

Multi-step prediction of dissolved oxygen in rivers based on random forest missing value imputation and attention mechanism coupled with recurrent neural network

Juan Huan^{a,*}, Mingbao Li^a, Xiangen Xu^b, Hao Zhang^a, Beier Yang^a, Jiang Jianming^a and Bing Shi^a

^a School of Computer Science and Artificial Intelligence, Changzhou University, Changzhou 213164, China

^b Changzhou Institute of Environmental Sciences, Changzhou 213022, China

*Corresponding author. E-mail: huanjuan@cczu.edu.cn

ABSTRACT

Accurately predicting dissolved oxygen is of great significance to the intelligent management and control of river water quality. However, due to the interference of external factors and the irregularity of its changes, this is still a ticklish problem, especially in multi-step forecasting. This article mainly studies two issues: we first analyze the lack of water quality data and propose to use the random forest algorithm to interpolate the missing data. Then, we systematically discuss and compare water quality prediction methods based on attention-based RNN, and develop attention-based RNN into a multi-step prediction for dissolved oxygen. Finally, we applied the model to the canal in Jiangnan (China) and compared eight baseline methods. In the dissolved oxygen single-step prediction, the attention-based GRU model has better performance. Its measure indicators MAE, RMSE, and R^2 are 0.051, 0.225, and 0.958, which are better than baseline methods. Next, attention-based GRU was developed into multi-step prediction, which can predict the dissolved oxygen in the next 20 hours with high prediction accuracy. The MAE, RMSE, and R^2 are 0.253, 0.306, and 0.918. Experimental results show that attention-based GRU can achieve more accurate dissolved oxygen prediction in single-neural network and multi-step predictions.

Key words: attention mechanism, dissolved oxygen prediction, missing value interpolation, multi-step prediction, recurrent neural network, rivers

HIGHLIGHTS

- Random forest was employed to complete the missing data of dissolved oxygen monitoring.
- Recurrent Neural Network was combined with attention mechanism.
- The prediction effect of the attention-based RNN model is better than that of the traditional seasonal model.
- The model was used for multi-step prediction of dissolved oxygen, and the result was reliable.

INTRODUCTION

Dissolved oxygen (DO) plays a critical role in regulating various biogeochemical processes and biological communities in rivers (Senlin & Salim 2020). Moreover, quantification of DO is important for evaluating surface water quality because it represents the level of pollution and the state of an aquatic ecosystem (Ouma *et al.* 2020). Therefore, accurate prediction of DO has great meaning for water quality management in rivers.

In recent years, most cities in China have begun to build water environment monitoring systems. However, water quality records often have missing values for various reasons, such as due to malfunctioning of equipment, network interruptions, and natural hazards (Mital *et al.* 2020). Through the practice of a large number of scholars, it is found that the numerical simulation of water quality usually needs a complete water quality time series and other meteorological records (e.g., temperature, precipitation, wind speed) as the input of the simulation (Yazdi & Moridi 2018; Gao *et al.* 2018; Jiang *et al.* 2018). Therefore, it is necessary to reconstruct or estimate the missing values accurately and establish a complete time series to provide data support for the accuracy of the eco-hydrological model (Schneider 2001). Past efforts for imputing missing values of a precipitation time series fall under three broad categories: deletion, imputation, and non-processing method of retaining original data information (Breiman 2021). With the popularity of machine learning, more and more scholars have adopted filling methods that make full use of data, including the regression filling method, K-nearest neighbor filling method, and

This is an Open Access article distributed under the terms of the Creative Commons Attribution Licence (CC BY 4.0), which permits copying, adaptation and redistribution, provided the original work is properly cited (<http://creativecommons.org/licenses/by/4.0/>).

decision tree filling method. However, these methods are susceptible to data distribution, and the predictive interpolation performance is poor when too many missing values are encountered.

Therefore, such methods have limited applicability when it comes to reconstructing a water quality time series. To solve this dilemma, random forest (RF) proposed by Breiman (2001) is not limited to data distribution. It is an emerging comprehensive decision tree prediction method and is widely used in various fields (Qian *et al.* 2016; Wang *et al.* 2016; Wu *et al.* 2019). RF is an ensemble learning method that reduces associated bias and variance, making predictions less prone to overfitting. A recent study showed that RF-based imputation is generally robust, and performance improves with the increasing correlation between the target and references (Tang & Ishwaran 2017). Based on the above considerations, in this work, we use RF to interpolate water quality missing data from automatic monitoring stations in Jiangnan, China.

The process of predicting water quality over a catchment is complex, nonlinear, and exhibits both temporal and spatial variability. The models developed to simulate the process can be categorized as empirical, conceptual, and physically-based distributed models, such as QUAL2 K, EFDC, WASP6 (Samuela & Christina 2010; Wu & Xu 2011; Zhang *et al.* 2012). These models require vast information and data on various hydrological subprocesses to arrive at the results, which create boundary conditions as limitations (Zhu & Heddam 2020). In recent years, methods such as time series analysis and machine learning have been proposed and developed rapidly (Solgi *et al.* 2017; Khazaee *et al.* 2019). However, they can easily fall into local minima and present issues with stability and reliability. Because of these problems, this paper presents a Recurrent Neural Network (RNN) prediction model based on the attention mechanism. According to the latest research, attention-based RNN can dynamically learn spatial-temporal relationships and then achieve the best results in single-step prediction of multivariate time series (Qin *et al.* 2017) and short-term prediction of sensory time series (Yu *et al.* 2018). Inspired by their works, we analyze the feature representation of water quality relationships in attention-based RNN. Furthermore, we developed state of the art attention-based RNN methods into a multi-step prediction for dissolved oxygen prediction. Finally, we applied this method in the actual data set and compared it with eight baseline models to illustrate the effectiveness of the model.

MATERIALS AND METHODS

Study area and data

Jiangnan (119°08'–120°12'E, 31°09'–32°04'N) is located in the Taihu Lake Plain, Yangtze River Delta, in the southeastern part of Jiangsu Province, China. It has a subtropical monsoon climate, with an annual average temperature of 16.5 °C and annual precipitation of 1063.71 mm (Gao *et al.* 2017).

In this study, the Canal River section in the Yangtze River Delta of Jiangnan (China) was selected as the experimental area. The Canal River section was equipped with an automatic water quality monitoring station, which collected data online every 4 hours. Station monitors multiple indicators, such as ammonia nitrogen ($\text{mg}\cdot\text{L}^{-1}$), total phosphorus ($\text{mg}\cdot\text{L}^{-1}$), CODMn ($\text{mg}\cdot\text{L}^{-1}$), total nitrogen ($\text{mg}\cdot\text{L}^{-1}$), pH, water temperature, dissolved oxygen ($\text{mg}\cdot\text{L}^{-1}$), conductivity ($\text{us}\cdot\text{cm}^{-1}$) and turbidity (NTU). In this paper, we collected 2190 data for a total of 365 days from January 1, 2019 to January 1, 2020. The first 1752 data were selected as the training set, and the last 438 data were used as the test set.

Random forecast

Random forest (RF) is an ensemble learning methodology and like other ensemble learning techniques, the performance of a number of weak learners (which could be a single decision tree, single perceptron, etc.) is boosted via a voting scheme (Ahmad *et al.* 2017).

An RF generates C number of decision trees from N number of training samples. For each tree in the forest, bootstrap sampling is performed to create a new training set, and the samples which are not selected are known as out-of-bag (OOB) samples (Jiang *et al.* 2009). This new training set is then used to fully grow a decision tree without pruning by using CART methodology. In each split of the node of a decision tree, only a small number of M features (input variables) are randomly selected instead of all of them (this is known as random feature selection). This process is repeated to create M decision trees in order to form a randomly generated forest. The output of all these randomly generated trees is aggregated to obtain one final prediction value which is the average value of all trees in the forest.

Attention mechanism

The attention mechanism in deep learning is similar to the human selective visual attention mechanism, which uses limited attention to select more critical information from numerous input features. Luong *et al.* (2015) divided the attention process into three stages, calculating weight, normalization, and cumulative summation. The structure is shown in Figure 1.

X_t is the input vector of attention in Figure 1, f is the function to calculate the distribution of attention, R_t is the weight of attention intermediate output, ∂ is the weight of standardized attention intermediate output, C is the hidden information of the previous moment, and Y is the output of attention. It is worth noting that the output of any model can be used as an input to the attention.

Firstly, the output weights R_t between the input vectors of attention and the intermediate candidate layers are calculated by Equation (1). Then, the weight matrix R_t is normalized by Equation (2). Finally, the normalized weight ∂_t multiplied into X_t and summed. The final output state of attention is represented by Equation (3):

$$R_t = f(C, X_t) \tag{1}$$

$$\partial_t = softmax(R_t) \tag{2}$$

$$Y = \sum \partial_t * x_t \tag{3}$$

Gated recurrent unit

The RNN network based on the encode-decoder structure has become one of the popular methods to solve the time-series prediction problem (Cho *et al.* 2014). The Gated Recurrent Unit (GRU) is a kind of RNN, with a typical three-layer structure: input layer, hidden layer, and output layer. Since the GRU structure can maintain long-term dependence through a linear flow of information in the cell mechanism and gate mechanism, it is used to encode the raw series into feature representation in the encoder stage, and it is also used to decode the encoded feature vector in the decoder stage. Therefore, we introduce the data mapping process of the common GRU model that is used in this study.

GRU contains two gates, which are reset gates and update gates. The unit structure of the GRU is shown in Figure 2. In the figure, $R_{(t)}$ is the reset probability distribution, $Z_{(t)}$ is the update probability distribution, σ is the sigmoid function, $\tilde{H}(t - 1)$ is

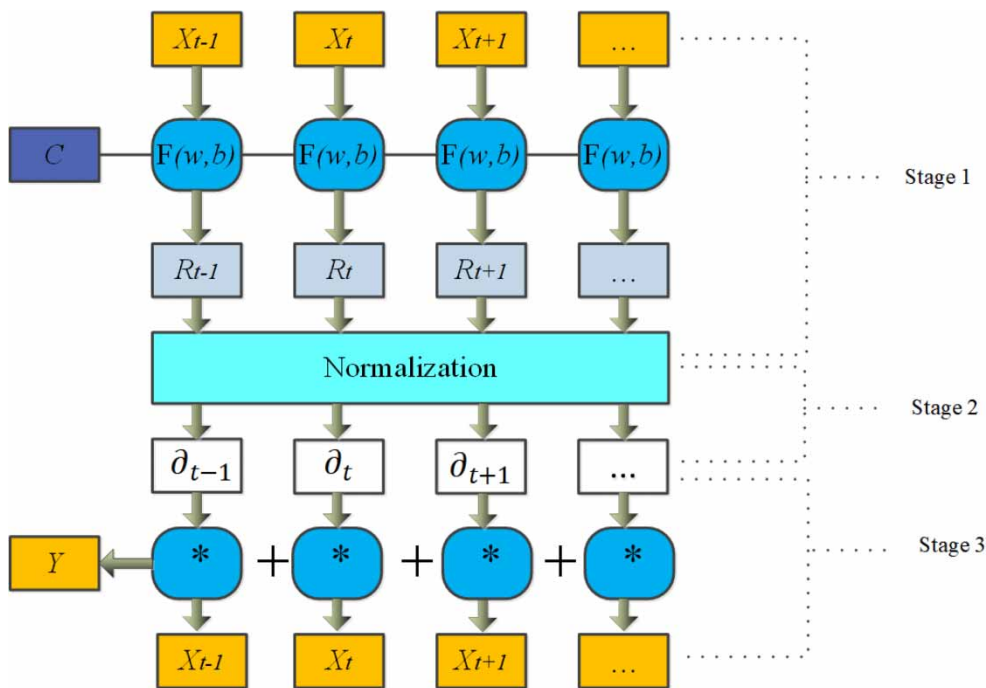


Figure 1 | Attention mechanism structure flow chart.

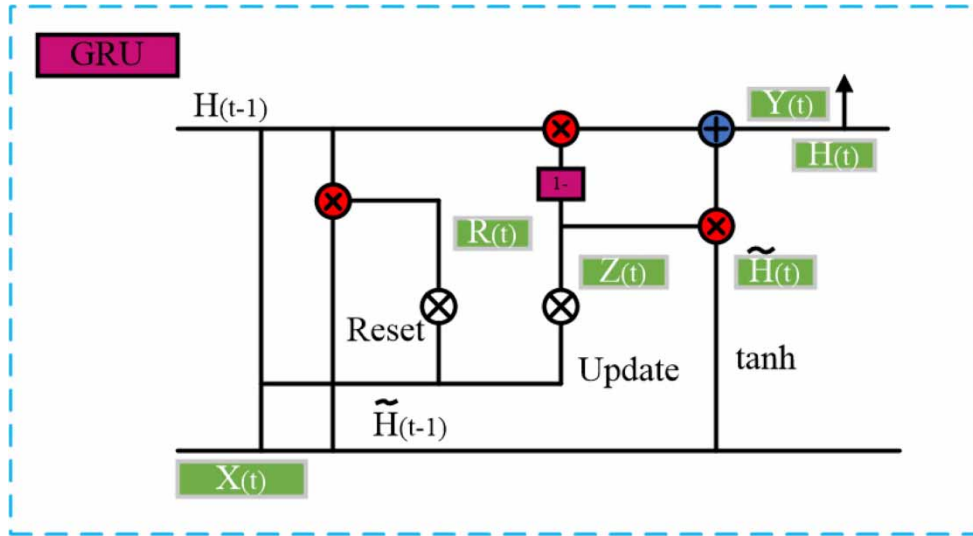


Figure 2 | GRU unit structure.

the potential state of the cell at the previous moment, $\tilde{H}(t)$ is the potential state of the cell at the current moment, and $H(t)$ is the output state of the cell at the current moment. The reset probability distribution $R_{(t)}$ and the updated probability distribution $Z_{(t)}$ can be calculated by Equations (4) and (5) respectively:

$$R_{(t)} = \sigma(W * [X_{(t)}, H_{(t-1)}] + b) \quad (4)$$

$$Z_{(t)} = \sigma(W * [X_{(t)}, H_{(t-1)}] + b) \quad (5)$$

The unit potential state of the cell $\tilde{H}(t-1)$ at the previous moment can be obtained by multiplying the input state after resetting and the unit state at the previous moment, as shown in Equation (6). The potential state $\tilde{H}(t)$ of the unit at the current time is the result of the input vector passing through the \tanh layer, as shown in Equation (7). The final output of the GRU unit is determined by the unit state $H_{(t-1)}$ at the previous moment, the potential state of the unit at the current moment, and the updated probability distribution. The final output state $H_{(t)}$ can be obtained by Equation (8):

$$\tilde{H}_{(t-1)} = H_{(t-1)} * R_{(t)} \quad (6)$$

$$\tilde{H}_{(t)} = \tanh(W * [X_{(t)}, \tilde{H}_{(t-1)}] + b) \quad (7)$$

$$H_{(t)} = H_{(t-1)} * (1 - Z_{(t)}) + Z_{(t)} * \tilde{H}_{(t)} \quad (8)$$

MODEL AND EXPERIMENT OF RF-ATTENTION-BASED GRU

Experimental set-up

In this paper, a multi-step prediction model of dissolved oxygen based on missing value completion and attention-based GRU is proposed. The scheme flow chart is shown in Figure 3.

The model can be divided into three parts: data processing, data set partition and, attention-based GRU model training. Data processing includes outlier filtering and missing value interpolation. Then the data set was divided into training set and prediction set according to the ratio of 8:2. In the final step, establishing attention-based GRU model and training. The processing is discussed below.

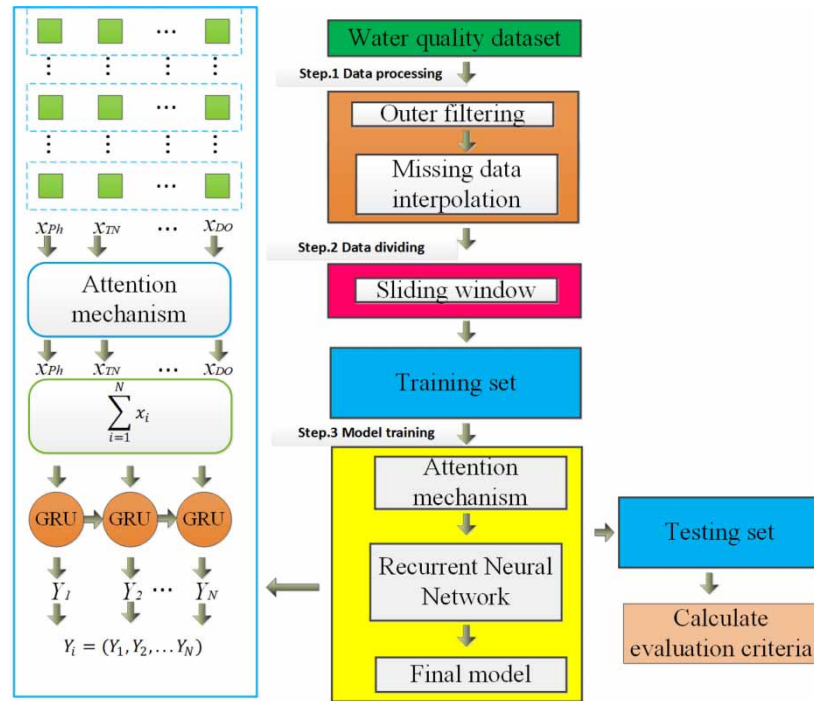


Figure 3 | Attention-based RNN dissolved oxygen multi-step prediction flow chart.

Step 1. Data processing

In this step, we use the density-based local outer factor (LOF) algorithm to filter outliers from water quality data (Lu et al. 2021). Then, the abnormal values are directly shaved and treated as missing values. Finally, the random forest algorithm is used to complete the missing value of water quality data.

Step 2. Data dividing

The current multi-step prediction methods are mostly single-step cycle iterative prediction. This method inputs the predicted value as the measured value into the model for re-predict to achieve the effect of multi-step prediction (Cho & Park 2019). To overcome the problem of single-step iterative prediction error superposition, this paper uses a time window sliding algorithm to effectively avoid the problem of error superposition. The sliding algorithm of the time window is shown in Figure 4, where X is a specific component, M is input width and N is forecast width. In this paper, according to the demand of water quality forecast in Jiangnan, the input sample window width M is set to $4*14$, and the prediction target window width N is set to 1, 2, 3, 4, and 5. Finally, this paper conducts 1-, 2-, 3-, 4-, and 5-step prediction experiments.

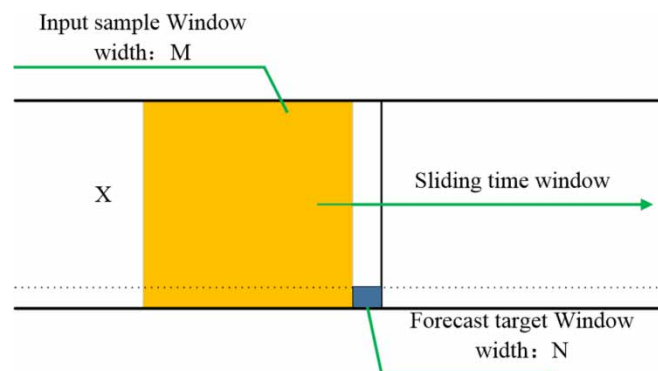


Figure 4 | Time window sliding algorithm.

Step 3. Attention-based GRU model training

In this paper, we use one attention layer, two GRU network layers (GRU), two fully connected layers (dense), one dropout layer (dropout), and one output layer (output) to build an attention-based GRU model, and then input data for training. Super parameter settings of the GRU model are shown in Table 1.

Evaluation measures

Appropriate evaluation criteria must be selected to perform a reasonable model. In water quality prediction, one of the most popular evaluation measures is RMSE, which is defined as the deviation between monitoring value and prediction value. Using RMSE can not be considered the sole criterion because it can not describe the overall performance of the model in the test set. Therefore, mean absolute error (MAE) and R^2 are also employed as performance measures. The calculation of each indicator is shown in Equations (9)–(11):

$$MAE = \frac{1}{m} \sum_{i=1}^m |Y_i - y_i| \quad (9)$$

$$RMSE = \sqrt{\frac{1}{m} \sum_{i=1}^m (Y_i - y_i)^2} \quad (10)$$

$$R^2 = 1 - \frac{\sum_{i=1}^m (Y_i - y_i)^2}{\sum_{i=1}^m (Y_i - \bar{Y})^2} \quad (11)$$

where m is the number of samples, Y_i is the predicted value, and y_i is the actual value of samples. The smaller the three evaluation indicators of RMSE and MAE, the smaller the error between the predicted value and the observed value, and the higher the prediction accuracy of the model. R^2 closer to 1 = the higher the overall prediction accuracy of the model on the test set.

In addition, to further evaluate the reliability of attention-based RNN models in the multi-step prediction of dissolved oxygen, we also use the probability diagram of probability integral transformation (PIT) value to assess the reliability of the model. PIT is calculated as follows:

$$PIT = F(y_i) = \int_{-\infty}^{y_i} p(Y_i) d(Y_i) \quad (12)$$

where $F(y_i)$ is the cumulative distribution function and Y_i is the forecast value. If the predictions are reliable, the PIT values obey uniform distribution between 0 and 1. PIT values for all test samples displayed in a uniform probability plot can clearly check whether it is subject to uniform distribution (Zhang *et al.* 2019).

RESULTS AND DISCUSSION

Interpolation of dissolved oxygen deletion value

In the process of monitoring dissolved oxygen in the river, due to the influence of some external factors, there was a phenomenon of missing data, which seriously affected the quality of the data used in the model. In this paper, RF is used to complete the missing values of dissolved oxygen. The interpolation result is shown in Figure 5.

Table 1 | Hyper-parameters of GRU

Predict step	GRU1	GRU2	Dense1	Dense2	Dropout	Learning	Optimizer
1	90	23	73	53	0.1745	0.0069	Adam
2	102	71	74	94	0.1807	0.0071	Adam
3	107	23	87	62	0.1745	0.0069	Adam
4	101	26	83	63	0.1818	0.0056	Adam
5	133	134	134	114	0.3560	0.0030	Adam

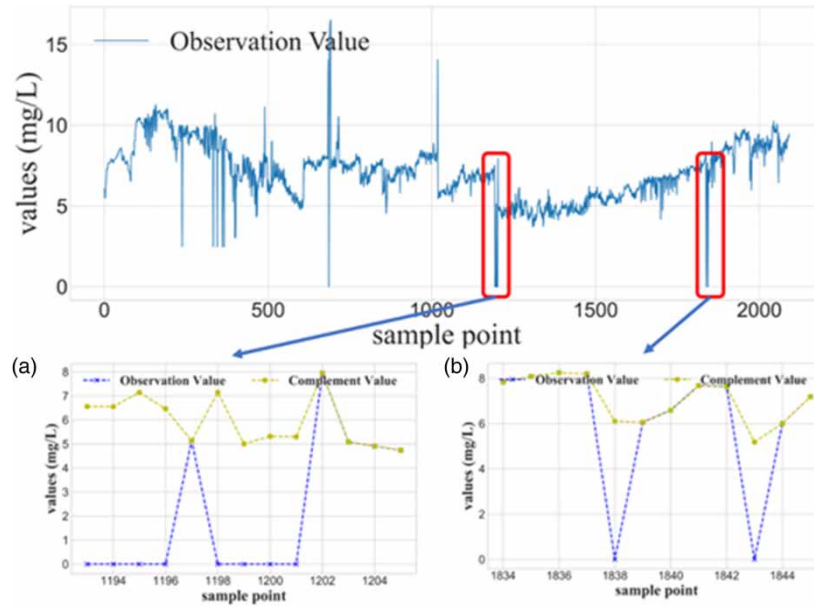


Figure 5 | Missing value interpolation of dissolved oxygen. *Note:* The missing value in the original data is replaced by 0 value.

Figure 5 shows the DO sample with a total of 2190 data points, including 10 missing data. Figure 5(a) shows the completion of missing values of dissolved oxygen sample points 1193–1205, and Figure 5(b) shows the completion of missing values of dissolved oxygen sample points 1834–1845. Figure 5(a) has many continuous missing values, which belong to continuous missing cases, Figure 5(b) has two missing values and the missing values are discontinuous, which belongs to discrete missing cases (Lin & Cai 2020). Combined with Figure 5(a) and 5(b), it can be found that the dissolved oxygen complement value in continuous loss and discrete loss conforms to its future development trend. Therefore, we have reason to believe that RF can effectively complete the dissolved oxygen deficiency data.

Attention-based RNN models comparison

In this section, we compare the effect of traditional RNN methods and attention-based RNN methods in single-step prediction of dissolved oxygen. To show our proposed methods more intuitively and clearly, we show their visualization results in Figure 6.

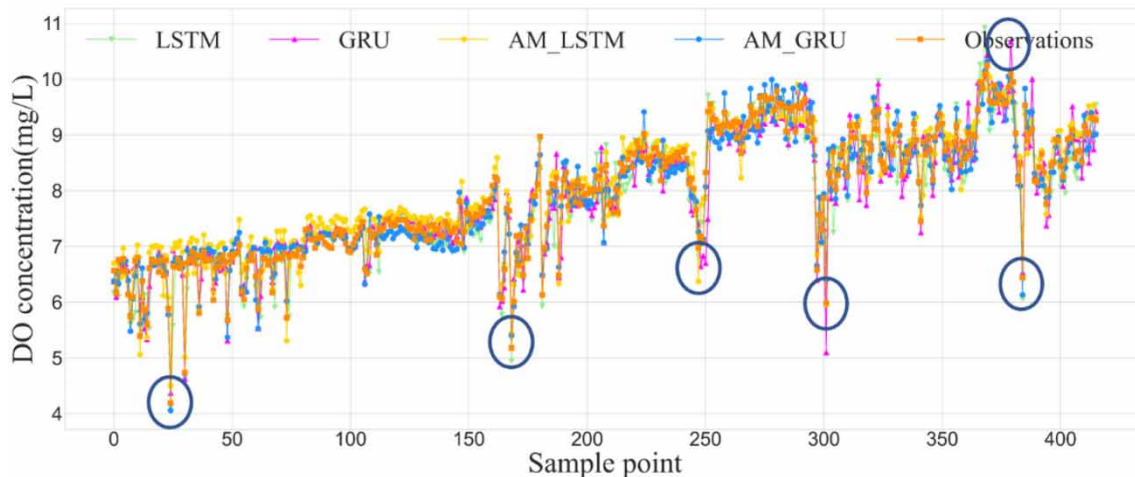


Figure 6 | DO prediction result of RNN models and attention-based RNN models.

In Figure 6, the green curve represents LSTM prediction results, the magenta curve represents GRU prediction results, the gold curve represents attention-based LSTM prediction results, the blue curve represents attention-based GRU prediction results, and the orange curve represents DO test samples. Comparing the prediction curves of attention-based LSTM and LSTM, we find that the LSTM prediction curve fits the DO test samples poorly. Comparing the prediction curves of attention-based GRU and GRU, we find that the prediction curve of attention-based GRU is closer to the DO test sample. It shows that the attention mechanism reduces a load of network units by assigning weights, and optimizes the prediction results of LSTM and GRU networks. This is reflected in the extreme values of DO samples (blue circle in Figure 6), such as 35, 168, 250, etc. At these data points, the deviation between the predicted value and the sample value of the attention-based GRU model is smaller.

Concerning evaluation measures, Table 2 shows the details of four models. Compared with the RNN model, attention-based RNN has smaller prediction errors. In addition, among the four models, the attention-based GRU model has the smallest prediction error, and the LSTM has the largest. Further comparison shows the R^2 of the attention-based GRU model is 0.037 larger than that of the GRU model, which shows that the attention-based GRU model outperforms better than the GRU model on the test set. At the same time, the RMSE and MAE of the attention-based GRU model are 0.084 and 0.044 smaller than those of the GRU model, respectively.

In face to face nonlinear and non-stationary long-time series, the LSTM and GRU model will have problems such as gradient disappearance or gradient explosion, which leads to a large deviation between the predicted value and the actual value when facing the extreme in the test set (Qin *et al.* 2017). The attention-based GRU model adds an attention layer, which greatly optimizes the GRU network structure, reduces the load of network units, and gives special weight to the extreme, to improve its importance in the model.

In Table 2 we compare attention-based RNN models. We know that the attention-based GRU model has a smaller error than the attention-based LSTM model. The R^2 of attention-based LSTM is 0.025 smaller than the attention-based GRU model. The MAE of attention-based LSTM is 0.03 higher than the attention-based GRU model and the RMSE of attention-based LSTM is 0.059 higher than the attention-based GRU model. This means that compared with the attention-based LSTM, the predicted DO value of the attention-based GRU model is closer to the DO observed value in the test set. Compared with the attention-based LSTM model, attention-based GRU has a simpler network unit structure and fewer parameters, which enables the attention-based GRU model to focus on the extreme during training. This may be one of the reasons why the R^2 , RMSE, and MAE of the attention-based GRU model are better than that of the attention-based LSTM model in the test set. Based on the above analysis, we found that the attention-based model has the best predictive effect, which is consistent with the research conclusions of Liu *et al.* (2019).

Comparison of baseline models

To evaluate the feasibility of the proposed DO prediction model, five models are selected for comparison: Adaptive Network-based Fuzzy Inference System (ANFIS), Artificial Neural Network (ANN), Extreme Learning Machine (ELM), Radial Basis Function (RBF-ANN), Support Vector Machines (SVM). In this paper, the hyperparameter of SVM and RBF-ANN are optimized by Grid Search (Li *et al.* 2010; Fayed & Atiya 2019). The prediction results of each model are shown in Figure 7.

As shown in Figure 7, the abscissa is the DO test value and the ordinate is the predicted value, which constitutes the data points in the figure. The more dense the data points are, the closer the test value is to the predicted value and the higher the fitting degree is with the curve in the figure, thus the better the prediction performance of the model is. The points in the ANFIS, RBF-ANN and ELM models are in a discrete state and have a low degree of fitting with the equation, among

Table 2 | Comparison of prediction performance of RNN and attention-based RNN models

Model	RMSE	MAE	R^2
LSTM	0.334	0.112	0.907
GRU	0.309	0.095	0.921
Attention-based LSTM	0.284	0.081	0.933
Attention-based GRU	0.225	0.051	0.958

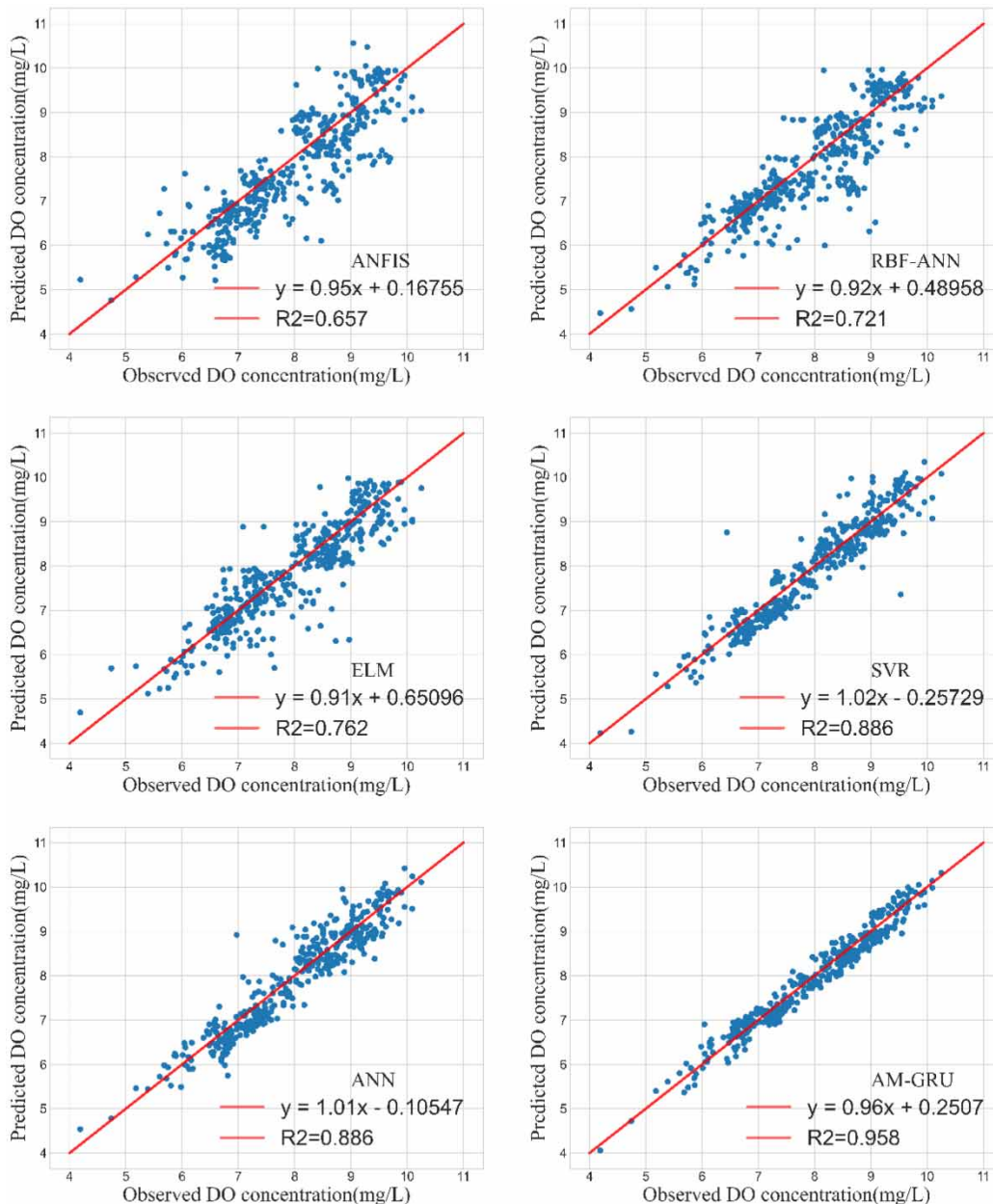


Figure 7 | DO prediction result of baseline models.

which ANFIS has the highest degree of dispersion. The data points in SVR and ANN models are concentrated on both sides of the equation, which has a high degree of fitting with the equation. Among the benchmark models, the ANFIS model has the worst prediction accuracy, while the ANN model and SVR model have the best prediction performance. By comparing the ANN model with the attention-based GRU model, we find that the data points in the attention-based GRU model are denser, which indicates that the predicted value of the attention-based GRU model is closer to the test value, and the prediction performance is better.

Regarding evaluation measures, Table 3 shows the details of five models. Compared with the five models, the neural network model has the best prediction, with a prediction accuracy of 0.886, which is 0.229 higher than that of ANFIS. To reflect the optimization of the proposed model, we choose the ANN model to compare with the attention-based GRU model, because ANN is the best performer among the five benchmark models. The R² of the attention-based GRU model is 0.072 larger than that of the ANN model. The MAE of the ANN model is 0.088 larger than that of the attention-based GRU

Table 3 | Comparison of five baseline models prediction performance

Model	RMSE	MAE	R ²
ANFIS	0.645	0.416	0.657
RBF-ANN	0.583	0.339	0.721
ELM	0.537	0.289	0.762
SVR	0.372	0.138	0.886
ANN	0.373	0.139	0.886
Attention-based GRU	0.225	0.051	0.958

model and the RMSE of the ANN model is 0.148 higher than that of the attention-based GRU model. It shows that the attention-based GRU model outperforms better than the ANN model on the test set.

During training, the ANN model only considers the relationship between data units at the current moment and ignores the historical characteristics of water quality monitoring data (Dhussa *et al.* 2014; Wang *et al.* 2017). As a result, when ANN meets the extreme, its predicted value deviates greatly from the actual value, resulting in a large error of ANN in the test set. Summarizing the above analysis, compared with the five baseline models, the attention-based GRU model is more suitable for dissolved oxygen prediction.

Multi-step DO prediction employ attention-based GRU model

After analyzing Figures 6 and 7 and Tables 2 and 3 in the text, we conclude that the attention-based GRU model performs best in the single-step prediction of dissolved oxygen. Therefore, to further investigate the applicability of the attention-based GRU model, we use it to conduct multi-step dissolved oxygen prediction research. The result is shown in Figure 8.

As can be seen from Figure 8, there is little difference between most observed values and predicted values. However, with the increase of prediction step size, outliers appear in the prediction results and deviate from the error line in the figure. In terms of evaluation indicators, Table 4 gives detailed information on this phenomenon.

In the prediction step 2, the R² of attention-based GRU is 0.95, which was up to 0.032 higher than the prediction step 4. As can be seen, the RMSE and MSE of prediction step 3 were up to 38%, 30% higher than the predicted step 2. Combined with Figure 9, with the superposition of prediction steps, the dispersion between the predicted value and the actual value of the attention-based GRU model begins to increase, which is the reason for the increase of the prediction error index of the attention-based GRU model. The superposition of prediction steps leads to an increase in the input sample area and the total consumption of the model, which reduces the accuracy of dissolved oxygen prediction (Majid *et al.* 2021).

Through the analysis of Table 4 and Figure 8, we find that the prediction accuracy of the attention-based GRU model is the lowest when the time step is 4, but according to the research results of scholars, its error is still within an acceptable range (Ji *et al.* 2017; Kisi *et al.* 2020; Lidija *et al.* 2020). The above analysis is based on the micro prediction results and to make the experimental results more convincing, we calculated the PIT values of the test results and the prediction results. The reliability of the experimental results is visualized by drawing the uniform probability diagram of the PIT value, as shown in Figure 9.

The PIT values of the four prediction steps are evenly distributed around the diagonal and its range evenly covers [0, 1]. All PIT points are located in the Kolmogorov 5% significance band, which indicates that predicted probability distribution functions (PDF) are not excessively high or low, or excessively wide or narrow (Ruder 2016). According to the PIT reliability results, we can say that the DO multi-step prediction of the attention-based GRU model is reliable and persuasive. Combined with Figures 8 and 9, we analyze and discuss the multi-step prediction results of dissolved oxygen from micro and macro perspectives respectively, and conclude that the attention-based GRU model has achieved ideal results in the multi-step prediction of river dissolved oxygen and can accurately predict the future dissolved oxygen.

CONCLUSIONS

Accurate prediction of dissolved oxygen has always been a challenge for researchers. Although many machine learning and deep learning models have been applied to the prediction of dissolved oxygen, the accuracy of these models is usually limited to a short advance step and ignores the quality of the original data. This paper first interpolates the missing dissolved oxygen

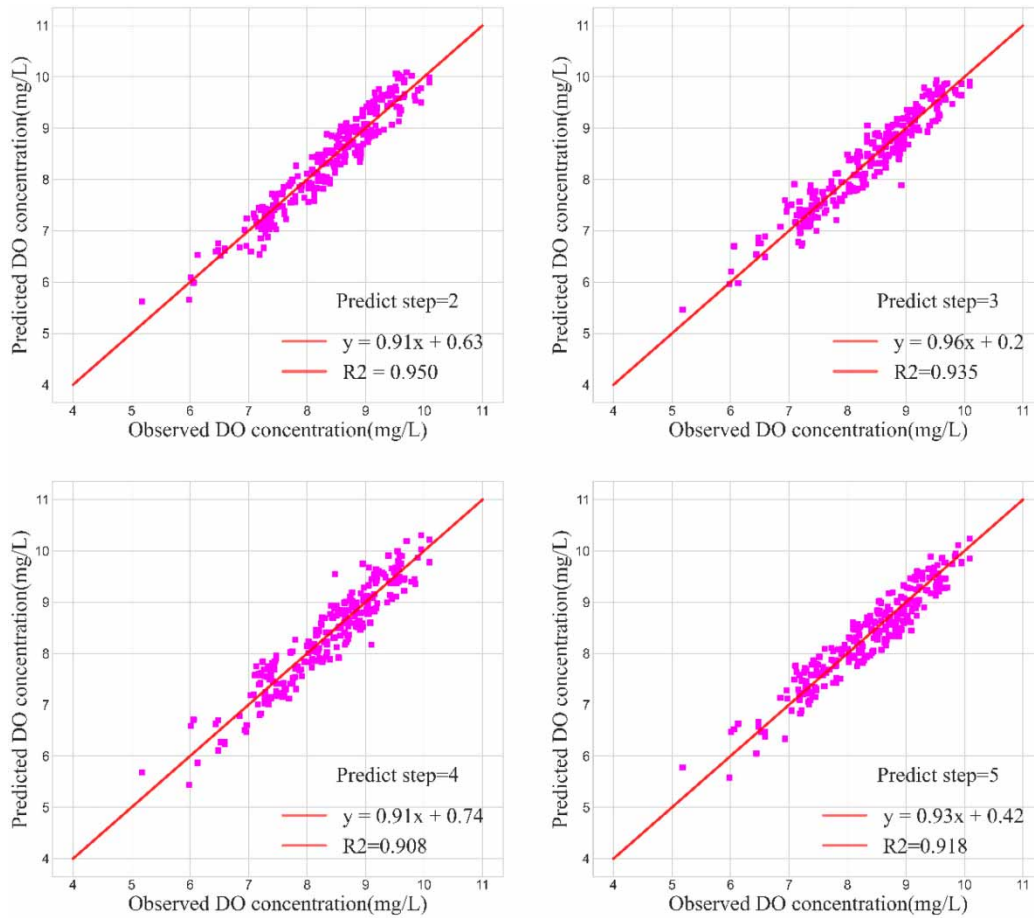


Figure 8 | Multi-step DO prediction results of attention-based GRU.

Table 4 | Multi-step DO prediction performance of attention-based GRU model

Predict step	RMSE	MAE	R ²
2	0.230	0.180	0.950
3	0.279	0.218	0.935
4	0.317	0.234	0.908
5	0.306	0.253	0.918

data by using random forest and then studies the effectiveness of the attention-based GRU method in single- and multi-step prediction of dissolved oxygen.

The RF algorithm can compensate for the loss of dissolved oxygen monitoring data, which is conducive to the construction of a high-quality water quality monitoring data set and improves the prediction accuracy of the model. The attention-based GRU showed good performance in the single-step prediction experiment of river dissolved oxygen, such as low RMSE, MAE and high R². With the superposition of the predicted step size, the dispersion between the predicted value of the attention-based GRU model and the actual value of dissolved oxygen begins to increase, but the prediction error is still within an acceptable range. The PIT analysis of the prediction results of dissolved oxygen based on the attention-based GRU model shows that the PIT values are all located in Kolmogorov 5% significant band, which indicates that the attention-based GRU model is effective in the prediction of dissolved oxygen in the river.

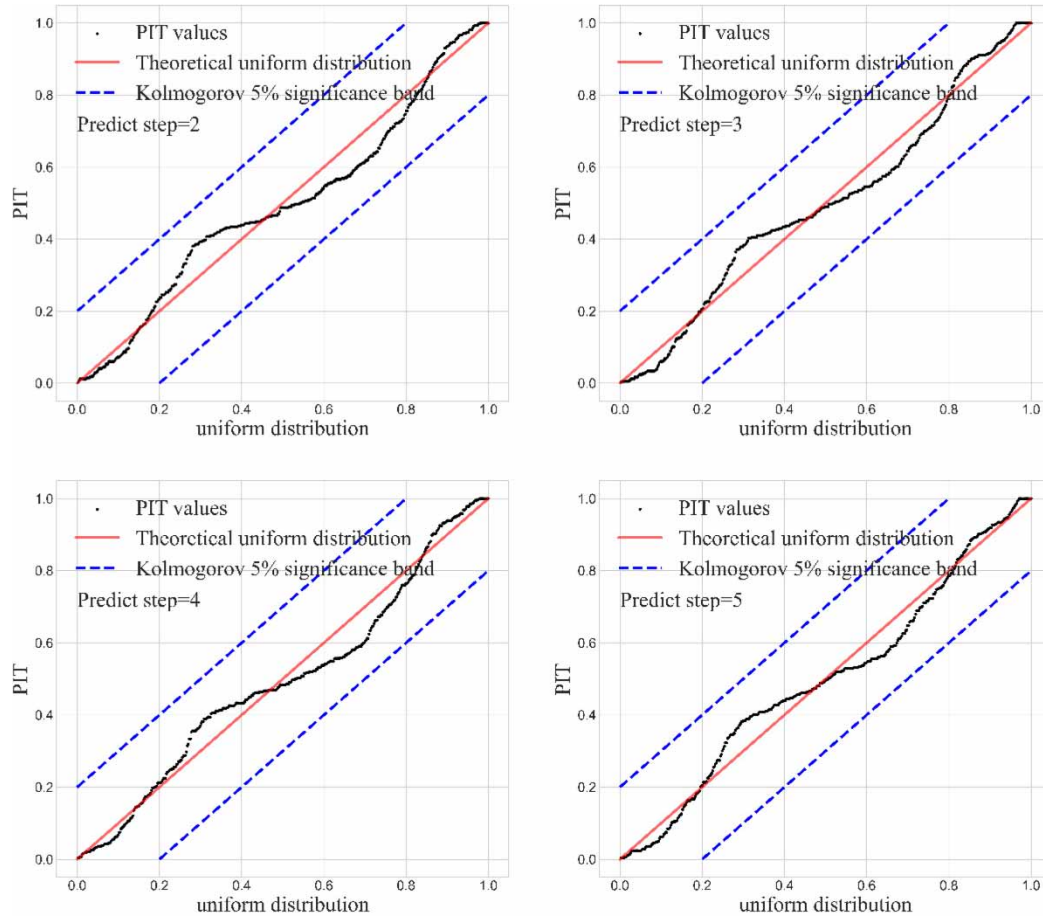


Figure 9 | Reliability test of attention-based GRU in multi-step DO prediction.

The application of the attention-based GRU model can predict dissolved oxygen in the next 20 hours and provide scientific references for river water quality management. Although the overall prediction accuracy of the established attention-based GRU model is relatively high, the prediction model does not consider the impact of rainfall on dissolved oxygen. Enriching research data and taking full account of rainfall conditions are the directions and priorities for further research.

ACKNOWLEDGEMENTS

The authors would like to acknowledge the financial support from the National Natural Science Foundation of China (61803050), the Changzhou Science and Technology Program (CE20205037) and Postgraduate Research & Practice Innovation Program of Jiangsu Province.

DATA AVAILABILITY STATEMENT

All relevant data are included in the paper or its Supplementary Information.

REFERENCES

- Ahmad, M. W., Mourshed, M. & Rezgui, Y. 2017 *Trees vs neurons: comparison between random forest and ANN for high-resolution prediction of building energy consumption*. *Energy and Buildings* **147**, 77–89.
- Breiman, L. 2001 *Random forests*. *Machine Learning* **45**, 5–32.
- Cho, H. & Park, H. 2019 Merged-LSTM and multistep prediction of daily chlorophyll-a concentration for algal bloom forecast. In: *IOP Conference Series: Earth and Environmental Science*, Vol. 351. IOP, Kaohsiung, Taiwan, China, p. 012020.

- Cho, K., Merriënboer, B. V., Gulcehre, C., Bahdanau, D., Bougares, F., Schwenk, H. & Bengio, Y. 2014 Learning phrase representations using RNN encoder-decoder for statistical machine translation. In: *Computer Science: Computation and Language*. Ithaca, USA, pp. 1508.04025. pp. 1724–1734.
- Dhussa, A. K., Sambhi, S. S., Kumar, S. S. & Kumar, S. 2014 Nonlinear autoregressive exogenous modeling of a large anaerobic digester producing biogas from cattle waste. *Bioresource Technology* **170**, 342–349.
- Fayed, H. A. & Atiya, A. F. 2019 Speed up grid-search for parameter selection of support vector machines. *Applied Soft Computing* **80**, 202–210.
- Gao, Y., Ma, L. & Liu, J. 2017 Constructing ecological networks based on habitat quality assessment: a case study of Changzhou, China. *Scientific Reports* **7**, 46073.
- Gao, Q., He, G. J., Fang, H. W., Bai, S. & Huang, L. 2018 Numerical simulation of water age and its potential effects on the water quality in Xiangxi Bay of Three Gorges Reservoir. *Journal of Hydrology* **566**, 484–499.
- Ji, X. L., Dahlgren, R. A. & Zhang, M. H. 2017 Prediction of dissolved oxygen concentration in hypoxic river systems using support vector machine: a case study of Wen-Rui Tang River, China. *Environmental Science and Pollution Research* **24**, 16062–16076.
- Jiang, R., Tang, W. W., Wu, X. B. & Fu, W. H. 2009 A random forest approach to the detection of epistatic interactions in case-control studies. *BMC Bioinformatics* **10** (1), 1–12.
- Jiang, L., Li, Y. P., Zhao, X., Tillotson, M. R., Wang, W. C., Zhang, S. S., Sarpong, L. D., Asmaa, O. & Pan, B. Z. 2018 Parameter uncertainty and sensitivity analysis of water quality model in Lake Taihu, China. *Ecological Modelling* **375**, 1–12.
- Khazaei, P. A., Shourian, M. & Ebrahimi, H. 2019 A comparative study of MLR, KNN, ANN and ANFIS models with wavelet transform in monthly stream flow prediction. *Water Resources Management* **33**, 2907–2923.
- Kisi, O., Alizamir, M. & Gorgij, A. D. 2020 Dissolved oxygen prediction using a new ensemble method. *Environmental Science and Pollution Research* **27**, 9589–9603.
- Li, C. H., Lin, C. T., Kuo, B. C. & Chu, H. S. 2010 An automatic method for selecting the parameter of the RBF kernel function to support vector machines. In: *International Conference on Technologies and Applications of Artificial Intelligence*. IEE, Hsinchu, Taiwan, China, pp. 226–232.
- Lidija, J., Sanja, M. K. & Vladanka, P. U. 2020 Prediction of nitrate concentration in Danube River water by using artificial neural networks. *Water Supply* **20** (6), 2119–2132.
- Lin, W. C. & Cai, Z. F. 2020 Missing value imputation: a review and analysis of the literature (2006–2017). *Artificial Intelligence Review* **53**, 1487–1509.
- Liu, Y. Q., Zhang, Q., Song, L. H. & Chen, Y. Y. 2019 Attention-based recurrent neural networks for accurate short-term and long-term dissolved oxygen prediction. *Computers and Electronics in Agriculture* **165**, 104964.
- Lu, S. W., Wu, X. L., Zheng, J., He, Z., Gu, J. & Han, H. G. 2021 Data cleaning method of urban sewage treatment process based on dynamic fusion lof. *Control and Decision* **37** (5), 1231–1240. (in Chinese).
- Luong, M. T., Pham, H. & Manning, C. D. 2015 Effective approaches to attention-based neural machine translation. In: *Computer Science: Computation and Language*. Ithaca, USA, pp. 1508.04025.
- Majid, G. S., Mouna, D. & Sophie, D. 2021 Short-term water demand predictions coupling an artificial neural network model and a genetic algorithm. *Water Supply* **21** (5), 2374–2386.
- Mital, U., Dwivedi, D., Brown, J. B., Faybisenko, B., Painter, S. L. & Steefel Carl, I. 2020 Sequential imputation of missing spatio-temporal precipitation data using random forests. *Frontiers in Water* **2020**, 2–20.
- Ouma, Y. O., Okuku, C. O., Njau, E. N. & Meštrović, A. 2020 Use of artificial neural networks and multiple linear regression model for the prediction of dissolved oxygen in rivers: case study of hydrographic basin of River Nyando, Kenya. *Complexity* **23**, 9570789.
- Qian, C., Chen, J. X., Luo, Y. B. & Dai, L. 2016 An interpolation method for missing data in high way tunnel operation based on random forest. *Transportation System Engineering and Information* **16** (3), 81–87 (in Chinese).
- Qin, Y., Song, D. J., Cheng, H. F., Wei, C. & Cottrell, G. 2017 A dual-stage attention-based recurrent neural network for time series prediction. In *International Joint Conference on Artificial Intelligence (IJCAI): Machine Learning*.
- Ruder, S. 2016 An overview of gradient descent optimization algorithms. In: *Computer Science: Machine Learning*. Ithaca, USA, pp. 1609.04747.
- Samuela, F. & Christina, W. T. 2010 Assessment of uncertainty sources in water quality modeling in the Niagara River. *Advances in Water Resources* **33** (4), 493–503.
- Schneider, T. 2001 Analysis of incomplete climate data: estimation of mean values and covariance matrices and imputation of missing values. *Journal of Climate* **14**, 853–871.
- Senlin, Z. & Salim, H. 2020 Prediction of dissolved oxygen in urban rivers at the Three Gorges Reservoir, China: extreme learning machines (ELM) versus artificial neural network (ANN). *Water Quality Research Journal* **55** (1), 106–118.
- Solgi, A., Pourhaghi, A., Bahmani, R. & Zarei, H. 2017 Improving SVR and ANFIS performance using wavelet transform and PCA algorithm for modeling and predicting biochemical oxygen demand (BOD). *Ecohydrology & Hydrobiology* **17** (2), 164–175.
- Tang, F. & Ishwaran, H. 2017 Random forest missing data algorithms. *Statistical Analysis and Data Mining* **10**, 363–377.
- Wang, Z., Ren, H. & Fang, Y. P. 2016 Application of random forest in carrier big data completion. *Telecommunications Science* **32** (12), 7–12.
- Wang, Y. Y., Zhou, J., Chen, K. J., Wang, Y. Y. & Liu, L. F. 2017 Water quality prediction method based on LSTM neural network. In: *12th International Conference on Intelligent Systems and Knowledge Engineering (ISKE)*. Cheng du, China, pp. 1–5

- Wu, G. Z. & Xu, Z. X. 2011 Prediction of algal blooming using EFDC model: case study in the Daoxiang Lake. *Ecological Modeling* **24**, 1245–1252.
- Wu, Y., Zhang, J. F., Fan, C. L. & Hu, W. D. 2019 Missing data interpolation of ship collision accident based on random forest. *Journal of Wuhan University of Technology* **43** (6), 1120–1124. (In Chinese).
- Yazdi, J. & Moridi, A. 2018 Interactive reservoir-watershed modeling framework for integrated water quality management. *Water Resources Management* **31** (7), 2105–2125.
- Yu, X. L., Song, Y. K., Jun, B. Z., Xiu, W. Y. & Yu, Z. 2018 GeoMAN: multi-level attention networks for geosensory time series prediction. In: *Twenty-Seventh International Joint Conference on Artificial Intelligence*. IJCAI, London, UK, pp. 3428–3434. pp. 3428–3434.
- Zhang, R. B., Xin, Q., Li, H. M., Yuan, X. C. & Rui, Y. 2012 Selection of optimal river water quality improvement programs using QUAL2 K: a case study of Taihu Lake Basin, China. *Science of the Total Environment* **431**, 278–285.
- Zhang, Z. D., Lei, Y., Qin, H., Liu, Y. Q., Wang, Z., Yu, X., Yin, X. L. & Li, J. 2019 Wind speed prediction method using shared weight long short-term memory network and Gaussian process regression. *Applied Energy* **247**, 270–284.
- Zhu, S. & Heddam, S. 2020 Prediction of dissolved oxygen in urban rivers at the Three Gorges Reservoir, China: extreme learning machines (ELM) versus artificial neural network (ANN). *Water Quality Research Journal* **55** (1), 106–118.

First received 7 September 2021; accepted in revised form 15 March 2022. Available online 31 March 2022