




OPEN

Application of hybrid improved temporal convolution network model in time series prediction of river water quality

Yankun Hu^{1,2}, Li Lyu^{1,2}, Ning Wang^{1,2}, Xiaolei Zhou^{1,2} & Meng Fang^{1,2}

Time series prediction of river water quality is an important method to grasp the changes of river water quality and protect the river water environment. However, due to the time series data of river water quality have strong periodicity, seasonality and nonlinearity, which seriously affects the accuracy of river water quality prediction. In this paper, a new hybrid deep neural network model is proposed for river water quality prediction, which is integrated with Savitaky-Golay (SG) filter, STL time series decomposition method, Self-attention mechanism, and Temporal Convolutional Network (TCN). The SG filter can effectively remove the noise in the time series data of river water quality, and the STL technology can decompose the time series data into trend, seasonal and residual series. The decomposed trend series and residual series are input into the model combining the Self-attention mechanism and TCN respectively for training and prediction. In order to verify the proposed model, this study uses opensource water quality data and private water quality data to conduct experiments, and compares with other water quality prediction models. The experimental results show that our method achieves the best prediction results in the water quality data of two different rivers.

Rivers are the most common and basic source of water for many organisms, and play an important role in domestic water use, agricultural irrigation, and industrial development¹. But due to the interaction between rivers and the surrounding environment, as well as the exchange of urban, industrial and agricultural pollutants along the way, the problem of river pollution is becoming more and more serious². Water quality prediction and assessment are essential for the protection of human and environmental health and for effective and sustainable water resource management³. Although water quality testing can be performed using traditional techniques and methods, such methods are usually time-consuming, expensive, and less accurate. In recent years, the technology of water quality modeling and prediction based on machine learning and deep learning methods has been widely used and achieved better prediction results⁴.

Traditional models based on statistical analysis methods mainly include multiple linear regression (MLR), Autoregressive (AR), Autoregressive integrated Moving Average (ARIMA) and SARIMA models, etc.^{5,6}. Jiang Wu et al.⁷ proposed a water quality prediction method combining ARIMA and clustering model, and taking the water quality monitoring data of a basin as a sample, the total phosphorus (TP) index of water quality was selected as the prediction object, and the water quality change in the basin was successfully predicted. Mohamed Elhag et al.⁸ used the adjusted ARIMA and SARIMA models to predict water quality parameters, and verified that the SARIMA model could effectively predict water quality parameters with seasonal characteristics. However, these models cannot capture the nonlinear characteristics in the data, resulting in low prediction accuracy and poor applicability of the model.

Machine learning regression analysis methods mainly include Support vector regression (SVR), Random forest regression (RFR), Bayesian Network (BN), Decision tree (DT), artificial neural network (ANN), BP neural network, etc.^{9,10}. Theyazn H et al.¹¹ carried out water quality prediction research based on machine learning algorithms such as Support vector Machine (SVM), K-Nearest Neighbor (K-NN) and Naive Bayes. S. Ayesha Jasmin et al.¹² developed dissolved oxygen prediction models using three popular machine learning algorithms including Random Forest (RF), Adaboost and deep neural networks. Such models can deal with nonlinear features in data to a certain extent, but the prediction ability of the model is limited.

¹Shenyang Institute of Computing Technology, Chinese Academy of Sciences, Shenyang 110168, Liaoning, China. ²University of Chinese Academy of Sciences, Beijing 100049, China. ✉email: wangning@sict.ac.cn

The application of deep learning models in water quality prediction mainly focuses on recurrent neural network (RNN), LSTM, Bi-LSTM, GRU, CNN models, etc. With the development of technology, more and more hybrid deep learning models have been developed and applied, such as CNN-LSTM, Attention-Bi-LSTM model, etc.¹³. Sakshi Khullar et al.¹⁴ used the deep learning Bi-LSTM method to predict the water quality of the Yamuna River in India. Yurong Yang et al.¹⁵ proposed a water quality prediction model combining convolutional neural network (CNN), Long Short-Term memory network (LSTM) and Attention mechanism, which has a strong ability to solve nonlinear time series prediction problems. Although the feature extraction ability and long-term memory ability of the model are further improved by integrating with CNN, Attention and other models, the problems of gradient disappearance and long training time are still not effectively solved.

In order to solve the problems existing in the above models and improve the accuracy of water quality prediction, we first used the SG filter¹⁶ to eliminate the noise in the original water quality data, and then decomposed the water quality data into trend, seasonal and residual series based on the STL time series decomposition method¹⁷. The decomposed trend and residual series will be separately used for model training and prediction, so as to better realize the extraction of features. Bai et al.¹⁸ proposed the TCN model in 2018, which introduces causal convolution, dilated convolution, and residual blocks. Compared with the RNN model, TCN does not have the gradient vanishing problem and has a longer memory capacity; TCN supports parallel computation, which means that each weight in each layer can be updated simultaneously at each time step, significantly improving the model computational efficiency¹⁹. The TCN model integrates the feature convolution processing capability of the CNN model and the time series information mining capability of the recurrent neural network, and has been widely studied and applied in a variety of time series forecasting problems such as load forecasting²⁰ and wind speed forecasting²¹. Current research has demonstrated the superiority of TCN models over traditional machine learning algorithms as well as LSTM, GRU and other models on a variety of tasks and datasets. Therefore, in our model, TCN is used to extract the long-term dependencies in the sequences. At the same time, in order to further improve the performance of TCN model and solve the problem of local information loss of TCN model, we combine the Self-attention mechanism with the TCN model²², so that the model can pay more attention to the features that contribute more to the output, so as to have stronger feature extraction ability. In this paper, our contributions can be summarized as follows:

1. We use the SG filter to smooth the time series of river water quality, thereby eliminating the strong noise in the data and enhancing the availability of the data.
2. Based on the characteristics of river water quality data, we use STL decomposition technology to decompose the original water quality data into three sub-series: trend, seasonality and residual, which better retains the seasonality in the series and improves the prediction accuracy of the model.
3. The TCN model is improved, and the Self-attention mechanism is added to the residual block structure of TCN to further improve the feature extraction ability of the model.
4. The proposed river water quality prediction model is tested in two real river water quality data sets, and compared with other commonly used water quality prediction models, which verifies that our method can achieve the best prediction effect.

Methodology

The overall structure and workflow of the model are shown in Fig. 1. In our proposed model structure, the collected water quality time series are firstly smoothed by the SG filter, so as to eliminate the strong noise in the data. Then, the smoothed data are decomposed by STL decomposition technology to obtain three sub-series: seasonality, trend and residual. Then, the trend series and residual series obtained by decomposition are input into the improved TCN model for feature extraction, and the trend series prediction value, residual series prediction value and retained seasonal series are fused to obtain the final prediction result of the model. Finally, the predicted value was compared with the real value, and the model was evaluated by a series of evaluation criteria.

Savitzky-Golay (SG) filter

Smoothing and denoising the original time series data can effectively improve the accuracy of prediction²³. Therefore, in this study, we used the SG filter to remove the noise in the original water quality time series. The SG filter is a filtering method based on the local polynomial least squares fitting in the time domain. The biggest feature of this filter is that it can ensure the shape and width of the signal are unchanged while filtering out the noise.

A subsequence x with window size $w = 2m + 1$ can be expressed as:

$$\{x_{s-m}, \dots, x_s, \dots, x_{s+m}\}, s \in [m + 1, T - m]. \quad (1)$$

The polynomial $p(i)$ of order R used to fit the data points in the window is defined as follows:

$$p(i) = \sum_{v=0}^R a_v i^v, i \in [-m, m], \quad (2)$$

where a_v denotes the v -th coefficient of the SG filter.

Then, use the least squares method to minimize the error ϵ .

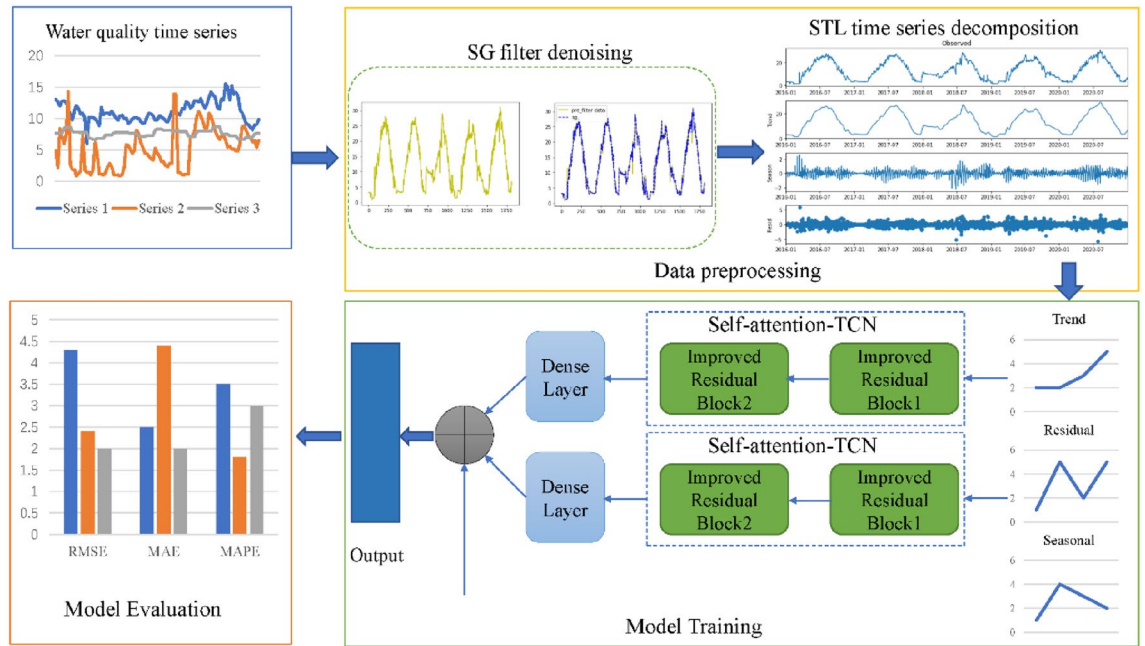


Figure 1. Overall structure and workflow of the model.

$$\epsilon = \sum_{i=-m}^m (p(i) - x_{s+i})^2. \tag{3}$$

Then we can find the best fit $p(0)$ of the window center point x_s by computing a_0 . By sliding the window, each point in the series x will become the center point in the window until all the values in the series are smoothed. Finally, we will get the smoothed sequence x_t .

STL time series decomposition method. In order to better extract the trend characteristics and nonlinear characteristics of river water quality time series, and retain the seasonal trend characteristics of the series, we used STL decomposition technology to decompose the original water quality time series. Seasonal and Trend decomposition using Loess (STL) is a very general and robust decomposition method for time series, where Loess is a method for estimating nonlinear relationships. STL aims to decompose the time series data Y_v at a certain time into trend (T_v), season (S_v) and residual (R_v), denoted as $Y_v = T_v + S_v + R_v$ ²⁴. The algorithm consists of an outer loop and an inner loop. The outer loop is mainly used to assign a robust weight to each data point through the residual, so as to reduce the influence of outliers. The inner loop is nested in the outer loop and mainly does trend fitting and periodic component calculation. The process of the k -th epoch in the inner loop is as follows²⁵:

1. Detrending. Remove the trending component from the original series, and get $Y_v - T_v^{(k)}$.
2. Cycle-subseries smoothing. Each cyclic-subseries obtained from step 1 is smoothed by Loess, resulting in a preliminary seasonal series denoted as $C_v^{(k+1)}$.
3. Low-Pass Filtering. The sequence $C_v^{(k+1)}$ obtained from step 2 is processed by moving average and Loess regression to obtain the result sequence $L_v^{(k+1)}$, which is equivalent to extracting the low-pass of the periodic subsequence.
4. Detrending of Smoothed Cycle-subseries. Calculate seasonal trends $S_v^{(k+1)} = C_v^{(k+1)} - L_v^{(k+1)}$.
5. Deseasonalizing. Subtract the periodic component, $Y_v - S_v^{(k+1)}$.
6. Trend Smoothing. Loess regression is performed on the cycled series obtained in step 5 to obtain the trend component $T_v^{(k+1)}$.

When the inner loop reaches the accuracy requirement, the outer loop starts, and the residual component $R_v^{(k+1)}$ is calculated by the estimated trend and seasonal components in the outer loop. The calculation method is as follows:

$$R_v^{(k+1)} = Y_v - T_v^{(k+1)} - S_v^{(k+1)}. \tag{4}$$

ImprovedTCN model. Temporal convolutional network (TCN) is based on the traditional one-dimensional fully convolutional neural network model, and combines causal convolution, dilated convolution and residual block structure, so that the model has the ability to extract features from time series and achieve prediction,

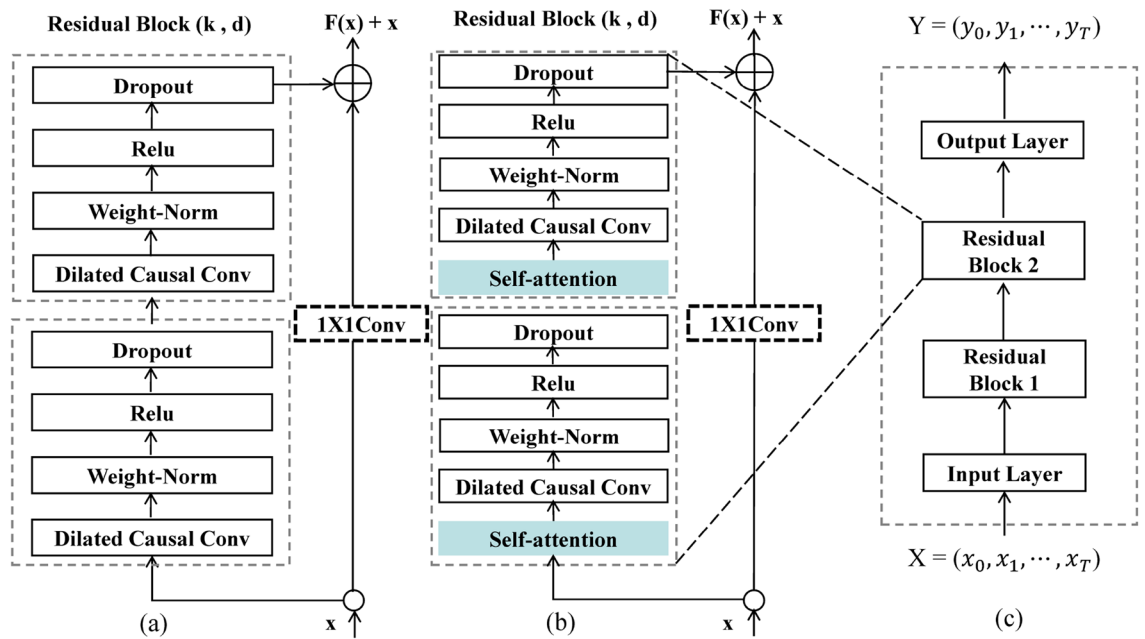


Figure 2. (a) The Standard TCN residual block structure, (b) The improved TCN residual block structure, (c) The TCN network structure we adopted.

and can effectively solve the performance degradation problem of deep networks in the process of network training²⁶. Figure 2a shows the standard TCN residual block structure, including dilated causal convolution, Weight Norm, ReLU activation function, and Dropout regularization. The dilated causal convolution is used for feature extraction, the Weight Norm can regulate the input of the hidden layer to counteract the gradient explosion problem of the network, the ReLU activation function introduces nonlinearity into the model, and the Dropout regularization operation can randomly drop neurons according to a certain probability, so as to prevent overfitting and accelerate the model training speed. If x is the input of the residual block, the output of the residual block o can be expressed as follows:

$$o = Activation(x + F(x)), \tag{5}$$

where $Activation$ is the activation function and $F(x)$ is the residual. Since the residual $F(x)$ is not 0 in practice, the stacked layers in the deep learning network can always learn new features, so the learning performance of the deep network will not degrade.

Causal Convolutions were originally proposed in the WaveNets network. Since the traditional CNN model cannot directly deal with the sequence problem, causal convolution can abstract the sequence into according to x_1, x_2, \dots, x_t and y_1, y_2, \dots, y_{t-1} to predict y_t and make y_t close to the actual value. Compared with recurrent neural networks (RNNS), models using causal convolutions do not use recurrent connections, thus allowing time series data to be input in parallel, which makes the network training faster and has greater advantages when the amount of data is large²⁷. However, in order to expand the receptive field of neural network neurons in standard causal convolution, many network layers need to be stacked or large convolution kernels need to be used. In order to solve the problem of limited receptive field of standard causal Convolution, TCN combines Dilated convolution with causal convolution, and uses Dilated Casual Convolution (DCC) to increase the receptive field of neurons without causing a significant increase in computational cost²⁸. The one-dimensional dilated causal convolution operation is expressed as follows:

$$F(s) = \sum_{i=0}^{k-1} f(i)x_{s-di}, \tag{6}$$

where x is the input sequence, $f(i)$ is the filter, also known as the convolution kernel, d is the dilation factor, k is the size of the convolution kernel, $s - di$ ensures that only past inputs can be convolved. Figure 3 shows the dilated causal convolution structure with dilation factors $d = 1, 2, 4$ as well as the convolution kernel $k = 2$.

Attention mechanism is mainly a simulation of the form of human brain attention allocation, and its essence is to assign weights to different features in the hidden layer, so as to highlight the influence of important features²⁹. The Self-attention mechanism is an improvement of the attention mechanism, which aims to capture the internal correlation of the data, so as to further improve the prediction ability of the model³⁰. The structure of self-attention mechanism is shown in Fig. 4, it consists of Q(Query), K(Key), and V(Value) vectors, which are obtained by multiplying the input data by three matrices $W_q, W_k,$ and W_v . In our proposed model, $W_q, W_k,$ and W_v are the transformations of water quality data, and the attention matrix used to determine the feature attention can be calculated by Eq. (7).

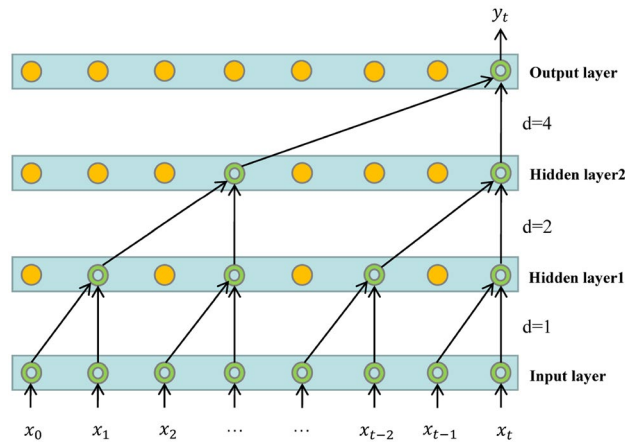


Figure 3. Dilated causal convolutions with dilation factors $d = 1, 2, 4$ and kernels $k = 2$.

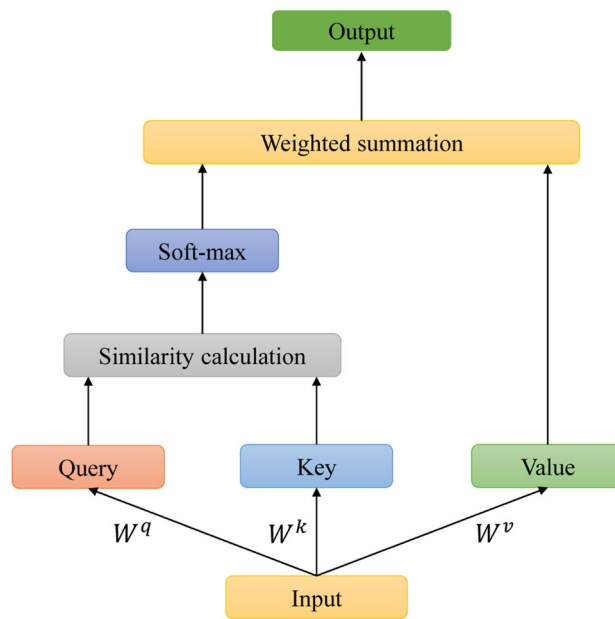


Figure 4. Self-attention structure diagram.

$$attention(Q, K, V) = softmax\left(\frac{QK^T}{\sqrt{d_K}}\right), \tag{7}$$

where T refers to the matrix transpose and d_K refers to the dimension of K .

We add Self-attention before each convolutional unit in the TCN residual block structure, and the improved TCN residual block structure is shown in Fig. 2b. Based on the above improved residual block structure, the basic structure of the deep TCN model adopted in this paper is shown in Fig. 2c. The model consists of an input layer, an improved residual block, and an output layer. The input layer mainly receives the decomposed water quality time series data. Two stacked residual block structures were used to increase the depth of the model and make the model training more adequate. The output layer is a fully connected layer, which receives the output vector of the TCN model and calculates the predicted value.

Empirical analysis

Dataset and model evaluation criteria. The opensource dataset uses the environmental estuary water quality monitoring data of the Burnett River in Queensland, Australia, which can be obtained from the Queensland Government Open Data portal. The raw data contains measurements of temperature, PH, dissolved oxygen, electrical conductivity, turbidity, and chlorophyll concentration in the water measured at 30-min intervals from 2014 to 2018³¹. Our experiment was mainly to predict the dissolved oxygen concentration value,

so we extracted the dissolved oxygen concentration data from the original data separately and resampled the data on a daily basis. However, there were missing data in the data after resampling, and we used the mean value to fill the missing data.

The private data set used in our experiment is from the Liao River automatic monitoring Station in China. The data set contains the dissolved oxygen concentration measurements recorded every 4 h from January 1, 2016 to August 31, 2022. Similarly, the original data set was resampled on a daily basis, and the missing values in the data were filled with the mean value.

To effectively evaluate our model, Root Mean Square Error (RMSE), Mean Absolute Error (MAE), and Mean Absolute Percentage Error (MAPE) were used to measure the difference between the predicted value and the true value. The smaller the error value is, the closer the predicted value is to the true value, and the higher the prediction accuracy of the model³². The calculation formula is as follows:

$$RMSE = \sqrt{\frac{1}{n} \sum_{t=1}^n (y_t - \hat{y}_t)^2}, \quad (8)$$

$$MAE = \frac{1}{n} \sum_{t=1}^n |y_t - \hat{y}_t|, \quad (9)$$

$$MAPE = 100 \times \frac{1}{n} \sum_{t=1}^n \left| \frac{\hat{y}_t - y_t}{y_t} \right|, \quad (10)$$

where n represents the sequence length, y_t and \hat{y}_t represent the true and predicted values of the model at time t , respectively.

Data smoothing. We will use the SG filter to smooth the noise present in the water quality data. Two important parameters in the SG filter determine the ability of the filter to remove noisy data, which are the window size M and the K value. It should be noted that the value of M must be a positive odd integer. If the value of M is set too large, it will remove the temporal features in the sequence, and too small, it will not be used to reduce the noise in the sequence. The value of K indicates that the points in the window are fitted by a k -order polynomial. If the value of K is too large, it will lead to least squares overfitting, and if it is too small, it will lead to underfitting. Therefore, in order to get a better SG filter, we set a variety of different parameter combinations, and determine the best parameter combination according to the RMSE, and the results are shown in Table 1.

According to Table 1, when the window size M is set to 7 and the K value is set to 5, the SG filter obtains the smallest RMSE value in both datasets, so we determine the best parameter combination $M=7$ and $K=5$ in this experiment.

STL data decomposition. Now, we will decompose the smoothed data using the STL technique to decompose the original time series into trend, seasonal, and residual series. In the experiment, the decomposition of time series is realized by calling the STL method in the statsmodels class library. The decomposition results of the water quality data of Burnett River are shown in Fig. 5a, and the decomposition results of the water quality data of Liao River are shown in Fig. 5b.

In order to verify that we used the correct time series decomposition method, we further observed the residual distribution and the mean value of the time series. According to Fig. 6a and b, the residual values of the two water quality data show an approximate normal distribution with the mean value of 0, so it can be shown that the decomposition method we adopted is correct.

Model training and prediction. Next, we will train our proposed model based on the trend series and residual series obtained after decomposition and make predictions. We take 80% of the data set as the training set, 20% as the test set, and the data set is normalized by the min–max normalization method. In the experiment, we take the water quality data of the first 7 days as the input of the model, and the output is the predicted value of the next day, that is, the sliding window size is 7, and the prediction step size is 1.

Window size (M)	K value size	Burnett River (RMSE)	Liao River (RMSE)
5	3	0.122	0.336
7	3	0.162	0.437
7	5	0.109	0.300
9	3	0.205	0.514
9	5	0.138	0.385
11	7	0.126	0.353

Table 1. Results of different parameter combinations of SG filter.

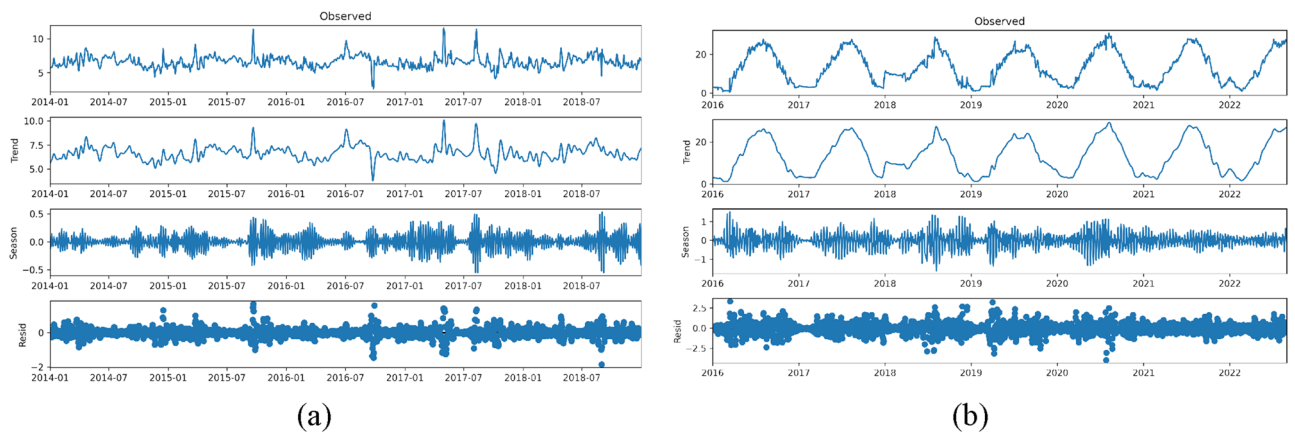


Figure 5. (a) Decomposition results of the water quality data of the Burnett River, (b) Decomposition results of the water quality data of the Liao River.

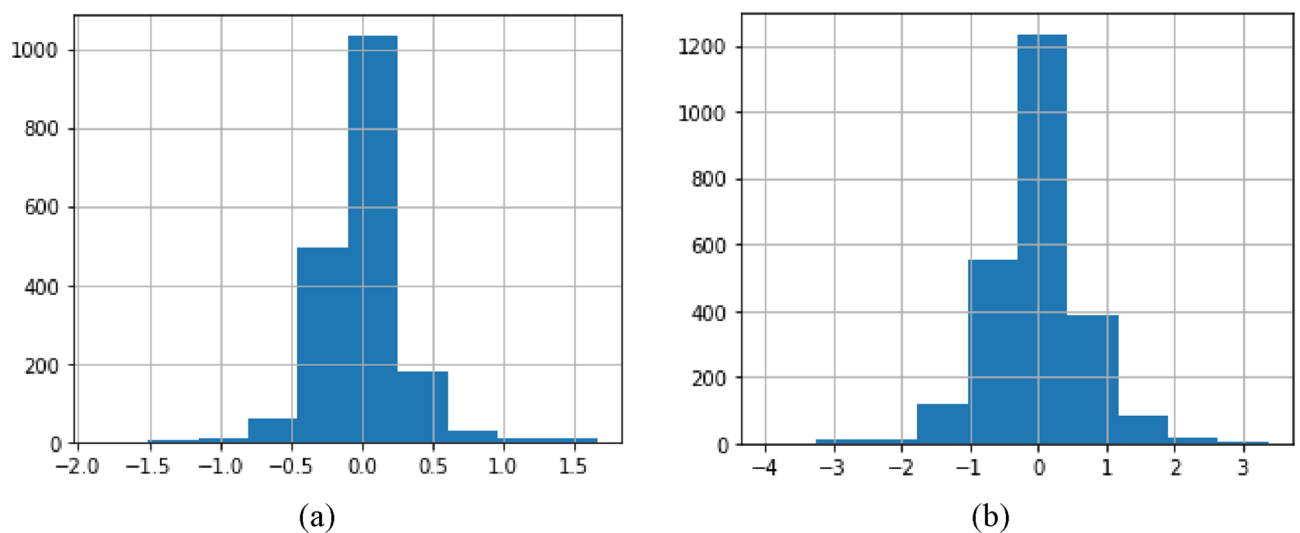


Figure 6. (a) Residual distribution of Burnett River water quality data, (b) Residual distribution of Liao River water quality data.

In order to prove the prediction ability of the proposed model, we set up baseline models to compare with our model, including Support Vector Regression (SVR), Long Short-Term Memory network (LSTM), Temporal Convolutional Network (TCN) and TCN+Self-attention model. All models were tested in the same environment, and the same evaluation criteria were adopted for comparative analysis. The main parameter Settings of each model in the test are shown in Table 2.

Model	Parameters
SVR	C = 10, kernel = 'rbf', epsilon = 0.1
LSTM	Units = 32, num_layer = 2, dense = 1
TCN	nb_filters = 64, kernel_size = 2, nb_stacks = 2
TCN + Self-attention	nb_filters = 64, kernel_size = 2, nb_stacks = 2, units = 32
Our model	M = 7, k = 5, nb_filters = 64, kernel_size = 2, nb_stacks = 2, units = 32

Table 2. Parameter Settings of each model.

Model	RMSE	MAE	MAPE
SVR	0.6818	0.5409	7.7094
LSTM	0.5729	0.3234	5.7459
TCN	0.4132	0.3018	4.2751
TCN + Self-attention	0.3042	0.2353	3.9408
Our model	0.2439	0.1901	2.9189

Table 3. Prediction results of water quality data of the Burnett River.

Results analysis

The prediction results of the Burnett River water quality data on various prediction models are shown in Table 3. According to Table 3, the results of our proposed model are the best among various evaluation indicators, which proves that the model has higher prediction accuracy and better prediction ability. In order to show the gap between the predicted results of various models and the real monitoring values more intuitively, we show the predicted values and the real values through the curve. The comparison of the prediction result curves of various prediction models on this data set is shown in Fig. 7.

Table 4 shows the prediction results of various comparison models on the Liao River water quality data. According to Table 4, The prediction result of SVR model is the worst, the prediction result of TCN model is better than LSTM, and the prediction ability of TCN model with the addition of Self-attention is greatly improved. The prediction results of our model on this data set are still the best, indicating that our model has stronger generalization ability, and the model has a good application prospect in the field of river water quality

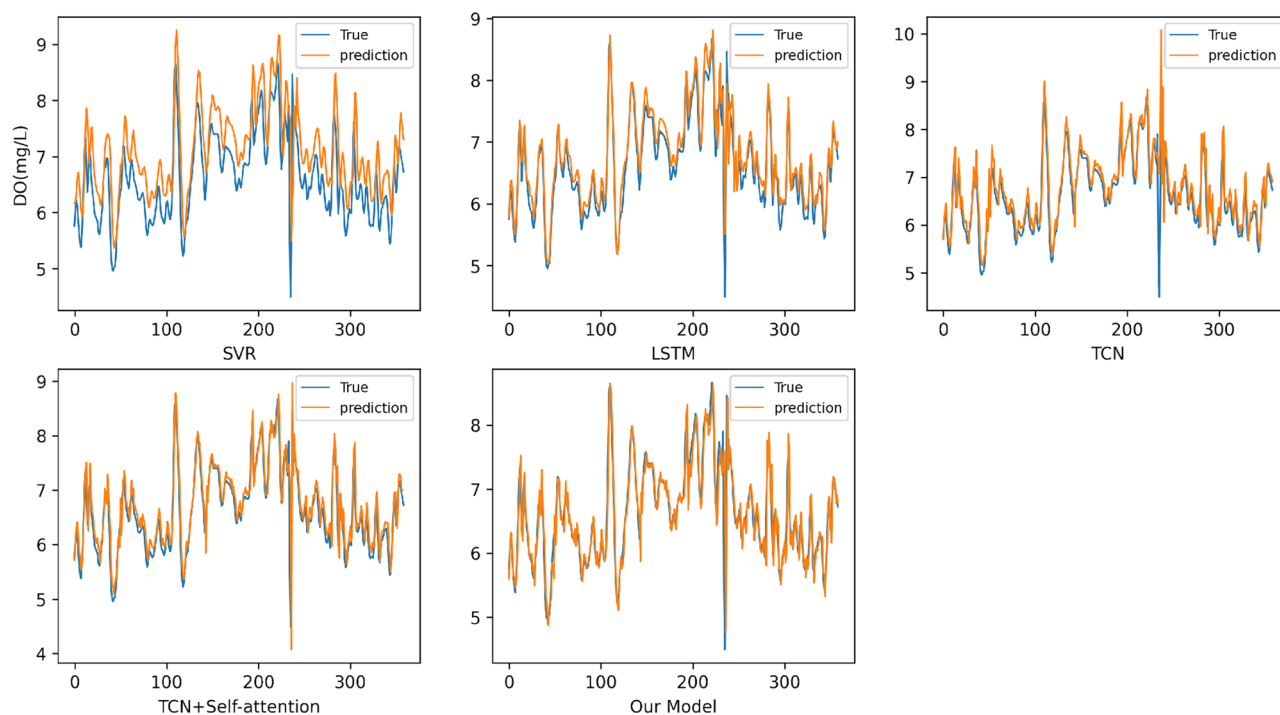


Figure 7. Prediction curves of water quality data of the Burnett River on various models.

Model	RMSE	MAE	MAPE
SVR	1.2836	1.0863	11.3279
LSTM	0.9274	0.7739	9.2364
TCN	0.7432	0.5583	6.3728
TCN + Self-attention	0.6908	0.4624	5.1906
Our model	0.4082	0.3299	4.5664

Table 4. Prediction results of water quality data of the Liao River.

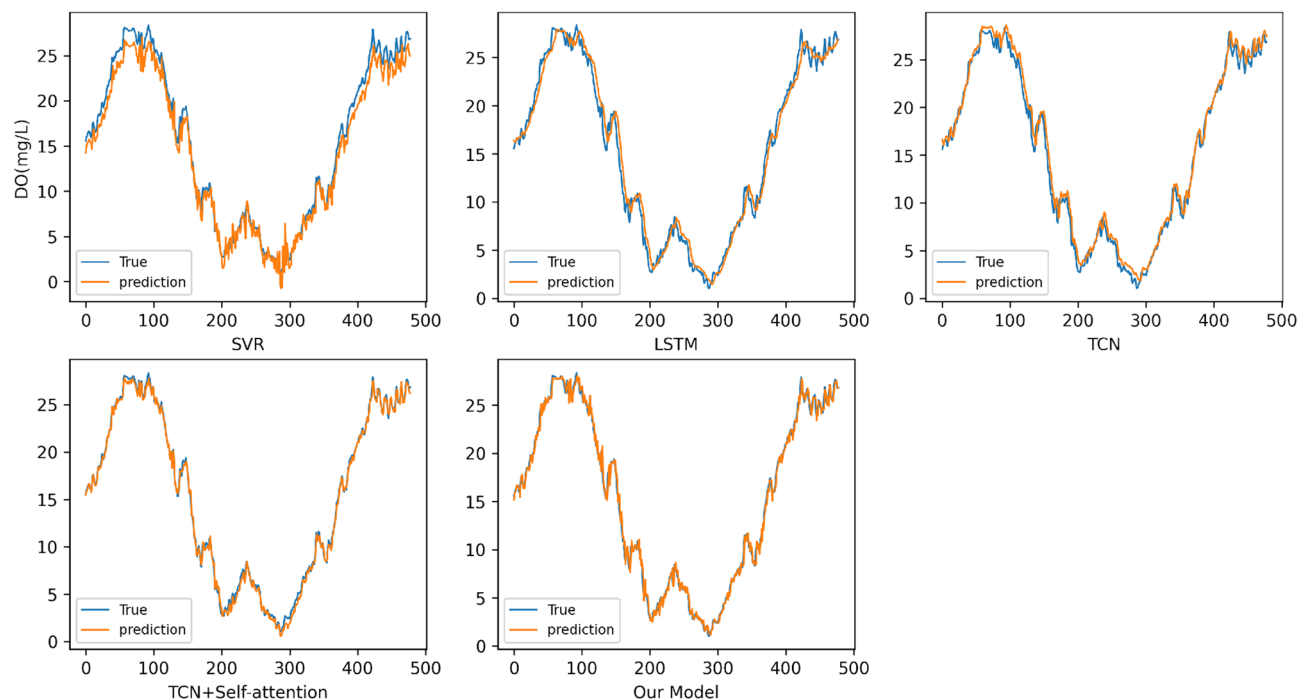


Figure 8. Prediction curves of water quality data of the Liao River on various models.

prediction. The comparison of the prediction curves of various prediction models on this data set is shown in Fig. 8.

Conclusion

Water quality prediction models are very important for water environment management. However, due to the complex and variable characteristics of water environment and the periodicity and nonlinearity of water quality data, it is a great challenge to accurately predict water quality. Therefore, in this study, we propose a hybrid improved temporal convolutional network model, which combines SG filter, STL time series decomposition method, TCN model and Self-attention mechanism. We use the SG filter to remove the noise in the water quality data, and use the STL decomposition method to decompose the water data into trend, seasonal and residual series. We improve the residual block structure of the standard TCN model and add the Self-attention mechanism to improve the prediction ability of the TCN model. Finally, we use open-source water quality data and private water quality data to conduct experimental verification. The results show that our proposed model has higher prediction accuracy than several other commonly used benchmark models. The water quality prediction model proposed in our present study can predict the water quality changes of rivers in the future period more accurately, thus helping managers to make early warnings of water pollution and take necessary measures in advance to deal with water pollution problems and better protect the water environment.

The limitation of this study is that our model only takes the historical water quality data as the model input, and does not consider the influence of other relevant features on the data. Therefore, in the next step, we will analyze the influence of different external factors on water quality changes and consider more features as the input of the model, so as to further improve the practicability of the model in real scenarios. In addition, we will continue to use water quality data of different rivers and different pollutants to evaluate the prediction ability of the model, further optimize the model structure, explore larger prediction steps, and carry out in-depth research on long-term series prediction.

Data availability

The data sets used and/or analyzed during the current study are available from the corresponding author on reasonable request.

Received: 23 March 2023; Accepted: 8 July 2023

Published online: 12 July 2023

References

1. Nouraki, A., Alavi, M., Golabi, M. & Albaji, M. Prediction of water quality parameters using machine learning models: A case study of the Karun River, Iran. *Environ. Sci. Pollut. Res.* **28**(40), 57060–57072 (2021).
2. Asadollah, S. B. H. S., Sharafati, A., Motta, D. & Yaseen, Z. M. River water quality index prediction and uncertainty analysis: A comparative study of machine learning models. *J. Environ. Chem. Eng.* **9**(1), 104599 (2021).
3. Kazakis, N. Corrigendum to “Improving prediction of water quality indices using novel hybrid machine-learning algorithms”. *Sci. Total Environ.* **742**, 141568 (2020).

4. Islam Khan, M. S., Islam, N., Uddin, J., Islam, S. & Nasir, M. K. Water quality prediction and classification based on principal component regression and gradient boosting classifier approach. *J. King Saud Univ. Comput. Inform. Sci.* **34**(8), 4773–4781 (2022).
5. Wang, X., Tian, W. & Liao, Z. Statistical comparison between SARIMA and ANN's performance for surface water quality time series prediction. *Environ. Sci. Pollut. Res.* **28**(25), 33531–33544 (2021).
6. Bi, J., Lin, Y., Dong, Q., Yuan, H. & Zhou, M. Large-scale water quality prediction with integrated deep neural network. *Inf. Sci.* **571**, 191–205 (2021).
7. Wu, J. *et al.* Application of time serial model in water quality predicting. *Comput. Mater. Continua* **74**(1), 67–82 (2023).
8. Elhag, M. *et al.* Time series analysis of remotely sensed water quality parameters in arid environments, Saudi Arabia. *Environ. Dev. Sustain.* **23**(2), 1392–1410. <https://doi.org/10.1007/s10668-020-00626-z> (2021).
9. Antanasijević, D., Pocajt, V., Perić-Grujić, A. & Ristić, M. Multilevel split of high-dimensional water quality data using artificial neural networks for the prediction of dissolved oxygen in the Danube River. *Neural Comput. Appl.* **32**(8), 3957–3966. <https://doi.org/10.1007/s00521-019-04079-y> (2020).
10. Abba, S. I. *et al.* Implementation of data intelligence models coupled with ensemble machine learning for prediction of water quality index. *Environ. Sci. Pollut. Res.* **27**(33), 41524–41539. <https://doi.org/10.1007/s11356-020-09689-x> (2020).
11. Aldhyani, T. H. H., Al-Yaari, M., Alkahtani, H. & Maashi, M. Water quality prediction using artificial intelligence algorithms. *Appl. Bionics Biomech.* **2020**, 6659314 (2020).
12. Ayesha Jasmin, S., Ramesh, P. & Tanveer, M. An intelligent framework for prediction and forecasting of dissolved oxygen level and biofloc amount in a shrimp culture system using machine learning techniques. *Expert Syst. Appl.* **199**, 117160. <https://doi.org/10.1016/j.eswa.2022.117160> (2022).
13. Tan, W. *et al.* Application of CNN and long short-term memory network in water quality predicting. *Intell. Autom. Soft Comput.* **34**(3), 1943–1958 (2022).
14. Khullar, S. & Singh, N. Water quality assessment of a river using deep learning Bi-LSTM methodology: Forecasting and validation. *Environ. Sci. Pollut. Res.* **29**(9), 12875–12889. <https://doi.org/10.1007/s11356-021-13875-w> (2022).
15. Yang, Y. *et al.* A study on water quality prediction by a hybrid CNN-LSTM model with attention mechanism. *Environ. Sci. Pollut. Res.* **28**(39), 55129–55139. <https://doi.org/10.1007/s11356-021-14687-8> (2021).
16. Bi, J., Zhang, X., Yuan, H., Zhang, J. & Zhou, M. A hybrid prediction method for realistic network traffic with temporal convolutional network and LSTM. *IEEE Trans. Autom. Sci. Eng.* **19**(3), 1869–1879. <https://doi.org/10.1109/TASE.2021.3077537> (2022).
17. Li, W. & Jiang, X. Prediction of air pollutant concentrations based on TCN-BiLSTM-DMAttention with STL decomposition. *Sci. Rep.* **13**(1), 4665. <https://doi.org/10.1038/s41598-023-31569-w> (2023).
18. Bai S, Kolter JZ, Koltun V. An empirical evaluation of generic convolutional and recurrent networks for sequence modeling. Preprint at <https://arxiv.org/quant-ph/180301271> (2018).
19. Fan, J., Zhang, K., Huang, Y., Zhu, Y. & Chen, B. Parallel spatio-temporal attention-based TCN for multivariate time series prediction. *Neural Comput. Appl.* <https://doi.org/10.1007/s00521-021-05958-z> (2021).
20. Wang, Y. *et al.* Short-term load forecasting for industrial customers based on TCN-LightGBM. *IEEE Trans. Power Syst.* **36**(3), 1984–1997 (2021).
21. Li, D., Jiang, F., Chen, M. & Qian, T. Multi-step-ahead wind speed forecasting based on a hybrid decomposition method and temporal convolutional networks. *Energy.* **238**, 121981. <https://doi.org/10.1016/j.energy.2021.121981> (2022).
22. Xiang, L., Liu, J., Yang, X., Hu, A. & Su, H. Ultra-short term wind power prediction applying a novel model named SATCN-LSTM. *Energy Convers. Manag.* **252**, 115036. <https://doi.org/10.1016/j.enconman.2021.115036> (2022).
23. Massaoudi, M., Refaat, S. S., Abu-Rub, H., Chihi, I. & Oueslati, F. S. PLS-CNN-BiLSTM: An end-to-end algorithm-based Savitzky-Golay smoothing and evolution strategy for load forecasting. *Energies* **13**(20), 5464 (2020).
24. Yin, H. *et al.* STL-ATTTLSTM: Vegetable price forecasting using STL and attention mechanism-based LSTM. *Agriculture* **10**(12), 612 (2020).
25. Xing, H. *et al.* Urban land surface temperature prediction using parallel STL-Bi-LSTM neural network. *J. Appl. Remote Sens.* **16**(3), 034529. <https://doi.org/10.1117/1.JRS.16.034529> (2022).
26. Guo, C., Kang, X., Xiong, J. & Wu, J. A new time series forecasting model based on complete ensemble empirical mode decomposition with adaptive noise and temporal convolutional network. *Neural Process. Lett.* <https://doi.org/10.1007/s11063-022-11046-7> (2022).
27. Liu, C., Zhang, L., Yao, R. & Wu, C. Dual attention-based temporal convolutional network for fault prognosis under time-varying operating conditions. *IEEE Trans. Instrum. Meas.* **70**, 1–10 (2021).
28. Hewage, P., Trovati, M., Pereira, E. & Behera, A. Deep learning-based effective fine-grained weather forecasting model. *Pattern Anal. Appl.* **24**(1), 343–366. <https://doi.org/10.1007/s10044-020-00898-1> (2021).
29. Yang, F., Zhang, H. & Tao, S. Travel order quantity prediction via attention-based bidirectional LSTM networks. *J. Supercomput.* **78**(3), 4398–4420. <https://doi.org/10.1007/s11227-021-04032-8> (2022).
30. Fu, E., Zhang, Y., Yang, F. & Wang, S. Temporal self-attention-based Conv-LSTM network for multivariate time series prediction. *Neurocomputing* **501**, 162–173. <https://doi.org/10.1016/j.neucom.2022.06.014> (2022).
31. Pantiskas L, Verstoep K, Bal H. Interpretable Multivariate Time Series Forecasting with Temporal Attention Convolutional Neural Networks. (2020)
32. Huang, L., Xie, G., Zhao, W., Gu, Y. & Huang, Y. Regional logistics demand forecasting: A BP neural network approach. *Complex Intell. Syst.* <https://doi.org/10.1007/s40747-021-00297-x> (2021).

Acknowledgements

This study was funded by the Applied Basic Research Program of Liaoning Province (Grant NO. 2022JH2/101300126) and the Science and Technology Program of Shenyang (Grant NO. 21-108-9-23).

Author contributions

Y.H.: Conceptualization, Methodology, Resources, Validation, Data curation, Writing—original draft; L.L.: Investigation, Writing—review & editing; N.W.: Conceptualization, Writing—review & editing, Supervision, Fund-ing acquisition; X.Z.: Conceptualization, Methodology; M.F.: Writing—review & editing, Supervision. All authors reviewed the manuscript.

Competing interests

The authors declare no competing interests.

Additional information

Correspondence and requests for materials should be addressed to N.W.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2023