The applicability of ASCS_LSTM_ATT model for water level prediction in small- and medium-sized basins in China

Ke Li, Dingsheng Wan, Yuelong Zhu, Cheng Yao, Yufeng Yu, Cunyou Si and Xiangchao Ruan

ABSTRACT

Water level prediction of small- and medium-sized rivers plays an important role in water resource management and flood control. Such a prediction is concentrated in the flood season because of the frequent occurrence of flood disasters in the plain area. Moreover, the flood in mountainous areas suddenly rises and falls, and the slope is steep. Thus, establishing a hydrological prediction model for small- and medium-sized rivers with high accuracy and different topographic features, that is, plains and mountains, is an urgent problem. A prediction method based on ASCS_LSTM_ATT is proposed to solve this problem. First, the important parameters are optimized by improving the cuckoo search algorithm. Second, different methods are used to determine the forecast factors according to various topographic features. Finally, the model is combined with the self-attention mechanism to extract significant information. Experiments demonstrate that the proposed model has the ability to effectively improve the water level prediction accuracy and parameter optimization efficiency.

Key words | correlation coefficient, cuckoo search algorithm, LSTM, self-attention mechanism, water level prediction

HIGHLIGHTS

- Different methods are proposed to determine the forecast factors according to different topographic features.
- The self-attention mechanism is combined with LSTM.
- An improved CS algorithm is proposed to optimize the parameters of ASCS_LSTM_ATT.
A timely and effective water level prediction is of considerable importance to reservoir operation and flood warning. The water level prediction and management of large rivers have been improved in recent years. However, many small- and medium-sized rivers exist in China, i.e. those with a drainage area of between 200 and 3,000 km² (Zhiyu 2012; Zhiyu et al. 2015), most of which are located in poorly-gauged areas of the mountains and plains. The intensity of flood and rainstorms is large, and necessary emergency detection means are lacking. Therefore, attention should be provided to the water level simulation of small- and medium-sized rivers with different topographies to improve the hydrological forecast accuracy and provide decision-making services for flood control.

At present, the two approaches for water level prediction are available: conceptually or physically based and data-driven models. Conceptually or physically based models utilize multiple-related variables, such as evaporation, infiltration rate, and soil moisture content, to obtain the physical parameters for the prediction task. Data-driven models directly find the relationship between precipitation and water level from the obtained hydrological data and increase data availability (Liu et al. 2020). Conceptually or physically based models have been widely used in many countries due to their high efficiency. However, these models exhibit some problems due to the lack of hydrological data in small- and medium-sized rivers. These conceptually or physically based models also exhibit difficulty in capturing highly nonlinear relationships due to the large number and wide distribution of small- and medium-sized rivers.

With information development, the neural network becomes a group of classic data-driven methods that can solve the problem of model construction caused by the lack of characteristic hydrological parameters to a large extent and simulation of the temporal and spatial nonlinear changes in hydrological systems (Chen et al. 2019). For example, the neural network has been used in reservoir operation (Anvari et al. 2014), water level prediction (Piasecki et al. 2017), water quality simulation (Chang et al. 2013), and precipitation forecasting (Akbari Asanjan et al. 2018). The recurrent neural network (RNN) efficiently performs in time series prediction and has been successfully applied to hydrology research (Bai & Shen 2019). However, the RNN is unable to learn and deal with the ‘long-term dependency’ tasks autonomously. After many epochs, the gradient tends to disappear or explode in most cases, thereby resulting in network performance degradation and prediction accuracy reduction. Therefore, Hochreiter & Schmidhuber (1997) proposed the long short-term memory (LSTM) to solve the problem of ‘long-term dependencies,’ gradient explosion, and vanishing.
The LSTM has extensive usage in certain subjects, such as speech recognition (Graves et al. 2013), machine translation (Cho et al. 2014), and stock forecasting (Nelson et al. 2017). Le et al. (2019) established an LSTM flood forecast model. This model respectively used daily flow and precipitation as input data and proved its effectiveness for runoff forecast in Vietnam River Basin. Zhang et al. (2018) used LSTM to predict the daily water level, and the proposed model was applied and evaluated in Hetao Irrigation District in arid northwestern China. The experiment proved that LSTM can effectively prevent overfitting. Liu et al. (2019) used LSTM to predict the water level at the upstream and downstream of the Gezhouba Hydropower Station. Experiments demonstrated that LSTM can overcome certain shortcomings, such as the unclear mechanism of heuristic algorithms. Mei et al. (2018) used the LSTM method to predict passenger flow and proved its superiority to multiple linear regression and error backpropagation (BP) models. The above-mentioned studies show that LSTM efficiently performs in hydrological prediction. Thus, the applicability of LSTM in small- and medium-sized rivers will be studied.

The present research adopted the hybrid model method to improve the neural network (Liu et al. 2020), mainly focusing on the parameter optimization algorithm and extracting key features, to improve the prediction performance. The parameter optimization algorithms include genetic algorithm (GA) (Yang & Honavar 2002), particle swarm optimization (PSO) algorithm (Liu et al. 2006), ant colony algorithm (Moeini & Afshar 2013), cuckoo search (CS) (Yang & Deb 2010), etc. Nevertheless, these algorithms often show local convergence and limited optimization when faced with complex multi-node networks. These conditions will lead to local optimal solutions during water level prediction. Therefore, the improved parameter optimization algorithm is of considerable importance to the water level prediction of the model.

The forecast factors also have an important influence on model performance. For example, the prior judgment method considerably relies on empirical judgment, and the neural network selection method is computationally expensive and inefficient. The above-mentioned forecast factor selection method will slow down the model calculation speed during water level prediction (Zhao & Yang 2011). Hence, different forecast factor selection methods are studied for small- and medium-sized rivers with various topographic features.

The attention mechanism enables neural networks to focus on its input (or feature) subset, thus assigning different weights to forecast factors. Rush et al. (2015) applied an attention model to text summarization to extract keywords from long sentences or paragraphs. Bahdanau et al. (2014) applied the attention model to machine translation. However, the information captured by the early attention mechanism is limited; thus, Google proposed the self-attention mechanism (Vaswani et al. 2017). The self-attention mechanism can effectively learn the internal structure of the sequence and then extract different aspects of information from the sequence. Given the limited hydrological data of small- and medium-sized rivers, this mechanism is used to allocate reasonable weights for the forecast factors to extract key features. Furthermore, the self-attention mechanism applications are still at their initial phase, with few studies on hydrological prediction in China. Studying the applicability of the self-attention mechanism in small- and medium-sized rivers in China is important.

Therefore, the hydrological data and deep learning technology are maximized to model the water level process of small- and medium-sized rivers with different topographic features. This study proposed the adaptive step size CS algorithm (ASCS) to the hydrological model by combining the LSTM and the self-attention mechanism, namely ASCS_LSTM_ATT. The contributions of this study are as follows:

1. The self-attention mechanism is combined with LSTM and can effectively capture the dependence between time series and extract key hydrologic features to address the problems of low accuracy and limited hydrological data.
2. Different methods are proposed to determine the forecast factors according to different topographic features to improve the accuracy of water level prediction of small- and medium-sized rivers.
3. An improved CS algorithm is proposed to optimize the parameters of ASCS_LSTM_ATT and improve global and local search capabilities. Such a task is performed to address the local convergence and limited optimization.
RELATED WORK

LSTM

LSTM uses three gating mechanisms to filter the amount of information flow. Different functions are used to calculate and obtain the hidden layer state value to learn the time-dependence relationship in the signal. Figure 1 shows the structure of the ‘memory cell’.

The gate control mechanism is the key to realize LSTM. Variable $g_t$ is the input gate that controls the input value; $f_t$ is the forget gate, which controls the preservation of the cell’s historical state; and $o_t$ is the output gate that controls the output value.

Self-attention mechanism

The self-attention mechanism can quickly lock the key points in the target from massive information and reduce the calculation burden of processing high-dimensional input data. Figure 2 shows the structure of the attention mechanism.

The attention mechanism is a process wherein the Encoder is responsible for learning the semantic code $C$ from the input sequence Source, and then the Decoder generates each output Target considering the semantic code $C$.

The self-attention mechanism is a special case of attention mechanism. In the Encoder–Decoder framework of general tasks, the source and target are different. The aforementioned mechanism refers to the attention calculation mechanism that occurs between the internal elements of the source or target to further obtain the internal correlation and improve data validity.

CS algorithm

CS (Yang & Deb 2010) is an optimization algorithm proposed by Yang to solve the problem by simulating the breeding strategy of cuckoos. In this algorithm, a hatched nest represents a group of solutions. If the fitness value of the new solution is satisfactory, then the new solution will replace the poor solution of the previous generation. The new nest location is generated by Levy flight, as Equation (1) shows.

$$x_{i}^{h+1} = x_{i}^{h} + \alpha \odot L(\beta)$$

According to the step factor in Yang & Deb (2010),

$$\alpha = \alpha_{0} \ast (x_{i}^{h} - x_{\text{best}})$$

where $x_{i}^{h+1}$ and $x_{i}^{h}$ are the positions of the $i(i = 1, 2, 3, \ldots n)$ nest in the $h + 1$ and $h$ generations, respectively; $\odot$ is the point-to-point multiplication; $\alpha$ is the step control quantity, $\alpha_{0} = 0.01$; $x_{\text{best}}$ is the current optimal solution position; $L(\beta)$ is the Levy random search path; and $L(\beta)$ and time $t$ obey Levy distribution.

$$L\hat{u} = t^{-\beta}$$

where $\beta$ is the heavy-tailed distribution.

Evaluation index

The evaluation index value is an important way to improve model prediction quality.
Root mean square error (RMSE) measures the overall performance across the entire dataset.

\[
RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^{n} (y_i - \hat{y}_i)^2}, \quad i = 1, 2, \ldots, n
\]  

(4)

R-square \((R^2)\) can judge the fitting degree of the algorithm, and its value range is \([0, 1]\). The algorithm fitting effect is satisfactory when the result is close to 1.

\[
R^2 = \left(1 - \frac{\sum_{i=1}^{n} (y_i - \hat{y}_i)^2}{\sum_{i=1}^{n} (y_i - \bar{y}_i)^2}\right)^2, \quad i = 1, 2, \ldots, n
\]  

(5)

Mean absolute error \((MAE)\) can efficiently reflect the deviation between the predicted and the actual values.

\[
MAE = \frac{1}{n} \sum_{i=1}^{n} |y_i - \hat{y}_i|, \quad i = 1, 2, \ldots, n
\]  

(6)

Mean squared error \((MSE)\) is the quotient of intra-group variation and error freedom in ANOVA.

\[
MSE = \frac{1}{n} \sum_{i=1}^{n} (y_i - \hat{y}_i)^2, \quad i = 1, 2, \ldots, n
\]  

(7)

Nash–Sutcliffe efficiency \((NSE)\) quantitatively describes the model output accuracy.

\[
NSE = 1 - \frac{\sum_{i=1}^{n} (y_i - \hat{y}_i)^2}{\sum_{i=1}^{n} (y_i - \bar{y}_i)^2}, \quad i = 1, 2, \ldots, n
\]  

(8)

where \(y_i, \hat{y}_i,\) and \(\bar{y}_i,\) respectively, represent the corresponding real, predicted, and average values.

**PROPOSED METHODS**

As previously mentioned, the CS can enhance the prediction model parameters. However, this algorithm still has room for improvement. The use of the self-attention mechanism can improve the relevance of data. The selection of forecast factors in different basins is also crucial to the prediction accuracy. Therefore, ASCS_LSTM_ATT is proposed to predict the water level of small- and medium-sized rivers.

The proposed model gains an advantage in mimicking highly nonlinear, complex systems, and building models without a priori information. Moreover, ASCS_LSTM_ATT acquires the relationship between input and output directly from existing hydrological data rather than the physically based model using a full set of mathematic equations for each part in the hydrological cycle (i.e., interception, infiltration, and evaporation). The physically based model of small- and medium-sized rivers is difficult to complete because of the lack of hydrological data. ASCS_LSTM_ATT is able to find hidden hydrologic rules from the existing hydrological data of small- and medium-sized rivers, which saves expensive computational costs and reduces plenty of data requirements and parameters to be estimated.

**ASCS algorithm**

CS randomly generates a step factor through the Levy flight. However, the fixed value of the step factor \(a_0 = 0.01\). Consequently, this algorithm is unable to adjust the step size and slows down the convergence speed. Therefore, the \(a_0\) of the algorithm is improved.

In the middle stage of optimization, the size of the \(a_0\) is reduced to improve the local search capability. The \(a_0\) in Equation (2) is replaced with Equation (9).

\[
a_0 = \frac{2}{1 + 10^{h_i/h_{\text{max}}}}
\]  

(9)

Therefore, the improved Levy flight update position is shown in the following equation:

\[
x_{i}^{h+1} = x_{i}^{h} + \frac{2}{1 + 10^{h_i/h_{\text{max}}}} r (x_{i}^{h} - x_{\text{best}}) L(\beta)
\]  

(10)

where \(h_i\) represents the current number of iterations, \(h_{\text{max}}\) represents the total number of iterations.

Equation (1) is used to calculate the location of hatched nests, and the random number \(r\) is generated. The probability of discovering \(r\) and foreign cuckoo eggs with the host \(p_a\) is compared. If \(r > p_a\), then the new nest location...
is obtained by the random walk search strategy, as shown below:

\[ x_{h}^{i+1} = x_{h}^{i} + \xi(x_{h}^{i} - x_{i}^{j}) \] (11)

where \( x_{h}^{i} \) and \( x_{h}^{j} \) are the position vectors of two randomly hatched nests of the \( h \) generation, and \( \xi \) is a uniform scaling factor that obeys \([0, 1]\).

The steps of the ASCS algorithm are as follows:

1. Initialize parameters, including the number of bird nest \( n \), the dimension \( dim \) of the solution, the range of the solution, and the maximum number of iterations \( time \). The fitness of all solutions \( (fitness) \) is calculated.

2. Each nest generates a new solution through Equation (10).

3. Compare the fitness of the new solution \( f_{new} \) and the old one \( fitness \). If \( f_{new} > fitness \), then the new solution will replace the old one; otherwise, some solutions are randomly eliminated according to the \( p_{e} \) of finding eggs of foreign birds and Equation (11) is used to generate a new nest location to replace the old one.

4. Whether the algorithm meets the end conditions is determined. If so, then output the final solution. Otherwise, steps (2)–(4) are repeated.

Figure 3 shows the flow of the ASCS algorithm.

**Forecast factor selection method**

The model performance is improved by removing irrelevant and redundant variables, thereby improving the accuracy and speed of the model. The main principles for selecting forecast factors are as follows.

1. Factors that have an impact on the flood occurrence must be considered when selecting forecast factors, such as precipitation, water level, evaporation, and other information.

2. Factors that can be obtained before the flood occurrence include precipitation and water level information a few hours before the current time.

Thus, the correlation between the input and output data of the model is calculated by the following equation:

\[ r = \frac{\sum_{i=1}^{n} (x_{i} - \bar{x})(y_{i} - \bar{y})}{\sqrt{\sum_{i=1}^{n} (x_{i} - \bar{x})^{2} \sum_{i=1}^{n} (y_{i} - \bar{y})^{2}}} \] (12)

where \( n \) represents the total number of time series, \( x_{i} \) represents the \( i \)th series, \( y_{i} \) represents the value to be predicted corresponding to the \( i \)th series, \( \bar{x} \) represents the average value of all series, and \( \bar{y} \) represents the average value of the value to be predicted corresponding to all series.

**In plain areas**

The small- and medium-sized rivers in the plain area have a gentle slope, and the terrain in the basin is mainly plain and hilly with low average elevation. The plain polder area is also low lying and has few water storage projects and weak flood detention capacity. The river terrain is flat, and its confluence speed is slow. In general, the farther the distance from the rain gauges to the outlet station of the basin, the longer the confluence time. Accordingly, the precipitation of rain gauges located in different geographical locations has diverse impacts on the target water level. A factor prediction method is proposed according to the precipitation and geographical characteristics of small- and medium-sized rivers in plain areas: The correlation between the water level of the target station in the specified forecast periods and the precipitation of rain gauges in the basin before the forecast period is calculated through Equation (12). The distance between the rain gauges and the target station is calculated by latitude and longitude.

The different rain gauges are arranged in ascending order according to the distance and the correlation coefficients of rain gauges are arranged in ascending order. A variable called \( T_{corr} \), which refers to the time corresponding to the correlation coefficient, is proposed. The steps to determine \( T_{corr} \) are as follows.
1. The \( T_{corr} \) of the rain gauges, which has a small distance, should be first considered.

2. The \( T_{corr} \) can be the same when the difference of the distance of the markable station is within the allowable range.

3. Finally, the time \( T_{corr} \) with the largest correlation is determined. If \( T_{corr} \) has been selected by other rain gauges, then the largest \( T_{corr} \) in the remaining correlation coefficient is selected.

All values from \( T_{corr} \) to current time \( T \) are selected as the forecast factors of the model. Figure 4 shows the flowchart of determining forecast factors for small- and medium-sized rivers in plain areas.

**In mountainous areas**

The small- and medium-sized rivers in the mountainous areas experience floods with extremely fast onset, large quantity, and violent fluctuation. Moreover, the terrain of the basin is mainly mountainous with high average elevation, large relative height difference, steep and narrow riverbed, large slope, complex confluence path and short runoff producing and converging process. Thus, the precipitation of rain gauges in different geographical locations slightly affects the target water level station. Hence, the precipitation of rain gauges in different geographical locations has a slight impact on the target water level station. A factor prediction method is proposed according to the characteristics of precipitation and topography of small- and medium-sized rivers in mountainous areas. The correlation between the water level of the target station in the specified forecast periods and the precipitation of rain gauges in the basin before the forecast period is calculated through Equation (12). The steps to determine \( T_{corr} \) are as follows.

1. Arrange the correlation coefficients of different rain gauges in descending order.
2. Select the time corresponding to the largest correlation coefficient as \( T_{corr} \).

All values from \( T_{corr} \) to current time \( T \) are selected as the forecast factors of the model according to the correlation time \( T_{corr} \) calculated by the correlation coefficient. This step is conducted to improve the correlation between the model prediction and the input data.

Figure 5 shows the flowchart of determining forecast factors for small- and medium-sized rivers in mountainous areas.

**ASCS_LSTM_ATT prediction model**

The input layer of the water level prediction model based on ASCS_LSTM_ATT takes the water level of the hydrological station and precipitation of rain gauges as independent variables, including data filling realization, normalization, and training set division. Data feature searching and parameter transfer are realized in the hidden layer. In the output layer, the water level data are reverse normalized, and the water level prediction results are output.

The self-attention mechanism is used to assign different weights to the forecast factors for improving the correlation between the input and output data. Additionally, the key features extracted from the self-attention mechanism are used as input of the first LSTM unit after regularization, and the regularization structure is used to avoid the overfitting of the model. Then, the key implicit features of the
hydrological data are acquired through multiple LSTM units in turn. The dropout layers are used to prevent the model from overfitting to increase the generalization ability of parameters to the dataset. Finally, the output of the last LSTM layer is the feature of the whole time series, and the layer is connected to a fully connected layer for water level prediction.

In the model training, the nonlinear modeling performance of the prediction model increases with the number of hidden layer nodes. However, the training time, complexity, and calculation amount of the model will increase. Therefore, the number of hidden layer nodes and learning rates are determined by the ASCS algorithm.

Figure 6 shows the flowchart of ASCS_LSTM_ATT.

The steps of establishing ASCS_LSTM_ATT model are as follows.

1. Process hydrological time series sample data.
2. Use the appropriate forecast factor method to select the appropriate forecast factors.
3. Assign different weights to input data through the self-attention mechanism.
4. Set the model’s structure and initialize parameters, which mainly include hidden_size, lr, and the parameters in ASCS.
5. Train the ASCS_LSTM_ATT and use the ASCS to optimize the parameters.
6. Use the optimal solution to establish the ASCS_LSTM_ATT and output the prediction water level value. Otherwise, return to step (5) and continue to optimize the parameters.

**EXPERIMENTAL RESULTS AND ANALYSIS**

**Data description and preprocessing**

**Data description**

The Qinhua basin is mainly located in the middle and lower reaches of the Yangtze River plain in China, and the basin covers an area of 2,631 km$^2$. The abdomen of this basin is the low-lying polder area. The average elevation is approximately 6–8 m, the plain occupies nearly 74.3% of the area with a gentle slope below 3°, and the river slope is around 0.1‰. The basin belongs to the small- and medium-sized rivers in the plain area (Xiaoying et al. 2019). The annual precipitation is approximately 1,047.8 mm and mainly concentrated from April to September, thereby accounting for 70.6% of the annual precipitation. The flow of the upper reaches of the basin and the flood duration are short, and the flood detention capacity is weak. The outlet section will be supported by the Yangtze River tide, and the drainage will be rough. Long-term heavy rain may damage the ecology and economy of Nanjing City. Therefore, the flood forecast is conducted for Qinhua (Xuan et al. 2018). Fourteen rain gauges and one hydrological station in Qinhua are tested, as shown in Figure 7(a). A total of 40,835 pieces of data are available, and the hydrological data are hourly data. Among the data, 32,000 are training sets, and 8,835 are test sets.

The Tunxi basin is located in the mountainous area of Southern Anhui between the Huangshan Mountains and Baiji Mountains in China, and the basin covers an area of 2,696.76 km$^2$ (Yao et al. 2014). The basin is dominated by mountains, which account for more than 60% of the total area. The maximum, minimum, and average elevation of the basin are 1,398, 116, and 127 m, respectively, with a large relative height difference (Cheng et al. 2012). Therefore, the Tunxi belongs to the small- and medium-sized rivers in the mountainous area. Flood is caused by rainstorm, which is characterized by rapid fluctuation, strong suddenness, and fast confluence time. Tunxi is prone to floods in spring and droughts in summer; thus, flood control operation is crucial for Tunxi (Shuaihong et al. 2019). Eleven rain gauges and one hydrological station in the Tunxi are
tested, as shown in Figure 7(b). A total of 26,130 pieces of data are available, and the hydrological data are hourly data. Among the data, 18,000 are training sets, and 8,130 are test sets.

Data preprocessing

The core steps are implemented to avoid the negative effects of the missing original hydrological data, multiple noises, and random values on the data prediction accuracy.

Data Interpolation

Interpolating the precipitation is necessary because the precipitation information in the two basins is incomplete. The methods commonly used in hydrology include mean and linear interpolations.

1. Mean interpolation

Mean interpolation calculates the mean value of the answer unit in each target variable and fills the average value of each group into the all missing items.

\[ \bar{y} = \frac{\sum_{i=1}^{n} a_i y_i}{n} \]  

(13)

where \( a_i \) is an indicative variable, \( a_i = 0 \) represents no answer, \( a_i = 1 \) means an answer is provided, and \( n \) is the total number of units with answers.

2. Linear interpolation

Linear interpolation calculates the value of an unknown quantity between two known quantities by connecting two known quantities of straight lines. Given the coordinates \((x_0, y_0)\) and \((x_1, y_1)\), the value of \( x \) on the line at some point in the interval \([x_0, x_1]\) is solved.

\[ y = y_0 + \frac{y_1 - y_0}{x_1 - x_0} (x - x_0) \]  

(14)

The precipitation with 50 sample size and 20% missing rate in Tunxi was used. The missing values were generated at 4, 10, 23, 24, 27, 32, 36, 38, 40, and 43 places. Table 1 shows the comparison of the two interpolation methods.

| No | Real value | Mean interpolation | Linear interpolation |
|----|------------|---------------------|----------------------|
| 4  | 6          | 1.39                | 3.35                 |
| 10 | 1          | 1.39                | 0.65                 |
| 23 | 1          | 1.39                | 1.33                 |
| 24 | 3          | 1.39                | 1.57                 |
| 27 | 5          | 1.39                | 2.5                  |
| 32 | 1.3        | 1.39                | 1.15                 |
| 36 | 3          | 1.39                | 2                    |
| 38 | 1          | 1.39                | 1.5                  |
| 40 | 2          | 1.39                | 1.25                 |
| 43 | 3          | 1.39                | 1.75                 |
| RMSE | 0.93     |                     | 0.75                 |
Table 1 demonstrates that the RMSE of linear interpolation is smaller than that of mean interpolation one, which indicates completion value obtained is more accurate. Therefore, the linear interpolation method is adopted to complete the missing precipitation value.

**Min-Max**

In sequence $X$, $X_{\text{max}}$, and $X_{\text{min}}$ are the maximum and minimum values, respectively. Thus, the value of each element in the resulting sequence $X'$ is within $[0,1]$, and each element $X_i$ has the following equation:

$$X_i = \frac{X_i - X_{\text{min}}}{X_{\text{max}} - X_{\text{min}}}$$  \hspace{1cm} (15)

**Comparison of LSTM_ATT and LSTM**

The self-attention mechanism assigns different weights to the input to improve the prediction accuracy. Hence, the LSTM is combined with the self-attention mechanism to predict the water level.

Table 2 summarizes the details of the LSTM and LSTM_ATT. The best training effect can be selected by changing several training options and parameters of the model.

Figures 8 and 9, respectively, show the water level prediction hydrographs of LSTM and LSTM_ATT. Table 3 shows the evaluation indexes under different models. The forecast period is 1 h.

Figures 8 and 9 demonstrate that the prediction curve trend of the two models is similar to the real value curve. However, Figures 8(b), 8(c), 9(b), and 9(c) reveal that LSTM_ATT efficiently performed on the trough and peak values of the time series. In combination with the self-attention mechanism, the input is assigned with different weights according to the correlation between the input and the output, and the key features of the input data are extracted. Thus, the noise data and the curve jitter, which is close to the real value, are reduced. Considering that the LSTM and LSTM_ATT inputs are the same, the results might imply the impact of the structural difference on the simulation.

Table 3 shows that the LSTM_ATT values in terms of RMSE, MSE, and MAE are smaller than those of LSTM, thereby reducing the prediction error. The LSTM values in terms of $R^2$ and NSE are also lower than those of LSTM_ATT. This result demonstrates that the combination of LSTM and self-attention mechanism can improve the prediction accuracy to a certain extent.

The two above-mentioned situations show that the combination of LSTM and self-attention mechanism can improve the accuracy of water level prediction to a certain extent. The LSTM_ATT can reduce prediction error and input noise data compared with LSTM.

**Comparison of ASCS and other optimization algorithms**

Figure 10 shows the $a_0$ of the ASCS. The figure demonstrates that the initial value of $a_0$ is large, which can avoid the algorithm falling into local optimum and expand the search range. With the increase of iterations, $a_0$ gradually decreases. In the early stage of optimization, the descent speed is faster than that in the later stage, thereby shortening the convergence time in the early stage. The change range of the $a_0$ is small in the later stage, thereby improving the search accuracy.
PSO, GA, CS, and ASCS are used to optimize the LSTM_ATT parameters (number of hidden layer nodes $\text{hidden}\_\text{size}$ and learning rate $lr$). Parameters of PSO include maximum iterations of population ($\text{max}\_\text{iter}$), particle number ($m$), dimensions of solution ($d$), inertia parameter ($w$), learning factor ($c_1, c_2$), and random number ($r_1, r_2$). Parameters of GA include mutation probability ($\text{mp}$), crossover probability ($\text{cp}$), and group size ($\text{size}$). Parameters of CS and ASCS include nest number ($n$) and detection probability ($p_a$).

The Dongshan Station of Qinhuai is taken as an example,
and RMSE is chosen as the evaluation criterion. The lr range is generally set as [0.001, 0.01], and the hidden size is within [40, 150]. Table 4 shows the details of PSO, GA, CS, and ASCS.

The remarkable feature in Figure 11 is the ASCS, which performs better than other models in optimizing the LSTM_ATT parameters. The performance of the four optimization algorithms showed a declining trend with the increase in the forecast period according to the number of iterations and the RMSE value. For example, in the 1–6 h model, ASCS converges through 8, 9, 10, and 12 iterations. Moreover, the number of iterations to achieve convergence gradually increases. The RMSE values of convergence are approximately 0.02, 0.025, 0.025, 0.03, 0.037, and 0.05. The criterion value gradually increases, thereby indicating the slow increase in error. The initial RMSE value simultaneously increases with the forecast period.

Figure 11 exhibits that the performance of different algorithms in the same prediction period also varies. The comparison result of the number of iterations and the RMSE of the four algorithms indicated that the performance of the PSO and CS algorithms is in the middle, while that of the GA algorithm is poor. The ASCS algorithm is better than the PSO and CS algorithms. For example, in the model with a prediction period of 1 h, the PSO and CS algorithms achieve convergence through 11 iterations, and the RMSE values are 0.42 and 0.4, respectively. The GA algorithm achieves convergence through 15 iterations, and the RMSE values are approximately 0.47. The ASCS algorithm achieves convergence through eight iterations, and the RMSE values are approximately 0.02.

The above-mentioned results indicate that the superior performance of ASCS to that of the three other algorithms is due to the decrease in step factor in ASCS with the increase in the number of iterations. This phenomenon shortens the convergence time in the early stage and reduces the change range of step in the later stage. The local search performance is also considered while improving global search performance. Thus, the optimal solution can be quickly searched. Therefore, using ASCS to optimize the parameters of LSTM_ATT is effective and feasible in improving the prediction accuracy and time efficiency.

**Determination of the forecast factors of ASCS_LSTM_ATT**

**Qinhuai basin**

In Qinhuai, Dongshan station is selected as the target forecast station. The early-stage water level of Dongshan and the early-stage precipitation of rain gauges in the basin are selected as the forecast factors. The forecast period is set from 1 to 6 h. The forecast period of 1 h \( (t + 1) \) is taken as an example. Table 5 shows the correlation coefficient value and distance of Dongshan station.

In Table 5, Dongshan z \( (t + 1) \) represents the water level of Dongshan station in the future time \( (t + 1) \), and Dongshan q \( (t - 1) \) indicates the precipitation in the first hour before the current time of Dongshan station. The different rain gauges are sorted in the ascending order according to the distance in the table, and the results are as follows: Dongshan, Qianhancun, Tuqiao, Wolongshan, Chishanzha, Zhaojun, Fangbian, Tianzhengqiao, Jurong, Zhongshan, Jurong Reservoir, Tianwangsi, Beishan Reservoir, and Ershengqiao.

**Table 4 | Parameters of different algorithms**

| Algorithm | Parameters details |
|-----------|--------------------|
| PSO       | \( \text{max_iter} = 20, m = 20, d = 2, w = 0.8, c1 = 2, c2 = 2, r1 = 0.6, r2 = 0.3 \) |
| GA        | \( \text{max_iter} = 20, mp = 0.01, cp = 0.8, size = 20 \) |
| CS        | \( n = 20, d = 2, pa = 0.25, time = 20 \) |
| ASCS      | \( n = 20, d = 2, pa = 0.25, time = 20 \) |
Figure 11 | Convergence chart of the optimal fitness of models in different prediction periods. The forecast periods are as follows: (a) 1 h, (b) 2 h, (c) 3 h, (d) 4 h, (e) 5 h, and (f) 6 h.

Table 5 | Correlation coefficient values of input and output variables (plain area)

| Station      | $T$ | $t-1$ | $t-2$ | $t-3$ | $t-4$ | $t-5$ | $t-6$ | $t+1$ | Distance (km) |
|--------------|-----|-------|-------|-------|-------|-------|-------|-------|----------------|
| Dongshan z   | 0.893 | 0.864 | 0.854 | 0.803 | 0.773 | 0.743 | 0.713 | 1     | 48.5           |
| Fangbian     | 0.382 | 0.400 | 0.407 | 0.415 | 0.413 | 0.408 | 0.397 | –     | 24.1           |
| Tuqiao       | 0.355 | 0.380 | 0.379 | 0.377 | 0.371 | 0.361 | 0.357 | –     | 45.1           |
| Wolongshan   | 0.574 | 0.555 | 0.579 | 0.632 | 0.621 | 0.605 | 0.589 | –     | 49.5           |
| Zhaojun       | 0.400 | 0.425 | 0.434 | 0.434 | 0.433 | 0.427 | 0.416 | –     | 62.5           |
| Zhongshan     | 0.371 | 0.362 | 0.369 | 0.370 | 0.386 | 0.390 | 0.397 | –     | 55.5           |
| Dongshan q    | 0.720 | 0.757 | 0.749 | 0.745 | 0.742 | 0.687 | 0.642 | –     | 0              |
| Tianshengqiao | 0.371 | 0.400 | 0.406 | 0.410 | 0.411 | 0.412 | 0.401 | –     | 49.5           |
| Qianhancun    | 0.710 | 0.743 | 0.743 | 0.685 | 0.657 | 0.603 | 0.579 | –     | 6.9            |
| Jurong Station| 0.332 | 0.310 | 0.341 | 0.346 | 0.351 | 0.358 | 0.358 | –     | 54.4           |
| Jurong Reservoir | 0.343 | 0.331 | 0.354 | 0.359 | 0.361 | 0.364 | 0.371 | –     | 58.0           |
| Chishanzha    | 0.397 | 0.402 | 0.413 | 0.415 | 0.417 | 0.415 | 0.399 | –     | 45.5           |
| Ershengqiao   | 0.360 | 0.344 | 0.349 | 0.354 | 0.357 | 0.361 | 0.365 | –     | 66.2           |
| Tianwangsi    | 0.370 | 0.351 | 0.359 | 0.365 | 0.366 | 0.369 | 0.372 | –     | 58.7           |
| Beishan Reservoir | 0.364 | 0.321 | 0.336 | 0.347 | 0.359 | 0.365 | 0.369 | –     | 62.5           |

Bold signifies the Tcorr finally selected by different rain-gauges according to the selection principle of the forecast factors in the “Forecast factor selection method” section.
The flood lasted for a long time because the Qinhuai is located in the plain area. The terrain is relatively flat, and the precipitation of rain gauges in different geographical locations has varying effects on the water level of the target station. Thus, the correlation of water level of the target station and the precipitation of rain gauges should be fully combined with the distance as forecast factors.

The final inputs of the model according to the selection principle of the forecast factors in the ‘Forecast factor selection method’ section are as follows: the current water level of Dongshan station, precipitation Dongshan q (t−1)–Dongshan q (t), precipitation Qianhancun (t−1)–Qianhancun (t), precipitation Tuqiao (t−2)–Tuqiao (t), precipitation Wolongshan (t−3)–Wolongshan (t), precipitation Chishanzha (t−3)–Chishanzha (t), precipitation Zhaocun (t−3)–Zhaocun (t), precipitation Fangbian (t−3)–Fangbian (t), precipitation Tianshengqiao (t−4)–Tianshengqiao (t), precipitation Jurong (t−5)–Jurong (t), precipitation Zhongshan (t−5)–Zhongshan (t), precipitation Jurong Reservoir (t−6)–Jurong Reservoir (t), precipitation Tianwangsi (t−6)–Tianwangsi (t), precipitation Beishan Reservoir (t−6)–Beishan Reservoir (t), and precipitation Ershengqiao (t−6)–Ershengqiao (t). Among the factors, Station (t−x)–Station (t) represents all precipitation or water level from time (t−x) to time t at Station.

Table 6 compares the errors of two methods to determine the forecast factor. Case 1 uses the correlation coefficient and geographical distance to determine the forecast factors. Case 2 uses the correlation coefficient to determine the forecast factors. The inputs are different at Tuqiao, Zhaocun, and Tianshengqiao (i.e., Tuqiao (t−1)–Tuqiao (t), Zhaocun (t−3)–Zhaocun (t), and Tianshengqiao (t−5)–Tianshengqiao (t), respectively), and the remaining ones are the same as those in Case 1. The prediction period of 1 h is taken as an example.

Table 6 demonstrates that the RMSE and MAE of Case 1 are smaller than those of Case 2. This finding indicates that the error of prediction results can be reduced by determining the input through the correlation coefficient and geographical distance. Simultaneously, the $R^2$ and NSE of Case 1 indicate that this case has higher prediction accuracy than that of Case 2. Solely relying on the correlation coefficient to select the forecast factors for the Qinhuai results in poor performance. The time of confluence has a certain relationship with distance in the plain area. Thus, the geographical location influence must also be considered.

The correlation coefficient and geographical distance method are used to determine the forecast factors in the Qinhuai. These factors can efficiently reduce the impact of the model input data on the prediction results and improve prediction accuracy.

### Tunxi basin

In Tunxi, Tunxi station is selected as the target forecast station. The water level in the early stage of Tunxi station and the precipitation of rain gauges in the basin in the early stage are selected as the forecast factors. The forecast period is set from 1 to 6 h. The forecast period of 1 h (t+1) is taken as an example. Table 7 shows the correlation coefficient and distance of Tunxi station.

In Table 7, Tun Xi° (t+1) represents the water level of Tunxi Station in the future time (t+1), and Tun Xi° (t−6) indicates the precipitation in the sixth hour before the current time of Tunxi station.

Table 7 demonstrates that the rain gauges at different geographical locations have reached the maximum correlation coefficient at time (t−6). Simultaneously, the terrain considerably fluctuates with a certain height difference. Thus, the precipitation of rain gauges in different geographical locations has a slight impact on the water level of the target hydrological station. Therefore, the correlation coefficient is considered in the selection of forecast factors in Tunxi.

The final forecast factors according to the selection principle of forecast factors in the mountainous area in the ‘Forecast factor selection method’ section are as follows: the water level Tun Xi (t−6)–Tun Xi (t), precipitation Tun Xi (t−6)–Tun Xi (t), precipitation Yan Qian (t−6)–Yan Qian (t), precipitation Xiu Ning (t−6)–Xiu Ning (t), precipitation Cheng Cun (t−6)–Cheng Cun (t), precipitation Shang

| Factors       | RMSE | MSE  | MAE  | $R^2$ | NSE  |
|---------------|------|------|------|-------|------|
| Case 1        | 0.022| 0.001| 0.011| 0.997 | 0.997|
| Case 2        | 0.029| 0.003| 0.016| 0.953 | 0.953|
Xikou (t – 6)–Shang Xikou (t), precipitation Wu Cheng (t – 6)–Wu Cheng (t), precipitation Shi Men (t – 6)–Shi Men (t), precipitation Yi Xian (t – 6)–Yi Xian (t), precipitation Ru Cun (t – 6)–Ru Cun (t), precipitation Zuo Long (t – 6)–Zuo Long (t), and precipitation Da Lian (t – 6)–Da Lian (t). Among the factors, Station (t – x)–Station (t) represents all precipitation or water level from time (t – x) to time t at Station.

The errors of the two methods are compared to determine the forecast factor. Case 3 uses the correlation coefficient to determine the forecast factors of the model. The forecast factors of Case 4 are determined by using the correlation coefficient and geographical location according to the principle in the ‘Forecast factor selection method’ section. The appropriate Tcorr is selected, and the distance is combined with the correlation coefficient. Thus, the forecast factors are as follows: water level Tun Xi (t – 6)–Tun Xi (t), precipitation Tun Xi (t – 6)–Tun Xi (t), precipitation Yan Qian (t – 4)–Yan Qian (t), precipitation Xi Ning (t – 5)–Xi Ning (t), precipitation Cheng Cun (t – 2)–Cheng Cun (t), precipitation Shang Xikou (t – 4)–Shang Xikou (t), precipitation Wu Cheng (t – 5)–Wu Cheng (t), precipitation Shi Men (t – 5)–Shi Men (t), precipitation Yi Xian (t – 3)–Yi Xian (t), precipitation Ru Cun (t – 4)–Ru Cun (t), precipitation Zuo Long (t – 2)–Zuo Long (t), and precipitation Da Lian (t – 3)–Da Lian (t). The prediction period of 1 h is taken as an example. Table 8 shows the results.

From Table 8, it can be found that RMSE, MSE, and MAE of Case 3 are smaller than those of Case 4, indicating that factors of Case 3 reduce the error of prediction results and make the results closer to the real values. Meanwhile, the $R^2$ and NSE of Case 3 are larger, which indicates that the forecast factors possessed by Case 3 make the prediction results more accurate.

An analysis is conducted on the basis of the results of the two above-mentioned basins. Different methods are adopted to determine the forecast factors of ASCS_LSTM_ATT. In Qinhuai, the time of confluence has a certain relationship with the distance in the plain area. The precipitation of rain gauges in different geographical locations has varying effects on the water level of the target station. Thus, the correlation coefficient and geographical location are comprehensively considered. However, the terrain of the Tunxi is mainly mountainous with high average elevation, steep and narrow riverbed, large slope, complex confluence path. Thus, the precipitation of rain gauges in different geographical locations has slight effects on the water level of the

---

**Table 7** | Correlation coefficient values of the input and output variables (mountain area)

| Station      | $t$  | $t – 1$ | $t – 2$ | $t – 3$ | $t – 4$ | $t – 5$ | $t – 6$ | $t + 1$ | Distance (km) |
|--------------|------|---------|---------|---------|---------|---------|---------|---------|---------------|
| Tun Xi       | 0.713| 0.743   | 0.773   | 0.803   | 0.834   | 0.864   | 0.893   | 1       | 0             |
| Yan Qian     | 0.313| 0.344   | 0.368   | 0.385   | 0.393   | 0.402   | 0.404   | –       | 29.6          |
| Xiu Ning     | 0.348| 0.377   | 0.400   | 0.416   | 0.427   | 0.433   | 0.434   | –       | 16.3          |
| Cheng Cun    | 0.270| 0.302   | 0.328   | 0.347   | 0.361   | 0.371   | 0.377   | –       | 52.8          |
| Shang Xikou  | 0.334| 0.361   | 0.382   | 0.397   | 0.408   | 0.413   | 0.415   | –       | 32.0          |
| Wu Cheng     | 0.371| 0.389   | 0.401   | 0.406   | 0.408   | 0.411   | 0.416   | –       | 18.6          |
| Shi Men      | 0.365| 0.378   | 0.383   | 0.388   | 0.391   | 0.393   | 0.395   | –       | 13.1          |
| Zuo Long     | 0.290| 0.315   | 0.337   | 0.354   | 0.367   | 0.376   | 0.381   | –       | 60.5          |
| Yi Xian      | 0.250| 0.279   | 0.309   | 0.335   | 0.354   | 0.366   | 0.372   | –       | 45.7          |
| Da Lian      | 0.309| 0.335   | 0.356   | 0.373   | 0.387   | 0.397   | 0.402   | –       | 48.4          |
| Ru Cun       | 0.270| 0.301   | 0.329   | 0.349   | 0.362   | 0.371   | 0.396   | –       | 37.2          |

The target hydrological station, including water level and precipitation. Bold signifies the Tcorr finally selected by different rain-gauges according to the selection principle of the forecast factors in the ‘Forecast factor selection method’ section.

**Table 8** | Error comparison of the forecast factors in Tunxi

| Factors          | RMSE | MSE  | MAE  | $R^2$ | NSE  |
|------------------|------|------|------|-------|------|
| Case 3           | 0.017| 0.008| 0.004| 0.995 | 0.995|
| Case 4           | 0.036| 0.015| 0.007| 0.889 | 0.889|
target station; consequently, only the correlation coefficient is considered.

Comparison of different models

ASCS_LSTM_ATT with BP, CNN, SVM, and WaveNet are compared. The five models are predicted and simulated by observing the evaluation index and the fitting curve. Table 9 shows the parameters of ASCS_LSTM_ATT in different basins.

The PSO is used to adjust the hyperparameters of SVM, and the kernel function is selected as RBF. Two hyperparameters need to be adjusted in SVM: one is the penalty coefficient $C$, and the other is the gamma. The parameter values (ranges) are shown in Table 10. The parameters of PSO are mentioned in the 'Comparison of ASCS and other optimization algorithms' section.

The Bayesian Optimization is used to adjust the hyperparameters of WaveNet, and select ReLU as the activation function and Adam as the optimizer. Four hyperparameters need to be adjusted in WaveNet: the learning rate ($lr$), the hidden layers ($l$), the value of dropout ($d$), and the node of fully connected layers ($fn$). The parameters values (ranges) are followed in Table 11.

The GA is used to adjust the hyperparameters of BP, and select ReLU as the activation function and Adam as the optimizer. Three hyperparameters need to be adjusted in BP: the learning rate ($lr$), the hidden layers ($l$), and the node of hidden layers ($n$). The parameters values (ranges) are followed in Table 12. The parameters of GA are mentioned in the 'Comparison of ASCS and other optimization algorithms' section.

The Random Searching is used to adjust the hyperparameters of CNN, and select ReLU as the activation function and Adam as the optimizer. Three hyperparameters need to be adjusted in CNN: the hidden layers ($l$), the value of kernel_size ($k$), and filters ($f$). The parameters values (ranges) are followed in Table 13.

Table 9 | Parameters of ASCS_LSTM_ATT

| Parameters                  | Qinhuai | Tunxi |
|-----------------------------|---------|-------|
| Learning rate               | 0.009   | 0.001 |
| Dropout                     | 0.3     | 0.3   |
| No. of inputs               | 54      | 72    |
| No. of outputs              | 1       | 1     |
| No. of hidden layers        | 4       | 6     |
| No. of hidden nodes         | 64      | 120   |
| Activation function and optimizer | ReLU, Adam | ReLU, Adam |

Table 10 | Parameters of SVM

| Parameters | Qinhuai | Tunxi |
|------------|---------|-------|
| max_iter   | 100     | 120   |
| $M$        | 20      | 20    |
| $D$        | 2       | 2     |
| $W$        | 0.5     | 0.2   |
| Learning factor | $c_1 = 1.2, c_2 = 1.5$ | $c_1 = 1.8, c_2 = 1.3$ |
| The range of $C$ | [1, 30] | [1, 30] |
| The range of $d$ | [0.2, 0.6] | [0.2, 0.7] |
| Final result | $C = 5.12, \gamma = 1.2$ | $C = 20, \gamma = 0.05$ |

Table 11 | Parameters of WaveNet

| Parameters                  | Qinhuai | Tunxi |
|-----------------------------|---------|-------|
| The range of $lr$           | [0.001, 0.1] | [0.001, 0.1] |
| The range of $l$            | [1, 3]  | [1, 4] |
| The range of $d$            | [0.2, 0.6] | [0.2, 0.7] |
| The range of $fn$           | [1, 100] | [1, 120] |
| Final result $lr = 0.001, l = 3, d = 0.5, fn = 54$ | $lr = 0.009, l = 4, d = 0.4, fn = 72$ |

Table 12 | Parameters of BP

| Parameters | Qinhuai | Tunxi |
|------------|---------|-------|
| max_iter   | 100     | 125   |
| $M_p$      | 0.01    | 0.06  |
| $C_p$      | 0.8     | 0.75  |
| $Size$     | 20      | 20    |
| The range of $lr$ | [0.001, 0.1] | [0.001, 0.1] |
| The range of $n$ | [40, 150] | [40, 150] |
| The range of $l$ | [1, 3]  | [1, 4] |
| Final result $lr = 0.001, l = 2, n = 64,32$ | $lr = 0.009, l = 3, n = 120,72,36$ |
Table 14 shows the NSE and RMSE values of LSTM and SVM, BP, CNN, and WaveNet in the two catchments. The results show that one distinct feature is the satisfying performance of all indicators of the ASCS_LSTM_ATT model during the simulation process of the two catchments to those of the other models.

What can be clearly seen in Table 14 is the NSE of LSTM is higher than four other models, indicating that LSTM is relatively stable in the water level prediction of these two catchments and the prediction accuracy is relatively high. The NSE of LSTM generally ranges from 0.82 to 0.97 and reveals that there has been a gradual decrease in the forecast period of 1–6 h, which means that the simulation accuracy of water level forecast of LSTM decreases gradually. Furthermore, the LSTM had a narrow range of RMSE between 0.012 and 0.042 relative to the four other models, which implies that the simulated water level of LSTM was close to observed values and the error is small. These results suggest that the LSTM has higher accuracy, a lower error of water level prediction, and a better simulation effect of water level prediction than the other four models.

Although SVM has been widely used in hydrological forecasting and the RMSE of SVM on water level prediction is similar to those of LSTM in the two catchments. LSTM tend to perform better than SVM on NSE, which shows that the water level simulation effect and prediction accuracy in the two catchments of SVM is weaker than that of LSTM.

By contrast, the RMSE of BP, CNN, and WaveNet are relatively large, which means that the predicted values of the above three models fluctuate greatly, resulting in large prediction errors. Meanwhile, Table 14 illustrates the BP, CNN, and WaveNet that tend to perform worse than LSTM on the value of NSE, indicating that the water level prediction accuracy of the above three models is relatively low in the two catchments.

Overall, the experiments suggest that the structure of LSTM enables a preferable illustration of the hydrological processes because the memory cell of the LSTM model provides a way to mimic the water level ability of the catchments.

Figures 12 and 13 show the simulated water level hydrographs of different models (from 9:00 on 28 May 2018 to 15:00 on 3 June 2018, from 6:00 on 14 July 2018 to 2:00 on 20 July 2018). The figures demonstrate that ASCS_LSTM_ATT can better predict the trend of water level hydrograph and the trough and peak value of the given watershed time series than the other models. Figures 14–16 reveal that ASCS_LSTM_ATT has a small error and
high precision, which are able to be used for timely and effective flood warning.

Figures 12 and 13 show that the lag phase between the predicted hydrograph by ASCS_LSTM_ATT and the observed one increases with the forecast time. In Qinhuai, the forecast hydrograph lags behind when the forecast periods are 5 and 6 h. The lag phenomenon in Tunxi also has a certain performance. This phenomenon manifests that the increase in the forecast period may lead to the accumulation of errors, thereby resulting in hysteresis. The noticeable characteristic of ASCS_LSTM_ATT is the gradual failure of the model to predict the trend of the observed time series with the increase in lead time, especially the low and moderate water levels. The other prediction models, such as BP and WaveNet, poorly performed in the two catchments. In the prediction period of 1–6 h, the prediction curves of BP, CNN, and WaveNet considerably fluctuated, and the prediction water level value error in the trough and peak is large. During frequent precipitation, the jitter of BP, CNN, and WaveNet is serious with the increase in the forecast period, thereby affecting the water level prediction in the basin.

Considering that the three models have the same input as the ASCS_LSTM_ATT model, the low prediction accuracy may be due to the effective hydrological simulation method provided by the LSTM model to deal with hydrological time series. Such a model has the ability to further extract the key features of each prediction factor through the self-attention mechanism. The CNN and WaveNet always use 1D convolution to deal with water level time series based on the convolution and dilated convolution layers. The flood events lead to considerable

Figure 12 | Comparison of different models in Qinhuai Basin. The prediction periods are as follows: (a) 1 h, (b) 2 h, (c) 3 h, (d) 4 h, (e) 5 h, and (f) 6 h.
uncertainty in the parameters of the CNN and WaveNet models, which may also be part of the reason for the aforementioned uncertainty.

The water level curves in Figure 12 show that the SVM overestimated the maximum level even though it could dynamically and accurately predict the change in water level. Moreover, ASCS_LSTM_ATT is superior to SVM. The curve of the SVM prediction process shows a certain degree of jitter with the increase in prediction time, and the error is large. During frequent precipitation, the SVM prediction curve gradually deviates from the trend of the real value curve, thereby reducing the prediction accuracy. By contrast, the range of the ASCS_LSTM_ATT jitter is small, and the overall trend fits the real value curve. Figure 13 shows that SVM performance becomes gradually worse with the increase in the forecast time. When the prediction period is 4 h, the SVM poorly performed in terms of predicting the trough and peak of the water level, and the
prediction curve tended to be gentle. This phenomenon reduced prediction accuracy. In the two above-mentioned cases, this phenomenon may be due to the limited data samples and the violent oscillation of water level time series in a certain period. Thus, SVM had low prediction accuracy.

Different evaluation criteria, which are important for the simulation of the model prediction process line, are applied in this study. However, the content of each evaluation index is limited. Thus, the comprehensive performance of the model must be considered before accepting or rejecting the model based on the evaluation criteria.

Figures 14 and 15 show five statistical measures of performance used to assess the model performance for the two catchments.

Figures 14 and 15 show that one distinct feature is the superior performance of all indicators of the ASCS_LSTM_ATT model during the simulation process of the two catchments to those of the other models. The performance of the prediction model varies with different prediction periods. The ASCS_LSTM_ATT is able to efficiently predict the water level of the two catchments with a prediction period of 6 h according to $R^2$ and NSE. The MSE, RMSE, and MAE indicated that the ASCS_LSTM_ATT model is capable of reducing the prediction error and further improve the prediction accuracy.

In Qinhuai, the $R^2$ and NSE values of ASCS_LSTM_ATT indicate the efficient performance of ASCS_LSTM_ATT in the forecast period of 1–6 h. The SVM efficiently performed in the prediction period of 1–4 h and decreased from
the fifth hour. In Tunxi, ASCS_LSTM_ATT efficiently performed in the forecast period of 1–6 h. The SVM model satisfactorily performed in the prediction period of 1–3 h, and its accuracy decreased after the fourth hour. The prediction error of SVM is close to that of ASCS_LSTM_ATT in terms of RMSE, MSE, and MAE in the Qinhuai. The errors in the Tunxi are larger than those of the Qinhuai. The characteristics of flood events vary considering the different topographic features of the two catchments. Therefore, the SVM performance in the two catchments is distinct.

The BP, WaveNet, and CNN performance is poor in the Qinhuai and Tunxi basins. Although the evaluation index values of the CNN and WaveNet models are similar in the Qinhuai, these values are lower than those of the other prediction models with large errors. Figures 14 and 15 reveal that the $R^2$ and NSE of the BP model began to decline from the third hour in the prediction period, and the values at the third hour in the prediction period are 0.78 and 0.75. The BP model has large errors and low precision in terms of RMSE, MSE, and MAE.

The water level prediction methods in other references are compared to further evaluate the performance of the proposed model. The $R^2$ is taken as the evaluation standard. The forecast period of 1 h is taken as an example, as shown in Table 15.

Table 15 demonstrates that the $R^2$ of the ASCS_LSTM_ATT model is larger than that of all other models, which indicates the high prediction accuracy of the model. The basins verified in Liang et al. (2018) and Hong et al. (2014) do not belong to small- and medium-sized rivers. Thus, in the water level prediction of the two catchments in this study, the model performance proposed in the references is not as good as that of ASCS_LSTM_ATT. The model proposed in Yang et al. (2017), Yanwei et al. (2011), and Quanyin & Junfeng (2009) has been verified on small- and medium-sized rivers. However, the prediction error of
Figure 16 | Regression scatter map of different basins. (a) Regression scatter map of Qinhuai and (b) Regression scatter map of Tunxi.
the model in the aforementioned references is larger than that of ASCS_LSTM_ATT. Therefore, the prediction accuracy obtained by the proposed model outperforms those obtained by the comparison methods. This finding demonstrates the excellent water level prediction performance in small- and medium-sized rivers of the proposed model.

Figure 16 shows that ASCS_LSTM_ATT efficiently performs in the Qinhuai and Tunxi basins. In the forecast period of 1–6 h, points are distributed along the regression line. However, the points began to have a widespread range from the line with an increase in the forecast period. Therefore, the correlation coefficient between the real and the predicted water levels decreased with the increase in the prediction period, and the model performance also gradually decreased.

In the comparative analysis of the above-mentioned models, the performance is as follows: ASCS_LSTM_ATT > SVM > BP > WaveNet > CNN. Figures 14–16 demonstrate that ASCS_LSTM_ATT has the ability to efficiently predict the water level in Tunxi and Qinhuai basins with relatively high precision. Thus, ASCS_LSTM_ATT is crucial in early warning and prediction.

## CONCLUSION

This study proposes an improved CS algorithm to optimize the hydrological time series prediction method based on the combination of LSTM and self-attention mechanism. Such an undertaking is performed to solve the problem of water level prediction of small- and medium-sized rivers with different topographic features in China. The experimental results show that the improved CS algorithm is able to quickly find the optimal solution by considering the global and local search performance. Various methods for determining the forecast factors of small- and medium-sized rivers are proposed on the basis of different topographic features to realize consistency between the predicted and the real value. Moreover, combining the self-attention mechanism and the LSTM helps further extract the key features by assigning different weights to the forecast factors. Accordingly, the LSTM could effectively capture the dependence between time series and improve the accuracy of the model prediction results.

Although some achievements have been made, many problems must still be solved. One problem is that the forecast factors only select water level and precipitation information. However, the model input should also include soil moisture, temperature, meteorological, and hydrological information. Using the model for predicting other hydrological factors, such as stream flow and precipitation, is also the focus of future research.

## FUNDING

This research has been supported by the National Key R&D Program of China (No. 2018YFC1508100).

## DATA AVAILABILITY STATEMENT

Data cannot be made publicly available; readers should contact the corresponding author for details.

## REFERENCES

Akbari Asanjan, A., Yang, T., Hsu, K., Sorooshian, S., Lin, J. & Peng, Q. 2018 Short-term precipitation forecast based on the PERSIANN system and LSTM recurrent neural networks. *Journal of Geophysical Research: Atmospheres* **123**, 12543–12563.

Anvari, S., Jamshid Mousavi, S. & Morid, S. 2014 Sampling/stochastic dynamic programming for optimal operation of multi-purpose reservoirs using artificial neural network-based ensemble streamflow predictions. *Journal of Hydroinformatics* **16** (4), 907–921.
Bahdanau, D., Cho, K. & Bengio, Y. 2014 Neural machine translation by jointly learning to align and translate. arXiv preprint, arXiv:1409.0473.

Bai, S. & Shen, X. 2019 PM2.5 prediction based on LSTM recurrent neural network. Computer Applications and Software 36 (01), 67–70 + 104.

Chang, F., Tsai, Y., Chen, P., Cowney, A. & Vachaud, G. 2015 Modeling water quality in an urban river using hydrological factors – data driven approaches. Journal of Environmental Management 151, 87–96.

Chen, N., Xiong, C., Du, W., Wang, C., Lin, X. & Chen, Z. 2019 An improved genetic algorithm coupling a back-propagation neural network model (IGA-BPNN) for water level predictions. Water 11 (9), 1795.

Cheng, Y., Yiqiu, J., Zhijia, L. & Kailei, L. 2012 Parameter estimation and application of grid-based Xin’anjiang model. Journal of Hohai University (Natural Sciences) 40 (01), 42–47.

Cho, K., van Merrienboer, B., Gülçehre, Ç., Bougares, F., Schwenk, H. & Bengio, Y. 2014 Learning phrase representations using RNN encoder-decoder for statistical machine translation. arXiv:1406.1078.

Graves, A., Mohamed, A. & Hinton, G. 2013 Speech recognition with deep recurrent neural networks. In: Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing, 26–31 May 2013, Vancouver, BC, Canada, pp. 6645–6649.

Hochreiter, S. & Schmidhuber, J. 1997 Long short-term memory. Neural Computation 9 (8), 1735–1780.

Hong, D., Dong, W., Xianghui, L. & Peiji, L. 2014 Neural network based on genetic algorithm and anneal algorithm for water level prediction. Journal of Liuzhou Teachers College 34 (04), 59–62.

Le, X.-H., Ho, H. V., Lee, G. & Jung, S. 2019 Application of long short-term memory (LSTM) neural network for flood forecasting. Water 11, 1387.

Liang, C., Li, H., Lei, M. & Du, Q. 2018 Dongting lake water level forecast and its relationship with the three gorges dam based on a long short-term memory network. Water 10, 1389.

Liu, B., Wang, L. & Jin, Y. 2006 An effective PSO-based memetic algorithm for TSP. Lecture Notes in Control & Information Sciences 345, 1151–1156.

Liu, Y., Fan, Q., Shang, Y., Fan, Q. & Liu, Z. 2019 Short-term water level prediction method for hydropower station based on LSTM neural network. Advances in Science and Technology of Water Resources 39 (02), 56–60 + 78.

Liu, M., Huang, Y., Li, Z., Tong, B., Liu, Z., Sun, M., Jiang, F. & Zhang, H. 2020 The applicability of LSTM-KNN model for real-time flood forecasting in different climate zones in China. Water 12 (2), 440.

Mei, L., jing, L., Zijian, W., Sida, W. & Laijinj, C. 2018 Short-time passenger flow forecasting at subway station based on deep learning LSTM structure. Urban Mass Transit 21 (11), 42–46 + 77.

Moeini, R. & Afshar, M. H. 2015 Extension of the constrained ant colony optimization algorithms for the optimal operation of multi-reservoir systems. Journal of Hydroinformatics 15 (1), 155–173.

Nelson, D. M. Q., Pereira, A. C. M. & de Oliveira, R. A. 2017 Stock market’s price movement prediction with LSTM neural networks. In Proceedings of the International Joint Conference on Neural Networks (IJCNN), 14–19 May 2017, Anchorage, AK, USA, pp. 1419–1426.

Piasecki, A., Juras, J. & Skowron, R. 2017 Forecasting surface water level fluctuations of lake Serwy (Northeastern Poland) by artificial neural networks and multiple linear regression. Journal of Environmental Engineering and Landscape Management 25, 379–388.

Quanyin, Z. & Junfeng, D. 2009 A water level prediction model of Hongze Lake. Computer Simulation 26 (04), 113–115 + 157.

Rush, A. M., Chopra, S. & Weston, J. 2015 A neural attention model for abstractive sentence summarization. arXiv preprint, arXiv:1509.0685.

Shuaihong, Z., Zhijia, L., Yingchun, H. & Qiaoling, L. 2019 Calculation method and application of river network regression coefficient Cs based on self-similarity of river network structure. Journal of Lake Sciences 31 (03), 788–800.

Vaswani, A., Shazeer, N. & Parmar, N. 2017 Attention is all you need. In: Advances in Neural Information Processing Systems (A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, L. Kaiser & I. Polosukhin, eds). NIPS, Long Beach, pp. 6000–6010.

Xiaoying, D., Youpeng, X., Zhixin, L., Qiang, W., Yu, X. & Bin, G. 2019 Analysis on variation and influencing factors of runoff in Qinhuaihe River Basin in the lower reaches of the Yangtze River. Research of Soil and Water Conservation 26 (04), 68–73.

Xuan, C., Hongwei, Y. & Xuan, S. 2013 Flood control situation analysis and Countermeasures of Qinhuai River Basin. Water Resources Planning and Design 10, 23–27.

Yang, X. S. & Deb, S. 2010 Engineering optimization by cuckoo search. International Journal of Mathematical Modelling and Numerical Optimization 1 (4), 330–343.

Yang, J. & Honavar, V. 2002 Feature subset selection using a genetic algorithm. IEEE Intelligent Systems & Their Applications 15 (2), 44–49.

Yang, Y., Fu, Q. & Wan, D. 2017 A prediction model for time series based on deep recurrent. Computer Technology and Development 27 (3), 35–38.

Yanwei, L., Huixue, M. & Ying, L. 2011 Application of support vector machine in flood peak water level prediction of Meichi station. Marine Forecasts 28 (01), 72–76.

Yao, C., Zhang, K., Yu, Z., Li, Z. & Li, Q. 2014 Improving the flood prediction capability of the Xinanjiang model in ungauged nested catchments by coupling it with the geomorphologic instantaneous unit hydrograph. Journal of Hydrology 517, 1035–1048.

Zhang, J., Zhu, Y., Zhang, X., Ye, M. & Yang, J. 2018 Developing a long short-term memory (LSTM) based model for predicting
water table depth in agricultural areas. *Journal of Hydrology* **561**, 918–929.

Zhao, S. & Yang, D. 2017 Mutual information-based input variable selection method for turnoff-forecasting neural network model. *Journal of Hydroelectric Engineering* **30** (01), 24–30.

Zhiyu, L. 2017 Research and application of early warning and forecasting technology for mountain torrents. *China Flood & Drought Management* **22** (02), 41–45 + 50.

Zhiyu, L., Aizhong, H. & Xiuqing, W. 2015 Flood forecasting for small- and medium-sized rivers based on distributed hydrological modeling. *Journal of China Hydrology* **35** (01), 1–6.

First received 4 March 2020; accepted in revised form 2 September 2020. Available online 15 October 2020