ISEIS

Journal of Environmental Informatics 43(2) 141-154 (2024)

Journal of Environmental Informatics

www.iseis.org/jei

Water Quality Prediction Based on an Innovated Physical and Data Driving Hybrid Model at Basin Scale

Y. L. A[1], G. Q. Wang[1, 2 *], Q. Z. Zhang[2], P. Z. Wang[3], B. L. Xue[1], Z. Y. Gao[2], and Y. B. Peng[2]

[1] *Faculty of Geographical Science, Beijing Normal University, Beijing 100875, China*
[2] *Environment Research Institute, Shandong University, Qingdao 266237, China*
[3] *China Water Resources Beifang Investigation, Design and Research Co. Ltd., Tianjin 300222, China*

Received 20 June 2022; revised 15 June 2023; accepted 27 June 2023; published online 05 January 2024

ABSTRACT. The prediction of basin water quality has become an urgent need for water environment management, where water pollution is on the increase. Currently, physical models are primarily used for water quality predictions, but the models are not adaptable for the automatic future prediction of watershed water quality owing to their non-automatic boundary setting. The development of big data, which has led to artificial intelligence (AI) technology, has remedied the deficiency of physical models and has been widely used in water quality prediction. However, the accuracy of AI models depends only on the quantity and quality of dataset, which is applied on specific and discrete sections with enough data and difficult to extend to regions with limited monitoring data. Thus, we constructed migration and distribution gates to express the spatial influence of different variables from different sections on the water quality of a specific and discrete section. The temporal processes were expressed by degradation equation. The migration gate, distribution gate and degradation equations were incorporated into Long Short-Term Memory Network (LSTM) to improve the operation mechanism of the LSTM algorithm to create the Im-LSTM model, which considers both the temporal influence of a specific section and the spatial influence of other sections on a specific section at basin scale. Compared to ANN, LSTM, Im-LSTM showed the best performance for basin water quality prediction, especially for mainstream sections at sudden pollution process. Thus, the proposed Im-LSTM provides a new approach for water environment supervision.

Keywords: long term, water quality prediction, physical equations, long short-term memory network, hybrid model

1. Introduction

Water pollution management was become a challenge with the increase occurrence of sudden and hidden water pollution events (Xu et al., 2021), accurate prediction of water quality based on different models is the foundation for effective pollution control (Wang et al., 2019b). Transport and transformation processes of pollutants in rivers are complex, which influenced by various factors, such as temperature, precipitation, runoff, physical chemistry, and land use (Elhatip and Komur 2008; Chang et al. 2016; Wang et al., 2019a). Presently, the main methods for water quality prediction are classified into physical models and non-physical models according to the theoretical basis of models (Yeon et al., 2008). Physical models, for example, the commonly used water quality prediction models are the Soil and Water Assessment Tool (SWAT) and Storm Water Management Model (SWMM), are based on the motion equation, continuity equation and energy equation in fluid mechanics to simulate the diffusion and migration law of water pollutants with space and time under the action of physics, chemistry and biology processes (Zhang et al., 2022). Obviously, the built model and calibrated parameters based on historical data cannot be applied to future unconventional water environment management (Xue et al., 2020). The construction of the model depends on the previous historical data, real-time monitoring data cannot help to optimize model by updating parameters, thus the future-oriented prediction error will expand because the parameter update is not timely with variable condition (Elhatip and Komur, 2008; Hong, 2012; Su et al, 2019), which is not suitable for automatic monitoring and prediction in basin scale. In addition, the future boundary conditions were set by history or manual, which is not suitable for future automatic perdition.

The development of big data, which has led to artificial intelligence (AI) technology, is a remedy for the deficiencies in water quality models (Palani et al., 2008; Yeon et al., 2008; Wang et al., 2017; Hu et al., 2019). Since 2008, artificial intelligence technology has been widely used in the field of water quality prediction (Yeon et al., 2008). Biochemical oxygen demand (BOD) predictions resulting from the use of artificial neural networks (ANN)are faster and more accurate than those from other conventional methods. Results have shown that chemical oxygen demand (COD) is the most effective variable for estimating biochemical oxygen demand through sensitivity

* Corresponding author: Tel.: 0086-01058807093.
E-mail address: wanggq@bnu.edu.cn (G. Q. Wang).

ISSN: 1726-2135 print/1684-8799 online
© 2024 ISEIS All rights reserved. doi:10.3808/jei.202400510.

analysis (Dogan et al., 2008; Dogan et al., 2009). Multilayer perceptron (MLP) and generalized regression neural network (GRNN) models, in which different inputs are selected through step-by-step construction methods, can predict suspended solids (SS) and dissolved oxygen (DO) via multistep prediction (Areerachakul et al., 2013). For time-series problems, LSTM (Long Short-Term Memory) has the best prediction ability in many studies owing to its unique gate structure, which enables the capture and transfer of long-term dependencies over time (Piotrowski et al., 2015; Chen et al., 2021; Prasad et al., 2022; Zhang et al., 2022). Even though AI methods showed high accuracy on different predict problems, the accuracy only depends on quantity and quality of the input dataset (Xie et al., 2021; Xie and Yao, 2022), which is suitable for specific and discrete sections with enough data and difficult to extend to regions with limited monitoring data (Palani et al., 2008). Since these models have no assumptions of the processes underlying the data, known laws or theory are ignored and this omission can lead to spurious and inaccurate predictions (Nayak et al., 2013; Lazer et al., 2014). Therefore, the coupling of physical mechanisms with artificial intelligence algorithms has been an increasingly interesting hotspot (Xie et al., 2021). The most common coupling method is the output of one model is used as the input of the other model, called loose coupling. For example, to improve physical model accuracy, some complex and dynamic parameters of the physical models were extracted based on AI model and then integrated into the physical model during simulation process (Yan et al., 2019; Koppa et al., 2022). Koppa et al. (2022) built a deep learning algorithm to obtain a variable parameter suitable for different regional scales, which then replaced the static parameters of the physical model for global ET prediction. In contrast, to deal with insufficient training data for AI model in limited data region, the output of a physical model is the input data for AI training (Xie et al., 2021; A et al., 2022). Even though loose coupling can deal with some drawbacks of physical and AI model that can improve trust in outcomes, their foundation drawbacks within themselves have no solved. According to the new modeling paradigm — "Theory-Guided Data Science" (Read et al., 2019; Liu et al., 2022), water temperature was predicted by integrating the energy balance equation into an AI model as a loss function (Read et al., 2019). However, the loss function focuses on the entire control of total prediction by constraining the final outputs. Few studies have focused on the physical constraints on AI processes; they include many nerve cells, layers, and connections, and appear black to researchers (Siljic et al., 2015; Chen et al., 2022), especially in the field of water quality prediction.

The water quality of a specific section is determined by different processes, such as the spatial transformation process among different sections and temporal degradation by oxidation. Thus, this study focused on the physical expression of spatial transformation and temporal degradation, and the expression was integrated into an AI model. Based on the near-real-time observation data of different sections, this study combines the mechanism equation and artificial intelligence algorithm to mine the resolution law of different pollutants in a specific section over time and the migration law between different sections over space driven by the outside world, and realizes the dynamic construction and prediction of the model, which can be used in long-term prediction practice at basin scale. Our objectives were to 1) construct physical equations to express spatial and temporal influences on a specific section, 2) incorporate the physical equations into a suitable AI model, and 3) train and evaluate the efficiency of the Hybrid Model for basin water quality prediction at different sections.

2. Model Data and Model Framework

2.1. Study Area

The Xiaoqing River is an important man-made excavated river in Shandong Province. Its main function is to provide flood control and waterlogging, supplemented by irrigation, navigation, and ecological landscapes. Starting from Mulizhuang, the Xiaoqing River flows eastward through Zhangqiu and Shouguang in the Licheng District of Jinan city to Yangjiaogou, where it reaches the Bohai Sea (as shown in Figure 1). The total length of the Xiaoqing River is 233 km, the drainage area is 10,336 km^2, the average channel slope is 0.15‰, and the river network density is 0.27 km/km^2. The upper reaches stretch from Mulizhuang to Huangtaiqiao, with a length of 22.2 km and a slope of 0.45‰. The middle reaches of the river stretch from Huangtaiqiao to Shichun, with a slope of approximately 0.14‰. From Shichun to its mouth, the river extends 77.5 km in length, with a slope ratio of approximately 0.064‰. The water replenishment of the Xiaoqing River Basin mainly depends on natural precipitation. In the basin, average surface water resources are approximately 1.33×10^8 m^3, average groundwater resources are 1.57×10^8 m^3, and average total water resources are 2.42×10^8 m^3 (excluding the amount of repeated calculation); the approximate water production modulus is 170,000 m^3/km^2. The cities through which the Xiaoqing River flows are important economic centers in Shandong Province and contain many industrial point sources and agricultural nonpoint sources of pollution in the basin.

2.2. Hybrid Model Framework

2.2.1 Basic Model Frame of LSTM

LSTM is based on RNN (Recurrent Neural Networks) innovation and development. The RNN infrastructure (Figure S1) is composed of the input layer, hidden layer and output layer; the hidden layer consists of the continuous temporal characteristics of the information contained in the learning process. The specific calculation equation is as follows:

$$o_t = g(Vs_t) \tag{1}$$

$$s_t = f(Ux_t + Ws_{t-1}) \tag{2}$$

x is the network input layer of the input data (vector), s is the output of hidden layer calculated data (vector), o is the output layer to obtain the final data (vector), U is the matching weight matrix of input data x, V is the matching weight matrix

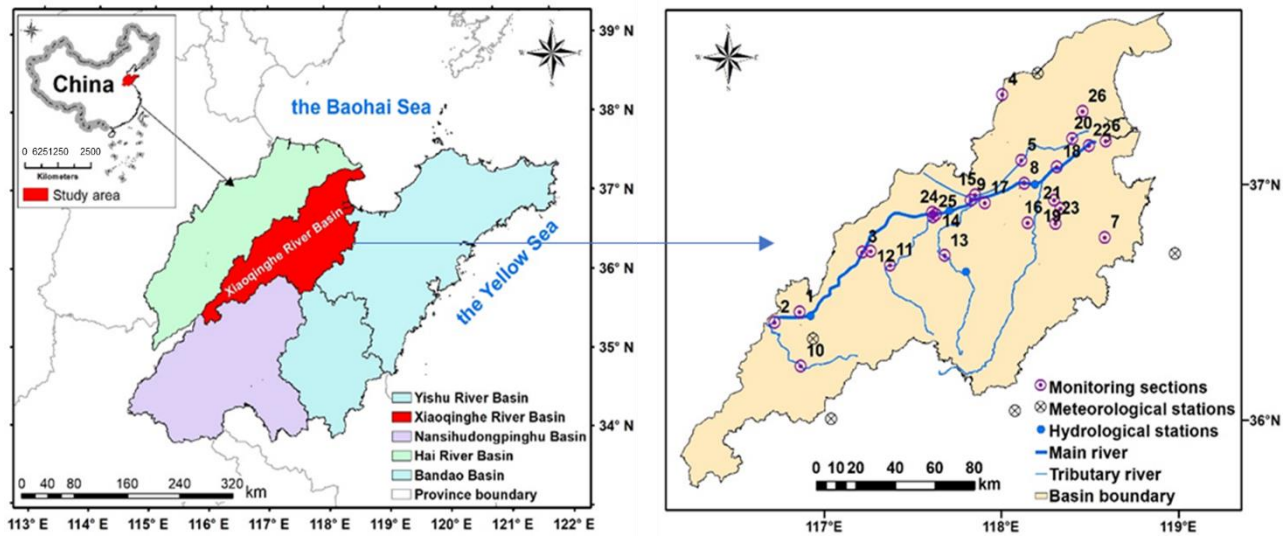

Figure 1. Location of Xiaoqing River basin and its monitoring stations.

of hidden layer *s*, and *W* is the matching weight matrix in the cycle. *g* () and *f* () are the activation function, *t* and *t* – 1 mean the time of current state and previous state.

The frame of LSTM was shown in Figure S2. the input value of LSTM has three values at time *t*: the input value of the current time is x_t; the output value of the previous moment is h_{t-1}; and the unit state of the previous moment is c_{t-1}. The LSTM output has two values: the current moment output value h_t and the current moment unit state c_t. LSTM realizes the continuous transmission of timing information features through the above variables and proposes a forget gate, input gate and output gate. The concept of the gate method (see Figure S3) is that the forget gate determines how much information of the united state c_{t-1} at the last moment is retained to the current moment c_t; the input gate determines how much information from the current input x_t is saved to the cell state c_t; and the output gate control unit state c_t determines how much information is output to the current output value h_t (Wang, et al., 2019b). The equations of LSTM's forget gate, input gate and output gate and other core algorithms are as follows:

$$f_t = \sigma(W_f \cdot [h_{t-1}, x_t] + b_f) \tag{3}$$

$$[W_f]\begin{bmatrix} h_{t-1} \\ x_t \end{bmatrix} = [W_{fh} \quad W_{fx}]\begin{bmatrix} h_{t-1} \\ x_t \end{bmatrix} \tag{4}$$

$$i_t = \sigma(W_i \cdot [h_{t-1}, x_t] + b_i) \tag{5}$$

$$c_t = \tanh(W_c \cdot [h_{t-1}, x_t] + b_c) \tag{6}$$

$$c_t = f_t \circ c_{t-1} + i_t \circ c_t \tag{7}$$

$$o_t = \sigma(W_o \cdot [h_{t-1}, x_t] + b_o) \tag{8}$$

$$h_t = o_t \circ \tanh(c_t) \tag{9}$$

For Equations (3) to (9), f_t is the forget gate; i_t is the input gate; c_t is the current input unit state; c_t is the unit state at the current moment; o_t is the output gate; h_t is the final output of LSTM; and W_t is the weight matrix of the forget gate, which includes W_{fh} and W_{fx}. W_{fh} is the weight matrix of the corresponding input term h_{t-1}; W_{fx} is the weight matrix of the corresponding input term X_t; $[h_{t-1}, X_t]$ is the joining of two vectors into a longer vector; b_f is the bias term of the forget gate; σ is an activation function; W_i is the weight matrix of the input gate; b_i is the bias term of the input gate; W_c is the weight matrix of the current input element state c_t; b_c is the bias term of the current input unit state c_t; W_o is the weight matrix of the output gate; b_o is the bias term of the output gate; *tanh* () is the activation function; · means elements multiplication, ◦ means matrix multiplication.

2.2.2. Traditional Description of Temporal and Spatial Change of Pollutants

In general, pollutant concentration of a specific section at the current time (*t*) was determined by its self-concentration at previous time (*t* – 1) and other drive process, such as self-degradation and migration process from other sections to calculated section, etc. (Maguire et al., 2019). Self-degradation process in time means the pollutant concentration decreases with time under water self-purification process, such as physical flocculation, chemical reaction, biological denitrification, which is represented by degradation coefficient. Assume that the concentration of pollutant *i* in the water body at time *t* is $C_i(t)$, and the degradation coefficient is *K* (0 < *K* < 1), which means pollutant concentration is reduced by *K* times per unit time. So, the change of pollutant concentration (ΔC_i) after time Δt is:

$$\Delta C_i = -K \cdot C_i(t-1) \cdot \Delta t \tag{10}$$

Then Equation (5) can be changed to:

$$\frac{\Delta C_i}{\Delta t} = -K \cdot C_i(t-1) \tag{11}$$

When $\Delta t \to 0$, the above Equation can be changed to:

$$\frac{dC_i}{dt} = -K \cdot C_i(t-1) \tag{12}$$

For Equations (10) ~ (12), $C_i(t)$ and $C_i(t-1)$ is the concentration of pollutant i at time t and $t-1$. K is the first-order degradation coefficient, which means the change rate of pollutant i. The degradation coefficient is a constant but varies with different pollutant types and water conditions, which was deeply studies in many regions. According to *Chinese Academy for Environmental Planning* published *The Technical Guidelines of National Water Environmental Capacity*, the initial value of K for COD and NH_3-N were set as 0.08 d^{-1}.

Generally, the spatial distribution of all natural elements (including pollutants) has structural and random characteristics. The structural distribution characteristics mean that the spatial aggregation or dispersion distribution characteristics of pollutants keeps relatively stable and can be extracted by all sections in the whole basin. And the pollutant concentration on the calculated section follows the whole distribution pattern at basin and is highly correlated with its around sections. In addition, pollutants concentration on calculated section is also influenced by spatial migration process with other sections. And as previous studies reported that (Flores et al., 2019; Xu et al., 2021) the closer distance between calculated section and a specific section the higher influencing degree of the specific section on calculated section. Thus, the spatial migration and distribution influences on calculated section from other sections was defined by Euclidean distance among water pollutants:

$$Eu(x', x) = \sqrt{\frac{\sum_{i=1}^{n}[C(x_i) - C(x_i')]^2}{n}} \tag{13}$$

where $Eu(x', x)$ is the Euclidean distance of water quality between the calculation section (x) and other monitoring sections (x'), $C(x_i)$ means the concentration C of pollutant i at section x, n means the number of pollutants.

2.2.3. Hybrid Model Framework

Based on the spatial migration and distribution influence of pollutants concentration, we construct a state unit that can express the spatial correlation of water quality at different sections referring to the LSTM cells Equations, named the migration gate and distribution gate. The migration gate determines how much water quality information is transferred from other monitoring section at the previous moment to the calculated monitoring section at the current moment. The distribution gate determines how much water quality information of calculated section should be reserved following the whole spatial distribution pattern, which extracted from all the other sections at current time. The specific calculation equation is as follows:

$$m_t = \max\left(\frac{1}{Eu(x', x)} W_m \cdot [X'_{t-1}, x_t] + b_m\right) \tag{14}$$

$$d_t = \sigma(W_d \cdot [X'_t, x_t] + b_d) \tag{15}$$

where m_t is the migration gate, d_t is the distribution gate, W_m is the weight matrix of the migration gate, W_d is the weight matrix of the distribution gate, b_m is the bias term of the migration gate, b_d is the bias term of the distribution gate. x_t is the input tensor of calculation section at the current time (t), X'_{t-1} is the input matrix of other monitoring sections besides calculation section at the previous moment ($t-1$). X'_t is the input matrix of other monitoring sections at the current time, $Eu(x', x)$ is the Euclidean distance of water quality between the calculation section (x) and other monitoring sections (x'). And *max*() and σ are activetion functions.

Equation (14) introduces the Euclidean distance of the spatial section into the weight optimization mechanism of the migration gate. The inverse of Euler distance ($1 / Eu(x', x)$) is used to represent the migration influencing degree of pollutants concentration from a specific section to calculation section, because the closer distance the more pollutants migration to calculation section. Through the *max*() activation function, the maximum migration influence of a specific section at previous time on calculation section is obtained.

As above described, the physical process of pollutants, such as degradation and migration, is important for water quality prediction but has not considered in data-driven models. The degradation coefficient in Equation (13) was incorporated into Equation (5), a modified new input gate equation can be obtained:

$$i_t = \sigma(-K \cdot W_i \cdot [h_{t-1}, x_t] + b_i) \tag{16}$$

Equations (14) and (15) into Equation (9), a modified new final output equation can be obtained. In addition, the final form of the equation (17) was decided by validation of calibration:

$$h_t = o_t \circ \tanh(c_t + m_t \circ d_t) \tag{17}$$

The hybrid model algorithm (improved LSTM model, Im-LSTM) is constructed based on Equations (3), (4), (6), (7), (8), (16), and (17). The frame of Im-LSTM was shown in Figure 2. In summary, the physical process related to water quality prediction including pollutants degradation and transferring processes etc., which have deeply studied by previous researchers and have not been considered in data-driven model. Referring to the equations and constructure of LSTM model, we constructed equations expressing the physical relationships as well as following the the LSTM frame, such as migration gate and distribution gate. Then, we use degradation coefficient of pol-

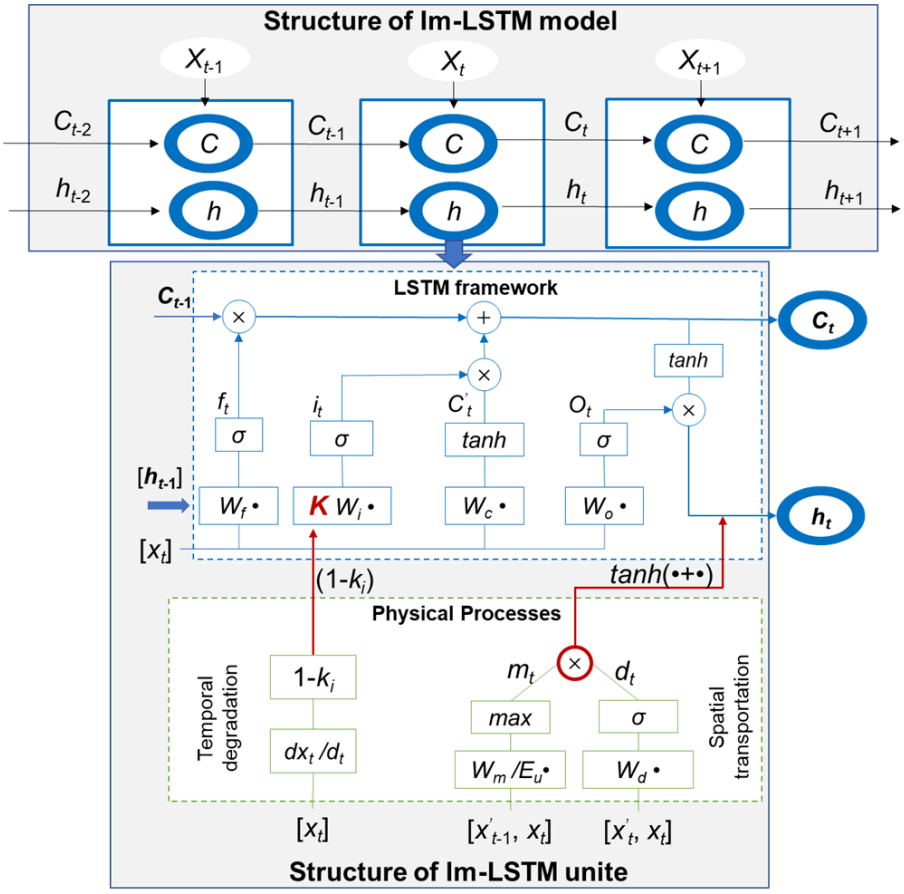

Figure 2. Frame diagram of physical process-data driving hybrid model (Im-LSTM model).

lutants as initial weight to optimize input gate, and incorporated migration gate and distribution gate into output gate to limit final output values. Finally, the data-driven training was guided or limited by physical relationships.

2.3. Model Simulation and Compilation

The simulation of Im-LSTM based on the backpropagation algorithm, including (1) the calculation of memory units for each neuron in the neural network, that is, the values of the forget gate (f_t), input gate (i_t), output gate (o_t), migration gate (m_t), distribution gate (d_t) and other vectors; (2) the error term (δ) of each neuron is calculated by reverse calculation, and then the error is reduced by iterative calculation. The reverse calculation method of the Im-LSTM error term continues the backpropagation mode of the original LSTM, that is, the error is calculated along the time axis of the time sequence information and the hierarchy of the neural network. Based on this, according to the corresponding error term, the gradient of each weight is calculated as following equations:

$$\delta_k^T = \Pi_{j=k}^{t-1}\delta_{o,j}^T W_{oh} + \delta_{f,j}^T W_{fh} + \delta_{i,j}^T W_{ih} + \delta_{c,j}^T W_{ch} + \delta_{m,j}^T W_{mh} + \delta_{d,j}^T W_{dh} \quad (18)$$

$$\frac{\partial E}{\partial net_t^{l-1}} = (\delta_{f,t}^T W_{fx} + \delta_{i,t}^T W_{ix} + \delta_{c,t}^T W_{cx} + \delta_{o,t}^T W_{ox} + \delta_{m,t}^T W_{mx} + \delta_{d,t}^T W_{dx}) \circ f'(net_t^{(l-1)}) \quad (19)$$

Equation (17) is the error term to be transmitted forward to any moment k, where $\delta^T_{o,i}$, $\delta^T_{f,i}$, $\delta^T_{i,i}$, $\delta^T_{c,i}$, $\delta^T_{m,i}$, and $\delta^T_{d,i}$ represent the error terms of the output gate, forget gate, input gate, cell state, migration gate, and distribution gate at time j, respectively. W_{oh}, W_{fh}, W_{ih}, W_{ch}, W_{mh}, and W_{dh} represent the output gate, forget gate, input gate, cell state, migration gate, and distribution gate, respectively, corresponding to the input h_{t-1} weight matrix. Equation (18) is the equation for error transfer to the upper layer, where X is the error function; W_{ix}, W_{fx}, W_{ox}, W_{cx}, W_{mx}, and W_{dx} represent the weight matrix of input X_t corresponding to output gate, forget gate, input gate, cell state, migration gate, and distribution gate, respectively; $net^{l-1}{}_t$ is the $l-1$ layer of $\sum_{ele=0}^{N} \omega_{ele} x_{ele}$; and $f'()$ is the derivative of the activation function. Equations in Table 1 determine the weight gradient of each parameter.

Java and Python languages were used to encode and realize the long-term prediction AI model of water quality, and the AI framework was built by the basic functional modules of theTensorFlow library. The LSTM module of the Keras library

Table 1. LSTM Algorithm Weight Gradient Equations

$\frac{\partial E}{\partial W_{oh}} = \sum_{j=1}^{t} \delta_{o,j} h_{j-1}^{T}$	$\frac{\partial E}{\partial W_{mh}} = \sum_{j=1}^{t} \delta_{m,j} h_{j-1}^{T}$	$\frac{\partial E}{\partial W_{ix}} = \delta_{i,t} X_t^{T}$	$\frac{\partial E}{\partial b_f} = \sum_{j=1}^{t} \delta_{f,j}$
$\frac{\partial E}{\partial W_{fh}} = \sum_{j=1}^{t} \delta_{f,j} h_{j-1}^{T}$	$\frac{\partial E}{\partial W_{dh}} = \sum_{j=1}^{t} \delta_{d,j} h_{j-1}^{T}$	$\frac{\partial E}{\partial W_{cx}} = \delta_{\tilde{c},t} X_t^{T}$	$\frac{\partial E}{\partial b_c} = \sum_{j=1}^{t} \delta_{\tilde{c},j}$
$\frac{\partial E}{\partial W_{ih}} = \sum_{j=1}^{t} \delta_{i,j} h_{j-1}^{T}$	$\frac{\partial E}{\partial W_{ox}} = \delta_{o,t} X_t^{T}$	$\frac{\partial E}{\partial b_o} = \sum_{j=1}^{t} \delta_{o,j}$	$\frac{\partial E}{\partial b_m} = \sum_{j=1}^{t} \delta_{m,j}$
$\frac{\partial E}{\partial W_{ch}} = \sum_{j=1}^{t} \delta_{\tilde{c},j} h_{j-1}^{T}$	$\frac{\partial E}{\partial W_{fx}} = \delta_{f,t} X_t^{T}$	$\frac{\partial E}{\partial b_i} = \sum_{j=1}^{t} \delta_{i,j}$	$\frac{\partial E}{\partial b_d} = \sum_{j=1}^{t} \delta_{d,j}$

was improved according to the method in Section 2.4, adding the Im-LSTM module to obtain the Hybrid Model algorithm in K. The core module of the AI model was a five-layer neural network (see Table 2), of which four layers were the Im-LSTM layers, and the whole connectivity layer (Dense) was finally connected. The activation function σ adopted in the model was exponential linear units (ELUs), which combined the advantages of sigmoid and ReLU to mitigate the disappearance of the model gradient. The activation function of the fully connected layer adopted a linear element, which made the model output conform to the order of magnitude of the sectional monitoring value. The activation function equation of the model is as follows:

$$ELU(x) = \begin{cases} x, & x \geq 0 \\ \alpha(e^{x} - 1), & x < 0 \end{cases} \tag{20}$$

$$\tanh(x) = \frac{e^{x} - e^{-x}}{e^{x} + e^{-x}} \tag{21}$$

$$maxout(x) = \max(w_1^T x + b_1, w_2^T x + b_2, \cdots, w_n^T x + b_n) \tag{22}$$

The loss function of the model compilation adopted the mean square error function (MSE), and the neural network optimization algorithm adopted Adam. The learning rate was set at 0.001, and the fuzzy factor was set at 10-8.

2.4. Model Training and Validation

The training and validation have three steps (see Figure 3). Firstly, the multiple statistical analysis was done to identify potential driving factors to help select predictive variables. Then, the historical data was used to training Im-LSTM model. Finally, the predictive performance was evaluated by different algorithms.

The dataset used in this study included hydrological, meteorological and water quality data. As shown in Figure 1, hydrological condition represented by daily runoff were collected from 4 hydrological stations in the Xiaoqing River Basin. The meteorological data including daily rainfall, daily evaporation and daily average temperature were collected from 6 meteorological stations around the Xiaoqing River Basin. Water quality data included daily automatic detection data of 26 monitoring sections and daily automatic detection data of 390 pollution sources in the river basin. Water quality data collected from river sections were target dataset, others were predictive dataset. To maintain data consistency, we used daily data from the overlapping period (2018 ~ 2020) for modeling simulations. Additionally, data cleaning and preprocessing were performed to remove null values and outliers. Then all datasets were normalized, and the normalization algorithm was min-max.

As shown in Figure 4, the consecutive monitoring data of 7 days setting as one sample. Then, we used random sampling to get 10,000 samples. The proportions of training samples and validation samples were 70% and 30%, respectively. The batch size for training data is set to 128. The stride of the sliding window for the input flow data is set to 1. In all the experiments, the size of the time window is set to 4, which means that 28-day historical data is used as a training unit. The number of iterations of the model was set to 1,000 based on the combination of computer computing ability and the stability of model evaluation parameter changes. The core module parameters were shown in Table 2.

Table 2. Core Module Framework

Network Layer	Output Dimensions	Activation Function
Im-LSTM	(N, 21, 490)	ELU
Im-LSTM	(N, 21, 350)	ELU
Im-LSTM	(N, 21 ,140)	ELU
Im-LSTM	(N, 21, 42)	ELU
Dense	(N, 7)	Linear

There are many methods that can predict water quality for specific and discrete sections. These range from multivariate regression in the traditional statistics period, to artificial neural network (ANN) in the data mining period, and then to RNN and LSTM in the deep learning period. Thus, we carry out experiments to investigate the prediction performance between

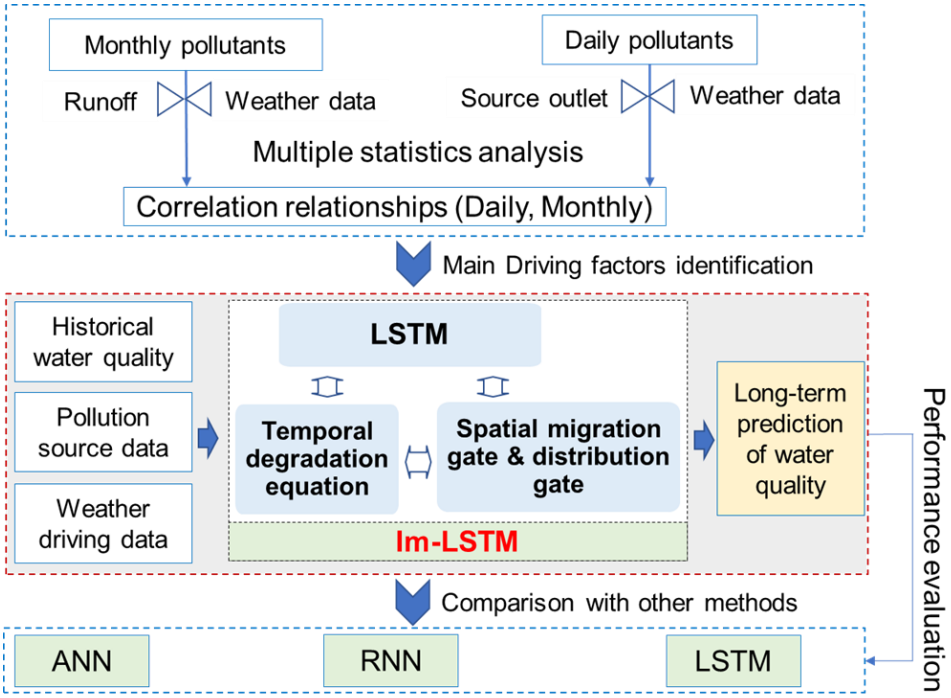

Figure 3. Technical flow of water quality prediction by proposed hybrid model (Im-LSTM).

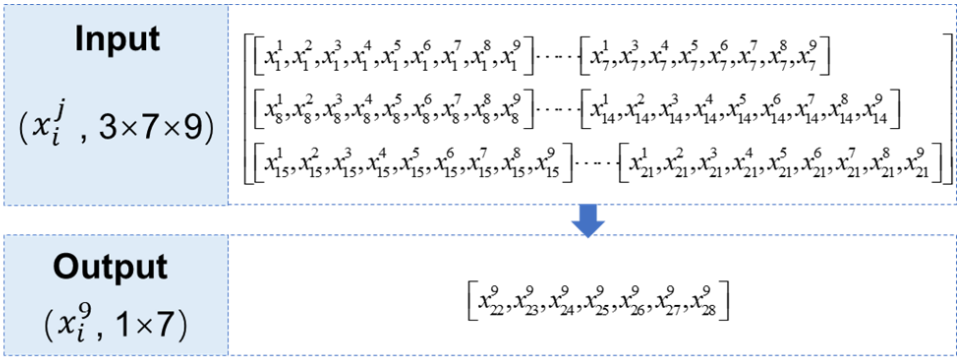

Figure 4. Sample of input and output in Im-LSTM. Note that x_i^j means input of variables (j) at different time (i), j from 1 to 8 means predict variable, such as temperature, precipitation, relative humidity, sunshine duration , potential evaporation, latitude, longitude, point sources discharge; x_i^9 means the concentration of the target pollutants (such as COD) in calculation section at different monitoring time (i). $x_1^9,\ldots,x_{21}^9$ means the input monitoring concentration of target pollutants; $x_{22}^9,\cdots,x_{28}^9$ means the prediction concentration.

the proposed model (Im-LSTM) and the existing methods for water quality prediction including ANN, RNN, and LSTM. In order to ensure the accurate evaluation of the simulation ability of different algorithms, the basic conditions such as the same operating environment, coding structure, input parameters and the number of iterations were adopted in this study. The ANN algorithm adopts a framework of one input layer, three hidden layers and one input layer. The RNN and LSTM algorithms adopt a five-layer structure, and the RNN() and LSTM() coding packages of Keras framework are selected for the first four layers. The final layer is still Dense() coded package, and the input and output parameters are consistent with the Im-LSTM model framework.

The determination coefficient and mean square error were used to evaluate the prediction results of water quality. The specific calculation method is as follows:

$$R^2 = \left[\frac{\sum_{n=1}^{N}(y_n-\bar{y}_n)\sum_{n=1}^{N}(\hat{y}_n-\bar{\hat{y}}_n)}{\sqrt{\sum_{n=1}^{N}(y_n-\bar{y}_n)^2}\sqrt{\sum_{n=1}^{N}(\hat{y}_n-\bar{\hat{y}}_n)^2}}\right]^2 \quad (23)$$

$$MSE = \frac{1}{n}\sum_{n=1}^{N}(\hat{y}_n - y_n)^2 \quad (24)$$

where R^2 is the determination coefficient, MSE is the mean square error, N is the number of samples, y_n is the actual value, and $\hat{y}_n$ is the predicted value.

3. Results and Discussions

3.1. Spatial Relationships of Pollutants

The Euclidean distance of water pollutants was calculated among different sections as shown in Figure 5. The smaller the Euclidean distance, the higher the correlation similarity. Figure 5 shows the correlation between the monitoring sections of Xiaoqing River basin and the external monitoring sections. As shown in the figure, the monitoring sections in Xiaoqing River basin have strong correlation similarity with the water quality of the surrounding basins, showing typical spatial focus, especially on the monitoring sections in the middle and lower reaches of the basin. For example, the square Euclidean distance of water quality of 14 monitored sections (Figure 5, red circle) in the middle reaches of Xiaoqing River Basin is less than 0.02, showing strong correlation similarity. Then, Moran's I was further used to quantify and test the significance of the correlation degree of different water quality indicators in and out of the watershed. Additionally, the values of Moran's I for NH_3-N and COD inside the Xiaoqing River Basin were 0.27 and 0.35, and those for NH_3-N and COD in its surrounding basins were 0.2 and 0.13, respectively. All of the above values of Moran's I reached the highly significant level both inside and surrounding the basin, indicating that the spatial correlation of water pollution factors did exist, and NH_3-N and COD were the most significant.

3.2. Water Quality Prediction by Im-LSTM

The prediction performance of Im-LSTM model for the typical pollutants (COD and NH_3-N) at 26 different sections of water quality monitoring in the Xiaoqing River Basin was evaluated and compared. Table 3 shows the evaluation results of the model for the simulation results of each monitoring section. As shown in Table 3, the mean square error of COD in each monitoring section simulated by the model ranged from 3.1582 to 490.357, and the mean square error of the basin was 38.9982. The mean square error of NH_3-N ranged from 0.0014 to 0.0937, and the mean square error of was basin is 0.0398. The COD determination coefficient of each monitoring section simulated by the model ranged from 0.53 to 0.93, and the mean determination coefficient of its basin was 0.82 of these, the maximum COD determination coefficient appeared at Xinfengzhuang station and the minimum at Zhangjianqiao section (No. 7). The determination coefficient of NH_3-N in all monitored sections simulated by the model ranged from 0.48 to 0.90, and the average determination coefficient of the basin was 0.79. Among them, the maximum determination coefficient of NH_3-N appeared at the Xinfengzhuang section and Wohushan section, and the minimum determination coefficient still appeared at the Zhangjianqiao section. In conclusion, the water quality prediction model based on the Hybrid Model had good simulation ability and could meet the basic needs of water quality supervision in the basin.

As shown in in Figure 6, the average value of the simulation decision coefficients of all non-mainstream monitoring sections is about 0.82, lower than that of all mainstream monitoring sections which is about 0.85, especially for the Lijin section (No. 4) and Zhangjianqiao section (No. 7) with lowest decision coefficients both for COD (Lijin section: 0.57, Zhang jianqiao section: 0.53) and NH_3-N (Lijin section: 0.52, Zhang jianqiao section: 0.48). This is because two monitoring sections located on the edge of Xiaoqing River Basin. So, there is no direct exchange of pollutants with other monitoring sections through water flow. As well as there is no obvious correlation with pollution sources. Thus, the simulation effect is poorest for two monitoring sections of pollutants. In contrast, it is obvious that the average value of the simulation decision coefficients of all main monitoring sections is higher than that of all tributary monitoring sections. For example, the COD simulation determination coefficient of the Xiahouqiao section (No. 11) located on the tributary Xinghua River is lower than that of Tangkouqiao section (No. 24) located at the confluence of the Xinghua River and the Xiaoqing River, and their determination coefficients are 0.84 and 0.91, respectively. The highest decision coefficients of COD and NH_3-N were appeared at the Xinfengzhuang section (No. 3), which also located in the mainstream. In addition, the NH_3-N determination coefficients of mainstream sections from upstream to downstream at Mulizhuang, Xinfengzhuang, Tangkouqiao, Xizha, Fanli, Wangdaozha and Yangkou were 0.88, 0.90, 0.81, 0.82, 0.84, 0.80, and 0.81, respectively, and their determination coefficients were all above 0.80. This is because the Xiaoqing River has close ties between each monitoring section, and there is contact with water for the sand silted tributaries monitoring section. However, the decision coefficients of Shukou section (No. 1, in Figure 7b) is much lower than other sections located in mainstream. Because Shukou section located in upper reaches of the Xiaoqing River, where surrounded by intensive industry-factories. There are 21 sewage treatment plants region and 2 point-sources within around 10 km^2 region directly discharged at of river pollution. Besides, the connection between Shukou section and other section were poor. Thus, the influence of these sources as well as poor connection cause the simulation effect of pollutants at Shukou section to be low. According to the above discussion, the Im-LSTM algorithm fully considers the spatial characteristics between data in the calculation process. So, the model is more suitable for connected sections, which is always appeared more complex exchange process, such as downstream rivers.

3.3. Comparison with Other Methods

As shown in Figure 8, the box plot shows the statistical values of the pollutant simulation results of 26 water quality monitoring sections in the basin of different models (ANN, RNN, LSTM, Im-LSTM) at training and validation stage. The

Table 3. Results of the Model Predicted Water Quality of Monitored Sections

No.	Name of Station	MSE of COD	MSE of NH_3-N	R^2 of COD	R^2 of NH_3-N
1	Shukou	4.3404	0.0014	0.64	0.57
2	Mulizhuang	4.0517	0.0111	0.86	0.88
3	Xinfengzhuang	50.3784	0.0526	0.93	0.90
4	Lijing	3.1582	0.0058	0.57	0.52
5	Chenqiao	23.4271	0.0573	0.87	0.88
6	Yangkou	490.357	0.0484	0.80	0.81
7	Zhangjianqiao	3.2186	0.0074	0.53	0.48
8	Fanli	12.4094	0.0451	0.91	0.84
9	Xizha	20.6733	0.0265	0.76	0.82
10	Wohushan	48.6294	0.0573	0.91	0.90
11	Xiahouqiao	16.4536	0.0328	0.84	0.80
12	Zhangqigou	43.1937	0.0588	0.88	0.86
13	Yuanjiaqiao	23.6192	0.0245	0.90	0.81
14	Ruxiaoqinghe	26.0205	0.0232	0.92	0.81
15	Daoxundu	23.8342	0.0308	0.82	0.85
16	302Jin	23.0768	0.0937	0.77	0.73
17	Ruyubeihe	31.7329	0.0659	0.82	0.78
18	Wangdaozha	15.9493	0.0465	0.80	0.80
19	Xishuimo	12.5783	0.0464	0.87	0.84
20	Xinshalu	21.6856	0.0512	0.78	0.74
21	Nanguoqiao	12.3569	0.0492	0.84	0.83
22	Bamianhe	12.9863	0.0489	0.89	0.86
23	Shumiao	12.0713	0.0374	0.80	0.79
24	Zhangguanzhuang	23.5639	0.0206	0.90	0.83
25	Dongbalu	28.3996	0.0632	0.70	0.71

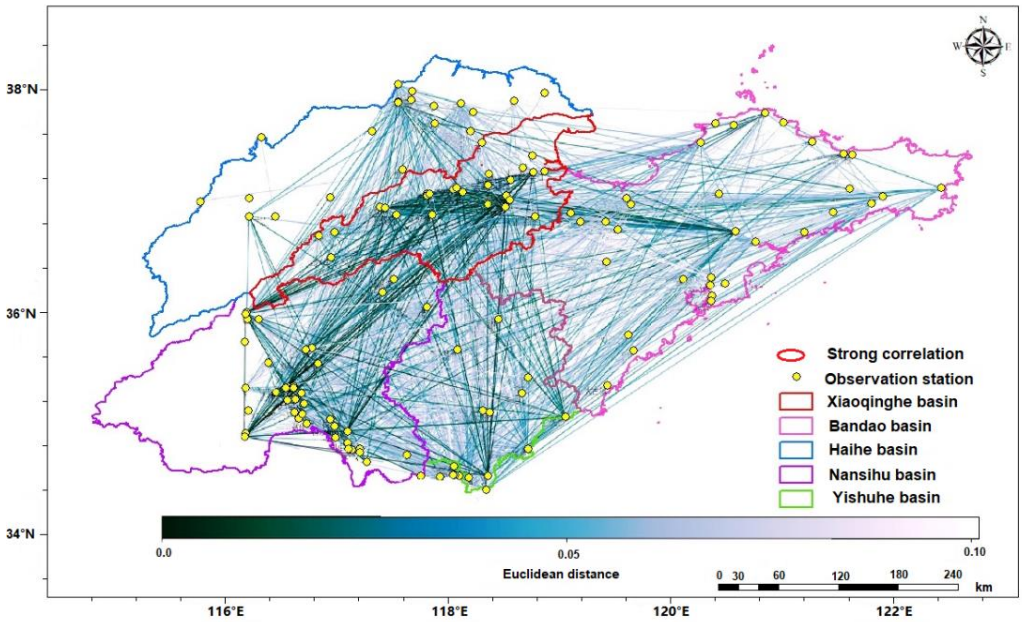

Figure 5. Correlation network of water quality based on Euclidean distance in and around Xiaoqinghe Basin.

average determination coefficients in training and verification stages for ANN, RNN, LSTM and Im-LSTM are 0.56 and 0.52, 0.75 and 0.74, 0.78 and 0.73, 0.81 and 0.80, respectively. As shown in Figure 8(b), the MSE of COD in the three stages presents the Test > Validation, and ANN > RNN > LSTM > ImLSTM are presented in four algorithms. Figure 8(c) shows that the MSE of NH_3-N simulated by Im-LSTM is significantly lower than that of the other three algorithms. The variation trend of different algorithms is same among different simulation stage. So, we're going to focus on comparing performances

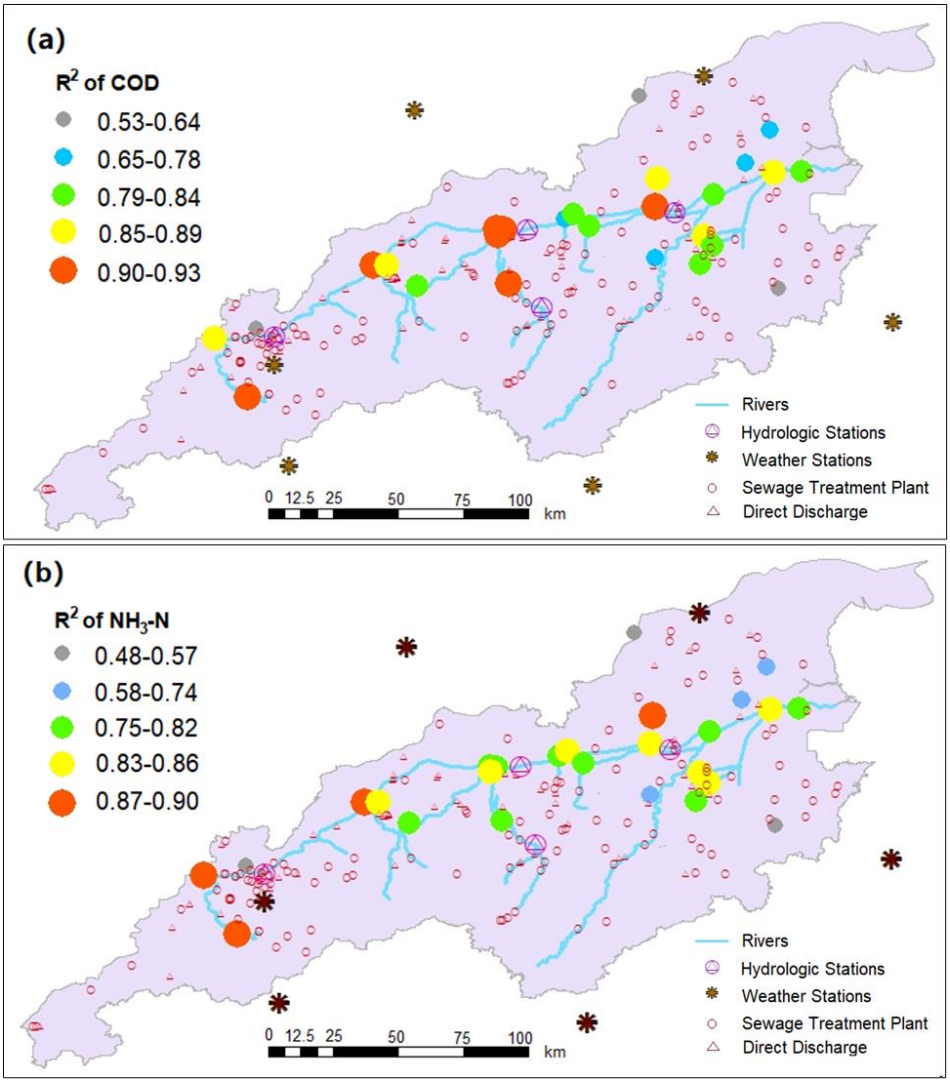

Figure 6. Distribution of water quality prediction determination coefficient (a-COD, b-NH_3-N).

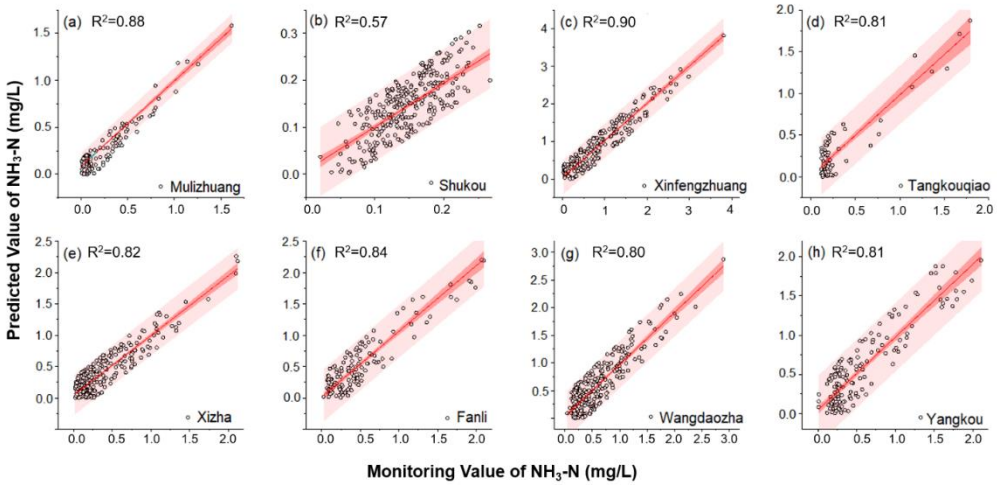

Figure 7. Simulation results of NH_3-N at mainstream sections from upstream to downstream.

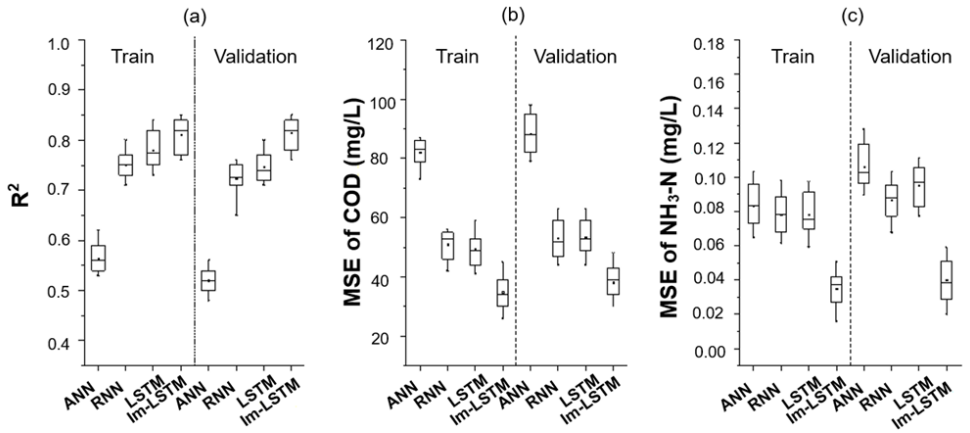

Figure 8. Evaluation of prediction models at different training, validation and testing stages.

of different models during the validation stage. ANN showed lowest decision efficient and highest error. The experimental results demonstrate that ANN has largest prediction error and lowest decision efficient both for COD and NH_3-N among the above prediction models. Compared to other models, ANN belongs to the feedforward neural network model and cannot process the sequence information in the capture and transmission of sequence data, which has approved by other researches (Pang et al., 2020). Compared to ANN, RNN and LSTM achieves a better and similar performance where R^2 have a roughly average 0.2, 0.22 increase, respectively. In addition, LSTM model has a little higher accuracy comparing with the RNN model. This illustrated both RNN and LSTM can well extract and transfer series information of water quality. However, RNN and LSTM is still unable to capture the other inherent features such as spatial features. As above discussed, with the incorporation of spatial relationship into Im-LSTM, the prediction of connected mainstream sections significantly increased, so that the total performance of all sections based on Im-LSTM were increased comparing to LSTM model. The improvement of LSTM by Im-LSTM increases the average simulation determination coefficient by 9.5%, and the mean square error of COD and NH_3-N decreases by 27.4 and 58.4%, respectively, on average. In concluding, Im-LSTM combined the temporal and spatial relationship of water quality so that showed best performance among all these models.

To illustrate the generalization of the proposed model, we investigate the prediction performance under different scenarios. the prediction results of pollutants by the four algorithms for different scenarios as shown in Figure 9. Taking the NH_3-N prediction results of Tangkouqiao section as example, it can also be observed that the proposed model (Im-LSTM) still outperforms the other algorithms with both two metrics (R^2 and MSE) for all prediction horizons, especially for sudden fluctuation scenarios. The sudden increasing loading of NH_3-N in middle stream section (Tangkouqiao section) came from upper inflow as well as agricultural leaching-out, which mainly influenced by rainfall event (Maguire et al., 2019). For the sudden process, the prediction of ANN missed one peak at the the 12th day (Figurer 9). Thus, the performance metrices were worst with lowest R^2 of 0.49 and highest MSE of 0.13. This is because ANN only consider the influence of current condition (Palani et al. 2008), such as the impact of a storm event. However, the temporal influence was not be considered, such as the delayed runoff peak induced by previous rainfall. Even though RNN and LSTM predicted the peaks of NH_3-N loading, the range of prediction error were much higher than that of Im-LSTM. Because Im-LSTM not only considered the temporal influence of current and previous time, but also considered the NH_3-N loading of inflows from upper streams, which constraint the change range of NH_3-H. The prediction of four models for low concentration of NH_3-N were showed fluctuation, which could be induced by variable driven condition (Zheng et al., 2021), such as change of temperature, runoff volume. However, the fluctuation as well as error range of Im-LSTM were lowest. In summary, RNN and LSTM algorithms of the deep learning methods are better than ANN algorithms due to the consideration of temporal information, and the proposal model (Im-LSTM) provides best universality for water quality prediction under different scenarios due to the consideration of temporal and spatial information.

3.4. Limitations and Directions

Because most of the existing models predict the future of the current section based on the historical simulation of the current section, they ignore the spatial correlation characteristics formed by pollutants driven by water flow and other factors, that is, certain concentrations of pollutants in the upstream section migrate to the downstream section and form corresponding concentrations under certain conditions. In this paper, the spatial correlation characteristics of different pollutants were revealed and quantified by combining spatial autocorrelation index and Euclidean distance, which were introduced into the LSTM model to achieve near-real-time water quality prediction of key sections. The prediction results of different sections show that if the predicted section is located in the main stream and is well connected with the tributaries, the prediction effect is better. If the section is located at the edge of the watershed

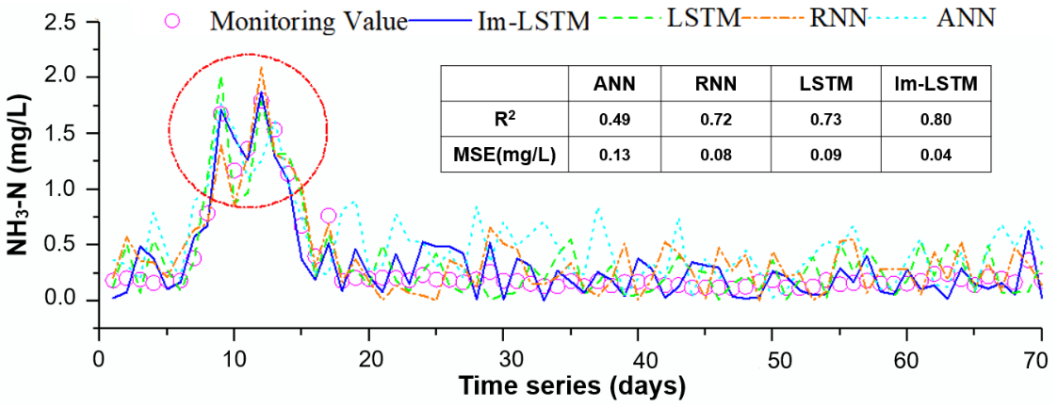

Figure 9. Comparison of prediction performance of different algorithms at Tangkouqiao section.

and has poor connection with the flow and material of other sections, the prediction effect is also poor. Because the Im-LSTM algorithm takes into account the spatial correlation characteristics of pollutants caused by water flow connection in the calculation process, that is, the water quality of the middle and lower reaches of the watershed is highly affected by the water quality of the upstream section, the identification and quantification of the driving factors of the water quality change of the upstream section are still insufficient. Thus, the model is more suitable for downstream and connective sections. For the later model improvement, different copula functions can be set for different sections or regions to carry out constraints.

Additionally, in order to support the intelligent early warning management of water environment, the prediction model not only needs to improve the accuracy of the prediction results, but also needs to improve the interpretability of the change process, which can not be accomplished independently by a single artificial intelligence model or mechanism model. In this study, the method and effect of the coupling between the mechanism experience and the intelligent algorithm were explored. The results show that the coupling between the mechanism experience relationship and the deep learning algorithm can not only improve the accuracy of water quality prediction, but also improve the interpretability of the prediction error sources in different sections. However, because the expression of mechanism process is inconsistent with the expression of underlying architecture and process logic of deep learning algorithm, the current coupling sum is limited to the transmission of data streams of the two, and the underlying logic expression mode integrating mechanism and data has not been truly developed to realize the tight coupling of mechanism data. Thus, it is needed to develop better incorporation mechanisms between physical mechanism and data analysis.

4. Conclusions

Water quality prediction at basin scale has become an urgent demand owing to the increase in water pollution, but the prediction is difficult to achieve because of the complex changing processes under variable boundaries. Currently, mechanism models are primarily used for water quality prediction. However, the future boundary conditions of this model can only be set by history or manual setting, so it cannot automatically predict the future watershed water quality. With the development of big data, artificial intelligence technology, especially Long short-term memory network (LSTM), has remedied the deficiency of mechanism models and has been widely used on water quality prediction. However, its reliability depends only on a large dataset, which is suitable for and discrete sections with enough data and difficult to extend to regions with limited monitoring data owing to without considering physical processes in space and time at basin scale.

In this study, the spatial correlation feature extraction mechanism and pollutant degradation parameters of discrete section water quality are introduced into the LSTM algorithm equation, and the Hybrid Model (Im-LSTM) is proposed. The main improvement is following: on the basis of ensuring the original characteristics of LSTM algorithm, IM-LSTM algorithm constructs a state unit at the front end of LSTM that can obtain the spatial correlation of water quality at different monitoring sections, namely Migration Gate and Distribution Gate. The ability to capture the spatial characteristics of water quality between different monitoring sections is enhanced. Then, the prediction performance was evaluated at different monitoring sections at Xiaoqing River basin. The result showed that the average performance coefficients of mainstream monitoring sections were better than that of non-mainstream. Because the connection and exchange of water flow were more significant among mainstream sections, which have been well expressed by Im-LSTM model. Finally, the comparison between the Im-LSTM and the traditional data mining algorithm (ANN, RNN and LSTM) was carried out to demonstrate the outperformance of the Water quality prediction, especially for the prediction of sudden water pollution events. And, Im-LSTM improved the simulation determination coefficient by 9.5%, and the mean square error of COD decreased by 27.4% and NH_3-N by 58.4% on average.

Acknowledgments. This study was supported by the National Science Fund for Distinguished Young Scholars (No. 52125901). National Natural Science Foundation of China (Grant No. 52209041).

References

A, Y.L., Wang, G.Q., Hu, P., Lai, X.Y., Xue, B.L. and Fang, Q.Q. (2022). Root-zone soil moisture estimation based on remote sensing data and deep learning. *Environ. Res.*, 212(B), 113278. https://doi.org/10.1016/j.envres.2022.113278

Areerachakul, S., Sophatsathit, P. and Lursinsap, C. (2013). Integration of unsupervised and supervised neural networks to predict dissolved oxygen concentration in canals. *Ecol. Modell.*, 261-262, 1-7. https://doi.org/10.1016/j.ecolmodel.2013.04.002

Chang, F.J., Chen, P.A., Chang, L.C. and Tsai, Y.H. (2016). Estimating spatio-temporal dynamics of stream total phosphate concentration by soft computing techniques. *Sci. Total Environ.*, 562, 228-236. https://doi.org/10.1016/j.scitotenv.2016.03.219

Chen, H., Huang, J.J., Dash, S.S., Wei, Y.Z. and Li, H. (2022). A hybrid deep learning framework with physical process description for simulation of evapotranspiration. *J. Hydrol.*, 606, 127422. https://doi.org/10.1016/j.jhydrol.2021.127422

Chen, Z., Xu, H., Jiang, P., Yu, S., Lin, G., Bychkov, I., Hmelnov, A., Ruzhnikov, G., Zhu, N. and Liu. Z. (2021). A transfer Learning-Based LSTM strategy for imputing Large-Scale consecutive missing data and its application in a water quality prediction system. *J. Hydrol.*, 602, 126573. https://doi.org/10.1016/j.jhydrol.2021.126573

Dogan, E., Ates, A., Yilmaz, E.C. and Eren, B. (2008). Application of artificial neural networks to estimate wastewater treatment plant inlet biochemical oxygen demand. *Environ. Prog.*, 27(4), 439-446. https://doi.org/10.1002/ep.10295

Dogan, E., Sengorur, B. and Koklu, R. (2009). Modeling biological oxygen demand of the Melen River in Turkey using an artificial neural network technique. *J. Environ. Manage.*, 90(2), 1229-1235. https://doi.org/10.1016/j.jenvman.2008.06.004

Elhatip, H., and Komur, M.A. (2008). Evaluation of water quality parameters for the Mamasin dam in Aksaray City in the central Anatolian part of Turkey by means of artificial neural networks. *Environ. Geol.*, 53, 1157-1164. https://doi.org/10.1007/s00254-007-0705-y

Flores, E., Zortea, M. and Scharcanski. J. (2019). Dictionaries of deep features for land-use scene classification of very high spatial resolution images. *Pattern Recognit.*, 89, 32-44 https://doi.org/10.1016/j.patcog.2018.12.019

Hong, Y. S. T. (2012). Dynamic nonlinear state-space model with a neural network via improved sequential learning algorithm for an online real-time hydrological modeling. *J. Hydrol.*, 468-469, 11-21. https://doi.org/10.1016/j.jhydrol.2012.08.001

Hu, Z.H., Zhang, Y.R., Zhao, Y.C., Xie, M.S., Zhong, J.Z., Tu, Z.G. and Liu, J.T. (2019). A water quality prediction method based on the deep LSTM network considering correlation in smart mariculture. *Sens.*, 19(6), 1420. https://doi.org/10.3390/s19061420

Koppa, A., Rains, D., Hulsman, P., Poyatos, R. and Miralles, D.G. (2022). A deep learning-based hybrid model of global terrestrial evaporation. *Nat. Commun.*, 13, 1912. https://doi.org/10.1038/s41467-022-29543-7

Lazer, D., Kennedy, R., King, G. and Vespignani, A. (2014). The parable of Google Flu: traps in big data analysis. *Sci.*, 343(6163), 1203-1205. https://doi.org/10.1126/science.1248506

Liu, W., Lai, Z., Bacsa, K. and Chatzi. E. (2022). Physics-guided Deep Markov Models for learning nonlinear dynamical systems with uncertainty. *Mech. Syst. Signal Process.*, 178, 109276. https://doi.org/10.1016/j.ymssp.2022.109276

Maguire, T.J., Spencer, C., Grgicak-Mannion, A., Drouillard, K., Mayer, B. and Mundle, S.O. (2019). Distinguishing point and non-point sources of dissolved nutrients, metals, and legacy contami-nants in the Detroit River. *Sci. Total Environ.*, 681, 1-8. https://doi.org/10.1016/j.scitotenv.2019.04.311

Nayak, P., Venkatesh, B., Krishna, B. and Jain, S.K. (2013). Rainfall-runoff modeling using conceptual, data driven, and wavelet based computing approach. *J. Hydrol.*, 493, 57-67. https://doi.org/10.1016/j.jhydrol.2013.04.016

Palani, S., Liong, S.Y. and Tkalich. P. (2008). An ANN application for water quality forecasting. *Mar. Pollut. Bull.*, 56(9), 1586-1597. https://doi.org/10.1016/j.marpolbul.2008.05.021

Pang, Z., Niu, F. and O'Neill, Z. (2020). Solar radiation prediction using recurrent neural network and artificial neural network: A case study with comparisons. *Renewable Energy*, 156, 279-289. https://doi.org/10.1016/j.renene.2020.04.042

Piotrowski, A.P., Napiorkowski, M.J., Napiorkowski, J.J. and Osuch, M. (2015). Comparing various artificial neural network types for water temperature prediction in rivers. *J. Hydrol.*, 529(1), 302-315. https://doi.org/10.1016/j.jhydrol.2015.07.044

Prasad, D. V.V., Venkataramana, L.Y., Kumar, P.S., Prasannamedha, G., Harshana, S., Srividya, S.J., Harrinei, K. and Indraganti, S. (2022). Analysis and prediction of water quality using deep learning and auto deep learning techniques. *Sci. Total Environ.*, 821, 153311. https://doi.org/10.1016/j.scitotenv.2022.153311

Read, J.S., Jia, X., Willard, J., Appling, A.P., Zwart, J.A., Oliver, S. K., Karpatne, A., Hansen, G. J.A., Hanson, P.C., Watkins, W., Steinbach, M. and Kumar. V. (2019). Process-Guided deep learning predictions of lake water temperature. *Water Resour. Res.*, 55(11), 9173-9190. https://doi.org/10.1029/2019WR024922

Siljic, A., Antanasijević, D., Perić-Grujić, A., Ristić, M. and Pocajt, V. (2015). Artificial neural network modelling of biological oxygen demand in rivers at the national level with input selection based on Monte Carlo simulations. *Environ. Sci. Pollut. Res.*, 22, 4230-4241. https://doi.org/10.1007/s11356-014-3669-y

Su, Y., Yang, C. and Qiao, J. (2019). A self-organizing cascade neural network for nonlinear system modeling. *Chinese Control Conference (CCC)*, Guangzhou, China, 1598-1603, https://doi.org/10.23919/ChiCC.2019.8866199.

Wang, G.Q., Li, J.W., Sun, W.C., Xue, B.L., A, Y.L. and Liu, T.X. (2019a). Non-point source pollution risks in a drinking water protection zone based on remote sensing data embedded within a nutrient budget model. *Water Res.*, 157, 238-246. https://doi.org/10.1016/j.watres.2019.03.070

Wang, P.Z., Yao, J.P., Wang, G.Q., Hao, F.H., Shrestha, S., Xue, B.L., Xie, G. and Peng. Y.B. (2019b). Exploring the application of artificial intelligence technology for identification of water pollution characteristics and tracing the source of water quality pollutants. *Sci. Total Environ.*, 693, 133440. https://doi.org/10.1016/j.scitotenv.2019.07.246

Wang, Y. Y., Zhou, J., Chen, K. J., Wang, Y. Y. and Liu. L. F. (2017). Water Quality Prediction Method Based on LSTM Neural Network. *12th International Conference on Intelligent Systems and Knowledge Engineering (ISKE)*, Nanjing, China, 1-5, https://doi.org/10.1109/ISKE.2017.8258814.

Xie, J. and Yao. B. (2022). Physics-constrained deep active learning for spatiotemporal modeling of cardiac electrodynamics. *Comput. Biol. Med.*, 146, 105586. https://doi.org/10.1016/j.compbiomed.2022.105586

Xie, K., Liu, P., Zhang, J.Y., Han, D.Y., Wang, G.Q. and Shen, C.P. (2021). Physics-guided deep learning for rainfall-runoff modeling by considering extreme events and monotonic relationships. *J. Hydrol.*, 603(C), 127043. https://doi.org/10.1016/j.jhydrol.2021.127043

Xu, J., Xu, M., Zhao, Y.X., Wang, S.F., Tao, M.H. and Wang, Y.G. (2021). Spatial-temporal distribution and evolutionary haracteristics of water environment sudden pollution incidents in China from 2006 to 2018. *Sci. Total Environ.*, 801, 149677. https://doi.org/10.1016/j.scitotenv.2021.149677

Xue, B. L., Wang, G., Xiao, J.F., Helman, D., Sun, W.C., Wang, J.H. and Liu, T.X. (2020). Global convergence but regional disparity in the hydrological resilience of ecosystems and watersheds to drought. *J. Hydrol.*, 591, 125589. https://doi.org/10.1016/j.jhydrol.2020.125589

Yan, J.Z., Xu, Z.B., Yu, Y.C., Xu, H.X. and Gao, K.L. (2019). Application of a hybrid optimized BP network model to estimate water quality parameters of Beihai Lake in Beijing. *Appl. Sci.*, 9 (9), 1863. https://doi.org/10.3390/app9091863

Yeon, I.S., Kim, J.H. and Jun, K.W. (2008). Application of artificial intelligence models in water quality forecasting. *Environ. Technol.*, 29(6), 625-631. https://doi.org/10.1080/09593330801984456

Zhang, Y.T., Li, C.L., Jiang, Y.Q., Sun, L., Zhao, R.B., Yan, K.F. and Wang, W.H. (2022). Accurate prediction of water quality in urban drainage network with integrated EMD-LSTM model. *J. Cleaner Prod.* 354, 131724. https://doi.org/10.1016/j.jclepro.2022.131724

Zheng, H.F., Lin, F, Feng, X.X. and Chen. Y.J. (2021). A Hybrid deep learning model with Attention-Based Conv-LSTM networks for Short-Term traffic flow prediction. *IEEE Trans. on Intell. Transp. Syst.*, 22(11), 6910-6920. https://doi.org/10.1109/TITS.2020.2997352.