Journal of Hydroinformatics

Journal of Hydroinformatics Vol 24 No 2, 367 doi: 10.2166/hydro.2022.116

A novel medium- and long-term runoff combined forecasting model based on different lag periods

Ping Ai [a,b], Yanhong Song[a,*], Chuansheng Xiong[b], Binbin Chen[a] and Zhaoxin Yue[c]

[a] College of Computer and Information Engineering, Hohai University, Nanjing 211100, China
[b] College of Hydrology and Water Resources, Hohai University, Nanjing 210098, China
[c] College of Computer and Software, Nanjing Vocational University of Industry Technology, Nanjing 210023, China
*Corresponding author. E-mail: syh826346006@126.com

PA, 0000-0003-3132-7236

ABSTRACT

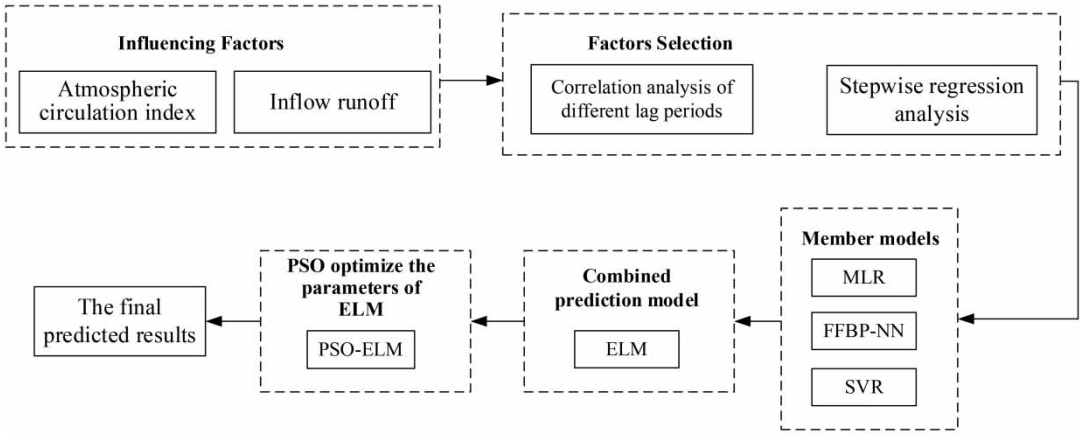
The accuracy of medium- and long-term runoff forecasting plays a significant role in several applications involving the management of hydrological resources, such as power generation, water supply and flood mitigation. Numerous studies that adopted combined forecasting models to enhance runoff forecasting accuracy have been proposed. Nevertheless, some models do not take into account the effects of different lag periods on the selection of input factors. Based on this, this paper proposed a novel medium- and long-term runoff combined forecasting model based on different lag periods. In this approach, the factors are initially selected by the time-delay correlation analysis method of different lag periods and further screened with stepwise regression analysis. Next, an extreme learning machine (ELM) is adopted to integrate each result obtained from the three single models, including multiple linear regression (MLR), feed-forward back propagation-neural network (FFBP-NN) and support vector regression (SVR), which is optimized by particle swarm optimization (PSO). To verify the effectiveness and versatility of the proposed combined model, the Lianghekou and Jinping hydrological stations from the Yalong River basin, China, are utilized as case studies. The experimental results indicate that compared with MLR, FFBP-NN, SVR and ridge regression (RR), the proposed combined model can better improve the accuracy of medium- and long-term runoff forecasting in the statistical indices of MAE, MAPE, RMSE, DC, U95 and reliability.

Key words: combined model, delay correlation analysis, extreme learning machine, medium- and long-term runoff forecasting, optimization algorithm

HIGHLIGHTS

- The delay correlation analysis with different lag periods was used to select the key factors.
- A novel medium- and long-term runoff combined forecasting model based on different lag periods was proposed, which performed better than other models.
- The lag period of physical factors delay can affect the accuracy of runoff forecasting.

GRAPHICAL ABSTRACT



INTRODUCTION

With the development of national economy and the adjustment of national water control policy, the gap between the existing hydrologic medium- and long-term runoff forecasting methods and the demand for production and application have been further widened. In addition, the increasing amount of hydrological data can introduce redundant and noisy information to the prediction feature or factor, which may deteriorate the performance of the mid- to long-term runoff prediction (Yue *et al.* 2020a). Therefore, runoff prediction is not only a management issue but also a scientific problem, and accurate runoff forecasts will be of vital importance to the policymakers.

Generally, the models to predict medium- and long-term runoff fall into two main categories: process-driven models and data-driven models. The process-driven model is based on the conception of hydrology, with which the discharge forecasting can be performed by simulation of the runoff variation and river channel evolution. Yet, the runoff process is highly complex and affected by many factors, such as runoff, atmospheric circulation index, climate and topography (Ekwueme & Agunwamba 2020). A precise construction of a physics-based model is very difficult to attain as an abundance of this data is required to fulfill the initial conditions (Yaseen *et al.* 2019). In comparison, the data-driven model can make the best use of existing data to achieve a predetermined model structure (Lu *et al.* 2021).

One of the most important steps in the data-driven model development process is determining the significant input variables. A subset of compact and informative inputs with the max-relevance and min-redundancy can significantly enhance the model performance and reduce the demand for training samples (Lyu *et al.* 2017). However, the determination of a suitable set of inputs is challenging due to the time lag of runoff response to the influence of large-scale atmospheric circulation (Cheng *et al.* 2019). The existing studies on the lag selection of runoff factors are mainly carried out for 1 or 2 years (Gao 2006; Yang *et al.* 2013; Li *et al.* 2019; Yue *et al.* 2020a). It is generally known that it is difficult to accurately determine the lag period since runoff flows through different field conditions to reach a point (He *et al.* 2011). To select the appropriate lag period, the 6, 12, 18 and 24 months as the different lag periods were considered sufficiently for capturing the input factors. The methods adopted for factors selection were mainly the correlation coefficient method (Yang *et al.* 2013; Shan *et al.* 2015), stepwise regression analysis (Shan *et al.* 2015), principal component analysis (PCA) (Lu *et al.* 2021), kernel principal component analysis (KPCA) (Li *et al.* 2019), mutual information (Lu & Zhou 2014) and partial mutual information (PMI) (He *et al.* 2011). Among them, the correlation coefficient method and stepwise regression analysis have the advantages of simplicity and quickness, and have been widely used in the field of hydrological prediction. Therefore, in this study, we propose a combined use of correlation coefficient and stepwise regression to determine the input factors of different lags.

Prediction model design is another important research content in mid- to long-term runoff prediction. A lot of approaches, including multiple linear regression (MLR) (Schilling & Wolter 2005; Wang *et al.* 2014), autoregressive (AR) techniques, autoregressive moving average (ARMA), autoregressive integrated moving average (ARIMA), seasonal autoregressive integrated moving average (SARIMA) (Sang *et al.* 2013; Valipour *et al.* 2013; Valipour 2015), feed-forward back propagation-neural network (FFBP-NN) (Wang *et al.* 2019), support vector regression (SVR) (Wei *et al.* 2012; Sahoo *et al.* 2018) and Elman neural network (Chu *et al.*

2017), have been widespread in medium- and long-term runoff forecasting. However, due to the intrinsic weaknesses of all models, and the uncertainty and complexity of runoff change, one forecasting model cannot improve runoff accuracy fundamentally, particularly when runoff data has nonlinear characteristics. Therefore, a combined model is used to exploit and integrate the diversity of skillful prediction from different models and reduce their uncertainty (Chen *et al.* 2019; Lu *et al.* 2021). It is divided into linear and nonlinear combination prediction, according to the relationship among the combined single-item prediction models. Linear combination prediction refers to the linear coupling of various single-item prediction methods. However, the possible emergence of negative weights and other controversial issues restricts the generalization of linear combination prediction to a certain extent. Nonlinear combination prediction refers to the nonlinear coupling of various single-item prediction methods, which overcome the limitation of linear combination prediction. The artificial neural networks have some unique and distinguishing features such as nonlinearity, parallel processing, robustness, adaptability and self-organization (Xu *et al.* 2021). An extreme learning machine (ELM) has the capability of approaching any complex continuous function and is easy to fit the nonlinearity of the predicted objects. It is proposed as a feed-forward neural network with a single hidden layer for determining weights related to the output. The ELM has been successfully applied for the estimation of reference evapotranspiration (Liu *et al.* 2017), river flow forecasting (Lima *et al.* 2016; Yaseen *et al.* 2019), mid- to long-term runoff forecasting (Yue *et al.* 2020b), longitudinal dispersion coefficients in water pipelines (Sabed-Movahed *et al.* 2020) and fast ionospheric delay (Zhao *et al.* 2021). Consequently, this paper proposes a combination prediction method using ELM to predict the medium- and long-term runoff for better prediction performance and stronger robustness.

The important point for solving the above problems is to select the optimal lag period of key factors that are closely related to the runoff and design a prediction model. Aiming to determine the optimal lag, this paper takes 6, 12, 18 and 24 months lag as the different lag periods. Moreover, to solve the problems caused by the low accuracy and poor stability of the single forecasting model, a combination prediction method using ELM to predict the medium- and long-term runoff is established, in which the random input weights and hidden biases in ELM are optimized by particle swarm optimization (PSO). The main goal of this study is achieved by (1) exploring the influence of atmospheric circulation factors on runoff in different lag periods and (2) proposing a combination prediction method using ELM to predict the medium- and long-term runoff.

METHODOLOGY

Construction of the combination method using ELM

The traditional combination forecasting methods just combine the forecasting results of several forecasting models together, average the weight coefficient in individual models or use an optimization algorithm to optimize the weight coefficient (Chen *et al.* 2019). A novel combined forecasting method is proposed in this paper, which puts the intermediate forecasting results of MLR, FFBP-NN and SVR into ELM. In this approach, this paper initially uses the time-delay correlation analysis and stepwise correlation analysis to select the optimal forecasting factor as the inputs. Then, three single prediction models, namely MLR, FFBP-NN and SVR, are established and used to predict runoff. And then, ELM was used to combine the intermediate forecasting results of the three single prediction models. At last, considering that the random input weights and hidden biases of ELM always have some influence on the training process, we use PSO to optimize the parameters to derive the resultant forecast. The flowchart of the combined model is shown in Figure 1; from the figure, we can conclude that the proposed method contains the following four steps.

Step 1. *Factors selection*. The input characteristic factors of different lag periods are obtained by the time-delay correlation analysis and stepwise regression analysis. The illustrative diagram is shown in Figure 1(a).

Step 2. *Single model learning*. The training set was used to train MLR, FFBP-NN and SVR, respectively, to select the parameters of the three models. The three trained models are applied to the training set and the test set, respectively, and the respective prediction results are obtained, which are used as the training set and test set of the combination forecasting model. The process of single model learning is represented in Figure 1(b).

Step 3. *Combined model forecasting*. MLR, FFBP-NN and SVR are regarded as approaches to forecast inflow runoff, respectively, and their intermediate forecasting results are used as input of ELM, and then train the ELM repeatedly to get the final forecasting results. Figure 1(c) shows the specific process of the combined model.

Step 4. *Parameters optimization*. Two parameters in ELM are very important in the forecasting process, so the random input weights and hidden biases are optimized by PSO, then the output weights could be calculated. Figure 1(d) gives the schematic diagram of the optimization procedure.

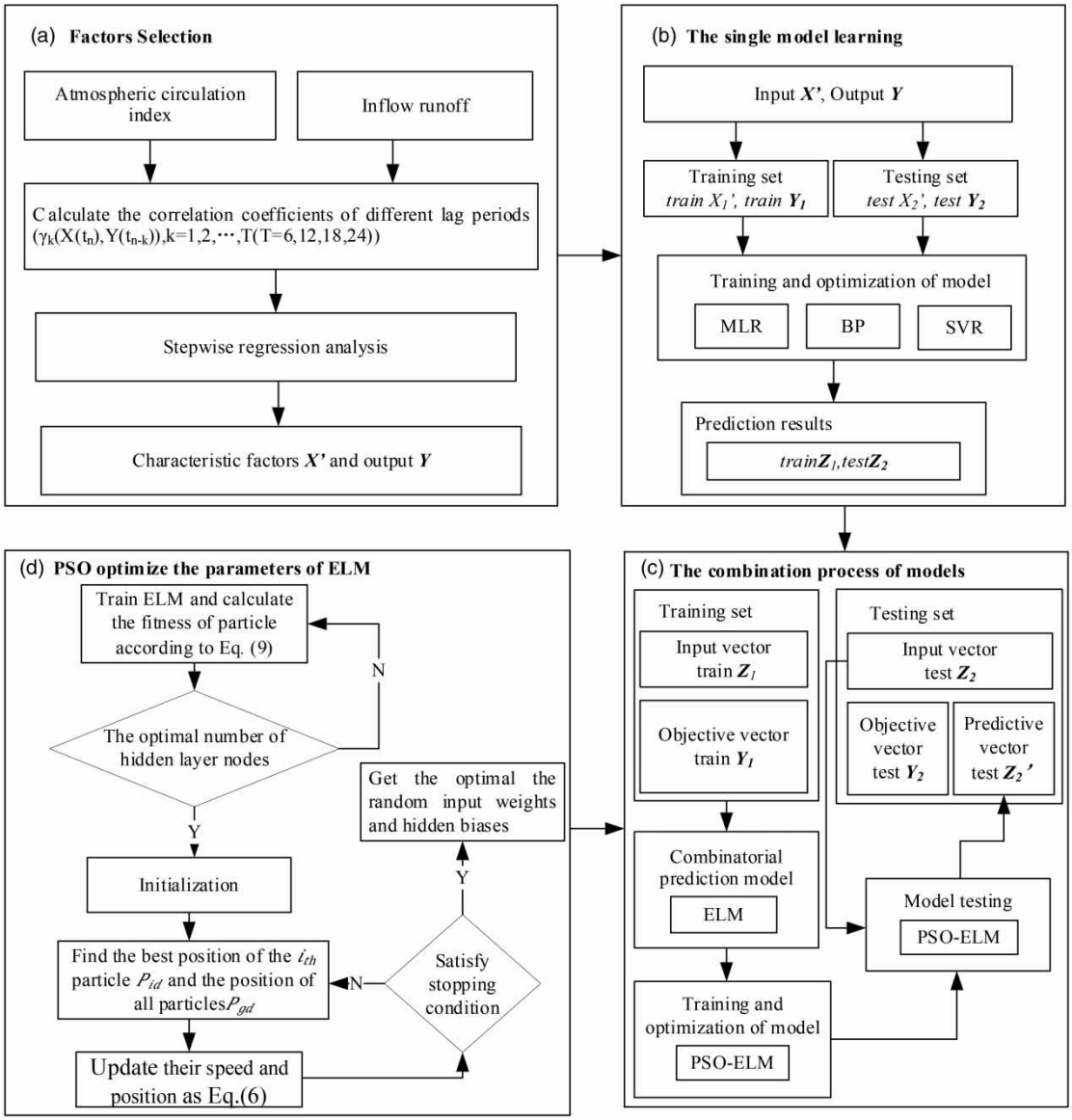


Figure 1 | Flowchart of the proposed combined model.

Time-delay correlation analysis and stepwise regression analysis

Time-delay correlation analysis with different lag periods

The time-delay correlation analysis is used to measure the similarity between two different time series and consider a certain range of time delay (Li *et al.* 2019). The correlation coefficient shows the intensity of the linear relationship between the two time series. The numerical range is from −1 to 1, and the closer the absolute value is to 0, indicating that the two sequences tend to be completely independent, and the closer their absolute values are to 1, indicating that there is a strong correlation between the two sequences. The calculation formula is as follows:

$$r_k(x_1, x_2) = \frac{c_k(x_1, x_2)}{\sigma_{x_1}\sigma_{x_2}} = \frac{\sum_{n=k+1}^{N+k}(x_1(t_n)-\bar{x}_1)(x_2(t_{n-k})-\bar{x}_2)}{\sqrt{\sum_{n=k+1}^{N+k}(x_1(t_n)-\bar{x}_1)^2\sum_{n=k+1}^{N+k}(x_2(t_{n-k})-\bar{x}_2)^2}} \quad (1)$$

where k is the lag period ($k = 1, 2, 3, \cdots, T$), T stands for different lag periods, $T = 6$, 12, 18, and 24 months; r_k is the correlation coefficient of two time series $x_1(t_n)$ and $x_2(t_n)$ under the time delay k; c_k is the covariance of two time series $x_1(t_n)$ and $x_2(t_n)$ under the time delay k; σ_{x_1} and σ_{x_2} are the mean square deviations of $x_1(t_n)$ and $x_2(t_n)$, respectively; $\bar{x}_1$ and $\bar{x}_2$ are the mean values of $x_1(t_n)$ and $x_2(t_n)$, respectively; and N is the length of time series.

Stepwise regression analysis

The main idea of stepwise regression analysis is to introduce predictors step by step. At each step in the process, after a new variable is added, a test is made to check if some variables can be deleted without appreciably increasing the residual sum of squares (RSS). The procedure terminates when the measure is (locally) maximized or when the available improvement falls below some critical value.

Extreme learning machine

ELM is an innovative machine learning algorithm that is characterized mainly by the fact that there is no need for a tuning of the model's internal parameters (i.e., the hidden neurons). In essence, ELM is an improved version of the conventional ANN model where it is able to solve regression problems with a reduced model execution time. This is because the input weights and biases are randomly generated so that the output weights have a unique least-squares solution solved by the Moore–Penrose generalized inverse function (Yaseen *et al.* 2019). The general structure of the ELM model can be visualized in Figure 2.

Mathematically, the ELM model can be summarized by setting up a set of training dataset samples (x_i, y_i). Where $x_i = [x_{i1}, x_{i2}, \cdots, x_{in}]^T$ present the predictor variables that are the results of runoff output from different prediction models and $y_i = [y_{i1}, y_{i2}, \cdots, y_{im}]^T$ are the predicted value of runoff. $x_i \in R^d$ and $y_i \in R^m$. The target function is defined as follows:

$$\sum_{i=1}^{M} \beta_i g(w_i \cdot x_j + b_i) = t_j, j = 1, 2, \cdots, N \tag{2}$$

where M is the number of hidden layer neurons; β_i stands for the weight factor connecting the ith hidden nodes and output node; $g(w_i \cdot x_j + b_i)$ denotes the activation function, $w_i \cdot x_j$ is the inner product of w_i and x_j, w_i and b_i are the hidden node parameters that are randomly determined; and t_j is the network output.

To achieve the training effect, the output error needs to be minimized, which can be expressed as follows:

$$\|H(\hat{W}_i, \hat{b}_i)\hat{\beta}_i - T\| = \min\|H(W_i, b_i)\beta_i - T\| \tag{3}$$

where H is the network hidden layer output, and T is the desired output. Equation (1) is equivalent to the loss minimization

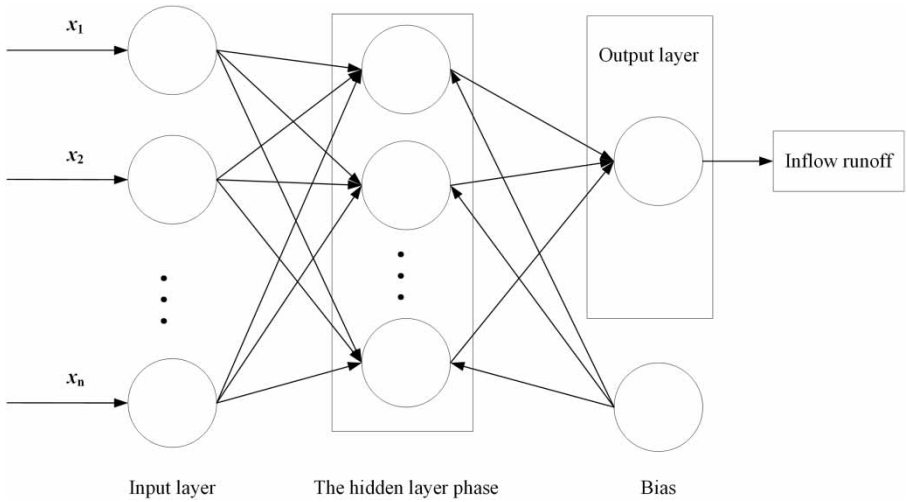


Figure 2 | Architecture of the extreme learning machine model.

function:

$$E = \sum_{j=1}^{N} \left(\sum_{i=1}^{M} \beta_i g(w_i \cdot x_j + b_i) - t_j \right)^2 \tag{4}$$

Equation (4) can be converted to solving a linear equation $H\beta = T$, and the output weight can be uniquely determined:

$$\hat{\beta} = H^{+}T \tag{5}$$

where H^{+} is the Moore–Penrose inverse of Hussain (H) matrix.

Particle swarm optimization

PSO is a population-based stochastic optimization technique inspired by the social behavior of bird flocking (Motahari & Mazandaranizadeh 2017). In PSO, all particles have a fitness value, which is determined by the fitness function. Each particle can save the best position to be searched, and the particle is also given a speed to determine the search direction and distance. In the process of searching, the particle tracks two extremes to complete the iteration of a particle search. The optimal solution of the particle search is called the individual extreme value pBest. The optimal solution of the whole particle swarm search is called the global extremum gBest.

In the process of iterative optimization, particles constantly update their speed and position through pBest and gBest, that is,

$$\begin{aligned} V_{id}^{m+1} &= \alpha V_{id}^{m} + c_1 t_1 (P_{id}^{m} - X_{id}^{m}) + c_2 t_2 (P_{gd}^{m} - X_{id}^{m}) \\ X_{id}^{m+1} &= X_{id}^{m} + V_{id}^{m+1} \end{aligned} \tag{6}$$

where m represents the current number of iterations; c_1 and c_2 stand for the cognitive and social constants, $c_1 = c_2 = 1.49445$; t_1 and t_2 show the random variables; α is the inertia term; V_{id} represents the current particle velocity; X_{id} stands for the current particle position; and P_{id} and P_{gd} represent the particle neighborhood best and the particle best, respectively.

Evaluation metrics

At present, various performance metrics are used to evaluate the forecasting performance of the model. But there is no single standard to determine which evaluation metric is the most accurate assessment method. Therefore, to estimate the performance of the proposed model, the evaluation metrics, such as bias index (BIAS) (Najafzadeh & Sattar 2015; Barzkar *et al.* 2021), scatter index (SI) (Najafzadeh *et al.* 2022), mean absolute percentage error (MAPE), root-mean-square error (RMSE), mean absolute error (MAE) and deterministic coefficient (DC) (Yue *et al.* 2020a, 2020b), are applied. Among them, four common evaluation metrics are adopted in this paper, including MAE, MAPE, RMSE and DC. The expressions are as follows:

$$\text{MAE} = \frac{1}{n} \sum_{t=1}^{n} |y_t - \hat{y}_t| \tag{7}$$

$$\text{MAPE} = \frac{1}{n} \sum_{t=1}^{n} \left| \frac{y_t - \hat{y}_t}{y_t} \right| \tag{8}$$

$$\text{RMSE} = \sqrt{\frac{1}{n} \sum_{t=1}^{n} (y_t - \hat{y}_t)^2} \tag{9}$$

$$\text{DC} = 1 - \frac{\sum_{t=1}^{n} (y_t - \hat{y}_t)^2}{\sum_{t=1}^{n} (y_t - \bar{y})^2} \tag{10}$$

where y_t is the observed value of runoff at time t, $m^3 \cdot s^{-1}$; $\hat{y}_t$ represents the predicted runoff value at time t, $m^3 \cdot s^{-1}$; $\bar{y}$ represents the mean of the observed runoff value, $m^3 \cdot s^{-1}$ and n is the total number of samples.

The MAPE is the most frequently used statistics index and is employed for examining the error between the predicted value and the observed value. The RMSE is used to measure the predicted value accuracy. An RMSE value of zero indicates a perfect match between the predicted value and the observed value. The MAE is the average of the absolute value of the deviation between all individual observations and the arithmetic mean, which can accurately reflect the actual prediction error. Likewise, the DC provides a measure of the capacity of the model to predict observed values. And, the smaller the values of the MAE, MAPE and RMSE are, the better forecasting performance is. The higher the DC value shows, the better forecasting performance is.

STUDY AREA AND DATA SOURCES

Study area

The Yalong River basin is located at 96°52′–102°48′ E, 26°32′–33°58′ N, with an area of about 136,000 km^2. The river is 1,571 km long with a drop of 3,830 m and is one of the rivers with the most abundant water energy resources in China (Yue *et al.* 2020b). The precipitation in the basin increases from north to south, and the east is greater than the west. The annual precipitation in the Heyuan area is 500–600 mm, and the annual precipitation in the middle and lower reaches is 900–1,300 mm. The runoff was mainly from precipitation, and the annual variation and regional distribution of runoff were basically consistent with the variation trend of precipitation. The annual runoff distribution can be roughly divided into the period from June to October, during which rainfall is the main replenishment and the water volume accounts for 76.8% of the whole year. The hydrographic stations in the Yalong River basin, Lianghekou and Jinping are shown in Figure 3.

Data sources

To prove that the combined model proposed in this paper can improve medium- and long-term runoff forecasting accuracy, the collected datasets include: (1) monthly inflow runoff data of Lianghekou and Jinping stations from January 1960 to December 2011 were provided by the Hydrographic Bureau of the Yangtze River Water Conservancy Commission; (2) 130 atmospheric circulation index data from the national climate center (https://www.ncc-cma.net/Website/index.php), the length of the sequence for January 1951–May 2020. And the research period spans over 50 years from January 1962 to December 2011, which is divided into two parts: a training set (from January 1962 to December 2001) and a forecasting test set (from January 2002 to December 2011) in this study. In relation to this, Figure 4 shows the dataset structure of Lianghekou and Jinping stations, which were selected as the forecast objects. It can be seen from Figure 4 that the abundance and drought of the three stations are very similar, indicating that the abundance and drought of the Yalong River basin are basically the same. In addition, we have done a statistical analysis of the three datasets. The statistical characteristics, including the maximum, minimum, means, standard and median, of the runoff series at the different time period divisions are summarized in Table 1.

RESULTS

Factors selection procedure and result

In our study, considering the time consistency of serial data and the missing term of atmospheric circulation index, the alternative candidate predictive factors include 95 teleconnection climate factors and antecedent runoff in Lianghekou and Jinping stations of the Yalong River basin. For the convenience of analysis, they are abbreviated as a1, a2,…, a96. In addition, due to the lag effect of climate-related factors on runoff (Yue *et al.* 2020a), this paper took 6, 12, 18 and 24-month lag as the different lag periods. Thus, the input factors are expressed as a1($t-1$), a1($t-2$),…, a1($t-T$),…, a96($t-1$), a96($t-2$),…, a96($t-T$) ($T=6$, 12, 18, 24), including (96*T) variables.

The input selection was initially conducted based on the time-delay correlation coefficient with different lag periods between each factor and runoff. To control the number of primary factors, the maximum correlation coefficient and the absolute value of the correlation coefficient are greater than 0.7. Thus, the input factors of the 6- and 12-month primary elections in the Lianghekou lag period are shown in Table 2. The primary factors at 18 and 24 months in the screen lag period are shown in Table 3. The correlation between the above factors and the runoff coefficient has passed the significance test with a confidence level of 95%.

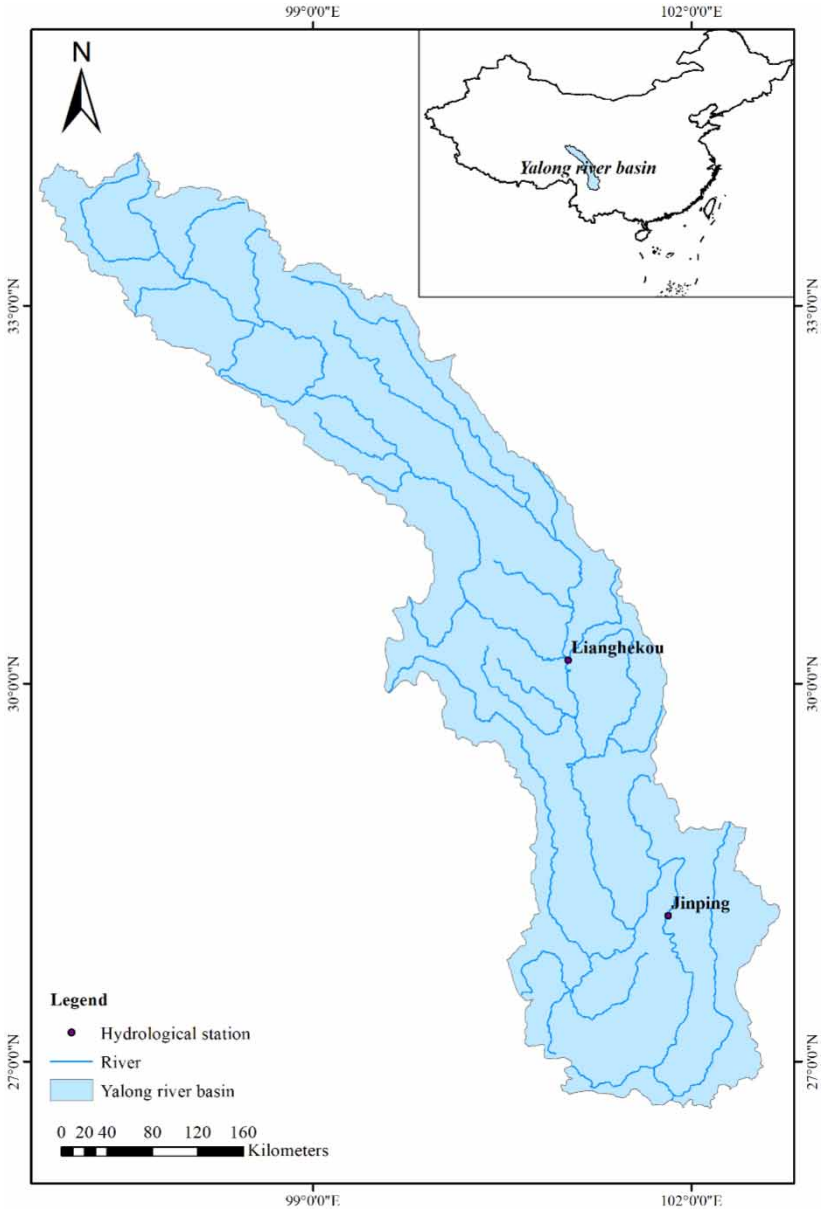


Figure 3 | Location of the study area.

And then, to eliminate multicollinearity among problem variables, the stepwise regression analysis was applied, and the results of the significance of the factors are presented in Tables 4–7. Tables 4 and 5 show that the stepwise regression analysis indicates that the factors selected at 6 and 12 months of the lag period of the Lianghekou hydrological station were both statistically significant in predicting month runoff (sig. < 0.05) with significant F values of 260.672 and 200.097 (Ekwueme & Agunwamba 2020). The factors affecting the change of runoff process of the Lianghekou station when the lag period is 6 months are mainly related to a43, a96, a29, a2, a3, a42 and a36; for a lag period of 12 months, the factors affecting the runoff change are mainly in connection with a36, a42, a64, a2, a43, a44, a96 and a31 (see the bold values in Table 2). As can be seen from Tables 6 and 7, the factors selected at 18 and 24 months of the lag period of the Jinping hydrological station were both statistically significant in predicting monthly runoff (sig. < 0.05) with significant F values of 264.787 and 276.324. The factors affecting the change of runoff process for a lag period of 18 months are mainly associated with a36, a43, a64, a96, a82, a44, a13 and a33; for a lag period of 24 months, the factors affecting the runoff change are mainly related to a36, a96, a43, a83, a18, a44, a82 and a33 (see the bold values in Table 3). A similar procedure was implemented to decide the input

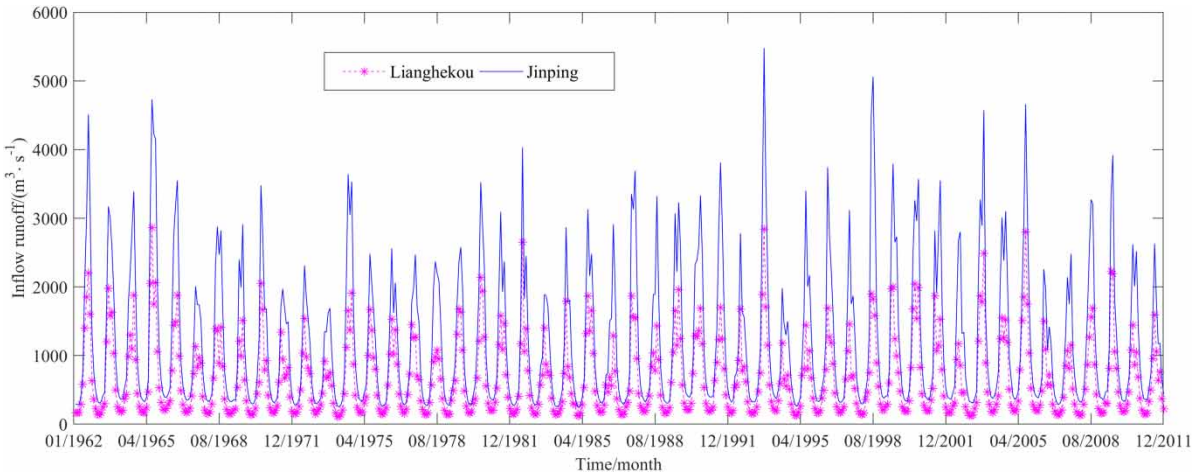


Figure 4 | Dataset structure of Lianghekou and Jinping stations.

Table 1 | Statistical indicators of the experimental data

				Statistical indicators ($m^3 \cdot s^{-1}$)				
Stations	**Period of records**	**Samples**	**Numbers**	**Max.**	**Min.**	**Mean**	**Std.**	**Median**
Lianghekou	1,962.1–2,011.12	All samples	600	2,860.00	108.00	661.65	558.23	445.50
	1,962.1–2,001.12	Training	480	2,860.00	108.00	662.87	552.81	446.00
	2,012.1–2,011.12	Testing	120	2,800.00	124.00	656.77	581.76	426.50
Jinping	1,962.1–2,011.12	All samples	600	5,480.00	243.00	1,209.33	1,034.72	739.00
	1,962.1–2,001.12	Training	480	5,480.00	243.00	1,209.43	1,039.66	739.00
	2,012.1–2,011.12	Testing	120	4,660.00	285.00	1,208.89	1,019.01	737.00

Table 2 | The results of input variables selected by the time-delay correlation analysis method in the lag period of 6 months and 12 months for the Lianghekou station

Lag period of 6 months				**Lag period of 12 months**			
Serial number	**Input factors**	**Correlation coefficient**	**Delay time/month**	**Serial number**	**Input factors**	**Correlation coefficient**	**Delay time/month**
1	**a2**	0.7039	1	1	**a2**	0.7039	1
2	**a3**	0.7032	1	2	a3	0.7032	1
3	**a29**	−0.8123	1	3	a8	0.7125	12
4	a30	−0.7554	1	4	a10	0.7064	12
5	a31	0.7421	1	5	a29	−0.8123	1
6	a33	−0.8048	1	6	a30	−0.7554	1
7	**a36**	0.8124	1	7	**a31**	0.7477	7
8	**a42**	0.7886	1	8	a33	−0.8048	1
9	**a43**	−0.8033	6	9	**a36**	0.8124	1
10	a44	−0.8004	6	10	**a42**	0.8059	12
11	**a96**	0.7261	1	11	**a43**	−0.8033	6
				12	**a44**	−0.8004	6
				13	**a64**	0.7282	11
				14	**a96**	0.7434	12

Table 3 | The results of input variables selected by the time-delay correlation analysis method in the lag period of 18 months and 24 months for the Jinping station

Lag period of 18 months				Lag period of 24 months			
Serial number	**Input factors**	**Correlation coefficient**	**Delay time/month**	**Serial number**	**Input factors**	**Correlation coefficient**	**Delay time/month**
1	a1	0.7171	1	1	a1	0.7171	1
2	a2	0.7476	1	2	a2	0.7476	1
3	a3	0.7654	1	3	a3	0.7654	1
4	a7	0.7319	1	4	a7	0.7319	1
5	a8	0.7497	1	5	a8	0.7497	1
6	a10	0.7502	1	6	a10	0.7502	1
7	**a13**	0.7338	1	7	a13	0.7338	1
8	a14	0.7248	1	8	a14	0.7248	1
9	a29	−0.8213	1	9	**a18**	0.7014	24
10	a30	−0.7673	1	10	a21	0.7068	24
11	a31	0.7459	7	11	a29	−0.8213	1
12	a32	0.7008	14	12	a30	−0.7673	1
13	**a33**	−0.8055	1	13	a31	0.7550	19
14	**a36**	0.8374	1	14	a32	−0.7008	14
15	a42	0.8220	1	15	a33	−0.8055	1
16	**a43**	−0.8127	18	16	**a36**	0.8374	1
17	**a44**	−0.8122	7	17	a42	0.8220	1
18	**a64**	0.7295	11	18	**a43**	−0.8127	18
19	**a82**	0.7411	16	19	**a44**	−0.8122	7
20	a83	0.7127	4	20	**a64**	0.7295	11
21	**a96**	0.7912	12	21	**a82**	0.7411	16
				22	**a83**	0.7127	4
				23	**a96**	0.8159	24

Table 4 | ANOVA of key factors in the lag period of 6 months for the Lianghekou station

ANOVA[a]

	Sum of squares	**df**	**Mean square**	***F***	**Sig.**
Regression	140,936,496.583	7	201,337,85.226	260.672	0.000[b]
Residual	457,248,77.917	592	77,237.969		
Total	186,661,374.500	599			

[a]The dependent variable: runoff.
[b]The predictor variable: (constant), a43, a96, a29, a2, a3, a42, a36.

Table 5 | ANOVA of key factors in the lag period of 12 months for the Lianghekou station

ANOVA[a]

	Sum of squares	**df**	**Mean square**	***F***	**Sig.**
Regression	136,329,141.467	8	17,041,142.683	200.097	0.000[b]
Residual	50,332,233.033	591	85,164.523		
Total	186,661,374.500	599			

[a]The dependent variable: Runoff.
[b]The predictor variable: (constant), a36, a42, a64, a2, a43, a44, a96, a31.

Table 6 | ANOVA of key factors in the lag period of 18 months for the Jinping station

ANOVA[a]

	Sum of squares	df	Mean square	F	Sig.
Regression	501,420,363.068	8	62,677,545.384	264.787	.000[b]
Residual	139,895,444.557	591	236,709.720		
Total	641,315,807.625	599			

[a]The dependent variable: Runoff.
[b]The predictor variable: (constant), a36, a43, a64, a96, a82, a44, a13, a33.

Table 7 | ANOVA of key factors in the lag period of 24 months for the Jinping station

ANOVA[a]

	Sum of squares	df	Mean square	F	Sig.
Regression	506,029,414.191	8	63,253,676.774	276.324	.000[b]
Residual	135,286,393.434	591	228,910.987		
Total	641,315,807.625	599			

[a]The dependent variable: Runoff.
[b]The predictor variable: (constant), a36, a96, a43, a83, a18, a44, a82, a33.

factors for the remaining different lag periods of Lianghekou and Jinping stations. The inputs retained in the final subset of Lianghekou and Jinping stations are formed in Table 8.

ELM combination method for runoff

Comparison models

To sufficiently demonstrate the superiority of the proposed model for medium- and long-term runoff forecasting, the current state-of-the-art data-driven models, including MLR, FFBP-NN, SVR and ridge regression (RR), are selected for the comparison. MLR is a statistical approach to modeling the relationship between a dependent variable and independent variables for runoff forecasting; FFBP-NN is a traditional shallow neural network; SVR is based on statistical learning theory and the structural risk minimization hypothesis to achieve good robustness and efficiency; RR is put forward on the basis of general regression integrated forecast. Among them, MLR, FFBP-NN and SVR are three single-item prediction models; MLR is a linear model; and BP and SVR are nonlinear machine learning algorithms constructed based on different theories. The RR model is one of the traditional combination forecasting methods, which can solve the problem of coefficient selection of a single model for a linear combination prediction model.

Table 8 | The final selection results of input factors of different lag periods for Lianghekou and Jinping stations

Lag length (months)	Station	Selected factors
$T = 6$	Lianghekou	a2($t-1$), a3($t-1$), a29($t-1$), a36($t-1$), a42($t-1$), a43(t − 6), a96($t-1$)
	Jinping	a1($t-1$), a13($t-1$), a14($t-1$), a36($t-1$), a43($t-6$), a44($t-6$), a82($t-4$), a96($t-1$)
$T = 12$	Lianghekou	a2($t-1$), a31($t-7$), a36($t-1$), a42($t-12$), a43($t-6$), a44($t-6$), a64($t-11$), a96($t-12$)
	Jinping	a13($t-1$), a14($t-1$), a29($t-1$), a36($t-1$), a43($t-7$), a44($t-7$), a64($t-11$), a82($t-4$), a96($t-12$)
$T = 18$	Lianghekou	a2($t-1$), a3($t-1$), a31($t-7$), a36($t-1$), a43($t-18$), a44($t-18$), a64($t-11$), a82($t-16$)
	Jinping	a13($t-1$), a33($t-1$), a36($t-1$), a43($t-18$), a44($t-7$), a64($t-11$), a82($t-16$), a96($t-12$)
$T = 24$	Lianghekou	a1($t-24$), a3($t-24$), a14($t-24$), a31($t-7$), a36($t-1$), a43($t-18$), a44($t-18$), a64($t-11$), a82($t-16$), a96($t-24$)
	Jinping	a18($t-24$), a36($t-1$), a43($t-18$), a44($t-7$), a64($t-11$), a82($t-16$), a83($t-4$), a96($t-24$)

The different sections of BP and SVR are mainly shown in (1) the optimization goal of the BP algorithm which is based on the empirical risk minimization criterion to minimize the training error between the network output and the ideal output. It means that the curves and surfaces fitted by the BP neural network go through the training sample points as much as possible, which will cause the BP algorithm to rely too much on learning samples and has over-learning problems, so it is difficult to obtain good generalization ability. The SVR algorithm is based on the structural risk minimization criterion. In order to minimize the expected risk, the empirical risk and the confidence range should be minimized at the same time. In the SVR learning algorithm for regression, we construct a regression estimation function to minimize the VC dimension of the function on the premise that the distance from the target value is less than ε. That is, keeping the training error fixed and minimizing the confidence range, which solves the over-learning problem and has a better ability to generalize the samples. (2) The idea of the BP algorithm is that the forward propagation of the signal and the backpropagation of the error are carried out repeatedly, so that the weights are adjusted continuously, so as to ensure that the training error of the network output is minimized. In essence, it is an iterative learning algorithm based on gradient descent, which has some defects such as slow learning convergence and easy to fall into local minima. The SVR algorithm reduces the above regression estimation problem to a convex quadratic programming problem with linear equality constraints and linear inequality constraints, which can ensure the global optimality of the algorithm and effectively overcome the curse of dimensionality.

Models structure and parameter selection

To verify the versatility and university of the proposed combined method, the stations are used as experiments. The specific parameters of the two stations have some differences, in the development of FFBP-NN, the training function is 'tansig', the learning function is 'logsig', the maximum training times are 1,000, the learning rate is 0.1, the model training adopts LM algorithm, the momentum factor is 0.9 and the expected error is 0.001. For Lianghekou station, when the lag period is 6 months, networks with 2–15 hidden neurons were evaluated to determine the optimal network. The performance measure, RMSE, in training and testing, is demonstrated in Figure 5. The improvement of model performance in training was found due to the increase of hidden neurons, but the performance degrades when hidden neurons were larger than 6. Similar trends were also generally observed for the testing. Thus, the structure of FFBP-NN is 7-6-1 (neurons in the input layer-neurons in the hidden layer-neurons in the output layer). A similar procedure was implemented to decide optimal architectures for different lag periods. Optimal networks for a lag period of 12, 18 and 24 months are 8-7-1, 8-4-1 and 10-11-1, respectively. For the SVR model, the kernel function is 'sigmoid', and we use the GridSearch grid to select the best parameters C (regularization parameters) and gamma which the traversal gamma is in the interval [0.1, 2] and C is in the interval [0.5, 5] (Liang *et al.* 2020). When the lag period is 6, 12, 18 and 24 months, the parameters (C and gamma) of the SVR model are (3.0, 0.23), (0.5, 0.10), (0.5, 0.10) and (0.5, 0.10), respectively. In the development of ELM, we choose the number of

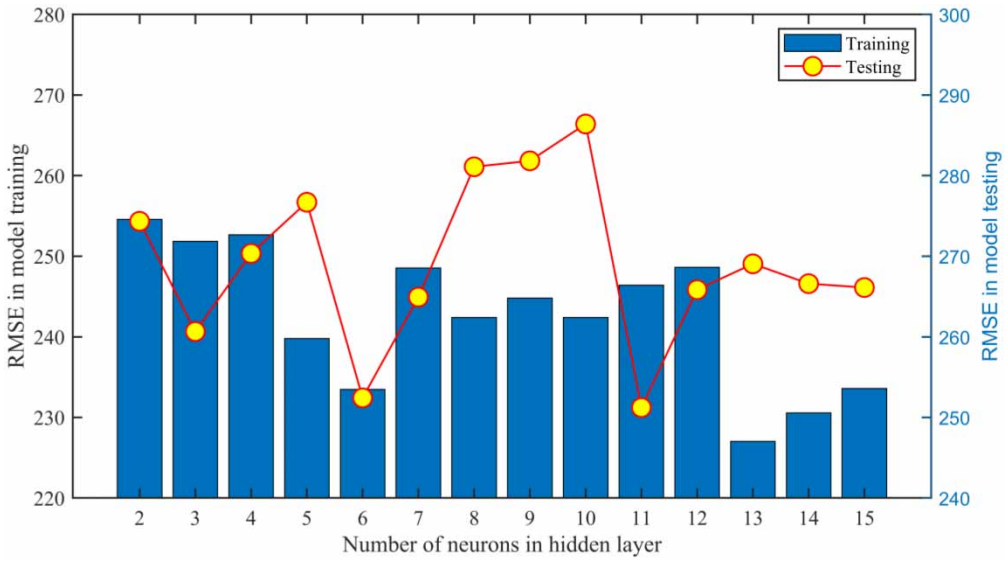


Figure 5 | Selection of hidden neurons in the lag period of 6 months at the Lianghekou station.

intermediate nodes in the same way as FFBP-NN. Thus, when the lag period is 6, 12, 18 and 24 months, the structure of ELM is 3-6-1, 3-7-1, 3-7-1 and 3-8-1, respectively; the combined prediction model selects the same structure as ELM.

A similar procedure was implemented to decide optimal architectures for the Jinping station. The optimal architecture of FFBP-NN, when the lag period is 6, 12, 18 and 24 months, is 8-10-1, 9-7-1, 8-8-1 and 8-11-1, respectively; the optimal parameters (C and gamma) of the SVR model are (0.5, 0.10), (5.0, 0.23), (1.0, 0.10) and (0.5, 0.10), respectively; the structure of ELM is 3-6-1, 3-6-1, 3-7-1 and 3-8-1, respectively; and the combined prediction model selects the same structure as ELM.

Training results

With the input variables and parameters selected, the runoff was forecasted using MLR, FFBP-NN and SVR, and then applied ELM to integrate the intermediate results of them. In addition, this experiment compares the presented combined model with the single models and the traditional combination forecasting method RR during the training period. The results of four evaluation metrics are listed in Table 9, and Table 9 shows the following:

(a) For the Lianghekou station, when the lag period is 6 months, compared with the other models, the presented combined model has the minimum MAE, MAPE and RMSE, with the values of 113.00, 0.1708 and 203.21 $m^3 \cdot s^{-1}$, respectively; the maximum DC value of 0.8646. Next, for the other three single models, SVR performs better than the other two models, and MLR performs the worst. When the lag period is 12, 18 and 24 months, the presented combined model with MAE values of 150.83, 155.64 and 148.09, respectively, is still the most accurate forecasting model. For the other three single models, there is not a unified law based on the performance metrics. For a lag period of 12 months, based on RMSE and DC, FFBP-NN performs better than the other two models; based on MAE and MAPE, SVR can get better performance. For a lag period of 18 months, based on MAE and MAPE, the order of the other three models from good to bad is SVR, MLR and FFBP-NN; based on RMSE and DC, the order of the other three models from good to bad is SVR, FFBP-NN and MLR. For a lag period of 24 months, in terms of four evaluation metrics, SVR can get better forecasting results than the other two models; among them, the performance of FFBP-NN is better based on MAE and MAPE, and the performance of MLR is better based on RMSE and DC. In the other two combination forecasting methods, ELM has a better fitting effect than the RR model from a lag period of 6–24 months.

(b) For the Jinping station, the presented combined model with MAPE values of 0.1364, 0.1948, 0.1963 and 0.1891, respectively, from a lag period of 6–24 months can get the most accurate forecasting results. For the other three single models, there is not a unified law based on the performance metrics. Based on MAE and MAPE, when the lag is 6–24 months, it is obvious that the value of SVR is lower than the other two single models. Based on RMSE and DC, when the lag period is 6 months, the order of the other three models from good to bad is SVR, FFBP-NN and MLR; when the lag period is 12, 18 and 24 months, FFBP-NN can get better performance, SVR is the second and MLR was the worst. For the other two combined forecasting methods, ELM performs better than the RR model when the lag is 6–24 months.

Runoff forecasting

This section mainly compares the presented combined model with other models in different lag stages of the forecast period during the test period, including the lag period of 6, 12, 18 and 24 months. The results of the four evaluation criteria are calculated in Table 10. From Table 10, the following result can be obtained:

(a) For the Lianghekou station, when the lag period is 6 months, compared with the other models, the presented combined model has the minimum MAE, MAPE and RMSE, with the values of 167.35, 0.2719 and 262.55 $m^3 \cdot s^{-1}$, respectively; the maximum DC value of 0.7946. Next, for the other three single models, there is not a unified law based on the performance metrics; based on MAE, FFBP-NN performs better than the other two models, and SVR gets the worst forecasting results; based on MAPE, FFBP-NN performs better than the other two models, MLR gets the worst forecasting results; and based on RMSE and DC, the order of the other three models from good to bad is FFBP-NN, MLR and SVR. When the lag period is 12, the presented combined model with an MAE value of 180.26 is still the most accurate forecasting model. For the other three single models, the SVR model has the minimum MAE, MAPE and RMSE, with the values of 181.83, 0.2865 and 308.31 $m^3 \cdot s^{-1}$, respectively, and the maximum DC value of 0.7168; the order of the other two models from good to bad is FFBP-NN and MLR. When the lag period is 18 and 24 months, it is very obvious that the proposed combined model gets the best forecasting results in terms of the four evaluation metrics. For the other three single models, their

Table 9 | Results of four evaluation metrics of the combined model and the three single models during the training period

		MAE				MAPE				RMSE ($m^3 \cdot s^{-1}$)				DC			
		Lag length (months)				Lag length (months)				Lag length (months)				Lag length (months)			
Station	Model	6	12	18	24	6	12	18	24	6	12	18	24	6	12	18	24
Lianghe kou	MLR	180.06	178.18	181.16	173.39	0.3201	0.2963	0.3087	0.2811	275.17	281.22	278.33	270.20	0.7517	0.7407	0.7460	0.7606
	FFBP-NN	142.03	162.34	184.51	167.89	0.2092	0.2369	0.3516	0.2462	238.50	266.25	272.32	271.60	0.8135	0.7675	0.7568	0.7581
	SVR	114.50	158.14	157.94	153.21	0.1712	0.2251	0.2301	0.2193	212.07	267.63	265.10	256.88	0.8525	0.7651	0.7696	0.7836
	RR	115.30	158.16	163.72	150.63	0.1918	0.2365	0.2747	0.2180	205.13	262.94	262.14	251.66	0.8620	0.7733	0.7747	0.7923
	ELM	113.28	155.15	160.33	150.72	0.1742	0.2253	0.2465	0.2209	204.35	260.76	259.70	251.84	0.8631	0.7777	0.7788	0.7920
	Proposed model	**113.00**	**150.83**	**155.64**	**148.09**	**0.1708**	**0.2231**	**0.2266**	**0.2108**	**203.21**	**253.98**	**253.10**	**239.03**	**0.8646**	**0.7885**	**0.7899**	**0.8127**
Jinping	MLR	298.23	314.45	311.00	300.52	0.2661	0.2831	0.2744	0.2461	471.79	491.42	481.99	471.58	0.7936	0.7761	0.7846	0.7938
	FFBP-NN	232.51	273.22	286.76	260.91	0.2175	0.2275	0.2457	0.2134	387.89	452.02	452.23	418.35	0.8605	0.8106	0.8104	0.8377
	SVR	162.68	270.99	269.63	258.92	0.1409	0.2045	0.2029	0.1975	326.60	466.58	460.84	436.41	0.9011	0.7982	0.8031	0.8234
	RR	165.89	269.05	271.61	249.07	0.1421	0.2239	0.2279	0.1943	318.76	446.99	438.13	407.66	0.9058	0.8148	0.8220	0.8459
	ELM	164.96	259.53	272.78	245.77	0.1383	0.1975	0.2253	0.1965	319.17	443.08	435.21	403.26	0.9056	0.8180	0.8244	0.8492
	Proposed model	**161.82**	**253.75**	**256.19**	**240.56**	**0.1364**	**0.1948**	**0.1963**	**0.1891**	**305.72**	**438.26**	**419.00**	**394.69**	**0.9134**	**0.8219**	**0.8372**	**0.8556**

Table 10 | Results of four evaluation criteria of the combined model and the three single models during the testing period

		MAE				MAPE				RMSE				DC			
		Lag length (months)				Lag length (months)				Lag length (months)				Lag length (months)			
Station	Model	6	12	18	24	6	12	18	24	6	12	18	24	6	12	18	24
Lianghekou	MLR	190.72	202.75	196.02	202.48	0.3568	0.3807	0.3480	0.3486	282.80	325.89	308.11	316.05	0.7617	0.6836	0.7172	0.7024
	FFBP-NN	172.32	197.68	229.33	251.46	0.2721	0.3134	0.5083	0.5096	273.12	319.90	328.86	366.48	0.7777	0.6951	0.6778	0.5998
	SVR	192.67	181.83	184.76	190.14	0.2936	0.2865	0.3072	0.3008	306.98	308.31	305.04	310.36	0.7192	0.7168	0.7228	0.7130
	RR	211.40	188.83	201.32	227.36	0.3573	0.3014	0.3838	0.4429	327.24	308.78	309.88	333.78	0.6809	0.7159	0.7139	0.6681
	ELM	208.85	189.47	187.84	226.16	0.3522	0.2988	0.3284	0.3646	321.05	311.08	303.79	365.92	0.6929	0.7117	0.7250	0.6710
	Proposed model	**167.35**	**180.26**	**169.32**	**193.49**	**0.2719**	**0.2745**	**0.2773**	**0.3255**	**262.55**	**298.34**	**290.95**	**306.65**	**0.7946**	**0.7348**	**0.7478**	**0.7198**
Jinping	MLR	298.18	316.76	286.40	306.39	0.2624	0.2971	0.2551	0.2668	468.11	501.48	498.95	505.16	0.7872	0.7558	0.7582	0.7522
	FFBP-NN	288.64	311.90	295.83	395.12	0.2108	0.2459	0.2140	0.3606	522.96	549.86	522.11	632.33	0.7344	0.7064	0.7353	0.6117
	SVR	284.93	268.17	257.93	278.34	0.2239	0.2182	0.1871	0.2113	487.86	476.41	468.00	504.84	0.7689	0.7796	0.7873	0.7525
	RR	297.25	288.84	265.33	342.34	0.2323	0.2384	0.2055	0.2908	513.06	508.83	478.94	572.53	0.7444	0.7486	0.7772	0.6817
	ELM	295.00	286.51	268.28	324.86	0.2306	0.2261	0.1872	0.2570	512.20	524.54	470.78	570.65	0.7452	0.7328	0.7773	0.6960
	Proposed model	**249.89**	**252.42**	**250.12**	**252.63**	**0.1843**	**0.1922**	**0.1870**	**0.2169**	**434.76**	**464.28**	**466.93**	**440.47**	**0.8164**	**0.7907**	**0.7883**	**0.8116**

order from good to bad is SVR, MLR and FFBP-NN based on the four evaluation metrics. In the other two combination forecasting methods, ELM has a better fitting effect than the RR model from a lag period of 6–24 months.

(b) For the Jinping station, when the lag period is 6 months, the proposed combined model with MAE, MAPE, RMSE and DC values of 249.89, 0.1843, 434.76 $m^3 \cdot s^{-1}$ and 0.8164, respectively, can get the most accurate forecasting results. Next, for the other three single models, there is not a unified law based on the performance metrics; based on MAE, SVR performs better than the other two models; in terms of the MAPE, FFBP-NN can get better forecasting results; and based on RMSE and DC, it is very obvious that the value of MLR is better than the other two single models. When the lag period is 12, the presented combined model gets the best forecasting results compared with the other three single models based on the four evaluation metrics. For the other three single models, there is not a uniform regularity based on the performance metrics; in terms of the MAE and MAPE, their order from good to bad is SVR, FFBP-NN and MLR; and based on the RMSE and DC, it is very obvious that the value of SVR is better than the other two single models, MLR followed, and FFBP-NN is the worst. For a lag period of 18 and 24 months, the proposed combined model with values of 0.1870 and 0.2169, respectively, is still the most accurate forecasting model. Next, for the other three single models, there is not a unified law based on the performance metrics; in terms of the MAE, RMSE and DC, their order from good to bad is SVR, FFBP-NN and MLR; based on the MAPE, in the lag period of 18 months, SVR performs better than the other two models, FFBP-NN followed, and MLR is the worst; and for a lag period of 24 months, their order from good to bad is SVR, MLR and FFBP-NN. For the other two combined forecasting methods, ELM performs better than the RR model when the lag is 6–24 months.

Performance of uncertainty and reliability analyses

To evaluate the performance of the proposed approach, the results of the above two experiments will be further discussed with uncertainty and reliability analyses.

Uncertainty analysis

The major goal of the uncertainty analysis is to restrict the expected range in which the true value of the outcome of an experiment lies. This estimated range is in the form of an interval and is called the uncertainty interval. It can be estimated based on the errors calculated for the measurement process of the experiment under consideration. U95 is a type of uncertainty analysis procedure to compute the uncertainty interval (Sabed-Movahed *et al.* 2020). To be more specific, U95 restricts the uncertainty of runoff at a 95% confidence level. Thus, the smaller the amount of U95, the more accurate the runoff value. It is defined as follows:

$$\mathrm{U95} = \frac{1.96}{n} * \sqrt{\sum_{t=1}^{n} (y_t - \hat{y}_t)^2 + \sum_{t=1}^{n} (y_t - \bar{y}_t)^2} \tag{11}$$

Reliability analysis

The reliability analysis is a statistical method for measuring the overall consistency of a model. It is designed based on the amount of random error from the measurement process. The greater the number of cases for which error is less than a certain threshold, the more reliable the overall consistency of the model will be. Generally, a metric for the reliability analysis is represented by the following equation:

$$\mathrm{Reliability} = \frac{\sum_{t=1}^{n} l_t}{n} * 100\% \tag{12}$$

where k_t is obtained through two steps. (1) The relative average error (RAE) is defined as a vector whose tth component is given as follows:

$$\mathrm{RAE}_t = \left| \frac{y_t - \hat{y}_t}{y_t} \right|$$

(2) If $RAE_t \leq \Delta$, then $l_t = 1$; otherwise, $l_t = 0$, where Δ is the threshold value of runoff forecast. In other words, k_t is defined as the number of times the value of RAE is less than or equal to that of Δ. The optimum value of Δ based on Chinese standards is 0.2 or equivalently is 20%. In this article, we take 20%.

Results of benchmarks analysis

The results of uncertainty and reliability analyses are presented in Table 11. From Table 11, the following result can be obtained:

(a) For the Lianghekou station, the proposed model represented the lowest value for uncertainty with a 95% confidence level, with uncertainty values of **52.6423**, 54.3781, 54.3454 and 53.8330, respectively, from a lag period of 6–24 months in the training phase when compared to the other models. Additionally, predictions of runoff provided by the proposed model were more reliable in comparison with the predictions made by the other models, with reliability values of **68.83**, 58.25, 56.67 and 62.71% from a lag period of 6–24 months, respectively. Moreover, the proposed combined model with a 6-month lag can provide better performance than other lag periods. Similar results were also generally observed for the testing stage, that is, the proposed model had the lowest uncertainty and the highest reliability when the lag is 6–24 months (see the bold values in Table 11).

(b) For the Jinping station, when compared with the other models, the presented model showed the lowest uncertainty, with U95 values of **96.8543**, 100.8462, 100.1888 and 99.3955, respectively, from a lag period of 6–24 months in the training phase. Additionally, considering reliability, it is conspicuous that the proposed model, with reliability values of **83.50**,

Table 11 | Comparisons of the performance results for the uncertainty and reliability analyses

		U95				Reliability (%)			
		Lag length (months)				Lag length (months)			
Station	Model	6	12	18	24	6	12	18	24
Lianghekou	Training datasets								
	MLR	55.1973	55.4406	55.3238	55.0002	37.92	45.42	42.50	44.58
	FFBP-NN	53.8142	54.8461	55.0840	55.0556	59.38	55.63	37.71	47.08
	SVR	52.9215	54.8996	54.8013	54.4873	63.75	57.08	57.29	60.00
	RR	52.7020	54.7179	54.6872	54.2918	64.58	54.17	46.67	61.04
	ELM	52.6777	54.6346	54.5940	54.2987	68.54	58.13	55.83	61.04
	Proposed model	**52.6423**	**54.3781**	**54.3454**	**53.8330**	**68.83**	**58.25**	**56.67**	**62.71**
	Testing datasets								
	MLR	115.3462	118.9300	117.4029	118.0773	37.50	36.67	44.17	40.00
	FFBP-NN	114.5970	118.4087	119.1918	122.6546	45.00	45.00	22.50	25.00
	SVR	117.3081	117.4198	117.1465	117.5925	45.83	47.50	45.17	43.17
	RR	119.0485	117.4596	117.5518	119.6284	36.67	47.50	35.83	30.00
	ELM	118.5077	117.6538	117.0422	122.6010	42.50	45.00	43.33	40.83
	Proposed model	**113.8033**	**116.5922**	**115.9934**	**117.2807**	**46.33**	**48.17**	**46.17**	**44.50**
Jinping	Training datasets								
	MLR	102.0498	102.7886	102.4304	102.0419	42.92	45.00	44.17	48.75
	FFBP-NN	99.1813	101.3313	101.3384	100.1671	55.00	55.63	52.71	57.29
	SVR	97.3983	101.8581	101.6487	100.7819	80.63	59.38	61.04	60.00
	RR	97.1902	101.1524	100.8418	99.8137	76.46	57.08	55.21	63.33
	ELM	97.2009	101.0146	100.7403	99.6705	81.04	63.33	53.54	61.67
	Proposed model	**96.8543**	**100.8462**	**100.1888**	**99.3955**	**83.50**	**63.54**	**62.92**	**64.58**
	Testing datasets								
	MLR	199.9499	202.5241	202.3238	202.8165	49.17	45.83	50.00	50.83
	FFBP-NN	204.2553	206.5045	204.1855	213.9284	67.50	60.00	63.33	40.00
	SVR	201.4559	200.5770	199.9422	202.7911	63.33	65.00	65.83	55.00
	RR	203.4498	203.1096	200.7696	208.4674	63.33	58.33	64.17	50.00
	ELM	203.3838	204.3848	201.6053	208.9196	60.00	60.00	69.17	55.00
	Proposed model	**197.5251**	**199.664**	**199.862**	**197.9293**	**77.50**	**70.00**	**69.27**	**58.33**

63.54, 62.92 and 64.58% from a lag period of 6–24 months, respectively, was the most reliable. Moreover, the proposed combined model with a 6-month lag was able to provide very good performance compared with the other lag periods. In the testing stage, there were also generally observed similar results, that is, the proposed model had the lowest uncertainty and the highest reliability when the lag is 6–24 months (see the bold values in Table 11). And, the effect of a lag period of 6 months was the best in terms of different lag periods.

DISCUSSION

Developing medium- and long-term runoff forecasting for water resource planning and management activities, such as water conservancy infrastructure operation, flood control and reservoir operation, is of great significance. In view of the influence of different lag periods on runoff and the uncertainty of a single model, a medium- and long-term runoff combined forecasting model based on different lag periods is proposed in this paper.

The experimental results showed that the input factors of different lag periods affect the prediction accuracy of runoff based on the analysis of the above results. In addition, the optimal lag period of the Lianghekou is 6 months, and the factors (Table 12) affecting the change of runoff process are mainly related to the lag time of 1 month of the Area Index of Northern Hemisphere Subtropical High (5E-360), 1 month of the North Africa Subtropical High-Intensity Index (20 W–60E), 1 month of the North Africa, Atlantic and North America Subtropical High-Intensity Index (110 W–60E), 1 month of the Central Intensity of the Northern Hemisphere Polar Vortex (JQ), 1 month of the Tibet Plateau (25N–35N, 80E–100E), 6 months of the Tibet Plateau (30N–40N, 75E–105E) and 1 month of the antecedent runoff. For the Jinping station, the optimal lag period is also 6 months. The factors affecting the change of runoff are shown in Table 12, which are no longer listed here. Among them, the Central Strength of the Northern Hemisphere Polar Vortex (JQ), Tibet Plateau (25N–35N, 80E–100E) and the antecedent runoff (see the bold values in Table 12) are three identical features of Lianghekou and Jinping stations. This indicates that we consider not only the influence of atmospheric circulation on runoff, but also the special climatic impact caused by the geographical location of the Yalong River, such as snow cover on the Tibet Plateau (Xu & Ma 2011).

The experimental results indicate that the proposed combined model performs better than other methods whether it's during training or testing. This is mainly because the classic MLR model is relatively easy to construct with the simplest type of parameters, and it can capture the global trend over an entire input space. Its accuracy, however, is not satisfactory, which may not meet the requirements of medium- and long-term runoff forecasting. The BP model can identify complex nonlinear relationships between input and output data, and its accuracy is satisfactory for runoff forecasting, but there is a risk of over-fitting. The SVR model is also appropriate for reproducing the nonlinear problem, which can provide a suitable mapping between input and output data in a higher-dimensionality feature space to improve the forecasting accuracy. Its parameters

Table 12 | The best input factor in the lag period of 6 months for Lianghekou and Jinping stations

Station	Lag time (months)	Predictive factors	Definition and interpretation
Lianghekou	$t-1$	a2	Area Index of North Africa Subtropical High (20 W – 60E)
	$t-1$	a3	North Africa, Atlantic and North America Subtropical High Area Index (110 W – 60E)
	$t-1$	a29	Polar Vortex Strength Index in Asia Region (Region 1, 60E – 150E)
	$t-1$	**a36**	**Central Strength of the Northern Hemisphere Polar Vortex (JQ)**
	$t-1$	a42	East Asian Trough Strength (CQ)
	$t-6$	**a43**	**Tibet Plateau (25N – 35N, 80E–100E)**
	$t-1$	**a96**	**The antecedent runoff**
Jinping	$t-1$	a1	Area Index of the Northern Hemisphere Subtropical High (5E-360)
	$t-1$	a13	North Africa Subtropical High Intensity Index (20 W–60E)
	$t-1$	a14	North Africa, Atlantic and North America Subtropical High Intensity Index (110 W–60E)
	$t-1$	**a36**	**Central Strength of the Northern Hemisphere Polar Vortex (JQ)**
	$t-6$	**a43**	**Tibet Plateau (25N–35N, 80E–100E)**
	$t-6$	a44	Tibet Plateau (30N–40N, 75E–105E)
	$t-4$	a82	IOWPA Indian Ocean Warm Pool Area Index
	$t-1$	**a96**	**The antecedent runoff**

need to be determined carefully since they significantly influence the accuracy of the SVR model (Chu *et al.* 2017). The RR model is one of the traditional combination forecasting methods, which can solve the problem of coefficient selection of a single model for a linear combination prediction model. The possible emergence of negative weights and other controversial issues restrict the generalization of linear combination prediction to a certain extent (Xu *et al.* 2021). However, the proposed combined model is easy to fit the nonlinearity of the predicted objects and meet the requirements of medium- and long-term runoff forecasting, which overcome the limitation of linear combination prediction.

Based on the accomplished findings, there are still some areas that need to be improved. In future research, more factors such as vegetation index, topography and electricity generation, especially the humans' activity, will be introduced to better describe the changing process of runoff. In the method of selecting key factors, we concentrate on only the linear relationship and neglect the nonlinear relationship between the impact factors. Thus, we will adopt the PMI method to select the key factors in the future study, which has the advantages of characterizing relationships of linear and nonlinear among factors. In addition, the deep learning approaches will be applied to runoff prediction because the shallow network is degenerated for the medium- and long-term runoff forecasting task.

CONCLUSION

With the rapid development of cascade hydropower stations in a big basin, the actual operation is becoming more and more in-demand for runoff prediction. Therefore, it is of great significance to develop medium- and long-term runoff forecasting for water resource planning and management activities such as water conservancy infrastructure operation, flood control, reservoir operation and drinking water distribution. Numerous studies that adopted combined forecasting models to enhance runoff forecasting accuracy have been proposed. Nevertheless, some models do not consider the effects of different lag periods on the selection of input factors.

Thus, this paper proposed a medium- and long-term runoff combined forecasting model based on different lag periods. The presented combined model initially uses the lagging correlation coefficient and stepwise regression to extract features as the optimal input factor. Then, the selected optimal factor is predicted by using three common single models, which include MLR, FFBP-NN and SVR. And then ELM is utilized to integrate the forecasting results of the three individual models. In addition, considering that the random input weights and hidden biases of ELM always have some influence on the training process, we use PSO to optimize the parameters. By simulating and experimenting with the runoff data of Lianghekou and Jinping stations from the Yalong River basin, and comparing the proposed combined method with the three individual forecasting methods and the traditional combination forecasting methods, the major findings of the proposed approaches and their applications in this paper are: (1) the lag period of physical factors delay can affect the accuracy of runoff forecasting. And, the optimal lag period of both the Lianghekou and Jinping stations is 6 months. (2) Whether it is during the training or testing, the proposed combined model performs better than the other three individual models and the traditional combination method. (3) The proposed combination model can be considered as a very reliable forecasting tool for runoff forecasting.

ACKNOWLEDGEMENTS

This work was supported by 'the key projects supported by the Major Research Plan of the National Natural Science Foundation of China' (Grant No. 91846203) and 'the school research fund of Nanjing Vocational University of Industry Technology' (Grant No. YK21-05-05).

AUTHOR CONTRIBUTIONS

A.P. is involved in conceptualization, writing – reviewing and editing, and funding acquisition. S.Y. is involved in the writing – original draft, methodology and formal analysis. X.C. is involved in the methodology and data curation. C.B. is involved in the visualization and data curation. Y.Z. is involved in the supervision and resources, and funding acquisition.

CONFLICT OF INTEREST

The authors declare no conflict of interest.

DATA AVAILABILITY STATEMENT

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

REFERENCES

Barzkar, A., Najafzadeh, M. & Homaei, F. 2021 Evaluation of drought events in various climatic conditions using data-driven models and a reliability-based probabilistic model. *Natural Hazards* 1–22. doi:10.1007/s11069-021-05019-7.

Chen, Y., He, Z., Shang, Z., Li, C., Li, L. & Xu, M. 2019 A novel combined model based on echo state network for multi-step ahead wind speed forecasting: a case study of NREL. *Energy Conversion and Management* **179**, 13–29. doi:10.1016/j.enconman.2018.10.068.

Cheng, Q., Zuo, X., Zhong, F., Gao, L. & Xiao, S. 2019 Runoff variation characteristics, association with large-scale circulation and dominant causes in the Heihe River Basin, Northwest China. *Science of the Total Environment* **688**, 361–379. doi:10.1016/j.scitotenv.2019.05.397.

Chu, H., Wei, J., Li, J., Qiao, Z. & Cao, J. 2017 Improved medium- and long-term runoff forecasting using a multimodel approach in the Yellow River headwaters region based on large-scale and local-scale climate information. *Water* **9** (8), 1–16. doi:10.3390/w9080608.

Ekwueme, B. N. & Agunwamba, J. C. 2020 Modeling the influence of meteorological variables on runoff in a tropical watershed. *Civil Engineering Journal* **6** (12), 2344–2351. doi:10.28991/cej-2020-03091621.

Gao, Z. C. 2006 GM(1, 2) time-lag model of mid-long term runoff forecasting on Lanhe River. *Journal of Taiyuan University of Technology* **37**, 71–73. https://doi.org/10.16355/j.cnki.issntyut.2006.01.021.

He, J., Valeo, C., Chu, A. & Neumann, N. F. 2011 Prediction of event-based stormwater runoff quantity and quality by ANNs developed using PMI-based input selection. *Journal of Hydrology* **400** (1-2), 10–23. doi:10.1016/j.jhydrol.2011.01.024.

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), 1113–1130. doi:10.3390/w11061113.

Liang, H., Huang, S., Meng, E. & Huang, Q. 2020 Runoff prediction based on multiple hybrid models. *Journal of Hydraulic Engineering* **51**, 112–125. https://doi.org/10.13243/j.cnki.slxb.20190434.

Lima, A. R., Cannon, A. J. & Hsieh, W. W. 2016 Forecasting daily streamflow using online sequential extreme learning machines. *Journal of Hydrology* **537**, 431–443. doi:10.1016/j.jhydrol.2016.03.017.

Liu, T. F., Ding, Y. S., Cai, X., Zhu, Y. F. & Zhang, X. F. 2017 extreme learning machine based on particle swarm optimization for estimation of reference evapotranspiration. In: *Paper Presented at the Proceedings of the 36th Chinese Control Conference (CCC 2017)*.

Lu, D. & Zhou, H. 2014 Medium and long-term runoff forecasting based on mutual information and BP neural network. *Journal of China Hydrology* **34** (4), 8–14, 67. doi:10.3969/j.issn.1000-0852.2014.04.002.

Lu, P., Lin, K., Xu, C.-Y., Lan, T., Liu, Z. & He, Y. 2021 An integrated framework of input determination for ensemble forecasts of monthly estuarine saltwater intrusion. *Journal of Hydrology* **598**, 1–15. doi:10.1016/j.jhydrol.2021.126225.

Lyu, H., Wan, M., Han, J., Liu, R. & Wang, C. 2017 A filter feature selection method based on the maximal information coefficient and Gram-Schmidt orthogonalization for biomedical data mining. *Computers in Biology and Medicine* **89**, 264–274. doi:10.1016/j.compbiomed.2017.08.021.

Motahari, M. & Mazandaranizadeh, H. 2017 Development of a PSO-ANN model for rainfall-runoff response in basins, case study: Karaj Basin. *Civil Engineering Journal* **3**, 35–44. doi:10.28991/cej-2017-00000070.

Najafzadeh, M. & Sattar, A. M. A. 2015 Neuro-fuzzy GMDH approach to predict longitudinal dispersion in water networks. *Water Resources Management* **29** (7), 2205–2219. doi:10.1007/s11269-015-0936-8.

Najafzadeh, M., Homaei, F. & Mohamadi, S. 2022 Reliability evaluation of groundwater quality index using data-driven models. *Environmental Science And Pollution Research* **29**, 8174–8190. doi:10.1007/s11356-021-16158-6.

Sabed-Movahed, F., Najafzadeh, M. & Mehrpooya, A. 2020 Receiving more accurate predictions for longitudinal dispersion coefficients in water pipelines: training group method of data handling using extreme learning machine conceptions. *Water Resources Management* **34** (2), 529–561. doi:10.1007/s11269-019-02463-w.

Sahoo, B. B., Jha, R., Singh, A. & Kumar, D. 2018 Application of support vector regression for modeling low flow time series. *KSCE Journal of Civil Engineering* **23** (2), 923–934. doi:10.1007/s12205-018-0128-1.

Sang, Y., Wang, Z. & Liu, C. 2013 Research progress on the time series analysis methods in hydrology. *Progress in Geography* **32**, 20–30. doi:10.3724/SP.J.1033.2013.00020.

Schilling, K. E. & Wolter, C. F. 2005 Estimation of streamflow, base flow, and nitrate-nitrogen loads in IOWA using multiple linear regression models. *Journal of the American Water Resources Association* **41**, 1333–1346. doi:10.1111/j.1752-1688.2005.tb03803.x.

Shan, L., Zhang, L., Liu, L. & Jia, J. 2015 Research on low-flow runoff forecasting based on multimethod optimized selection factors and artificial neural network coupled model. *Engineering Journal of Wuhan University* **48** (06), 758–763. doi:10.14188/j.1671-8844.2015-06-004.

Valipour, M. 2015 Long-term runoff study using SARIMA and ARIMA models in the United States. *Meteorological Applications* **22** (3), 592–598. doi:10.1002/met.1491.

Valipour, M., Banihabib, M. E. & Behbahani, S. M. R. 2013 Comparison of the ARMA, ARIMA, and the autoregressive artificial neural network models in forecasting the monthly inflow of Dez dam reservoir. *Journal of Hydrology* **476**, 433–441. doi:10.1016/j.jhydrol.2012.11.017.

Wang, Q., Zhang, T., You, H., Chang, J. & Cao, Y. 2014 Mid-long term runoff forecasting model for Dahuofang reservoir based on multiple regression analysis. *Hydroelectric Power* **40** (5), 17–20. doi:10.13243/j.cnki.slxb.20190434.

Wang, W., Li, W., Xu, D. & Li, Q. 2019 Runoff prediction based on GM-BP model calibration against Markov chain. *South-to-North Water Transfers and Water Science & Technology* **17** (5), 44–49. doi:10.3969/j.issn.0559-9342.2014.05.005.

Wei, T., Ma, G. & Huang, W. 2012 Runoff forecast based on weighted support vector machine regression model. *Journal of Hydroelectric Engineering* **31** (6), 35–38 + 43. doi:CNKI:SUN:SFXB.0.2012-06-007.

Xu, L. & Ma, G. 2011 Monthly runoff forecast of Ertan hydropower station based on meteorological analysis. *Transactions of the CSAE* **27** (S2), 80–85. https://doi.org/10.3969/j.issn.1002-6819.2011.z2.016.

Xu, S., Alturki, R., Rehman, A. U. & Tariq, M. U. 2021 BP neural network combination prediction for big data enterprise energy management system. *Mobile Networks & Applications* **26** (1), 184–190. doi:10.1007/s11036-020-01698-x.

Yang, L., Tian, F. & Hu, H. 2013 Modified ESP with information on the atmospheric circulation and teleconnection incorporated and its application. *Journal of Tsinghua University (Science and Technology)* **53**, 606–612. doi:10.16355/j.cnki.

Yaseen, Z. M., Sulaiman, S. O., Deo, R. C. & Chau, K.-W. 2019 An enhanced extreme learning machine model for river flow forecasting: state-of-the-art, practical applications in water resource engineering area and future research direction. *Journal of Hydrology* **569**, 387–408. doi:10.1016/j.jhydrol.2018.11.069.

Yue, Z., Ai, P., Xiong, C., Hong, M. & Song, Y. 2020a Mid- to long-term runoff prediction by combining the deep belief network and partial least-squares regression. *Journal of Hydroinformatics* **22** (5), 1283–1305. doi:10.2166/hydro.2020.022.

Yue, Z., Ai, P., Yuan, D. & Xiong, C. 2020b Ensemble approach for mid-long term runoff forecasting using hybrid algorithms. *Journal of Ambient Intelligence and Humanized Computing* (2), 1–20. doi:10.1007/s12652-020-02345-9.

Zhao, T., Pan, S., Gao, W., Qing, Z., Yang, X. & Wang, J. 2021 Extreme learning machine-based spherical harmonic for fast ionospheric delay modeling. *Journal of Atmospheric and Solar-Terrestrial Physics* **216**, 1–10. doi:10.1016/j.jastp.2021.105590.

First received 17 August 2021; accepted in revised form 7 February 2022. Available online 21 February 2022