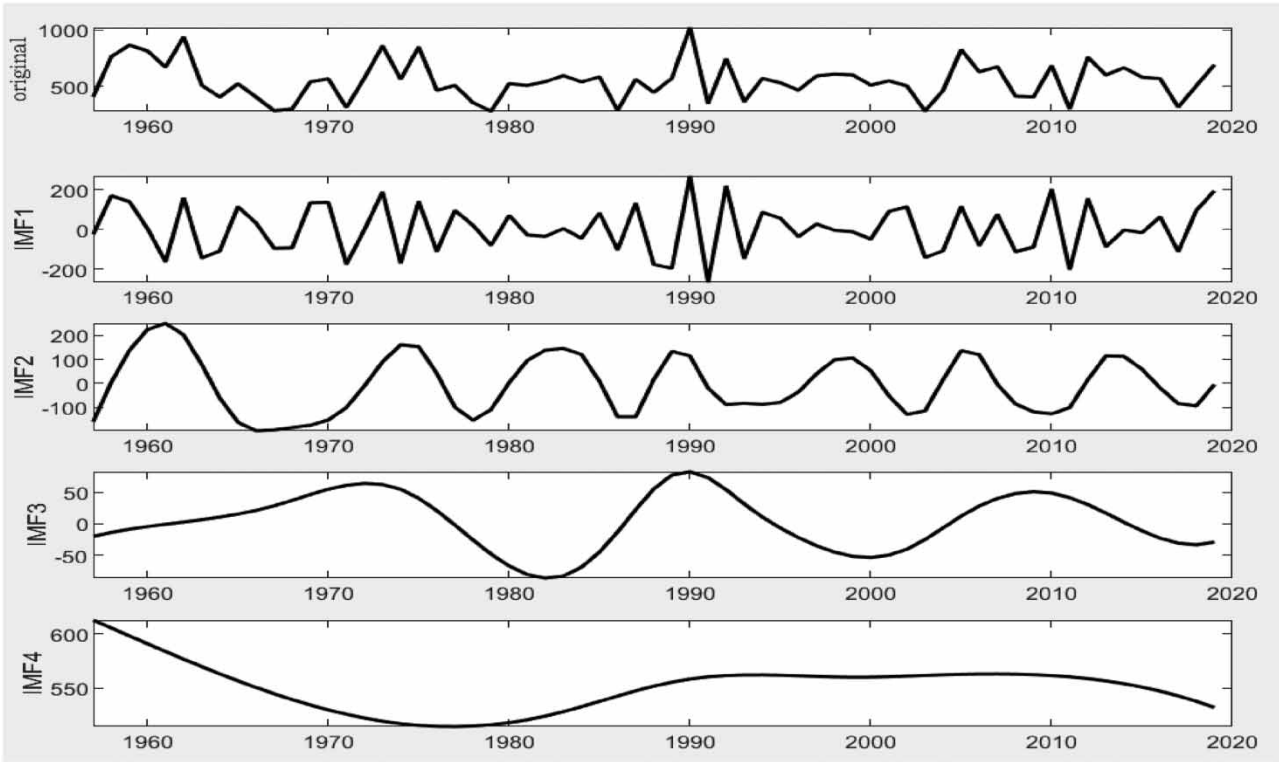


Figure 5 | CEEMD decomposition of annual runoff series of Shizhu Station.

Table 1 | Correlation coefficient and information entropy of IMF1–IMF4 decomposition sequence and original sequence

	Original sequence	IMF1	IMF2	IMF3	IMF4
Correlation coefficient	1.000	0.695	0.639	0.057	0.224
Information entropy	0.130	0.142	0.093	0.071	0.065

4. DISCUSSION

In order to verify the performance of CEEMD-PSO-ANFIS model, the ANFIS model and EEMD-ANFIS model are compared with it this time. The comparison results of each model are shown in Figure 7.

The evaluation results of the simulation and prediction stages of each model are shown in Table 3.

5. CONCLUSION

(1) The CEEMD model is decomposed by adding positive and negative white noise to reduce the deviation in the process of reconstructing the original sequence. First, the original runoff series is decomposed into multiple intrinsic components, then the PSO-ANFIS model is used to predict each component, and finally, the predicted components are reconstructed into the original runoff series. Through decomposition, the runoff series with strong nonstationarity is changed into more stable multiple intrinsic components, thus increasing the prediction accuracy.

(2) The average error of training results of CEEMD-PSO-ANFIS model is 0.039, and the determination coefficient is 0.973. Its model evaluation results are better than the ANFIS model and the EEMD-ANFIS model, which indicates that the model is feasible for annual runoff prediction.

(3) However, in the process of calculation, it is found that for IMF1 with strong nonstationary, if you want to obtain better prediction accuracy, you must increase the number of membership function families of the ANFIS model, which will make the calculation speed of the model slower than IMF2-IMF4 to a certain extent. In the process of calculation, the

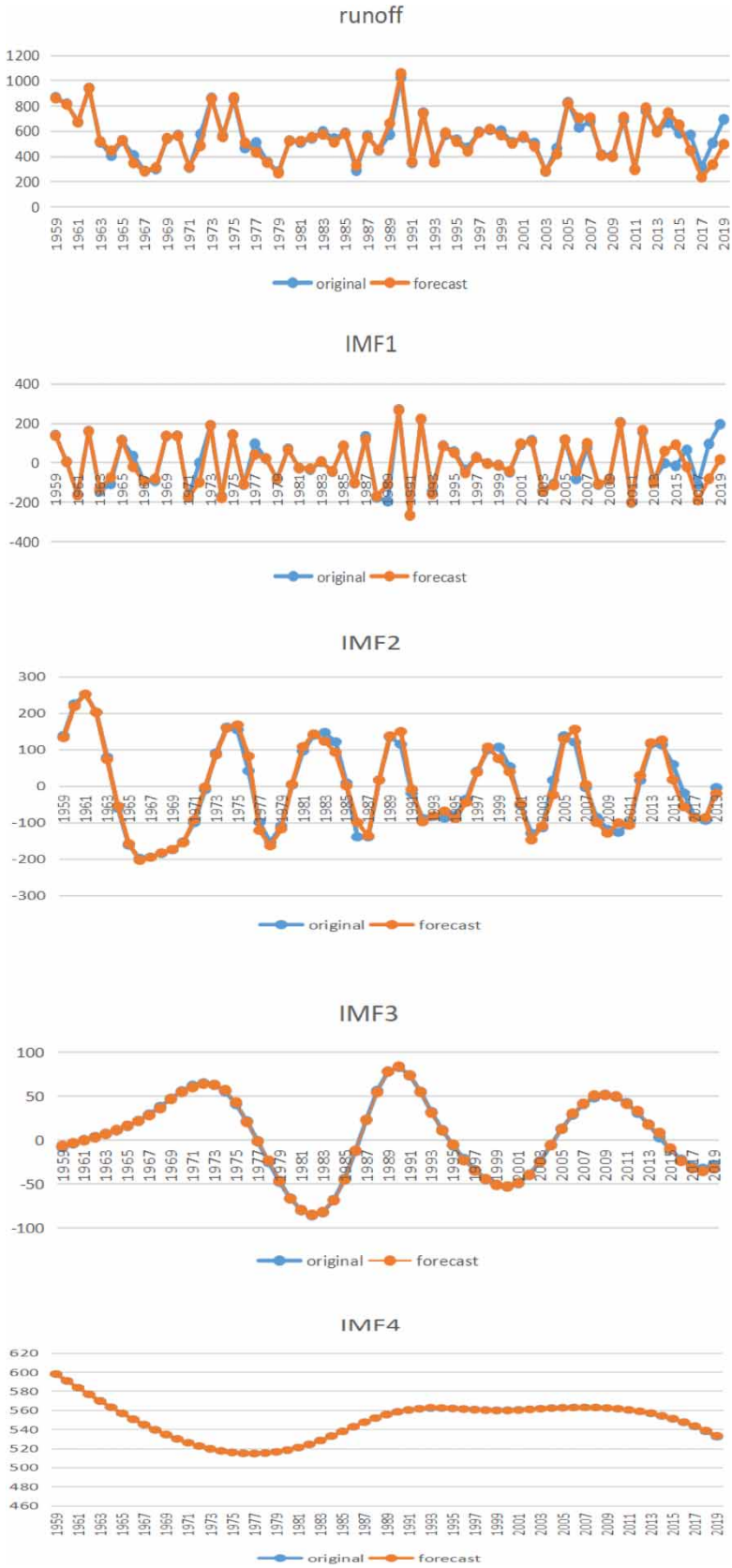


Figure 6 | Simulation prediction results.

Table 2 | Evaluation table of simulation prediction results

Sequence	Stage	M	R	M′	R^2
Original	Simulation	0.039	19.780	29.668	0.973
	Forecast	0.223	119.971	129.655	0.446
IMF1	Simulation	1.217	10.723	22.513	0.967
	Forecast	4.581	114.317	123.477	0.191
IMF2	Simulation	0.255	11.372	15.886	0.983
	Forecast	1.011	19.825	24.823	0.888
IMF3	Simulation	0.038	0.687	0.930	1.000
	Forecast	0.413	2.751	3.056	0.961
IMF4	Simulation	0.000	0.080	0.147	1.000
	Forecast	0.001	0.577	0.692	0.991

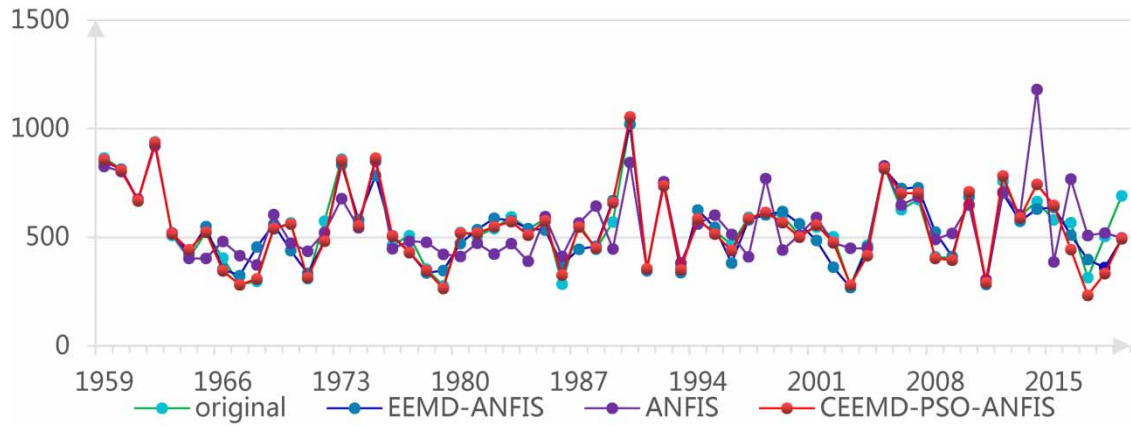


Figure 7 | Comparison of simulation and prediction results of various models.

Table 3 | Comparison of evaluation results of each model

Model	Stage	M	R	M′	R^2
CEEMD-PSO-ANFIS	Simulation	0.039	19.780	29.668	0.973
	Forecast	0.223	119.971	129.655	0.446
EEMD-ANFIS	Simulation	0.090	41.899	57.586	0.886
	Forecast	0.182	96.026	111.635	−0.138
ANFIS	Simulation	0.149	68.715	91.970	0.609
	Forecast	0.399	219.062	264.449	0.013

number of PSO population and the number of families of ANFIS membership function mainly depend on experience. In the future, it is necessary to increase the research of relevant aspects to determine the parameters of the model more accurately and scientifically.

AUTHOR CONTRIBUTION

All authors contributed to the study conception and design. Huifang Guo and Yuan Fang contributed to writing and editing; Shixia Zhang contributed to chart editing; and Lihui Chen contributed to preliminary data collection. All authors read and approved the final manuscript.

FUNDING

This research was supported by a Provincial cultivation project of the Natural Science Foundation of Zhejiang Province (Grant No. LZJWY22E090005), Zhejiang Tongji Vocational College of Science and Technology (Grant No. FRF21PY001), Science and Technology Project of Zhejiang Water Resources Department in 2021 (Grant No. RC2110), and Science and Technology Project of Zhejiang Water Resources Department in 2020 (Grant No. RC2031).

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

Daneshfaraz, R., Abam, M., Heidarpour, M., Abbasi, S., Seifollahi, M. & Abraham, J. 2022 The impact of cables on local scouring of bridge piers using experimental study and ANN, ANFIS algorithms. *Water Supply* **22** (1), 1075–1093. https://doi.org/10.2166/ws.2021.215.

Dongmei, X., Yaqin, W. & Wenchuan, W. 2022 Monthly precipitation prediction model based on VMD-TCN. *Hydrology* **42** (02), 13–18. doi: 10.19797/j.cnki.1000-0852.20210101.

Ghasempour, R., Azamathulla, H. M. & Roushangar, K. 2021 EEMD- and VMD-based hybrid GPR models for river streamflow point and interval predictions. *Water Supply* **21** (7), 3960–3975. https://doi.org/10.2166/ws.2021.151.

Han, H. & Morrison, R. R. 2022 Improved runoff forecasting performance through error predictions using a deep-learning approach. *Journal of Hydrology* **608**, 127653.

Hao, W., Yang, L., Liliang, R., Jianhua, W., Denghua, Y. & Fan, L. 2015 Hydrological model uncertainty and overall framework of ensemble simulation. *Water Resources and Hydropower Technology* **46** (06), 21–26. doi:10.13928/j.cnki.wrahe.205.06.004.

Jianyi, L. & Chuntian, C. 2006 Application of support vector machine in medium- and long-term runoff prediction. *Journal of Water Conservancy* **37** (06), 681–686.

Norouzi, R., Sihag, P., Daneshfaraz, R., Abraham, J. & Hasannia, V. 2021 Predicting relative energy dissipation for vertical drops equipped with a horizontal screen using soft computing techniques. *Water Supply* **21** (8), 4493–4513. doi: https://doi.org/10.2166/ws.2021.193.

Nourani, V. & Komasi, M. 2013 A geomorphology-based ANFIS model for multi-station modeling of rainfall–runoff process. *Journal of Hydrology* **490**, 41–55.

Rui, F., Guangwei, F., Jianyi, L., Baozhen, Y. & Qinghai, G. 2021 Enhanced long short-term memory model for runoff prediction. *Journal of Hydrologic Engineering* **26** (2), doi: 10.1061/(ASCE)HE.1943-5584.0002035.

Ruiming, F. 2019 Wavelet based relevance vector machine model for monthly runoff prediction. *Water Quality Research Journal* **54** (2), 134–141.

Shabnam, V. & Morteza, M. S. 2022 Rainfall–runoff modeling using adaptive neuro-fuzzy inference system (ANFIS) and genetic algorithm (GA). *Water Supply* **22** (10), 7460–7475.

Shahabi, S., Khanjani, M.-J. & Kermani, M. H. 2016 Hybrid wavelet-GMDH model to forecast significant wave height. *Water Supply* **16** (2), 453–459. https://doi.org/10.2166/ws.2015.151.

Talei, A., Chua, L. H. C. & Wong, T. S. W. 2010 Evaluation of rainfall and discharge inputs used by adaptive network-based fuzzy inference systems (ANFIS) in rainfall–runoff modeling. *Journal of Hydrology* **391** (3), 248–262.

Tao, L., Hangpeng, L., Shenshen, Z., Chenglin, Z. & Tongkuan, M. 2021 Research on short-term wind speed prediction based on SSA-PSO-ANFIS. *Journal of Solar Energy* **42** (03), 128–134. doi: 10.19912/j.0254-0096.tynxb.2018-1184.

Tian, Z. 2020 A combined prediction approach based on wavelet transform for crop water requirement. *Water Supply* **20** (3), 1016–1034. https://doi.org/10.2166/ws.2020.024.

Tongtie Gang, Z., Dawen,, Y., Ximing, C. & Yong, C. 2012 Study on runoff prediction in dry season in the upper reaches of the Yangtze River based on stochastic forest model. *Journal of Hydropower* **31**, 18–24, 38.

Vousoughi, F. D. 2023 Wavelet-based de-noising in groundwater quality and quantity prediction by an artificial neural network. *Water Supply* **23** (3), 1333–1348. ws2023021. https://doi.org/10.2166/ws.2023.021

Wu, D. & Wu, C. 2022 Research on the time-dependent split delivery green vehicle routing problem for fresh agricultural products with multiple time windows. *Agriculture* **12**, 793. https://doi.org/10.3390/agriculture12060793.

Xin, J., Jungang, L., Shangyao, Z. & Na, W. 2022 Runoff forecasting model based on variational mode decomposition and artificial neural networks. *Mathematical Biosciences and Engineering: MBE* **19** (2), 1633–1648.

Xinmin, X., Yunzhong, J., Yubo, S., Yong, D. & Chunxiao, Y. 1999 Research on real-time prediction of river runoff based on artificial neural network. *Water Conservancy and Hydropower Technology* **30** (09), 1–4.

Yeh, J. R., Shieh, J. S. & Huang, N. E. 2010 Complementary ensemble empirical mode decomposition: a novel noise enhanced data analysis method. *Advances in Adaptive Data Analysis* **2** (2), 135–156.

Yudong, C. & Linsheng, Y. 1995 Artificial neural network method for long-term runoff prediction. *Progress in Water Science* **6** (01), 61–65.

Zhang, J., Zhao, Y. & Lin, X. 2017 Uncertainty analysis and prediction of river runoff with multi-time scales. *Water Supply* **17** (3), 897–906. doi: 10.2166/ws.2016.190.

Zhang, X., Shi, J., Zhu, G., Xiao, Y. & Chen, H. 2022a Study of regional monthly precipitation based on CEEMD-BILSTM coupled model. *Water Supply* **22** (11), 8036–8049. https://doi.org/10.2166/ws.2022.321.

Zhang, X., Duan, B., He, S., Wu, X. & Zhao, D. 2022b A new precipitation forecast method based on CEEMD-WTD-GRU. *Water Supply* **22** (4), 4120–4132. https://doi.org/10.2166/ws.2022.037.

First received 7 August 2022; accepted in revised form 28 February 2023. Available online 14 March 2023