



Monthly runoff forecasting based on LSTM–ALO model

Xiaohui Yuan^{1,2} · Chen Chen¹ · Xiaohui Lei³ · Yanbin Yuan⁴ · Rana Muhammad Adnan¹

Published online: 22 May 2018

© Springer-Verlag GmbH Germany, part of Springer Nature 2018

Abstract

Accurate runoff forecasting plays an important role in management and utilization of water resources. This paper investigates the accuracy of hybrid long short-term memory neural network and ant lion optimizer model (LSTM–ALO) in prediction of monthly runoff. As the parameters of long short-term memory neural network (LSTM) have influence on the prediction performance, the parameters of the LSTM are calibrated by using ant lion optimizer. Then the selection of suitable input variables of the LSTM–ALO is discussed for monthly runoff forecasting. Finally, we decompose root mean square error into three parts, which can help us better understanding the origin of differences between the observed and predicted runoff. To test the merits of the LSTM–ALO for monthly runoff forecasting, other models are employed to compare with the LSTM–ALO. The scatter-plots and box-plots are adopted for evaluating the performance of all models. In the case study, simulation results with the historical monthly runoff of the Astor River Basin show that the LSTM–ALO model has higher accuracy than that of other models. Therefore, the proposed LSTM–ALO provides an effective method for monthly runoff forecasting.

Keywords Monthly runoff forecasting · Long short-term memory neural network · Ant lion optimizer · Errors decomposition

1 Introduction

Runoff forecasting is extremely important for various activities of water resources planning and management, such as operation of water infrastructures, flood control, dam planning, reservoir operation, distribution of drinking water and planning for navigation. Among them, monthly runoff prediction have attracted more attentions because

the storage-yield sequences are generally related to monthly periods. Meanwhile, long-term forecasting based on monthly time scales is useful in reservoir operations and irrigation management decisions (Chen et al. 2014; Liang et al. 2017; Yuan et al. 2015a).

Since the nonlinear characteristics of runoff time series, it is very challenging to predict runoff accurately. In the past decades, researchers have devoted efforts to monthly runoff prediction. Generally, the methods of monthly runoff prediction can be divided into two categories: process-driven methods and data-driven methods. Process-driven methods are composed of experimental formulae based on the physical phenomena (Yuan et al. 2014). While these methods can not reflect the stochastic volatility of medium- and long-term runoff series. Therefore, data-driven methods are used as substitute or supplement to process-based ones in the field of runoff forecasting. Data-driven methods are fundamentally black-box methods, which are convenient as they can be applied easily without considering the internal physical mechanism of watershed system (Adamowski and Sun 2010). The multiple linear regression (MLR) was adopted to predict the runoff of the Ogoki River located in Province of Ontario, Canada (Seidou and

✉ Xiaohui Yuan
yxh71@163.com

✉ Xiaohui Lei
leixiaohui_sky@163.com

¹ School of Hydropower and Information Engineering, Huazhong University of Science and Technology, Wuhan 430074, China

² College of Electrical Engineering and New Energy, China Three Gorges University, Yichang 443002, China

³ State Key Laboratory of Simulation and Regulation of Water Cycle in River Basin, China Institute of Water Resources and Hydropower Research, Beijing 100038, China

⁴ School of Resource and Environmental Engineering, Wuhan University of Technology, Wuhan 430070, China

Ouarda 2007). Anderson et al. (2013) applied autoregressive moving average (ARMA) based Gaussian function to get the prediction intervals of monthly runoff. In order to overcome the shortcomings of MLR and ARMA models for runoff forecasting, such as difficulty of identification model's parameters and not easily capturing nonlinear characteristics of runoff time series (Kong et al. 2015; Yaseen et al. 2015), artificial neural network (ANN), adaptive network-based fuzzy inference system (ANFIS) and support vector machine (SVM) are applied to runoff forecasting. Wang et al. (2006) selected hybrid ANNs as univariate time series models to forecast runoff series. Kalteh (2013) employed the ANN based wavelet transform to improve the accuracy of monthly runoff forecasting. El-Shafie et al. (2007) put forward ANFIS model for inflow forecasting of the Nile river at Aswan high dam. In order to improve the performance of the ANFIS model, combination of ANFIS with loading simulation program C++ (LSPC) model was proposed to forecast runoff (Sharma et al. 2015). The particle swarm optimization (PSO) based parameters of SVM was presented to forecast monthly runoff (Sudheer et al. 2013). Application of wavelet analysis in improving monthly streamflow forecasting performance of the SVM was studied by Kalteh and Chen (2014). To solve the complex computation in SVM, least squares support vector machine (LSSVM) was proposed to predict runoff series in the Tahtali and Gordes Watersheds (Okkan and Serbes 2012). Kisi (2015) investigates the ability of LSSVM and ANFIS in forecasting monthly streamflow. In the above mentioned methods, ANNs depend on the structure of neural networks and exist over-fitting phenomenon. ANFIS is hard to handle noise. SVM and LSSVM are difficult to determine the parameters of kernel function in runoff forecasting.

Recently, long short-term memory (LSTM) neural networks has been developed (Hochreiter and Schmidhuber 1997), which has merits of nonlinear predictive capability, faster convergence and capturing long-term correlation of time series. Therefore, the LSTM model has been successfully applied to speed recognition, handwriting recognition, voice recognition, traffic flow forecasting and so on (Greff et al. 2016; Bukhari et al. 2017). The previous researches have demonstrated that the monthly runoff time series show the characteristics of long-term memory, so this study applies the LSTM model to prediction of monthly runoff series. To further improve the forecasting accuracy, a hybrid model (LSTM–ALO) is put forward by combining the LSTM with ant lion optimizer (ALO). In the proposed LSTM–ALO model, the ALO is utilized to optimize the number of hidden layer neurons and learning rate of the LSTM. To the best of the authors' knowledge, this is the first time to investigate the application of ALO in parameters tuning of the LSTM model. The LSTM–ALO

not only inherits the advantages of the LSTM, but also utilizes the superiority of the ALO for solving optimization problems, where the process of population evolution can interact with the LSTM in the training process, and automatically calibrate the parameters of the LSTM model. Therefore, the LSTM–ALO can both solve the problem that the optimal parameters are difficult to be calibrated in the LSTM modeling and further improve the prediction performance of the LSTM model. The original LSTM, LSTM–ALO, and hybrid of LSTM and PSO (LSTM–PSO) are compared to verify the feasibility and effectiveness of the ALO for optimizing the parameters of the LSTM model. At the same time, the ALO is adopted to estimate the parameters of Back Propagation (BP) Neural Network, Radial Basis Neural Network (RBNN) and LSSVM respectively, which can further validate the superiority of the LSTM–ALO in monthly runoff forecasting. The monthly runoff data of the Astor River Basin in Northern Pakistan is chosen to test the prediction effect of the LSTM–ALO model. For better understanding the origin of the errors between prediction and observed runoff series, root mean square error (*RMSE*) is decomposed into three parts: bias, the amplitude error and the phase error. Finally, scatter-plots and box-plots are adopted to show monthly runoff forecasting of the LSTM–ALO more intuitively.

The rest of the paper is organized as follows. Section 2 overviews the basic principle of LSTM model. Section 3 briefly presents the main steps of ALO. Section 4 describes the process of the runoff forecasting with the LSTM–ALO model. Empirical results are reported in Sect. 5. Section 6 gives the conclusions.

2 Overview of long short-term memory neural network

The LSTM model belongs to a kind of deep learning neural network, which involves input layer, hidden layer, recurrent layer, and output layer. In the LSTM model, the output of the hidden layer fed back to the hidden layer through a buffer layer is called the recurrent layer. A special structural unit in the recurrent layer—memory block can make the input layer, the hidden layer and the output layer interact with each other. Therefore, the recurrent layer makes the LSTM model to supervise learning about time and spatial information.

The memory block is consisted of one or more connection memory cells and three multiplier units (input gate, forget gate and output gate) (Hochreiter and Schmidhuber 1997). Three gate units have the same input information, and each of them has a different activation function. When the input gate is high-active, the input information will be stored in the memory cell. While the output gate is high-

active, the information stored in the memory cell is released to the next neuron. If the forget gate is high-active, the information will be deleted from the memory cell. Since there is a memory block in LSTM, where errors can flow back forever, and making error flow outside the cell tends to decay exponentially (Gers et al. 2000). The basic working principle of the LSTM model is shown in Fig. 1.

In Fig. 1, x_t is the input vector at time t ; y_t is the output vector at time t ; h_t is the recurrent output vector at time t ; f_g is the forget gate vector; i_g is the input gate vector; o_g is the output gate vector; c_t is the t th iteration state of the memory cell. First, the input information is passed to the hidden layer through the input layer, while the output of the previous generation hidden layers is transferred to the current-generation hidden layer through the recurrent layer. Then, the information of the hidden layer and the information stored in the previous generation memory cell are transferred to the next generation of the memory cell through the input gate and the forget gate. Finally, the information stored in the new generation memory cell will be transferred to the recurrent layer through the output

gate. Similarly, the information is repeatedly trained between the hidden layer and the recurrent layer until the end of the iteration, and will be transferred to the output layer to get the final output. As space is limited, the detailed procedure for implementing the LSTM model can be found in the literature (Gers et al. 2000).

3 Overview of ant lion optimizer

In recent years, a new optimization method known as ALO proposed by Mirjalili (2015) has become a candidate for optimization application due to its flexibility and efficiency, which is based on the hunting behavior of antlions and entrapment of ants in antlions' traps. There are two elements in ALO, one is the ant and the other one is the antlion. The ants represent the individuals, which are required to move over the search space using different random walks. The antlions are allowed to hunt them and become fitter using traps. There are mainly four operations

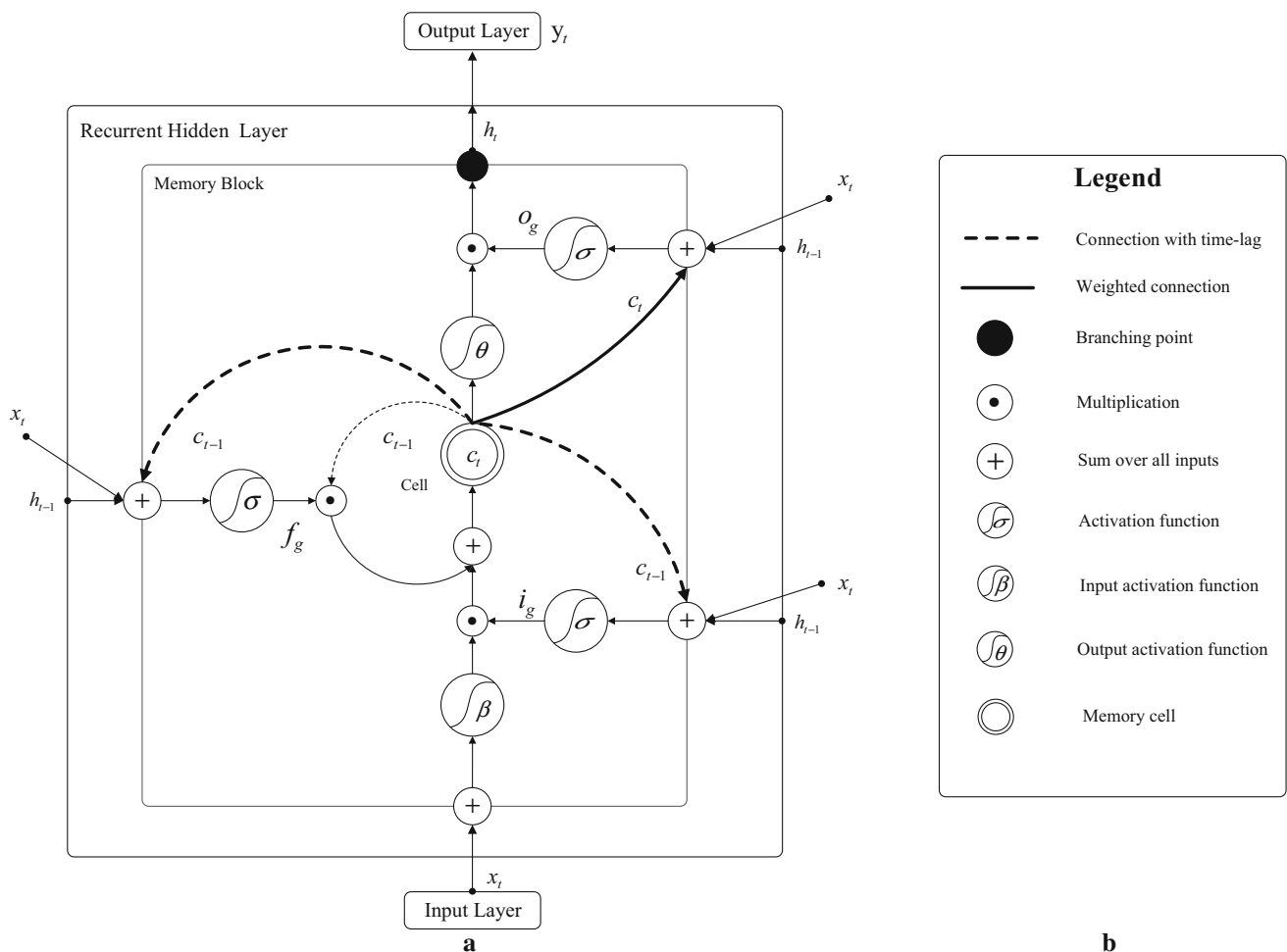


Fig. 1 The working principle of LSTM model (a the flowchart of LSTM model; b the legend of LSTM model)

in ALO, namely random walk of ants, update the position of ant, sliding ants towards antlion and catching preys.

The main implementation steps of the ALO are as follows:

- Step 1 Initialize the population of ants and antlions randomly;
- Step 2 Calculate the fitness of ants and antlions;
- Step 3 Find the best antlions and assume it as the elite;
- Step 4 Building trap. The antlion individuals are selected based on their fitness by using roulette wheel technique;
- Step 5 Sliding ants towards ant lion. Once the selection of antlion is over, update the ants by decreasing the radius of ant’s random walks;
- Step 6 Trapping of ants in antlion’s pits. As the search space of ants is shrinking around the antlion with the number of iterations increasing, the position of the updated ants will tend to the selected antlion;
- Step 7 Catching prey and re-building the pit. Calculate the fitness of all updated ants. If the fitness of the ant caught by the antlion is better than the antlion’s, the positions of antlions are replaced with the captured ants. Then update the fitness of all antlions, choose the best antlion and compare with elite individual. If the fitness of the best antlion is better than elite individual, then replace elite individual with the antlion; otherwise proceed to the next iteration;
- Step 8 Repeat Steps 3–7 until the maximum number of iterations is reached, and output elite individual as optimal solution.

4 The monthly runoff forecasting based on LSTM–ALO model

4.1 Hybrid LSTM–ALO model

The key of the proposed LSTM–ALO model is to make full use of the advantages of the LSTM model and the ALO algorithm, respectively. The ALO is utilized to optimize the key parameters of the LSTM model, which establishes a hybrid LSTM–ALO model for monthly runoff prediction. There are two key parameters directly influencing the output of the LSTM model, namely, the number of hidden layers (HN) and the learning rate of weight coefficient updating (α). Therefore, the ALO is adopted to find optimal values of HN and α in the LSTM model. In the LSTM–ALO model, the individuals of populations Ant and Antlion are consisted of the parameters (HN, α) which need to be optimized in the LSTM model. By building the corresponding LSTM model with each Ant and Antlion, the mean square root error ($RMSE$) of the training data set is used as the fitness of the LSTM Model, which can help optimally search and eventually find the optimal values of

the parameters (HN, α). Thus, the LSTM–ALO prediction model can be determined. The main steps of the LSTM–ALO implementation are as follows:

1. Initialize population of individuals
 Populations of Ant and Antlion are represented as $Ant = [A_1^1, A_2^1, \dots, A_{NP}^1]$, $Antlion = [Al_1^1, Al_2^1, \dots, Al_{NP}^1]$, where NP denotes the size of population.
 Each individual in the populations of Ant and Antlion consists of parameters (HN, α), namely A_i^1 is composed of hn_i^1 and α_i^1 , and Al_i^1 is composed of $h\tilde{n}_i^1$ and $\tilde{\alpha}_i^1$ respectively. The initial populations of Ant and Antlion are generated randomly.
 First, construct the corresponding LSTM model from the individuals in the initial Ant and Antlion populations. Assume that the HN and α of the A_i^1 are hn_i^1 and α_i^1 respectively, and the initial weight coefficient matrix of the memory block neurons in the corresponding LSTM model denotes as W_i^1 . W_i^1 is determined by the hn_i^1 and the dimension of each structure in memory block neurons, where each initial weight element consists of a random number between 0 and 1.
2. Calculate the fitness of individuals

In the evolution process of ALO, let A_i^{gen} is the i th individual of population Ant in generation gen , and Al_i^{gen} is the i th individual of population Antlion in generation gen . The corresponding LSTM model established by A_i^{gen} denotes as $LSTM_i^{gen}$ and the HN and α of $LSTM_i^{gen}$ express as hn_i^{gen} and α_i^{gen} respectively. The input vector of $LSTM_i^{gen}$ at time t is x_t , and the number of inputs is M . The input of memory block at time t is \hat{c}_t . The recurrent weight matrices of $LSTM_i^{gen}$ are W_i^{gen} , W_f^{gen} , W_o^{gen} and W_c^{gen} in $\mathbb{R}^{hn_i^{gen} \times (M+hn_i^{gen})}$ and the corresponding bias terms are b_i^{gen} , b_f^{gen} , b_o^{gen} and b_c^{gen} respectively. The hidden layer vector h_t at time t of the $LSTM_i^{gen}$ in the forward training process can be expressed as:

$$\begin{cases} ig_t = \sigma(W_i^{gen}H + b_i^{gen}) \\ fg_t = \sigma(W_f^{gen}H + b_f^{gen}) \\ og_t = \sigma(W_o^{gen}H + b_o^{gen}) \\ \hat{c}_t = \tanh(W_c^{gen}H + b_c^{gen}) \\ c_t = fg_t \odot c_{t-1} + ig_t \odot \hat{c}_t \\ h_t = \tanh(og_t \odot c_t) \end{cases} \quad (1)$$

where ig_t , fg_t and og_t denote the activation of the input gate, forget gate and output gate at time t , respectively; \hat{c}_t represents the input of memory block at time t ; $\sigma(\cdot)$ is the sigmoid function, which is defined in Eq. (2), and $\tanh(\cdot)$ denotes the \tanh function, which is defined in Eq. (3); \odot is the pointwise multiplication of two vectors and H in $\mathbb{R}^{M+hn_i^{gen}}$ represents the concatenation

of the input x_t and the previous hidden vector h_{t-1} ,

$$H = \begin{bmatrix} Ix_t \\ h_{t-1} \end{bmatrix}$$

$$\sigma(x) = \frac{1}{1 + e^{-x}} \tag{2}$$

$$\tanh(x) = \frac{e^x - e^{-x}}{e^x + e^{-x}} \tag{3}$$

So, the output yf_t at time t of the $LSTM_i^{gen}$ in the forward training process can be expressed as:

$$yf_t = \sigma(Wy_i^{gen}h_t + by_i^{gen}) \tag{4}$$

where Wy_i^{gen} in $\mathbb{R}^{hn_i^{gen}}$ is the weight matrices in the output layer; by_i^{gen} is the bias terms.

When the forward training process is completed, then the backward training. Calculate training losses Δ_t at time t :

$$\Delta_t = (Ty_t - yf_t)^2 \tag{5}$$

where Ty_t is the observed value of the training sample at time t .

Following the work of Hochreiter and Schmidhuber (1997), the updated weight coefficients bias terms with Δ_t and learning rate α_i^{gen} , then go on the next training of the $LSTM_i^{gen}$ at time $t + 1$. After the training of the $LSTM_i^{gen}$ is finished, we can get the training result $Y_i^{gen} = (y_i^{gen}(1), y_i^{gen}(2), \dots, y_i^{gen}(M))$, where $y_i^{gen}(M)$ is the training result of the $LSTM_i^{gen}$ at time M .

We use *RMSE* as an optimization criterion for the LSTM–ALO model, from which we calculate the fitness of Ant and Antlion. The fitness function ($fit(A_i^{gen})$) of the i th Ant in the gen th iteration is expressed as:

$$fit(A_i^{gen}) = \sqrt{\frac{1}{N} \sum_{n=1}^N (y_i^{gen}(n) - Ty_t)^2} \tag{6}$$

where N is the number of training set. In each iteration, the error (*RMSE*) is calculated by the output of each corresponding LSTM model and the observed sample, the smaller the *RMSE* of the LSTM model is, the better performance of the corresponding Ant individual.

Calculate the fitness $fit(A)$ of each individual in population Ant according to Eq. (6). Similarly, calculate the fitness $fit(Al)$ of each individual in population Antlion, and then determine the elite individual and its corresponding LSTM model.

3. Update individuals

Update the individuals according to the ALO’s working principle. The current updated individual of the Ant population is expressed as: $A_i^{gen+1} = [(hn_1^{gen+1}, \alpha_1^{gen+1}), (hn_2^{gen+1}, \alpha_2^{gen+1}), \dots, (hn_{NP}^{gen+1}, \alpha_{NP}^{gen+1})]$

4. Calculate the fitness of all individuals in population Ant

The fitness of each individual in population Ant is calculated and compared with its corresponding individuals in population Antlion. The Antlions are replaced with their corresponding Ants, if the fitness of the updated individual in Ants is greater than that in Antlions. Then update the corresponding LSTM model.

5. Update the elite individual

Choose the best individual from the updated Antlion population as the elite individual, and update the elite individual and the corresponding LSTM model.

6. Repeat Steps (2–5) until the termination criterion is satisfied. The optimal parameter combination (HN^*, α^*) and the corresponding optimal LSTM* model can be obtained from the elite individual.

7. The LSTM–ALO model is validated by test data set based on the best LSTM* model. Thus, a hybrid model LSTM–ALO is obtained for predicting the monthly runoff series.

The flow chart of the LSTM–ALO model for monthly runoff prediction is given in Fig. 2.

4.2 Normalization of the original runoff series data

In order to avoid data patterns and attributions with large numerical ranges dominating the role of the smaller numerical ranges and increase the prediction flexibility, we normalize the input data as follows:

$$\tilde{Q}_t = \frac{Q_t - Q_{\min}}{Q_{\max} - Q_{\min}} \tag{7}$$

where Q_{\min} and Q_{\max} denote the minimum and maximum value of runoff series; Q_t and \tilde{Q}_t are the observed and normalized runoff at time t , respectively.

4.3 The evaluation performance index of runoff prediction model

In order to test the performance of prediction models, normalized mean absolute error (*NMAE*), coefficient of correlation (R^2) and Nash–Sutcliffe efficiency coefficient (*NSEC*) (McCuen et al. 2006) are used as evaluation indexes. The definitions of these indexes are given as follows.

$$NMAE = \left(\frac{1}{N} \sum_{i=1}^N |y_p^i - y_T^i| \right) \tag{8}$$

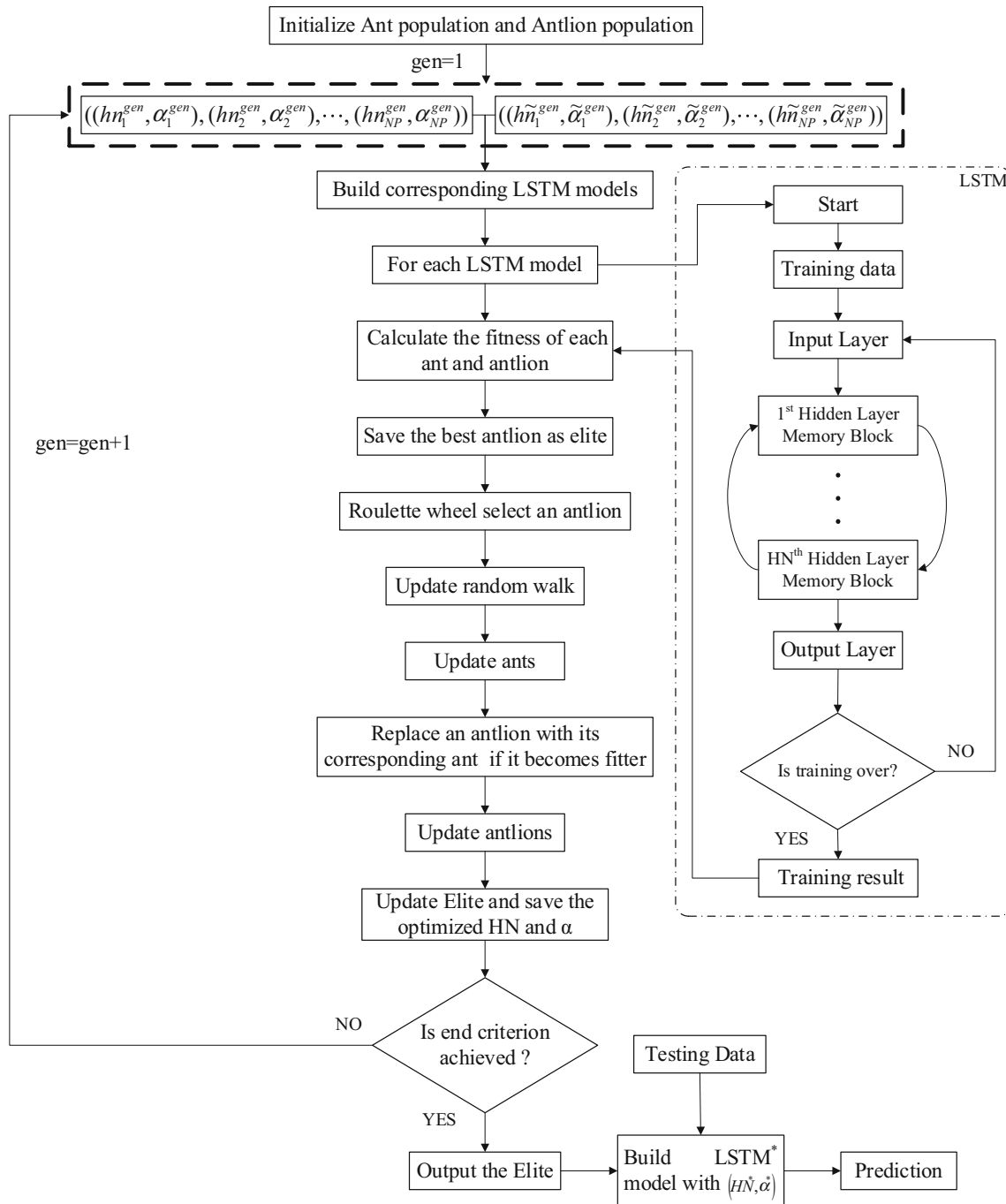


Fig. 2 Flow chart of the LSTM-ALO model for monthly runoff prediction

$$R^2 = \frac{\sum_{i=1}^N (y_T^i - \bar{y}_T)^2 (y_P^i - \bar{y}_P)^2}{\sum_{i=1}^N (y_T^i - \bar{y}_T) \sum_{i=1}^N (y_P^i - \bar{y}_P)} \tag{9}$$

$$NSEC = 1 - \left[\frac{\sum_{i=1}^N (y_T^i - y_P^i)^2}{\sum_{i=1}^N (y_T^i - \bar{y}_T)^2} \right] \tag{10}$$

where y_P^i is the i th prediction; \bar{y}_T is the average of the observed runoff; \bar{y}_P is the average of forecast runoff. If the value of $RMSE$ and $NMAE$ are close to 0, the error of

prediction is smaller. Moreover, if the value of R^2 and $NSEC$ are close to 1, the correlation between the observed and predicted runoff series is stronger.

To further evaluate the performance of the model, the $RMSE$ can be decomposed into three terms (De Giorgi et al. 2014): the bias, the amplitude error (SD_{bias}) and the phase error ($DISP$). SD_{bias} reflects the deviation of the predicted results from the observed values. $DISP$ is due to a time shift of the predicted values in respect to the observed

data that occurs if the amplitude of the forecast is right, but arrives too early or too late. The mathematical expressions are as follows:

$$RMSE^2 = bias^2 + SD_{bias}^2 + DISP^2 \tag{11}$$

$$bias = \frac{1}{N} \sum_{i=1}^N y_P^i - y_T^i \tag{12}$$

$$SD_{bias} = \sigma_T - \sigma_P \tag{13}$$

$$DISP = \sqrt{2\sigma_T\sigma_T(1 - R_{TP})} \tag{14}$$

where σ_T is the standard deviation of the observed runoff; σ_P is the standard deviation of the prediction runoff; R_{TP} is the cross-correlation coefficient between the observed and prediction runoff. The closer SD_{bias} is to 0, the smaller the error amplitude fluctuation is. If $DISP > 0$, the phase of predicted runoff lags behind the observed phase.

4.4 The implementation process of LSTM–ALO for runoff prediction

The runoff prediction process based on the LSTM–ALO model is implemented as follows.

1. Divide the runoff series data into training set Q_T and prediction set Q_P . Then the data are normalized according to Eq. (7);
2. Set parameters of the LSTM–ALO model and randomly generate initial population (HN, α) for ALO;
3. Train LSTM–ALO model: the training data set Q_T is used as the input of the LSTM–ALO model, and the corresponding output is obtained by the LSTM–ALO model;
4. Obtain the optimal parameters combination (HN, α) for LSTM–ALO model by using ALO;
5. Build LSTM–ALO prediction model with test data set Q_P ;
6. Get the runoff prediction.

5 Case study

5.1 Study area

The Astor River Basin is situated in the high mountains of Hindukush–Karakoram–Himalaya (HKH) region at the North of Pakistan. The geographical location of the region is approximately between longitudes $74^{\circ}24'$ and $75^{\circ}14'E$, and between latitudes $34^{\circ}45'$ to $35^{\circ}38'N$. The Astor River covers drainage area of 3990 km^2 . Water and power development authority (WAPDA) has one gauge station, i.e. Doyian hydrometric station in this area for the flow record. The elevation of this gauge station is 1583 m and

its geographical location in the basin is $35^{\circ}33'$ latitude and $74^{\circ}42'$ longitude. The mean annual flow of the river is $137 \text{ m}^3/\text{s}$, i.e. it is equivalent to 1084 mm of water depth (Tahir et al. 2015). The flow data at Astor River was collected by WAPDA under surface water hydrology project (SWHP), Pakistan. This gauge station was installed in 1974 and has recorded monthly runoff since 1974. The historical monthly runoff data from 1974 to 2009 in the basin is shown in Fig. 3.

The average value (\bar{Q}), standard deviation (S_d), Kurtosis criterion (K_c) and Skewness criterion (S_c) of the monthly runoff data in the Astor River Basin are $137.0, 140.94 \text{ m}^3/\text{s}, 1.433$ and 1.235 respectively. From these statistical values, we can see that the monthly runoff series in this basin have more dynamic and stronger randomness. Both K_c and S_c are not zero, and it means that the monthly runoff series does not obey the normal distribution. Therefore, the historical runoff series data of the Astor River Basin are chosen as a case study to test the feasibility and effectiveness of the LSTM–ALO model for monthly runoff prediction.

5.2 Selection of model structure and input variables

In this study, the monthly mean runoff data in the Astor River Basin are divided into two parts: calibration data set and validation data set. The calibration period covered the data are from 1974 to 2002 (336 months) and have been chosen to establish LSTM–ALO, BP, RBNN and LSSVM models. The validation period covered the runoff data are from 2003 to 2009 (84 months) and have been used to evaluate the performance of these models.

First, according to the Eq. (7), the input runoff data is normalized. Subsequently, in order to find out the influence

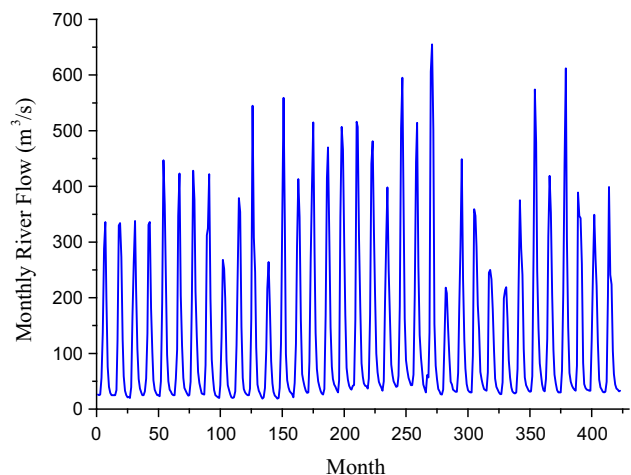


Fig. 3 Monthly runoff series from 1974 to 2009 in the Astor River Basin

of suitable time lags runoff on the current time interval runoff, the following five cases of input variable (M1, M2, M3, M4, M5) were designed with lagged t ranged from one up to 5 months before the current time for runoff forecasting.

1. M1 : $Q(t + 1) = f(Q(t));$
2. M2 : $Q(t + 1) = f(Q(t - 1), Q(t));$
3. M3 : $Q(t + 1) = f(Q(t - 2), Q(t - 1), Q(t));$
4. M4 : $Q(t + 1) = f(Q(t - 3), Q(t - 2), Q(t - 1), Q(t));$
5. M5 : $Q(t + 1) = f(Q(t - 4), Q(t - 3), Q(t - 2), Q(t - 1), Q(t)).$

where $Q(t + 1)$ is the target output or predicted runoff; $Q(t)$ is the 1-month lagged $Q(t + 1)$, $Q(t - 1)$ is the 2-month lagged $Q(t + 1)$. The rest can be done in the same manner. $f(\cdot)$ represents the models' type (i.e., LSTM, BP, RBNN or LSSVM).

5.3 Monthly runoff forecasting based on LSTM, LSTM-PSO and LSTM-ALO

As we know, the parameters have a significant influence on the prediction performance of the LSTM model. Therefore, the LSTM-ALO model was employed to forecast monthly runoff. The maximum number of iterations is set 500, the number of hidden layer neurons (HN) is set in the range [1, 50] and the learning rate (α) is set in the range [0, 1]. In order to verify the efficiency and stability of the ALO optimizing the parameters of the LSTM model, PSO is also applied to optimize the parameters of the LSTM model

(LSTM-PSO). At the same time, LSTM-ALO is compared with the original LSTM model and the LSTM-PSO model. Five kinds of input variables (M1, M2, M3, M4, M5) are performed 20 successive random trials by using the three models. The statistical analysis of the best, average and worst values of the performance metrics over 20 trials are shown in Table 1.

As can be seen from Table 1, the best, average and worst values of each metric of the LSTM-ALO model and the LSTM-PSO model have improved in different degrees compared with the LSTM model, which indicates that the parameters optimization can improve the forecasting performance of the LSTM model. Four evaluation metrics of the LSTM-ALO are better than those of LSTM-PSO in the whole, which means that the ALO is better than PSO for the parameters optimization of the LSTM model. So, the parameters optimization LSTM model based ALO is more suitable for the monthly runoff forecasting. In addition, the computation time for the execution of the LSTM-ALO, LSTM-PSO and LSTM models with five cases of input variables (M1, M2, M3, M4, M5) gives in Table 2. All the

Table 2 The computational time of LSTM-ALO, LSTM-PSO and LSTM models (unit: second)

	M1	M2	M3	M4	M5
LSTM-ALO	266	276	300	337	366
LSTM-PSO	452	506	550	598	643
LSTM	152	176	194	228	260

Table 1 Comparisons of the performance of calibration results for different models

Input	Metrics	<i>NMAE</i>			<i>RMSE</i>			<i>R²</i>			<i>NSEC</i>		
		B	A	W	B	A	W	B	A	W	B	A	W
M1	ALO-	0.043	0.045	0.047	0.072	0.079	0.086	0.943	0.931	0.918	0.889	0.877	0.855
	PSO-	0.048	0.050	0.052	0.085	0.089	0.092	0.918	0.912	0.906	0.843	0.838	0.836
	LSTM	0.051	0.054	0.059	0.085	0.093	0.101	0.905	0.901	0.890	0.818	0.809	0.737
M2	ALO-	0.039	0.040	0.041	0.071	0.072	0.075	0.943	0.942	0.938	0.889	0.882	0.873
	PSO-	0.042	0.045	0.049	0.077	0.080	0.083	0.933	0.929	0.924	0.870	0.866	0.857
	LSTM	0.045	0.047	0.048	0.079	0.080	0.083	0.931	0.927	0.922	0.866	0.855	0.813
M3	ALO-	0.038	0.039	0.039	0.065	0.069	0.072	0.951	0.946	0.942	0.910	0.893	0.883
	PSO-	0.039	0.040	0.042	0.072	0.074	0.076	0.942	0.940	0.937	0.888	0.882	0.862
	LSTM	0.043	0.044	0.046	0.074	0.077	0.081	0.938	0.934	0.927	0.880	0.871	0.847
M4	ALO-	0.038	0.039	0.039	0.071	0.071	0.077	0.944	0.943	0.935	0.891	0.885	0.874
	PSO-	0.039	0.040	0.043	0.070	0.074	0.078	0.946	0.940	0.931	0.895	0.875	0.865
	LSTM	0.050	0.051	0.053	0.085	0.087	0.094	0.919	0.913	0.901	0.844	0.835	0.816
M5	ALO-	0.039	0.039	0.040	0.069	0.071	0.074	0.947	0.944	0.940	0.897	0.886	0.872
	PSO-	0.042	0.043	0.044	0.073	0.076	0.079	0.941	0.937	0.931	0.885	0.869	0.852
	LSTM	0.050	0.052	0.056	0.081	0.086	0.095	0.928	0.917	0.899	0.860	0.843	0.806

ALO- denotes LSTM-ALO model, PSO- denotes LSTM-PSO model, B denotes the best value, A denotes the average value and W denotes the worst value

Table 3 Comparison of Wilcoxon rank sum test for different models

Input	Metrics	LSTM–PSO		LSTM	
		<i>p</i>	<i>h</i>	<i>p</i>	<i>h</i>
M1	<i>NMAE</i>	1.45E–11	1	1.45E–11	1
	<i>RMSE</i>	5.80E–11	1	2.90E–11	1
	<i>R</i> ²	1.45E–11	1	1.45E–11	1
	<i>NSEC</i>	1.45E–11	1	5.80E–11	1
M2	<i>NMAE</i>	1.45E–11	1	1.45E–11	1
	<i>RMSE</i>	1.45E–11	1	1.45E–11	1
	<i>R</i> ²	1.45E–11	1	1.45E–11	1
	<i>NSEC</i>	1.45E–11	1	1.45E–11	1
M3	<i>NMAE</i>	1.45E–11	1	1.45E–11	1
	<i>RMSE</i>	2.90E–11	1	1.45E–11	1
	<i>R</i> ²	5.80E–11	1	1.45E–11	1
	<i>NSEC</i>	1.10E–05	1	1.45E–11	1
M4	<i>NMAE</i>	6.53E–10	1	1.45E–11	1
	<i>RMSE</i>	2.00E–06	1	1.45E–11	1
	<i>R</i> ²	8.00E–03	1	1.45E–11	1
	<i>NSEC</i>	3.03E–08	1	1.45E–11	1
M5	<i>NMAE</i>	1.45E–11	1	1.45E–11	1
	<i>RMSE</i>	1.02E–10	1	1.45E–11	1
	<i>R</i> ²	5.80E–11	1	1.45E–11	1
	<i>NSEC</i>	1.00E–06	1	1.45E–11	1

simulations are implemented in MATLAB R2017a environment running on the personal computer (Windows 10 operation system; CPU: Intel (R) Core (TM) i5-4200H

CPU @ 2.80 GHz; RAM: 8 GB). The total execution time of the LSTM–ALO is less than that of the LSTM–PSO. Therefore, the execution time of the LSTM–ALO is acceptable for practical application of the monthly runoff prediction. According to the computational time associated with the three models, the model can be chosen for the monthly runoff prediction based on the performance and computational cost trade-off.

In order to further verify the effectiveness of the LSTM–ALO model in monthly runoff forecasting, the Wilcoxon rank sum test with significance level 0.05 has been applied to evaluate the statistical significance of the difference of four indexes between the LSTM–ALO model and the compared models (LSTM–PSO, LSTM) respectively (Ji et al. 2014; Yuan et al. 2015b). The results are shown in Table 3. The *p* < 0.05 and *h* = 1 imply that the LSTM–ALO model significantly outperforms the compared models in statistically.

As can be seen from Table 3, the *p* of all performance indexes are less than 0.05 and the *h* are equal to one with all five different input variables. The results show that the Wilcoxon rank sum test rejects all of null hypothesis in the difference between the LSTM–ALO and the compared models (LSTM–PSO, LSTM) on the monthly runoff forecasting, which means that there is a significant difference between the results of the LSTM–ALO model, the LSTM–PSO and LSTM models. So, as can be seen from Tables 1 and 3, it demonstrates that the LSTM–ALO is superior to

Table 4 Validation performance of four models for different inputs

Model input	Indicator	LSTM–ALO	BP–ALO	RBNN–ALO	LSSVM–ALO
M1	<i>NMAE</i>	0.048	0.113	0.086	0.110
	<i>RMSE</i>	0.085	0.162	0.134	0.155
	<i>R</i> ²	0.932	0.727	0.636	0.748
	<i>NSEC</i>	0.867	0.520	0.353	0.557
M2	<i>NMAE</i>	0.043	0.069	0.062	0.065
	<i>RMSE</i>	0.079	0.102	0.097	0.096
	<i>R</i> ²	0.940	0.901	0.912	0.911
	<i>NSEC</i>	0.884	0.811	0.828	0.829
M3	<i>NMAE</i>	0.040	0.058	0.059	0.053
	<i>RMSE</i>	0.074	0.090	0.091	0.087
	<i>R</i> ²	0.948	0.924	0.923	0.929
	<i>NSEC</i>	0.899	0.850	0.849	0.862
M4	<i>NMAE</i>	0.041	0.066	0.056	0.054
	<i>RMSE</i>	0.076	0.095	0.096	0.087
	<i>R</i> ²	0.945	0.913	0.914	0.929
	<i>NSEC</i>	0.893	0.833	0.832	0.861
M5	<i>NMAE</i>	0.041	0.077	0.066	0.053
	<i>RMSE</i>	0.075	0.120	0.104	0.088
	<i>R</i> ²	0.947	0.860	0.901	0.926
	<i>NSEC</i>	0.897	0.735	0.803	0.857

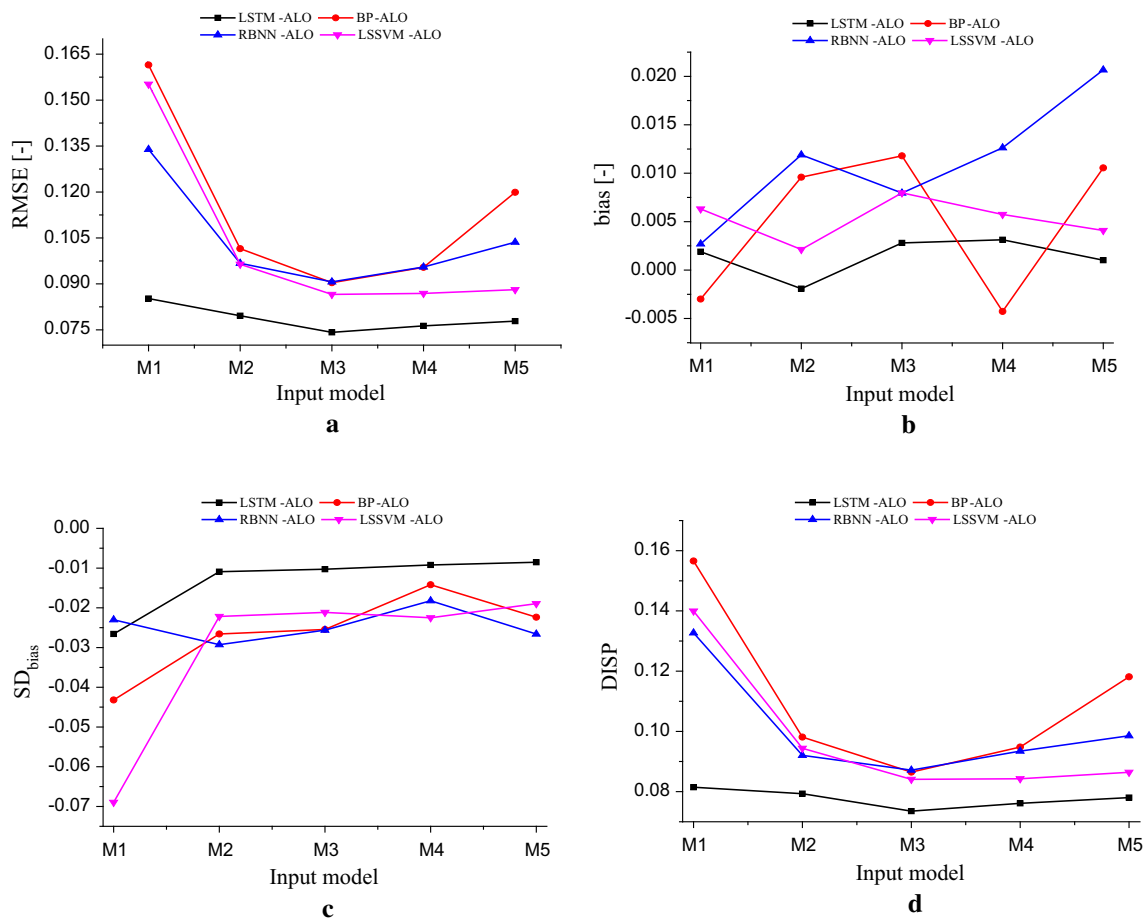


Fig. 4 $RMSE$, $bias$, SD_{bias} and $DISP$ of LSTM–ALO, BP–ALO, RBNN–ALO and LSSVM–ALO models with different inputs (**a** $RMSE$; **b** $bias$; **c** SD_{bias} ; **d** $DISP$)

the LSTM–PSO and the LSTM in monthly runoff forecasting.

5.4 Monthly runoff forecasting based on different models optimized by ALO

To verify the performance of the LSTM–ALO, other three approaches are also adopted to forecast the monthly runoff series. The BP, RBNN and LSSVM models are employed to compare with the LSTM–ALO model. We optimize the parameters of hidden layer and output layer weight coefficient of the BP neural network, diffusion degree and maximum neuron number of radial basis functions in the RBNN model and the penalty factor of the LSSVM model by using ALO respectively. Three-fourths of the datasets are adopted to calibration while the validation process employs one-fourths of the dataset. The forecast results of four models (LSTM–ALO, BP–ALO, RBNN–ALO, LSSVM–ALO) with five cases of input variables are given in Table 4 in terms of the performance indices of $NMAE$, $RMSE$, R^2 , and $NSEC$.

In Table 4, we can see that there is a general trend of the prediction accuracy increasing as the time lag reached three, and then the trend is stable. Compared with BP–ALO, RBNN–ALO and LSSVM–ALO models, the LSTM–ALO performs much better. It can be seen that the change trend of results with the influence by inputs of the LSTM–ALO model are relatively stable, of which reason is that the LSTM–ALO can utilize a series of memory cells to handle input data and enhance the learning ability of runoff series.

In order to better understanding the origin of the errors between the observed and predicted runoff, the $RMSE$ has been decomposed into three components, namely the forecast bias ($bias$), the amplitude error SD_{bias} and the phase error $DISP$. The results are presented as Fig. 4.

As shown in Fig. 4, the $RMSE$ of the LSTM–ALO is much smaller than other models with each input model, which shows the prediction error of the LSTM–ALO is smaller than other models. The $bias$ of the LSTM–ALO model is much closer to zero than other models with the same input, which demonstrates the variation of prediction

Table 5 Runoff prediction for LSTM–ALO, BP–ALO, RBNN–ALO and LSSVM–ALO models

Time (month)	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17
Observed runoff	72	223	375	288	241	108	57	43	38	33	31	32	59	281	574	479	248
LSTM–ALO	73	240	388	328	186	99	57	36	31	27	27	35	83	218	410	441	256
BP–ALO	67	192	400	330	207	115	16	46	31	53	53	57	66	149	426	471	211
BNN–ALO	49	230	355	389	179	111	47	40	59	59	44	44	53	174	400	384	262
LSSVM–ALO	59	216	370	402	139	127	52	40	40	53	46	50	60	165	403	427	268
Time (month)	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34
Observed runoff	145	82	61	44	34	32	54	126	310	419	353	206	106	62	46	34	30
LSTM–ALO	105	70	52	43	36	32	37	89	304	470	375	231	105	65	48	41	36
BP–ALO	196	77	46	46	39	38	54	130	297	382	256	175	97	46	38	35	36
RBNN–ALO	148	65	55	83	51	38	48	154	317	375	333	175	110	53	52	62	34
LSSVM–ALO	127	71	59	68	38	33	49	146	309	386	316	182	109	61	47	43	28
Time (month)	35	36	37	38	39	40	41	42	43	44	45	46	47	48	49	50	51
Observed runoff	30	30	69	208	440	612	306	147	65	49	42	38	35	33	78	389	347
LSTM–ALO	30	32	66	219	407	415	343	142	71	55	46	40	33	35	70	249	514
BP–ALO	46	60	64	181	391	375	414	170	105	44	35	53	60	60	61	197	446
RBNN–ALO	38	46	45	219	354	437	349	147	70	33	71	66	60	54	50	246	479
LSSVM–ALO	40	53	55	206	369	452	337	102	90	43	51	56	57	55	54	230	432
Time (month)	52	53	54	55	56	57	58	59	60	61	62	63	64	65	66	67	68
Observed runoff	343	261	83	45	38	35	34	33	33	123	245	349	271	221	119	53	39
LSTM–ALO	329	225	125	61	45	41	35	29	34	86	308	405	304	185	94	57	39
BP–ALO	269	153	131	10	68	37	58	64	64	66	300	393	247	188	120	40	19
RBNN–ALO	256	145	132	38	39	59	54	55	52	54	364	294	341	165	107	60	18
LSSVM–ALO	354	198	137	49	36	43	54	58	57	60	353	314	313	123	114	61	17
Time (month)	69	70	71	72	73	74	75	76	77	78	79	80	81	82	83	84	
Observed runoff	35	31	30	31	49	203	399	241	224	104	63	44	37	35	32	33	
LSTM–ALO	35	34	31	40	98	239	371	364	185	96	58	37	32	31	32	41	
BP–ALO	24	51	54	58	66	121	390	385	181	103	12	49	29	45	59	58	
RBNN–ALO	51	53	40	43	50	133	392	417	142	94	41	57	51	53	56	45	
LSSVM–ALO	30	49	45	51	59	130	406	434	135	117	41	54	34	45	56	50	

error for the LSTM–ALO is smaller than others. The amplitude error (SD_{bias}) of the LSTM–ALO model is less than other models, and SD_{bias} is increasing while the time lags increasing but there is little difference among M2 to M5. The SD_{bias} of the BP–ALO and the RBNN–ALO are best when the time lag is four. The SD_{bias} of the LSSVM–ALO model is increasing with the time lags increasing. The $DISP$ of all models gets the best when the time lags are three. From the overall point of view of Table 4 and Fig. 4, when the time lags is three, the SD_{bias} and $DISP$ of all models tend to be saturated, and the LSTM–ALO performs best. Therefore, we choose M3 as the inputs of the monthly runoff forecasting models. The observed and predicted runoff of four models for the validation period are shown in Table 5.

To further analyze the prediction performance of the LSTM–ALO, Fig. 5 shows the scatter plots of the observed

and predicted runoff in comparison with the LSTM–ALO, BP–ALO, RBNN–ALO and LSSVM–ALO models. In contrast to the wide deviations by BP–ALO, RBNN–ALO and LSSVM–ALO models, the LSTM–ALO model leads to much stronger correlations between the observed and predicted runoff, which is consistent with the highest R^2 of LSTM–ALO as shown in Table 4, and it can be seen that the predictions by the LSTM–ALO have the minimal deviations and the strongest correlation. This is mainly due to the fact that the LSTM–ALO model can capture the long-term dependence of runoff series.

To compare among these models with more detailed analysis, the box-plots in Fig. 6 are utilized to indicate the degree of overall spread in the observed and predicted runoff series. We generate box-plots with the respective distributions of the original monthly runoff series and all approaches. Based on the box-plots, it can be seen that the

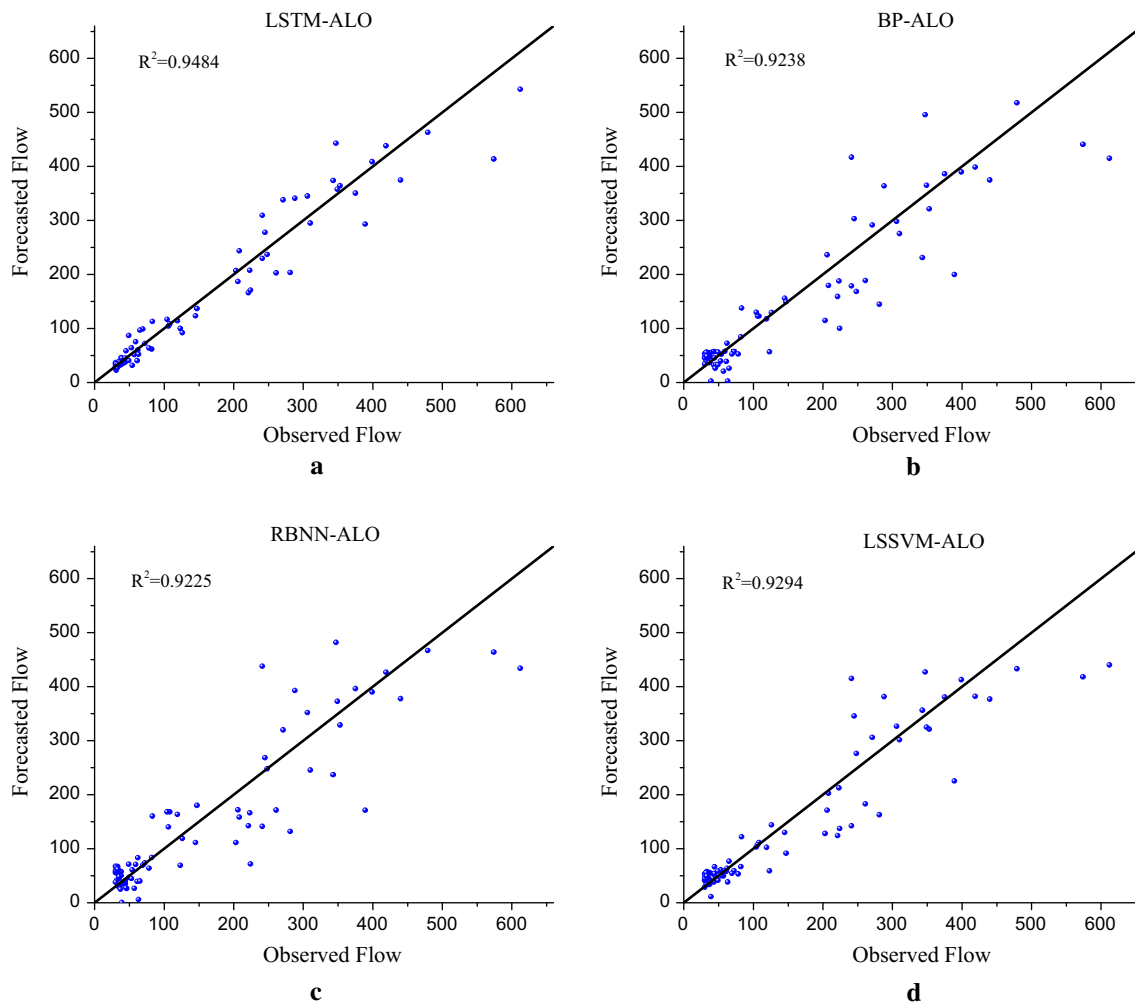


Fig. 5 The scatter-plots of runoff forecasting using LSTM-ALO, BP-ALO, RBNN-ALO and LSSVM-ALO models

spread of predictions based on the three LSTM models (LSTM-ALO, LSTM-PSO, LSTM) closely resembled the observed runoff between the 25th percentiles to the 75th percentiles. On the other hand, BP-ALO, RBNN-ALO and LSSVM-ALO models show that poorer predictive coverage for the lower flows when compared to the observed runoff and the LSTM models. In the three LSTM models, LSTM-ALO is closer to the observed runoff between the 50th percentiles to the 75th percentiles than other two LSTM models (LSTM-PSO, LSTM), and the outliers of LSTM-ALO are closest to the observed runoff. Therefore, the LSTM-ALO is the most effective model in terms of forecasting monthly runoff accurately for the validation set.

6 Conclusions

The accuracy of LSTM-ALO model had been investigated for monthly runoff forecasting in this study. Different combinations of runoff data in the Astor River Basin of

Northern Pakistan were discussed for choosing as the input variables of the LSTM-ALO. The simulation results by the LSTM-ALO were compared with those of the LSTM and LSTM-PSO. The comparisons show that the ALO could increase the accuracy of the LSTM model in forecasting monthly runoff with different inputs of the models. The BP, RBNN and LSSVM models were also compared with three LSTM models for monthly runoff forecasting. All the parameters of these models are optimized by ALO and evaluated by the indexes of *NMAE*, *RMSE*, R^2 and *NSEC*. We find that the performances of all models are becoming better with time lags increasing and the LSTM-ALO is the best among these models. The prediction ability of the LSTM-ALO with few time lags is much better than other models. Meanwhile, it should be noticed that the time lags have little influence on the LSTM-ALO. For further analysis, the bias, the amplitude error and phase error of the LSTM-ALO were less than other models. Therefore, the LSTM-ALO model can improve the prediction accuracy of the monthly runoff forecasting.

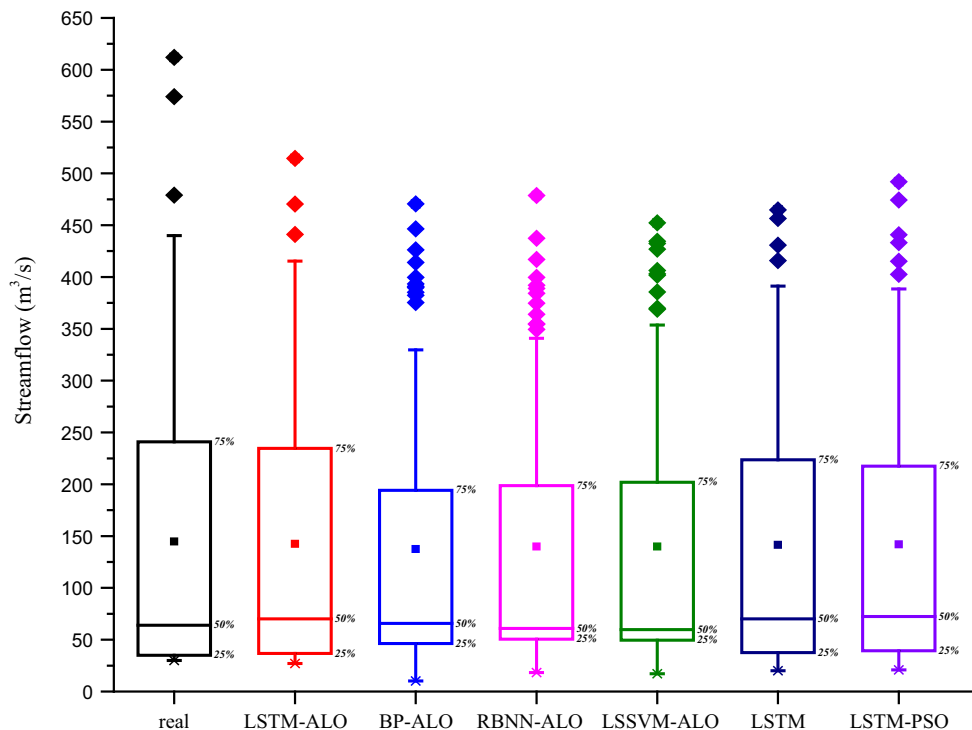


Fig. 6 Box-plots of runoff forecasting using LSTM-ALO, BP-ALO, RBNN-ALO, LSSVM-ALO, LSTM and LSTM-PSO models

Acknowledgements This work was supported by National Natural Science Foundation of China (Nos. U1765201, 41571514, 51379080), Hubei Provincial Collaborative Innovation Center for New Energy Microgrid in China Three Gorges University, and the Fundamental Research Funds for the Central Universities (No. 2017KFYXJJ204).

References

- Adamowski J, Sun K (2010) Development of a coupled wavelet transform and neural network method for flow forecasting of non-perennial rivers in semi-arid watersheds. *J Hydrol* 390(1):85–91
- Anderson PL, Meerschaert MM, Zhang K (2013) Forecasting with prediction intervals for periodic autoregressive moving average models. *J Time Ser Anal* 34(2):187–193
- Bukhari D, Wang Y, Wang H (2017) Multilingual convolutional long short-term memory, deep neural networks for low resource speech recognition. *Proc Comput Sci* 107:842–847
- Chen ZH, Yuan XH, Ji B, Wang PT, Tian H (2014) Design of a fractional order PID controller for hydraulic turbine regulating system using chaotic non-dominated sorting genetic algorithm II. *Energy Convers Manag* 84:390–404
- De Giorgi MG, Campilongo S, Ficarella A (2014) Comparison between wind power prediction models based on wavelet decomposition with least-squares support vector machine (LS-SVM) and artificial neural network (ANN). *Energies* 7(8):5251–5272
- El-Shafie A, Taha MR, Noureldin A (2007) A neuro-fuzzy model for inflow forecasting of the Nile River at Aswan high dam. *Water Resour Manag* 21(3):533–556
- Gers FA, Schmidhuber J, Cummins F (2000) Learning to forget: continual prediction with LSTM. *Neural Comput* 12(10):2451–2471
- Greff K, Srivastava RK, Koutník J (2016) LSTM: a search space odyssey. *IEEE Trans Neural Netw Learn Syst* 99:1–11
- Hochreiter S, Schmidhuber J (1997) Long short-term memory. *Neural Comput* 9(8):1735–1780
- Ji B, Yuan XH, Li XS, Huang YH, Li WW (2014) Application of quantum-inspired binary gravitational search algorithm for thermal unit commitment with wind power integration. *Energy Convers Manag* 87:589–598
- Kalteh AM (2013) Monthly river flow forecasting using artificial neural network and support vector regression models coupled with wavelet transform. *Comput Geosci* 54:1–8
- Kalteh AM, Chen G (2014) Evaluating a coupled discrete wavelet transform and support vector regression for daily and monthly streamflow forecasting. *J Hydrol* 519:2822–2831
- Kisi O (2015) Streamflow forecasting and estimation using least square support vector regression and adaptive neuro-fuzzy embedded fuzzy c-means clustering. *Water Resour Manag* 29(14):5109–5127
- Kong XM, Huang GH, Fan YR (2015) Maximum entropy-gumbel-hougaard copula method for simulation of monthly streamflow in Xiangxi river, China. *Stoch Environ Res Risk Assess* 29(3):833–846
- Liang J, Yuan XH, Yuan YB, Chen ZH, Li YZ (2017) Nonlinear dynamic analysis and robust controller design for Francis hydraulic turbine regulating system with a straight-tube surge tank. *Mech Syst Signal Process* 85:927–946
- Mccuen RH, Knight Z, Cutter AG (2006) Evaluation of the Nash-Sutcliffe efficiency index. *J Hydrol Eng* 11(6):597–602
- Mirjalili S (2015) The ant lion optimizer. *Adv Eng Softw* 83:80–98
- Okkan U, Serbes ZA (2012) Rainfall-runoff modeling using least squares support vector machines. *Environmetrics* 23(6):549–564
- Seidou O, Ouarda TBMJ (2007) Recursion-based multiple change point detection in multiple linear regression and application to river streamflows. *Water Resour Res* 43(7):W07404. <https://doi.org/10.1029/2006WR005021>

- Sharma S, Srivastava P, Fang X (2015) Performance comparison of adoptive neuro fuzzy inference system (ANFIS) with loading simulation program C++ (LSPC) model for streamflow simulation in El Nino southern oscillation (ENSO)-affected watershed. *Expert Syst Appl* 42(4):2213–2223
- Sudheer CH, Anand N, Panigrahi BK (2013) Streamflow forecasting by SVM with quantum behaved particle swarm optimization. *Neurocomputing* 101:18–23
- Tahir AA, Chevallier P, Arnaud Y, Ashraf M, Bhatti MT (2015) Snow cover trend and hydrological characteristics of the Astore River basin (Western Himalayas) and its comparison to the Hunza basin (Karakoram region). *Sci Total Environ* 505:748–761
- Wang W, Van Gelder PH, Vrijling JK (2006) Forecasting daily streamflow using hybrid ANN models. *J Hydrol* 324(1):383–399
- Yaseen ZM, El-Shafie A, Jaafar O (2015) Artificial intelligence based models for stream-flow forecasting: 2000–2015. *J Hydrol* 530:829–844
- Yuan XH, Ji B, Zhang SQ, Tian H, Chen ZH (2014) An improved artificial physical optimization algorithm for dynamic dispatch of generators with valve-point effects and wind power. *Energy Convers Manag* 82:92–105
- Yuan XH, Tian H, Yuan YB, Huang YH, Ikram RM (2015a) An extended NSGA-III for solution multi-objective hydro-thermal-wind scheduling considering wind power cost. *Energy Convers Manag* 96:568–578
- Yuan XH, Ji B, Yuan YB, Ikram RM, Zhang XP, Huang YH (2015b) An efficient chaos embedded hybrid approach for hydro-thermal unit commitment problem. *Energy Convers Manag* 91:225–237