

Article

A Water Quality Prediction Model Based on Modal Decomposition and Hybrid Deep Learning Models

Shuo Zhao ¹, Ruru Liu ¹, Yahui Liu ¹, Tao Zeng ¹, Chunpeng Chen ¹ and Liping Xu ^{2,*}

¹ School of Information Science and Technology, Shihezi University, Shihezi 832003, China; zs0536@stu.shzu.edu.cn (S.Z.); liururu@stu.shzu.edu.cn (R.L.); lyh@shzu.edu.cn (Y.L.); 20222508002@stu.shzu.edu.cn (T.Z.); chenchunpeng@stu.shzu.edu.cn (C.C.)

² School of Sciences, Shihezi University, Shihezi 832003, China

* Correspondence: xlpalw@163.com

Abstract: When the total nitrogen content in water sources exceeds the standard, it can promote the rapid proliferation of algae and other plankton, leading to eutrophication of the water body and also causing damage to the ecological environment of the water source area. Therefore, making timely and accurate predictions of water quality at the source is of vital importance. Since water quality data exhibit non-stationary characteristics, predicting them is quite challenging. This study proposes a novel hybrid deep learning model based on modal decomposition, ERSCB (EMD-RBMO-SVMD-CNN-BiGRU), to enhance the accuracy of water quality forecasting. The model first employs Empirical Mode Decomposition (EMD) technology to decompose the original water quality data. Subsequently, it quantifies the complexity of the subsequences obtained from EMD using Sample Entropy (SE) and further decomposes the most complex subsequences using Sequential Variational Mode Decomposition (SVMD). To address the matter of selecting balanced parameters in SVMD, this study introduces the Red-Billed Blue Magpie Optimization (RBMO) algorithm to optimize the hyperparameters of SVMD. On this basis, a forecasting model is constructed by integrating Convolutional Neural Networks (CNN) and Bidirectional Gated Recurrent Unit (BiGRU) networks. The experimental results show that, compared to existing water quality prediction models, the ERSCB model has an improved prediction accuracy of 4.0% and 3.1% for the KaShi River and GongNaiSi River areas, respectively.



Academic Editor: Craig Allan

Received: 17 December 2024

Revised: 4 January 2025

Accepted: 6 January 2025

Published: 10 January 2025

Citation: Zhao, S.; Liu, R.; Liu, Y.; Zeng, T.; Chen, C.; Xu, L. A Water Quality Prediction Model Based on Modal Decomposition and Hybrid Deep Learning Models. *Water* **2025**, *17*, 184. <https://doi.org/10.3390/w17020184>

Copyright: © 2025 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

Keywords: prediction of water quality concentration; modal decomposition; Red-billed Blue Magpie Optimization (RBMO); Bidirectional Gated Recurrent Unit (BiGRU); Successive Variational Modal Decomposition (SVMD); Convolutional Neural Network (CNN)

1. Introduction

Water resources are vital to the Earth's ecosystem, serving as the foundation for the survival of all life. Stable water quality in water source areas not only ensures the survival of living organisms but also maintains the balance of the local ecology, providing humans with essential clean water for daily life and agriculture irrigation. Currently, more than half of the global population faces annual water scarcity, with some regions experiencing severe water shortages. When the total nitrogen content in water bodies exceeds the standard, it can lead to the overgrowth of phytoplankton and trigger eutrophication. This condition poses a serious threat to the ecological environment of water sources. Therefore, timely and accurate prediction of total nitrogen in water source quality, along with the implementation of appropriate improvement measures, is of paramount importance for

ensuring water resource safety, maintaining ecological balance, safeguarding public welfare, and promoting sustainable development.

The prediction of water quality is to determine in advance the value of a specific water quality index in a certain period of time by using prediction models and data analysis methods [1]. At present, most of the water quality prediction models are based on deep learning, and artificial neural network models project the input information to the output domain through nonlinear transformation, and these models can deal with complex nonlinear data. However, their dependence on a large number of data, the difficulty of adjusting model parameters and the opacity of prediction results limit their application in small-scale sample data prediction. Water quality is also affected by many external factors such as climate, landform and human activities, and the interaction between these factors is more complex [2]. There is also a process of mutual conversion between various substances and organisms in the water body environment, which increases the dynamic changes in the water body, causing the water quality data to show unstable characteristics, thus making water quality prediction more difficult.

In light of this, this study aims to use modal decomposition algorithms combined with deep learning technology to address the instability of water quality data and develop an accurate water quality prediction model to predict the total nitrogen index of water sources. This plays a critical role in enhancing the efficiency of water environment management in water source basins, ensuring water resource security, improving water treatment technology, leading pollution prevention strategies, and promoting scientific research and technological innovation.

Water quality prediction is a time series prediction problem using historical series data to infer the trend of future value changes. Currently, water quality prediction models primarily consist of mechanism models, regression analysis models, traditional machine learning, and deep learning models. Mechanism prediction models rely on the physical, chemical, and biological characteristics of the aquatic environment to establish the mass balance equation of pollutants in water. By simulating the migration, transformation, and decay of pollutants in water, water quality prediction can be realized. Commonly used mechanical models include QUAL [3], WASP [4], SWAT [5] and others.

Due to the large data demands of mechanism model construction, the complexity is higher and more environmental and chemical knowledge is required. In contrast, regression analysis models do not rely on complex expertise, and the modeling process is relatively simple. The statistical models of water quality prediction mainly include the autoregressive moving average model (ARIMA) [6], seasonal autoregressive moving average model (SARIMA) [7] and others. Regression analysis models face challenges in handling complex nonlinear features of time series, and incorporating machine learning models can more effectively tackle these issues. Common machine learning models include boosted regression trees (BRT), random forest (RF), and support vector machines (SVM). Alnahit et al. use RF and BRT models for water quality prediction in selected watersheds [8]. Their results show that the RF algorithm is relatively easier to train and has better robustness to overfitting the model. Adil et al. used water quality data from the UCI Machine Learning Repository covering the years 2000–2021 for the Southern Bug River in Ukraine. They applied ten machine learning models, including Artificial Neural Networks (ANN), Support Vector Regression (SVR), Extreme Learning Machine (ELM), Decision Tree Regressor (DTR), RF, AdaBoost (AB), Gradient Boosting Regressor (GBR), XGBoost Regressor (XGBR), Gaussian Process (GP), and K-Nearest Neighbors (KNN) to estimate the Water Quality Index (WQI). The results indicated that the Gaussian Process (GP) model performed best, followed by the XGBoost Regressor (XGBR), while ANN and AB showed relatively weaker performance [9].

With the expanding scale and complexity of data, machine learning cannot automatically extract data features, and the ability to learn long-term dependencies is limited, so deep learning models have been widely used. The models that use deep learning for water quality sequence prediction mainly include Convolutional Neural Networks (CNN), Long Short Term Memory Networks (LSTM), and Gated Recurrent Units (GRU). Wang et al. used LSTM to predict water quality time series data. Their experimental results show that the LSTM model has significant advantages in the processing of time-dependent water quality data, and its prediction performance is significantly better than that of traditional machine learning methods [9]. In addition, Setshedi et al. used multi-layer perceptron (MLP) neural networks and radial basis function (RBF) neural networks to develop optimal models for predicting water quality parameters [10]. Their results show that MLP has the best prediction ability in water quality prediction. Rustam et al. used an artificial neural network (ANN) composed of a hidden layer and multiple exclusion and activation layers, and tested it on two datasets, achieving good water quality prediction results [11].

In order to combine the advantages of different models, some scholars have begun to study mixed models to further reduce the error of water quality prediction. Baek et al. combined CNN and LSTM to predict water quality [12]. Their results show that the CNN and LSTM models perform well in simulating water quality parameters, such as total nitrogen, total phosphorus and total organic carbon, and can accurately predict the trend of water quality. Chen et al. combined the attention mechanism with the LSTM model to achieve more accurate water quality time series prediction, demonstrating the broad application prospects of deep learning in water quality prediction [13].

Since the non-stationary characteristics of water quality data reduce the predictive performance of models, many scholars propose to use the signal decomposition method to reduce a model's stationarity and enhance its predictive precision. Wang et al. proposed a short-term water quality prediction model based on variation mode decomposition (VMD) and improved Grasshopper optimization algorithm (IGOA), which optimized LSTM and showed good performance in short-term prediction [14]. Empirical mode decomposition (EMD) effectively captures local characteristics and nonlinear oscillations of signals. Zhang et al. used EMD to deal with nonlinear and non-stationarity problems in time series data when predicting water quality in urban drainage networks [15]. The integrated EMD-LSTM model achieved the highest R^2 value and the lowest root mean square error (RMSE) in the prediction of chemical oxygen demand (COD), total phosphorus (TP), total nitrogen (TN), ammonia nitrogen ($\text{NH}_3\text{-N}$), and biochemical oxygen demand (BOD), which was a significant improvement compared with traditional preprocessing and data-driven models. Therefore, in this study, EMD was used to decompose the original TN concentration sequence, and Sample Entropy (SE) was employed to assess the complexity of each subsequence. Sample Entropy is a statistical measure used to assess the complexity of time series. It calculates the probability of data point patterns repeating at a specified length and tolerance level. Sample Entropy measures the likelihood that similar patterns remain similar when the pattern length is increased by one data point. It effectively quantifies the regularity and complexity of time series, which is very useful for analyzing and comparing the characteristics of different subsequences.

In order to increase the accuracy of prediction, some scholars use quadratic decomposition technology to extract the features of data at a deeper level. Tan et al. used the secondary decomposition structure to predict the water quality of Taihu Lake in China [16]. VMD technology was used to decompose the data, and the characteristic term and self-term were respectively processed. The self-term was further optimized by secondary mode decomposition to predict the dissolved oxygen content in the lake. Their results show that, compared with other single models and mixed models, the proposed WOA-CVMD-

CBiLSTM-AT-F model can effectively improve the accuracy of dissolved oxygen prediction. Although the VMD decomposition method has advantages in mathematical stability and local extreme point problem processing, in the decomposition process, it is more sensitive to the initial value of the center frequency of the mode. If the number of decompositions is too large, there will be excessive decomposition phenomena. Successive variational mode decomposition (SVMD) [17] does not need to set the number of modes K in advance, which greatly reduces the computational complexity. Compared with the VMD algorithm, SVMD is more robust to the initial value of the center frequency of the mode. However, the balance parameter $\max\alpha$ in SVMD will affect its performance, and manual tuning is cumbersome. Because of its strong global search capability, the RBMO algorithm is used in this study to optimize SVMD parameters.

The innovative contributions of this study are highlighted as follows:

- (1) Considering the challenges posed by the non-stationary nature of water quality sequences, this paper employs the EMD (Empirical Mode Decomposition) algorithm to address this issue. The EMD's capability to break down complex data into simpler Intrinsic Mode Functions (IMFs) significantly reduces non-stationarity, thereby enhancing the predictive power of the model.
- (2) This paper introduces a two-tiered decomposition strategy in which the EMD (Empirical Mode Decomposition) is first used to perform an initial decomposition, selecting the subsequence with the highest complexity. This most complex subsequence is then further decomposed using SVMD (Singular Value Mode Decomposition). This innovative approach further reduces the non-stationarity of the data. To overcome the arbitrariness and suboptimality of manually selecting SVMD balance factors, this paper utilizes an RBMO (Random Binary Matrix Optimization) algorithm. The RBMO algorithm automates the optimization of SVMD parameters, ensuring a more accurate and adaptive decomposition process.
- (3) This study introduces a pioneering hybrid model for water quality prediction, known as the ERSCB model. To demonstrate the model's robustness and versatility, we conducted experiments using data from two distinct monitoring stations in the Xinjiang region, namely the KaShi River and the GongNaiSi River. The selection of these stations, which exhibit a variety of environmental conditions, allows us to showcase the model's effectiveness under different geographical and hydrological settings.

2. Dataset Description and Method Introduction

2.1. Dataset

The dataset for this study was selected from the National Surface Water Quality Automatic Monitoring real-time data release system (<https://szddjc.cnemc.cn:8070/GJZ/Business/Publish/Main.html>, accessed on 15 September 2024), covering water quality data from various monitoring stations nationwide. The data is managed by the China National Environmental Monitoring Center and uses the software CNEMC (Version 2.5.1). The sampling equipment is the ET-8000D surface water sampler (Changzhou Mont Instrument Manufacturing Co., Ltd., Changzhou, China), and the analytical equipment is an online water quality analyzer (AppliTek, Belgium). To demonstrate the advantages and robustness of the proposed model, the data of the KaShi River and GongNaiSi River monitoring stations in Ili Kazak Autonomous Prefecture of Xinjiang were selected for the experiment.

Data from 12,814 sample points at two monitoring stations between 30 April 2020 and 27 August 2024 were selected as the dataset, and the sample points were divided into a training set and a test set in a ratio of 8:2. The training set consisted of 10,251 samples, while the test set comprised 2563 samples. Tables 1 and 2 display the statistical characteristics of the training set and test set data from these two monitoring stations, respectively.

Table 1. The statistical characteristics of the training set data.

Observatory	Feature Name	Mean	Min	Max
KaShi	TN	7.32	2.92	14.97
	TP	0.54	0.17	1.19
	DO	5.89	1.32	9.82
	NH ₃	3.75	1.12	6.94
	Conductivity	374.63	143.25	596.20
GongNaiSi	TN	4.58	2.26	10.01
	TP	0.15	0.12	0.37
	DO	5.67	0.73	12.38
	NH ₃	1.32	0.81	2.53
	Conductivity	238.43	130.43	337.23

Table 2. The statistical characteristics of the test set data.

Observatory	Feature Name	Mean	Min	Max
KaShi	TN	7.33	3.02	15.01
	TP	0.49	0.14	1.08
	DO	5.73	1.14	9.73
	NH ₃	3.84	1.08	7.24
	Conductivity	391.34	147.38	573.26
GongNaiSi	TN	5.32	2.34	9.73
	TP	0.14	0.08	0.41
	DO	5.76	0.67	12.97
	NH ₃	1.37	0.73	2.48
	Conductivity	218.43	137.26	327.54

2.2. Empirical Mode Decomposition Method

EMD is an innovative method to deal with non-stationary time series signals efficiently. The core of EMD is to refine the complex water quality sequence $X(t)$ into a series of intrinsic mode functions (IMF) with unique physical significance [18]. This process is realized by gradually stripping the features of different time scales in the original signal, and each IMF component carries the local fluctuation characteristics of the original water quality sequence in a specific time range and the inherent structural complexity of the dynamic changes in water quality. The EMD decomposition steps are as follows:

- (1) When processing the time series x_t , we first identify and extract the local maximum and minimum points, and then use the cubic spline interpolation technique to construct two smooth curves based on these extreme points: One is the upper envelope $I_{\max}(t)$, which closely fits above all local maximum points. The other is the lower envelope line $I_{\min}(t)$, which is closely below all local minimum points;
- (2) We calculate the average envelope line $m_{(t)}$ of the upper and lower envelope lines

$$m_{(t)} = \frac{I_{\min}(t) + I_{\max}(t)}{2} \quad (1)$$

- (3) We calculate the new sequence $h_1(t)$.

$$h_1(t) = x(t) - m(t) \quad (2)$$

- (4) We repeat step (3) to screen $h_1(t)$ until the decomposed signal is stable.

- (5) Finally, it is decomposed into k IMFs and a residual component $r_1(k)$.

$$\begin{cases} r_2(t) = r_1(t) - h_2(t) \\ r_3(t) = r_2(t) - h_3(t) \\ r_k(t) = r_{k-1}(t) - h_k(t) \end{cases} \quad (3)$$

- (6) The original signal is reconstructed as follows:

$$x(t) = \sum_{i=1}^k h_i(t) + r_k(t) \quad (4)$$

2.3. Successive Variational Mode Decomposition Optimized by the Red-Billed Blue Magpie Optimization Algorithm

2.3.1. Red-Billed Blue Magpie Optimization

RBMO is an optimization algorithm that emulates the collective behavior of the red-billed blue magpie, simulating its foraging and social behaviors in nature [19]. In RBMO, the algorithm structure is divided into three layers, namely, L layer (Leader), F layer (Follower), and S layer (Scout). As the leader of the whole group, the L-layer red-billed magpie has the best food source information currently found, that is, the optimal solution to the optimization problem. The F layer follows the L layer closely, searching carefully around it in the hope of finding a better solution. The S-layer red-billed blue Finch is responsible for exploring a wider area to discover new potential high-quality solutions. RBMO operates iteratively, with the L, F, and S tiers updating their positions according to their specific behavioral rules in each generation. The L-layer leads the whole group in a better direction, the F-layer searches locally around the L-layer, and the S-layer explores globally through random or heuristic methods. This layered cooperation mechanism enables RBMO to maintain its exploration ability and also to have efficient development ability, so as to effectively approach the optimal solution of the optimization problem.

2.3.2. Successive Variational Mode Decomposition

Successive Variational Mode Decomposition (SVMD), as an advanced tool, has been widely used in the analysis and decomposition of complex signals. It is designed to simplify the structure of non-stationary and nonlinear signals by extracting multiple Intrinsic Mode Functions (IMFs), thereby improving the accuracy of subsequent processing or prediction [20]. However, a key challenge of the VMD method is how to properly determine the number of modes (K value) obtained by decomposition, which directly affects the decomposition effect and the accuracy of subsequent analysis. In order to solve this problem, this paper adopts SVMD decomposition without pre-setting the specific number of modal decompositions, but gradually builds and optimizes the decomposition results through the iterative process until the energy of the residual signal is below a certain threshold or reaches a predetermined number of iterations. The solving process of the algorithm is as follows.

- (1) For the time series data $f(t)$, this article decomposes it into two parts: the L-order mode and the residual signal:

$$f(t) = u_L(t) + f_r(t) \quad (5)$$

In the equation, the residual signal represents the remaining portion of the data $f(t)$ that has not been fully covered by the L previously extracted modal components, denoted as $f_r(t)$. This part of the signal is the difference between the original data $f(t)$ and the sum of all determined modal components $\sum_{i=1}^{L-1} u_i(t)$.

- (2) The constraint conditions for minimizing the L-order mode are as follows.

$$J_1 = \left\| \partial_t \left[\left(\delta(t) + \frac{j}{\pi t} \right) * u_L(t) \right] e^{-j\omega_L t} \right\|_2^2 \quad (6)$$

In the formula, the center frequency of the L-order mode is denoted as ω_L .

- (3) When the effective component of $u_L(t)$ exists, the residual signal $f_r(t)$ should be minimized. To achieve stable constraints, it is necessary to select a suitable filter with a frequency response:

$$\hat{\beta}_L(\omega) = \frac{1}{\alpha(\omega - \omega_L)^2} \quad (7)$$

The constraint conditions are as follows:

$$J_2 = \|\beta_L(t) * f_r(t)\|_2^2 \quad (8)$$

- (4) When the above two constraints are not sufficient to distinguish between L-order and L-1 modes, the idea of the J_2 constraint can be used to further identify these two modes by selecting filters with specific frequency responses:

$$\hat{\beta}_i(\omega) = \frac{1}{\alpha(\omega - \omega_L)^2}, \quad i = 1, 2, \dots, L-1 \quad (9)$$

The constraint conditions are as follows:

$$J_3 = \sum_{i=1}^{L-1} \|\beta_i(t) * u_L(t)\|_2^2 \quad (10)$$

The final constraint ensures the complete reconstruction of the signal, expressed as follows:

$$f(t) = u_L(t) + \sum_{i=1}^{L-1} u_i(t) + f_u(t) \quad (11)$$

Therefore, modal component extraction can be regarded as a minimization problem under the three constraints mentioned above:

$$\begin{cases} \min\{aJ_1 + J_2 + J_3\} \\ u_L(t) + f_r(t) = f(t) \end{cases} \quad (12)$$

In the above equation, α represents the equilibrium of J_1 , J_2 , and J_3 .

2.3.3. RBMO-SVMD

In the application of the SVMD algorithm, it is necessary to specify the balance parameter number α artificially, which has large subjectivity and blindness. Therefore, RBMO is used to optimize maxAlpha, the equilibrium parameter of SVMD, and the fitness function adopts EE (Envelope Entropy). Envelope Entropy is a statistical measure that describes the complexity of a signal, used to measure the entropy of a symbol sequence, which is the weighted sum of the entropies of the individual symbols within the sequence. In the process of optimizing SVMD parameters, Envelope Entropy is used as the fitness function; the smaller the Envelope Entropy value, the better the decomposition effect, implying that the signal decomposition is more precise and has lower complexity. It is then optimized and updated to determine the final optimal parameters. The flowchart of RBMO-SVMD is shown in Figure 1.

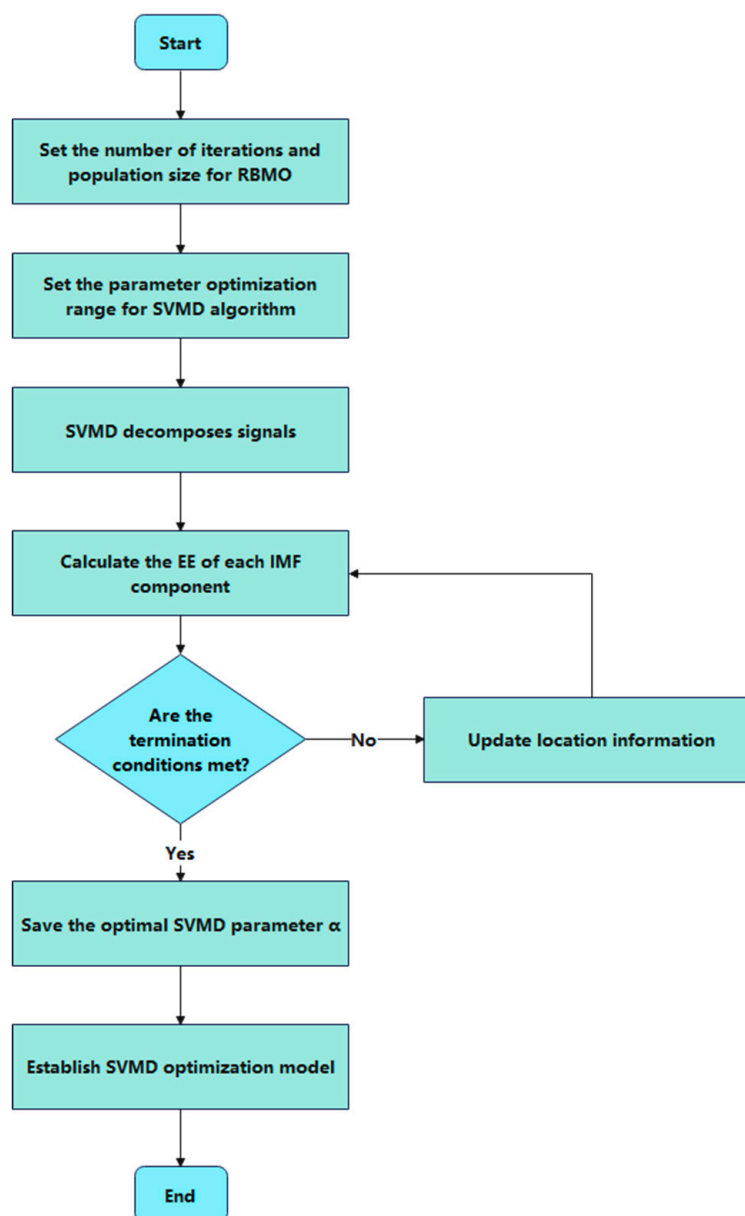


Figure 1. The flowchart of RBMO-SVMD.

2.4. Convolutional Neural Network

CNN is a feedforward neural network that implements automatic mapping from input to output [21], as shown in Figure 2, without the need to predefine input–output relationships. It simplifies the model through local connections and weight sharing, reducing parameters and optimizing the network. In TN sequence prediction, a 1D CNN is employed to capture features. The convolution kernel is the core of the CNN, performing convolution operations to extract intrinsic features from the data. It is represented as follows:

$$y_i = f\left(\sum_{k=1}^K x_{i+k-1} \cdot \omega_k + b\right) \quad (13)$$

The output y_i is calculated by applying the activation function $f(\cdot)$ to the weighted sum of the input element x_{i+k-1} and its corresponding weight ω_k , plus the bias term b . This operation will continuously slide and repeat along the input sequence to generate feature maps.

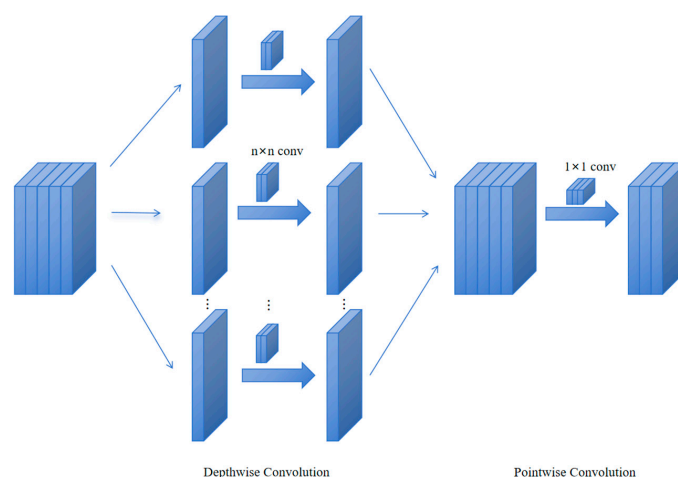


Figure 2. CNN network structure diagram.

2.5. Bidirectional Gated Recurrent Unit

The GRU (Gated Recurrent Unit) is a variant of Recurrent Neural Networks (RNN) designed to address the vanishing gradient problem in backpropagation, with functionality similar to LSTM [22]. Compared to LSTM, GRU has a simpler structure, with only two gating mechanisms, fewer parameters, lower training complexity, and reduced sample requirements. When the hyperparameters are fully optimized, GRU can achieve performance comparable to LSTM.

BiGRU consists of two independent GRU layers: one processes the input sequence in the forward direction, while the other processes it in the reverse direction [23]. Each GRU unit updates its state and generates output based on the current input and the previous hidden state. By leveraging both forward and backward processing, BiGRU captures long-range dependencies within sequences, enhancing the model's performance. Its network structure is shown in Figure 3.

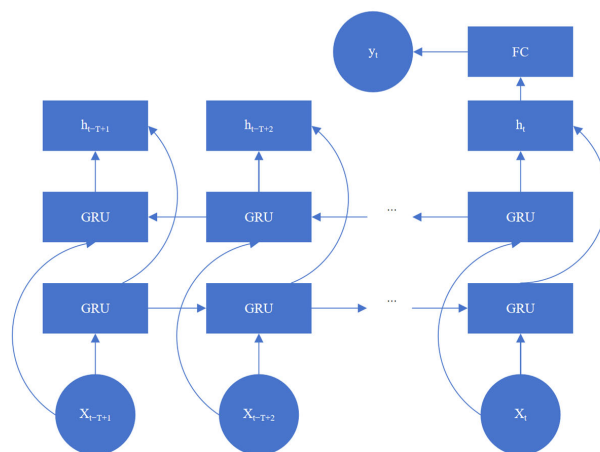


Figure 3. BiGRU network architecture diagram.

2.6. Prediction Model

This research introduces the ERSCB hybrid model framework for TN concentration prediction, which consists of three components, starting with data preprocessing. In this phase, the original TN concentration data, key water quality indicators (TP, DO, NH_3 , Conductivity), and temperature parameters are processed for missing value imputation and standardization to ensure data quality. The second component involves model design, in which a multi-step optimization strategy is applied. Initially, EMD is used to decompose the TN series into multiple IMFs and RES. The SE values of these components are then

calculated to evaluate their complexity. Subsequently, RBOA is introduced to optimize the balance parameter α of SVMD. The subsequence with the highest SE value undergoes a second decomposition, generating a series of VMFs to enhance the precision of feature extraction. The third part is the prediction of results. The TN values of the subsequences are reconstructed, and the CNN-BiGRU prediction model is used to obtain the final short-term TN forecast. The model's performance and reliability are thoroughly evaluated using a series of metrics. The flowchart of ERSCB is shown in Figure 4.

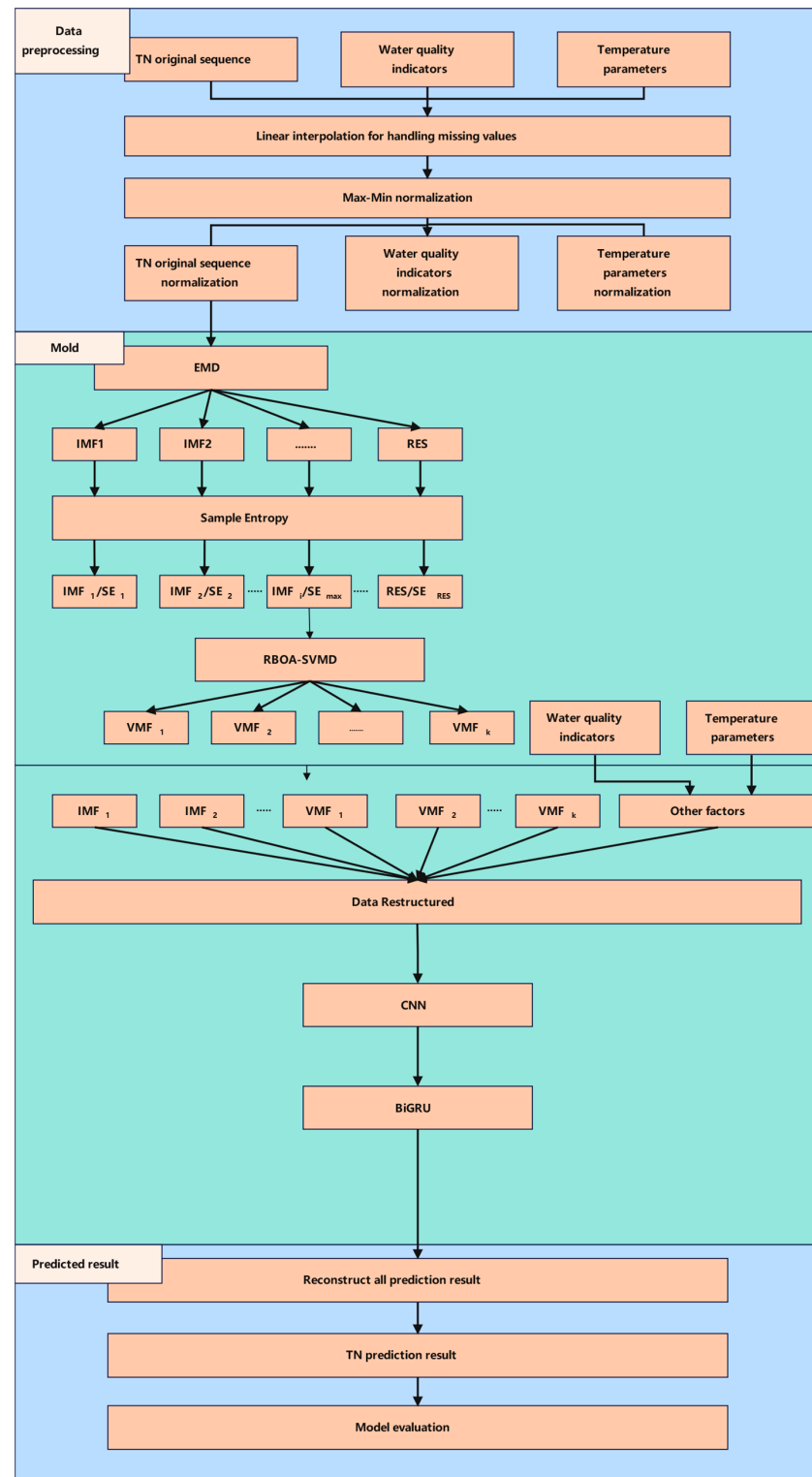


Figure 4. Flow diagram for the model.

2.7. Experimental Analysis and Setup

2.7.1. EMD Decomposition

First, the TN concentration data collected from two monitoring stations were decomposed using EMD, with the results visually presented in Figure 5. The SE values after EMD decomposition are shown in Table 3. By observing Figure 5, it is evident that the TN concentration sequences from the KaShi and GongNaiSi River monitoring stations have been effectively decomposed into 11 Intrinsic Mode Functions (IMF1 to IMF11), along with a residual component (RES). Analysis shows IMF1 from the KaShi River has the highest SE (0.3314), indicating its high complexity, while IMF4 from the GongNaiSi River has the highest SE (0.3291). Further decomposition is applied to these two subsequences.

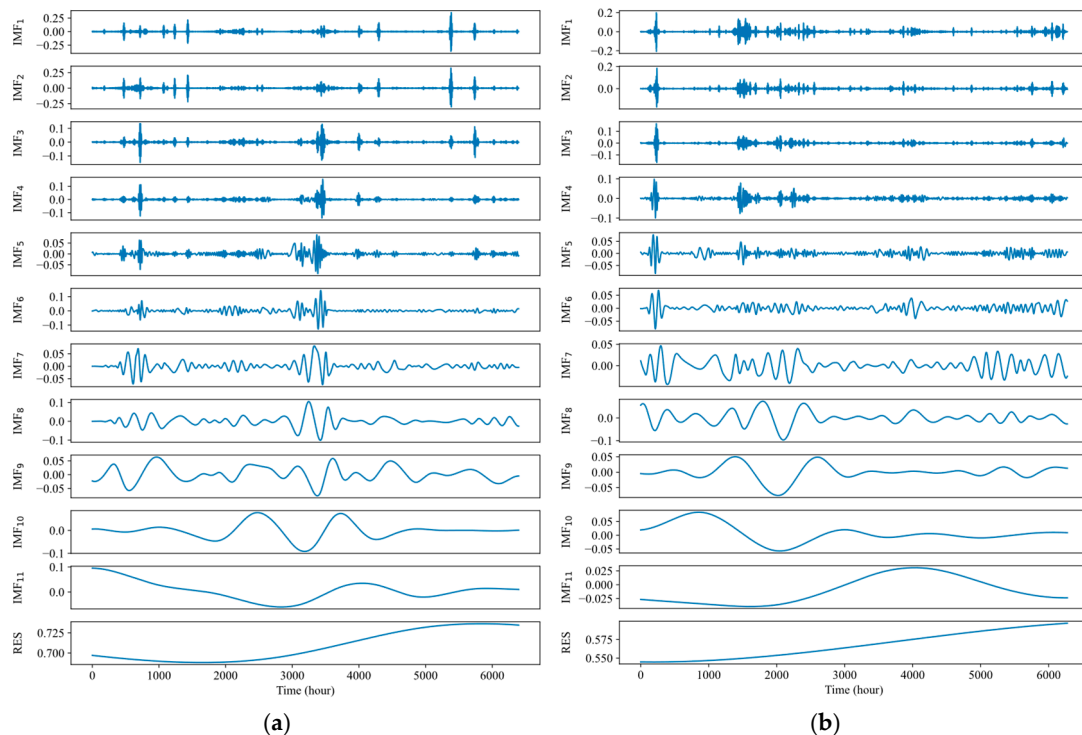


Figure 5. EMD decomposition results of KaShi River and GongNaiSi River data. (a) KaShi; (b) GongNaiSi.

Table 3. SE values of each subsequence decomposed by EMD.

KaShi	SE	GongNaiSi	SE
IMF1	0.3314	IMF1	0.2965
IMF2	0.2286	IMF2	0.3007
IMF3	0.2761	IMF3	0.2849
IMF4	0.2359	IMF4	0.3291
IMF5	0.2306	IMF5	0.2726
IMF6	0.1085	IMF6	0.1671
IMF7	0.0795	IMF7	0.0839
IMF8	0.0427	IMF8	0.0310
IMF9	0.0255	IMF9	0.0105
IMF10	0.0071	IMF10	0.0047
IMF11	0.0043	IMF11	0.0024
RES	0.0010	RES	0.0005

2.7.2. RBMO-SVMD of the Highest Complexity Subsequence

Using the RBMO-optimized SVMD algorithm, the IMF1 and IMF4 components from the two observation stations, obtained through EMD decomposition, were further decom-

posed. RBMO was used to optimize the balancing parameter of the SVM. The fitness function used was Envelope Entropy (EE). Both the population size and the number of iterations in the RBMO algorithm were set to 10. The iterative curve is shown in Figure 6, where the horizontal axis denotes the number of iterations and the vertical axis indicates the fitness function value. The graph shows that the fitness values stabilized after 2 and 9 iterations, respectively. The best balancing parameter values are listed in Table 4, with the optimal balancing parameters for the two monitoring stations being 1000 and 1385, respectively. Since only one parameter is optimized, the iteration curve is almost identical to the parameter variation graph.

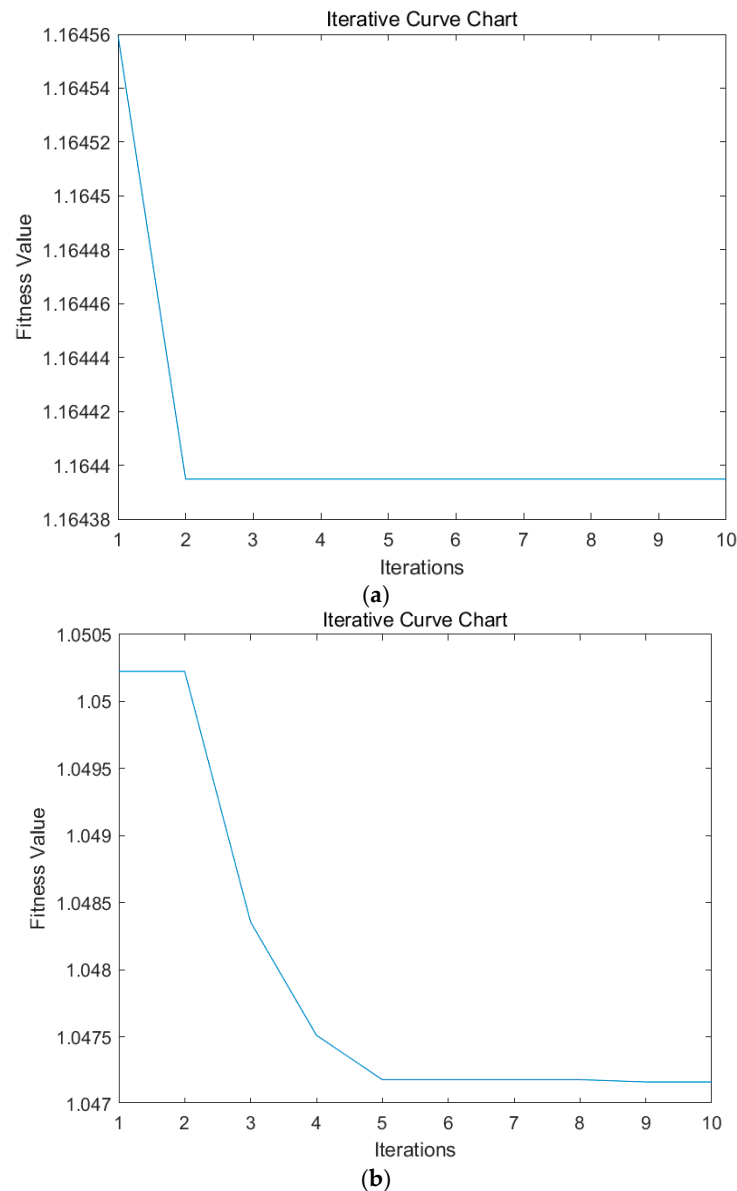


Figure 6. Iterative curve of RBMO-SVMD. (a) KaShi River iterative curve; (b) GongNaiSi River iterative curve.

Table 4. Optimal parameters of RBMO-SVMD.

Observatory	α	Fitness
KaShi	1000	1.1644
GongNaiSi	1385	1.0472

The SVMd algorithm optimized by RBMO was applied to perform secondary decomposition on IMF1. The decomposition results are shown in Figure 7. For the Kashgar River monitoring station and the Gongnasi River monitoring station, IMF1 and IMF4 were decomposed into four subsequences, respectively. As can be seen from the 3D decomposition diagram in Figure 7, the complexity is further reduced after the secondary decomposition of SVMd.

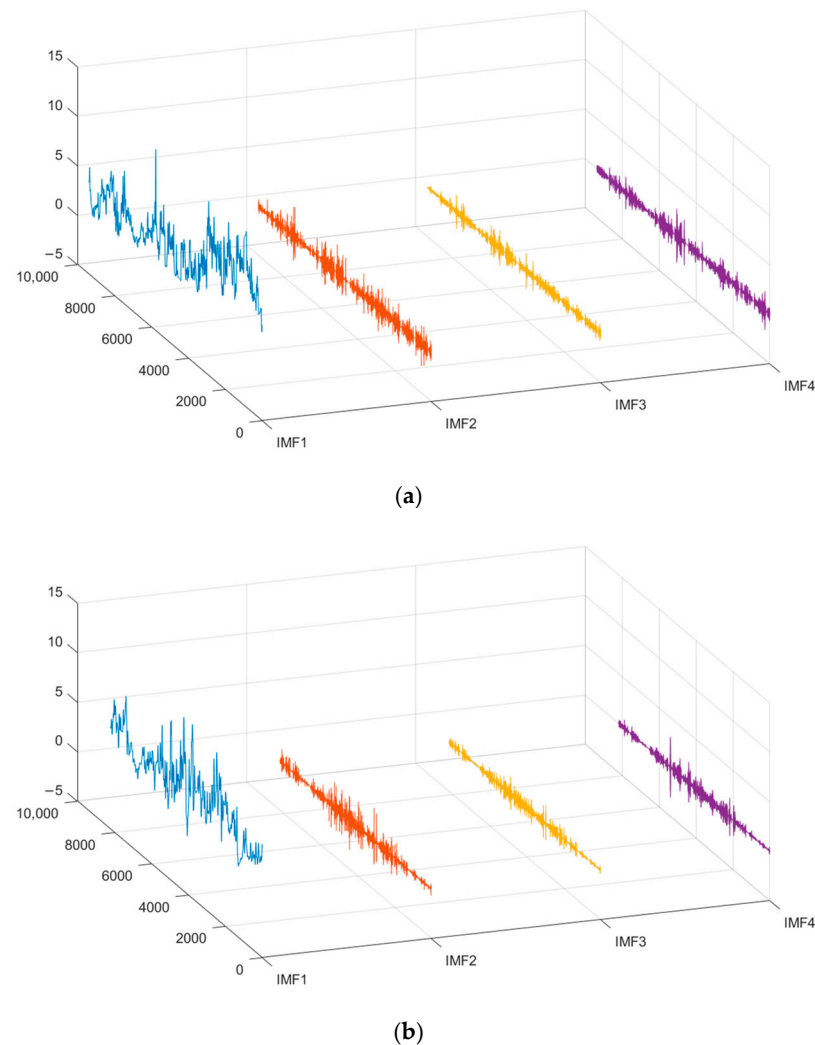


Figure 7. Three-dimensional decomposition diagram after secondary decomposition. (a) The second decomposition results for KaShi River data. (b) The second decomposition result for GongNaiSi River data.

2.7.3. Experimental Setup

In this research, min-max normalization was applied for data standardization. The hybrid model was created with the PyTorch framework, featuring 64 neurons. During training, hyperparameters were continually optimized for improved performance. Table 5 presents the specific hyperparameter settings of the hybrid model.

Table 5. Hyperparameter settings for hybrid models.

Hyperparameter Name	Setting Value
Activation function	relu
Loss function	mse
Optimization function	adam
Bathch size	64
epoch	200
learning rate	0.001
Convolution kernel	[3, 5]

2.8. Evaluation Metrics

This research assessed the model's performance using three key metrics: Mean Absolute Error (MAE), Root Mean Square Error (RMSE), and the Coefficient of Determination (R^2). These metrics together assess the accuracy of the model's predictions, sensitivity to outliers, and the fit between predicted and actual values, providing a detailed evaluation of the model's performance.

$$\text{MAE}(y, \hat{y}) = \frac{1}{n} \sum_{i=1}^n |y_i - \hat{y}_i| \quad (14)$$

$$\text{RMSE}(y, \hat{y}) = \sqrt{\frac{1}{n} \sum_{i=1}^n (|y_i - \hat{y}_i|)^2} \quad (15)$$

$$R^2(y, \hat{y}) = 1 - \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{\sum_{i=1}^n (y_i - \bar{y})^2} \quad (16)$$

The above formulas reflect the differences between the actual values y of sample points and the predicted values $\hat{y}$ from the model, using various statistical metrics for quantitative analysis. RMSE and MAE are crucial indicators for assessing prediction accuracy. Lower values of these metrics indicate greater consistency between the model's predictions and actual outcomes, reflecting a notable enhancement in predictive performance. Additionally, R^2 is used to evaluate the goodness of fit. The closer its value is to 1, the smaller the deviation between predicted and observed values, further indicating a strong fit of the model to the data.

3. Results and Discussion

3.1. The Prediction Results of the Model

The decomposed components were combined with the water quality factors and temperature parameters, and CNN and BiGRU were used to predict TN concentration data, forming the model's final predicted values. This approach yielded TN concentration predictions for the KaShi River and GongNaiSi River monitoring stations. Figure 8 shows the predictive performance of ERSCB at the two monitoring stations, indicating that the proposed ERSCB model demonstrates a high degree of fitting accuracy.

The predictive performance metrics of the ERSCB model at the two monitoring stations are presented in Table 6. For the convenience of comparison with other models, the errors selected in this paper are all normalized results. The results indicate that both stations exhibit similar performance in terms of MAE, with relatively low values, confirming that the model's prediction errors are small. In terms of RMSE, the value for the KaShi River station is significantly lower than that for the GongNaiSi River station, which may be attributed to the smaller data fluctuations at the former station. Overall, both stations

demonstrate high R^2 values, indicating that the model's predicted values closely align with the actual observed values.

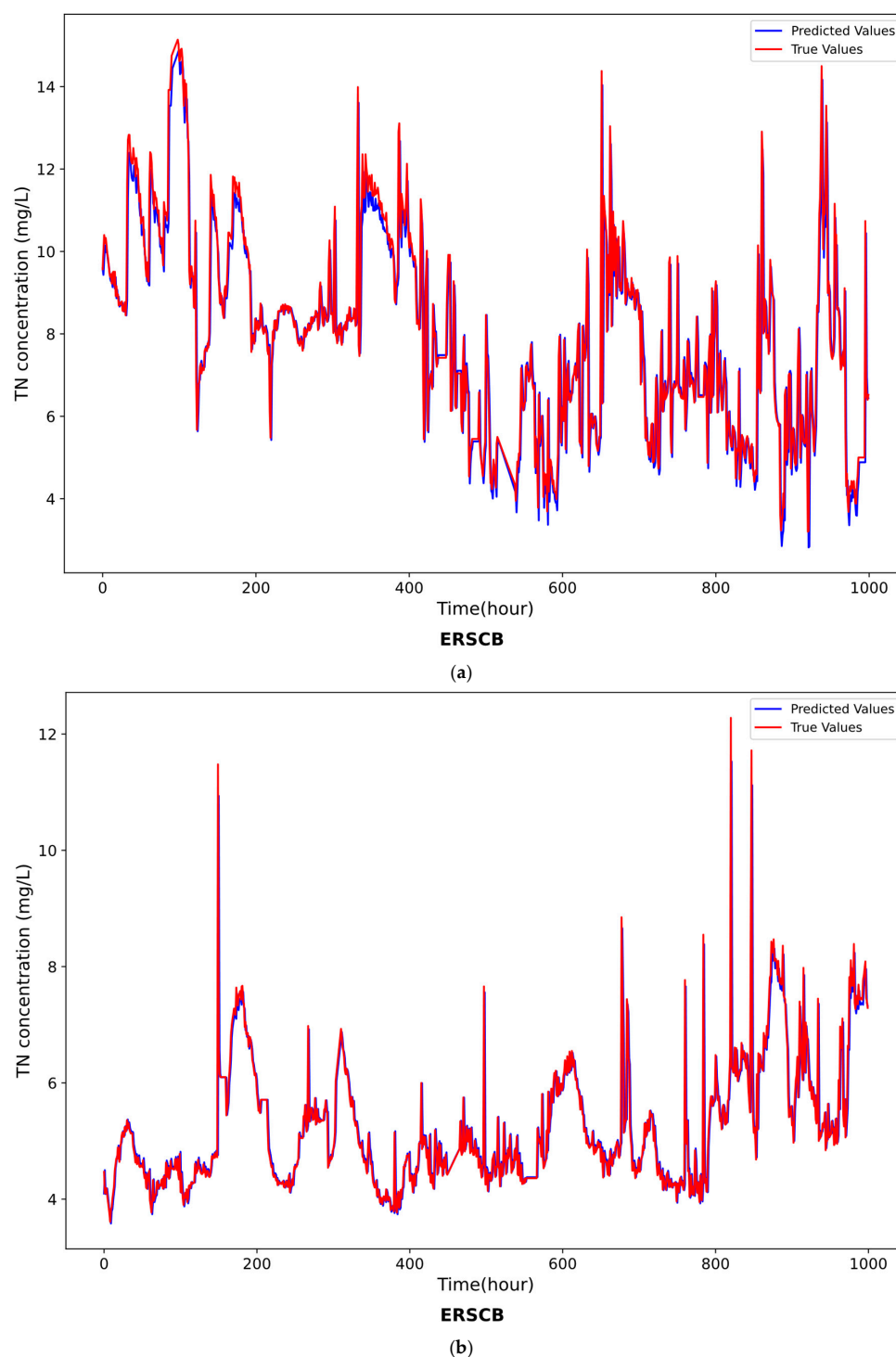


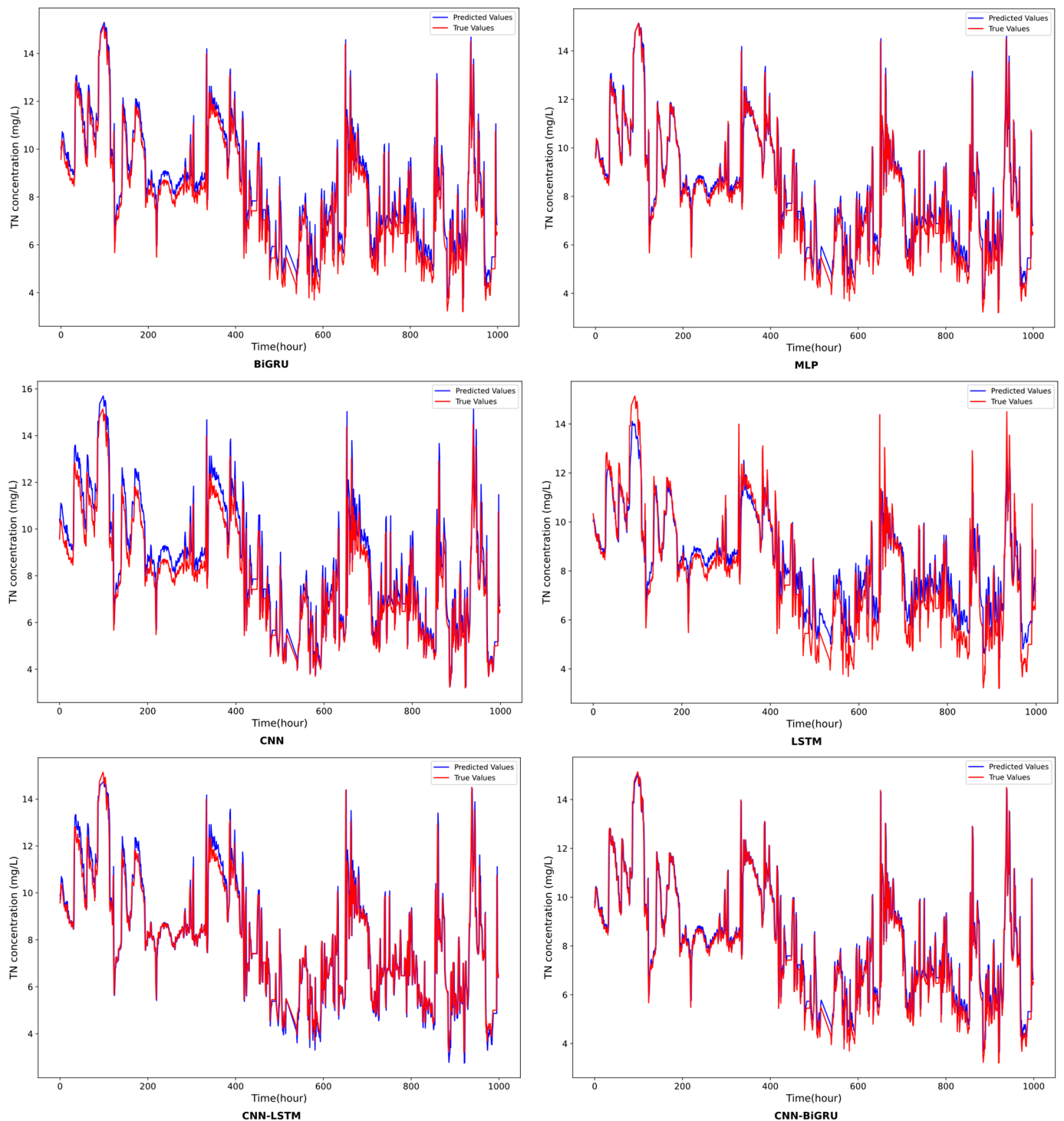
Figure 8. The proposed ERSCB model's line charts for TN concentration predictions at the KaShi and GongNaiSi River monitoring stations. (a) KaShi; (b) GongNaiSi.

Table 6. The proposed model ERSCB performs well in predicting TN concentration.

Observatory	RMSE	MAE	R^2
KaShi	0.0452	0.0316	0.9742
GongNaiSi	0.0637	0.0379	0.9677

3.2. Comparison of the Proposed Model's Predictions with Those of Traditional Deep Learning Models

In this study, several common deep learning models (MLP, CNN, BiGRU, LSTM, CNN-LSTM, and CNN-BiGRU) were compared with the proposed ERSCB model and validated using the KaShi and GongNaiSi River monitoring station data. To ensure more accurate predictions, the input parameters for all models included water quality factors and temperature parameters. The prediction results are shown in Figure 9, and the corresponding performance metrics are detailed in Table 5.



(a)

Figure 9. Cont.

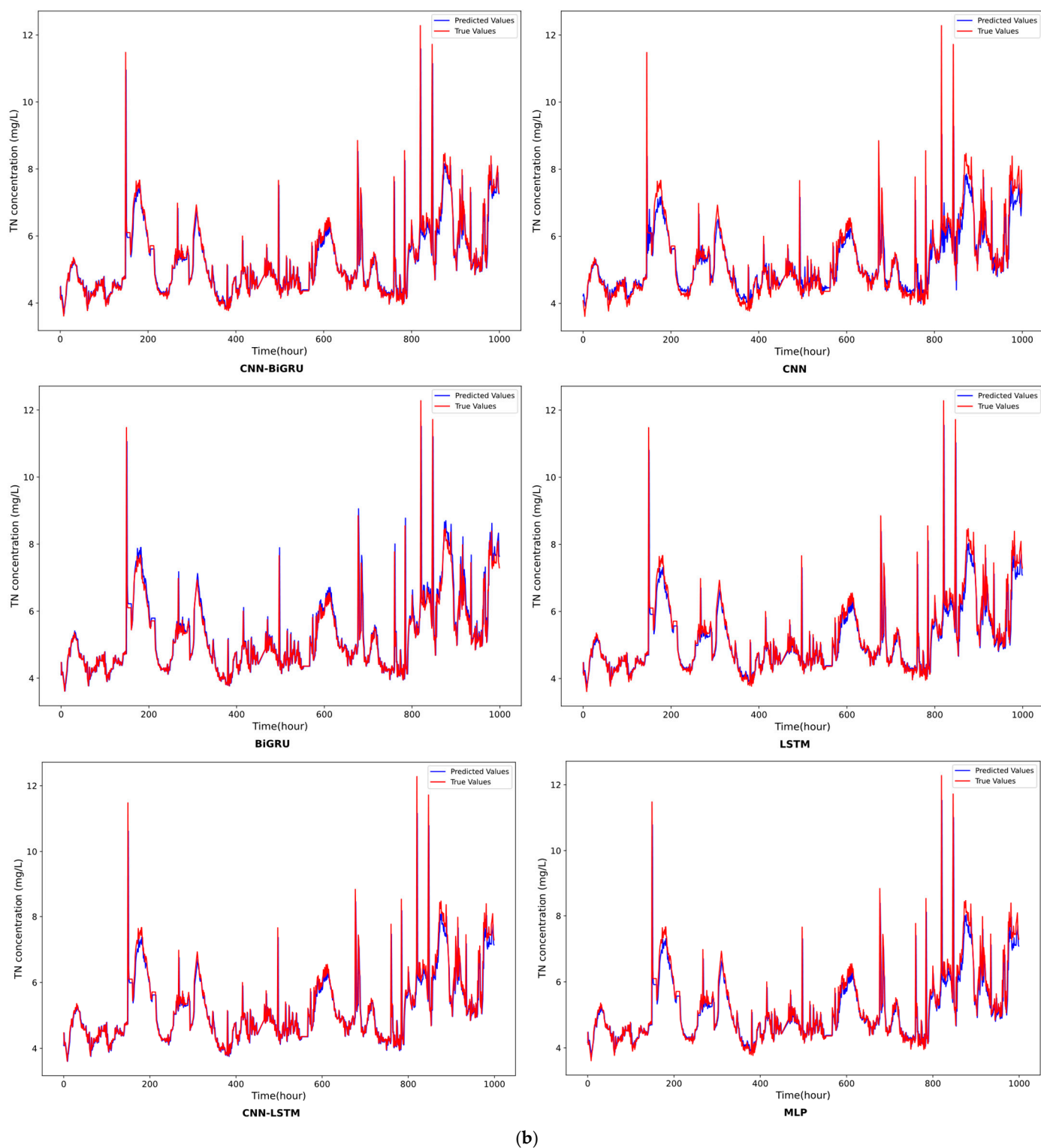


Figure 9. Comparison of prediction results of different deep learning models using data from two monitoring stations. (a) KaShi; (b) GongNaiSi.

As shown in Figure 9, the prediction results of traditional deep learning models exhibit a poor fit with the actual data and fail to accurately reflect the TN concentration trends. In contrast, the ERSCB model delivers better predictive performance, effectively capturing the TN concentration trends.

According to Table 7, at the KaShi River monitoring station, the LSTM model shows a high MAE value, indicating a significant prediction error. Additionally, the CNN model has the largest RMSE value, suggesting a considerable discrepancy between its predictions and the actual values. At the GongNaiSi River monitoring station, the CNN model performs the worst among the six models in terms of RMSE, MAE, and R^2 , highlighting its poor fitting capability. In comparison, the ERSCB model outperforms all other models across RMSE, MAE, and R^2 at both monitoring stations, demonstrating higher predictive accuracy than traditional deep learning models.

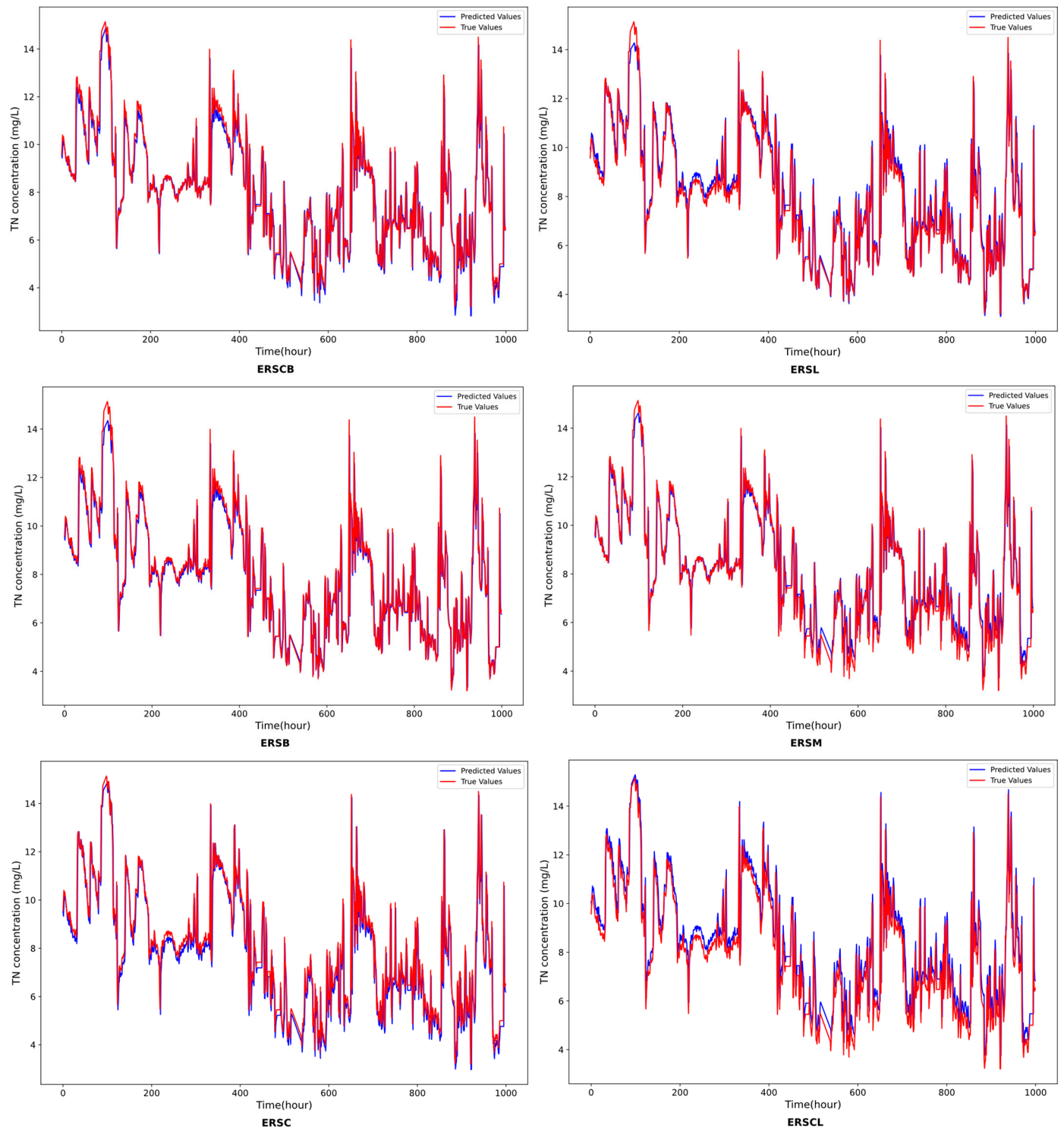
Table 7. Evaluation of the performance of the ERSCB with other deep learning models at the KaShi and GongNaiSi River monitoring stations.

Observatory	Model	RMSE	MAE	R^2
KaShi	MLP	0.0543	0.0324	0.9275
	CNN	0.0621	0.0341	0.8616
	LSTM	0.0606	0.0407	0.9077
	CNN-LSTM	0.0549	0.0350	0.9195
	BiGRU	0.0533	0.0318	0.9285
	CNN-BiGRU	0.0530	0.0286	0.9364
	ERSCB	0.0452	0.0316	0.9742
GongNaiSi	MLP	0.0730	0.0384	0.9270
	CNN	0.0939	0.0637	0.8571
	LSTM	0.0794	0.0624	0.8613
	CNN-LSTM	0.0678	0.0424	0.9094
	BiGRU	0.0710	0.0469	0.9356
	CNN-BiGRU	0.0687	0.0442	0.9388
	ERSCB	0.0637	0.0379	0.9677

3.3. Comparison Results of Different Models Combined with EMD-RBMO-SVMD

To further validate the accuracy of the EMD-RBMO-SVMD decomposition technique combined with the CNN-BiGRU hybrid model proposed in this study, several hybrid models (EMD-RBMO-SVMD-MLP (ERSM), EMD-RBMO-SVMD-CNN (ERSC), EMD-RBMO-SVMD-CNN-LSTM (ERSCL), EMD-RBMO-SVMD-LSTM (ERSL), and EMD-RBMO-SVMD-BiGRU (ERSB)) were compared with the proposed EMD-RBMO-SVMD-CNN-BiGRU (ERSCB) model. To demonstrate the effectiveness of the secondary decomposition method, it was applied to all the comparison models. The results with and without secondary decomposition were compared, and the performance metrics for each model are presented in Table 8. The prediction results are shown in Figure 10.

As seen in Table 8, the prediction accuracy of each model improved after applying secondary decomposition. The CNN-BiGRU model achieved the best performance and serves as the foundational model for ERSCB. The four performance metrics of the ERSCB model are the most optimal. For the KaShi River monitoring station data, the ERSCB model has RMSE, MAE, and R^2 values of 0.0452, 0.0316, and 0.9742, respectively. The corresponding metrics for the GongNaiSi River station are 0.0637, 0.0379, and 0.9677.



(a)

Figure 10. Cont.

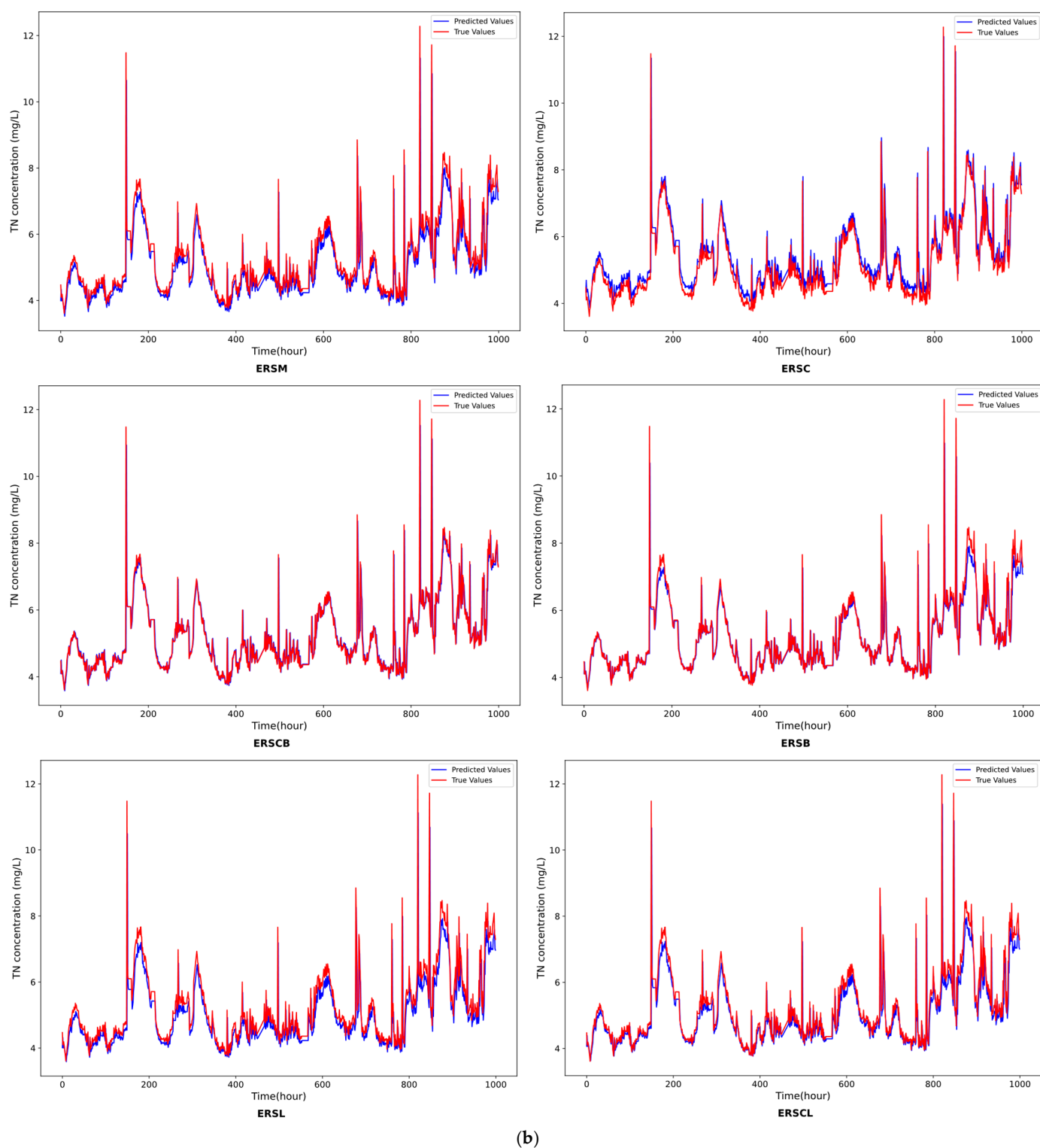


Figure 10. Comparison of prediction outcomes for hybrid models with secondary decomposition. (a) KaShi; (b) GongNaiSi.

Table 8. The performance of the proposed ERSCB model is compared with that of the hybrid model that combines EMD-RBMO-SVMD.

Observatory	Model	RMSE	MAE	R ²
KaShi	ERSM	0.0530	0.0324	0.9278
	ERSC	0.0531	0.0319	0.9227
	ERSCL	0.0549	0.0350	0.9195
	ERSL	0.0545	0.0341	0.9284
	ERSB	0.0528	0.0318	0.9299
	ERSCB	0.0452	0.0316	0.9742
GongNaiSi	ERSM	0.0650	0.0383	0.9370
	ERSC	0.0710	0.0466	0.8691
	ERSCL	0.0669	0.0410	0.9335
	ERSL	0.0649	0.0392	0.9101
	ERSB	0.0651	0.0384	0.9408
	ERSCB	0.0637	0.0379	0.9677

In the secondary decomposition hybrid models, both monitoring stations showed improvements after applying secondary decomposition compared to the CNN-BiGRU model. The percentage improvements are 3.78% and 2.89%, respectively, indicating that the use of secondary decomposition further reduced the complexity of the TN series and improved prediction accuracy, confirming the effectiveness and practicality of the proposed model.

The results after secondary decomposition are shown in Figure 10. Compared with other models, the ERSCB model is more in line with the actual values, indicating that it achieved the highest predictive performance.

3.4. Comparisons with Existing Models

In order to evaluate the effectiveness of the proposed model, this paper compared ERSCB with existing water quality prediction models, as presented in Table 9. Both the VMD-IGOA-LSTM model and the proposed model used the same dataset. At the KaShi and GongNaiSi River monitoring stations, the proposed model showed improved R² values of 4.0% and 3.1%, respectively, over the VMD-IGOA-LSTM model. Additionally, its RMSE and MAE values were lower than those of the VMD-IGOA-LSTM model. Therefore, the proposed model achieved higher accuracy in TN concentration predictions compared to the existing models.

Table 9. Comparison between the existing models and the model presented in this article.

Model	Time	Observation Station	RMSE	MAE	R ²
The proposed model	4h	KaShi	0.0452	0.0316	0.9742
		GongNaiSi	0.0637	0.0379	0.9677
VMD-IGOA-LSTM	4h	KaShi	0.0595	0.0408	0.9340
		GongNaiSi	0.0657	0.0389	0.9367

3.5. Discussion

SVMD is an adaptive, fully non-recursive method for modal decomposition and signal processing, offering excellent decomposition performance. However, it requires manual selection of the number of balancing parameters, which can affect the decomposition results. RBMO-SVMD automatically determines the optimal parameters for SVMD, enabling efficient signal decomposition and improving decomposition performance.

Due to the non-stationary and nonlinear nature of the TN concentration series, the complexity of the subsequences generated by EMD decomposition remains high. The subsequence with the highest complexity was subjected to secondary decomposition using RBMO-SVMD to reduce its complexity and enhance prediction accuracy.

To verify the accuracy of the proposed model, experiments were carried out using six comparison models using datasets from two monitoring stations in the Ili region. The results indicate that the proposed model significantly surpasses the other prediction models.

4. Conclusions

This study presents a novel hybrid model for predicting TN concentration, called ERSCB, which combines EMD, RBMO-SVMD, and CNN-BiGRU. The key conclusions are as follows:

- (1) A modified SVMD, named RBMO-SVMD, based on the RBMO optimization algorithm is proposed, which removes the necessity for manual selection of the balancing parameter α .
- (2) SE was used to measure the complexity of the subsequences obtained from the first EMD decomposition. The subsequences with the highest complexity were further decomposed to reduce the complexity of the TN concentration sequence.
- (3) The ERSCB hybrid model was proposed and tested using datasets from two monitoring stations in the Ili region: the KaShi and GongNaiSi River. In comparison to other deep learning models and hybrid models that integrate EMD-RBMO-SVMD, the proposed model exhibited significantly superior predictive performance across all comparison models.

The model has shown significant practical application value in the field of TN concentration prediction: (1) The prediction results of the model provide key data support for relevant departments to help formulate more accurate water pollution control strategies. (2) This study offers a fresh perspective on water quality prediction and holds reference value for future related research. (3) The proposed model is not only applicable to predicting water quality pollutants but can also be used in other domains.

Author Contributions: Conceptualization, S.Z.; Methodology, S.Z.; Software, S.Z., T.Z. and C.C.; Validation, S.Z., R.L., Y.L. and L.X.; Investigation, R.L.; Data curation, S.Z., Y.L., T.Z. and C.C.; Writing—original draft, S.Z.; Writing—review and editing, L.X.; Visualization, S.Z.; Supervision, Y.L. and L.X.; Project administration, R.L.; Funding acquisition, L.X. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported by the National Natural Science Foundation of China (No. 32460290), the Ecological Security Diagnosis and Evaluation Project of Xinjiang Production and Construction Corps Based on Multi source Heterogeneous Data (No. 2023CB008-23).

Data Availability Statement: The data used in this study were sourced from the National Real-time Automatic Monitoring and Data Publishing System for Surface Water Quality, accessible via the website <https://szzdjc.cnemc.cn:8070/GJZ/Business/Publish/Main.html> (accessed on 15 September 2024).

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Haghiabi, A.H.; Nasrolahi, A.H.; Parsaie, A. Water quality prediction using machine learning methods. *Water Qual. Res. J.* **2018**, *53*, 3–13. [[CrossRef](#)]
2. Wang, X.; Li, Y.; Qiao, Q.; Tavares, A.; Liang, Y. Water Quality Prediction Based on Machine Learning and Comprehensive Weighting Methods. *Entropy* **2023**, *25*, 1186. [[CrossRef](#)] [[PubMed](#)]

3. Choi, J.; Chung, S.; Ryoo, J. Comparative evaluation of QUAL2E and QUAL-NIER models for water quality prediction in eutrophic river. *J. Korean Soc. Water Environ.* **2008**, *24*, 54–62.
4. Moses, S.A.; Janaki, L.; Joseph, S.; Joseph, J. Water quality prediction capabilities of WASP model for a tropical lake system. *Lakes Reserv. Res. Manag.* **2015**, *20*, 285–299. [[CrossRef](#)]
5. Noori, N.; Kalin, L.; Isik, S. Water quality prediction using SWAT-ANN coupled approach. *J. Hydrol.* **2020**, *590*, 125220. [[CrossRef](#)]
6. Wang, J.; Zhang, L.; Zhang, W.; Wang, X. Reliable Model of Reservoir Water Quality Prediction Based on Improved ARIMA Method. *Environ. Eng. Sci.* **2019**, *36*, 1041–1048. [[CrossRef](#)]
7. Xu, R.; Xiong, Q.; Yi, H.; Wu, C.; Ye, J. Research on water quality prediction based on SARIMA-LSTM: A case study of Beilun Estuary. In Proceedings of the 2019 IEEE 21st International Conference on High Performance Computing and Communications; IEEE 17th International Conference on Smart City; IEEE 5th International Conference on Data Science and Systems (HPCC/SmartCity/DSS), Zhangjiajie, China, 10–12 August 2019; IEEE: Piscataway, NJ, USA, 2019; pp. 2183–2188.
8. Alnahit, A.O.; Mishra, A.K.; Khan, A.A. Stream water quality prediction using boosted regression tree and random forest models. *Stoch. Environ. Res. Risk Assess.* **2022**, *36*, 2661–2680. [[CrossRef](#)]
9. Masood, A.; Niazkari, M.; Zakwan, M.; Piraei, R. A Machine Learning-Based Framework for Water Quality Index Estimation in the Southern Bug River. *Water* **2023**, *15*, 3543. [[CrossRef](#)]
10. Setshedi, K.J.; Mutingwende, N.; Ngqwala, N.P. The use of artificial neural networks to predict the physicochemical characteristics of water quality in three district municipalities, Eastern Cape Province, South Africa. *Int. J. Environ. Res. Public Health* **2021**, *18*, 5248. [[CrossRef](#)] [[PubMed](#)]
11. Rustam, F.; Ishaq, A.; Kokab, S.T.; Diez, I.d.l.T.; Mazón, J.L.V.; Rodríguez, C.L.; Ashraf, I. An Artificial Neural Network Model for Water Quality and Water Consumption Prediction. *Water* **2022**, *14*, 3359. [[CrossRef](#)]
12. Baek, S.-S.; Pyo, J.; Chun, J.A. Prediction of water level and water quality using a CNN-LSTM combined deep learning approach. *Water* **2020**, *12*, 3399. [[CrossRef](#)]
13. Chen, H.; Yang, J.; Fu, X.; Zheng, Q.; Song, X.; Fu, Z.; Wang, J.; Liang, Y.; Yin, H.; Liu, Z.; et al. Water Quality Prediction Based on LSTM and Attention Mechanism: A Case Study of the Burnett River, Australia. *Sustainability* **2022**, *14*, 13231. [[CrossRef](#)]
14. Wang, Z.; Wang, Q.; Wu, T. A novel hybrid model for water quality prediction based on VMD and IGOA optimized for LSTM. *Front. Environ. Sci. Eng.* **2023**, *17*, 88. [[CrossRef](#)]
15. Zhang, Y.; Li, C.; Jiang, Y.; Sun, L.; Zhao, R.; Yan, K.; Wang, W. Accurate prediction of water quality in urban drainage network with integrated EMD-LSTM model. *J. Clean. Prod.* **2022**, *354*, 131724. [[CrossRef](#)]
16. Tan, R.; Wang, Z.; Wu, T.; Wu, J. A data-driven model for water quality prediction in Tai Lake, China, using secondary modal decomposition with multidimensional external features. *J. Hydrol. Reg. Stud.* **2023**, *47*, 101435. [[CrossRef](#)]
17. Wang, Y.; Sun, S.; Chen, X.; Zeng, X.; Kong, Y.; Chen, J.; Guo, Y.; Wang, T. Short-term load forecasting of industrial customers based on SVM and XGBoost. *Int. J. Electr. Power Energy Syst.* **2021**, *129*, 106830. [[CrossRef](#)]
18. Wang, T.; Zhang, M.; Yu, Q.; Zhang, H. Comparing the applications of EMD and EEMD on time–frequency analysis of seismic signal. *J. Appl. Geophys.* **2012**, *83*, 29–34. [[CrossRef](#)]
19. Fu, S.; Li, K.; Huang, H.; Ma, C.; Fan, Q.; Zhu, Y. Red-billed blue magpie optimizer: A novel meta-heuristic algorithm for 2D/3D UAV path planning and engineering design problems. *Artif. Intell. Rev.* **2024**, *57*, 134. [[CrossRef](#)]
20. Nazari, M.; Sakhaei, S.M. Successive variational mode decomposition. *Signal Process.* **2020**, *174*, 107610. [[CrossRef](#)]
21. Anand, M.V.; Sohitha, C.; Saraswathi, G.N.; Lavanya, G.V. Water quality prediction using CNN. In *Journal of Physics: Conference Series*; IOP Publishing: Bristol, UK, 2023; Volume 2484, p. 012051.
22. Tu, J.; Yang, X.; Chen, C.; Gao, S.; Wang, J.; Sun, C. Water Quality Prediction Model Based on GRU hybrid network. In Proceedings of the 2019 Chinese Automation Congress (CAC), Hangzhou, China, 22–24 November 2019; IEEE: Piscataway, NJ, USA, 2019; pp. 1893–1898.
23. Gai, R.; Yang, J. Research on water quality spatiotemporal forecasting model based on ST-BIGRU-SVR neural network. *Water Sci. Technol.* **2023**, *88*, 530–541. [[CrossRef](#)] [[PubMed](#)]

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.