Effects of stacking LSTM with different patterns and input schemes on streamflow and water quality simulation

Yucong Hu[a,b], Yan Jiang[a,b*], Huiting Yao[a,b], Yiping Chen[a,b], Xuefeng Wu[a,b], Xuyong Li[a,b]

[a] State Key Laboratory of Urban and Regional Ecology, Research Center for Eco-Environmental Sciences, Chinese Academy of Sciences, Beijing 100085, China

[b] University of Chinese Academy of Sciences, Beijing 100049, China

Highlights

- Effects of LSTM with different patterns and input schemes are analyzed.
- Sliding windows is a more appropriate pattern compared to lags.
- Predicting streamflow by meteorological data is limited especially with low-volume flow.
- Predicting WQs by streamflow is reliable.
- Separately adding historical streamflow and WQs into LSTM can increase accuracy.

Abstract

Streamflow and water quality parameters (WQs) are commonly forecasted by mechanism models and statistics models. However, these models are challenged due to computational time costs, redundant parameters and several other uncertainties. Long short-term memory (LSTM) neural networks are a powerful deep learning method that provide the potential to minimize these deficiencies in a data-driven way,

* Corresponding author at Research Center for Eco-Environmental Sciences, Chinese Academy of Sciences, Beijing 100085, China.

E-mail address: yanjiang@rcees.ac.cn; Tel/Fax: +86 10-62849428

especially when stacking is used. Therefore, a stacking LSTM model for enhanced capability was applied to simulate streamflow and eight WQs in the present study. Generally, two patterns, lags and sliding windows, can be applied to LSTM models, leading to different accuracies of simulation with various input architectures. The simulation effects of each pattern were first studied, and sliding windows was detected as the pattern with the higher and more stable accuracy for both streamflow and water quality forecast. Similarly, different input schemes resulted in different simulation accuracy. Predicting streamflow with only meteorological input failed to capture peaks and was accuracy-restricted for significantly increased window sizes (from 10 to 30), with Nash-Sutcliffe efficiency coefficient (NSE) around 0.6 in high-volume rivers and much lower accuracy (<0.1) in low-volume rivers. Adding historical streamflow into the input data set caused an increase to NSE ca. 0.9 in both mainstreams and tributaries. The value is only slightly below the NSE of a mechanism (Delft 3D) model, (ca. 0.99), implying that historical streamflow should be included into input data in further forecasts. Effects of the various input variables (i.e., meteorological factors, streamflow, other influential WQs) as well as their respective combinations were individually studied in predicting WQs at each station with each LSTM model of specifically searched six hyperparameters (e.g., neurons of each layer, etc.). Our results document that WQs could be predicted by such alternative input schemes, with schemes including streamflow dominating the ideal schemes of above cases. A reliable performance level of relative error (RE) below 30% was achieved, despite of a weak capture of trends. Adding antecedent WQs into the input data caused a drop of the average value of the ideal RE of all stations by at least 48.80%. This method slightly impaired the accuracy compared to the results of Delft 3D model, but is still acceptable with RE reaching 17% at most, thus validating the modified

input schemes in WQs forecasts. Our study documents that the LSTM model with appropriate pattern and input data is an effective method for daily streamflow and water quality forecasts.

Key words: Long short-term memory (LSTM), Streamflow forecast, Water quality forecast, pattern, input schemes

1 Introduction

Human lifestyles, industry development and vegetation growth largely depend on the availability of water. Two important features of water, the streamflow and the water quality, are of specific significance, especially in commonly densely-populated river areas. River discharge has changed dramatically over the last years in 24% of the world's main rivers (Li et al., 2020). Also, human activities have strongly degraded the cleanliness of the water, and the water quality of rivers is adversely affected by harmful pollutants (Jaffar et al., 2022). Without reliable control and management systems, the increasing water pollution and the fluctuating river volume can unexpectedly endanger adjoining areas. Therefore, accurate and reliable prediction of streamflow and water quality are of fundamental significance to regional water security and provide guidelines for regional water management.

Generally, mechanism models and statistical models are the two methods commonly used for prediction. To achieve reliable forecasts of the streamflow, numerous hydrological models have been developed and classified including empirical, conceptual, and process-based models (Cho and Kim, 2022; Dadson et al., 2019; Peng et al., 2022). Likewise, prediction models for the water quality were significantly improved in recent years from single factor, steady-state model to multiple factors and multiple dimensional models, with concomitant increase of accuracy and complexity (Wang et al., 2011). Although the improved models are more effective and closer to the reality, exemplified by process-based models, some deficiencies still persist insolvable with the state-of-the-art knowledge and technology, such as computational time costs, redundant parameters and uncertainties of model structures (Alizadeh et al., 2021; Wan et al., 2022). Although statistical models, including the Regression Trees (Stidson et al., 2012), and the family of ARIMA (Auto-regressive integrated moving average) models (Valipour et al., 2013), have been reasonably utilized in the prediction of sequence data, such as streamflow and water quality parameters or indicators (WQs for short), their application is limited by the incapability of highly non-linear processes (Najah et al., 2013). Nevertheless, despite of these deficiencies, prediction of streamflow has been demonstrated with a high precision, due to the easily measured long-term data (Yaseen et al., 2016). In contrast, none of the mechanism models or statistical models for predicting WQs work as well as in predicting the streamflow. And it is especially prominent for some intricate parameters related to nitrogen and phosphorous, which encouraged further studies, focusing on alternative methods to address those remaining problems.

Purely simulating relations between input and output data, Artificial Intelligence (AI) has thoroughly overturned the state of modeling and significantly improved computational efficiency, with minor prior knowledge required (Bai et al., 2021). Since its implementation into natural sciences, AI essentially improved the trade-off between time costs and precision in many disciplines, also in hydrology. Specifically, machine learning, as a part of AI, comprises several innovative methods to solve major problems. Among those methods, multiple linear regression and non-linear models, such as Support Vector Regression (SVR), Bayesian Neural Network (BNN), and Gaussian Process (GP) have become popular in streamflow forecasts (Rasouli et al., 2012). Likewise, Artificial Neural Networks (ANN) and Adaptive Neuro-Fuzzy Inference System (ANFIS) have become the most utilized methods for WQ simulation, with popular investigated WQs such as biochemical oxygen demand, chemical oxygen demand, and dissolved oxygen being modelled (Ighalo et al., 2021). Although machine learning is very effective, some drawbacks, including learning divergence, poor generalizing performance, local minimum and over-fitting problems must be considered (Ghimire et al., 2021). To address these problems in highly non-linear processes, deep learning, the subdomain of machine learning, is a more powerful tool (Barzegar et al., 2020). Specifically for sequence data, Recurrent Neural Networks (RNN) yields reliable results in the short-term period, whereas on the other hand, in the long run, the results are comparably weak, due to gradient exploding and vanishing (Chen et al., 2018). Based on these findings, a further improved model, the Long short-term Memory Neural Networks (LSTM) has been designed to deal with longer data by maintaining constant errors in gradient calculations. Compared to other machine learning models or other more simple versions of RNN, like Gated Recurrent Neural Networks (GRU), LSTM yields more reliable results in hydrology (Kratzert et al., 2018; Shen, 2018), and has been successfully applied in streamflow forecasts (Lees, 2022) and WQs simulations (Wang et al., 2017).

In terms of streamflow, apart from the general prediction with antecedent values, machine learning models have further explored for its capabilities. However, it is widely acknowledged that machine learning fails in simulating streamflow in the absence of historical streamflow, and thus models such as SVR, ANN and Random Forest (RF) are improved by coupling with other methods, like flow separation method as a function of precipitation, temperature and potential evapotranspiration (Tongal and Booij, 2018). LSTM was originally introduced to rainfall-runoff models to explore its potential in predicting a variety of catchments (Kratzert et al., 2018). Subsequently, it was gradually extended to be compared and combined with other mechanism models (Xiang et al., 2020), validating capabilities of reasonably capturing trends but more biases (Cho and Kim, 2022), and also applied to explorations of fixed time steps and lead times of sliding window approach in sample generation as well as effects of adding extra meteorological data (Hu et al., 2020). However, in an hourly prediction of short-term runoff, the optimal time step of LSTM was detected to be unnecessary for the gradually stable accuracy, as time steps increase (Gao et al., 2020). Concerning the input data, imbalanced mass conservation was recognized in LSTM in the relationship of precipitation and flow discharge in the snow melting period (Yokoo et al., 2022), indicating its incapability besides energy conservation (Jia et al., 2019). Thus, concerning streamflow, samples are mainly generated on short-term runoff data in hours, with rare consideration on other units, like days. Correspondingly, the capability of LSTM with only meteorological factors (temperature plus precipitation) as input, and the difference between main streams and tributaries, have not been systematically studied in different situations so far.

Likewise, concerning WQs forecast, machine learning models, represented by ANN, have been well utilized in predicting certain WQs with various input variables, not only considering influences of other WQs but also taking meteorological factors and streamflow into account (Najah Ahmed et al., 2019). LSTM has also been applied in simulating WQs, such as chlorophyll, total phosphorous and dissolved oxygen (Wang et al., 2017). In addition, it was used in comparison (Liang et al., 2020) and combination (Zhang et al., 2022) with complicated water quality models, for its capability in dealing with non-linear processes. However, comparative studies among different WQs (Li et al., 2023) and different input schemes are rare. In essence, the effects of each type of input variable and their combinations, as well as the historical WQs data have not been systematically studied, in clear contrast with the progresses with machine learning methods (Zhu et al., 2022) .

LSTM is a well-developed model with various hyper-parameters to be tuned. Moreover, two patterns can be selected, in addition to specific strategies like Dropout (Srivastava et al., 2014) and Early Stopping (Prechelt, 1998) in case of overfitting, as well as stacking (Zhang et al., 2020) in the purpose of improving the simulation capabilities. The hyper-parameters, consisting of model structures, were subject of several studies (Wang et al., 2017). However, the patterns of different input architectures, i.e., lags and sliding windows of various time steps and lead times, were poorly studied, leading to the ignorance on reasonably choosing appropriate pattern and direct decisions on sliding windows in sample generation (Chen et al., 2022; Muzaffar and Afshari, 2019; Solgi et al., 2021). For a LSTM model that is to simulate two different and complex non-linear processes, i.e., streamflow and water quality, it is essential to appropriately select the running pattern for each process with variation of time steps and lead times. In terms of input data, effects of separate and compositional input factors pertaining to each mechanical process, also require systematic studies on the basis of suitable pattern.

In our study, LSTM models are well devised by hyperparameters searched 30 times with 5-folds cross-validation, to predict streamflow and water quality parameters with the two patterns and different input schemes in the middle reach of the Yangtze River in China. The objectives of our research are twofold: (i) Choice of the appropriate pattern of LSTM in the prediction of streamflow and water quality. (ii) Based on the proper pattern, proving the feasibility of LSTM with different input schemes through a comparison with results of a mechanism model, Delft 3D model, so as to explore the plausibility in streamflow and water quality forecasts with alternative input schemes.

2 Materials and data

2.1 Study area

The research watershed is situated in the Wuhan Metropolitan Area (110°38'N-115°60'N, 29°45'E-32°5'E) (Fig. 1), an urban cluster consisting of nine cities located in central China. The studied watershed constitutes the major water source for this region and is characterized by predominantly flat terrain surrounded by mountainous areas. The forest coverage within the watershed exceeds 25%. It encompasses the main channel of the Yangtze River and various tributaries such as the Han River and the Dongting Lake. The watershed extends from Luoshan town in Honghu city (Hubei Province) to Jiujiang city (Jiangxi Province), with a total water resource volume of 410.73 M km^3 (in 2021). Geographically, the watershed is situated in a transitional zone between the mid-latitude northwest wind belt and the low-latitude easterly wind belt, exhibiting typical characteristics of a subtropical humid monsoon climate. This climate is characterized by high temperatures with distinct seasonal variations and concurrent rainfall. The average annual temperature ranges from 16.3 to 16.8°C, with occasional extreme temperatures > 40 ℃. Annual precipitation ranges from 1130 to 1600mm, with the highest rainfall from April to October and lower rainfall from November to March. The mean annual streamflow recorded during the period from

2004 to 2019 was approximately 6084.79 m^3 per day. Key indicators of water pollutant levels in the watershed include total phosphorus, ammonia nitrogen, chemical oxygen demand, and permanganate index. Industrial sources contribute 12% and 11% to the total emissions of ammonia nitrogen and total phosphorus, amounting to 2.86 M tons/year and 4.34 M tons/year, respectively. Meanwhile, domestic sources contribute 87% and 89% to these emissions individually (Chong et al., 2023).

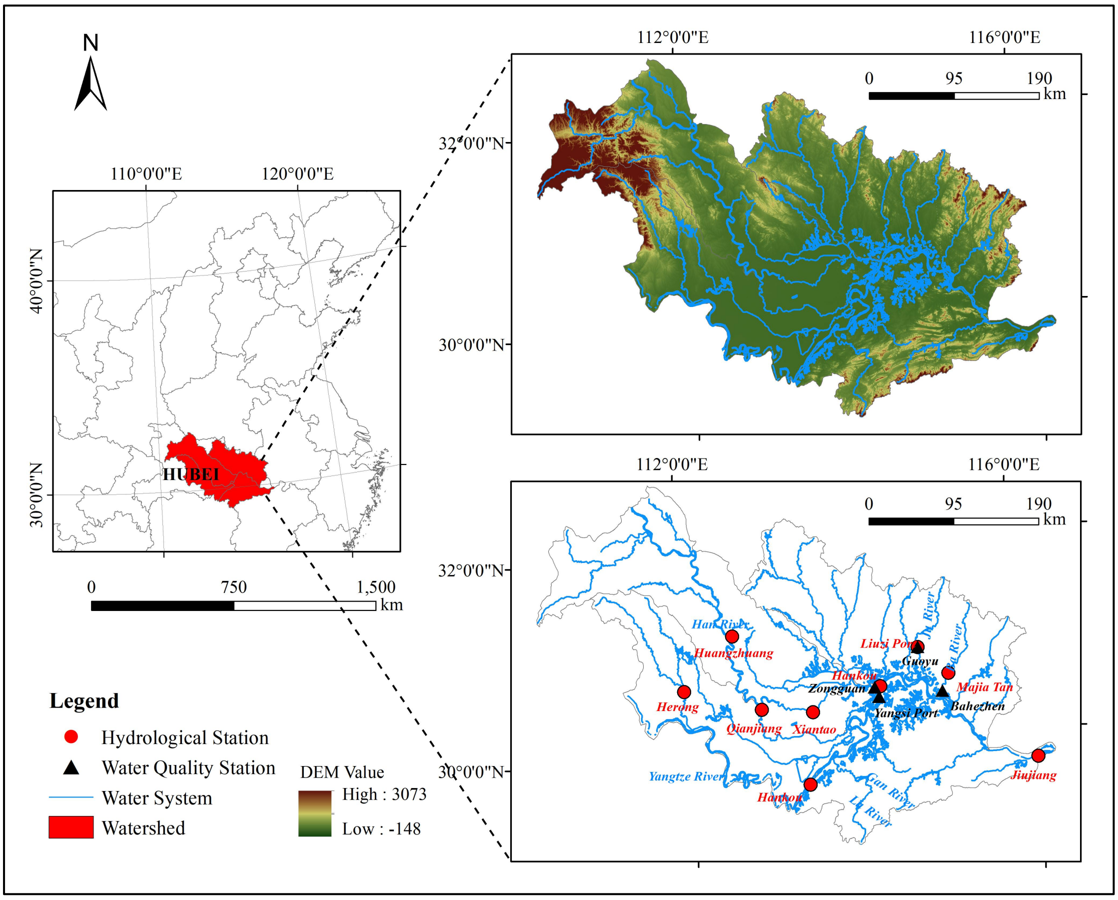

Fig. 1. Location and hydrology of the research area with the studied hydrological and water quality stations

2.2 Data collection

Samples of streamflow used in this study were collected at nine hydrological stations (Fig. 1) daily from 2004 to 2019, encompassing a total of 5,844 days. At four water quality monitoring stations within the research area, samples were collected for a period ranging from November 1, 2018, to December 1, 2019, in a total of 396 days. The study included eight water quality parameters, namely pH, dissolved oxygen (DO), ammonia nitrogen (NH_3-N), permanganate index (COD_{Mn}), turbidity (TU), electronic conductivity (EC), total nitrogen (TN), and total phosphorus (TP). In instances where data was missing, linear interpolation was employed. Fig. 1 depicts the position of the nine hydrological stations and four water quality stations, and Table 1 comprises the corresponding matches. The raw data for each water quality parameter at the Yangsi Port station is presented in Fig. 2, and Fig. 3 displays the raw data for streamflow at the Hankou station, with the training and test sets outlined. Due to the limited amount of available water quality data, only the training and test sets were utilized in our study to assess the feasibility, roughly one year for training and one month for testing. The validation set will be incorporated in the model architecture for hyperparameter optimization.

Table 1

Gauging stations and their relation

River Type	Rive name	Hydrological stations	Water quality stations
Main stream	The Yangtze River	Luoshan	Yangsi Port
		Jiujiang	
Tributaries	Han River	Huangzhuang	
		Xiantao	
		Hankou	Zongguan
	Juzhang River	Herong	
	Dongjing River	Qianjiang	
	Ju River	Liuzi Port	Guoyu
	Ba River	Majiatan	Bahezhen

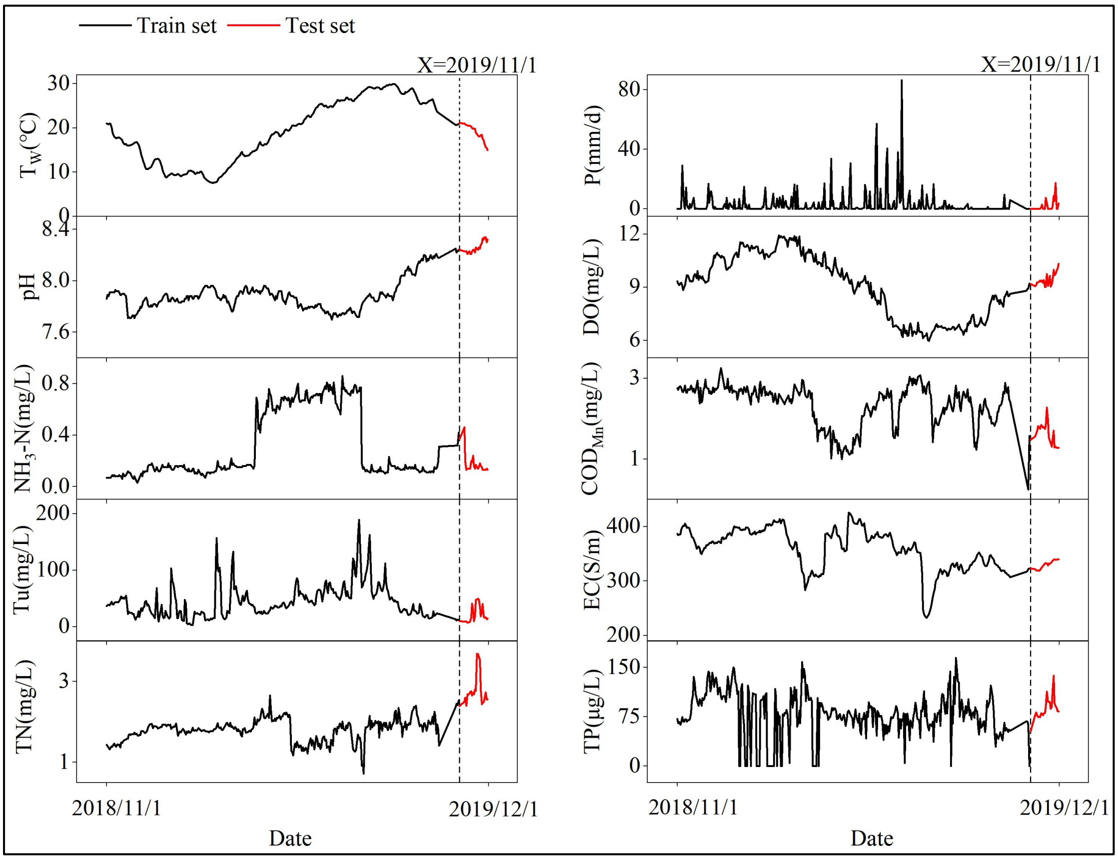

Fig. 2. Meteorological raw data (precipitation, P; water temperature, T_w) and water quality parameters at station Yangsi Port

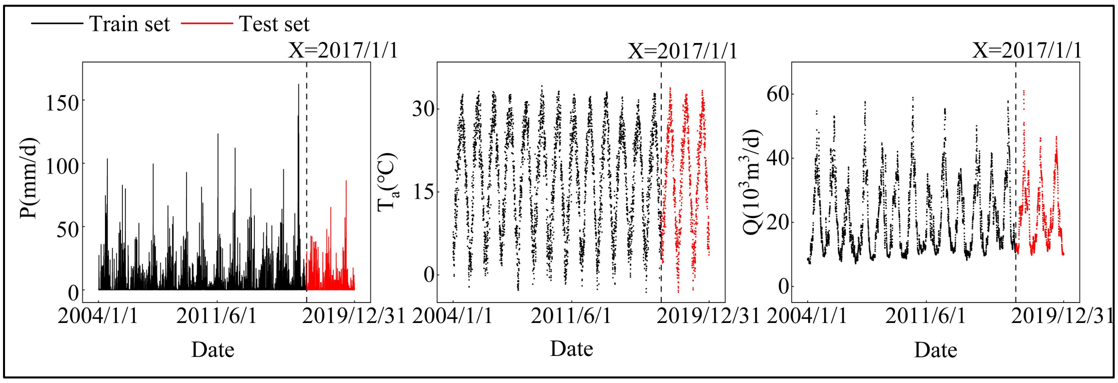

Fig. 3. Meteorological input data (precipitation, P; air temperature, Ta) and streamflow of station Hankou.

2.3 Data pre-processing

To eliminate units' affection on the learning ability of models, the input data is standardized by the z-score method and is scaled according to the mean and variance of the training sets. The formula is shown in Eq. (1):

$$x_{ijnew} = \frac{x_{ij} - \overline{x_j}}{S_j} \tag{1}$$

where, $\overline{x_j}$ is the mean value of the j column of input data and S_j is the standard deviation of the j column of input data.

3 Methods

3.1 LSTM and its two patterns

Sequence data possess distinct characteristics. The knowledge it encompasses at a specific moment can be significantly influenced by adjacent information or strongly depend on information received a considerable time earlier. Consequently, conventional backpropagation neural networks exhibit limited learning capabilities in handling such data. To address this limitation, Recurrent Neural Networks (RNN) have been developed. It is structured as interconnected units, wherein each unit not only receives input data at the present time step but also obtains output data from the preceding unit in the sequence (Fig. 4).

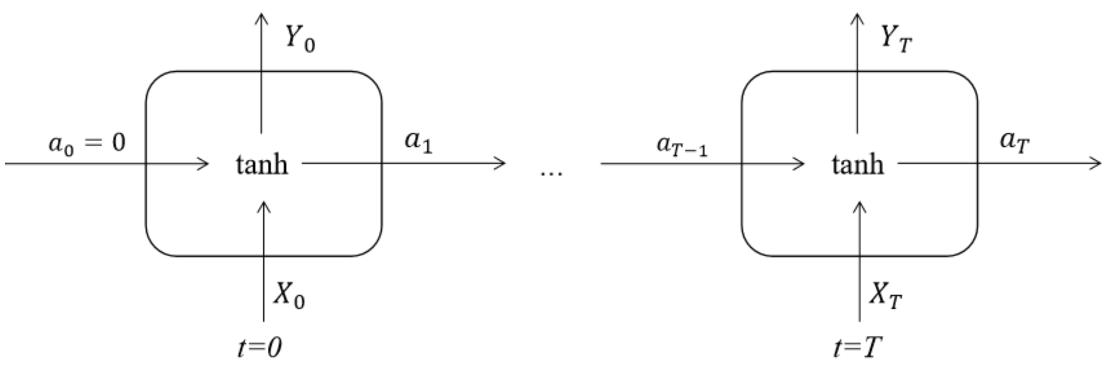

Fig. 4 Structures of RNN

$$a_T = \tanh(U[X_T, a_{T-1}] + b_a) \tag{2}$$

$$Y_T = g(Va_T + b_Y) \tag{3}$$

Where X is sequence data divided by T times. a is the hidden output from previous moments, and b_a , b_Y are both biases. Through multiplication of weight matrices U, V and activation of nonlinear functions, *tanh* and g, the output Y at a certain time is obtained. Although the use of the *tanh* function can reduce the speed of gradient descent, which strengthens the ability to effectively learn information from longer time, RNN is still difficult to cope with the increasing amount of sequence data. This results in gradient vanishing or exploding, which, therefore, falls short of requirements of training longer data (Pascanu et al., 2013). However, by adding constant error carousel, LSTM can keep error streamflow linear, which could learn more than 1000 lag steps (Hochreiter and Schmidhuber, 1997), and by adding forgetting gates, it can adapt to changes in the longer run(Gers et al., 2000). The basic structure of LSTM is shown in Fig.5, with related equations 4 to 9.

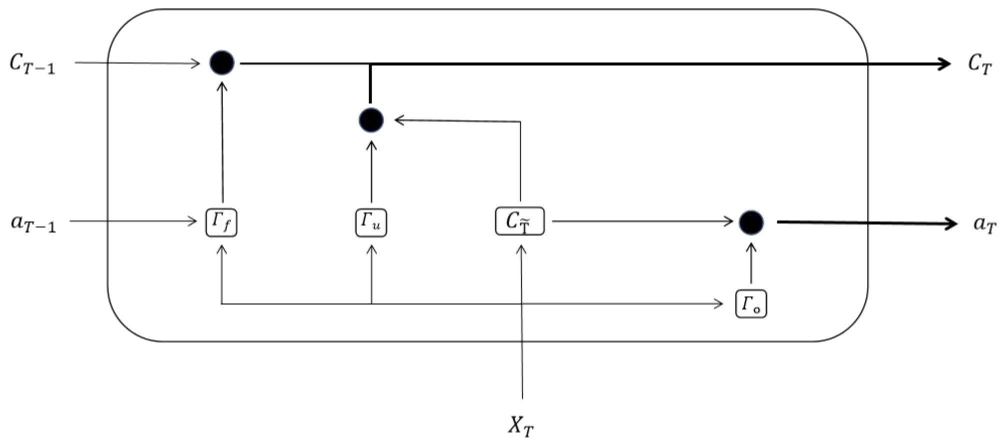

Fig. 5. basic unit of LSTM

$$C_{\tilde{T}} = \tanh(W_C[X_T, a_{T-1}] + b_C) \tag{4}$$

$$\Gamma_f = \sigma(W_f[X_T, a_{T-1}] + b_f) \tag{5}$$

$$\Gamma_u = \sigma(W_u[X_T, a_{T-1}] + b_u) \tag{6}$$

$$\Gamma_o = \sigma(W_o[X_T, a_{T-1}] + b_o) \tag{7}$$

$$C_T = \Gamma_u \cdot C_{\tilde{T}} + \Gamma_f \cdot C_{T-1} \tag{8}$$

$$a_T = \Gamma_o \cdot \tanh(C_T) \tag{9}$$

X_T is the input data of time series, X at T moment. a_{T-1} is the hidden output of time T-1, while b_C, b_f, b_u, and b_o are all biases, and W_C, W_f, W_u, as well as W_o are weight matrices of all time. Activated by logistic sigmodal function σ, the four gates, $C_{\tilde{T}}$, Γ_f, Γ_u and Γ_o, all react to X_T and a_{T-1}. Forget gate Γ_f controls the extent of the information to be left out, and update gate Γ_u controls the extent of the information to be saved. $\cdot$ represents element-wise multiplication of matrices. C_T is the final information defined by update gate and forget gate. Through output gate Γ_o, it would become hidden output a_{T-1}, passed to next $T+1$ time. Single LSTM layer could ensure that gradient descent proceeds without gradient disappearance or gradient explosion, but multiple stacked LSTM have been reported to be more efficient for learning potential patterns between data (Lipton et al., 2015).

LSTM can be used to predict a single target or many target at each time step, and we only focus on one target calculated at one time step for better clarity in our study. Two patterns, sliding windows and lags are often separately applied in time series forecasts. The former utilizes the combined past *k* time-steps (X_{t0}-X_{tk}) data, also termed as window size, as input data to predict the next *m* lead-times target (Chen et al., 2022). In contrast, the latter sets lags to construct one-to-one relationship between input from the previous *k* lag and target output at the next time step (Muzaffar and Afshari, 2019; Solgi et al., 2021). To compare the many-to-one architecture and one-to-one architecture (Camps-Valls et al., 2021), the lead times in sliding windows are not considered. Therefore, output only target at the next one-time step, rather than the next few steps. Since both patterns yield reasonable results but are not compared before research, a preference should be made prior to the experiments.

3.2 Improved model framework

Based on the progress of deep learning, it was recognized that models with complex structures and many parameters tend to provide stronger learning capabilities (Collins et al., 2016). However, in case of limited available data is limited and the excessive number of parameters, overfitting becomes a considerable concern. Various effective techniques have been developed to address this problem. In our study, we selected two commonly used methods, dropout (Srivastava et al., 2014) and early stopping (Prechelt, 1998), to minimize the risk of overfitting. By adding a dropout layer with a given proportion, some neurons of the last layer become stochastically forced not to update their weights in the next layer, thus contributing to prevent over-reliance on specific features and promoting the development of more robust representations. Early stopping monitors the changes in error or other evaluation metrics between consecutive training epochs. In case of a drop of the difference below the predefined threshold for a certain number of epochs, the model is considered to have attained a stable state and training is stopped to prevent further overfitting.

LSTM comprises multiple hyper-parameters to be adjusted for the ideal performance. Searching for grids to test every combination of all hyper-parameters within the ranges is demanding. We therefore use RandomizedSearchCV (Buitinck et al., 2013), an API in scikit-learn package from Python, to randomly detect the ideal combination of hyper-parameters in ranges, a method that has been reported to be more efficient than grid search (Bergstra and Bengio, 2012). To reduce the mean squared error (MSE) (Wallach and Goffinet, 1989) for each combination of hyperparameters, we applied 30 times iteration and 5-folds cross validation, accounting for 10% in the training set, while searching and learning in iterations. After that, the input data would be trained 500 epochs with those selected hyperparameters. Those hyper-parameters contain neurons of layers (L), proportion of each dropout layer (D) and lags or size of sliding window (m). Searching ranges of each kind of hyper-parameters remain the same (Table 2).

Table 2

Ranges of hyper-parameters

	Layer	D	m
Ranges	[8,12,16,20,24,28,32,36,40,44,48,52,56,60,64]	[0.1,0.2,0.3,0.4]	[2,3,4,5,6,7,8,9]

Streamflow and water quality parameters are both strongly affected by previous periods and are characterized by periodic changes. To learn more about the sequence data, the model structure is devised with stacked layers to decode more relations (Zhang et al., 2020), with two dropout layers sandwiched between three stacked LSTM layers and a fully connected layer in the tail to improve the learning ability (Fig. 6). Our code is programmed with Python, and calculated on the AMD EPYC™ ROME 7H12 CPU with 128 cores, 2G RAM per core. Loss function is defined as Equation (10):

$$Loss = \sum_{i=1}^{T} (y_i \times \lg(y_i^{hat}) + (1 - y_i \times (1 - \lg(y_i^{hat})))) \tag{10}$$

where y_i is measured value at time i, and y_i^{hat} is the prediction value at time i.

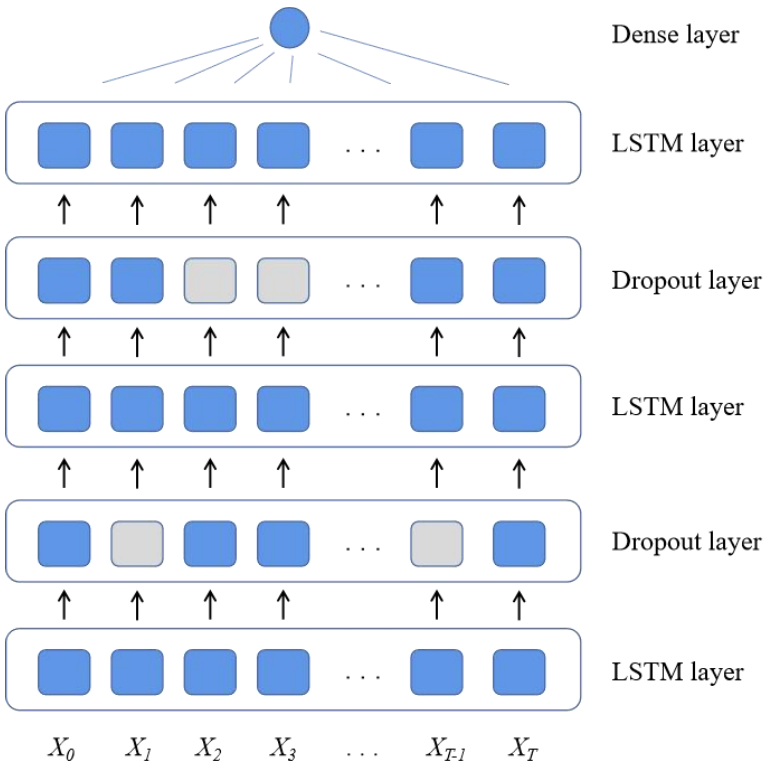

Fig. 6. Structure of the LSTM

3.3 Model comparison and performance evaluation

To test the performance of our LSTM model, we compared it with the Delft 3-Dimension (Delft 3D) model, a mechanism model used to simulate variations in hydrodynamic conditions and water quality (Roelvink and Van Banning, 1995). Based on river topography, boundary conditions of water level and etc., and roughness data, Delft 3D constructs a hydrodynamic module that generates essential information regarding flow and water levels within the simulation area, as well as at specified cross sections. Following the calibration of the hydrodynamic module, the water quality module can be constructed by selecting the appropriate process library, allowing the simulation of specific water quality indicators such as NH_3-N, TN, and TP. Simulating hydrodynamic conditions and water quality using Delft 3D has been documented as an efficient and credible method in several studies (Bai et al., 2022). Therefore, it is settled as the appropriate benchmark to evaluate the performance of

LSTM without any improvement.

We adopted Nash-Sutcliffe efficiency coefficient (NSE) and Relative error (RE) for the evaluation in our study. NSE is often utilized to evaluate hydrological models, ranging from $-\infty$ to 1, with a better accuracy close to 1. Contrarily, the increase of RE causes decrease of the performance. Related formulae are Eq (11&12):

$$NSE = 1 - \frac{\sum_{i=1}^{N}(y_i - y_i^{hat})^2}{\sum_{i=1}^{N}(y_i - \overline{y})^2} \tag{11}$$

$$RE = \frac{\sum_{i=1}^{N}\left|y_i - y_i^{hat}\right|}{\sum y_i} \times 100\% \tag{12}$$

where y_i and y_i^{hat} is the observed and simulated data at time i respectively， with mean value of observation as $\overline{y}$.

3.4 Impacts of model complexity on model efficiency

3.4.1 Two patterns

The decision of the pattern used for predicting streamflow and water quality prior to the experiments is essential, since they are of different complexity. Therefore, comparisons were performed with lags and the length of sliding window unevenly increasing as 1, 2, 3 and 7, which was based on empirical values at first (schemes are shown in Table 3). The optimal value and prediction error of each scheme applied to each station would be recorded to decide the ideal pattern for simulating streamflow and WQs separately.

Table 3

Schemes of testing effects of lags and sliding window

Input-output	Lags / the length of sliding window
streamflow-streamflow(a)	1/2/3/7

streamflow-streamflow(b)
water quality parameter -water quality parameter(a)
water quality parameter-water quality parameter(b)
Tips：a-lags，b-sliding window

3.4.2 Different input variables

Based on the derived appropriate pattern, our experimental schemes for predicting streamflow and water quality factors consist of two parts. The first part focuses on streamflow and involves the testing of the performance of meteorological factors, specifically air temperature and precipitation, as the input data. We varied the window sizes to assess their impact on the predictions. The second part considers the separate and combined effects of three types of factors: meteorological factors, streamflow, and other influential water quality parameters. Meteorological factors, including water temperature and precipitation, have a significant impact on both streamflow and WQs (Hu et al., 2020; Choi et al., 2021). Streamflow itself also strongly influences the concentrations of WQs, and certain WQs can also affect each other (Yousefi et al., 2018).

In order to obtain accurate measurements in the laboratory, in our study, it is crucial to avoid interference from other WQs. And in our study, we focused on three key parameters: COD_{Mn}, TN, and TP. We considered the effects of pH and DO on COD_{Mn} (Meng et al., 2015; Wang et al., 2022), pH on TP (Li et al., 2013), as well as turbidity and NH_3-N on TN (Huang et al., 2017). These factors and their combinations were included in the input data to predict the intricacy of the three selected WQs. To evaluate the effects of these factors and their combinations, we generated nine different schemes (A_2, A_3 and B_2-B_8) as outlined in Table 4, apart from original benchmark (A_1 and B_1).

Table 4

Experimental schemes

Target	Model	Schemes	Input data	Output data
streamflow	LSTM	A_1	S	S
		A_2	MF	S
		A_3	MF, S	S
WQs	LSTM	B_1	WQ	WQ
		B_2	MF	WQ
		B_3	S	WQ
		B_4	Other WQs	An affected WQ
		B_5	MF, S	WQ
		B_6	MF, other WQs	An affected WQ
		B_7	S, other WQs	An affected WQ
		B_8	MF, S, other WQs	An affected WQ

MF: meteorological factors, S: represents streamflow, WQ: water quality

4 Results and Analyses

4.1 Fault-tolerance capability of LSTM

Either lags or sliding windows become inaccurate as time effect occurs. Therefore, the choice of the appropriate pattern and the time steps for each prediction are crucial. The training set for streamflow data consists of daily data from 2004/1/1 to 2016/12/31, while the test set covered daily data from 2017/1/1 to 2019/12/31 (Fig. 2). Regarding the WQs datasets, with limited data from 2018/11/1 to 2019/12/1, the training set included data from 2018/11/1 to 2019/11/1 to capture changes throughout the year. The remaining data was used as the test set to evaluate the learning ability of the model. Stochastic gradient decent (SGD) was chosen to minimize loss function with the Adam optimizer, which had a default learning rate of 0.001 (Loshchilov and Hutter, 2018) and epochs of 100. Early stopping was implemented if differences within 20 epochs was less than 0.0001. Initially, a fixed structure of LSTM was used for eight hydrological stations and four water quality stations. Neurons of three layers were set 64, 64, and 32 respectively, while dropout proportion is both set to be 0.5, with settings of LSTM summarized in Table 2 to prevent possible overfitting. The goals were to determine the ideal values for lags and sliding windows that produce a minimum loss, and to choose the appropriate pattern, which thus required no efforts in searching hyperparameters. The ideal value of both patterns was 1 for simulating streamflow, with roughly the same high accuracy between lags and sliding windows in Luoshan and Hankou but weak accuracy at Liuzi Port and Majia Tan (Table 5a). Sliding windows yielded a better accuracy for most WQs (Table 5b). Considering the non-instantaneous changes in WQs in general, it is more reliable to incorporate more previous data. Consequently, the LSTM model with sliding windows (without lead time in this study) is thus selected as the ideal pattern for both streamflow and WQs.

Table 5a

Results of simulating streamflow with A_1 scheme and two patterns in 8 stations

Watershed	Station name	Lags			Sliding window		
		RE (%)	NSE	Ideal value	RE (%)	NSE	Ideal value
The Yangtze River	Luoshan	3.302	0.977	1	2.462	0.986	1
	Jiujiang	3.870	0.972	1	2.616	0.989	1
Han River	Hankou	3.376	0.976	1	3.551	0.979	1
	Xiantao	9.455	0.878	1	10.420	0.953	1
Ju River	Liuzi Port	23.179	0.71	1	30.319	0.800	1
Ba River	Majia Tan	49.814	0.490	1	56.631	0.303	1
JuZhang River	Herong	29.773	0.232	1	14.638	0.517	1
Dongjing River	Qianjiang	26.024	0.833	1	21.871	0.936	1

Note: The ideal value of the pattern is 1

Table 5b

Results of simulating eight WQs with A_1 scheme and two patterns in Yangsi Port

Station	WQs	lags		Sliding window	
		Ideal value	RE (%)	Ideal value	RE(%)
Yangsi Port	pH	1	0.942	1	1.249
	DO	1	2.367	1	2.464
	NH_3-N	1	14.851	3	24.392
	COD_{Mn}	1	6.933	1	7.226
	EC	3	0.502	1	0.452
	TU	1	38.447	1	44.002
	TN	1	18.789	4	19.750
	TP	1	8.510	4	8.026

4.2 Performances of experimental schemes

4.2.1 Simulation of streamflow

The same fixed structure of LSTM models was used, which had been tested previously. In accordance with schemes from A_1 to A_3, effects of input data in the Hankou, Luoshan and Jiujiang stations are illustrated in Fig.7 with the help of the ideal size of sliding windows. The predicted trends of different input schemes, calculated from LSTM, are compared with results using Delft 3D to evaluate the reliability of the LSTM model (Fig. 7). The accuracy of the streamflow simulated from the basic LSTM model (A_1 scheme) is comparable to that of the Delft 3D model, with the NSE value approaching 0.99. However, the peaks of streamflow modeled with the A_2 scheme are flattened. In addition to these high-volume stations, the low-volume stations Xiantao and Herong, monitored less than 20% and 1% respectively, of those three previous stations, were added in Table 6 to illustrate differences between high-volume and low-volume stations. Effects of A_3 scheme are also exhibited in Table 6. The data show that implementation of meteorological factors into streamflow (S+MF) slightly reduces the original accuracy, but the NSE value remained 0.9 in most stations. On the contrary, replacing streamflow entirely with meteorological factors causes a significantly loss of accuracy. The simulation capabilities with the A_2 scheme were limited, with NSE values around 0.6 for the three high-volume stations, and much worse values in Xiantao and Herong. Our results indicate that predicting streamflow with only meteorological factors is less accurate than with historical streamflow. Nevertheless, the addition of streamflow (A_3 scheme) could be considered in practice and in certain cases, which yields more reliable results.

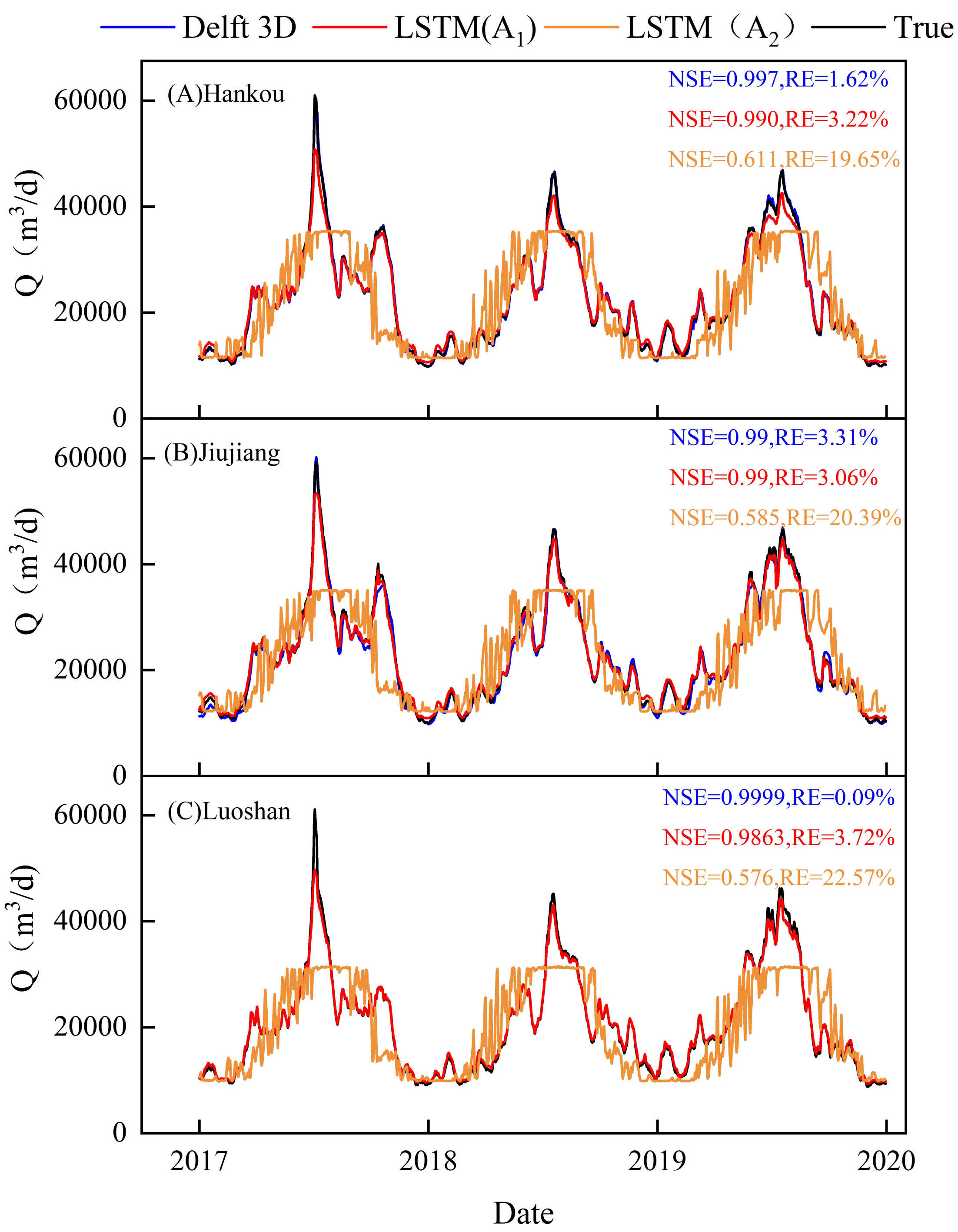

Fig. 7 Comparison of predicted streamflow between LSTM (A_1 scheme: red, A_2 scheme: brown), Delft 3D (blue) and the observed data (true).

Table 6

Comparison among three schemes of simulating streamflow at five stations

Watershed	Stations	S – S (A_1)		MF – S (A_2)		S + MF – S (A_3)	
		RE	NSE	RE	NSE	RE	NSE

The Yangtze River	Luoshan	2.46%	0.986	22.57%	0.576	5.26%	0.975
	Jiujiang	2.61%	0.988	20.39%	0.585	2.92%	0.987
Han River	Hankou	3.55%	0.979	19.65%	0.611	2.65%	0.989
	Xiantao	10.42%	0.952	44.60%	0.012	7.71%	0.951
Juzhang River	Herong	14.63%	0.516	68.42%	0.084	23.37%	0.600

4.2.2 Simulation of water quality parameters

Considering the different mechanical processes involved in various WQs, individual LSTM was established for each scheme and each WQ parameter at each station. Thus, this approach differs from the fixed LSTM model used for streamflow forecasting. Within the ranges specified in Table 2, hyperparameters for each LSTM model were randomly searched and resampled 30 times to detect the ideal combination with the minimum MSE. A five-fold cross-validation was performed, with the validation set comprising 10% of the training set. Table 7 presents the ideal combination of hyperparameters for one particular station, namely Yangsi Port in the Yangtze River. It is important to note that RandomizedSearchCV, the method used to detect the optimal parameters, may only randomly conduct combinations within repetitions. Therefore, for WQs with more complex processes, such as COD_{Mn}, TN, and TP, it is possible that the loss function could not converge stably within the limited search times. However, since the focus of our study is not to detect globally optimal values, but to explore the feasibility of alternative input data, the instability of some hyperparameters has no negative impact on the reliability of our findings.

Table 7

Ideal hyper-parameters of LSTM model for B_1 scheme at Yangsi Port

	m	Layer3	Layer2	Layer1	D2	D1
pH	6	24	8	12	0.2	0.3
DO	6	24	8	12	0.2	0.3
NH_3-N	2	44	28	24	0.3	0.1
COD_{Mn}	6	24	8	12	0.2	0.3
TU	6	24	8	12	0.2	0.3

EC	6	24	8	12	0.2	0.3
TN	2	24	36	28	0.3	0.1
TP	2	44	28	24	0.3	0.1

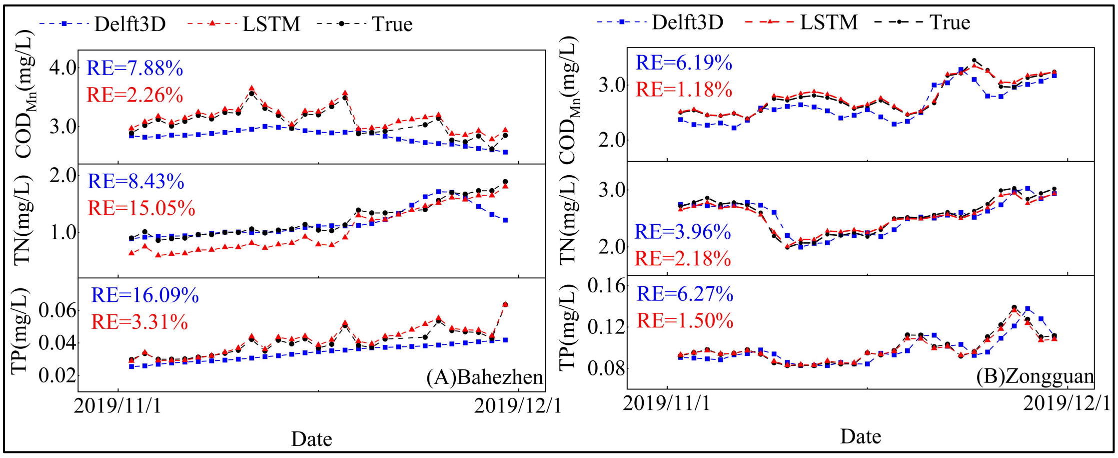

Figure. 8. Comparative performances between LSTM (B_1 scheme) and Delft 3D model at three stations.

Basic performances of LSTM, using the results of B_1 scheme are compared with that of Delft 3D model for two stations, Zongguan and Bahezhen in Fig. 8. The figure illustrates more accurate results for LSTM in capturing trends for the three main pollutants COD_{Mn}, TN and TP in the Yangtze River, thus, defining the benchmark for the remaining schemes. Fig. 9 and Table. 8 summarize the performances of B_1 scheme and the following conclusions are drawn for the basic simulation ability of LSTM: (i) NH_3-N, TN, and TP are considered as complex WQs and their accurate prediction was challenging, even with the B_1 scheme, which generally yielded the most precise predictions. (ii) Some easily measurable parameters, such as pH and DO, were accurately predicted at most stations. However, other parameters including NH_3-N, TU and TP exhibited significant fluctuation between different stations in the prediction accuracy, according to standard deviation. These findings demonstrate the complexity and variability in predicting different WQs, and constitute a major benchmark for comparison with the following results from LSTM models, accompanied by Delft 3D models.

Table 8

RE of all WQs implementing B_1 scheme with ideal hyper-parameters at the studied stations

Stations	pH	DO	NH_3-N	COD_{Mn}	TU	E.C	TN	TP
Yangsi Port	0.91%	2.92%	16.87%	10.27%	42.22%	1.61%	9.47%	9.59%
Zongguan	0.41%	2.00%	10.73%	3.91%	7.44%	0.95%	4.35%	6.76%
Guoyu	0.72%	3.21%	21.74%	7.40%	35.97%	4.39%	3.24%	27.06%
Bahezhen	2.49%	3.73%	35.44%	5.47%	24.56%	1.84%	8.75%	10.51%
Standard deviation	0.009	0.007	**0.105**	0.027	**0.152**	0.015	0.031	**0.091**

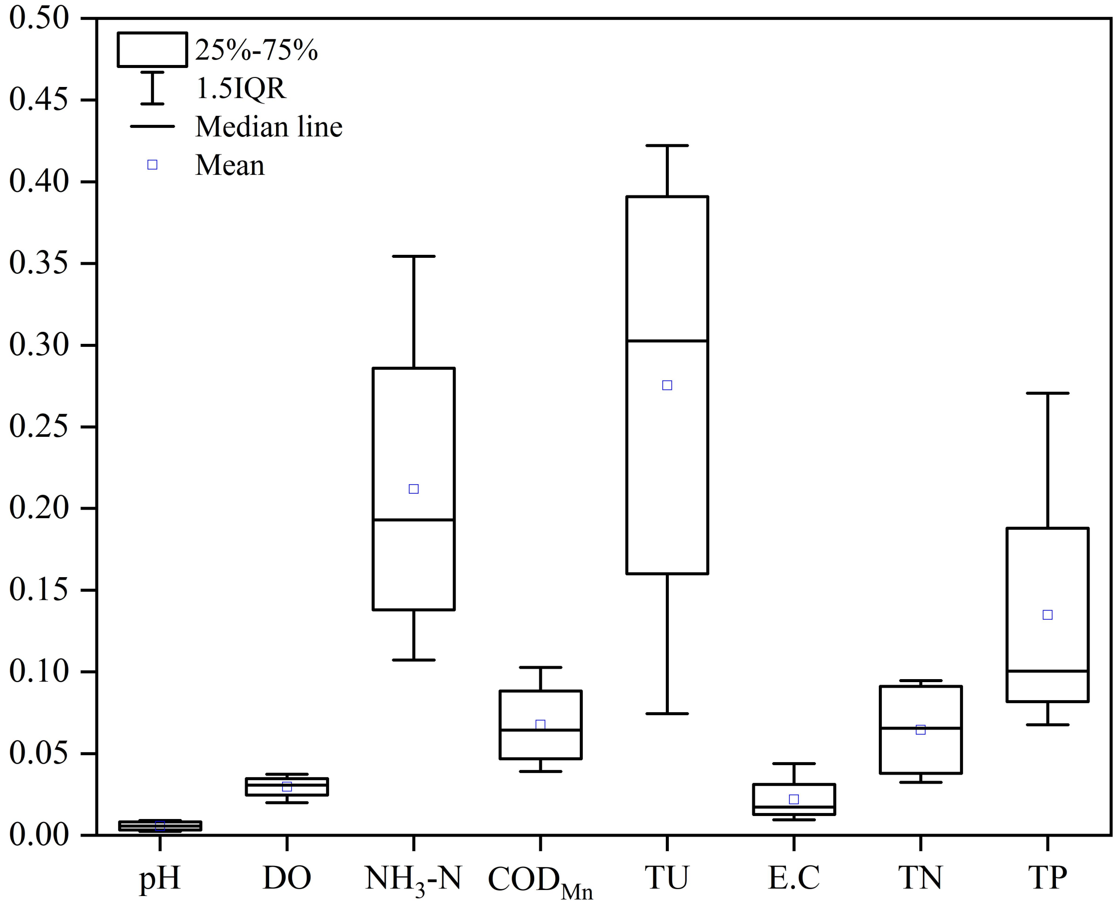

Figure. 9 Boxplot of all studied WQs among stations

Based on the defined benchmarks, all schemes (B_2-B_8) were tested with double selection, where the hyperparameters of each model were selected first, followed by the selection of the ideal scheme for each station. Table 9 summarizes the three main pollutants in the study area, with the ideal scheme and the minimum RE marked bold.

The results document that among all the ideal schemes, B_3 (using only streamflow as input data) accounted for the largest proportion at 50%. Additionally, schemes that included streamflow (B_3, B_5, B_7, B_8) made up to 75% of the selected schemes, indicating the significant influence of streamflow in predicting WQs. Moreover, the data document that streamflow played a key role in forecasting COD_{Mn}, TN, and TP, with proportions > 50% for each WQs. Concerning the accuracy, errors in the predicted COD_{Mn}, TN and TP generally increase less than 10%, which is practical for usage. Fig. 10 compares results of LSTM and Delft 3D model for the Zongguan and Bahezhen stations. Despite of the LSTM model's weaker ability in capturing trends, it provides acceptable precision, with median RE for the three WQs below 17%. The comparison indicates acceptable errors. Considering the costs of measurement and time consumption, our results confirm that historical WQ are not required anymore as input data in the LSTM model, and should be replaced by some influential factors with acceptable precision, albeit the captured trends tend to be flattened. In addition, the ideal scheme varies at each case due to the heterogeneities.

Table 9

Ideal schemes and prediction accuracies at four selected stations (minimum RE are marked bold for Zongguan and Bahezhen stations)

Stations	COD_{Mn}	TN	TP
Yangsi Port	B3(26.73%)	B2(24.42%)	B3(11.41%)
Zongguan	**B3(11.53%)**	**B3(12.44%)**	**B6(13.75%)**
Guoyu	B3(22.47%)	B3(9.69%)	B4(17.60%)
Bahezhen	**B5(6.69%)**	**B5(12.84%)**	**B7(16.43%)**
Average	16.85%	14.85%	14.79%

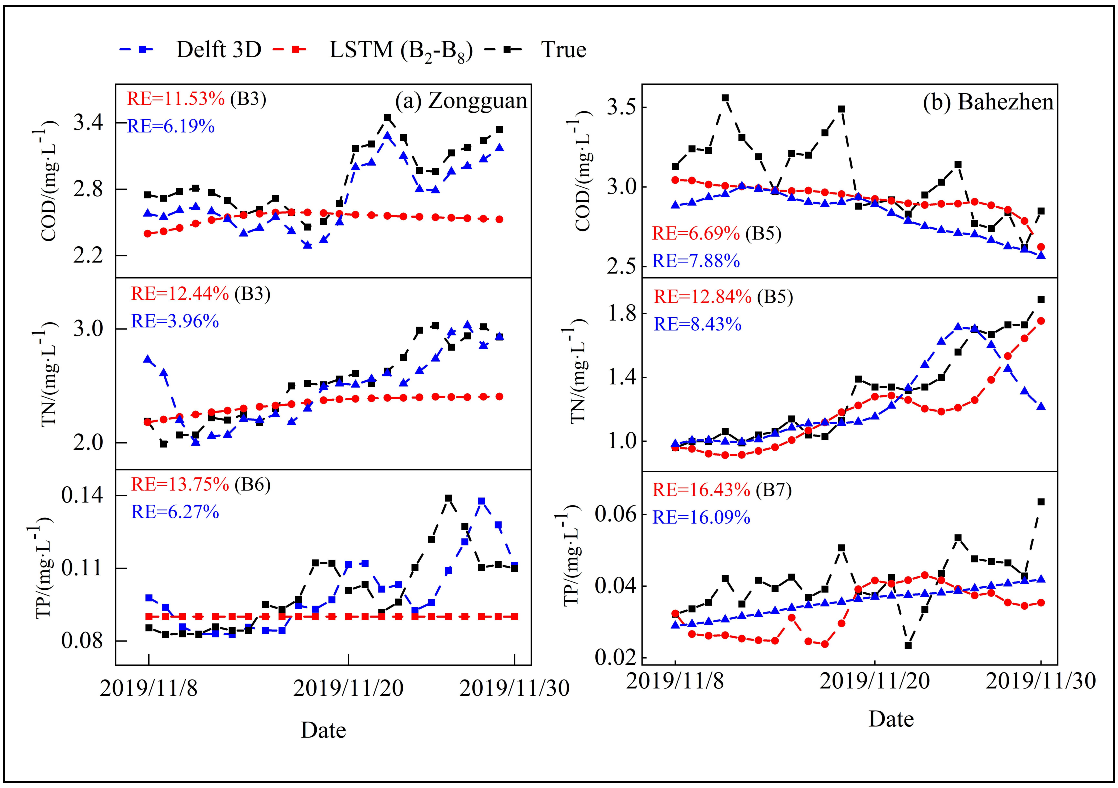

Figure. 10 Comparison between the performance of LSTM (the ideal scheme, red) and Delft 3D (blue) for three WQs at the Zongguan and Bahezhen stations.

5 Discussions

5.1 Parameters controlling the simulation of streamflow

Apart from results of streamflow (Table 6 & Fig. 7), we also detected that for meteorological factors as the only input data, despite of partly unsatisfying precision, the RE and NSE increase as window size rise from 1, 2, 3 to 7. Several mechanism models have constrained equations between meteorological factors and streamflow: For example, the Weather Research and Forecasting hydrological modeling system (WRF-Hydro model) was applied for a lake basin with NSE of 0.93, highly close to the observed streamflow (Cho and Kim, 2022). Such successful simulation just raised the question that whether the precision of deep learning model could be more reliable, for example, with NSE > 0.8, as more previous data be provided? We conducted further studies, with the size of sliding windows empirically rising from 10, 20 to 30 (Fig. 11). Results show no obvious increase of the accuracy with expanding window size. At stations of ample streamflow, like Jiujiang and Luoshan, and the station of minor volume, namely Xiantao, accuracies attained to the zenith at the size of 20, and decrease after that, while precisions of the other two stations show a valley-type variation. Thus, the positive effect of longer sliding windows is limited in improving the simulation of streamflow, consistent with findings of a previous research (Gao et al., 2020). However, since NSE and R^2 were almost in the identical format of equations in simulation, a R^2 around 0.6 was concluded to be effective in some cases (Bai et al., 2021; Moriasi et al., 2007; Yokoo et al., 2022), which is contrary to our study, due to differing accuracy requirements. In addition to such hyper-parameters, input data has a major impact on the performance of the proposed scheme at different stations. The LSTM tends towards underestimation and poor capability in case of low water volumes (Cho and Kim, 2022), which explains the weak performances at the Herong station situated at a tributary, compare to stations like Hankou and Luoshan located at the main stream of the Yangtze River. Although LSTM models with historical streamflow (A_1&A_3) performed as well as Delft 3D models, they are less reliable as mechanism models, when antecedent data could not be involved in. However, since deep learning models require no boundary conditions, and is a well-known and promising method, the LSTM is appropriate to further study the potential of streamflow forecast.

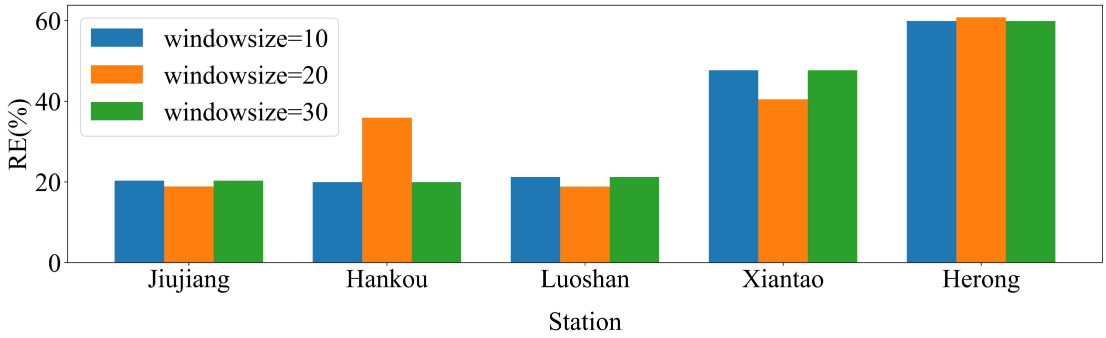

Figure. 11 Accuracy of LSTM (A_2 scheme) with rising size of sliding windows, from 10, 20 to 30 in five typical water quality stations.

5.2 Analyses on simulating water quality parameters

We have simulated eight WQs with B_1 scheme, and three WQs from B_2 to B_8 schemes in the study area. The former showed reliable results in predicting pH, DO, COD_{Mn}, and E.C, with RE < 20%, but less reliable performances in NH_3-N, TU and TP forecasts. The latter results of COD_{Mn}, TN and TP are partly consistent with previous studies that reported better predictions in COD_{Mn} and TP in the Haihe river basin in Northen China (Li et al., 2023). By comparing average RE at all stations among B_2-B_8 schemes, we further analyzed the average effects of different schemes among stations (Table 10). The data document a less accurate performance of the three WQs for schemes including streamflow, such as B_5, B_7, and B_8, compared to B_1 schemes. However, the RE of 19.62%, 32.21% and 27.99% with B_3 scheme are still acceptable, compared with the other two schemes (B_2 and B_4), thus corroborating streamflow as the most dominant factor among the three input factors, consistent with findings of the previous research (Patil et al., 2022). Adding historical data has a major impact: After recalculation of this new scheme, comparisons were made between the ideal scheme (bold data in Table 10) and the scheme containing historical WQ and all three factors. Results in Table 11 document that among these typical WQs, holding historical data dramatically declines the RE value by 56.68%, 62.53%, and 48.80% for the three WQs, and thus excels those selected scheme with the ideal performances. Compared with the Delft 3D model that requires many pre-requisites and generally causes larger errors than basic LSTM models (B_1 scheme), altering input data with more accessible variables in LSTM is not only more cost-effective, but also produces acceptable results for practical usage.

Apart from accuracy achievements, our results show some deficiencies. Performances of LSTM on NH_3-N and EC with B_1 scheme are weak (Table 8), compared with other WQs. The reason remains uncertain, despite those dynamic structures had been provided. In terms of hyperparameters, while constructing LSTM models, the optimal time step of the appropriate pattern is different in various cases. Through comparison, window sliding is chosen in our study, and is taking different optimal time steps for each WQs at each station. However, the specific pattern and value of timesteps should be settled down according to each case in further study.

Table 10

Average RE of all scheme at all stations

	COD_{Mn}	TN	TP
B1	8.43%	9.28%	12.02%
B2	29.21%	45.74%	35.21%
B3	19.62%	32.21%	27.99%
B4	33.35%	35.57%	26.91%
B5	30.81%	36.58%	25.70%
B6	29.53%	31.21%	27.07%
B7	25.98%	32.40%	26.08%
B8	26.71%	34.71%	23.87%

Table 11 Comparisons between the minimum average RE of the ideal scheme (except B_1) and the average RE of the scheme, containing all factors, at all stations

	COD_{Mn}	TN	TP
The ideal scheme	19.62%	32.21%	27.99%

Histrocial WQs & all factors	8.50%	12.07%	14.33%

6 Conclusion

We studied the capability of LSTM model with the appropriate pattern and input schemes for the prediction of streamflow and water quality in the middle reach of the Yangtze River, in China. The comparison with results from a mechanism model (Delft 3D) allows to evaluate the efficacy and reliability of our models and to draw conclusions as follows.

(1) Sliding windows is the more appropriate pattern than lags as the pattern of LSTM in simulating both streamflow and water quality parameters.

(2) Only using meteorological factors as input data reduces the performance of forecasting streamflow, even with the longer sliding windows, with NSE of ca. 0.6 in main stream and much worse values in tributaries of lower water volumes. Adding historical streamflow into input data slightly improves the performance, which could achieve similar accuracy as the Delft 3D models with NSE reaching 0.9.

(3) Only implementing meteorological factors, streamflow and other influential WQs as input data could achieve lower ability for capturing trends but acceptable precision of median RE below 17% in the prediction of COD_{Mn}, TN and TP. Streamflow is detected as the most dominant factor. Adding historical WQs into the input data increases accuracy by 48.8% at least among all proposed alternative schemes.

Author contribution

Yucong Hu: Conceptualization, Methodology, Software, Formal analysis, Writing -

original draft.

Yan Jiang: Writing - review & editing, Supervision, Funding acquisition.

Huiting Yao: Software, Data curation, Writing - original draft.

Yiping Chen: Visualization.

Xuefeng Wu: Data Curation, Visualization.

Xuyong Li: Writing - review & editing.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgement

The research was supported by the National Science Foundation of China (2019YFB2102902). The authors would like to express their gratitude to EditSprings (https://www.editsprings.cn) for the expert linguistic services provided.

Reference

Alizadeh, B. et al., 2021. A novel attention-based LSTM cell post-processor coupled with bayesian optimization for streamflow prediction. Journal of Hydrology, 601. DOI:10.1016/j.jhydrol.2021.126526

Bai, J., Zhao, J., Zhang, Z., Tian, Z., 2022. Assessment and a review of research on surface water quality modeling. Ecological Modelling, 466: 109888.

Bai, P., Liu, X., Xie, J., 2021. Simulating runoff under changing climatic conditions: A comparison of the long short-term memory network with two conceptual hydrologic models. Journal of Hydrology, 592. DOI:10.1016/j.jhydrol.2020.125779

Barzegar, R., Aalami, M.T., Adamowski, J., 2020. Short-term water quality variable prediction using a hybrid CNN–LSTM deep learning model. Stochastic Environmental Research and Risk Assessment, 34(2): 415-433. DOI:10.1007/s00477-020-01776-2

Bergstra, J., Bengio, Y., 2012. Random search for hyper-parameter optimization. Journal of machine learning research, 13(2).

Buitinck, L. et al., 2013. API design for machine learning software: experiences from the scikit-learn project. arXiv preprint arXiv:1309.0238.

Camps-Valls, G., Tuia, D., Zhu, X.X., Reichstein, M., 2021. Deep learning for the Earth Sciences: A comprehensive approach to remote sensing, climate science and geosciences. John Wiley & Sons, 114 pp.

Chen, C. et al., 2022. Forecast of rainfall distribution based on fixed sliding window long short-term memory. Engineering Applications of Computational Fluid Mechanics, 16(1): 248-261. DOI:10.1080/19942060.2021.2009374

Chen, Y., Cheng, Q., Cheng, Y., Yang, H., Yu, H., 2018. Applications of recurrent neural networks in environmental factor forecasting: a review. Neural computation, 30(11): 2855-2881.

Cho, K., Kim, Y., 2022. Improving streamflow prediction in the WRF-Hydro model with LSTM networks. Journal of Hydrology, 605. DOI:10.1016/j.jhydrol.2021.127297

Chong, L., Zhong, J., Sun, Z., Hu, C., 2023. Temporal variations and trends prediction of water quality during 2010–2019 in the middle Yangtze River, China. Environmental Science and Pollution Research, 30(11): 28745-28758.

Dadson, S.J., Hirpa, F., Thomson, P., Konar, M., 2019. Monitoring and modelling hydrological processes. Water Science, Policy, and Management: A Global Challenge: 117-137.

Gao, S. et al., 2020. Short-term runoff prediction with GRU and LSTM networks without requiring time step optimization during sample generation. Journal of Hydrology, 589: 125188.

Gers, F.A., Schmidhuber, J., Cummins, F., 2000. Learning to forget: Continual prediction with LSTM. Neural computation, 12(10): 2451-2471.

Ghimire, S. et al., 2021. Streamflow prediction using an integrated methodology based on convolutional neural network and long short-term memory networks. Sci Rep, 11(1): 17497. DOI:10.1038/s41598-021-96751-4

Hochreiter, S., Schmidhuber, J., 1997. Long short-term memory. Neural computation, 9(8): 1735-1780.

Hu, Y., Yan, L., Hang, T., Feng, J., 2020. Stream-flow forecasting of small rivers based on LSTM. arXiv preprint arXiv:2001.05681.

Huang, W., zhou, J., Zhang, Q., 2017. Influencing factors and improvement of the methods for determination of total nitrogen in water. Environmental Monitoring and Forewarning, 9(6): 45-47.

Ighalo, J.O., Adeniyi, A.G., Marques, G., 2021. Artificial intelligence for surface water quality monitoring and assessment: a systematic literature analysis. Modeling Earth Systems and Environment, 7(2): 669-681.

Jaffar, A., M. Thamrin, N., Megat Ali, M.S.A., Misnan, M.F., Mohd Yassin, A.I., 2022. Water Quality Prediction Using Lstm-Rnn: A Review. Journal of Sustainability

Science and Management, 17(7): 204-225. DOI:10.46754/jssm.2022.07.015

Jia, X. et al., 2019. Physics guided RNNs for modeling dynamical systems: A case study in simulating lake temperature profiles, Proceedings of the 2019 SIAM international conference on data mining. SIAM, pp. 558-566.

Kratzert, F., Klotz, D., Brenner, C., Schulz, K., Herrnegger, M., 2018. Rainfall–runoff modelling using Long Short-Term Memory (LSTM) networks. Hydrology and Earth System Sciences, 22(11): 6005-6022. DOI:10.5194/hess-22-6005-2018

Lees, T., 2022. Deep learning for hydrological modelling: from benchmarking to concept formation, University of Oxford.

Li, H., Liu, L., Li, M., Zhang, X., 2013. Effects of pH, temperature, dissolved oxygen, and flow rate on phosphorus release processes at the sediment and water interface in storm sewer. Journal of analytical methods in chemistry, 2013.

Li, L. et al., 2020. Global trends in water and sediment fluxes of the world's large rivers. Science Bulletin, 65(1): 62-69.

Li, Q., Yang, Y., Yang, L., Wang, Y., 2023. Comparative analysis of water quality prediction performance based on LSTM in the Haihe River Basin, China. Environ Sci Pollut Res Int, 30(3): 7498-7509. DOI:10.1007/s11356-022-22758-7

Liang, Z. et al., 2020. Simulate the forecast capacity of a complicated water quality model using the long short-term memory approach. Journal of Hydrology, 581. DOI:10.1016/j.jhydrol.2019.124432

Lipton, Z.C., Berkowitz, J., Elkan, C., 2015. A critical review of recurrent neural networks for sequence learning. arXiv preprint arXiv:1506.00019.

Loshchilov, I., Hutter, F., 2018. Fixing weight decay regularization in adam.

Meng, F., Li, H., Su, F., Wang, T., 2015. Analyses of Diversion Water Input's Influence on Water Quality of Dahuofang Reservoir. Nature Environment and Pollution Technology, 14(4): 989.

Moriasi, D.N. et al., 2007. Model evaluation guidelines for systematic quantification of accuracy in watershed simulations. Transactions of the ASABE, 50(3): 885-900.

Muzaffar, S., Afshari, A., 2019. Short-term load forecasts using LSTM networks. Energy Procedia, 158: 2922-2927.

Najah, A., El-Shafie, A., Karim, O.A., El-Shafie, A.H., 2013. Application of artificial neural networks for water quality prediction. Neural Computing and Applications, 22: 187-201.

Najah Ahmed, A. et al., 2019. Machine learning methods for better water quality prediction. Journal of Hydrology, 578. DOI:10.1016/j.jhydrol.2019.124084

Pascanu, R., Mikolov, T., Bengio, Y., 2013. On the difficulty of training recurrent neural networks, International conference on machine learning. Pmlr, pp. 1310-1318.

Patil, R., Wei, Y., Pullar, D., Shulmeister, J., 2022. Effects of change in streamflow patterns on water quality. Journal of Environmental Management, 302: 113991.

Peng, A., Zhang, X., Xu, W., Tian, Y., 2022. Effects of training data on the learning performance of LSTM network for runoff simulation. Water Resources Management, 36(7): 2381-2394.

Prechelt, L., 1998. Early Stopping - But When? In: Orr, G.B., Müller, K.-R. (Eds.), Neural Networks: Tricks of the Trade. Springer Berlin Heidelberg, Berlin, Heidelberg, pp. 55-69. DOI:10.1007/3-540-49430-8_3

Rasouli, K., Hsieh, W.W., Cannon, A.J., 2012. Daily streamflow forecasting by machine learning methods with weather and climate inputs. Journal of Hydrology, 414: 284-293.

Roelvink, J., Van Banning, G., 1995. Design and development of DELFT3D and application to coastal morphodynamics. Oceanographic Literature Review, 11(42): 925.

Shen, C., 2018. A Transdisciplinary Review of Deep Learning Research and Its Relevance for Water Resources Scientists. Water Resources Research, 54(11): 8558-8593. DOI:10.1029/2018wr022643

Solgi, R., Loáiciga, H.A., Kram, M., 2021. Long short-term memory neural network (LSTM-NN) for aquifer level time series forecasting using in-situ piezometric observations. Journal of Hydrology, 601. DOI:10.1016/j.jhydrol.2021.126800

Srivastava, N., Hinton, G., Krizhevsky, A., Sutskever, I., Salakhutdinov, R., 2014. Dropout: a simple way to prevent neural networks from overfitting. The journal of machine learning research, 15(1): 1929-1958.

Stidson, R.T., Gray, C.A., McPhail, C.D., 2012. Development and use of modelling techniques for real-time bathing water quality predictions. Water and Environment Journal, 26(1): 7-18. DOI:10.1111/j.1747-6593.2011.00258.x

Tongal, H., Booij, M.J., 2018. Simulation and forecasting of streamflows using machine learning models coupled with base flow separation. Journal of Hydrology, 564: 266-282. DOI:10.1016/j.jhydrol.2018.07.004

Valipour, M., Banihabib, M.E., Behbahani, S.M.R., 2013. Comparison of the ARMA, ARIMA, and the autoregressive artificial neural network models in forecasting the monthly inflow of Dez dam reservoir. Journal of hydrology, 476: 433-441.

Wallach, D., Goffinet, B., 1989. Mean squared error of prediction as a criterion for evaluating and comparing system models. Ecological modelling, 44(3-4): 299-306.

Wan, H. et al., 2022. A novel model for water quality prediction caused by non-point sources pollution based on deep learning and feature extraction methods. Journal of Hydrology, 612. DOI:10.1016/j.jhydrol.2022.128081

Wang, L. et al., 2022. Nutrients and Environmental Factors Cross Wavelet Analysis of River Yi in East China: A Multi-Scale Approach. International Journal of Environmental Research and Public Health, 20(1): 496.

Wang, Q. et al., 2011. Water quality model establishment for middle and lower reaches of Hanshui River, China. Chinese Geographical Science, 21(6): 646-655. DOI:10.1007/s11769-011-0504-y

Wang, Y., Zhou, J., Chen, K., Wang, Y., Liu, L., 2017. Water quality prediction method based on LSTM neural network, 2017 12th international conference on intelligent systems and knowledge engineering (ISKE). IEEE, pp. 1-5.

Xiang, Z., Yan, J., Demir, I., 2020. A Rainfall‐Runoff Model With LSTM‐Based Sequence‐to‐Sequence Learning. Water Resources Research, 56(1). DOI:10.1029/2019wr025326

Yaseen, Z.M., Kisi, O., Demir, V., 2016. Enhancing Long-Term Streamflow Forecasting and Predicting using Periodicity Data Component: Application of Artificial Intelligence. Water Resources Management, 30(12): 4125-4151. DOI:10.1007/s11269-016-1408-5

Yokoo, K. et al., 2022. Capabilities of deep learning models on learning physical relationships: Case of rainfall-runoff modeling with LSTM. Sci Total Environ, 802: 149876. DOI:10.1016/j.scitotenv.2021.149876

Zhang, Q., Zhang, J., Zou, J., Fan, S., 2020. A novel fault diagnosis method based on stacked lstm. IFAC-PapersOnLine, 53(2): 790-795.

Zhang, Y. et al., 2022. Accurate prediction of water quality in urban drainage network with integrated EMD-LSTM model. Journal of Cleaner Production, 354. DOI:10.1016/j.jclepro.2022.131724

Zhu, M. et al., 2022. A review of the application of machine learning in water quality evaluation. Eco-Environment & Health, 1(2): 107-116. DOI:10.1016/j.eehl.2022.06.001