

## Article

# A River Water Quality Prediction Method Based on Dual Signal Decomposition and Deep Learning

Yifan Bai<sup>1,2</sup>, Menghang Peng<sup>3</sup> and Mei Wang<sup>1,3,\*</sup>

<sup>1</sup> Hebei Province Key Laboratory of Sustained Utilization & Development of Water Recourse, Hebei Geo University, Shijiazhuang 050030, China; beiyf124@gmail.com

<sup>2</sup> Yellow River Engineering Consulting Corporation Limited, Zhengzhou 450003, China

<sup>3</sup> School of Water Conservancy and Transportation, Zhengzhou University, Zhengzhou 450001, China; pmh2424706150@gs.zzu.edu.cn

\* Correspondence: maywang226@zzu.edu.cn

**Abstract:** Traditional single prediction models struggle to address the complexity and nonlinear changes in water quality forecasting. To address this challenge, this study proposed a coupled prediction model (RF-TVSV-SCL). The model includes Random Forest (RF) feature selection, dual signal decomposition (Time-Varying Filtered Empirical Mode Decomposition, TVF-EMD, and Sparrow Search Algorithm-Optimized Variational Mode Decomposition, SSA-VMD), and a deep learning predictive model (Sparrow Search Algorithm-Convolutional Neural Network-Long Short-Term Memory, SSA-CNN-LSTM). Firstly, the RF method was used for feature selection to extract important features relevant to water quality prediction. Then, TVF-EMD was employed for preliminary decomposition of the water quality data, followed by a secondary decomposition of complex Intrinsic Mode Function (IMF) components using SSA-VMD. Finally, the SSA-CNN-LSTM model was utilized to predict the processed data. This model was evaluated for predicting total phosphorus (TP), total nitrogen (TN), ammonia nitrogen (NH<sub>3</sub>-N), dissolved oxygen (DO), permanganate index (COD<sub>Mn</sub>), conductivity (EC), and turbidity (TB), across 1, 3, 5, and 7-d forecast periods. The model performed exceptionally well in short-term predictions, particularly within the 1–3 d range. For 1-, 3-, 5-, and 7-d forecasts, R<sup>2</sup> ranged from 0.93–0.96, 0.79–0.87, 0.63–0.72, and 0.56–0.64, respectively, significantly outperforming other comparison models. The RF-TVSV-SCL model demonstrates excellent predictive capability and generalization ability, providing robust technical support for water quality forecasting and pollution prevention.

**Keywords:** feature selection; secondary decomposition; SSA; LSTM; water quality prediction



**Citation:** Bai, Y.; Peng, M.; Wang, M. A River Water Quality Prediction Method Based on Dual Signal Decomposition and Deep Learning. *Water* **2024**, *16*, 3099. <https://doi.org/10.3390/w16213099>

Academic Editors: Tianxiang Wang and Alex Neumann

Received: 30 August 2024

Revised: 22 October 2024

Accepted: 25 October 2024

Published: 29 October 2024



**Copyright:** © 2024 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/>).

## 1. Introduction

With the rapid development of the global economy and industrialization, water pollution has become increasingly severe. The deterioration of river water quality has become a focus of global attention. Water pollution not only threatens the safety of drinking water for humans but also leads to significant degradation of aquatic ecosystems, affecting biodiversity and ecological balance. Given the complexity and widespread nature of water pollution, relying solely on post-event remediation is insufficient to control the resulting damage effectively. Therefore, establishing an effective water quality prediction system is essential. By conducting long-term forecasts of water pollutants, potential risk areas can be identified in a timely manner, thereby reducing the impact of water pollution on the environment and human health. This is of critical importance for the sustainable management of global water resources and environmental protection.

To understand future trends in water quality changes, a lot of research has been done on methods for predicting water quality. Traditional water quality prediction models, like the Autoregressive Moving Average (ARMA) and the Autoregressive Integrated Moving Average (ARIMA), perform well when dealing with short-term linear data. However, they

struggle when facing non-linear or complex systems [1,2]. Additionally, these models often rely on a large amount of hydrological data, increasing the burden of data collection and computation. They are also quite sensitive to missing data, which can lead to increased forecast errors [3,4]. In recent years, with the rapid advancement of artificial intelligence and machine learning technologies, deep learning neural network models have shown new potential in long-term water quality forecasting. Many researchers have begun to apply techniques such as Long Short-Term Memory networks (LSTM) [5–8], Gated Recurrent Units (GRU) [9–11], and Temporal Convolutional Networks (TCN) [12,13] to construct water quality prediction models.

However, due to the complex nonlinear relationships among various factors influencing water quality, including meteorological conditions, geological features, and pollution sources, traditional single models have shown clear limitations in dealing with this complexity and are often unable to accurately capture the changes in time series data. Multi-feature prediction models that combine Convolutional Neural Networks and Long Short-Term Memory networks (CNN-LSTM) have been successfully applied in water quality forecasting due to their advantages in feature extraction, long-term dependency capture, fault tolerance, and prediction accuracy [14–17]. For example, in a study on water quality at the Beilun Estuary in Guangxi, China, Yang et al. (2021) used a CNN-LSTM model to predict pH and ammonia nitrogen (NH<sub>3</sub>-N), demonstrating the model's excellent performance in handling nonlinear time series predictions and effectively capturing long-term dependencies through the attention mechanism [17]. Additionally, Barzegar et al. (2020) applied a CNN-LSTM model to predict dissolved oxygen (DO) and chlorophyll-a (Chl-a) concentrations in Small Prespa Lake, Greece, and compared it with traditional machine learning models like support vector regression and decision tree [14]. The results showed that the CNN-LSTM hybrid model excelled in capturing both high and low concentrations of water quality variables, particularly outperforming other models in DO prediction.

Nevertheless, relying solely on neural network models still presents certain limitations, especially in terms of long-term forecasting capabilities, which urgently need further optimization. By combining signal decomposition methods such as Empirical Mode Decomposition (EMD), Complete Ensemble Empirical Mode Decomposition with Adaptive Noise (CEEMDAN), Variational Mode Decomposition (VMD), and Time-Varying Filtered Empirical Mode Decomposition (TVF-EMD) with neural network models to construct an ensemble model, prediction accuracy can be significantly improved. Among these methods, EMD may produce mode mixing problems when handling nonlinear and nonstationary signals [18]. Although CEEMDAN alleviates this issue to some extent by adding white noise, it has higher computational complexity and is sensitive to outliers [19,20]. In contrast, TVF-EMD and VMD methods are computationally simpler and less prone to mode mixing during decomposition, making them more effective in improving model performance [21–23].

However, existing studies [18,24,25] mainly focus on single signal decomposition, which, although it improves the prediction performance of the model to some extent, still leaves the Intrinsic Mode Function (IMF) components containing a large amount of complex signals. Directly using these IMFs for prediction may weaken the model's predictive ability.

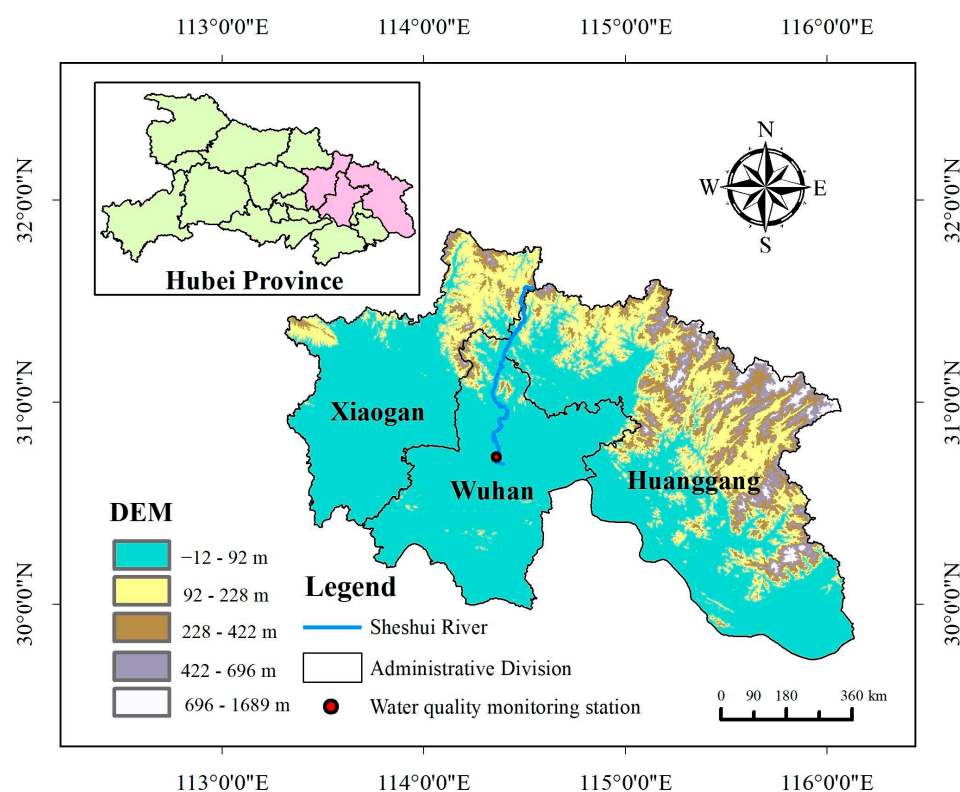
To address this issue, this study proposes a coupled prediction model that integrates multi-level decomposition and deep learning techniques, including RF, TVF-EMD, VMD, Sparrow Search Algorithm (SSA), CNN, and LSTM. The model aims to effectively resolve the incomplete decomposition problem in traditional models and enhance water quality prediction accuracy. Specifically, the study first employs Random Forest (RF) [26] for feature selection, followed by initial signal decomposition of water quality data using TVF-EMD. The most complex IMF components, with the highest noise, are then subjected to a secondary decomposition using VMD. Finally, the CNN-LSTM model is applied to predict the high and low-frequency decomposed data, with the parameters of VMD and CNN-LSTM optimized using the SSA [27]. This paper evaluates the prediction performance and generalization capability of the proposed model using the total Phosphorus (TP), total

Nitrogen (TN), ammonia nitrogen (NH<sub>3</sub>-N), DO, permanganate index (COD<sub>Mn</sub>), electrical conductivity (EC), and turbidity (TB) indicators from the Sheshui River in the Yangtze River Basin, across different lead times of 1, 3, 5, and 7 days.

## 2. Materials and Methods

### 2.1. Data Source

The water quality data utilized in this study were obtained from the Shekou monitoring station (Figure 1), located on the Sheshui River in Wuhan, China. The dataset, provided by the China National Environmental Monitoring Centre (<https://szzdjc.cnemc.cn:8070/GJZ/Business/Publish/Main.html>), covers the period from 1 January 2019 to 31 December 2021. The monitored water quality indicators encompass nine parameters: temperature (Temp), pH, DO, EC, TB, COD<sub>Mn</sub>, NH<sub>3</sub>-N, TP, and TN. A total of 1050 daily data entries were collected, with 80% of the data used for model training and the remaining 20% used for model testing to evaluate performance. To better assess the model’s generalization ability, this study focused on predicting seven key water quality indicators: TP, TN, NH<sub>3</sub>-N, DO, COD<sub>Mn</sub>, EC, and TB. The statistical characteristics of these indicators are presented in Table 1.



**Figure 1.** Location of Shekou monitoring station.

**Table 1.** Statistical characteristics of each indicator sequence.

Indicator	Sequence Length/d	Maximum	Minimum	Mean	Standard Deviation
TP (mg/L)	1050	0.20	0.01	0.07	0.02
TN (mg/L)	1050	8.45	0.49	1.60	1.05
NH <sub>3</sub> -N (mg/L)	1050	3.93	0.05	0.52	0.51
COD <sub>Mn</sub> (mg/L)	1050	6.70	2.16	4.07	0.83
DO (mg/L)	1050	15.54	0.76	7.00	2.99
EC (μS/cm)	1050	407.83	108.00	244.12	54.83
TB (NTU)	1050	125.00	6.32	30.94	17.94

The Sheshui River, a tributary of the Yangtze River, originates from Sanjiao Mountain in Dawu County, Hubei Province, and flows through Dawu County, Hong'an County, and Huangpi District before merging into the Yangtze River near Wuhan. The Sheshui River is 142.14 km long and has a watershed area of 2312 km<sup>2</sup>. As a crucial drinking water source, the Sheshui River supplies potable water to approximately 700,000 people along its banks. Therefore, predicting the temporal variations in water quality indicators in the Sheshui River is of significant importance for understanding water quality dynamics and supporting regional economic development.

### 2.2. Feature Selection Based on Random Forest (RF)

In complex water quality prediction tasks, selecting appropriate features is crucial for improving the predictive performance of the model. RF is widely used for feature selection due to its advantages in handling high-dimensional data and capturing complex relationships between variables [28,29]. RF is a simple and efficient ensemble learning method based on decision trees, and its feature importance calculation module can determine the importance of each feature and rank them, providing the advantage of reduced risk of overfitting during the prediction process [30]. Compared to boosting techniques (such as LightGBM, XGBoost, and CatBoost), RF utilizes its out-of-bag (OOB) evaluation mechanism, which more effectively avoids model bias, making the feature selection results more representative. Additionally, RF can identify the contribution of each feature to the overall model prediction, and this approach is more stable than other gradient boosting methods, especially in cases where water quality data are commonly imbalanced and noisy.

By using RF, the importance of other water quality indicators relative to the target indicator can be calculated, enabling the identification and utilization of the most contributive features for water quality prediction, thereby laying a solid foundation for subsequent model construction [31]. The implementation steps of the RF algorithm are as follows.

Step 1: Randomly sample with replacement from dataset  $NNN$ , selecting  $n$  ( $n < N$ ) samples for a total of  $k$  iterations to generate  $k$  training sets, with the remaining data forming the OOB dataset.

Step 2: Construct and train a decision tree for each of the  $k$  training sets, resulting in  $K$  decision trees, and calculate the voting results of each decision tree on its corresponding OOB dataset.

Step 3: Sequentially shuffle the features  $X_i$  ( $i = 1, 2, 3, \dots, M$ , where  $M$  is the number of features in the dataset) in the OOB dataset to create new OOB samples, then use the established RF model to vote on the new OOB samples and obtain the voting results.

Step 4: Calculate the importance score  $IMP_i$  for each feature  $X_i$  according to Equation (1) and select features with  $IMP_i > 0.1$  as the input variables.

$$IMP_i = \frac{1}{K} \sum_{k=1}^K (S_k + S_{k,i}) \quad (1)$$

where  $IMP_i$  is the importance score,  $K$  is the number of decision trees,  $S_k$  is the voting result of each decision tree on its corresponding OOB dataset, and  $S_{k,i}$  is the voting result after the features have been randomly shuffled in the OOB dataset [32].

### 2.3. Time-Varying Filtered Empirical Mode Decomposition (TVF-EMD)

In the preprocessing of water quality data, issues such as signal non-stationarity and noise often lead to a decline in prediction accuracy. The TVF-EMD offers an effective solution for handling these complex time series data. TVF-EMD effectively addresses the challenges of mode mixing and low efficiency at low sampling rates encountered in the traditional Empirical Mode Decomposition (EMD) process [33,34]. It does not require manual setting of the mode function  $K$  and can filter signals during the decomposition process. The TVF-EMD method utilizes non-uniform B-spline approximations as time-varying filters to select the cutoff frequencies, improves the stopping criteria, and offers

complete adaptability [35]. It is suitable for analyzing linear and non-stationary signals and has gradually been applied to the decomposition of water quality data. The steps for implementing TVF-EMD are as follows.

Step 1: Use B-spline approximations to calculate the local cutoff frequencies, which can be expressed as:

$$\varphi'_{bis}(t) = \frac{\varphi'_1(t) + \varphi'_2(t)}{2} \tag{2}$$

where  $\varphi'_{bis}(t)$  is the cutoff frequency, and  $\varphi'_1(t)$  and  $\varphi'_2(t)$  represent the instantaneous frequency components. To address the mode mixing issue, the local cutoff frequencies  $\varphi'_{bis}(t)$  need to be adjusted.

Step 2: Reconstruct the signal  $\varphi'_{bis}(t)$ , obtaining a new signal  $g(t)$ , and further filter the input signal using time-varying filters to derive the local mean function.

Step 3: Check whether the residual signal meets the stopping criterion. When the stopping criterion  $\theta_t$  (Equation (3)) is less than or equal to the set threshold  $\varepsilon$ , an IMF can be determined, indicating that the stopping condition is satisfied.

$$\theta_t = \frac{B_{Loughlin}(t)}{\varphi_{avg}(t)} \tag{3}$$

where  $B_{Loughlin}(t)$  represents Loughlin’s instantaneous bandwidth, and  $\varphi_{avg}(t)$  is the weighted average of the instantaneous frequencies.

#### 2.4. Sparrow Search Algorithm (SSA)

The SSA, as an emerging swarm intelligence optimization algorithm, has gained significant attention due to its performance in global search and local optimization [36]. The core concept of SSA is to divide the sparrow population into three roles: discoverers, joiners, and scouters. Each role has different tasks within the algorithm, and they work together to complete the optimization process [37–39].

The primary task of the discoverers is to search for food, providing foraging areas and food sources for the entire sparrow population. In each iteration of the algorithm, discoverers continuously explore new solutions by updating their positions. The position update is as shown in Equation (4).

$$X_{i,j}^{t+1} = \begin{cases} X_{i,j}^t \cdot \exp(-\frac{i}{\alpha t_{max}}) & \text{if } R_2 < ST \\ X_{i,j}^t + QL & \text{if } R_2 \geq ST \end{cases} \tag{4}$$

where  $X_{i,j}^t$  represents the position information of the  $j$ -th dimension of the  $i$ -th sparrow in the  $t$ -th iteration;  $t_{max}$  is the maximum number of iterations;  $\alpha$  is a random number in the range  $[0, 1]$ ;  $ST$  is a constant within the range  $[0.5, 1]$ , representing the safety threshold;  $R_2$  is a random value within the range  $[0, 1]$ , representing the alert value. When  $R_2 < ST$ , the environment is safe for foraging, and the searchers continue searching for food. Conversely, when  $R_2 \geq ST$ , the scouters issue an alert, and the searchers stop searching and immediately flee to a safe area;  $Q$  is a random number following a normal distribution, and  $L$  is a  $1 \times d$  matrix where all elements are 1 [40].

The main task of the joiners is to follow the discoverers to obtain food and adjust their positions during the process to seek the optimal solution. The position update for joiners is as shown in Equation (5).

$$X_i^{t+1} = \begin{cases} Q \cdot \exp(\frac{X_{worst}^t - X_{i,j}^t}{i^2}) & \text{if } i > \frac{n}{2} \\ X_p^{t+1} + |X_{i,j}^t - X_p^{t+1}| \cdot A^+ \cdot L & \text{if } i \leq \frac{n}{2} \end{cases} \tag{5}$$

where  $X_{worst}^t$  is the worst individual in the  $t$ -th iteration;  $X_p^t$  is the position of the best searcher in the  $(t + 1)$ -th iteration;  $A$  is a  $1 \times d$  matrix with elements randomly assigned as

1 or  $-1$ ;  $A^+$  is a matrix that satisfies  $A^+ = AT(AAT)^{-1}$ ;  $n$  is the number of joiners. When  $i > n/2$ , it indicates that the current sparrow position is poor and it needs to fly farther to forage. When  $i \leq n/2$ , it indicates that the current sparrow position is good, and it only needs to move closer to the best-positioned sparrow [21].

The task of the scouts is to issue alerts when the group is in danger and lead the entire group to safety. The movement of the scouts is as shown in Equation (6).

$$X_{i,j}^{t+1} = \begin{cases} X_{best}^t + \beta \left| X_{i,j}^t - X_{best}^t \right| & \text{if } f_i > f_g \\ X_{i,j}^t + K \left( \frac{|X_{i,j}^t - X_{worst}^t|}{f_i - f_w + \epsilon} \right) & \text{if } f_i = f_g \end{cases} \tag{6}$$

where  $X_{best}^t$  represents the position of the best individual in the  $t$ -th iteration;  $\beta$  is a random number following a normal distribution with a mean of 0 and a standard deviation of 1;  $K$  is a random number in the range  $[-1, 1]$ , where the sign indicates the direction of the sparrow’s movement and the magnitude controls the step size;  $f_i$  is the fitness value of the current individual;  $f_g$  is the maximum fitness value in the current group; and  $f_w$  is the minimum fitness value in the current group;  $\epsilon$  is a small constant to prevent the denominator from being zero [41,42].

2.5. Variational Mode Decomposition (VMD)

VMD is an adaptive and fully non-recursive method for decomposing modes, primarily used to process complex, nonlinear, and non-stationary time series signals. VMD adaptively determines the number of modes to be decomposed, effectively separating the signal components and reducing the non-stationarity of the signal. The performance of the decomposition mainly depends on the number of modes  $K$  and the penalty factor  $\alpha$  [43].

The primary task of VMD is to decompose the original signal into multiple mode components, each corresponding to a specific frequency range. VMD achieves this by formulating a variational problem, with the steps as follows (Equations (7) and (8)).

$$\min_{\{u_k\}, \{w_k\}} \left\{ \sum_k \left\| \partial_t [(\delta_t + j/\pi t) * u_k(t)] e^{-jw_k t} \right\|_2^2 \right\} \tag{7}$$

$$s.t. \sum_{K=1}^K u_k(t) = f(t) \tag{8}$$

where  $K$  is the number of modes to be decomposed;  $u_k(t)$  is the  $k$ -th mode component after decomposition;  $w_k$  is the center frequency of the  $k$ -th mode component;  $\delta_t$  is the Dirac delta function;  $*$  denotes the convolution operator;  $f(t)$  is the original signal;  $[(\delta_t + j/\pi t) * u_k(t)]$  is the analytic signal of the original signal;  $e^{-jw_k t}$  shifts the spectrum to the base frequency band, and  $j$  is the imaginary unit [44].

To obtain the optimal solution to the VMD problem, the Lagrange multiplier  $\lambda$  and penalty factor  $\alpha$  are introduced, transforming the constrained variational problem into an unconstrained variational problem. The augmented Lagrangian function is shown in Equation (9). On this basis, the Alternating Direction Method of Multipliers (ADMM) is used to iteratively update the mode components and center frequencies, as shown in Equations (10) and (11).

$$L(\{u_k, w_k, \lambda\}) = \alpha \sum_k \left\| \partial_t [(\delta_t + j/\pi t) * u_k(t)] e^{-jw_k t} \right\|_2^2 + \left\| f(t) - \sum_k u_k(t) \right\|_2^2 + \langle \lambda(t), f(t) - \sum_k u_k(t) \rangle \tag{9}$$

$$\hat{u}_k^{n+1}(t) = IFFT \left\{ \frac{\hat{f}(w) - \sum_{i \neq k} \hat{u}_i(w) + \frac{\hat{\lambda}(w)}{2}}{1 + 2\alpha(w - w_k)^2} \right\} \tag{10}$$

$$w_k^{n+1}(t) = \frac{\int_0^\infty w |\hat{u}_k(w)|^2 dw}{\int_0^\infty |\hat{u}_k(w)|^2 dw} \tag{11}$$

where  $\hat{u}_k^{n+1}(w)$  is the Wiener filter of the residual signal;  $w$  is the frequency; *IFFT* is the inverse Fourier transform;  $\hat{f}(w)$  and  $\hat{\lambda}(w)$  are the Fourier transforms of the original signal and the  $k$ -th mode component, respectively [45].

A key limitation of VMD lies in its sensitivity to the selection of initial parameters, such as the number of modes and penalty factors, which directly affect the quality of the decomposition results. To address this issue, we employed the SSA to optimize VMD parameters. As an optimization algorithm based on swarm intelligence, SSA has strong global search capabilities, effectively avoiding local optima and ensuring the discovery of the globally optimal parameter combination for VMD, thus improving the accuracy of the decomposition. Additionally, the automated parameter optimization process using SSA reduces the subjectivity and limitations of manual tuning, further enhancing the accuracy and robustness of the model.

2.6. Neural Networks (CNN)

CNN has become a crucial tool in the field of deep learning due to their outstanding performance in image recognition and time series processing. In water quality prediction, CNN can enhance model accuracy by extracting local features [8]. The convolutional and pooling layers in a CNN model are specially designed data processing layers that alternate to better extract local features from the data and reduce the dimensionality of the features [46]. The principle of CNN is illustrated in Figure 2, and the convolution calculation formula is as shown in Equation (12).

$$T(i, j) = \sum_{k=1}^n (X_k, W_k)(i, j) \tag{12}$$

where  $T$  represents the time feature sequence after convolution;  $n$  is the length of the input data;  $X_k$ , and  $W_k$  is the convolution kernel function sequence [17].

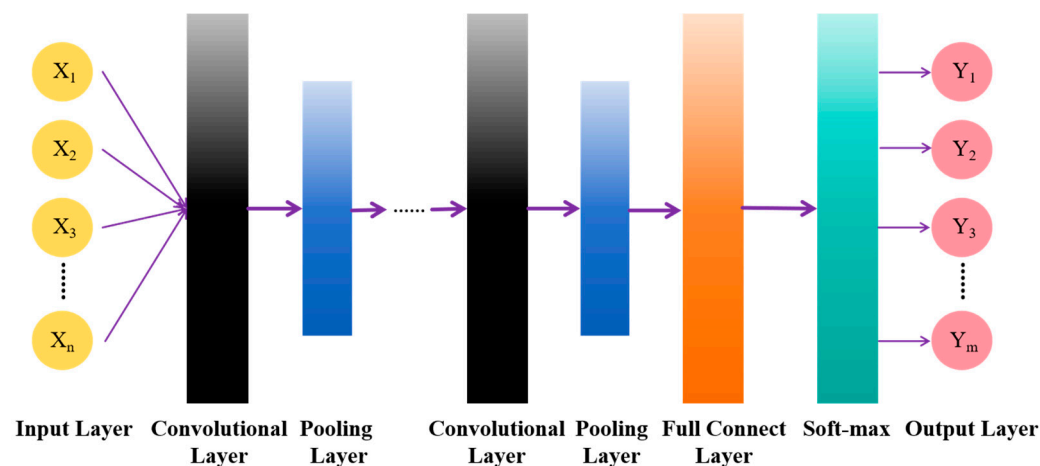


Figure 2. CNN structure.

2.7. Long Short-Term Memory (LSTM)

LSTM is a special type of Recurrent Neural Network (RNN) with excellent performance in capturing long-term dependencies in time series data. Compared to traditional RNNs, LSTM introduces three gating units (forget gate, input gate, and output gate) and a cell state, which effectively address the issues of vanishing and exploding gradients, allow-

ing the network to maintain stable training performance over longer time sequences [47]. The structure of the LSTM hidden layer is shown in Figure 3.

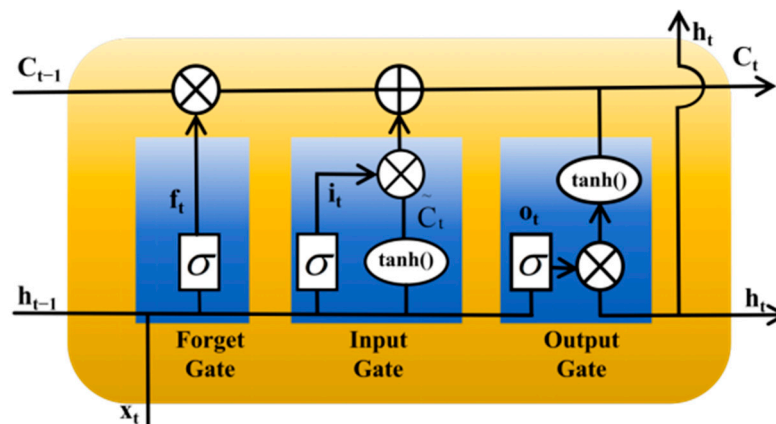


Figure 3. LSTM structure.

The forget gate determines which information to discard from the cell state. It reads the previous output  $h_{t-1}$  and the current input  $x_t$ , then generates a vector  $f_t$  through a Sigmoid function, with values ranging from 0 to 1. Each value in this vector represents the degree to which the information is retained, where values closer to 0 indicate more information is discarded, and values closer to 1 indicate more information is retained. This vector is then multiplied by the cell state  $C_{t-1}$  to determine which information should be forgotten. The forget gate calculation is expressed by Equation (13).

$$f_t = \sigma(W_f \cdot [h_{t-1}, x_t] + b_f) \tag{13}$$

where  $W_f$  is the weight matrix of the forget gate,  $b_f$  is the bias vector of the forget gate, and  $\sigma$  denotes the Sigmoid activation function.

The input gate determines how new information will be updated into the cell state. It consists of two parts: the first part is a Sigmoid layer, called the “input gate layer,” which decides which values will be updated; the second part is a  $\tanh$  layer, which generates a new candidate value vector  $\tilde{C}_t$  that will be added to the cell state. The input gate calculation process is shown in Equations (14) and (15):

$$i_t = \sigma(W_i \cdot [h_{t-1}, x_t] + b_i) \tag{14}$$

$$\tilde{C}_t = \tanh(W_c \cdot [h_{t-1}, x_t] + b_c) \tag{15}$$

where  $W_i$  and  $W_c$  are the weight matrices for the input gate and the candidate values, respectively;  $b_i$  and  $b_c$  are the corresponding bias vectors; and  $\tanh$  represents the hyperbolic tangent activation function.

The cell state is the key to LSTM, as it carries the accumulated information from a sequence of time steps. The new cell state  $C_t$  is derived by multiplying the previous cell state  $C_{t-1}$  by the forget gate’s output  $f_t$ , and then adding the result of the input gate’s output  $i_t$  multiplied by the candidate value  $\tilde{C}_t$ . This represents the memory update at the current time step, as expressed by Equation (16).

$$C_t = f_t \cdot C_{t-1} + i_t \cdot \tilde{C}_t \tag{16}$$

The output gate determines the content of the hidden state  $h_t$  at the current time step. It first computes the output based on the current cell state, then passes it through a Sigmoid layer to decide which parts of the information will be output. Finally, the cell state



is processed by a tanh function and multiplied by the output of the Sigmoid layer to obtain the final output  $h_t$ . This process is described by Equations (17) and (18).

$$o_t = \sigma(W_o[h_{t-1}, x_t] + b_o) \quad (17)$$

$$h_t = o_t \cdot \tanh(C_t) \quad (18)$$

where  $W_o$  is the weight matrix of the output gate, and  $b_o$  is the bias vector of the output gate [48].

In this study, CNN was first employed to process the data and extract important local feature relationships, reducing the dimensionality of the features. The processed data was then input into LSTM for time-series prediction. Additionally, SSA was used to optimize the parameters of LSTM, further enhancing the ability of the CNN-LSTM combination to handle complex nonlinear data. This approach improves adaptability and robustness, making the model capable of coping with dynamic water quality environments.

### 2.8. Evaluation Metrics and Model Workflow

To describe the accuracy of the model's predictions, regression model evaluation metrics were used: the coefficient of determination  $R^2$  (goodness of fit), mean absolute error (MAE), and root mean square error (RMSE) (Equations (19)–(21)).

$$R^2 = 1 - \frac{\sum_i (y'_i - y_i)^2 / n}{\sum_i (\bar{y}_i - y_i)^2 / n} \quad (19)$$

$$MAE = \frac{\sum_{i=1}^m |y_i - y'_i|}{m} \quad (20)$$

$$RMSE = \sqrt{\frac{\sum_{i=1}^m (y'_i - y_i)^2}{m}} \quad (21)$$

where  $y_i$ ,  $y'_i$  and  $\bar{y}_i$  represent the actual value, the predicted value and the mean of the actual values, respectively;  $i$  is the sample index, and  $n$  is the number of samples.

The water quality prediction coupled model developed in this study consists of three parts: feature selection (RF), dual signal decomposition (TVF-EMD and SSA-VMD), and prediction (SSA-CNN-LSTM), referred to as RF-TVSV-SCL. The workflow of the model is illustrated in Figure 4, covering all stages of data processing, model training, and evaluation. The specific steps are as follows: First, RF is used for feature selection to extract key features relevant to water quality prediction. Then, in the data preprocessing stage, TVF-EMD is applied to initially decompose the water quality data. The resulting IMF components, which may contain complex signals, are further decomposed using the SSA-VMD. Finally, the SSA is used to optimize the CNN-LSTM to predict the data processed in the previous step.

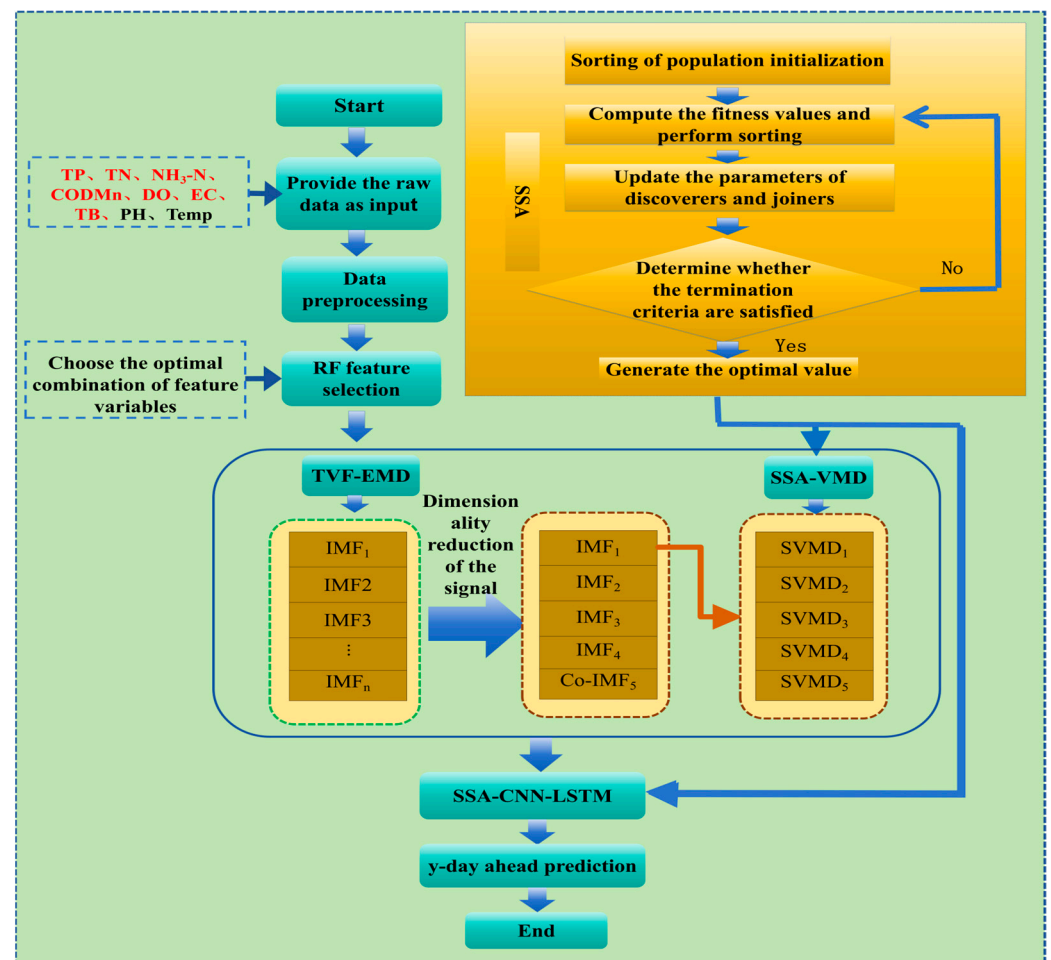


Figure 4. Workflow of the RF-TVSV-SCL model.

### 3. Results and Discussion

All models in this study were run on a Windows 11 operating system with an Intel Core i5 processor and 8GB of RAM, using Python 3.6 and the PyTorch machine learning library for model predictions.

#### 3.1. RF Feature Selection Results

Before training the model, we used the RF algorithm to calculate the importance of various features in relation to the target variable [49]. In the feature selection process, the number of decision trees (n-estimators) was set to 200, the maximum number of features (max-features) was set to 8, the minimum number of samples required to split a node (min-samples-split) was set to 2, and the minimum number of samples required for each leaf node (min-samples-leaf) was set to 1, with all other hyperparameters set to default values.

Based on the importance evaluation, features that were irrelevant or redundant to the target variable were removed, and those with an importance value greater than 0.1 were selected as the optimal input feature set for the model. The results of the feature selection are shown in Table 2.

#### 3.2. TVF-EMD Decomposition and Dimensionality Reduction

The target sequence was decomposed using the TVF-EMD algorithm. This algorithm adaptively decomposes the target sequence into IMFs of different frequencies. During the decomposition, the B-spline order was set to 26, the smoothing factor  $\alpha$  was set to 0.5, and the tolerance sift-tol was set to  $10^{-5}$ , with all other parameters kept at their default values.

**Table 2.** Feature selection results.

Indicator	Optimal Feature Set
TP	COD <sub>Mn</sub> , EC, NH <sub>3</sub> -N, TN, TP
TN	EC, NH <sub>3</sub> -N, TN
NH <sub>3</sub> -N	EC, TN, NH <sub>3</sub> -N
COD <sub>Mn</sub>	Temp, EC, TP, COD <sub>Mn</sub>
DO	Temp, pH, DO
EC	NH <sub>3</sub> -N, TP, pH, EC
TB	NH <sub>3</sub> -N, DO, TB

Sample entropy is an algorithm used to characterize the complexity of a signal’s time series [50]. A sample entropy value closer to 0 indicates that the time series is simpler and more regular. A larger sample entropy value indicates that the time series is more prone to change, meaning it is more complex, less stationary, and less predictable. After the TVF-EMD decomposition, the sample entropy of the IMFs generated for each indicator was calculated, with the results shown in Table 3. The complexity of the IMFs for each indicator gradually decreased and eventually stabilized. To improve the efficiency of the prediction model and reduce the total number of prediction sequences, dimensionality reduction was performed. Based on the entropy values, IMFs obtained from IMF5 to IMF8 (or IMF5 to IMF7) were reconstructed into Co-IMF5. The decomposition results of the indicators after signal dimensionality reduction are shown in Figure 5.

**Table 3.** Sample entropy values.

IMF Component	TP	TN	NH <sub>3</sub> -N	COD <sub>Mn</sub>	DO	EC	TB
IMF1	0.972	0.726	0.891	1.055	1.044	0.863	0.871
IMF2	0.594	0.511	0.460	0.592	0.723	0.522	0.508
IMF3	0.514	0.180	0.321	0.454	0.551	0.444	0.504
IMF4	0.414	0.135	0.257	0.289	0.418	0.195	0.305
IMF5	0.098	0.061	0.093	0.097	0.091	0.086	0.096
IMF6	0.076	0.032	0.048	0.018	0.046	0.022	0.053
IMF7	0.029	0.003	0.026	0.004	0.008	0.002	0.022
IMF8	0.001	-	0.001	-	-	-	0.006

### 3.3. SSA-VMD Decomposition Results

As shown in Table 3, among the IMFs generated by the decomposition of each indicator, IMF1 has the highest entropy value, indicating the highest complexity in the signal. To improve the predictive performance of the model and reduce the complexity of the time series, IMF1 was subjected to a secondary decomposition using SSA-VMD to reduce its entropy value [51]. The minimum sample entropy of each IMF was used as the fitness function to optimize the VMD parameters, including the penalty factor  $\alpha$  and the number of IMF components  $k$ , through the SSA.

The main parameter settings were as follows: the SSA algorithm used a population size of 20 and 50 iterations; the VMD algorithm had a penalty factor  $\alpha$  range of [500, 3000],  $k$  was set to [3, 12], the noise tolerance  $\tau$  was set to 0, the direct current component was set to 0, the initial modal center frequency was set to 1, and the control error tolerance was set to  $10^{-7}$ . Using SSA-VMD, IMF1 of all indicators was further decomposed into 5 IMFs. The decomposition results are shown in Figure 6.

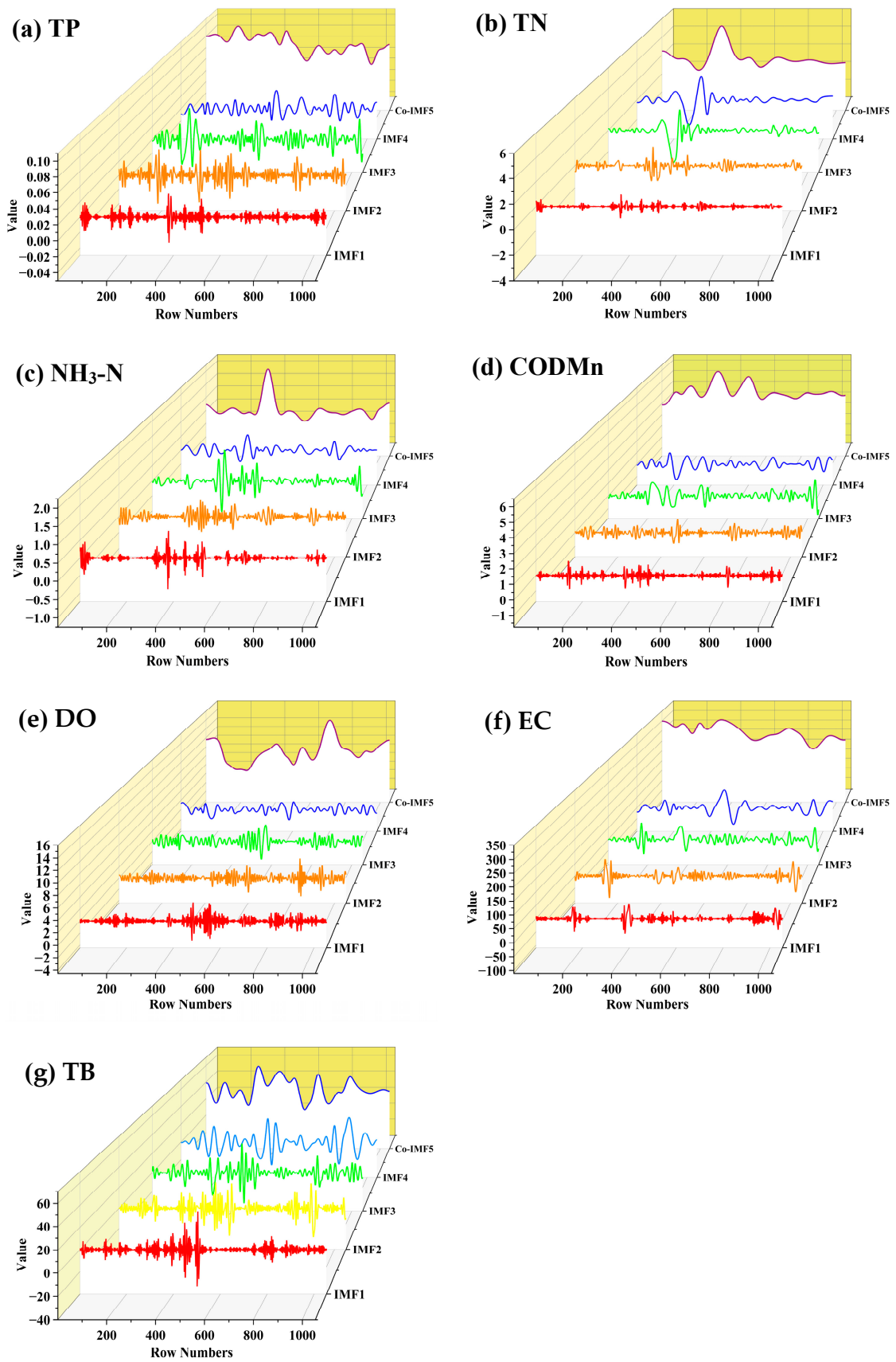


Figure 5. TVF-EMD decomposition results.

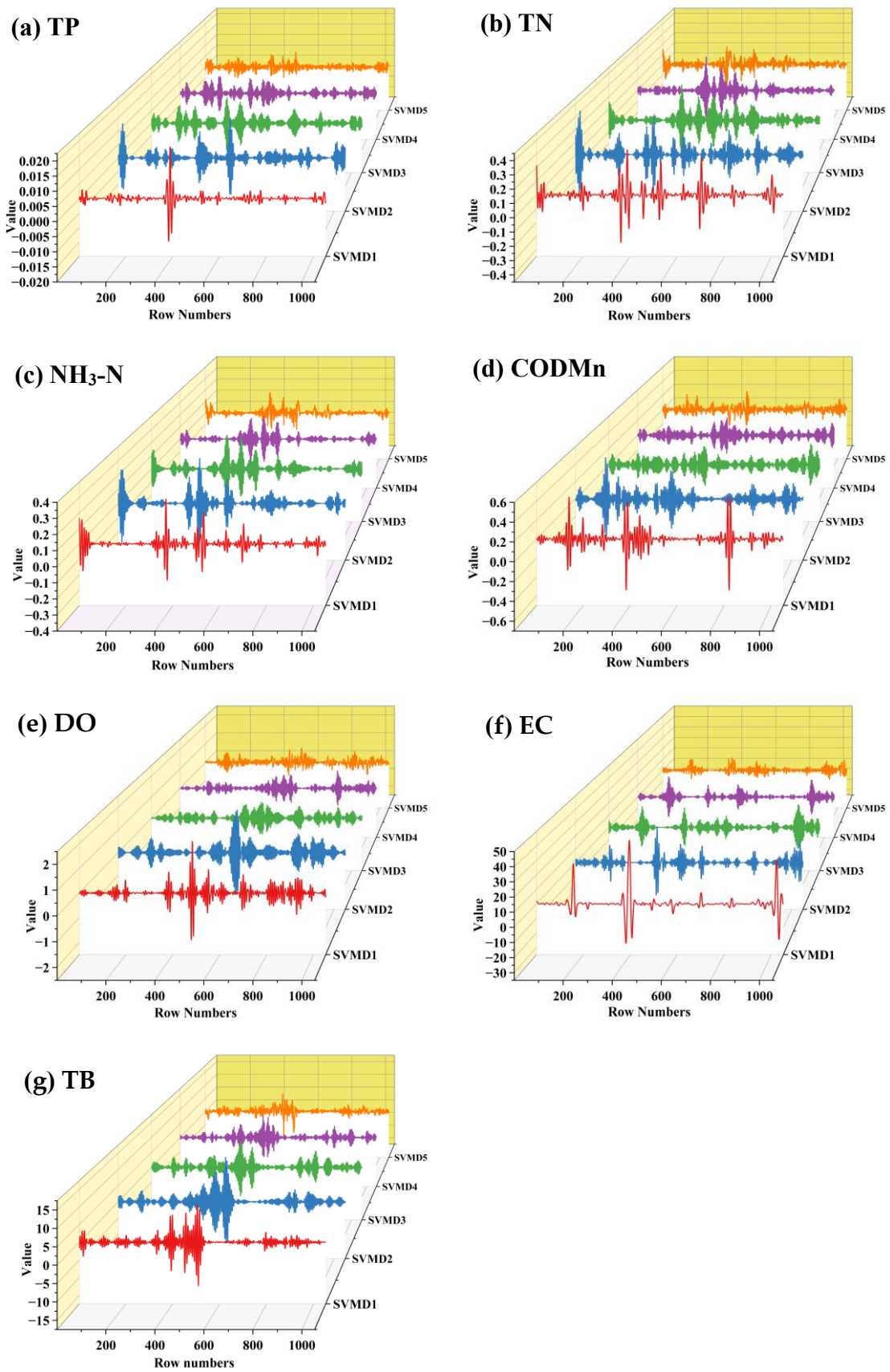


Figure 6. SSA-VMD decomposition results.

### 3.4. RF-TVSV-SCL Prediction Results

The water quality indicators after feature selection, dual decomposition and normalization, were input into the SSA-CNN-LSTM (SCL) model to predict for forecast periods of 1 d, 3 d, 5 d, and 7 d. In the CNN part of the SCL model, the fully connected layer had an output dimension of 64, the activation function was Relu, and the convolution kernel size was 9. In the LSTM part, the LSTM network and fully connected layer were used for time series data modeling, with the loss function set to Huber and the optimizer to Adam. For the SSA part, the population size was set to 20, and the maximum number of iterations was set to 300. The training loss parameters for the CNN-LSTM models of each indicator are shown in Table 4, and the hyperparameter ranges for the SSA-optimized LSTM model are provided in Table 5.

**Table 4.** Training loss parameters for CNN-LSTM models of each indicator.

Indicators	TP	TN	NH <sub>3</sub> -N	DO	COD <sub>Mn</sub>	EC	TB
Training loss threshold	0.0015	0.001	0.2	0.3	0.04	0.5	0.4

**Table 5.** SSA-Optimized LSTM model parameters.

Model	Modal Component	Count of Neurons	Dropout Rate	Batch Size	Learning Rate	Maximum Number of Training Iterations
TVF-EMD (IMF1)	S-VMD1	[32, 256]	[0, 0.3]	[8, 32]	0.001	200
	S-VMD2	[32, 256]	[0, 0.3]	[8, 32]	0.001	200
	S-VMD3	[32, 256]	[0, 0.3]	[8, 32]	0.005	150
	S-VMD4	[32, 256]	[0, 0.3]	[8, 32]	0.005	150
	S-VMD5	[32, 256]	[0, 0.3]	[8, 32]	0.005	150
TVF-EMD (IMF2)	IMF2	[32, 256]	[0, 0.3]	[8, 32]	0.005	100
TVF-EMD (IMF3)	IMF3	[16, 128]	[0, 0.3]	[8, 32]	0.01	100
TVF-EMD (IMF4)	IMF4	[16, 128]	[0, 0.3]	[8, 32]	0.01	100
TVF-EMD (Co-IMF5)	Co-IMF5	[16, 128]	[0, 0.3]	[8, 32]	0.01	100

Table S1 in Supplementary Information summarizes the predictive performance of different models on various indicators across different forecast periods. The proposed RF-TVSV-SCL model achieved  $R^2$  values of 0.93–0.96 for 1-d, 0.79–0.87 for 3-d, 0.63–0.72 for 5-d, and 0.56–0.64 for 7-d forecast periods across different water quality indicators. The MAE and RMSE values also showed similar trends, indicating that as the forecast period increases, the model's predictive accuracy gradually decreases (with increasing error). Generally, an  $R^2$  value of 0.7 or above is considered acceptable in water quality prediction. Therefore, it is evident that the model performs exceptionally well in water quality predictions for 1–3 day forecast periods, with high predictive accuracy. The high  $R^2$  values indicate that the model effectively captures water quality variations at these time scales. However, the model shows certain limitations in water quality predictions for 5-d and 7-d forecast periods.

### 3.5. Comparison of Model Performance

To fully validate the performance of the RF-TVSV-SCL model, five comparison models (TVSV-SCL, RF-SCL, RF-TV-SCL, RF-SV-SCL, and RF-CE-SCL) were established. In these model names, RF, TVSV, and SCL represent the Random Forest feature selection module, dual decomposition module based on TVF-EMD and SSA-VMD, and the prediction module based on SSA-CNN-LSTM, respectively. TV, SV, and CE represent TVF-EMD, SSA-VMD, and CEEMDAN, each being a single decomposition module. The parameters for

CEEMDAN were set as follows: standard deviation threshold = 0.2, iteration count = 500, maximum iteration count = 5000, with all other parameters set to default. All other models retained their parameter settings the same as those of the RF-TVSV-SCL model. Taking TP as an example, Figures 7 and 8 illustrate the test set predictions of different models for 1, 3, 5, and 7-day forecast periods. The prediction results for other indicators (TN, NH<sub>3</sub>-N, DO, COD<sub>Mn</sub>, EC, and TB) are presented Figures S1–S6 and Table S1 in Supplementary Information.

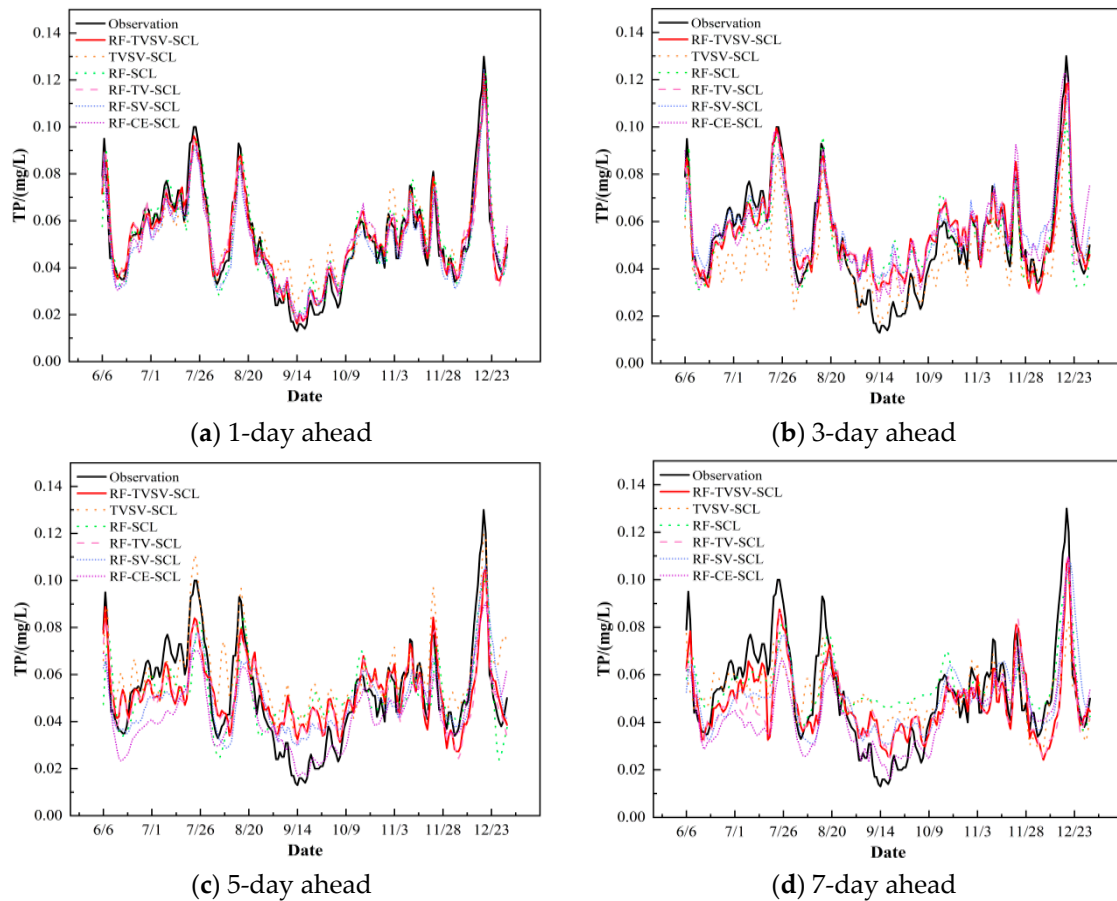


Figure 7. Water quality prediction of TP by different models.

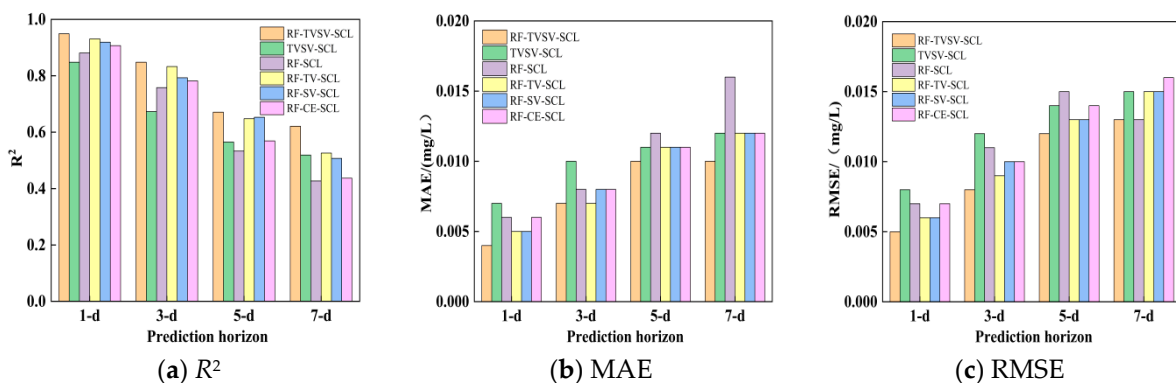


Figure 8. Comparison of different models' prediction performance for TP.

### 3.5.1. Validating the Effectiveness of Feature Selection

To reduce the training difficulty, the RF module was first used to select features from the original water quality indicators before signal decomposition and prediction.

To validate the effect of RF feature selection on improving model prediction accuracy, a comparison was conducted between the RF-TVSV-SCL model and the TVSV-SCL model.

As shown in Figure 7, it is evident that the TP prediction curve of the RF-TVSV-SCL model more closely matches the actual observation values. Additionally, as shown in Figure 8, the RF-TVSV-SCL model, after feature selection, achieved significantly higher prediction accuracy for the TP indicator at a 1-d forecast period ( $R^2 = 0.95$ , MAE = 0.004, RMSE = 0.005) compared to the TVSV-SCL model without feature selection ( $R^2 = 0.848$ , MAE = 0.007, RMSE = 0.008). The RF-TVSV-SCL model also performed better at 3-d, 5-d, and 7-d forecast periods. This demonstrates that RF feature selection effectively improves model prediction accuracy by removing irrelevant or redundant features.

### 3.5.2. Validating the Effectiveness of Dual Decomposition

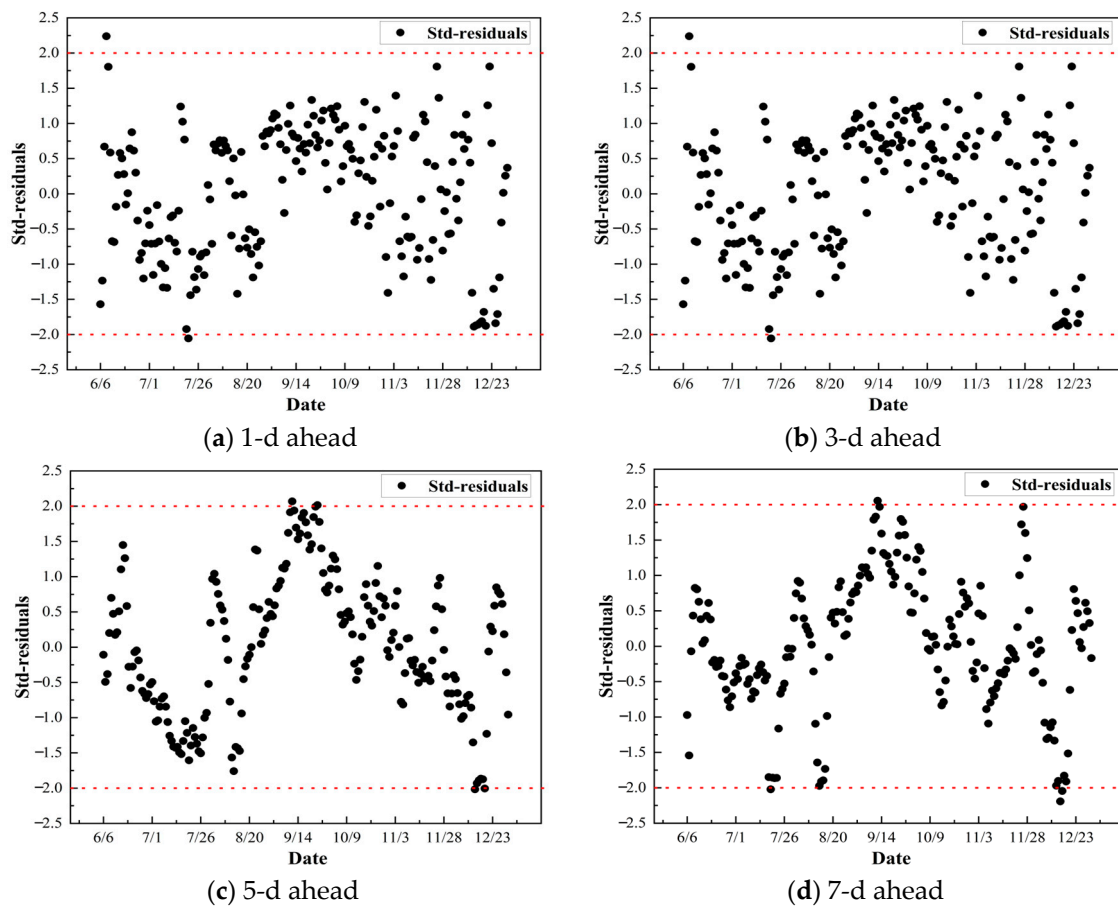
To validate the effectiveness of dual decomposition, the RF-TVSV-SCL model was compared with models containing single decomposition (RF-TV-SCL, RF-SV-SCL, and RF-CE-SCL) and a model without decomposition (RF-SCL). Overall, the TP prediction accuracy was ranked as follows: dual decomposition models > single decomposition models > no decomposition model. Compared to RF-SCL, the RF-TVSV-SCL model's  $R^2$  values improved by 7.72%, 11.87%, 25.89%, and 45.43% for 1-d, 3-d, 5-d, and 7-d forecast periods, respectively. Among the single decomposition models, the prediction accuracy for TP was ranked as follows: RF-TV-SCL > RF-SV-SCL > RF-CE-SCL. This indicates that, in terms of signal decomposition, TVF-EMD performed the best, followed by SSA-VMD, with CEEMDAN performing the weakest. TVF-EMD excels in handling nonlinear and non-stationary signals, making it the most effective in extracting information relevant to water quality prediction, leading to the highest prediction accuracy. Although SSA-VMD also has good decomposition capabilities, its performance may be affected by over-decomposition or computational complexity. While CEEMDAN provides a basic decomposition method, its weaker noise resistance and mode separation capabilities result in relatively lower prediction accuracy.

The RF-TVSV-SCL model, with dual decomposition, combines TVF-EMD and SSA-VMD, enabling more comprehensive feature extraction, effective noise suppression, and the capture of complex nonlinear relationships, thereby enhancing the model's robustness. These advantages make it significantly superior to single decomposition models and the no decomposition model in terms of prediction accuracy.

Figure 9 shows the distribution of standardized residuals for the RF-TVSV-SCL model in predicting TP across different forecast periods. In all four forecast periods, more than 95% of the residuals fell within the confidence interval of  $[-2, 2]$  (i.e., the area between the two red dashed lines in Figure 9), indicating that the model performed well in all forecast periods. However, the model demonstrates higher accuracy and stability in short-term forecasts (1–3 d). As the forecast period extends, while the majority of residuals remain within the confidence interval, the errors gradually increase, and the model's predictive stability declines, particularly in the 5-d and 7-d forecast periods.

Furthermore, based on the comparison of predicted and actual observation values in Figures S1–S6 and the prediction accuracy shown in Table S1 in Supplementary Information, the RF-TVSV-SCL model exhibited the best prediction performance among all tested models for the other six water quality indicators (TN,  $\text{NH}_3\text{-N}$ , DO,  $\text{COD}_{\text{Mn}}$ , EC, and TB). It also excelled in water quality predictions for 1–3 day forecast periods. Specifically, for the 1-d forecast period, the  $R^2$  values for these six indicators were 0.951, 0.958, 0.949, 0.953, 0.930, and 0.941, respectively; for the 3-d forecast period, the  $R^2$  values were 0.810, 0.798, 0.836, 0.820, 0.786, and 0.870, respectively; and for the 5-d forecast period, the  $R^2$  values exceeded 0.7 for all indicators except DO and EC.





**Figure 9.** Standardized residual plots of TP under different forecast periods.

Thus, the RF-TVSV-SCL model developed in this study demonstrates strong generalization capabilities, making it suitable for predicting most conventional water quality indicators and providing reliable decision support for water pollution prevention in forecast periods of at least 3 days.

### 3.6. Discussion

The RF-TVSV-SCL model developed in this study demonstrated excellent performance in water quality prediction, particularly in short-term forecasts (1-d and 3-d periods). It significantly outperformed other models in predicting key water quality indicators such as TP, TN,  $\text{NH}_3\text{-N}$ , DO,  $\text{COD}_{\text{Mn}}$ , EC, and TB. The model's superior performance in short-term forecasts can be attributed to several key factors.

First, the RF feature selection module effectively eliminated redundant or irrelevant features, simplifying the data and reducing noise interference [28]. This improved the model's training efficiency and prediction accuracy. As water quality data tends to exhibit more stability over short periods, the RF feature selection enabled the model to identify the most relevant variables associated with the prediction target, ensuring high accuracy in short-term forecasts.

Second, the dual signal decomposition methods (TVF-EMD and SSA-VMD) significantly enhanced the model's noise resistance and decomposition accuracy. TVF-EMD adaptively decomposed the nonlinear and non-stationary water quality data to extract key features [52], while SSA-VMD further optimized the decomposition results, reducing mode mixing and noise interference [53]. This dual decomposition approach effectively captured the complex relationships within the water quality data, allowing the model to maintain high prediction accuracy in the short term.

However, the model's performance declined in the 5-d and 7-d forecast periods, which may be attributed to the increasing unpredictability of external environmental factors (e.g., climate change, pollution inputs, fluctuations in river flow) as the forecast period extends. These factors have a greater impact on water quality variability over longer timeframes, reducing the model's prediction accuracy in long-term forecasts. Additionally, the performance of the signal decomposition methods in long-term forecasts presents certain limitations. As the time series extends, the increasing nonlinearity and complexity of the data may lead to decreased decomposition accuracy or over-decomposition, thereby affecting prediction precision.

To further enhance long-term prediction performance, future research could incorporate additional external environmental data, such as meteorological data, watershed hydrological data and watershed pollution sources, to improve the model's ability to capture long-term trends. Moreover, continued optimization of the signal decomposition algorithms would help reduce noise interference and further improve the model's long-term prediction accuracy.

#### 4. Conclusions

To address the nonlinearities and complexities in time-series water quality data, this paper proposes a hybrid model, RF-TVSV-SCL, to enhance the accuracy of surface water quality predictions. TP, TN, NH<sub>3</sub>-N, DO, COD<sub>Mn</sub>, EC, and TB data from the Sheshui River in the Yangtze River Basin were used for 1, 3, 5, and 7-d forecast periods, and the results were compared with five other hybrid models. The comparison reveals the following advantages and features of the proposed model: (1) The RF feature selection module effectively eliminates irrelevant or redundant features, improving the prediction accuracy of TP by 11.9%, 26.0%, 18.8%, and 19.7% for 1, 3, 5, and 7-day forecast periods, respectively, compared to models without feature selection. (2) The dual signal decomposition process enhances the model's noise resistance, decomposition accuracy, and reduces the risk of mode mixing, resulting in an increase in TP prediction accuracy by 7.72%, 11.87%, 25.89%, and 45.43%. (3) The model precisely captures short-term water quality changes, demonstrating superior prediction performance for the 1-d and 3-d forecast periods. (4) The model is applicable to multiple water quality indicators, showing excellent generalization performance. This model compensates for the limitations of single algorithms in water quality prediction and demonstrates strong adaptability and application potential, providing valuable decision support for pollution control and water environment planning.

However, the model still has some limitations, such as slightly weaker long-term forecasting ability (5 and 7-d periods). The nonlinearities and complexities in water quality data are relatively stable in the short term, but as the forecast period extends, more unpredictable external factors (e.g., climate change, pollution sources) may arise, making it difficult for the model to capture these changes, which leads to decreased prediction accuracy over the long term. In the future, external environmental factors, such as meteorological, hydrological, and pollution sources, could be incorporated as input variables to enhance the model's ability to identify long-term trends. Further optimization of signal decomposition algorithms would also help reduce noise and improve accuracy.

**Supplementary Materials:** The following supporting information can be downloaded at: <https://www.mdpi.com/article/10.3390/w16213099/s1>.

**Author Contributions:** Methodology, Y.B. and M.P.; validation, Y.B. and M.W.; formal analysis, Y.B. and M.P.; data curation, M.W.; writing—original draft preparation, Y.B. and M.P.; writing—review and editing, M.W.; supervision, M.W.; funding acquisition, M.W. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research was funded by the National Key R&D Program of China (2023YFC3205600) and the Open Fund for Hebei Province Key Laboratory of Sustained Utilization & Development of Water Recourse (HSZYL2022001).

**Data Availability Statement:** The data presented in this study are available upon request from the corresponding author. The data are not publicly available due to station observation data in the research area being somehow confidential, so some site data are not allowed to be open to the public. For those data and software that can be opened to the public, the URL links were shown in the manuscript. Thank you very much for your kind understanding.

**Conflicts of Interest:** Author Yifan Bai was employed by Yellow River Engineering Consulting Corporation Limited. The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

## References

1. Du, X.; Wu, G.; Xu, D. Prediction Methods Analysis of the Water Quality Based on the ARMA Model. *Chin. Agric. Sci. Bull.* **2013**, *29*, 221–224.
2. Luo, X.; He, Y.; Liu, P.; Li, W. Water Quality Prediction Using an ARIMA-SVR Hybrid Model. *J. Yangtze River Sci. Res. Inst.* **2020**, *37*, 21–27.
3. He, M.; Wu, S.; Huang, B.; Kang, C.; Gui, F. Prediction of Total Nitrogen and Phosphorus in Surface Water by Deep Learning Methods Based on Multi-Scale Feature Extraction. *Water* **2022**, *14*, 1643. [[CrossRef](#)]
4. Song, C.; Chen, X.H. Performance Comparison of Machine Learning Models for Annual Precipitation Prediction Using Different Decomposition Methods. *Remote Sens.* **2021**, *13*, 1018. [[CrossRef](#)]
5. Cho, K.; Kim, Y. Improving streamflow prediction in the WRF-Hydro model with LSTM networks. *J. Hydrol.* **2022**, *605*, 127297. [[CrossRef](#)]
6. Sun, K.; Hu, L.; Sun, J.; Cao, X. Enhancing groundwater level prediction accuracy at a daily scale through combined machine learning and physics-based modeling. *J. Hydrol. Reg. Stud.* **2023**, *50*, 101577. [[CrossRef](#)]
7. Zhang, Y.; Gu, Z.; The, J.V.G.; Yang, S.X.; Gharabaghi, B. The Discharge Forecasting of Multiple Monitoring Station for Humber River by Hybrid LSTM Models. *Water* **2022**, *14*, 1794. [[CrossRef](#)]
8. 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](#)]
9. Song, J.; Meng, H.; Kang, Y.; Zhu, M.; Zhu, Y.; Zhang, J. A method for predicting water quality of river basin based on OVMD-GAT-GRU. *Stoch. Environ. Res. Risk Assess.* **2024**, *38*, 339–356. [[CrossRef](#)]
10. Li, W.; Wu, H.; Zhu, N.; Jiang, Y.; Tan, J.; Guo, Y. Prediction of dissolved oxygen in a fishery pond based on gated recurrent unit (GRU). *Inf. Process. Agric.* **2021**, *8*, 185–193. [[CrossRef](#)]
11. Tian, Q.; Luo, W.; Guo, L. Water quality prediction in the Yellow River source area based on the DeepTCN-GRU model. *J. Water Process Eng.* **2024**, *59*, 105052. [[CrossRef](#)]
12. Fu, Y.; Hu, Z.; Zhao, Y.; Huang, M. A Long-Term Water Quality Prediction Method Based on the Temporal Convolutional Network in Smart Mariculture. *Water* **2021**, *13*, 2907. [[CrossRef](#)]
13. Sheng, S.; Lin, K.; Zhou, Y.; Chen, H.; Luo, Y.; Guo, S.; Xu, C.Y. Exploring a multi-output temporal convolutional network driven encoder-decoder framework for ammonia nitrogen forecasting. *J. Environ. Manag.* **2023**, *342*, 118232. [[CrossRef](#)]
14. Barzegar, R.; Aalami, M.T.; Adamowski, J. Short-term water quality variable prediction using a hybrid CNN-LSTM deep learning model. *Stoch. Environ. Res. Risk Assess.* **2020**, *34*, 415–433. [[CrossRef](#)]
15. Wang, Z.; Man, Y.; Hu, Y.; Li, J.; Hong, M.; Cui, P. A deep learning based dynamic COD prediction model for urban sewage. *Environ. Sci.-Water Res. Technol.* **2019**, *5*, 2210–2218. [[CrossRef](#)]
16. Yan, J.; Gao, Q.; Yu, Y.; Chen, L.; Xu, Z.; Chen, J. Combining knowledge graph with deep adversarial network for water quality prediction. *Environ. Sci. Pollut. Res.* **2023**, *30*, 10360–10376. [[CrossRef](#)]
17. Yang, Y.; Xiong, Q.; Wu, C.; Zou, Q.; Yu, Y.; Yi, H.; Gao, M. A study on water quality prediction by a hybrid CNN-LSTM model with attention mechanism. *Environ. Sci. Pollut. Res.* **2021**, *28*, 55129–55139. [[CrossRef](#)]
18. 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](#)]
19. Zhang, M.; Zhang, X.; Qiao, W.; Lu, Y.; Chen, H. Forecasting of runoff in the lower Yellow River based on the CEEMDAN-ARIMA model. *Water Supply* **2023**, *23*, 1434–1450. [[CrossRef](#)]
20. Zhu, X.; Guo, H.; Huang, J.J.; Tian, S.; Zhang, Z. A hybrid decomposition and Machine learning model for forecasting Chlorophyll-a and total nitrogen concentration in coastal waters. *J. Hydrol.* **2023**, *619*, 129207. [[CrossRef](#)]
21. Zhang, X.; Wang, X.; Li, H.; Sun, S.; Liu, F. Monthly runoff prediction based on a coupled VMD-SSA-BiLSTM model. *Sci. Rep.* **2023**, *13*, 13149. [[CrossRef](#)] [[PubMed](#)]
22. Zuo, H.; Gou, X.; Wang, X.; Zhang, M. A Combined Model for Water Quality Prediction Based on VMD-TCN-ARIMA Optimized by WSWOA. *Water* **2023**, *15*, 4227. [[CrossRef](#)]
23. Yang, X.; Chen, Z.; Qin, M. Monthly Runoff Prediction Via Mode Decomposition-Recombination Technique. *Water Resour. Manag.* **2023**, *38*, 269–286. [[CrossRef](#)]
24. Cao, W.; Su, Y.; Zeng, Y.; Liu, H.; Liu, L. Water Quality Prediction Model Based on EEMD-LSTM-SVR. *Syst. Eng.* **2023**, *41*, 1–12.

25. Jiao, J.; Ma, Q.; Huang, S.; Liu, F.; Wan, Z. A hybrid water quality prediction model based on variational mode decomposition and bidirectional gated recursive unit. *Water Sci. Technol.* **2024**, *89*, 2273–2289. [[CrossRef](#)] [[PubMed](#)]
26. Wang, H.; Sun, J.; Sun, J.; Wang, J. Using Random Forests to Select Optimal Input Variables for Short-Term Wind Speed Forecasting Models. *Energies* **2017**, *10*, 1522. [[CrossRef](#)]
27. Song, C.; Yao, L.; Hua, C.; Ni, Q. A water quality prediction model based on variational mode decomposition and the least squares support vector machine optimized by the sparrow search algorithm (VMD-SSA-LSSVM) of the Yangtze River, China. *Environ. Monit. Assess.* **2021**, *193*, 363. [[CrossRef](#)]
28. Breiman, L. Random forests. *Mach. Learn.* **2001**, *45*, 5–32. [[CrossRef](#)]
29. Li, J.; Alvarez, B.; Siwabessy, J.; Tran, M.; Huang, Z.; Przeslawski, R.; Radke, L.; Howard, F.; Nichol, S. Application of random forest, generalised linear model and their hybrid methods with geostatistical techniques to count data: Predicting sponge species richness. *Environ. Model. Softw.* **2017**, *97*, 112–129. [[CrossRef](#)]
30. Miao, D.; Yao, K.; Wang, W.; Liu, L.; Sui, X. Risk prediction of coal mine rock burst based on machine learning and feature selection algorithm. In *Georisk: Assessment and Management of Risk for Engineered Systems and Geohazards*; Taylor & Francis: Abingdon, UK, 2024. [[CrossRef](#)]
31. Wang, Z. Research on Feature Selection Methods based on Random Forest. *Teh. Vjesn. Teh. Gaz.* **2023**, *30*, 623–633. [[CrossRef](#)]
32. Lei, L.; Shao, S.; Liang, L. An evolutionary deep learning model based on EWKM, random forest algorithm, SSA and BiLSTM for building energy consumption prediction. *Energy* **2024**, *288*, 129795. [[CrossRef](#)]
33. Li, H.; Li, Z.; Mo, W. A time varying filter approach for empirical mode decomposition. *Signal Process.* **2017**, *138*, 146–158. [[CrossRef](#)]
34. Jamei, M.; Ali, M.; Malik, A.; Karbasi, M.; Rai, P.; Yaseen, Z.M. Development of a TVF-EMD-based multi-decomposition technique integrated with Encoder-Decoder-Bidirectional-LSTM for monthly rainfall forecasting. *J. Hydrol.* **2023**, *617*, 129105. [[CrossRef](#)]
35. Jamei, M.; Ali, M.; Karbasi, M.; Karimi, B.; Jahannemaei, N.; Farooque, A.A.; Yaseen, Z.M. Monthly sodium adsorption ratio forecasting in rivers using a dual interpretable glass-box complementary intelligent system: Hybridization of ensemble TVF-EMD-VMD, Boruta-SHAP, and eXplainable GPR. *Expert Syst. Appl.* **2024**, *237*, 121512. [[CrossRef](#)]
36. Feng, B.-F.; Xu, Y.-S.; Zhang, T.; Zhang, X. Hydrological time series prediction by extreme machine learning and sparrow search algorithm. *Water Supply* **2022**, *22*, 3143–3157. [[CrossRef](#)]
37. Guo, N.; Wang, Z. A combined model based on sparrow search optimized BP neural network and Markov chain for precipitation prediction in Zhengzhou City, China. *Aqua-Water Infrastruct. Ecosyst. Soc.* **2022**, *71*, 782–800. [[CrossRef](#)]
38. He, J.; Liu, S.-M.; Chen, H.-T.; Wang, S.-L.; Guo, X.-Q.; Wan, Y.-R. Flood Control Optimization of Reservoir Group Based on Improved Sparrow Algorithm (ISSA). *Water* **2023**, *15*, 132. [[CrossRef](#)]
39. Li, B.-J.; Sun, G.-L.; Li, Y.-P.; Zhang, X.-L.; Huang, X.-D. A hybrid model of variational mode decomposition and sparrow search algorithm-based least square support vector machine for monthly runoff forecasting. *Water Supply* **2022**, *22*, 5698–5715. [[CrossRef](#)]
40. Zhen, L.; Barbulescu, A. Comparative Analysis of Convolutional Neural Network-Long Short-Term Memory, Sparrow Search Algorithm-Backpropagation Neural Network, and Particle Swarm Optimization-Extreme Learning Machine Models for the Water Discharge of the Buzău River, Romania. *Water* **2024**, *16*, 289. [[CrossRef](#)]
41. Yao, Z.; Wang, Z.; Cui, X.; Zhao, H. Research on multi-objective optimal allocation of regional water resources based on improved sparrow search algorithm. *J. Hydroinformatics* **2023**, *25*, 1413–1437. [[CrossRef](#)]
42. Zhang, C.; Fu, S.; Ou, B.; Liu, Z.; Hu, M. Prediction of Dam Deformation Using SSA-LSTM Model Based on Empirical Mode Decomposition Method and Wavelet Threshold Noise Reduction. *Water* **2022**, *14*, 3380. [[CrossRef](#)]
43. Guo, S.; Sun, S.; Zhang, X.; Chen, H.; Li, H. Monthly precipitation prediction based on the EMD-VMD-LSTM coupled model. *Water Supply* **2023**, *23*, 4742–4758. [[CrossRef](#)]
44. He, X.; Luo, J.; Zuo, G.; Xie, J. Daily Runoff Forecasting Using a Hybrid Model Based on Variational Mode Decomposition and Deep Neural Networks. *Water Resour. Manag.* **2019**, *33*, 1571–1590. [[CrossRef](#)]
45. Qi, J.; Su, X.; Zhang, G.; Zhang, T. Research on monthly runoff prediction of VMD-LSTM model in different forecast periods. *Agric. Res. Arid Areas* **2022**, *40*, 258–267.
46. Pan, D.; Zhang, Y.; Deng, Y.; The, J.V.G.; Yang, S.X.; Gharabaghi, B. Dissolved Oxygen Forecasting for Lake Erie’s Central Basin Using Hybrid Long Short-Term Memory and Gated Recurrent Unit Networks. *Water* **2024**, *16*, 707. [[CrossRef](#)]
47. Pang, J.; Luo, W.; Yao, Z.; Chen, J.; Dong, C.; Lin, K. Water Quality Prediction in Urban Waterways Based on Wavelet Packet Denoising and LSTM. *Water Resour. Manag.* **2024**, *38*, 2399–2420. [[CrossRef](#)]
48. Wang, T.; Chen, W.; Tang, B. Water quality prediction using ARIMA-SSA-LSTM combination model. *Water Supply* **2024**, *24*, 1282–1297. [[CrossRef](#)]
49. Alshawabkeh, S.; Wu, L.; Dong, D.; Cheng, Y.; Li, L.; Alanaqreh, M. Automated Pavement Crack Detection Using Deep Feature Selection and Whale Optimization Algorithm. *Cmc-Comput. Mater. Contin.* **2023**, *77*, 63–77. [[CrossRef](#)]
50. Richman, J.S.; Moorman, J.R. Physiological time-series analysis using approximate entropy and sample entropy. *Am. J. Physiol.-Heart Circ. Physiol.* **2000**, *278*, H2039–H2049. [[CrossRef](#)]
51. Gao, X.; Guo, W.; Mei, C.; Sha, J.; Guo, Y.; Sun, H. Short-term wind power forecasting based on SSA-VMD-LSTM. *Energy Rep.* **2023**, *9*, 335–344. [[CrossRef](#)]

- 
52. Huang, N.E.; Shen, Z.; Long, S.R.; Wu, M.L.C.; Shih, H.H.; Zheng, Q.N.; Yen, N.C.; Tung, C.C.; Liu, H.H. The empirical mode decomposition and the Hilbert spectrum for nonlinear and non-stationary time series analysis. *Proc. R. Soc. A-Math. Phys. Eng. Sci.* **1998**, *454*, 903–995. [[CrossRef](#)]
  53. Dragomiretskiy, K.; Zosso, D. Variational Mode Decomposition. *IEEE Trans. Signal Process.* **2014**, *62*, 531–544. [[CrossRef](#)]

**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.