ISSN 1330-3651 (Print), ISSN 1848-6339 (Online)
https://doi.org/10.17559/TV-20230519000647
Preliminary communication

Water Quality Prediction Method Based on OVMD and Spatio-Temporal Dependence

Haitao MENG, Jinling SONG*, Liming HUANG, Yijin ZHU, Meining ZHU, Jingwu ZHANG

Abstract: Water quality changes at one monitoring spot are not only related to local historical data but also spatially to the water quality of the adjacent spots. Additionally, the non-linear and non-stationary nature of water quality data has a significant impact on prediction results. To improve the accuracy of water quality prediction models, a comprehensive water quality prediction model has been initially established that takes into account both data complexity and spatio-temporal dependencies. The Optimal Variational Mode Decomposition (OVMD) technology is used to effectively decompose water quality data into several simple and stable time series, highlighting short-term and long-term features and enhancing the model's learning ability. The component sequence and spot adjacency matrix are used as the input of Graph Convolutional Network (GCN) to extract the spatial characteristics of the data, and the spatio-temporal dependencies of water quality data at different spots are obtained by combining GCN into the neurons of Gated Recurrent Unit (GRU). The attention model is added to automatically adjust the importance of each time node to further improve the accuracy of the training model and obtain a multi-step prediction output that more closely aligns with the characteristics of water quality change. The proposed model has been validated with real monitoring data for ammonia nitrogen (NH3-N) and total phosphorus (TP), and the results show that the proposed model is better than ARIMA, GRU and GCN+GRU models in terms of prediction results and it shows obvious advantages in the benchmark comparison experiment, which can provide reliable evidence for water pollution source traceability or early warning.

Keywords: attention model; GCN; GRU; optimal variational modal decomposition; water quality

1 INTRODUCTION

Water pollution is currently a serious issue that hinders sustainable development in society [1]. Timely and effective monitoring of water environments is crucial for preventing and managing water pollution. With the rise of artificial intelligence, using neural networks for water quality prediction [2] has become a hot topic in research, shifting the focus from post-pollution prevention to pre-warning [3]. Although research on water quality prediction using quasi-recurrent neural network Echo State Networks (ESN) [4], recurrent neural networks Long Short-Term Memory (LSTM) [5] and Gated Recurrent Unit (GRU) [6] to predict time series relationships at the individual spot is prevalent, water quality at one spot is not only closely related to its historical conditions but also affected by pollutants from upstream and tributaries. Therefore, water quality prediction should not only take into account the time-relatedness of data at the individual spot, but also the spatial dependencies between data at different spots. Additionally, water quality data is non-linear and non-stationary, its high complexity also greatly increases the training difficulty of the prediction model. To improve the prediction effect of time series, the combination models of LSTM (GRU) and single module have begun to be applied in fields such as atmospheric pollutant prediction [7], air quality prediction [8], transportation prediction [9-11], and parking lot prediction [12] in recent years. Although these achievements have achieved good prediction results in these fields, they have not been able to effectively solve the above-mentioned problems in water quality prediction. Inspired by these studies, we attempt to integrate a variety of techniques to initially build a water quality prediction model. To reduce noise and improve the training effect, linear decomposition of water quality time series is performed with optimal variational mode decomposition (OVMD), which can reduce the data complexity. A good performance graph convolutional network (GCN) [13] and GRU [14] are selected to model the spatio-temporal complexity dependence of the monitoring spot data. At the same time, in order to better capture the influence of each time point on the current prediction value, an attention model is introduced to adjust the importance of each time point and obtain multi-step prediction output for each component. The experimental results show that the proposed comprehensive prediction model can obtain more accurate results than the traditional time series prediction model ARIMA, single GRU prediction model, and GCN+GRU composite prediction model.

2 DEFINITION OF THE PROBLEM

2.1 Problem Description

Water quality prediction is based on historical data of monitoring indicators from each station. Water quality monitoring indicators generally contain potassium permanganate (KMnO4), dissolved oxygen (DO), ammonia nitrogen (NH3-N), total phosphorus (TP), etc. A spatiotemporal relationship model of water quality data can be established by some techniques, so the water quality data in a future certain time of each station can be predicted.

Definition 1 (Site map G): Determine the topology $G = (V, E)$ of water quality monitoring sites in a basin, of which $V = \{v_1, v_2, \ldots, v_N\}$ is the site set and N is the number of sites; E is the boundary set between sites, representing the upstream and downstream relationship between the two sites, and the edge weight is 0 if there is no upstream and downstream relationship between sites. To upstream or downstream sites, considering that spots far away are less affected when pollutants propagate along the water network, the distance reciprocal between water network sites is taken as the edge weight, and the weighted adjacency matrix $A \in R^{N\times N}$ will be constructed. The site map G can model the complex dependency of monitoring data between sites.

Definition 2 (Feature matrix): Let the input feature matrix of the model be $X^{N\times F\times T_{in}}$, of which N is the number of sites, F is the input feature dimension, and T_{in} is the length of the input window. X_i is the input feature of the model when the time slice i.

Definition 3 (Water quality prediction task): Learn a mapping function $f(\bullet)$ on site map G and input feature matrix X to predict water quality data for all sites for a future period based on real historical monitoring data.

$$\left[X_{t+1}, ..., X_{t+T_{out}}\right] = f\left(G; \left(X_{t-T_{in}}, ..., X_{t-1}, X_t\right)\right) \quad (1)$$

where, T_{in} is the input window length and T_{out} is the prediction length.

2.2 Water Quality Prediction Process

The specific steps to build a comprehensive water quality prediction model based on a variety of techniques are as follows: Firstly, the historical monitoring data of each site is processed for missing value and normalization, and then the data is decomposed into optimal variational modes. Secondly, a weighted adjacency matrix is built according to the distribution of monitoring sites on the water network and water network distance to represent the upstream and downstream relationships, with monitoring data on each time slice of the site as a node attribute. Thirdly, to obtain the multi-step prediction output for each station, GCN is used to learn and aggregate the spatial dependencies between neighboring nodes, and GRU is used to capture the temporal dependencies between input sequences and output values, where the importance of each time node is adjusted using attention mechanism. Finally, the output of each component is summed and inverse normalized to obtain the actual prediction results. The water quality prediction process is shown in Fig. 1.

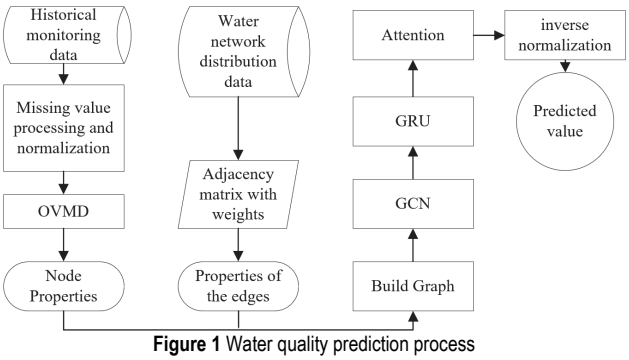


Figure 1 Water quality prediction process

3 RESEARCH METHODOLOGY

3.1 Missing Value Processing and Normalization

The lengths of various data sequences need to be consistent as input features of the water quality prediction model, and the water quality time series of all spots are supplemented based on the maximum length, where the missing data and abnormal values are supplemented as missing values. The missing value completion method combines multiple imputation (MI) and the Bi-GRU algorithm [15]. Multiple imputation can compensate for the shortcomings of GRU in predicting horizontal data, while GRU can learn the features of data in the time dimension, which can compensate for the noise generated by multiple imputation. The missing value completion process is shown in Fig. 2, and the complemented data of each spot can be used as the basic data for model training.

$$Z = \frac{X - \bar{X}}{S} \quad (2)$$

X is the raw data, $\bar{X}$ is the sample mean, S is the sample standard deviation, Z is the normalized data.

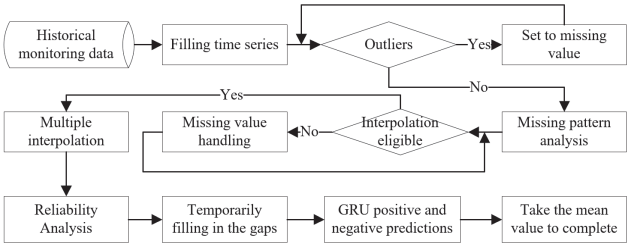


Figure 2 Missing value completion process

3.2 Mode decomposition

Currently, popular mode decomposition methods include Empirical Mode Decomposition (EMD) and Variational Mode Decomposition (VMD) [16]. Optimal Variational Mode Decomposition (OVMD) is an optimized version of VMD. Tab. 1 compares these methods. OVMD has the advantages of high temporal and frequency resolution and high robustness to noise [17], as well as the ability to optimize the number of modes (K) and the update step size (tau). It has high recognition accuracy and low computational and implementation complexity. Therefore, OVMD is selected to decompose water quality monitoring data that contain interference noise to reduce data complexity.

The residual index [18] is calculated according to the Eq. (3).

$$REI = \min \frac{1}{N} \sum_{i=1}^{N} \left| \sum_{k=1}^{k} U_k - f \right|_i \quad (3)$$

U is the number of each decomposition mode, f is the original signal, and N is the number of signals.

Table 1 Comparison of modal decomposition methods

Methods	Identification accuracy	Computational complexity	Implementation complexity	Pros	Cons
EMD	High	High	High	Able to capture stable patterns well in signals	Endpoint effects may occur in the decomposition
VMD	Medium	Medium	Medium	Non-recursive variational mode decomposition method for signal decomposition, can specify the number of modes desired	Mode overlapping may occur if K value selection is not appropriate
OVMD	High	Medium	Medium	It can effectively capture stable patterns in signals, can be used for non-stationary data, and can identify time-based modes	It requires optimization of K value and tau, and the implementation process is slightly more complex compared to VMD

Table 2 Comparison of graph neural network

Network	Computational complexity	Implementation complexity	Scope of application	Pros	Cons
GNN	High	High	Average	Information in its vicinity can be represented at any depth	Influenced by the number of neighboring nodes when computing graph convolution
GraphSAGE	Lower	High	Suitable for large graph data	Capable of handling large graph data with low computational complexity	May not capture the fine-grained information in the graph data well enough to handle weighted graphs
GCN	High	High	Average	Capable of capturing structural information in graph data very well	High computational complexity, may not be suitable for large graph data

Table 3 Comparison of recurrent neural network methods

Methods	Computational complexity	Implementation complexity	Speed of training	Pros	Cons
ESN	Relatively low	Simpler	Faster	Fast training speed and good modeling of one-dimensional time series	Difficult to determine the storage pool connection structure and weights, poor prediction accuracy for non-stationary and non-linear data, and a large number of parameter adjustments
RNN	Relatively low	Simpler	Faster	Simple, relatively easy to implement; lower computational complexity, faster training	Prone to gradient disappearance or gradient explosion problems, difficult to deal with long-term dependency problems
LSTM	Higher	More complex	Slower	Flexible structure, can adjust the size of the door according to the need	Higher computational complexity and slower training speed; more complex implementation
GRU	Moderate	Simpler	Slower	Simple structure and relatively low computational complexity	Slightly inferior to RNN in training speed

3.3 Graph Neural Network

The basic graph neural networks mainly include GNN, GraphSAGE, and GCN, and the comparative analysis of these networks is shown in Tab. 2. GCN is the application of CNN [19] on graph data to effectively capture structural information in the graph, which can resolve GNN's sensitivity to the number of neighbouring nodes and Graph SAGE's inability to handle weighted graphs. The spot graph data used in this experiment is relatively small and the water network distance is taken as edge weights, so GCN is more suitable for modelling the spatial relationships of water quality data among different spots.

3.4 Recurrent Neural Network

The common recurrent neural networks include RNN (Recurrent Neural Network) [20], LSTM and GRU. ESN is a quasi-recurrent neural network which has some characteristics of recurrent neural networks. Therefore, the performance of these four networks is compared in Tab. 3. Because ESN requires setting more parameters during training, most studies on time series prediction use recurrent neural networks at present. Furthermore, LSTM and GRU have comparable prediction effects and both are superior to RNN, but GRU has fewer parameters, a simpler structure and easier training than LSTM, therefore, GRU is chosen to learn the temporal dependencies of water quality data

3.5 Attention Model

Currently, available attention models can be categorized into various types such as soft and hard attention, global and local attention, self-attention, etc. Among them, soft and hard attention is a common pair of models. Hard attention can only discover single important relations which are difficult to train, whereas soft attention allows the model to focus on multiple positions simultaneously and captures various important relations between different positions to improve prediction accuracy effectively. The performance comparison between the two models is shown in Tab. 4. It is necessary to consider various relations in historical data to predict future values in water quality prediction, so soft attention is more suitable.

Table 4 Comparison of attentional models

Models	Computational efficiency	Explainability	Learning ability	Pros	Cons
Soft attention	Medium	Medium	High	The interpretability of the attention weights is strong, and it can automatically learn the position of important information	Computationally expensive
Hard attention	High	Low	High	The computation cost is low and it can ensure that every element is processed correctly	The interpretability of the attention weights is poor and the computation efficiency is low

The implementation of the soft attention model for time series $x_i (i = 1, 2, \ldots, n)$ is as follows: for the hidden state $H = h_i (i = 1, 2, \ldots, n)$ of GRU in different time steps, the multi-layer perceptron (MLP) in Eq. (4) is adapted to calculate the weights of each hidden state, and then the Softmax normalization exponential function in Eq. (5) is used to calculate the feature weights a_i. Then the context vector (C_t) that describes the global water quality changes

can be calculated by the attention function in Eq. (6).

$$e_i = w_{(2)}(w_{(1)}H + b_{(1)}) + b_{(2)} \tag{4}$$

$$\alpha_i = \frac{\exp(e_i)}{\sum_{k=1}^{n}\exp(e_k)} \tag{5}$$

$$C_t = \sum_{i=1}^{n}\alpha_i \cdot h_i \tag{6}$$

$w_{(1)}$ and $b_{(1)}$ are the weight and deviation of the first layer, $w_{(2)}$ and $b_{(2)}$ are the weight and deviation of the second layer.

4 WATER QUALITY PREDICTION MODEL FRAMEWORK

Several components can be obtained from water quality data after OVMD decomposition. The corresponding prediction results can be obtained with GCN+GRU for each component, and then the final prediction results can be obtained by merging all component predictions. The water quality prediction model framework is shown in Fig. 4. Among them, GCN is combined in the neurons of GRU, and the original input feature is replaced by the graph convolution operation with adjacency matrix A and input feature X_t. Therefore, the reset and update gate values of GRU can be calculated according to Eq. (7) and Eq. (8), and the hidden state value of GRU can be calculated according to Eq. (9) and Eq. (10), which makes the hidden state contain the spatio-temporal dependencies of data. Then, the hidden state is input into the attention model to calculate the weights (a_{t-n}, …, a_{t-1}, a_t) of all time nodes, and the context vector C_t of global water quality information is calculated with a weighted sum. Finally, the prediction result is output with the fully connected layer.

$$r_t = \sigma\left(W_r \cdot \left[gc\left(A, X_t\right)\right] + U_r h_{t-1} + b_r\right) \tag{7}$$

$$z_t = \sigma(W_z \cdot \left[gc\left(A, X_t\right)\right] + U_z h_{t-1} + b_z) \tag{8}$$

$$\tilde{h}_t = \tanh(W \cdot \left[gc\left(A, X_t\right)\right] + U \cdot (r_t \cdot h_{t-1}) + b) \tag{9}$$

$$h_t = (1 - z_t) \cdot h_{t-1} + z_t \cdot \tilde{h}_t \tag{10}$$

In the above equations, $gc(A, X_t)$ means that the graph convolution operation is performed on the adjacent matrix A and the input feature X_t to obtain a feature matrix containing spatial dependencies. The U and W with indices represent the weights between corresponding layers, b is the bias, σ is the sigmoid activation function, * represents element-wise multiplication, and tanh represents the hyperbolic tangent function. Eqs. (7) and Eq. (8) convert the hidden state h_{t-1} at time t - 1 and the spatial feature matrix represented by $gc(A, X_t)$ into the reset gate and update gate of GRU with the sigmoid function for information reset, forgetting and memory control. Eq. (9) selectively retains the spatial feature matrix represented by $gc(A, X_t)$ and the resetting data of hidden state h_{t-1} through activation function tanh to obtain the internal candidate state $\tilde{h}_t$. Eq. (10) is to forget and selectively choose some dimensions information of h_{t-1} and $\tilde{h}_t$ to obtain the output state h_t and update the memory. Through the above operations, GCN + GRU can capture both short-term and long-term spatio-temporal dependencies in the data sequence, enabling an understanding of the dynamic changes and complex topology of the data. To enhance the learning effect of the GCN module, the GCN of 2 layers is used to aggregate spatial features effectively and avoid over-smoothing, and the identity matrix of the node adjacent matrix is adjusted to enhance the self-connection weights. The above adjustments are helpful to improve the performance of the prediction model.

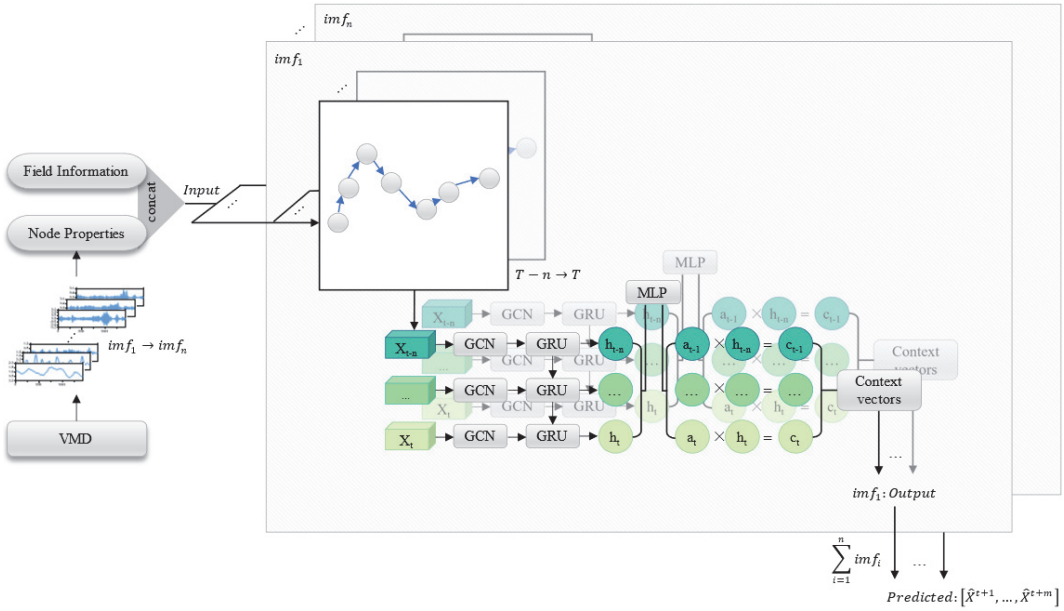


Figure 4 Water quality prediction model structure

5 EXPERIMENT AND ANALYSIS

5.1 Datasets

The water quality data in our study were derived from actual monitoring data from seven monitoring spots in the Mulan Creek Basin, Fujian Province. The distribution of the spots is shown in Fig. 5, and the pins represent each station. The selected water quality indicators were NH3-N and TP, with recording 6 monitoring sessions per day (for four hours) during January 2021 and August 2021 at each station, and data volumes of 1440 for each water quality indicator. Because sensor failure and other problems lead to abnormal or missing values in the monitoring data, the monitoring data outside the effective value range in the original data are considered abnormal values and participate in the missing value processing based on the effective value range of NH3-N and TP in the national standard for surface water quality [21].

5.2 Missing Value Processing

(1) Multiple imputation of data with SPSS. The dataset of NH3-N and TP at each station were analyzed for missing patterns, as shown in Fig. 6 (variables indicate station numbers), and the result showed that there was no full missing pattern, which was eligible for multiple imputation. Multiple imputation was implemented based on IBM SPSS Statistics 27, where the number of interpolations was set to 20 and the interpolation method was MCMC, the maximum number of iterations was 50, and the scalar variable model type was set to PMM. The interpolation result with the highest Cronbach's coefficient alpha was selected to fill the missing values according to the confidence range of Cronbach's coefficient alpha given by David L. Streiner [22]. A comparison of NH3-N and TP dataset before and after interpolation in Jinfengqiao station is shown in Fig. 7 and Fig. 8.

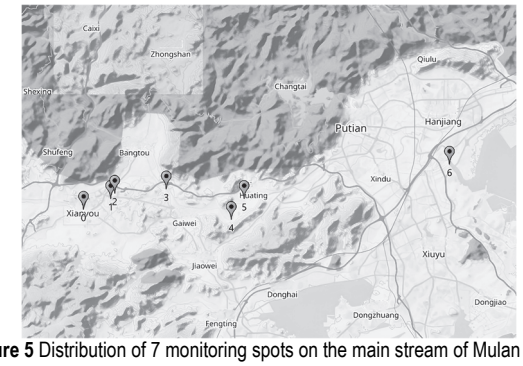
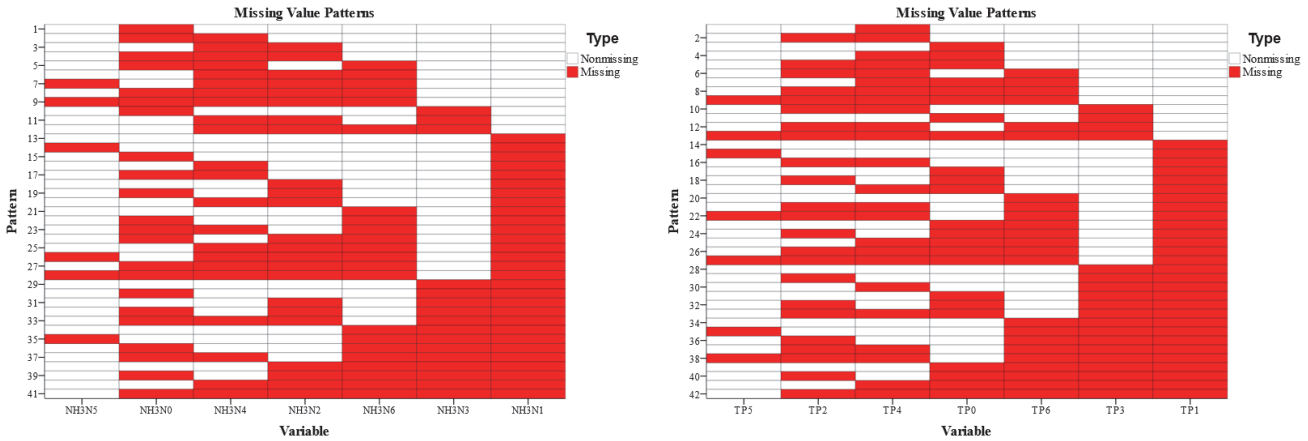


Figure 5 Distribution of 7 monitoring spots on the main stream of Mulan Creek basin in Fujian province



(a) NH3-N (b) TP

Figure 6 Analysis of the pattern of missing values of NH3-N and TP

(a) Before multiple interpolation (b) After multiple interpolation

Figure 7 NH3-N dataset comparison before and after multiple interpolation at Jinfengqiao station

(a) Before multiple interpolation (b) After multiple interpolation

Figure 8 TP dataset comparison before and after multiple interpolation at Jinfengqiao station

(a) NH3-N (b) TP

Figure 9 Jinfengqiao station dataset after bidirectional GRU prediction

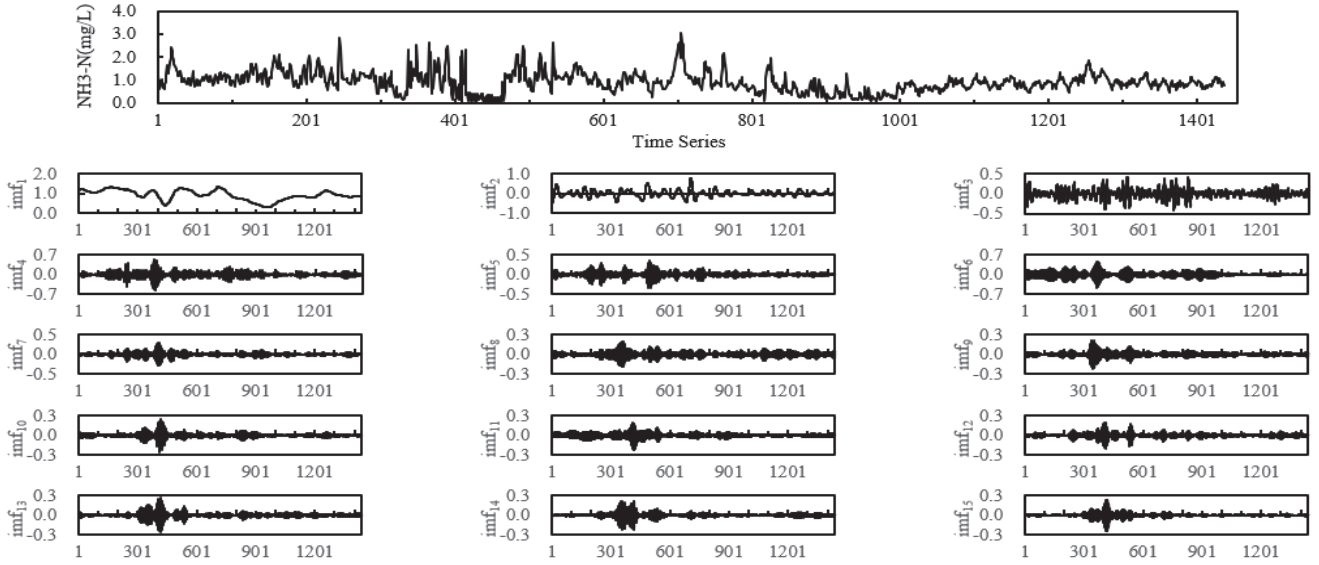


Figure 10 Modal decomposition of NH3-N dataset at Jinfengqiao station

(2) Hayati Rezvan et al. [23] pointed out that the error of multiple interpolation is larger when the missing rate is large. Due to the large missing rate of some sites in this experiment, the time series is complemented by forward and reverse-order GRU prediction after multiple interpolation. The settings of GRU: batch size was 32, hidden layer size to 128, Adam optimizer is adopted, learning rate (*lr*) to 0.002 and training epochs to 50, taking the average of bidirectional GRU prediction results to complete the corresponding missing values in the original sequence. The completion results of the NH3-N and TP dataset at Jinfengqiao station are shown in Fig. 9.

5.3 OVMD Data Enhancement

In OVMD the decomposition level K is determined with the central frequency method, and the update step (tau) is determined with the residual exponent index (REI) in Eq. (3), so the *K* value is 15 and tau is 0.57 according to the experimental dataset. In addition, the bandwidth constraint (alpha) is set to 2000, and the convergence tolerance criterion (tol) is set to 1e−7. Taking Jinfengqiao station as an example, the visualization results of the modal decomposition of NH3-N dataset is shown in Fig. 10, and the $imf_i (i = 1, 2, \ldots, 15)$ represents every mode. The raw data for NH3-N is unstable, the trend in the data is clear from the imf_1 component and the peak information is clear from the imf_2 after OVMD decomposition. The hidden information can be extracted better and the accuracy of the prediction model can be improved with the OVMD processing enhancement.

5.4 Experimental Setup

The water quality prediction model is written in the Python language under the PyTorch deep learning framework. The model is constructed with the Geometric Temporal extension library and trained on a computer with an Intel Core i5 CPU. The batch size of the water quality prediction model is set to 32 and the hidden layer size is set to 128. The adaptive moment estimation (Adam) optimizer is used for the first half of training and the stochastic gradient descent (SGD) optimizer is used for the second half. The initial value of the dynamic learning rate (*lr*) for the ReduceLROnPlateau method is 0.002 and the reduction factor is 0.9, the patience is 10 and the training number epochs is 50. The model is trained three times independently to reduce randomness, with the average of scores from the three trials as the final score. Based on mission objectives, the input characteristics of the prediction model are historical data of 6 water indicators in 7 spots to output 6 future steps data. 1435 data samples were generated with a sliding window size of 6 and step of 1, the datasets were divided into training and test sets in chronological order by a ratio of 8:2. The model parameters were optimized in the training set and the performance of the model was verified in the test set. The purpose of model training is to reduce the error between the real value and the predicted value. The loss function is used to adjust and optimize the model parameters on the training set, and the dispersion level between the prediction value and the real value is lower when the loss function value is smaller. That is the error is smaller and the prediction result is more reliable. The mean-square error (MSE) of the predicted value of all spots is set as loss function, which is shown in Eq. (11).

$$loss = \frac{1}{MN}\sum_{j=1}^{M}\sum_{i=1}^{N}(y_i^j - \widehat{y_i^j})^2 \tag{11}$$

y and $\hat{y}$ are the true and predicted values at different monitoring spots at time slice i, N is the station number, and M is the time window size. The mean absolute error (MAE), root mean square error (RMSE), and Nash-Sutcliffe efficiency coefficient (NSE) are selected as evaluating indicators of the prediction model, as shown in Eq. (12) to Eq. (14). MAE and RMSE have a range of [0, +∞), and a larger value indicates a bigger error. NSE represents data correlation and measures the ability of the forecast results to represent actual data, with a range of [0, 1] (0, 0.3] represents weak correlation, (0.3, 0.5] low correlation, (0.5, 0.8] significant correlation, (0.8, 1] high correlation, and 1 full correlation). A higher NSE value indicates a better prediction result.

$$MAE = \frac{1}{MN}\sum_{j=1}^{M}\sum_{i=1}^{N}\left|\hat{y}_i^j - y_i^j\right| \tag{12}$$

$$RMSE = \sqrt{\frac{1}{MN}\sum_{j=1}^{M}\sum_{i=1}^{N}(y_i^j - \hat{y}_i^j)^2} \tag{13}$$

$$NSE = 1 - \frac{\sum_{j=1}^{M}\sum_{i=1}^{N}(y_i^j - \hat{y}_i^j)^2}{\sum_{j=1}^{M}\sum_{i=1}^{N}(y_i^j - \bar{Y})^2} \tag{14}$$

y_i^j and $\hat{y}_i^j$ are the actual and predicted data for time window j of station i. N is the station number, M is the size of the time window and $\bar{Y}$ is the average of sample Y.

Table 5 Evaluation indicators of water quality prediction models for NH3-N and TP

Contaminants	Evaluation Indicators	4h	8h	12h	16h	20h	24h
NH3-N	RMSE	0.0819	0.0896	0.1001	0.1107	0.1191	0.1269
	MAE	0.0489	0.0530	0.0581	0.0631	0.0670	0.0707
	NSE	0.9228	0.9077	0.8846	0.8589	0.8366	0.8143
TP	RMSE	0.0254	0.0267	0.0288	0.0318	0.0341	0.0361
	MAE	0.0167	0.0175	0.0187	0.0204	0.0219	0.0232
	NSE	0.8812	0.8687	0.8475	0.8145	0.7868	0.7611

5.5 Analysis of Experimental Results

In our experiment, the NH3-N and TP dataset of all spots were predicted in six-time steps (i.e., 4, 8, 12, 16, 20, 24 hours). A comprehensive analysis of the evaluation indicators of each prediction model shows that RMSEs and MAEs in different prediction steps are relatively small, that is the prediction error is small. NSEs are more than 0.8 basically, that is the true value and the predicted value are highly correlated, indicating that the proposed water quality prediction method is good feasibility. The evaluation indicators of each prediction model are shown in Tab. 5 (the best results in bold). The experimental results show that the prediction effect of each NH3-N prediction model is better than TP, and NSEs under 4h and 8h steps are more than 0.9. Synthetically, the MAE and RMSE increase while the NSE decreases as the prediction step increases, indicating that the accuracy of the prediction model decreases with the increase of the prediction step. The prediction results and raw data of NH3-N and TP for 8, 16, 24h at each station were visualized in Fig. 11 in multiline graphs (X-axis is a time series and Y-axis is in mg/L). It is obvious that the predictions for different steps are highly consistent with the actual values, indicating that our model is effective in multistep prediction.

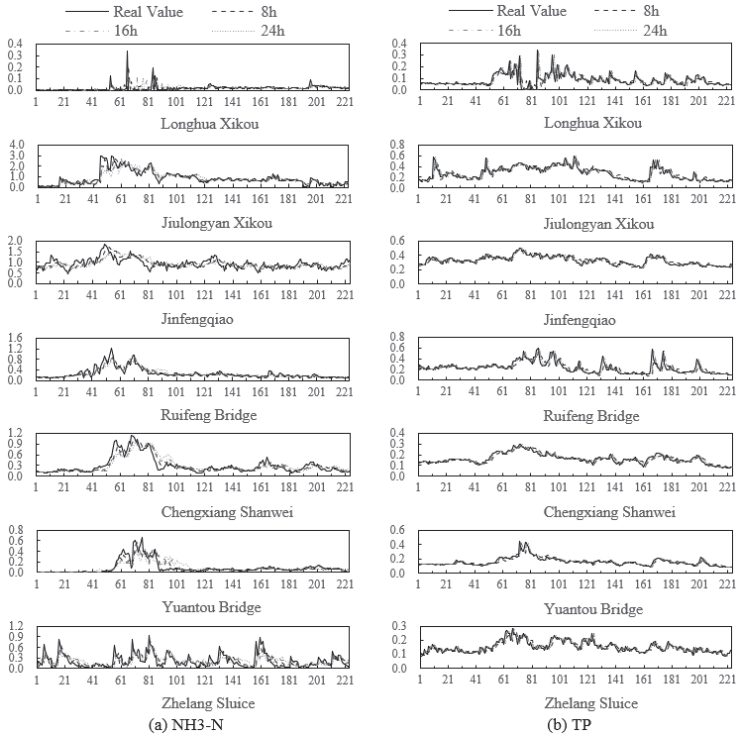


Figure 11 The prediction curve of 8, 16, 24 hours

6 MODEL VALIDATION

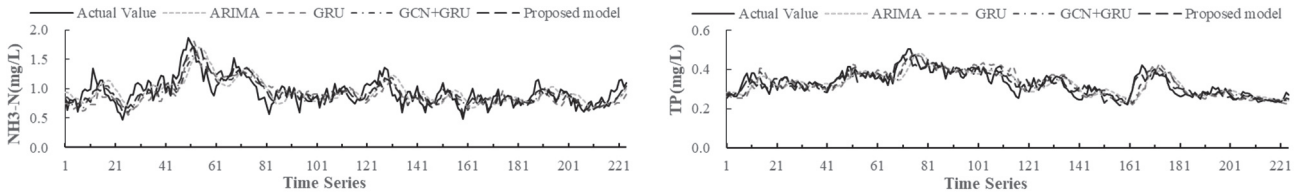
To verify the effectiveness of the proposed comprehensive water quality prediction model, we compare it with the traditional time series prediction model ARIMA, a single prediction model GRU, and a combined prediction model GCN + GRU (which includes soft attention). We select the evaluation indicators of 8, 16, and 24 h prediction steps for different models in Tab. 6 (bold data represent the best results). It can be seen from Tab. 6 that the performance of the ARIMA model is the worst among the four models. The RMSE and MAE of the GRU model are slightly lower than that of the ARIMA, and the NSE is slightly higher than that of the ARIMA, indicating that the performance of the GRU model has been improved compared to the ARIMA. The RMSE and MAE of the GCN+GRU model are lower than that of the GRU and the NSE is higher than that of the GRU, indicating that the prediction accuracy of the GCN + GRU model has been improved compared with GRU. Furthermore, compared with GCN+GRU, each evaluation index of the comprehensive prediction model has been significantly improved, which demonstrates that our model is significantly better than traditional time series models, single recurrent neural network models, and simple combination models.

Table 6 Evaluation indicators of four methods

Time step	Contaminants	Evaluation Indicators	ARIMA	GRU	GCN+GRU	Our method
8h	NH3-N	RMSE	0.1920	0.1840	0.1584	0.0896
		MAE	0.1021	0.0983	0.0830	0.0530
		NSE	0.5762	0.6107	0.7115	0.9077
	TP	RMSE	0.0517	0.0460	0.0444	0.0267
		MAE	0.0315	0.0288	0.0262	0.0175
		NSE	0.5078	0.6096	0.6372	0.8687
16h	NH3-N	RMSE	0.2122	0.2027	0.1834	0.1107
		MAE	0.1140	0.1095	0.0990	0.0631
		NSE	0.4818	0.5273	0.6129	0.8589
	TP	RMSE	0.0559	0.0504	0.0498	0.0318
		MAE	0.0348	0.0321	0.0303	0.0204
		NSE	0.4260	0.5329	0.5450	0.8145
24h	NH3-N	RMSE	0.2276	0.2166	0.2012	0.1269
		MAE	0.1229	0.1186	0.1104	0.0707
		NSE	0.4029	0.4590	0.5330	0.8143
	TP	RMSE	0.0593	0.0537	0.0535	0.0361
		MAE	0.0376	0.0347	0.0334	0.0232
		NSE	0.3569	0.4727	0.4768	0.7611

Table 7 Evaluation indicators of four methods

Comparison Model	Contaminants	RMSE Alteration	MAE Alteration	NSE Alteration
ARIMA	NH3-N	53.3% ⇩	48.1% ⇩	57.5% ⇧
	TP	48.3% ⇩	44.5% ⇩	71.0% ⇧
GRU	NH3-N	51.3% ⇩	46.1% ⇩	48.6% ⇧
	TP	42.0% ⇩	39.2% ⇩	42.5% ⇧
GCN + GRU	NH3-N	43.4% ⇩	36.1% ⇩	27.6% ⇧
	TP	39.8% ⇩	33.1% ⇩	36.3% ⇧



(a) NH3-N (b) TP

Figure 12 Comparison of prediction results of four methods

In comparison with a benchmark prediction step of 8 hours, the evaluation indicators comparison of the comprehensive prediction model with the other 3 models are shown in Tab. 7. It can be seen that the errors of the present model have decreased significantly and NSE has increased significantly compared to ARIMA, GRU, and GCN + GRU, indicating that our model has obtained better prediction results for both water quality indicators and the prediction accuracy is significantly improved. The performance improvement of the integrated prediction model compared to GCN + GRU is mainly due to the contribution of OVMD. The reason is that OVMD reduces the noise of the component series and makes the features of the component series more obvious, thus the training difficulty is reduced, the model stability is enhanced, over-fitting is reduced, and the accuracy of the prediction model is improved effectively. In addition, taking the Xianyou Jinfengqiao monitoring spot as an example, the future 8 hour prediction results of ammonia and total phosphorus for the four prediction models are visualized in Fig. 12. The curve in the figure shows that ARIMA can only fit the basic trend of the actual value and is overall seriously deviating from the actual value, so ARIMA has the lowest model fitting accuracy. GRU model has an obvious deviation from the actual value where the data fluctuates greatly, with lower model fitting accuracy. The predicted value of the GCN + GRU model is relatively close to the actual value and the accuracy of the model is improved because

GCN + GRU introduces spatial correlation, but the prediction result of the GCN + GRU model has a certain lag, indicating that the prediction accuracy is general. The proposed model can not only fit the actual value well but also make the peak part of the data well with the data enhancement effect of OVMD, which shows that our prediction model has good accuracy and feasibility.

7 CONCLUSION

A comprehensive water quality prediction model is proposed, which incorporates optimal variational mode decomposition to denoise and simplify historical water quality data, and uses GCN, GRU and attention model to make multi-step predictions based on the spatio-temporal dependencies of water quality data, and the GCN model is adjusted. Through comparison experiments with real-monitoring data, it is shown that the proposed model has good prediction ability and significant advantages compared to the baseline model, which can provide a decision-making basis for water pollution traceability or water pollution early warning. To fully verify the feasibility of this comprehensive prediction model, more similar models will be selected for comparative analysis in future work. Although the presented model has achieved a good level of prediction accuracy, the prediction ability of the peak value needs to be improved. To further improve the robustness of the comprehensive water quality prediction model, more monitoring spots should be built on tributaries besides on the main streams of Mulan River to expand the map dataset, and other graph neural networks should be adopted to better learn the spatial dependencies of different spots.

ACKNOWLEDGMENT

The work is supported by S & T Program of Hebei (Grant No.: 21370103D and 21373301D) and Research on Social Sciences Development in Hebei Province (Grant No.: 20210201445).

8 REFERENCES

[1] Obradovic, V. & Vulevic, A. (2023). Water Resources Protection and Water Management Framework in Western Balkan Countries in Drina River Basin. *Acadlore Transactions on Geosciences*, *2*(1), 24-32. https://doi.org/10.56578/atg020103

[2] Latif, S. D., Azmi, M. S. B. N., Ahmed, A. N., Fai, C. M., & El Shafie, A. (2020). Application of artificial neural network for forecasting nitrate concentration as a water quality parameter: A case study of Feitsui reservoir, Taiwan. *International Journal of Design & Nature and Ecodynamics*, *15*(5), 647-652. https://doi.org/10.18280/ijdne.150505

[3] Ding, X., Dong, X., Hou, B., Fan, G., & Zhang, X. (2021). Visual platform for water quality prediction and pre-warning of drinking water source area in the Three Gorges Reservoir Area. *Journal of Cleaner Production*, *309*, 127398. https://doi.org/10.1016/j.jclepro.2021.127398

[4] Kang, Y., Song, J., Lin, Z., Huang, L., Zhai, X., & Feng, H. (2022). Water quality prediction based on SSA-MIC-SMBO-ESN. *Computational Intelligence and Neuroscience*, 2022, 1264385. https://doi.org/10.1155/2022/1264385

[5] 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. *Environmental Science and Pollution Research*, *30*(3), 7498-7509. https://doi.org/10.1007/s11356-022-22758-7

[6] Li, W., Wu, H., Zhu, N., Jiang, Y., Tan, J., & Guo, Y. (2021). Prediction of dissolved oxygen in a fishery pond based on gated recurrent unit (GRU). *Information Processing in Agriculture*, *8*(1), 185-193. https://doi.org/10.1016/j.inpa.2020.02.002

[7] Scarselli, F., Gori, M., Tsoi, A. C., Hagenbuchner, M., & Monfardini, G. (2008). The graph neural network model. *IEEE Transactions on Neural Networks*, *20*(1), 61-80. https://doi.org/10.1109/TNN.2008.2005605

[8] Zhao, Y. M. (2020). Spatial-temporal correlation-based LSTM algorithm and its application in PM2.5 prediction. *Revue d'Intelligence Artificielle*, *34*(1), 29-38. https://doi.org/10.18280/ria.340104

[9] Omar, N. (2022). ResNet and LSTM based accurate approach for license plate detection and recognition. *Traitement du Signal*, *39*(5), 1577-1583. https://doi.org/10.18280/ts.390514

[10] Muthukumar, P., Cocom, E., Nagrecha, K., Comer, D., Burga, I., Taub, J., Calvert, C.F., Holm, J., Pourhomayoun, M. (2021). Predicting PM2. 5 atmospheric air pollution using deep learning with meteorological data and ground-based observations and remote-sensing satellite big data. *Air Quality, Atmosphere & Health*, *15*, 1221-1234. https://doi.org/10.1007/s11869-021-01126-3

[11] Tian, Y.H., Zheng, B., Li, Z.Y., Zhang, Y., & Wu, Q. (2020). Online car-hailing supply-demand forecast based on deep learning. *Ingénierie des Systèmes d'Information*, *25*(1), 21-26. https://doi.org/10.18280/isi.250103

[12] Diao, Z., Wang, X., Zhang, D., Liu, Y., Xie, K., & He, S. (2019). Dynamic spatial-temporal graph convolutional neural networks for traffic forecasting. *Proceedings of the AAAI Conference on Artificial Intelligence*, *33*(1), 890-897. https://doi.org/10.1609/aaai.v33i01.3301890

[13] Kipf, T. N. & Welling, M. (2016). Semi-supervised classification with graph convolutional networks.

[14] Cho, K., Van Merriënboer, B., Gulcehre, C., Bahdanau, D., Bougares, F., Schwenk, H., & Bengio, Y. (2014). Learning phrase representations using RNN encoder-decoder for statistical machine translation. https://doi.org/10.3115/v1/D14-1179

[15] Wang, G., Tao, T., Ma, J., Li, H., Fu, H., & Chu, Y. (2021). An improved ensemble learning method for exchange rate forecasting based on complementary effect of shallow and deep features. *Expert Systems with Applications*, 184, 115569. https://doi.org/10.1016/j.eswa.2021.115569

[16] Dragomiretskiy, K. & Zosso, D. (2014). Variational mode decomposition. *IEEE Transactions on Signal Processing*, *62*(3), 531-544. https://doi.org/10.1109/TSP.2013.2288675

[17] Wang, Y. & Markert, R. (2016). Filter bank property of variational mode decomposition and its applications. *Signal Processing*, *120*, 509-521. https://doi.org/10.1016/j.sigpro.2015.09.041

[18] Huang, N., Wu, Y., Cai, G., Zhu, H., Yu, C., Jiang, L., Zhang, Y., Zhang, J., & Xing, E. (2019). Short-term wind speed forecast with low loss of information based on feature generation of OSVD. *IEEE Access*, *7*, 81027-81046. https://doi.org/10.1109/ACCESS.2019.2922662

[19] Bruna, J., Zaremba, W., Szlam, A., & Le Cun, Y. (2013). Spectral networks and locally connected networks on graphs.

[20] Frasconi, P., Gori, M., & Sperduti, A. (1998). A general framework for adaptive processing of data structures. *IEEE transactions on Neural Networks*, *9*(5), 768-786. https://doi.org/10.1109/72.712151

[21] Su, J., Ji, D., Lin, M., Chen, Y., Sun, Y., Huo, S., Zhu, J., & Xi, B. (2017). Developing surface water quality standards in China. *Resources, Conservation and Recycling*, *117*, 294-303.

https://doi.org/10.1016/j.resconrec.2016.08.003
[22] Streiner, D. L. (2003). Starting at the beginning: an introduction to coefficient alpha and internal consistency. *Journal of Personality Assessment*, *80*(1), 99-103. https://doi.org/10.1207/S15327752JPA8001_18
[23] Hayati Rezvan, P., White, I. R., Lee, K. J., Carlin, J. B., & Simpson, J. A. (2015). Evaluation of a weighting approach for performing sensitivity analysis after multiple imputation. *BMC Medical Research Methodology*, *15*, 83. https://doi.org/10.1186/s12874-015-0074-2

Contact information:

Haitao MENG
School of Mathematics and Information Technology,
Hebei Agricultural Data Intelligent Perception and Application Technology Innovation Center,
Hebei Normal University of Science & Technology,
Qinhuangdao 066004, China

Jinling SONG
(Corresponding author)
School of Mathematics and Information Technology,
Hebei Agricultural Data Intelligent Perception and Application Technology Innovation Center,
Hebei Normal University of Science & Technology,
Qinhuangdao 066004, China
Hebei Key Laboratory of Ocean Dynamics, Resources and Environments,
Qinhuangdao 066004, China
E-mail: songjinling99@126.com

Liming HUANG
School of Business Administration,
Hebei Normal University of Science & Technology,
Qinhuangdao 066004, China

Yijin ZHU
Research Institute of International Chinese Language Education,
Beijing Language and Culture University,
Beijing 100083, China

Meining ZHU
School of Mathematics and Information Technology,
Hebei Agricultural Data Intelligent Perception and Application Technology,
Innovation Center,
Hebei Normal University of Science & Technology,
Qinhuangdao 066004, China

Jingwu ZHANG
School of Mathematics and Information Technology,
Hebei Agricultural Data Intelligent Perception and Application Technology Innovation Center,
Hebei Normal University of Science & Technology,
Qinhuangdao 066004, China