Check for updates

Improved monthly runoff time series prediction using the SOA–SVM model based on ICEEMDAN–WD decomposition

Dong-mei Xu[a], Xiang Wang[a], Wen-chuan Wang [a],*, Kwok-wing Chau[b] and Hong-fei Zang[a]

[a] College of Water Resources, Henan Key Laboratory of Water Resources Conservation and Intensive Utilization in the Yellow River Basin, North China University of Water Resources and Electric Power, Zhengzhou 450046, People's Republic of China
[b] Department of Civil and Environmental Engineering, The Hong Kong Polytechnic University, Hung Hom, Kowloon, Hong Kong, People's Republic of China
*Corresponding author. E-mail: wangwen1621@163.com; wangwenchuan@ncwu.edu.cn

W-cW, 0000-0003-1367-5886

ABSTRACT

In runoff prediction, the prediction accuracy is often affected by the non-linear and non-stationary characteristics of the runoff series. In this study, a coupled forecasting model is proposed that decomposes the original runoff series by an improved complete ensemble Empirical Mode Decomposition (EMD) (ICEEMDAN) combined with a wavelet decomposition (WD) and then forecasts the monthly runoff using a support vector machine (SVM) optimized by the seagull optimization algorithm (SOA). In this method, a series of Intrinsic Mode Function (IMF) and a Residual (Res) are obtained by decomposing the original runoff series with ICEEMDAN. The WD method is used to perform quadratic decomposition of high-frequency components decomposed by the ICEEMDAN method to make the runoff series as smooth as possible. Then the decomposed components are input into the SOA-SVM model for prediction. Finally, the prediction results of each component are superimposed and reconstructed to obtain the final monthly runoff prediction results. RMSE, Mean Absolute Percentage Error (MAPE), Nash-Sutcliffe Efficiency Coefficient (NSEC), and R are selected to evaluate the prediction results and the model is compared with SOA-SVM model, EMD-SOA-SVM model and CEEMDAN-SOA-SVM model other models. The proposed model is applied to the monthly runoff forecast of the Hongjiadu and Manwan Reservoirs. When compared with other benchmarking models, the ICEEMDAN-WD-SOA-SVM model attains the smallest Root Mean Square Error (RMSE) and MAPE and the largest NSEC and R. The ICEEMDAN-WD-SOA-SVM model has the best prediction effect, the highest prediction accuracy, and the lowest prediction error.

Key words: Improved complete ensemble EMD (ICEEMDAN), monthly runoff prediction, quadratic decomposition, seagull optimization algorithm (SOA), support vector machine (SVM), wavelet decomposition (WD)

HIGHLIGHTS

- The ICEEMDAN–WD model is used to decompose the original runoff series.
- The proposed ICEEMDAN–WD model can effectively reduce the complexity of the runoff series.
- The proposed SOA–SVM model can effectively improve the prediction accuracy of runoff series.
- The proposed model can provide high prediction accuracy and consistency.

ACRONYMS LIST

ANN   Artificial Neural Network
ARMA: Auto-Regressive and Moving Average Model
EMD: Empirical Mode Decomposition
EEMD: Ensemble Empirical Mode Decomposition
ELM: Extreme Learning Machine
GBRT: Gradient Boosting Regression Tree
IMF: Intrinsic Mode Function
LSSVM: Least Squares Support Vector Machine
LSTM: Long Short-Term Memory
MAPE: Mean Absolute Percentage Error
NSEC: Nash-Sutcliffe Efficiency Coefficient

R: correlation coefficient
Res: Residual
RMSE: Root Mean Square Errorl
VMD: Variational mode decomposition

1. INTRODUCTION

Runoff is the result of a comprehensive influence of the environment, including various climatic factors and human activities in a basin. Runoff changes with the environment as combined effect of climate change and human activities (Yong *et al.* 2017; Luo *et al.* 2019; Shao *et al.* 2021). Since fluctuations of runoff are mainly due to environmental change, the needs arise to have higher runoff prediction standards and requirements. Runoff prediction and prevention of natural disasters caused by climate change should be considered as key engineering measures. The prediction needs to account for the actual regional environmental factors and produce accurate, stable and effective results (Wang *et al.* 2021). It is important to know that accurate runoff forecasts are essential for water supply, flood control, drought relief, hydropower and shipping (Fang *et al.* 2019; Xu *et al.* 2022). Due to the influence of climate change, landforms, geographical locations, human activities and other environmental factors, the characteristics of runoff series are often non-linear, non-stationary, complex, uncertain, multi-scale, etc. At present, there is no perfect model to accurately describe its evolution process. Therefore, careful analysis of runoff sequence is needed to further improve the accuracy of the runoff prediction model (Zhao & Chen 2015). Traditional methods of the runoff prediction include the empirical correlation, mathematical statistics, probability theory, genetic analysis methods and so on. Nowadays, with the continuous progress of computer technology and algorithms, many new prediction methods are gradually adopted. For example, artificial neural network, support vector machine (SVM), relevance vector machine, deep recursive neural network, long short-term memory, least squares SVM, Elman neural network, data-augmented neural network model, extreme learning machine, graph neural network, auto-regressive integrated moving average model and so on (Okkan & Serbes 2012; Jajarmizadeh *et al.* 2014; Wang *et al.* 2015b; Niu *et al.* 2018; Yuan *et al.* 2018; Büyükşahin & Ertekin 2019; Li *et al.* 2019; Ruiming 2019; Bi *et al.* 2020; Zhang *et al.* 2021; Liu *et al.* 2022). In addition, there are examples of new hybrid prediction methods, such as the ML model developed by combining SVM, Artificial Neural Network (ANN) and Long Short-Term Memory (LSTM) (Essam *et al.* 2022b). So far, SVM is widely used. This method has been proven to have good performance in regression and time series prediction and can better solve practical problems such as small samples, non-linearity and local minima, it is also considered to be able to replace the auto-regressive and moving average models (Thissen *et al.* 2003). In the application of the model, the parameter setting is particularly important and different parameter values will bring different prediction results. Therefore, selecting the correct algorithm and parameter tuning process for machine learning problems is crucial to achieve the expected results. In previous studies, some scholars have used Bayesian and forest-based algorithms to optimize the parameters of three neural networks (Chong *et al.* 2022b). Of course, SVM parameters also need to be optimized. For example, Xing *et al.* (2016) used the bat algorithm to optimize SVM to seek its optimal learning parameters so as to obtain runoff prediction results, improve the effect of runoff prediction and meet the accuracy requirements of runoff prediction.

The characteristics of the runoff process, such as non-linearity, non-stationary, complexity, uncertainty and multi-scale, render it difficult to predict runoff. The runoff series can be decomposed into a series of relatively stable components by using decomposition technology. It was proven that decomposition technology can effectively improve the accuracy of runoff prediction. Therefore, many scholars combined decomposition technology with a runoff prediction model to form the 'decomposition-prediction-reconstruction' method (Wang *et al.* 2013, 2015a). The basic idea is as follows: firstly, the runoff series is divided into several stable components by decomposition technology. The model is then used to predict each decomposed component. Finally, the prediction results of each component are combined to obtain the prediction results of the whole runoff series (Ji *et al.* 2021). For example, He *et al.* (2020) adopted the method of coupling variational mode decomposition with a gradient propulsive regression model. First, the Variational mode decomposition (VMD) was used to decompose the original monthly runoff sequence into several sub-sequences. Then, the optimal number of input variables was selected and the Gradient Boosting Regression Tree (GBRT) model was used for prediction. Finally, the prediction results of each sub-sequence were aggregated to obtain the ensemble prediction results. It was trained and tested with 50 years of runoff data in the Weihe River Basin and results showed that it obtained better and more accurate runoff prediction results. Zhao *et al.* (2017) established a prediction model coupled with empirical mode decomposition and chaotic least squares SVM to predict the annual runoff of four hydrological stations in the upper reaches of the Fenhe River Basin. First, the original

annual runoff sequence was decomposed into a finite number of intrinsic mode functions (IMFs) and a trend term to make the sequence stationary. Then, if the IMF component had chaotic characteristics, Least Squares Support Vector Machine (LSSVM) was used for prediction; if the IMF component did not have chaotic characteristics, a polynomial method was used for simulation. In addition, the gray model was used to predict the retention trend item. Finally, by combining the prediction results of IMFs and trend terms, runoff prediction results with less error and higher precision were obtained. Niu *et al.* (2019) analyzed the changing trend of runoff at Three Gorges Hydrological Station during the past 50 years. The selected method not only combined Ensemble Empirical Mode Decomposition (EEMD) decomposition and the Extreme Learning Machine (ELM) model but also used an improved gravitational search algorithm (IGSA) to optimize the extreme learning machine. The method first used EEMD to decompose the original runoff data into a finite number of sub-sequences and residuals. Then, the ELM model was used to predict the sub-sequence, residuals and IGSA (based on an elite guided evolution strategy, selection operator and mutation operator) was used to optimize the parameters of the ELM model. Finally, all forecasting results were summarized to get the final forecast results, which proved further improvement in the precision of runoff prediction. It was known that in the prediction of runoff, a hybrid model in general outperformed the single prediction model (Zhao & Chen 2015; Niu *et al.* 2019; Liu *et al.* 2022; Zhang *et al.* 2022).

This article adopts the model of 'decomposition-predictor-reconstruction'. Different from most studies, this paper has two novel points. The first point is the decomposition part, which combines the improved complete ensemble EMD (ICEEMDAN) and the wavelet decomposition (WD) to form the second decomposition. The second point lies in the prediction part. In this part, a seagull optimization algorithm (SOA) is used to optimize SVM, forming a new SOA–SVM prediction model that has never been used before. The effect of the model is verified in the monthly runoff prediction of Hongjiadu Hydropower Station in the Wujiang River Basin. Results show that the hybrid model can effectively improve the accuracy of monthly runoff prediction and offer a novel method of runoff prediction. All methods and technologies adopted in this paper are implemented in the MATLAB software.

The rest of the paper is arranged as follows: Section 2 introduces the basic theories and algorithms of the ICEEMDAN, WD, SOA and SVM and describes the construction of the runoff prediction model. In Section 3, four performance evaluation indexes are outlined. In Section 4, the study area, data are introduced and the results of runoff prediction are described, compared, analyzed and discussed. Finally, Section 5 states the conclusion.

2. METHODOLOGY

2.1. ICEEMDAN

The ICEEMDAN, which was proposed by Colominas *et al.* (2014), is an improved algorithm for complete ensemble empirical mode decomposition with adaptive noise. It mainly solves the problems of residual noise and spurious models. The specific computation steps are as in the following Equations (1)–(7):

(1) Add EMD to the original signal to obtain special noise:

$$x^{(i)} = x + \beta_0 E_1[\omega^{(i)}] \tag{1}$$

where x is the original signal, $E_1[\omega^{(i)}]$ is the special noise, $\omega^{(i)}$ is the ith added white noise ($i = 1, 2, \ldots, N$) and β_0 is the expected signal-to-noise ratio of the first decomposed signal.

(2) The EMD algorithm is used to obtain the local mean value of the reconstructed signal and the first residual is obtained by taking their mean value to compute the IMF value:

$$r_1 = \frac{1}{N} \sum_{i=1}^{N} M[x^{(i)}] \tag{2}$$

$$\tilde{c}_1 = x - r_1 \tag{3}$$

where r_1 is the residual of the first decomposition, $M[\cdot]$ is the operator to compute the local mean and $\tilde{c}_1$ is the value of the first IMF.

(3) Compute the value of the second IMF:

$$r_2 = \frac{1}{N}\sum_{i=1}^{N} M\{r_1 + \beta_1 E_2[\omega^{(i)}]\} \tag{4}$$

$$\tilde{c}_2 = r_1 - r_2 \tag{5}$$

where r_2 is the residual of the second decomposition and $\tilde{c}_2$ is the value of the second IMF.

(4) Continue using the equation according to the above computation steps to find the value of the kth IMF:

$$r_k = \frac{1}{N}\sum_{i=1}^{N} M\{r_{k-1} + \beta_{k-1} E_k[\omega^{(i)}]\}, k = 2, 3, \ldots N \tag{6}$$

$$\tilde{c}_k = r_{k-1} - r_k \tag{7}$$

where r_k is the residual of the kth decomposition and $\tilde{c}_k$ is the value of the kth IMF. The IMF of the original signal can be decomposed accurately through iterative computation of this equation.

2.2. WD quadratic decomposition

Previous studies mentioned that different input variables affected the prediction accuracy of the model (Othman *et al.* 2020). As the effect of high-frequency components in the ICEEMDAN model is not ideal from previous studies, a method of combining the ICEEMDAN and VMD was to compose a quadratic decomposition (Wen *et al.* 2019). Therefore, in this paper, WD is employed to perform secondary decomposition of high-frequency components, further reducing the complexity of high-frequency components and rendering them more stable. WD uses the wavelet transform to decompose a high-frequency component into high and low-frequency signals of different scales and then makes model predictions for each high and low-frequency signal. Finally, the prediction results of different levels are combined into the predicted value of the high-frequency component (Ren *et al.* 2013). A wavelet can be expressed as a waveform with a zero mean of finite length and a discrete point. A wavelet has the characteristics of irregularity, short duration and asymmetry. The basic equation of wavelet transform is the wavelet function. Different wavelets have great differences in the waveform and similar wavelets form a wavelet family. The local property of a wavelet is that the value is not 0, only in a finite interval. This feature can be used well to represent signals with sharp, discontinuous signals (Chong *et al.* 2022a). The equation of wavelet transform is in the following equation:

$$\partial = W^T f \tag{8}$$

where ∂ represents the wavelet coefficient obtained by the transformation, W is the orthogonal matrix and f is the input signal.

A specific wavelet function consists of a set of specific wavelet filtering coefficients. When the wavelet function is selected, the corresponding wavelet filter coefficients are known. In this paper, the db4 wavelet basis function is used to decompose high-frequency components decomposed by the ICEEMDAN and get the trend sequence.

2.3. SOA

SOA is a swarm intelligence optimization algorithm proposed by Dhiman & Kumar (2019). The algorithm is mainly inspired by the migration and aggressive behaviors (foraging behaviors) of seagulls in nature and the algorithm has been proven to be able to solve challenging large-scale constraint problems. Compared with other optimization algorithms, it has strong competitiveness owing to its optimization capabilities and simple computation. SOA accepts real-number code and is thus directly identified through the solution. The count of dimensions is equivalent to the stability of the solution (Lavanya *et al.* 2022).

The mathematical model of SOA is as follows:

(1) Migration behavior

In the migration behavior, the algorithm simulates the seagull individual exploring from one location to another. At this stage, there are three things to watch out for: avoiding collisions, movement toward the best neighbor's direction and remaining close to the best search agent.

In the migration process, each seagull's position is different. In order to avoid collisions between neighbors (surrounding seagulls), an additional variable A is adopted to compute the new seagull position:

$$C_s(x) = A \times P_s(x) \tag{9}$$

where $C_s(x)$ is the new position of seagulls (no collision with other seagulls), $P_s(x)$ is the current position of seagulls, x is the current iteration number and A is the migration behavior of seagulls in the given search space. The value of A is determined
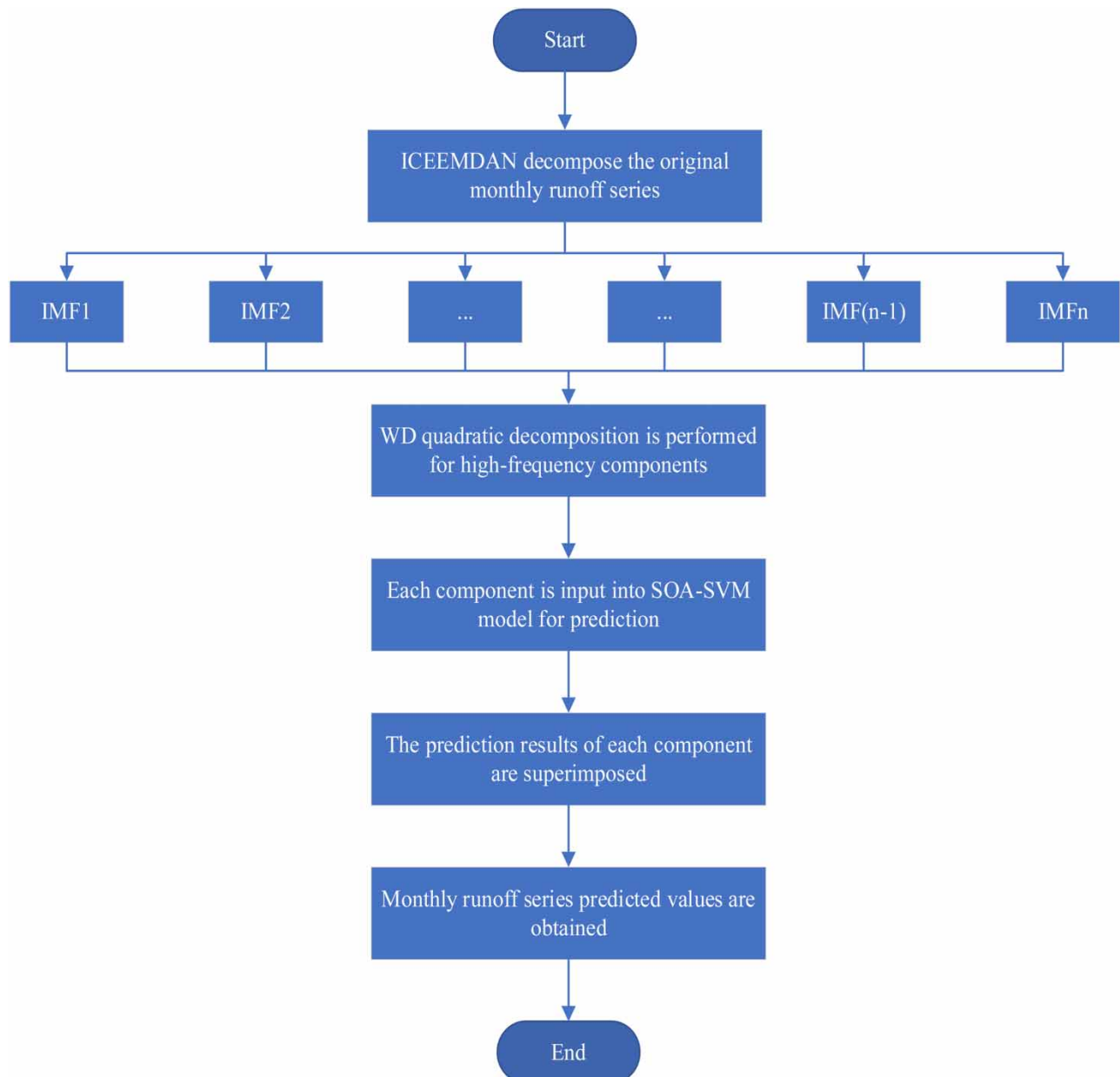
Figure 1 | Framework of monthly runoff prediction model.

by the following equation:

$$A = f_c - \left(x \times \left(\frac{f_c}{MAX_{iteration}} \right) \right) \qquad (10)$$

where f_c controls the frequency of variable A and the value of A decreases linearly from 2 to 0.

After having ensured that no individual gulls collide with each other, all gulls are brought close to the best position:

$$M_s(x) = B \times (P_{best}(x) - P_s(x)) \qquad (11)$$

where $M_s(x)$ is the convergence direction of the individual to the optimal seagull, B is the random number balancing global and local search with its expression as in the following equation:

$$B = 2 \times A^2 \times r_d \qquad (12)$$

where r_d is a random number in the range [0,1].

After having computed the convergence direction of each individual seagull, the seagull starts to move toward a new position, described in the following equation:

$$D_s(x) = |C_s(x) + M_s(x)| \qquad (13)$$

where $D_s(x)$ is the new position of the seagull.

(2) Aggressive behavior

During migration, seagulls can constantly change angle and speed to attack, maintain height through wings and weight and display spiraling aggressive behavior when foraging. The motion behaviors in three-dimensional spaces x, y and z are
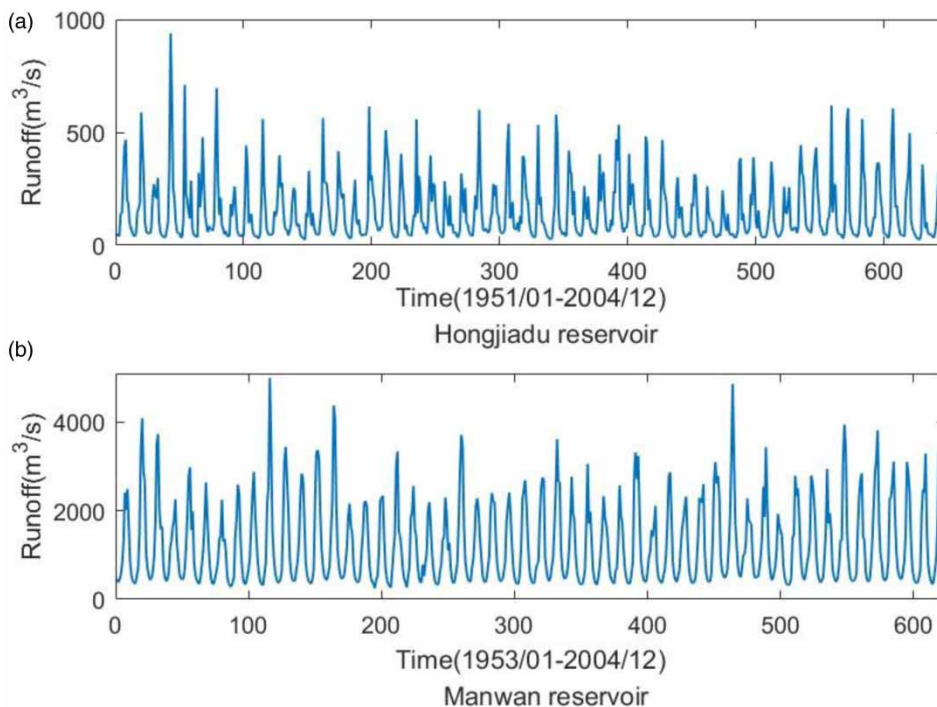

Figure 2 | Original monthly runoff series.

described as in the following Equations (14)–(17).

$$x = r \times \cos(\theta) \tag{14}$$

$$y = r \times \sin(\theta) \tag{15}$$

$$z = r \times \theta \tag{16}$$

$$r = u \times e^{\theta v} \tag{17}$$

where r is the spiral radius, θ is the random number within the range $[0, 2\pi]$, u and v are the correlation constants of the spiral shape (usually 1) and e is the base of the natural logarithm.

Combined with the new position of the seagull, the overall seagull position update equation is as in the following equation:

$$P_s(x) = D_s(x) \times x \times y \times z + P_{best}(x) \tag{18}$$

where $P_s(x)$ is the attack position of seagulls (the optimal solution of search space).

The application process of the seagull algorithm is as follows:

Step 1: Initialize parameters.

Step 2: Compute the fitness value of each seagull;

Step 3: Compute the new position $D_s(x)$ of seagulls according to the equation in migration behavior;

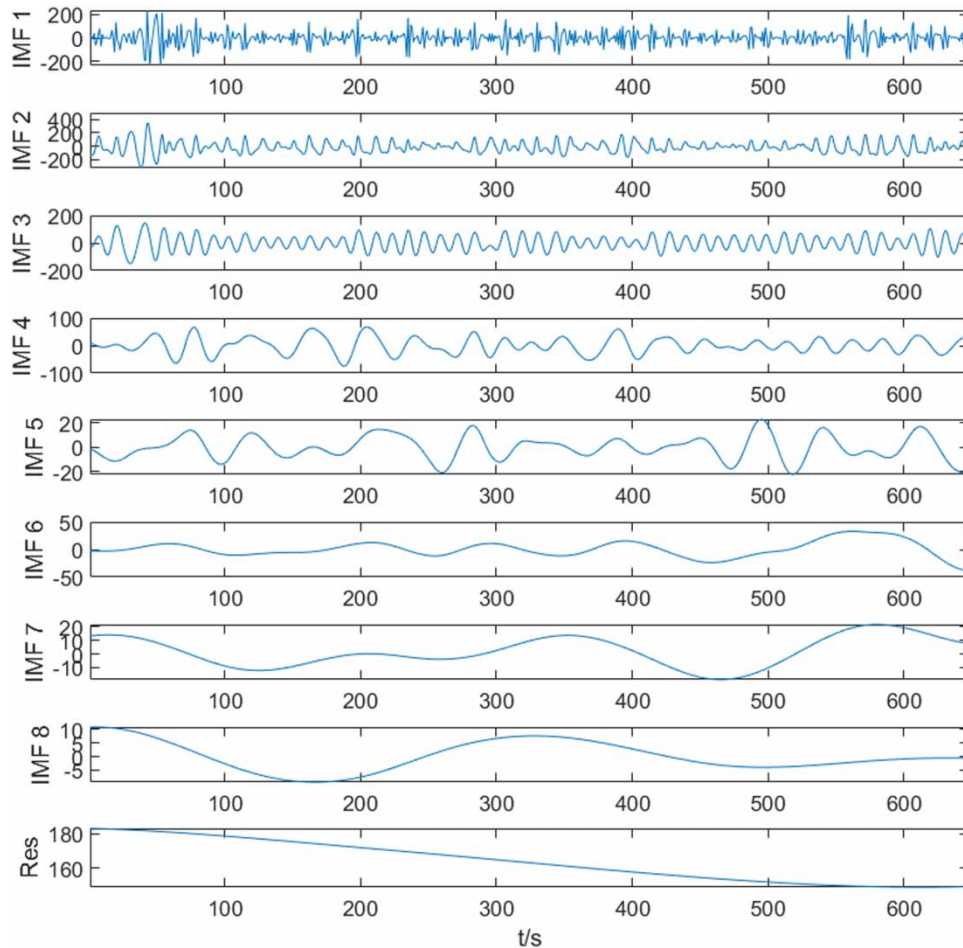Step 4: Compute the attack position $P_s(x)$ of seagulls according to the equation in attack behavior;

Figure 3 | Monthly runoff series of the Hongjiadu Reservoir decomposed by the ICEEMDAN.

Step 5: Update the location information and adaptive value of the best seagull (*iteration* = *iteration* + 1).

Step 6: If *iteration* > *MAX*$_{iteration}$, go to Step 7; otherwise, go to Step 3.

Step 7: Output the best position and adaptive value of seagulls.

2.4. SVM

SVM is a learning machine proposed by Corinna & Vladimir (1995) to solve two-group of classification problems. SVM represents a kernel-based algorithm. The advantages of SVM are that it has a small tendency to overfit, has good generalization ability, can simultaneously minimize estimation errors and can linearly separate inputs in mapped high-dimensional feature spaces. At the same time, it also shows its unique advantages in solving small sample, non-linear and high-dimensional pattern recognition. The SVM algorithm is very powerful. It not only supports linear and non-linear classification but also linear and non-linear regression. For runoff prediction, this paper adopts the regression prediction method, namely support vector regression (SVR). SVR typically estimates learning data through the definition of functions and determines linear separation functions to achieve realistic results that reflect their statistical learning theory (Essam *et al.* 2022a). SVM has been applied in different engineering fields. Considering the rationality and complexity of training samples comprehensively, data sets are given based on the principle of risk minimization (Roushangar *et al.* 2021). $(x_i, y_i), x_i \in R_n, i = 1, 2, \ldots, m$ denote the given data sets, x_i represent input vectors and y_i represent corresponding inputs and the decision function of SVM is in the following Equation (19):

$$f(x) = \sum_{i=1}^{m} (a_i - a_i^*)K(x_i, x) + b \tag{19}$$
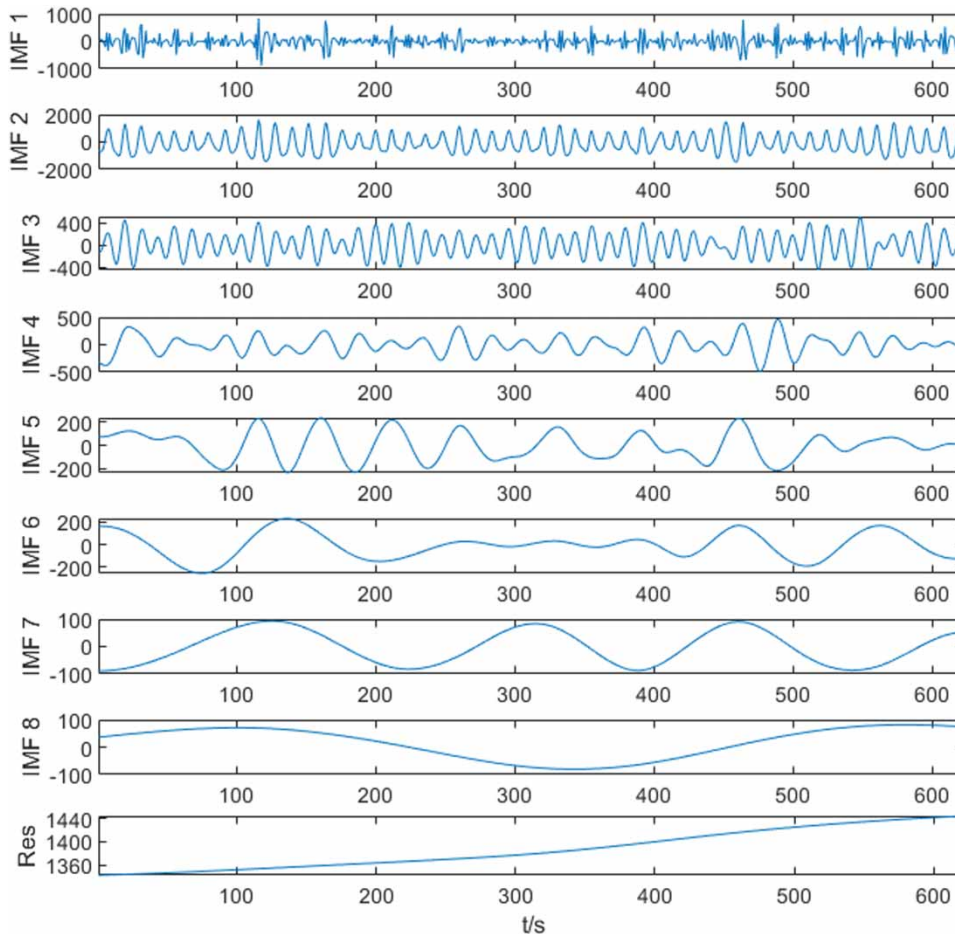

Figure 4 | Monthly runoff series of the Manwan Reservoir decomposed by the ICEEMDAN.

where samples (x_i, y_i) corresponding to the coefficient $(a_i - a_i^*) \neq 0$ are the support vectors, a_i is the Lagrange factor, b is the threshold determined by training samples, $K(x_i, x)$ is the kernel function that meets Mercer's conditions and $K(x_i, x) = \phi(x)^T \phi(x_i)$.

Cheng et al. (2017) mentioned that the linear kernel, polynomial kernel and radial basis function are the three most widely used principal kernel functions at present. Moreover, the linear kernel is mainly used in the case of linear separability. It can be seen that the dimension from the feature space to the input space is the same, its parameters are few and the speed is fast and the classification effect is ideal for linear separable data. However, the series data of monthly runoff are obviously complex, but they are not linearly separable data and the effect of a linear kernel is obviously not good for use here. A polynomial kernel function can realize mapping input space of low dimension to feature space of high latitude. However, the polynomial kernel function has many parameters. When the order of the polynomial is relatively high, the element value of the kernel matrix will tend to infinity or the infinitesimal and the computational complexity will be too large to compute, so it is not used. Both linear kernel function and polynomial kernel function belong to the global kernel function. Their generalization performance is good, but their learning ability is weak and they are affected by the distance between samples. However, radial basis functions belong to local kernel functions, which have strong non-linear mapping abilities. This kernel function is the most widely used one. It has relatively good performance in both large and small samples and it has fewer parameters compared with the polynomial kernel function. Therefore, in most cases when the kernel function is unknown, the Gaussian radial basis kernel function is generally preferred and most previous studies have also used the radial basis function (Roushangar & Shahnazi 2021; Roushangar et al. 2022). Therefore, the radial basis function is adopted in the article and its expression is as in the following equation:

$$K(x_i, x) = \exp\left(-g\|x_i - x\|^2\right) \tag{20}$$

where g is the kernel parameter and $g > 0$.

2.5. Construction of a monthly runoff prediction model

The basic framework of the runoff prediction model proposed in this article is shown in Figure 1. The detailed steps of the model are as follows:

Step 1: Decomposition: The original monthly runoff time series is decomposed into several IMFs and one residual (Res) by the ICEEMDAN method.
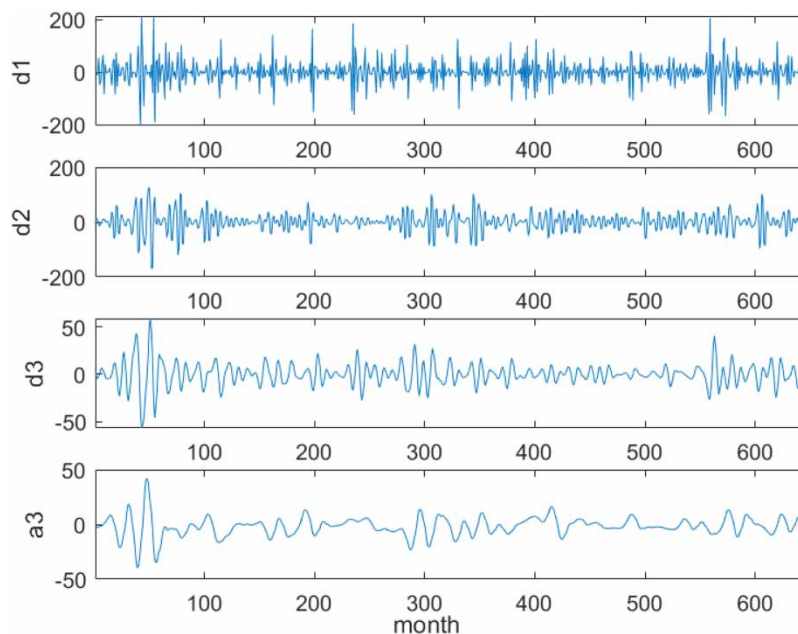

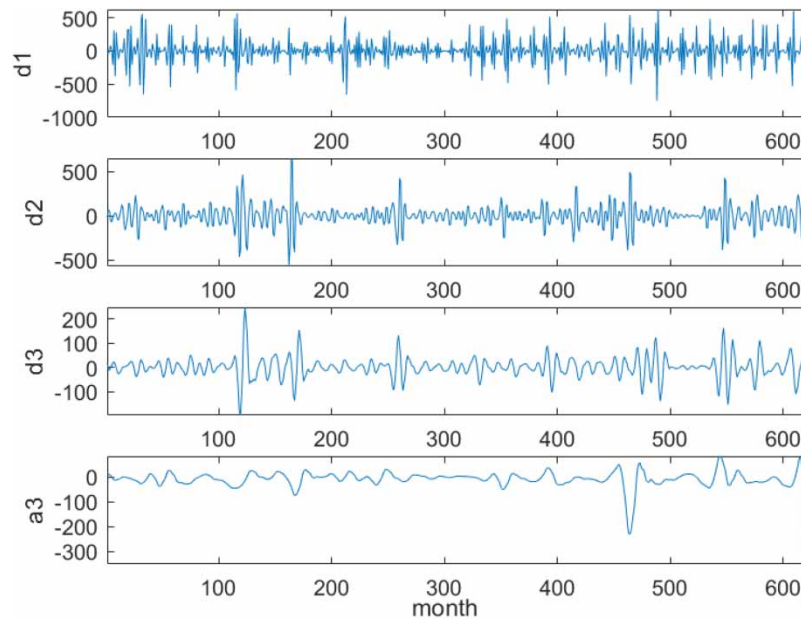**Figure 5** | WD results of IMF1 at the Hongjiadu Reservoir.

Figure 6 | WD results of IMF1 at the Manwan Reservoir.

Step 2: Quadratic decomposition: WD method is adopted to carry out WD of IMF high-frequency components in Step 1 to obtain a more stable trend sequence.

Step 3: Data preprocessing: All monthly runoff data are divided into training and test sets and both of them are normalized to the interval [0,1]

$$x'_i = \frac{x_i - \min\limits_{1 \le i \le n} \{x_i\}}{\max\limits_{1 \le i \le n} \{x_i\} - \min\limits_{1 \le i \le n} \{x_i\}} \tag{21}$$

where n is the number of samples, x_i and x'_i are the ith sample of the original and normalized data sequences, respectively.

Step 4: Model optimization: This paper uses SOA to optimize SVM and the unprecedented SOA–SVM prediction model is finally established.

Table 1 | Optimal parameters of the Hongjiadu Reservoir

model			Hongjiadu		
			c	g	p
SOA–SVM			135.674	4.53286	0.107157
ICEEMDAN–WD–SOA–SVM	IMF1	d1	1000	21.182	1
		d2	73.1989	4.70714	0.000188029
		d3	95.0232	2.89931	0.0462714
		a3	13.769	3.24761	0.000507261
	IMF2		901.85	2.0286	0.901826
	IMF3		136.066	8.86452	0.0561698
	IMF4		184.905	4.65206	0.273467
	IMF5		41.1947	10.1849	0.0330585
	IMF6		969.004	14.9353	0.00242618
	IMF7		78.1709	31.9604	0.0041999
	IMF8		44.3877	5.03156	0.00629108
	R		1,000	32.8223	0.000112655

Table 2 | Optimal parameters of the Manwan Reservoir

model			Manwan		
			c	g	p
SOA–SVM			1,000	1.60807	0.503161
ICEEMDAN–WD–SOA–SVM	IMF1	d1	1,000	2.09022	0.0174761
		d2	302.244	5.18232	0.0224248
		d3	17.7589	3.92211	0.0507203
		a3	109.443	6.87224	0.0733407
	IMF2		1,000	1.40956	0.00877693
	IMF3		200.185	4.07874	0.0436708
	IMF4		989.414	2.59465	0.0922121
	IMF5		157.769	11.2888	0.138869
	IMF6		920.794	11.3256	0.0880676
	IMF7		1000	2.76988	0.0205433
	IMF8		208.315	47.2232	0.0134492
	R		939.963	24.3069	0.000144657

Step 5: Model application: Each component decomposed by the ICEEMDAN and WD is input into the SOA–SVM model for training and the result of each component is predicted, respectively. Finally, the prediction results of each sub-sequence are superimposed to obtain the final monthly runoff prediction value.

3. EVALUATION INDICATORS

Four evaluation indicators, namely RMSE, MAPE, NSEC and R, are used to test the prediction ability of different prediction models. RMSE can effectively measure the prediction performance of the model when tracking high traffic. MAPE can better reflect the actual situation of the predicted value error. NSEC is the deviation between the predicted value and its average
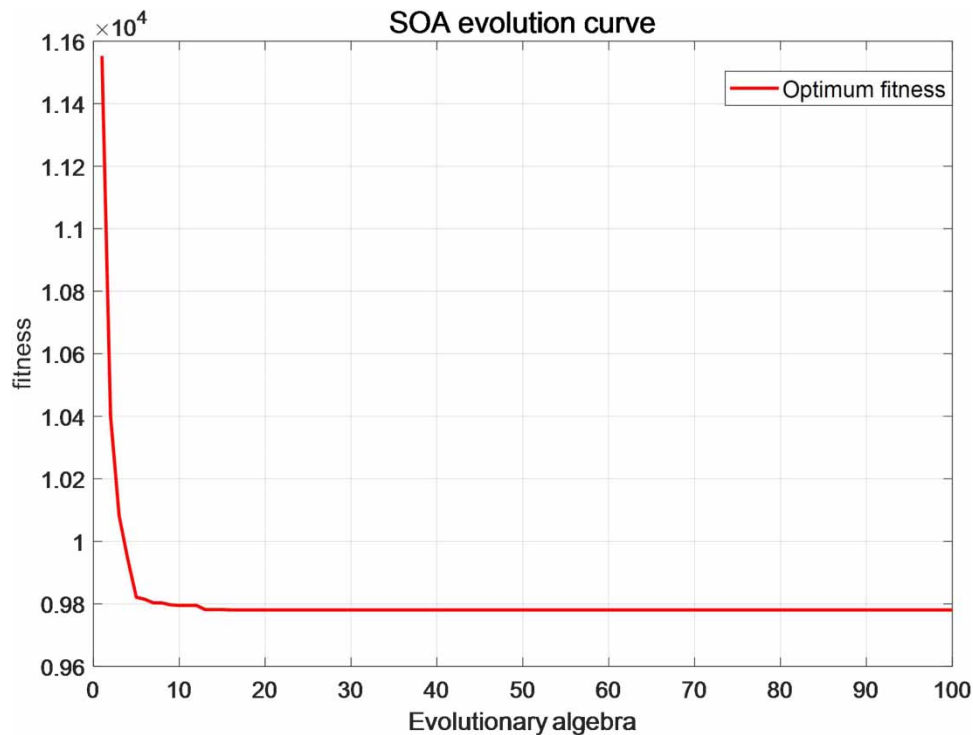

Figure 7 | Fitness function optimization process diagram of original monthly runoff series for the SOA–SVM model in the Hongjiadu Reservoir.
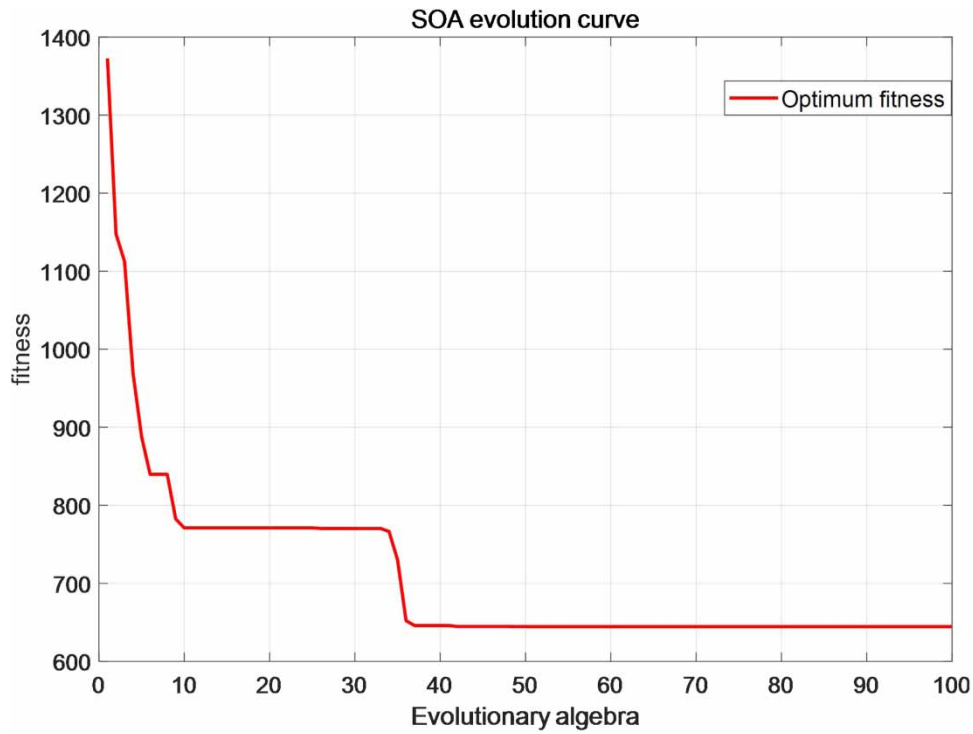
Figure 8 | Flow chart of fitness function optimization of the first component in the SOA–SVM model of the Hongjiadu Reservoir.
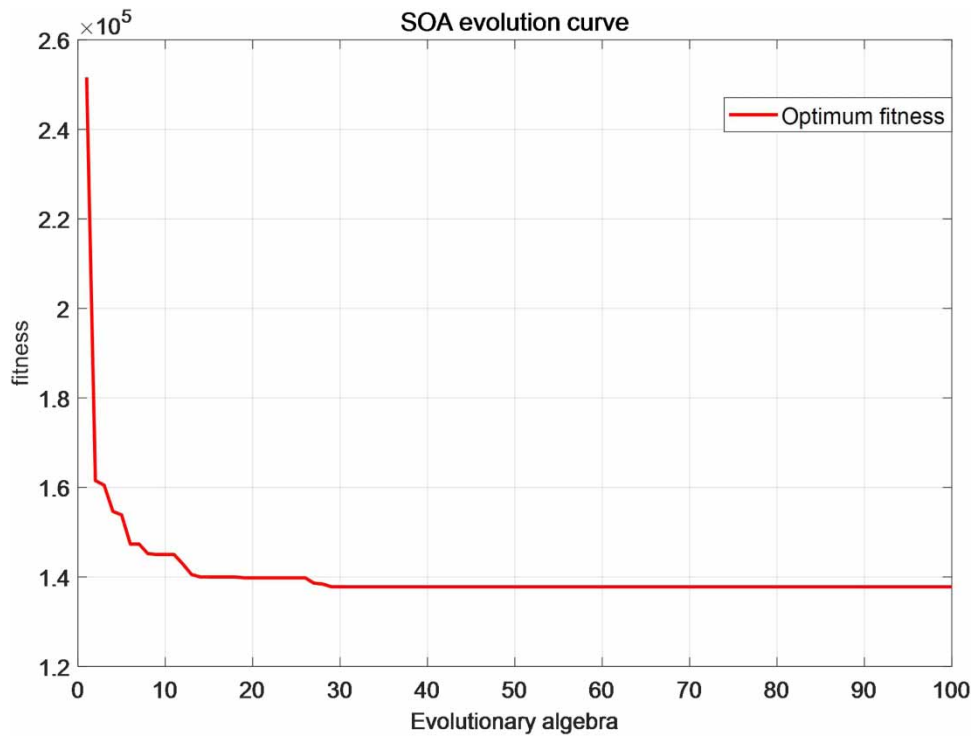

Figure 9 | Fitness function optimization process diagram of original monthly runoff series for the SOA–SVM model in the Manwan Reservoir.
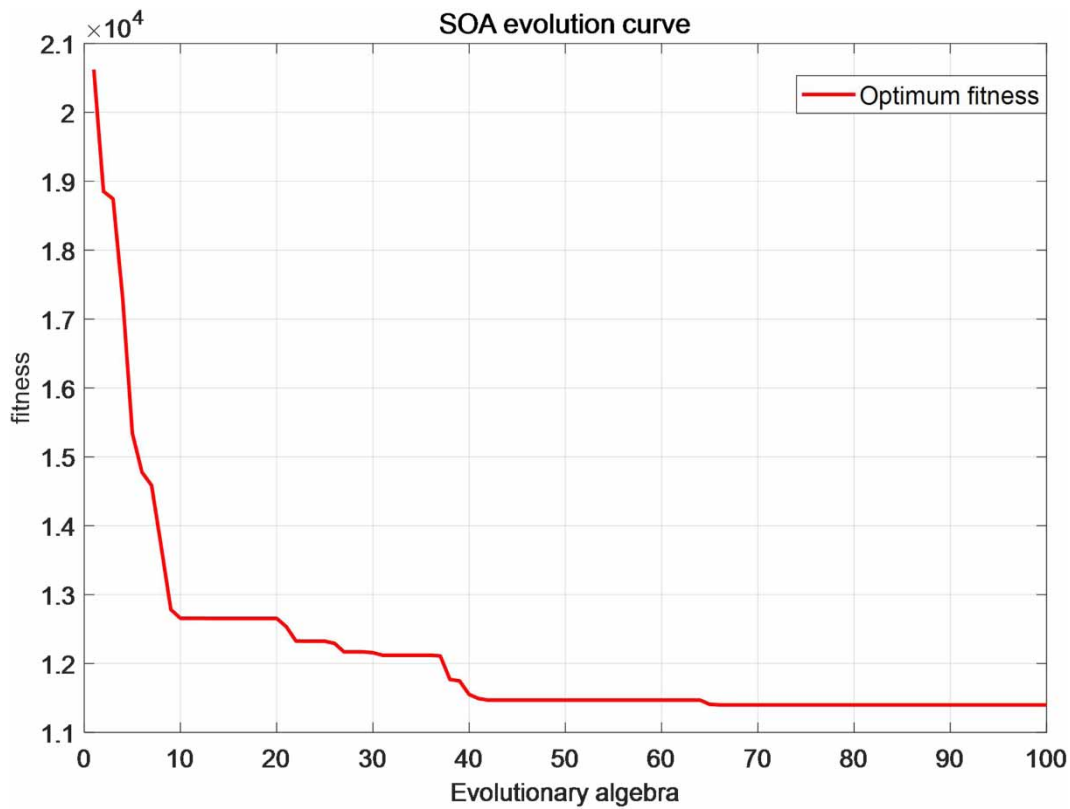
Figure 10 | Flow chart of fitness function optimization of the first component in the SOA–SVM model of the Manwan Reservoir.
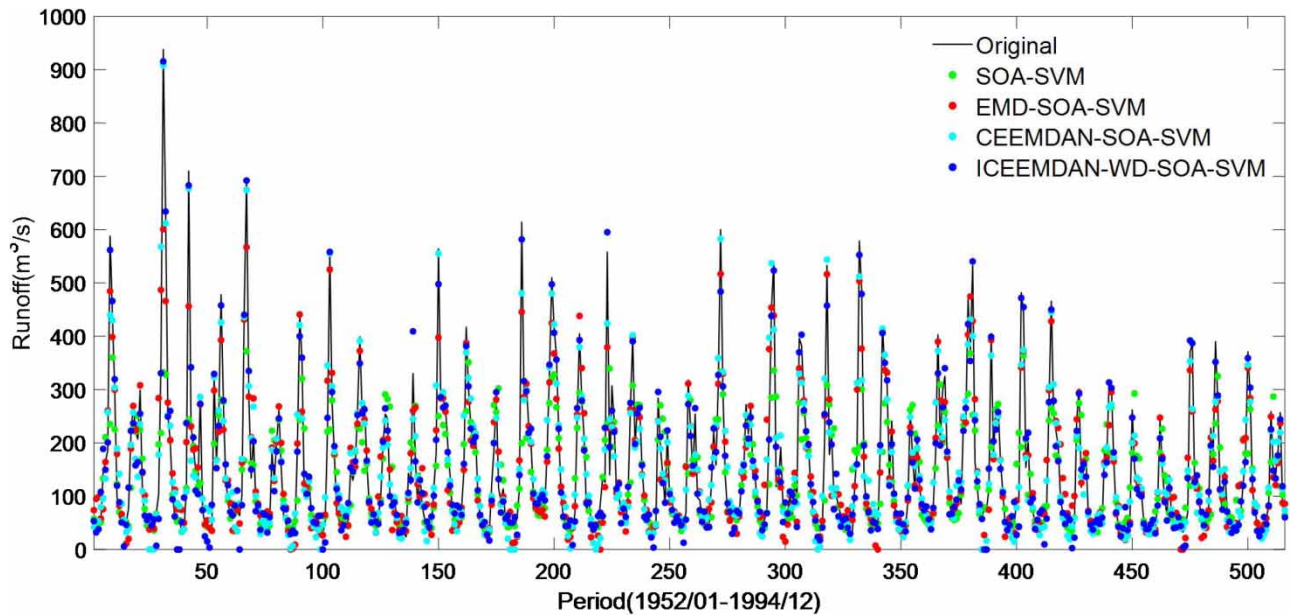

Figure 11 | Forecasting results of four models for the Hongjiadu Reservoir during the training period.

value, which is used to verify hydrological model simulation results. R reflects the linear correlation between the observed data and predicted results. Generally speaking, the better the effect of the runoff prediction model is, the smaller the value of the first two indexes (RMSE and MAPE) are and the larger the value of the last two indexes (NSEC and R) are (with a
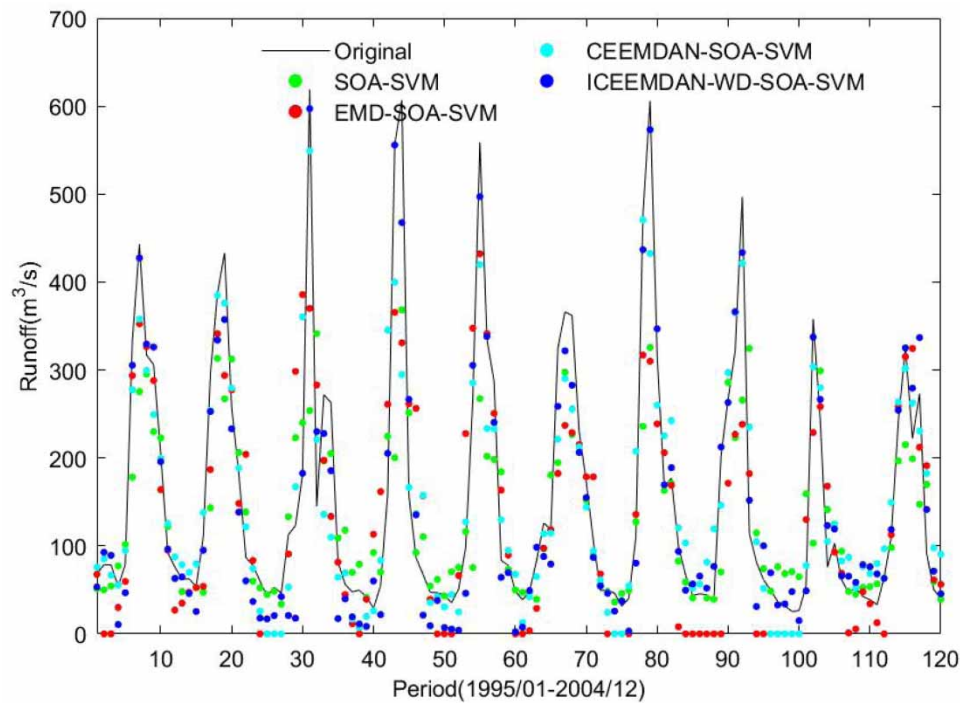
Figure 12 | Forecasting results of four models for the Hongjiadu Reservoir during the testing period.

maximum value of 1) (Wang *et al.* 2009; Adnan *et al.* 2018; Feng *et al.* 2020). The four indicators are defined as in the following equations:

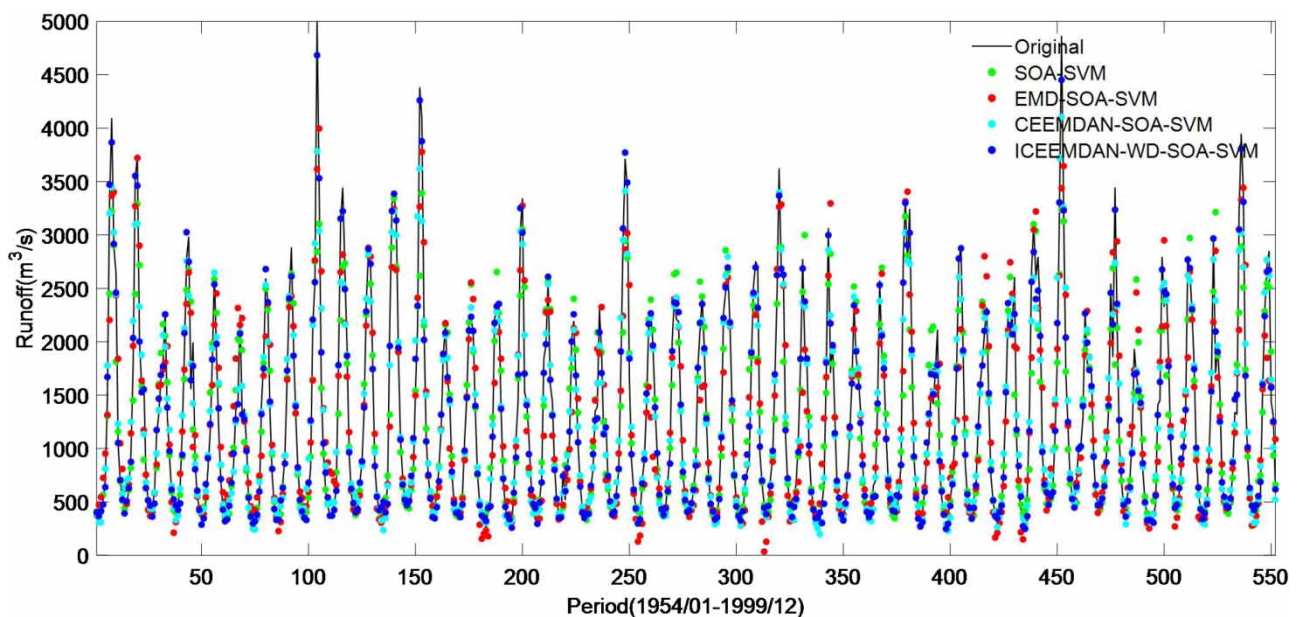$$\text{RMSE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2} \tag{22}$$

Figure 13 | Forecasting results of four models for the Manwan Reservoir during the training period.
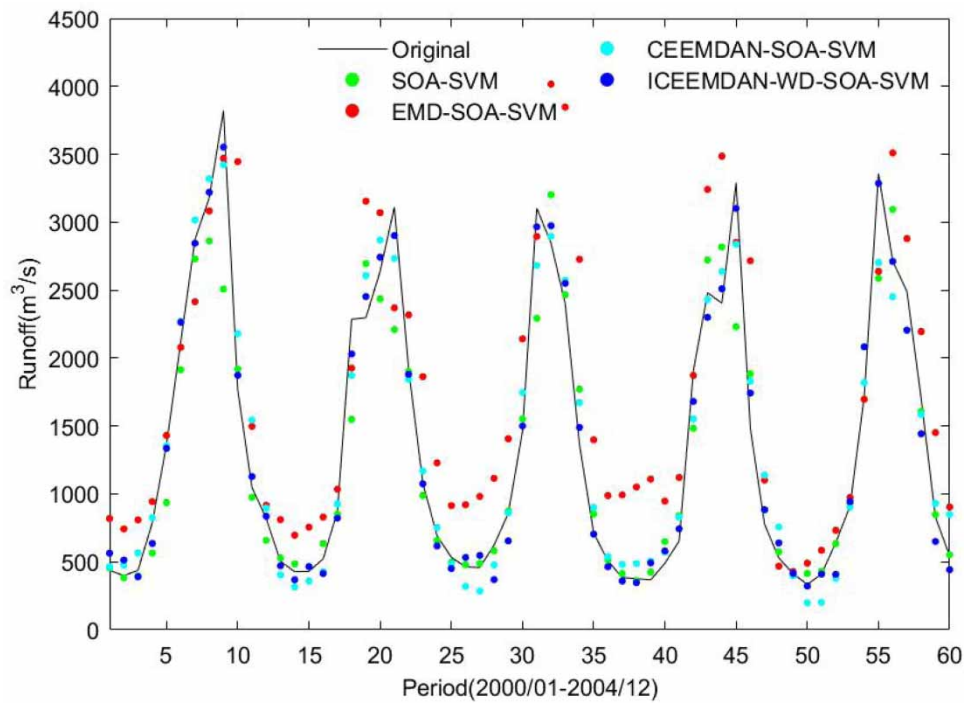
Figure 14 | Forecasting results of four models for the Manwan Reservoir during the testing period.

Table 3 | Comparison of prediction errors of different models for the Hongjiadu Reservoir

	Training				Testing			
Model	**RMSE (m³/s)**	**MAPE**	**NSEC**	**R**	**RMSE (m³/s)**	**MAPE**	**NSEC**	**R**
ARMA (Wang *et al.* 2009)	91.56	46.62	0.521	0.727	94.34	48.03	0.584	0.786
ANN (Wang *et al.* 2009)	91.16	46.25	0.526	0.725	91.07	46.15	0.612	0.786
SVM (Wang *et al.* 2009)	89.89	28.25	0.539	0.753	87.57	33.77	0.641	0.823
SOA–SVM	84.30	24.07	0.5943	0.7920	97.39	39.83	0.5564	0.7822
EMD–SOA–SVM	46.99	28.23	0.8739	0.9366	101.24	93.06	0.5206	0.7869
CEEMDAN–SOA–SVM	41.96	28.66	0.8995	0.9484	65.17	45.67	0.8013	0.8980
ICEEMDAN–WD–SOA–SVM	26.09	21.38	0.9612	0.9804	39.32	33.28	0.9277	0.9659

Table 4 | Comparison of prediction errors of different models for the Manwan Reservoir

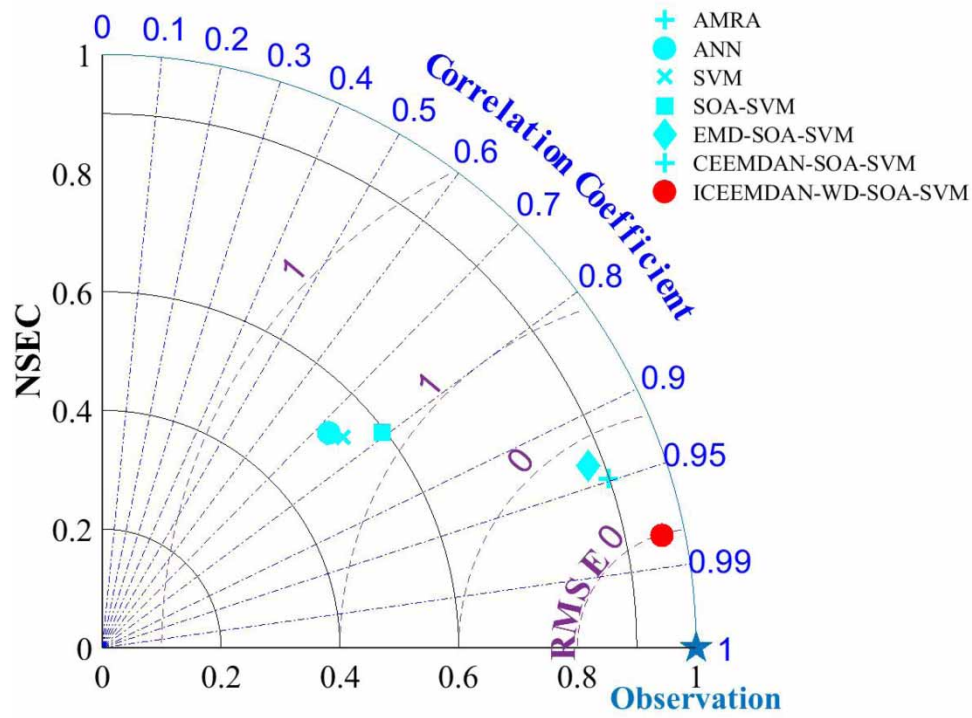|	Training				Testing			
Model	**RMSE (m³/s)**	**MAPE**	**NSEC**	**R**	**RMSE (m³/s)**	**MAPE**	**NSEC**	**R**
ARMA (Wang *et al.* 2009)	354.27	16.77	0.849	0.922	354.53	15.63	0.869	0.928
ANN (Wang *et al.* 2009)	346.31	16.16	0.856	0.925	345.37	14.01	0.867	0.9320
SVM (Wang *et al.* 2009)	334.07	12.49	0.866	0.9315	332.86	12.49	0.8836	0.9410
SOA–SVM	342.76	13.35	0.8585	0.9280	350.07	12.68	0.8783	0.9422
EMD–SOA–SVM	378.18	21.54	0.8278	0.9103	606.47	52.66	0.6348	0.8856
CEEMDAN–SOA–SVM	228.15	13.55	0.9373	0.9691	234.21	17.97	0.9455	0.9727
ICEEMDAN–WD–SOA–SVM	110.17	7.52	0.9854	0.9927	140.26	11.11	0.9805	0.9903

Figure 15 | Taylor diagram of seven models for the Hongjiadu Reservoir during the training period.
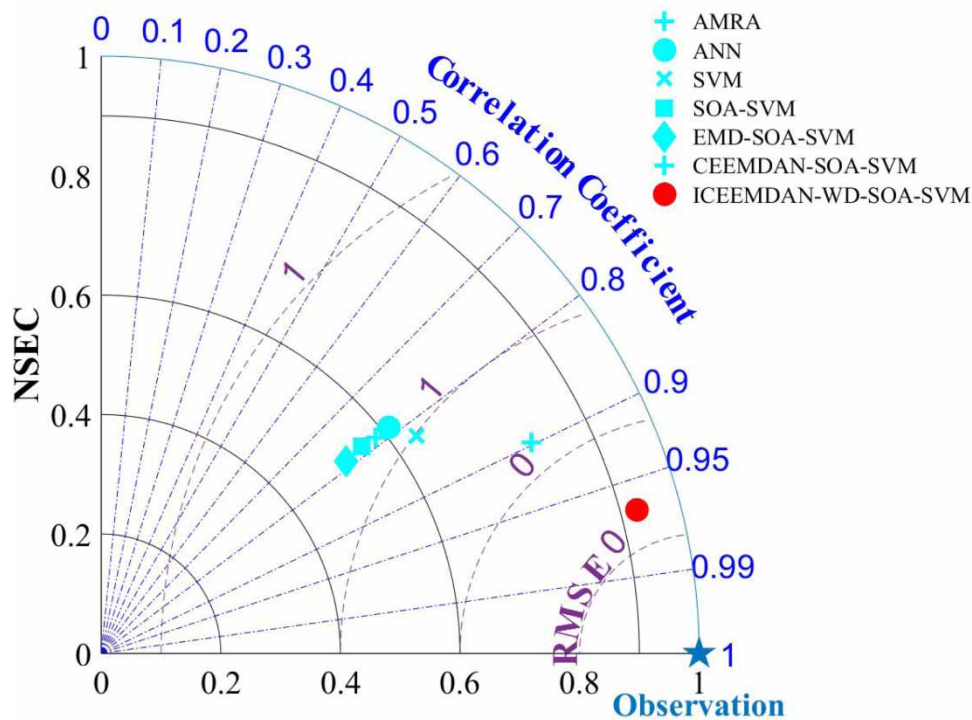

Figure 16 | Taylor diagram of seven models for the Hongjiadu Reservoir during the testing period.
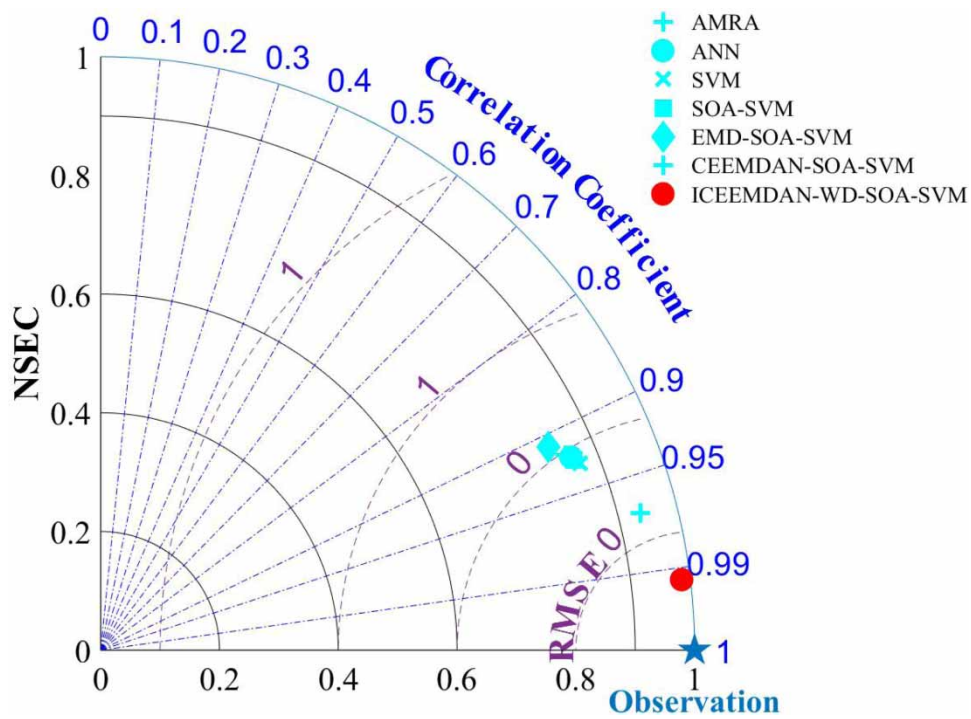
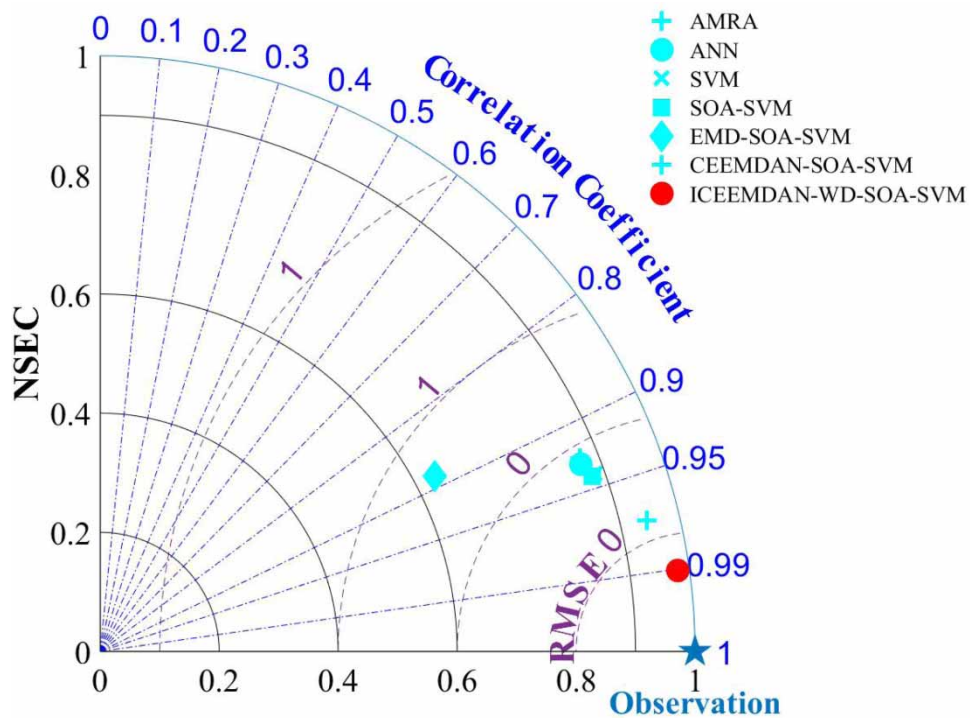**Figure 17** | Taylor diagram of seven models for the Manwan Reservoir during the training period.

Figure 18 | Taylor diagram of seven models for the Manwan Reservoir during the testing period.
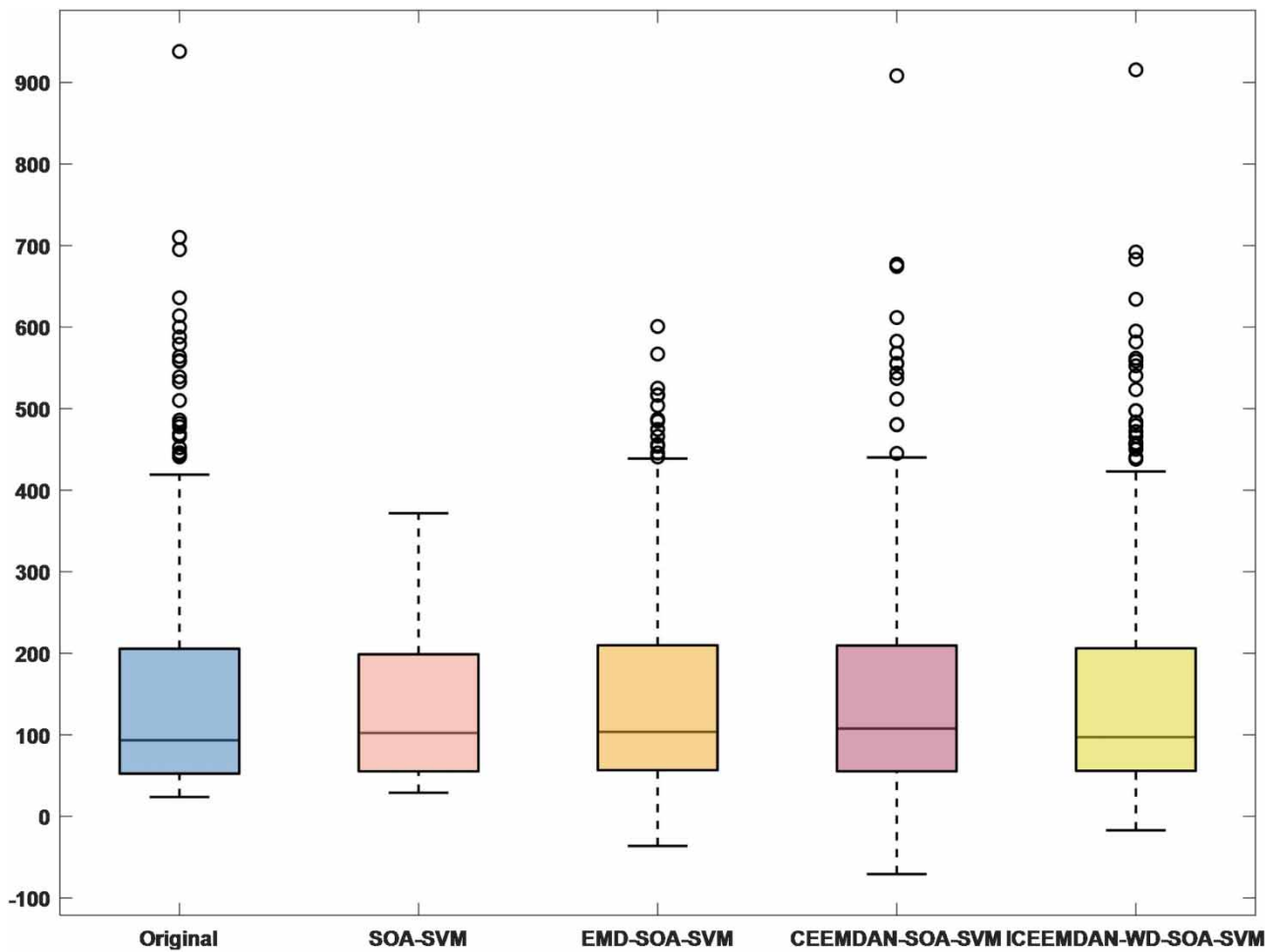
Figure 19 | Boxplot of four models for the Hongjiadu Reservoir during the training period.

$$\text{MAPE} = \frac{100}{n} \sum_{i=1}^{n} \left| \frac{y_i - \hat{y}_i}{y_i} \right| \tag{23}$$

$$\text{NSEC} = 1 - \frac{\sum\limits_{i=1}^{n} (y_i - \hat{y}_i)^2}{\sum\limits_{i=1}^{n} (y_i - y_{\text{arg}})^2} \tag{24}$$

$$R = \frac{\sum\limits_{i=1}^{n} [(y_i - y_{\text{arg}})(\hat{y}_i - \hat{y}_{\text{arg}})]}{\sqrt{\sum\limits_{i=1}^{n} (y_i - y_{\text{arg}})^2 (\hat{y}_i - \hat{y}_{\text{arg}})^2}} \tag{25}$$

where y_i is the observed value of the ith sample, $\hat{y}_i$ is the predicted value of the ith sample, y_{arg} is the average value of all observed values and $\hat{y}_{\text{arg}}$ is the average value of all predicted values.

In addition, boxplot, violin plot and Taylor diagram are used to analyze the final prediction results of different prediction models. The Taylor diagram can centrally represent the relevant statistical information of multiple models and display the results of three evaluation indicators on one chart (Taylor 2001). The three evaluation indexes used in the Taylor diagram of this paper are RMSE (normalization), NSEC and R. A boxplot is a simple tool for describing statistics, mainly used to

reflect the central location and spread range of one or more sets of continuous quantitative data distribution and to identify outliers in the data (Tareen *et al.* 2019). A violin plot is a combination of boxplot and density plot, which is used to show the distribution state of multiple groups of data and probability density. The boxplot is located inside the violin plot, flanked by the density map of the data, showing multiple details of the data (Tanious & Manolov 2022).

4. CASE STUDIES

4.1. Study area and dataset

Hongjiadu Reservoir in the upper reaches of the Wujiang River in Guizhou Province and Manwan Reservoir in the middle reaches of the Lancang River in Yunnan Province are selected as case studies in this paper. Hongjiadu Reservoir is a multi-year regulating reservoir with a mixed type of mountain canyon and lake. The total reservoir capacity is 4.947 billion m^3, the regulated storage capacity is 3.361 billion m^3, the total installed capacity is 0.6 million kW, the basin area is about 9,900 km^2, the annual average flow is 155 m^3/s and the annual average runoff is 4.89 billion m^3. The Hongjiadu Hydropower Station is the only power station in the Wujiang River that has the capacity to regulate water quantity for many years. It is mainly used for power generation and also has comprehensive functions of flood control, water supply, breeding, navigation, tourism and ecological protection. The dam site of Manwan Reservoir is located in a narrow valley, which is a seasonal regulating reservoir. The total reservoir capacity is 920 million m^3, the regulated storage capacity is 258 million m^3, the total installed capacity is 1.5 million kW, the basin area is about 114,500 km^2, the annual average flow is 1,230 m^3/s and the annual average runoff is 38.8 billion m^3. Manwan Hydropower Station is the first phase of the development of the mainstream of the Lancang
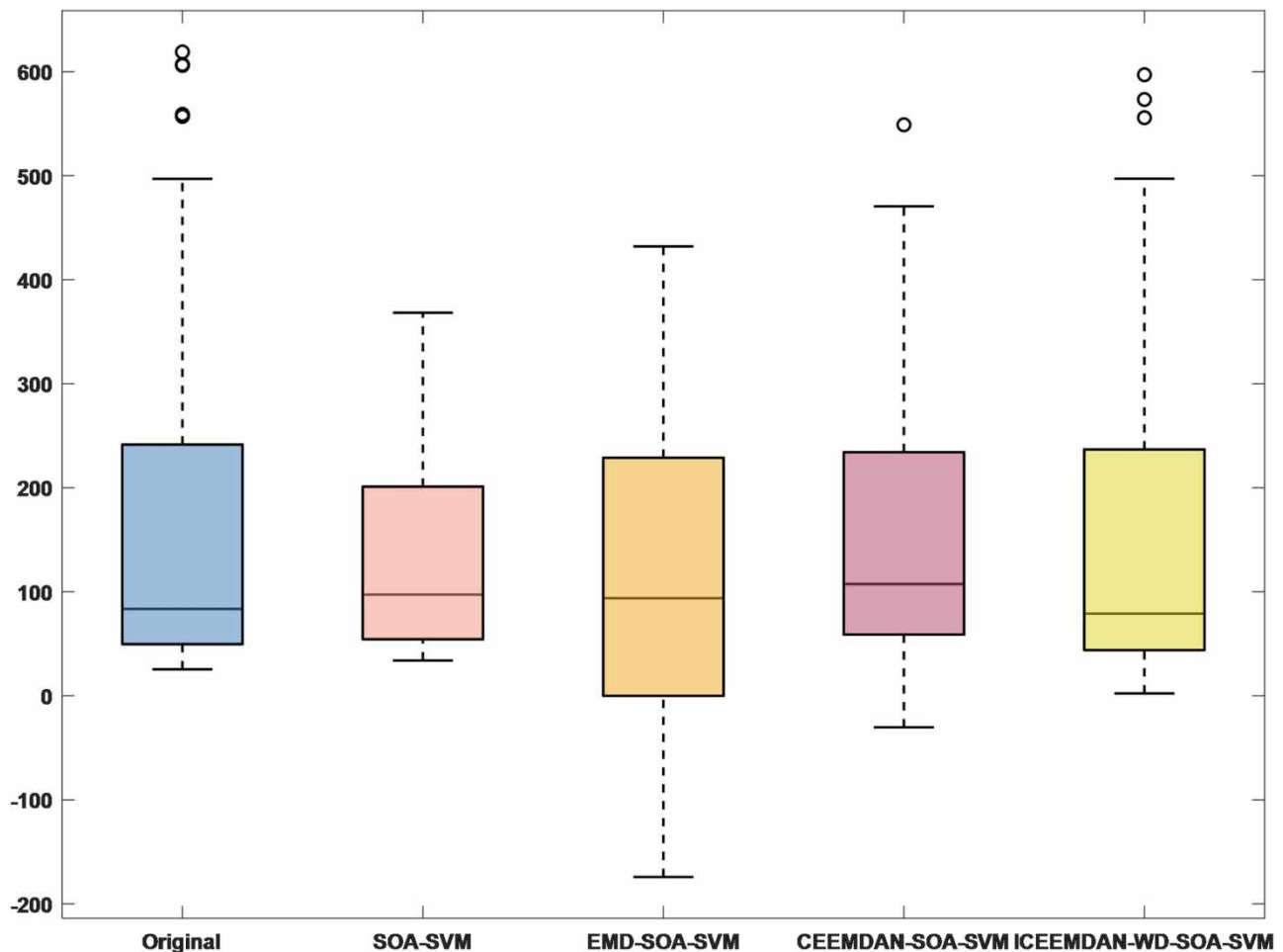

Figure 20 | Boxplot of four models for the Hongjiadu Reservoir during the testing period.

River, the completion of which plays a vital role in the economic development of Yunnan Province. In recent years, under the influence of climate change and human activities, natural disasters occur more and more frequently, so it is of great practical significance to establish an accurate monthly runoff prediction model.

This paper selects the same data as Wang *et al.* (2009), namely the monthly runoff data of the Hongjiadu Reservoir from 1951 to 2004 (648 data in total) and the Manwan Reservoir from 1953 to 2004 (624 data in total). The time span of the selected runoff series is long enough and the precision of the data is high enough to be representative. The original monthly runoff data series of the Hongjiadu and Manwan reservoirs are shown in Figure 2. Results are compared with those of Wang *et al.* (2009). In order to control the consistency of other conditions, the same data with input variables and output variables are selected. The monthly runoff series (528 data points in total) of the Hongjiadu Reservoir from 1951 to 1994 is used as the training set to train the model. The monthly runoff series from 1995 to 2004 (120 data points in total) was selected as the test set to verify the model. The monthly runoff series (564 data points in total) of the Manwan Reservoir from 1953 to 1999 is selected as the training set to train the model. The monthly runoff series from 2000 to 2004 (a total of 60 data points) is selected as the test set to verify the model.

4.2. ICEEMDAN decomposition results

Based on the ICEEMDAN method, the monthly runoff data of the Hongjiadu Reservoir for 54 years and the Manwan Reservoir for 52 years are decomposed, respectively, to reduce the volatility of the original monthly runoff series and obtain more stable trend series and component data. The basic parameters of ICEEMDAN are as follows: $Nstd = 0.2$, $NR = 100$, $MaxIter = 5000$, $SNRFlag = 2$. The decomposition results are shown in Figures 3 and 4,
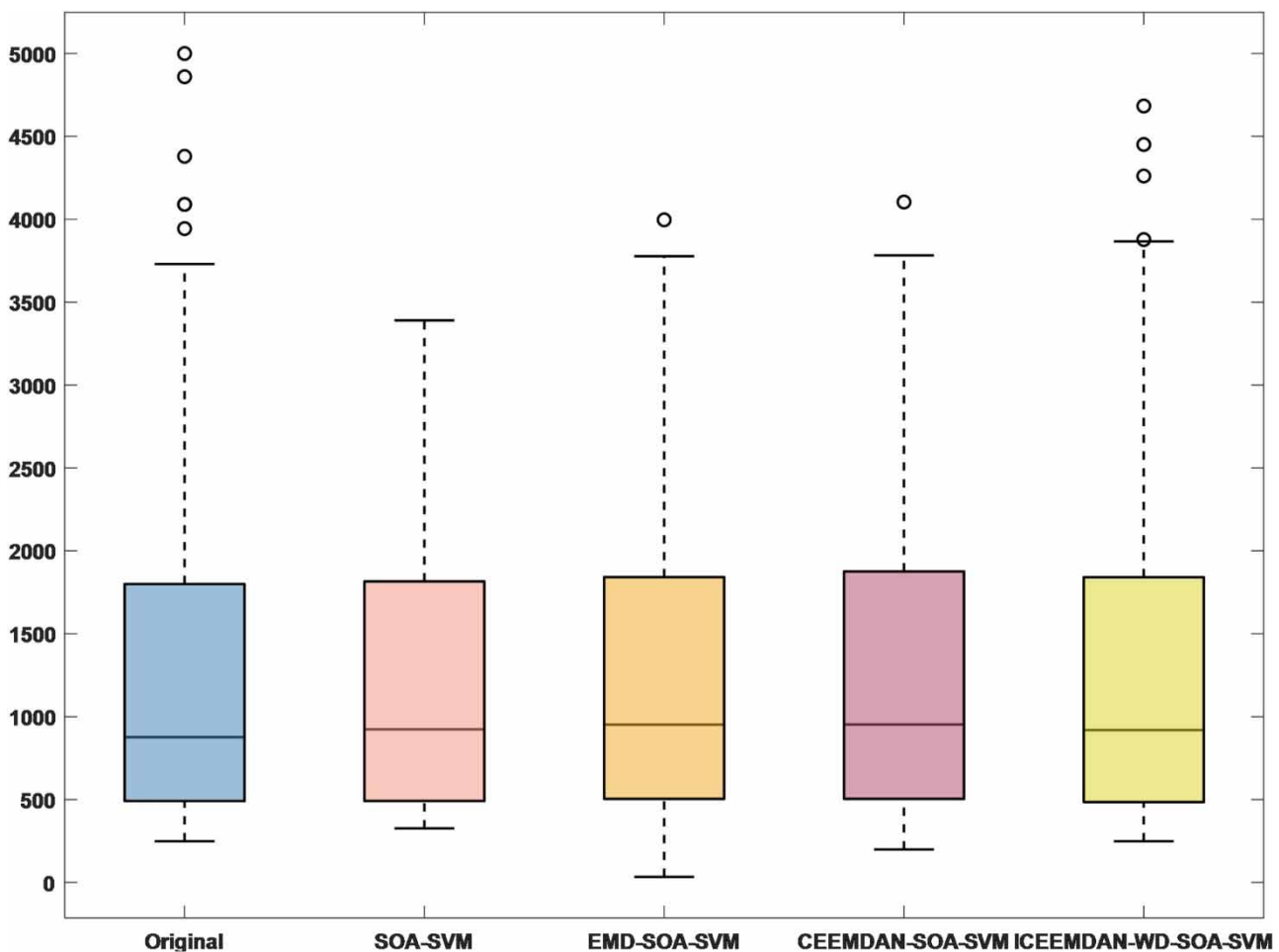

Figure 21 | Boxplot of four models for the Manwan Reservoir during the training period.

respectively. The original monthly runoff data series of the two stations are decomposed into nine components, namely eight IMF and one Res. The frequency and amplitude of the decomposed components from IMF1 to Res decrease gradually, while the wavelength and stability increase gradually. It can be observed from Figures 3 and 4 that the runoff of the Hongjiadu Reservoir generally presents a downward trend, while that of the Manwan Reservoir generally presents an upward trend. In addition, the volatility and complexity of IMF1 of the two stations are still very high, belonging to the high-frequency component, which affects the accuracy of runoff prediction and requires secondary decomposition.

4.3. Results of the quadratic decomposition of WD

As it can be seen from the decomposition results of the ICEEMDAN, IMF1 of the two stations has the largest fluctuation and highest complexity, belonging to high-frequency component, requiring secondary decomposition to reduce its volatility and render it more stable. In this paper, the db4 wavelet basis function is used to carry out a three-layer WD for IMF1 of the two stations and the decomposition results are shown in Figures 5 and 6, respectively. It can be seen from Figures 5 and 6 that the WD divides IMF1 of the two stations into high-frequency and low-frequency parts and the obtained component sequence is more stable. At this point, the decomposition part of 'decomposition-prediction-reconstruction' is completed. After ICEEM-DAN–WD, the original monthly runoff series of the Hongjiadu and Manwan Reservoirs are decomposed into 12 components, respectively. Then, in order to further prove the superiority of the ICEEMDAN–WD method, the EMD method and the CEEMDAN method are used to decompose the original monthly runoff series. Among them, 8 IMFs are obtained for
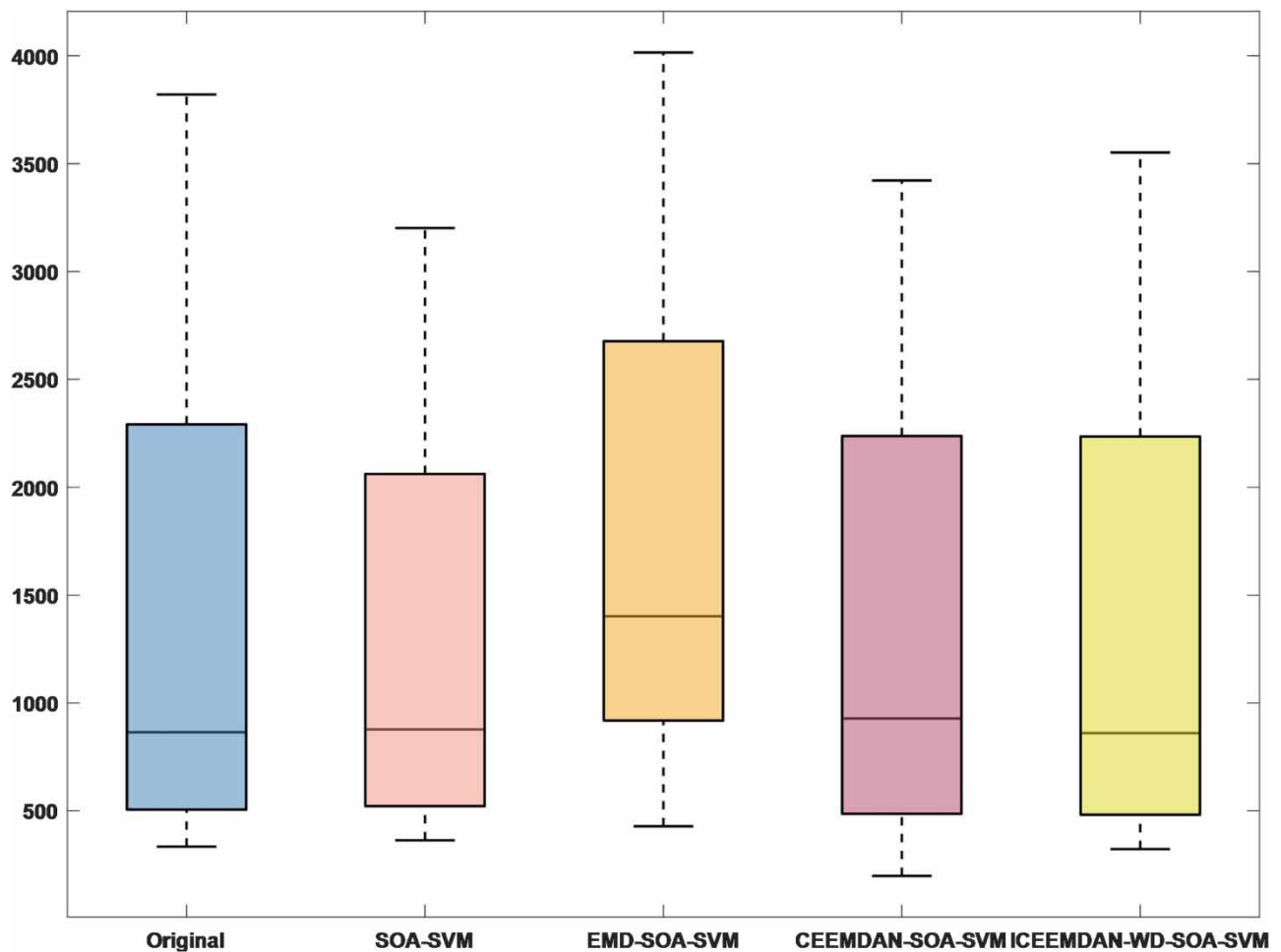

Figure 22 | Boxplot of four models for the Manwan Reservoir during the testing period.

Hongjiadu and Manwan Reservoirs using the EMD method and 10 IMFs are obtained for Hongjiadu and Manwan Reservoirs using the CEEMDAN method, respectively.

4.4. Application of the SOA–SVM runoff prediction model

This paper uses SOA to optimize SVM, which establishes the SOA–SVM model. The population size, location, speed and iteration number of SOA are initialized. The basic parameters of SOA are selected as follows: population number $N = 20$, maximum iteration coefficient $T = 100$, frequency coefficient $f_c = 2$ and cross validation $V = 5$. Three parameters of SVM are optimized by SOA, including the penalty factor C (in the interval of [0.0001, 1000]), kernel function parameter g (in the interval of [0.0001,1000]) and insensitive coefficient p (in the interval of [0.0001,1]). The SVM parameters adopted by Wang et al. (2009) are $C = 0.5045, p = 5.0814e - 004, \sigma = 0.6623$ in Hongjiadu and $C = 19,9373$, $p = 8.7775e - 004, \sigma = 1.2408$ in Manwan. Among them, $g = 1/2\sigma^2$. The 12 IMFs decomposed by the ICEEMDAN–WD method are input into the SOA–SVM model to obtain prediction results of each IMF. Finally, the prediction results of each sub-component are reconstructed to obtain the final monthly runoff prediction value. The optimized parameter results of the Hongjiadu and Manwan Reservoirs are shown in Tables 1 and 2, respectively. Meanwhile, the optimization process of the SOA–SVM model in terms of fitness function is shown in Figures 7–10.

5. RESULTS AND DISCUSSION

Four models, namely the SOA–SVM, EMD–SOA–SVM, CEEMDAN–SOA–SVM and the ICEEMDAN–WD–SOA–SVM, are used to predict monthly runoff. Prediction results of the Hongjiadu and Manwan Reservoirs are shown in Figures 11–14 and
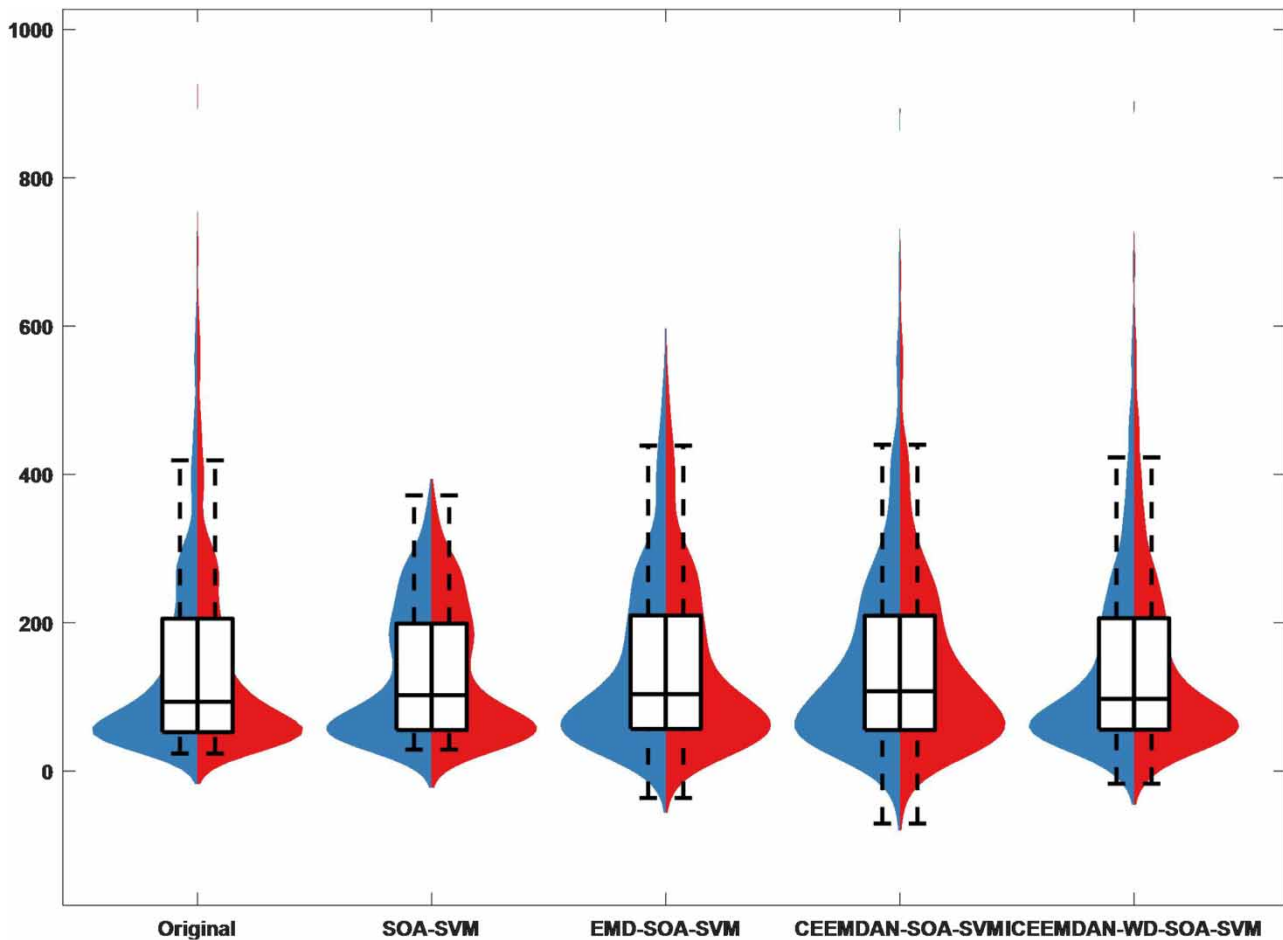

Figure 23 | Violin plot of four models for the Hongjiadu Reservoir during the training period.

the computation results of each evaluation index (together with those of the ARMA and ANN models as benchmarks for comparison) are shown in Tables 3 and 4. Taylor diagrams are shown in Figures 15–18, box plots are shown in Figures 19–22 and violin plots are shown in Figures 23–26.

It can be seen from Figures 11–14 that the prediction value obtained by the single SOA–SVM model without any decomposition has a large deviation from the real value and the prediction effect is the worst. The prediction effect of the EMD–SOA–SVM model is not ideal and the predicted runoff series only shows a general change trend. Although the prediction effect of the CEEMDAN–SOA–SVM is obviously better than the first two methods, the fitting effect of the peak value of the measured runoff series is not good and the error is very obvious. The predicted value and real value obtained by the ICCEMDAN–WD–SOA–SVM model have the best fitting effect and the predicted runoff series is closest to the measured runoff series, which reflects the accuracy and superiority of the runoff prediction model.

The prediction results during the test period can better reflect the performance of runoff prediction models. According to the nature of statistical indicators, the smaller the value of RMSE or MAPE, the better the result, while the larger the value of NSEC or R, the better the result. Tables 3 and 4 list the statistical results of the evaluation indicators for the final result data of the Hongjiadu and Manwan Reservoirs predicted by seven models. In addition, the general percentage comparison method is used to compute the statistical results of its evaluation indicators. It can be observed that:

(1) Comparing results of the ICEEMDAN–WD–SOA–SVM model with those of single the ARMA model during the test period at the Hongjiadu Reservoir they are as follows: RMSE is reduced by 58.32%; MAPE is decreased by 30.70%; NSEC is increased by 58.85%; and R is increased by 22.89%. Results obtained at the Manwan Reservoir are as follows: RMSE is reduced by 60.44%; MAPE is decreased by 28.93%. NSEC is increased by 12.83%; R is increased by 6.72%.
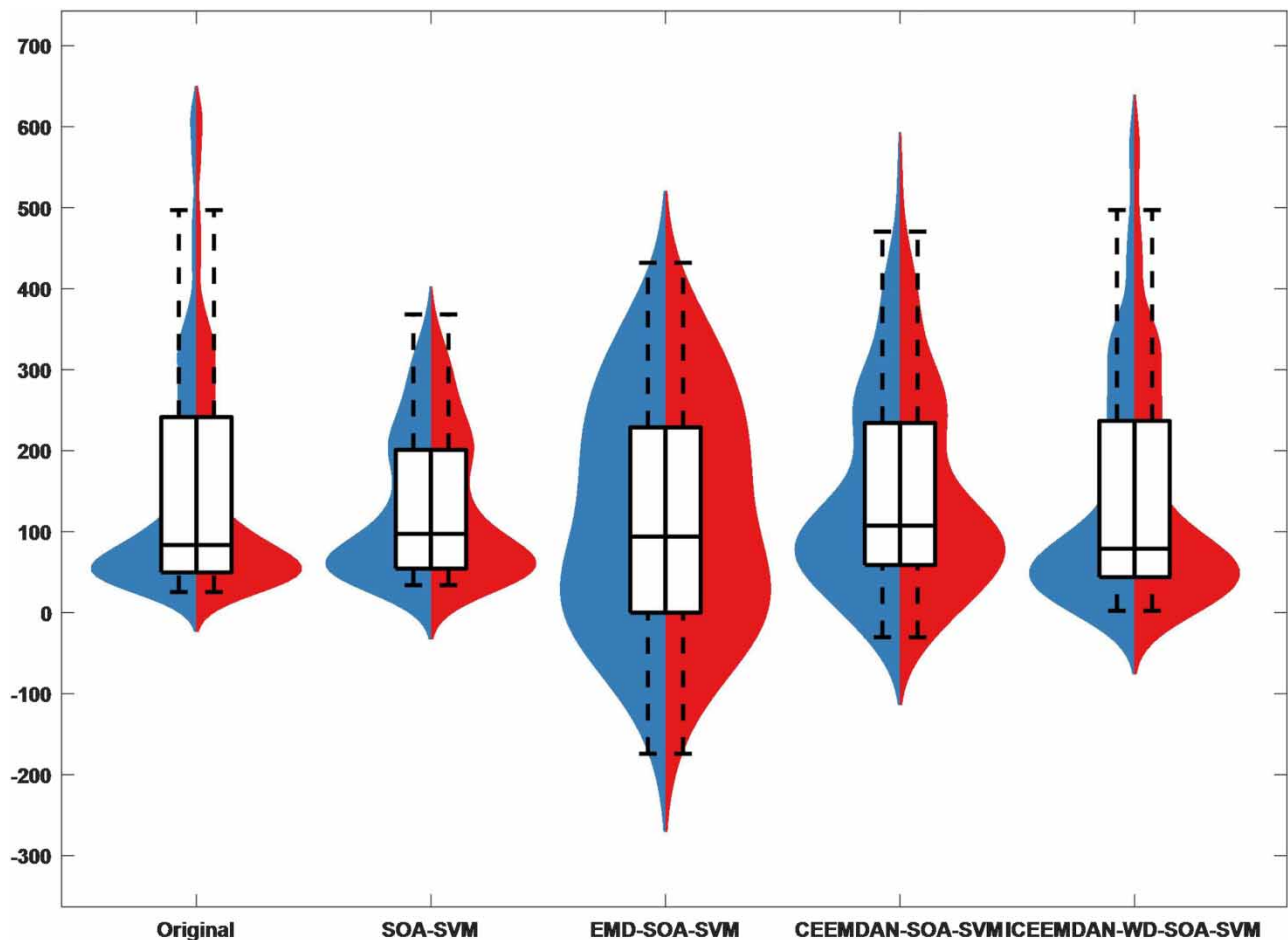

Figure 24 | Violin plot of four models for the Hongjiadu Reservoir during the testing period.

(2) Comparing results of the ICEEMDAN–WD–SOA–SVM model with those of the single ANN model during the test period at the Hongjiadu Reservoir they are as follows: RMSE is reduced by 56.82%; MAPE is decreased by 27.88%; NSEC is increased by 51.58%; and R is increased by 22.89%. Results obtained at the Manwan Reservoir are as follows: RMSE is reduced by 59.39%; MAPE is decreased by 20.71%. NSEC is increased by 13.09%; R is increased by 6.26%.

(3) Comparing results of the ICEEMDAN–WD–SOA–SVM model with those of a single SVM model during the test period at the Hongjiadu Reservoir they are as follows: RMSE is reduced by 55.10%; MAPE is decreased by 1.44%; NSEC is increased by 44.73%; and R is increased by 17.36%. Results obtained at the Manwan Reservoir are as follows: RMSE is reduced by 57.86%; MAPE is decreased by 11.06%. NSEC is increased by 10.96%; R is increased by 5.24%.

(4) Comparing results of the ICEEMDAN–WD–SOA–SVM model with those of the single SOA–SVM model during the test period at the Hongjiadu Reservoir they are as follows: RMSE is reduced by 59.63%; MAPE is decreased by 16.44%; NSEC is increased by 66.74%; and R is increased by 23.49%. Results obtained at the Manwan Reservoir are as follows: RMSE is reduced by 59.93%; MAPE is decreased by 12.40%. NSEC is increased by 11.63%; R is increased by 5.10%.

(5) Comparing results of the ICEEMDAN–WD–SOA–SVM model with those of the EMD–SOA–SVM model during the test period at the Hongjiadu Reservoir they are as follows: RMSE is reduced by 61.16%; MAPE is decreased by 64.23%; NSEC is increased by 78.19%; and R is increased by 22.75%. Results obtained at the Manwan Reservoir are as follows: RMSE is reduced by 76.87%; MAPE is decreased by 78.91%. NSEC is increased by 54.45%; R is increased by 11.82%.

(6) Comparing results of the ICEEMDAN–WD–SOA–SVM model with those of the CEEMDAN–SOA–SVM model during the test period at the Hongjiadu Reservoir they are as follows: RMSE is reduced by 39.67%; MAPE is
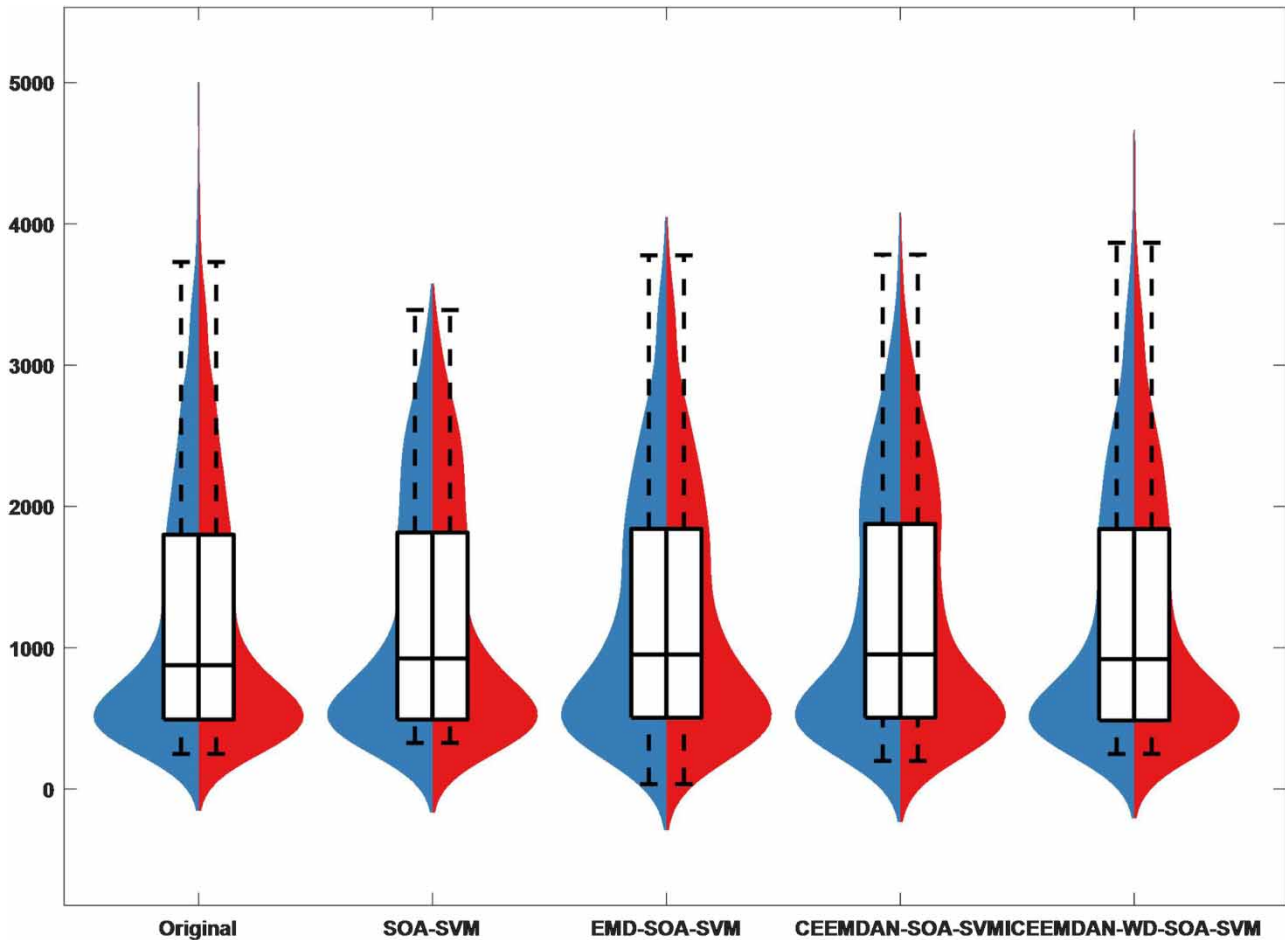

Figure 25 | Violin plot of four models for the Manwan Reservoir during the training period.

decreased by 27.13%; NSEC is increased by 15.77%; and R is increased by 7.56%. Results obtained at the Manwan Reservoir are as follows: RMSE is reduced by 40.11%; MAPE is decreased by 38.17%. NSEC is increased by 3.69%; R is increased by 1.82%.

It can be intuitively observed from Figures 15–18 that the RMSE, NSEC and R computed by the ICEEMDAN–WD–SOA–SVM model in the four Taylor charts are significantly superior to other models. It can be directly concluded that the predicted value obtained by the ICEEMDAN–WD–SOA–SVM model is closest to the real value and has the highest accuracy. At the same time, it can be seen that the prediction effects of the ARMA, ANN, SVM, SOA–SVM and the EMD–SOA–SVM models without any decomposition are very unsatisfactory and the performances of the three evaluation indicators are very poor. The prediction effect of the CEEMDAN–SOA–SVM model is obviously better than those of the previous five, but it is obviously worse than that of the ICEEMDAN–WD–SOA–SVM model.

It can be seen from the four boxplots in Figures 19–22 that the boxes of each model in the figure are not very different from the actual values (the upper quartile, median and lower quartile are approximately the same). By observing upper and lower limits, the comprehensive evaluation shows that the ICEEMDAN–WD–SOA–SVM model suits most with the actual value. More critically, through the comparison of outliers (highlighted in the figure), it can be directly concluded that the predicted outliers obtained by the ICEEMDAN–WD–SOA–SVM model are closest to the real outliers, which proves that the ICEEMDAN–WD–SOA–SVM model has the best prediction results.

It can be seen from Figures 23–26 that the the violin plot is similar to the box plot in that it focuses less on outliers and more on the distribution of the data, including the distribution profile and the distribution region, that is, the most concentrated region of data. As can be seen from the four violin plots, the data set regions of the four models are almost parallel and
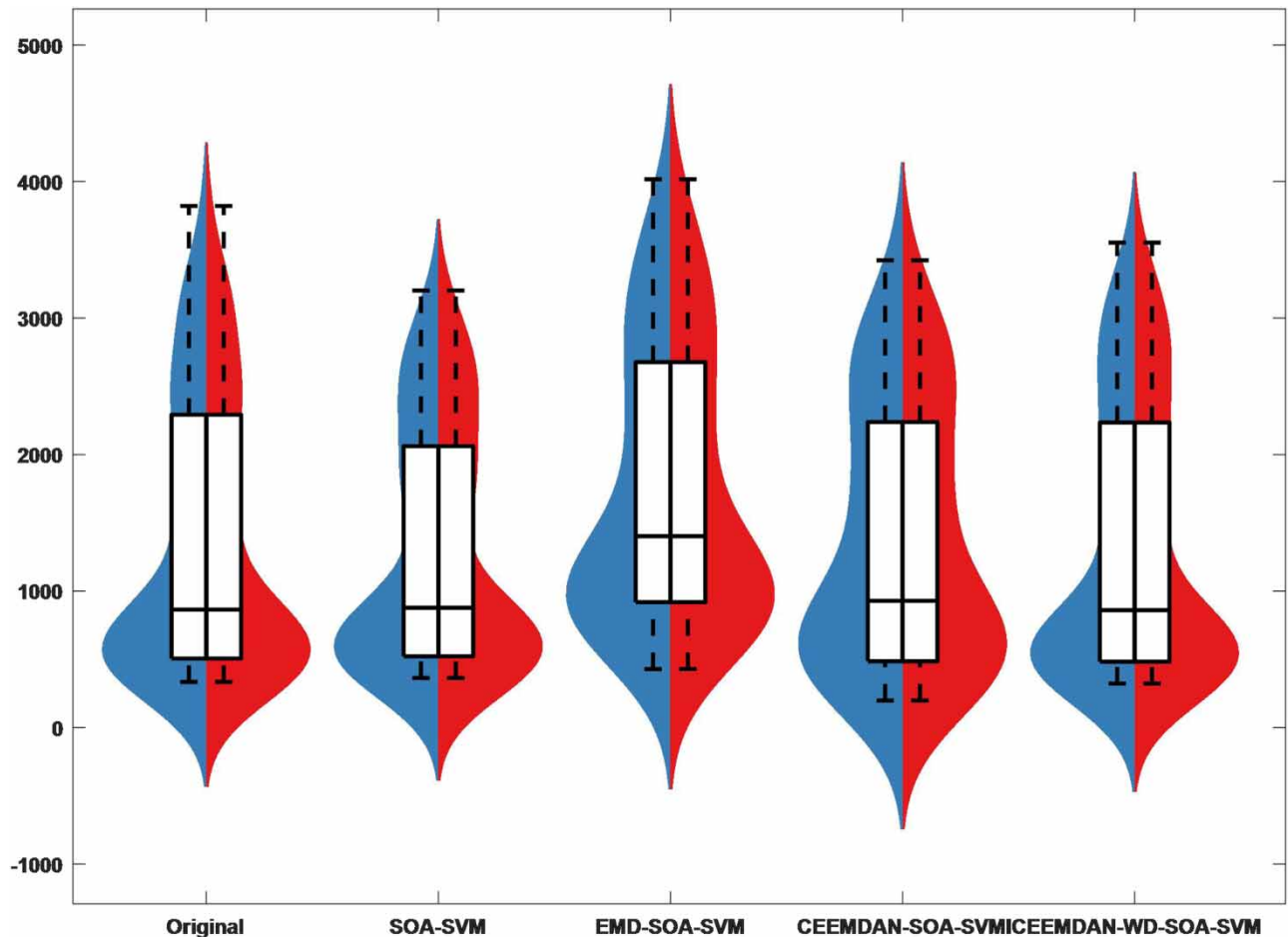

Figure 26 | Violin plot of four models for the Manwan Reservoir during the testing period.

are basically near the same value. At the same time, it can be seen that the data distribution shape of the ICEEMDAN–WD–SOA–SVM model in the four figures is the closest to the actual data distribution shape. By combining the boxplot and violin diagram, it can be seen that the ICEEMDAN–WD–SOA–SVM model is superior in both data distribution and outliers.

In summary, the ICEEMDAN–WD–SOA–SVM model proposed in this paper has the highest prediction accuracy results and the best fitting effect. This indicates that the runoff prediction performance of the ICEEMDAN–WD–SOA–SVM model is superior to those of the ARMA, ANN, SVM, SOA–SVM, EMD–SOA–SVM and the CEEMDAN–SOA–SVM models. This model has high feasibility, reliability and is suitable for predicting monthly runoff time series.

6. CONCLUSION

Medium- and long-term runoff prediction is of great significance to the rational development and utilization of water resources. In order to improve the accuracy of monthly runoff prediction, this paper proposes an ICEEMDAN–WD–SOA–SVM runoff prediction model. The main conclusions are as follows:

(1) The ICEEMDAN–WD method has advantages in dealing with non-stationary, non-linear runoff time series. The ICEEM-DAN method can effectively reduce the complexity of the original runoff series and make the runoff series as smooth as possible. The WD method can deal with high-frequency components of the runoff series decomposed by the ICEEMDAN method. After the decomposed data is input into the prediction model, the prediction result is more accurate and accurate. Thus, a novel method for runoff series decomposition is proposed.

(2) Compared with the single SOA–SVM model without decomposition, the prediction accuracy of the ICEEMDAN–WD–SOA–SVM model adopted in this paper is much higher, which proves the necessity of decomposing the runoff series. It is proved by the fact that runoff prediction using the 'decomposition-prediction-reconstruction' model can further improve the prediction accuracy and get more accurate prediction results.

(3) Based on the comparison of images and evaluation indicators, the prediction effect of the ICEEMDAN–WD–SOA–SVM model proposed in this paper is better than those of the ARMA, ANN, SVM, SOA–SVM, EMD–SOA–SVM and the CEEMDAN–SOA–SVM models. The peak fitting degree is higher and the error is smaller. The reliability and validity of the prediction ability of the proposed model are verified, which provides a reference for monthly runoff prediction and related research in other basins.

(4) The ICEEMDAN–WD–SOA–SVM model proposed in this paper effectively improves the accuracy of runoff prediction and is a feasible, reliable and novel method. However, this paper only considers the prediction effect of monthly runoff time series and runoff series of different time scales can be predicted in the future to verify the universal applicability of this runoff prediction model.

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

Adnan, R. M., Yuan, X., Kisi, O., Adnan, M. & Mehmood, A. 2018 Stream flow forecasting of poorly gauged mountainous watershed by least square support vector machine, fuzzy genetic algorithm and M5 model tree using climatic data from nearby station. *Water Resources Management* **32** (14), 4469–4486. doi:10.1007/s11269-018-2033-2.

Bi, X.-y., Li, B., Lu, W.-l. & Zhou, X.-z. 2020 Daily runoff forecasting based on data-augmented neural network model. *Journal of Hydroinformatics* **22** (4), 900–915. doi:10.2166/hydro.2020.017.

Büyükşahin, Ü. Ç. & Ertekin, Ş. 2019 Improving forecasting accuracy of time series data using a new ARIMA-ANN hybrid method and empirical mode decomposition. *Neurocomputing* **361**, 151–163. doi:10.1016/j.neucom.2019.05.099.

Cheng, N.-S., Chen, X. & Fang, Y. 2017 Estuary salinity prediction using a coupled GA-SVM model: a case study of the Min River Estuary, China. *Water Supply* **17** (1), 52–60. doi:10.2166/ws.2016.097.

Chong, K. L., Huang, Y. F., Koo, C. H., Najah Ahmed, A. & El-Shafie, A. 2022a Spatiotemporal variability analysis of standardized precipitation indexed droughts using wavelet transform. *Journal of Hydrology* **605**. doi:10.1016/j.jhydrol.2021.127299.

Chong, K. L., Huang, Y. F., Koo, C. H., Sherif, M., Ahmed, A. N. & El-Shafie, A. 2022b Investigation of cross-entropy-based streamflow forecasting through an efficient interpretable automated search process. *Applied Water Science* **13** (1). doi:10.1007/s13201-022-01790-5.

Colominas, M. A., Schlotthauer, G. & Torres, M. E. 2014 Improved complete ensemble EMD: a suitable tool for biomedical signal processing. *Biomedical Signal Processing and Control* **14**, 19–29. doi:10.1016/j.bspc.2014.06.009.

Corinna, C. & Vladimir, V. 1995 Support-vector networks. *Machine Learning* **20** (3), 273–297.

Dhiman, G. & Kumar, V. 2019 Seagull optimization algorithm: theory and its applications for large-scale industrial engineering problems. *Knowledge-Based Systems* **165**, 169–196. doi:10.1016/j.knosys.2018.11.024.

Essam, Y., Huang, Y. F., Birima, A. H., Ahmed, A. N. & El-Shafie, A. 2022a Predicting suspended sediment load in Peninsular Malaysia using support vector machine and deep learning algorithms. *Scientific Reports* **12** (1), 302. doi:10.1038/s41598-021-04419-w.

Essam, Y., Huang, Y. F., Ng, J. L., Birima, A. H., Ahmed, A. N. & El-Shafie, A. 2022b Predicting streamflow in Peninsular Malaysia using support vector machine and deep learning algorithms. *Scientific Reports* **12** (1), 3883. doi:10.1038/s41598-022-07693-4.

Fang, W., Huang, S., Ren, K., Huang, Q., Huang, G., Cheng, G. & Li, K. 2019 Examining the applicability of different sampling techniques in the development of decomposition-based streamflow forecasting models. *Journal of Hydrology* **568**, 534–550. doi:10.1016/j.jhydrol.2018.11.020.

Feng, Z.-k., Niu, W.-j., Tang, Z.-y., Jiang, Z.-q., Xu, Y., Liu, Y. & Zhang, H.-r. 2020 Monthly runoff time series prediction by variational mode decomposition and support vector machine based on quantum-behaved particle swarm optimization. *Journal of Hydrology* **583**. doi:10.1016/j.jhydrol.2020.124627.

He, X., Luo, J., Li, P., Zuo, G. & Xie, J. 2020 A hybrid model based on variational mode decomposition and gradient boosting regression tree for monthly runoff forecasting. *Water Resources Management* **34** (2), 865–884. doi:10.1007/s11269-020-02483-x.

Jajarmizadeh, M., Kakaei Lafdani, E., Harun, S. & Ahmadi, A. 2014 Application of SVM and SWAT models for monthly streamflow prediction, a case study in South of Iran. *KSCE Journal of Civil Engineering* **19** (1), 345–357. doi:10.1007/s12205-014-0060-y.

Ji, Y., Dong, H.-T., Xing, Z.-X., Sun, M.-X., Fu, Q. & Liu, D. 2021 Application of the decomposition-prediction-reconstruction framework to medium- and long-term runoff forecasting. *Water Supply* **21** (2), 696–709. doi:10.2166/ws.2020.337.

Lavanya, S., Annlin Jeba, S. V., Bhuvaneswari, P. & Shajin, F. H. 2022 Support vector machine classifier optimized with seagull optimization algorithm for brain tumor classification. *Concurrency and Computation: Practice and Experience* **35** (1). doi:10.1002/cpe.7396.

Li, C., Zhu, L., He, Z., Gao, H., Yang, Y., Yao, D. & Qu, X. 2019 Runoff prediction method based on adaptive elman neural network. *Water* **11** (6). doi:10.3390/w11061113.

Liu, G., Ouyang, S., Qin, H., Liu, S., Shen, Q., Qu, Y., Zheng, Z., Sun, H. & Zhou, J. 2022 Assessing spatial connectivity effects on daily streamflow forecasting using Bayesian-based graph neural network. *Science of the Total Environment* 158968. doi:10.1016/j.scitotenv.2022.158968.

Luo, M., Pan, C. & Zhan, C. 2019 Diagnosis of change in structural characteristics of streamflow series based on selection of complexity measurement methods: Fenhe River Basin, China. *Journal of Hydrologic Engineering* **24** (2). doi:10.1061/(asce)he.1943-5584.0001748.

Niu, W.-j., Feng, Z.-k., Cheng, C.-t. & Zhou, J.-z. 2018 Forecasting daily runoff by extreme learning machine based on quantum-behaved particle swarm optimization. *Journal of Hydrologic Engineering* **23** (3). doi:10.1061/(asce)he.1943-5584.0001625.

Niu, W.-j., Feng, Z.-k., Zeng, M., Feng, B.-f., Min, Y.-w., Cheng, C.-t. & Zhou, J.-z. 2019 Forecasting reservoir monthly runoff via ensemble empirical mode decomposition and extreme learning machine optimized by an improved gravitational search algorithm. *Applied Soft Computing* **82**. doi:10.1016/j.asoc.2019.105589.

Okkan, U. & Serbes, Z. A. 2012 Rainfall-runoff modeling using least squares support vector machines. *Environmetrics* **23** (6), 549–564. doi:10.1002/env.2154.

Othman, F., Alaaeldin, M. E., Seyam, M., Ahmed, A. N., Teo, F. Y., Ming Fai, C., Afan, H. A., Sherif, M., Sefelnasr, A. & El-Shafie, A. 2020 Efficient river water quality index prediction considering minimal number of inputs variables. *Engineering Applications of Computational Fluid Mechanics* **14** (1), 751–763. doi:10.1080/19942060.2020.1760942.

Ren, L., Xiang, X.-Y. & Ni, J.-J. 2013 Forecast modeling of monthly runoff with adaptive neural fuzzy inference system and wavelet analysis. *Journal of Hydrologic Engineering* **18** (9), 1133–1139. doi:10.1061/(asce)he.1943-5584.0000514.

Roushangar, K. & Shahnazi, S. 2021 Insights into the prediction capability of roughness coefficient in current ripple bedforms under varied hydraulic conditions. *Journal of Hydroinformatics* **23** (6), 1182–1196. doi:10.2166/hydro.2021.161.

Roushangar, K., Majedi Asl, M. & Shahnazi, S. 2021 Hydraulic performance of PK weirs based on experimental study and kernel-based modeling. *Water Resources Management* **35** (11), 3571–3592. doi:10.1007/s11269-021-02905-4.

Roushangar, K., Shahnazi, S. & Sadaghiani, A. A. 2022 An efficient hybrid grey wolf optimization-based KELM approach for prediction of the discharge coefficient of submerged radial gates. *Soft Computing*. doi:10.1007/s00500-022-07614-7.

Ruiming, F. 2019 Wavelet based relevance vector machine model for monthly runoff prediction. *Water Quality Research Journal* **54** (2), 134–141. doi:10.2166/wcc.2018.196.

Shao, Y., Mu, X., He, Y. & Chen, K. 2021 Variations in runoff, sediment load, and their relationship for a major sediment source area of the Jialing River Basin, southern China. *Hydrological Processes* **35** (7). doi:10.1002/hyp.14297.

Tanious, R. & Manolov, R. 2022 Violin plots as visual tools in the meta-analysis of single-case experimental designs. *Methodology* **18** (3), 221–238. doi:10.5964/meth.9209.

Tareen, A. D. K., Nadeem, M. S. A., Kearfott, K. J., Abbas, K., Khawaja, M. A. & Rafique, M. 2019 Descriptive analysis and earthquake prediction using boxplot interpretation of soil radon time series data. *Applied Radiation and Isotopes* **154**, 108861. doi:10.1016/j.apradiso.2019.108861.

Taylor, K. E. 2001 Summarizing multiple aspects of model performance in a single diagram. *Journal of Geophysical Research: Atmospheres* **106** (D7), 7183–7192. doi:10.1029/2000jd900719.

Thissen, U., van Brakel, R., de Weijer, A. P., Melssen, W. J. & Buydens, L. M. C. 2003 Using support vector machines for time series prediction. *Chemometrics and Intelligent Laboratory Systems* **69** (1–2), 35–49. doi:10.1016/s0169-7439(03)00111-4.

Wang, W.-C., Chau, K.-W., Cheng, C.-T. & Qiu, L. 2009 A comparison of performance of several artificial intelligence methods for forecasting monthly discharge time series. *Journal of Hydrology* **374** (3–4), 294–306. doi:10.1016/j.jhydrol.2009.06.019.

Wang, W.-c., Xu, D.-m., Chau, K.-w. & Chen, S. 2013 Improved annual rainfall-runoff forecasting using PSO-SVM model based on EEMD. *Journal of Hydroinformatics* **15** (4), 1377–1390. doi:10.2166/hydro.2013.134.

Wang, W.-c., Chau, K.-w., Xu, D.-m. & Chen, X.-Y. 2015a Improving forecasting accuracy of annual runoff time series using ARIMA based on EEMD decomposition. *Water Resources Management* **29** (8), 2655–2675. doi:10.1007/s11269-015-0962-6.

Wang, Y., Guo, S., Xiong, L., Liu, P. & Liu, D. 2015b Daily runoff forecasting model based on ANN and data preprocessing techniques. *Water* **7** (12), 4144–4160. doi:10.3390/w7084144.

Wang, X., Wang, Y., Yuan, P., Wang, L. & Cheng, D. 2021 An adaptive daily runoff forecast model using VMD-LSTM-PSO hybrid approach. *Hydrological Sciences Journal* **66** (9), 1488–1502. doi:10.1080/02626667.2021.1937631.

Wen, X., Feng, Q., Deo, R. C., Wu, M., Yin, Z., Yang, L. & Singh, V. P. 2019 Two-phase extreme learning machines integrated with the complete ensemble empirical mode decomposition with adaptive noise algorithm for multi-scale runoff prediction problems. *Journal of Hydrology* **570**, 167–184. doi:10.1016/j.jhydrol.2018.12.060.

Xing, B., Gan, R., Liu, G., Liu, Z., Zhang, J. & Ren, Y. 2016 Monthly mean streamflow prediction based on bat algorithm-support vector machine. *Journal of Hydrologic Engineering* **21** (2). doi:10.1061/(asce)he.1943-5584.0001269.

Xu, Z., Mo, L., Zhou, J., Fang, W. & Qin, H. 2022 Stepwise decomposition-integration-prediction framework for runoff forecasting considering boundary correction. *Science of the Total Environment* **851** (Pt 2), 158342. doi:10.1016/j.scitotenv.2022.158342.

Yong, B., Basheer, A. K., Wang, W. & Omer, A. 2017 Integrated assessment of the impacts of climate variability and anthropogenic activities on river runoff: a case study in the Hutuo River Basin, China. *Hydrology Research* **48** (2), 416–430. doi:10.2166/nh.2016.229.

Yuan, X., Chen, C., Lei, X., Yuan, Y. & Muhammad Adnan, R. 2018 Monthly runoff forecasting based on LSTM–ALO model. *Stochastic Environmental Research and Risk Assessment* **32** (8), 2199–2212. doi:10.1007/s00477-018-1560-y.

Zhang, J., Chen, X., Khan, A., Zhang, Y.-k., Kuang, X., Liang, X., Taccari, M. L. & Nuttall, J. 2021 Daily runoff forecasting by deep recursive neural network. *Journal of Hydrology* **596**. doi:10.1016/j.jhydrol.2021.126067.

Zhang, H., Peng, Z., Tang, J., Dong, M., Wang, K. & Li, W. 2022 A multi-layer extreme learning machine refined by sparrow search algorithm and weighted mean filter for short-term multi-step wind speed forecasting. *Sustainable Energy Technologies and Assessments* **50**. doi:10.1016/j.seta.2021.101698.

Zhao, X.-h. & Chen, X. 2015 Auto regressive and ensemble empirical mode decomposition hybrid model for annual runoff forecasting. *Water Resources Management* **29** (8), 2913–2926. doi:10.1007/s11269-015-0977-z.

Zhao, X., Chen, X., Xu, Y., Xi, D., Zhang, Y. & Zheng, X. 2017 An EMD-based chaotic least squares support vector machine hybrid model for annual runoff forecasting. *Water* **9** (3). doi:10.3390/w9030153.