Long-term Water Quality Prediction based on Multimodal Fusion and Temporal 2D-variation

Xiangxi Wu
College of Computer Science
Beijing University of Technology
Beijing, China
Wuxiangxi7@emails.bjut.edu.cn

Ziqi Wang
College of Computer Science
Beijing University of Technology
Beijing, China
ziqi_wang@emails.bjut.edu.cn

Yibo Li
College of Computer Science
Beijing University of Technology
Beijing, China
liyibo1@emails.bjut.edu.cn

Jing Bi
College of Computer Science
Beijing University of Technology
Beijing, China
bijing@bjut.edu.cn

Haitao Yuan
School of Automation Science and Electrical Engineering
Beihang University
Beijing, China
yuan@buaa.edu.cn

2024 International Conference on Networking, Sensing and Control (ICNSC) | 979-8-3503-6522-1/24/$31.00 ©2024 IEEE | DOI: 10.1109/ICNSC62968.2024.10760019

***Abstract*—The global water environment confronts numerous challenges, *e.g.*, water pollution, overexploitation, and ecological degradation. Comprehensive protection and management are imperative for sustainable water resource utilization. Water quality predictions provide timely warning of future water quality problems and enable early action to avoid deterioration. As science and technology are increasingly applied in comprehensive water environment management, a diverse array of multimodal data is gathered from various sources, including remote sensing images and hydrological time series. However, current water quality prediction methods, *e.g.*, statistical, machine learning, and deep learning methods fail to utilize multimodal data to enhance their accuracy of water quality prediction. To solve the above problem, this work proposes a multi-factor and long-term water quality prediction model based on multimodal data fusion named Low-rank Multimodal Fusion TimesNet (LMF-TimesNet). It first extracts features from hydrological time series and remote sensing images, respectively. Then, they are fused with the low-rank multimodal fusion network to extract diverse information. Finally, TimesNet is adopted to integrate fused multimodal water environment information for water quality prediction. Experimental results on a real-world dataset show that LMF-TimesNet achieves higher prediction accuracy and generalization ability than its state-of-the-art peers.**

***Index Terms*—Multimodal fusion, water quality prediction, temporal 2D-variation, feature extraction.**

I. INTRODUCTION

Water environment is impacted by natural and social factors, which is critical in sustaining the well-being of human society and ecosystems. However, water resources face pollution and scarcity with the rapid development of emerging manufacturing and information industries. Therefore, water quality prediction technology is vital for real-time assessment, dynamic control of pollution sources, and comprehensive management of water resources. Furthermore, water quality prediction models have a progressive evolution, such as statistical methods [1], machine learning models [2], and deep learning models [3], as monitoring devices and algorithms evolve.

The water environment is influenced by multiple factors, *e.g.*, climate change, alterations in land basins, and human activities. Changes in water quality indicators are characterized by complexity and nonlinearity. Traditional mechanistic models demand robust theoretical knowledge in biology and environmental science, making them unsuitable for real-time water quality prediction. In addition, statistical models fail to capture correlations and nonlinear relationships among water quality indicators. They lack the flexibility to adapt to dynamic changes in the water environment, thus constraining their accuracy and applicability in water quality prediction.

Deep learning methods adopt deep neural networks to deal with intricate data relationships compared to traditional machine learning ones. They possess robust generalization capabilities in time series prediction, *e.g.*, convolutional neural networks [4], long short-term memory networks [5], graph convolutional neural networks [6], and Transformer [7]. They are widely employed in water quality prediction. However, the massive multimodal data is generated from different monitoring stations and satellites. Most water quality prediction models rely solely on hydrological time series data and overlook the potential interrelationships among multimodal data. Moreover, current works concentrate on short-term and single-step predictions of individual water quality factors, failing to fully consider the intricate interactions and impacts among different water quality indicators. Therefore, they cannot capture the long-term trends and dynamic patterns of water quality changes. Thus, achieving real-time, multi-step, and multi-factor long-term water quality prediction is imperative.

To solve the above problems, this work proposes a multi-factor and long-term water quality prediction model named Low-rank Multimodal Fusion TimesNet (LMF-TimesNet). It

This work was supported by the National Natural Science Foundation of China under Grants 62473014 and 62173013, the Beijing Natural Science Foundation under Grants L233005 and 4232049, and in part by Beihang World TOP University Cooperation Program.

integrates multimodal water environment information using multimodal fusion [8] and considers the impact of precipitation observed in remote sensing images on hydrological time series. It first extracts features from hydrological time series and remote sensing images. Then, the Low-rank Multimodal Fusion (LMF) [9] network incorporates the distinctive features of each modality and efficiently fuses them. Next, TimesNet [10] utilizes the fused features for final water quality prediction. Specifically, the fused time series features are converted from 1D to 2D space by Fast Fourier Transform (FFT) [11]. Then, the stacked 2D convolutional layers are employed to extract features from multimodal data for water quality prediction, enhancing the ability to capture changes within interperiod and intraperiod variation of the time series features. Comparative experiments show that LMF-TimesNet achieves higher prediction accuracy than its typical peers.

II. Proposed Methodology

A. Low-rank Multimodal Fusion (LMF)

As the number of input modalities increases, both the dimension of the tensor and the size of the weight tensor undergo exponential growth. This results in a significant and substantial computational burden and potential model overfitting [12] problems, *i.e.*, the trained model performs well on training data but struggles to generalize to new data. This work employs the LMF network to address the tensor-based fusion challenge. For two modal (remote sensing images and hydrological time series) inputs, a one-dimensional expansion method can account for the feature correlation between the two modes while preserving the information of each mode. The input tensor formed by the unimodal representation is formulated as:

$$\mathcal{Z}=\bigotimes_{m=1}^{M} z_m,\ \ z_m \in \mathbb{R}^{d_m} \tag{1}$$

where $\bigotimes_{m=1}^{M}$ represents the tensor outer product, z_m denotes the tensor input expanded by one dimension. M represents the number of modalities, and m is the current index, identifying each specific input vector z_m in the set. This operation results in a higher-dimensional tensor by combining each vector z_m. Then, the input tensor $\mathcal{Z}\in\mathbb{R}^{d_1\times d_2\times\ldots\times d_M}$ is fed into the linear layer $g(\cdot)$ to produce a vector representation, *i.e.*,

$$h=g\left(\mathcal{Z};\mathcal{W},b\right)=\mathcal{W}\cdot\mathcal{Z}+b,\ \ h,b\in\mathbb{R}^{d_y} \tag{2}$$

where $\mathcal{W}$ denotes the weight of the layer and b denotes the bias term. Decomposing the weight tensor $\mathcal{W}$ into a set of modality-specific and low-rank factors can enhance the computational efficiency [13]. For an order-M tensor $\widetilde{W}_k\in\mathbb{R}^{d_1\times\ldots\times d_M}, k=1,\ldots,d_h$, there always exists an exact vector decomposition of the form, *i.e.*,

$$\widetilde{W}_k=\sum_{r=1}^{R}\bigotimes_{m=1}^{M} w_{m,k}^{(i)},\ \ w_{m,k}^{(i)}\in\mathbb{R}_m^d \tag{3}$$

where R denotes the smallest value that ensures the validity of the decomposition, and it is the rank of the tensor. The collection of vectors $\{\{w_{m,k}^{(i)}\}_{m=1}^{M}\}_{i=1}^{R}, k=1,\ldots,d_h$, constitutes the decomposition factors of the original tensor with rank R.

In LMF, a rank is initially set to r, and it is utilized to parameterize the model with r decomposition factors $\{\{w_{m,k}^{(i)}\}_{m=1}^{M}\}_{i=1}^{r}, k=1,\ldots,d_h$, enabling the reconstruction of a low-rank version of $\widetilde{W}_k$. Each $\widetilde{W}_k$ contributes to one-dimension in the output vector h, *i.e.*, $h_k=\widetilde{\mathcal{W}}_k\cdot\mathcal{Z}$. Next, these vectors are recombined and concatenated into M modality-specific and low-rank factors. Let $\mathbf{w}_m^{(i)}=\left[w_{m,1}^{(i)},w_{m,2}^{(i)},\ldots,w_{m,d_h}^{(i)}\right]$, where $\{w_m^{(i)}\}_{i=1}^{r}$ represents the corresponding low-rank factors for each modality m. Then, a low-rank weight tensor is defined as:

$$\mathcal{W}=\sum_{i=1}^{r}\bigotimes_{m=1}^{M}\mathbf{w}_m^{(i)} \tag{4}$$

Thus, (2) can be recalculated as:

$$h=\left(\sum_{i=1}^{r}\bigotimes_{m=1}^{M}\mathbf{w}_m^{(i)}\right)\cdot\mathcal{Z} \tag{5}$$

Fig. 1 shows the process of the two-modality and low-rank weight decomposition of hydrological time series and remote sensing images.

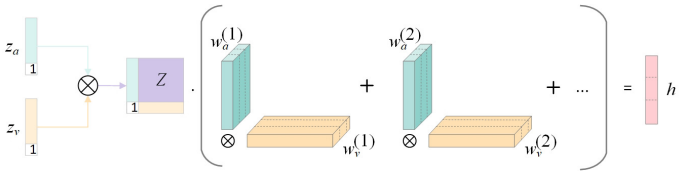


Fig. 1. Bimodal low-rank weight decomposition.

After that, $\mathcal{Z}=\bigotimes_{m=1}^{M} z_m$ is input into (5), *i.e.*,

$$\begin{aligned}\mathrm{h}&=\left(\sum_{i=1}^{r}\bigotimes_{m=1}^{M}\mathbf{w}_m^{(i)}\right)\cdot\mathcal{Z}\\&=\sum_{i=1}^{r}\left(\bigotimes_{m=1}^{M}\mathbf{w}_m^{(i)}\cdot\mathcal{Z}\right)\\&=\sum_{i=1}^{r}\left(\bigotimes_{m=1}^{M}\mathbf{w}_m^{(i)}\cdot\bigotimes_{m=1}^{M}z_m\right)\\&=\bigwedge_{m=1}^{M}\left[\sum_{i=1}^{r}\mathbf{w}_m^{(i)}\cdot z_m\right]\end{aligned} \tag{6}$$

where $\bigwedge_{m=1}^{M}$ denotes the element-wise product over a sequence of tensors.

Then, the input of (6) is illustrated in Fig. 1, *i.e.*,

$$\begin{aligned}\mathrm{h}&=\left(\sum_{i=1}^{r}\mathbf{w}_a^{(i)}\otimes\mathbf{w}_v^{(i)}\right)\cdot\mathcal{Z}\\&=\left(\sum_{i=1}^{r}\mathbf{w}_a^{(i)}\cdot z_a\right)\circ\left(\sum_{i=1}^{r}\mathbf{w}_v^{(i)}\cdot z_v\right)\end{aligned} \tag{7}$$

where z_a and z_v denote the weight tensors of remote sensing images and hydrological time series, respectively. $\mathbf{w}_a$ and $\mathbf{w}_v$ represent the low-rank factors obtained through their weight decompositions. Moreover, $\circ$ denotes the Hadamard product, indicating that the elements within two vectors are multiplied element-wise.

Then, h can be computed from the input representation z_m, without explicitly creating the tensor $\mathcal{Z}$. This is achieved using the core idea of parallel decomposition of $\mathcal{Z}$ and M. In the simplified calculation of h, the modalities of remote sensing images and hydrological time series are decoupled, allowing them to be parameterized by the tensor representing the two modalities rather than a set of vectors. This approach reduces the computational complexity of the model and enhances its fusion performance. In addition, different modalities are decoupled during the calculation process, allowing it to be easily generalized to a different number of modalities. In that case, adding a new modality can be performed by adding another set of modality-specific factors and extending (7). After the LMF operation, the two information modalities, including remote sensing images and water quality time series, are well integrated. It retains unimodality characteristics and extracts complementary valuable water quality information from different modes. Therefore, the fused features can provide assistance in the decision-making of the next temporal 2D-variation water quality prediction model, further improving prediction accuracy.

B. TimesNet

External natural and human factors interfere with changes in hydrological time series data, resulting in a complex pattern of variation. This work analyzes time series from a multi-period perspective. The time series commonly exhibit multiple periodicities, including daily, monthly, and annual changes in water environment monitoring values. These overlapping periods interact, rendering variation modeling challenging. For each period, the variation of individual time points is influenced by the temporal pattern of their immediate vicinity and the variation in adjacent periods. Changes occurring within a single day can be perceived as short-term temporal patterns. Conversely, the aggregation of these daily changes can be regarded as a long-term trend across successive periods. Therefore, these two types of temporal variation are named intraperiod and interperiod variations. Within each period, the variation of individual time points is influenced by the temporal pattern of their adjacent regions and correlated with the fluctuation and trend of neighboring periods. To tackle this challenge, the temporal variation is extended into 2D space. In that case, by converting the 1D time series into a collection of 2D tensors, the limitation of representation capacity in the original 1D space can be overcome. Intraperiod and interperiod variation can be effectively integrated into the 2D domain, thereby capturing temporal variation in a 2D space.

The original 1D arrangement for time series data is denoted as $\mathbf{X}_{\mathrm{1D}} \in \mathbb{R}^{T\times C}$, where T denotes the length and C denotes recorded variates. The time series is analyzed in the frequency domain using the FFT to identify trends and patterns in the interperiod variation. This process is shown as:

$$\mathbf{A}=\mathrm{Avg}\left(\mathrm{Amp}\left(\mathrm{FFT}(\mathbf{X}_{\mathrm{1D}})\right)\right) \tag{8}$$

$$\{f_1,\cdots,f_k\}=\mathrm{N}(\mathbf{A}),\ f_*\in\{1,\cdots,[\frac{T}{2}]\} \tag{9}$$

$$p_i=\left\lceil\frac{T}{f_i}\right\rceil,\ i\in\{1,\cdots,k\} \tag{10}$$

where $\mathrm{FFT}(\cdot)$ and $\mathrm{Amp}(\cdot)$ represent the FFT and the calculation of amplitude values, respectively. $\mathbf{A}\in\mathbb{R}^T$ denotes the amplitude calculated at each frequency, which is obtained by the average value $\mathrm{Avg}(\cdot)$ from the C dimension. $\mathrm{N}(\cdot)$ represents the process of selecting periods k. In addition, due to the sparsity of the frequency domain and to reduce noise introduced by insignificant high frequencies, the top-k amplitude values are selected, where k is a parameter that needs to be determined.

Based on the chosen top-k frequencies $\{f_1,\cdots,f_k\}$ and their corresponding period lengths $\{p_1,\cdots,p_k\}$, the 1D time series $\mathbf{X}_{\mathrm{1D}}\in\mathbb{R}^{T\times C}$ can be transformed into multiple 2D tensors, *i.e.*,

$$\mathbf{X}^i_{\mathrm{2D}}=\mathrm{S}_{p_i,f_i}\left(\mathrm{P}(\mathbf{X}_{\mathrm{1D}})\right),i\in\{1,\cdots,k\} \tag{11}$$

where P represents padding, and $\mathrm{P}(\cdot)$ expands the time series along the temporal dimension with a value of zero to make it uniform and compatible with the S_{p_i,f_i}. S represents the reshape operation that fills time series data into a 2D tensor. p_i and f_i represent the number of rows and columns of the 2D tensor, respectively.

Finally, by leveraging the selected frequency and estimated period, a set of tensors $\mathbf{X}^1_{\mathrm{2D}},\cdots,\mathbf{X}^k_{\mathrm{2D}}$ is obtained through the fusion of remote sensing images and hydrological time series. These tensors represent the k distinct temporal 2D-variation generated across different periods.

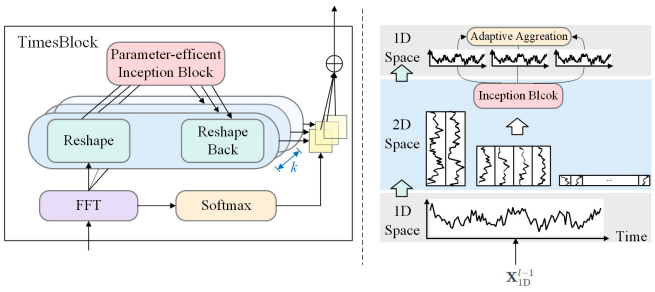


Fig. 2. Structure of TimesBlock.

Fig. 2 shows the structure of TimesBlock. It is constructed as a residual connection. For layer l of TimesNet with input $\mathbf{X}^{l-1}_{\mathrm{1D}}$, the connection process can be represented as:

$$\mathbf{X}^l_{\mathrm{1D}}=\mathrm{O}\left(\mathbf{X}^{l-1}_{\mathrm{1D}}\right)+\mathbf{X}^{l-1}_{\mathrm{1D}} \tag{12}$$

where $\mathrm{O}(\cdot)$ denotes the TimesBlock module.

For the TimesBlock l, data processing can be viewed as two consecutive parts. The first is to capture the temporal 2D-variation after passing through the transformation from 1D to

2D. Second, the 2D features are integrated into 1D as input to the TimesBlock l+1 by adaptively integrating data from different periods.

The Parameter-efficient Inception Block [14] includes a multi-scale 2D convolutional kernel. A series of convolutional kernels of different sizes are utilized through the Inception Block convolutional network to capture feature information at different scales in temporal 2D-variation efficiently. This process can be formalized as:

$$\widehat{\mathbf{X}}_{\mathrm{2D}}^{l,i}=\mathrm{I}\left(\mathbf{X}_{\mathrm{2D}}^{l,i}\right),\ i\in\{1,\cdots,k\} \tag{13}$$

where $\mathbf{X}_{\mathbf{2D}}^{l,i}$ is the 2D tensor of the i-th transformation. $\mathrm{I}(\cdot)$ represents the Inception Block module, responsible for feature extraction and characterization learning from the 2D tensor.

Then, the learned 2D feature representation $\mathbf{X}_{\mathbf{2D}}^{l,i}$ needs to be transformed back into the one-dimensional space $\widehat{\mathbf{X}}_{\mathrm{1D}}^{l,i}$. The padding sequence of length $p_i \times f_i$ is truncated to the original length T using $\mathrm{H}(\cdot)$, which denotes the truncation function. This process is shown as:

$$\widehat{\mathbf{X}}_{\mathrm{1D}}^{l,i}=\mathrm{H}\left(\mathrm{R}_{1,(p_i\times f_i)}\left(\widehat{\mathbf{X}}_{\mathrm{2D}}^{l,i}\right)\right),i\in\{1,\cdots,k\} \tag{14}$$

Finally, the fused k different 1D-representations are input to the next layer of TimesBlock. Since the frequency and period corresponding to the peak of amplitude $\mathbf{A}$ are representative, the 2D tensor obtained after the transformation at this amplitude is considered important. Thus, the 1D-representations are aggregated based on the amplitudes, *i.e.*,

$$\begin{aligned}\widehat{\mathbf{A}}_{f_1}^{l-1},\cdots,\widehat{\mathbf{A}}_{f_k}^{l-1}&=\mathrm{Softmax}\left(\mathbf{A}_{f_1}^{l-1},\cdots,\mathbf{A}_{f_k}^{l-1}\right)\\ \mathbf{X}_{\mathrm{1D}}^{l}&=\sum_{i=1}^{k}\widehat{\mathbf{A}}_{f_i}^{l-1}\times\widehat{\mathbf{X}}_{\mathrm{1D}}^{l,i}\end{aligned} \tag{15}$$

C. Architecture of LMF-TimesNet

Fig. 3 illustrates the structure of LMF-TimesNet. The LMF module receives inputs of remote sensing images X_r and hydrologic time series X_t. After feature extraction of the two modalities inputs by ResNet101 [15], the two modalities are fused in tensor form by the LMF. Next, the obtained fused time series features are embedded as inputs to TimesNet. It consists of TimesBlocks stacked in residual connection. First, TimesBlocks derives the periods of the time series from the FFT. Then, temporal 2D-variation features are extracted from the reshaping tensor from 1D to 2D space, utilizing a Parameter-efficient Inception block containing several multi-scale 2D convolutional layers. The amplitude is used as the weight value after passing through the Softmax function to reshape the output features of the convolutional layer, aggregating the 2D space back to 1D. Following the processes above, the final prediction result is obtained.

In summary, after multimodal fusion, the water environment information is transformed from a 1D form to multiple well-distributed 2D tensors, representing intraperiod and interperiod variation. In turn, TimesBlock can simultaneously and adequately capture the temporal 2D-variation of hydrologic time series data fused precipitation information. Therefore, the LMF-TimesNet integrates abundant water environment information through multimodal fusion, facilitating more efficient representation learning compared to direct extraction from 1D time series.

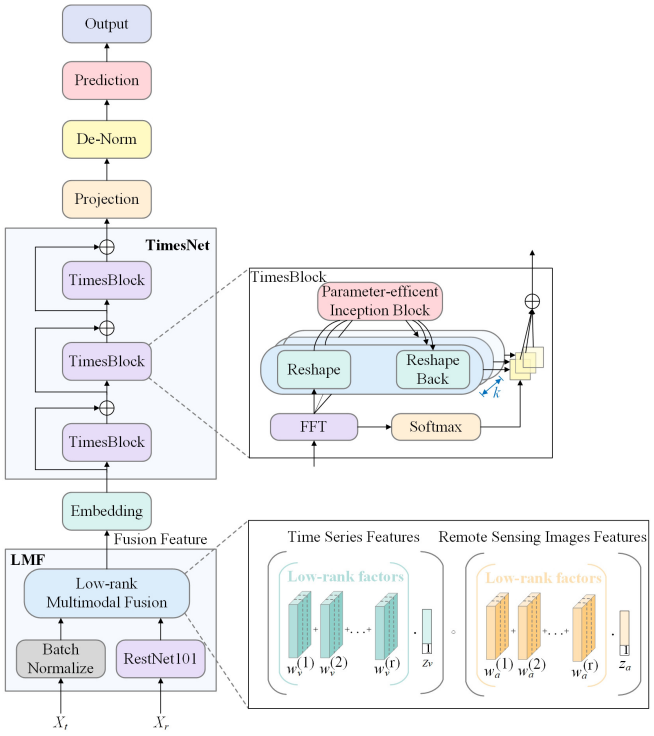


Fig. 3. Overall architecture of LMF-TimesNet.

III. Experimental Results and Discussion

A. Dataset Selection and Processing

The hydrological time series dataset is obtained from the public information officially released by China's national automatic surface water quality monitoring stations. It contains nine water quality monitoring stations distributed in the Haihe River Basin of the Beijing-Tianjin-Hebei (BTH) region. The remote sensing images dataset is obtained from publicly available precipitation remote sensing images from NASA's Goddard Earth Science Data and Information Service Center, containing global precipitation information. The period of the two datasets is selected from Jan 1, 2018 to Dec 31, 2023, *i.e.*, the samples from 2018 to 2023 are used as the dataset, and there are a total of 10,956 monitoring values.

Hydrologic time series data is collected from each water quality monitoring station at 0:00, 4:00, 8:00, 12:00, 16:00, and 20:00. The remote sensing images are sourced from satellite monitoring stations at 0:00 every day and automatically sampled every 30 minutes, resulting in a total of 48 monitoring values per day. To ensure the temporal alignment with the time series data, the remote sensing image data is also selected for 0:00, 4:00, 8:00, 12:00, 16:00, and 20:00 daily. Moreover, the latitude and longitude ranges of the remote sensing images

of precipitation are intercepted to the extent of the Haihe River Basin in the BTH region, which covers the water quality monitoring area of the nine monitoring stations. This ensures the spatial alignment with the monitoring stations.

B. Comparative Experiments

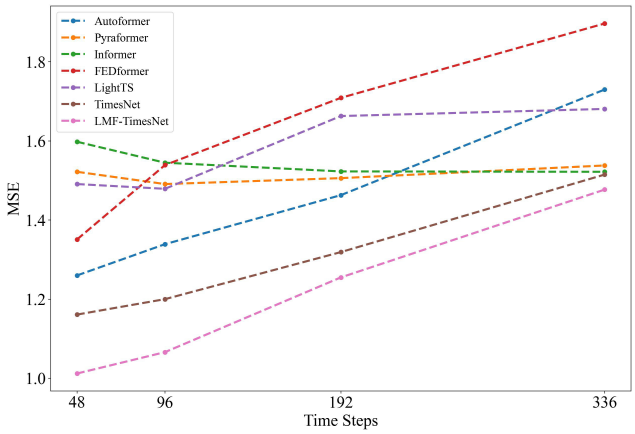


Fig. 4. MSE of different models under different time steps.

To verify the effectiveness of LMF-TimesNet, six baseline models are adopted for comparison, including Autoformer, Pyraformer [16], Informer, FEDformer [17], LightTS [18], and TimesNet. In addition, Mean Squared Error (MSE) is adopted as the evaluation metric to evaluate the accuracy of the prediction result. The MSE is more sensitive to significant errors, which helps to capture the slight deviation of the predicted values from the true values.

Table I shows the comparative results of MSE on the seven models for four prediction steps (48, 96, 192, and 336). It is shown that LMF-TimesNet outperforms other benchmark models in all prediction spans. The MSE errors are reduced by 9.1%–36.6% compared to the benchmark models, which proves that LMF-TimesNet has high accuracy and stability in long-term water quality prediction. In addition, the multimodal fusion reduces the MSE by 12.8%, 11.2%, 11.6%, and 9.1%, respectively, compared with TimesNet at different prediction steps. It verifies that the multimodal fusion of remote sensing images and hydrological time series can better capture the trend of water quality changes and significantly improve prediction accuracy. Fig. 4 shows the results of the MSE comparison between LMF-TimesNet and benchmark models in long-term water quality prediction. It is illustrated that LMF-TimesNet achieves the best performance and lowest MSE on all prediction steps, proving that the predicted values obtained by LMF-TimesNet are closer to the true values of water quality indicators.

Figs. 5(a)-5(f) illustrate the curves comparing the predicted with the true values for each model on the dissolved oxygen indicators. They show a more intuitive understanding of the strengths and weaknesses of different models for long-term water quality prediction. Fig. 5(a) and 5(b) show that the

TABLE I
MULTI-FACTOR AND LONG-TERM PREDICTION RESULTS UNDER DIFFERENT TIME STEPS

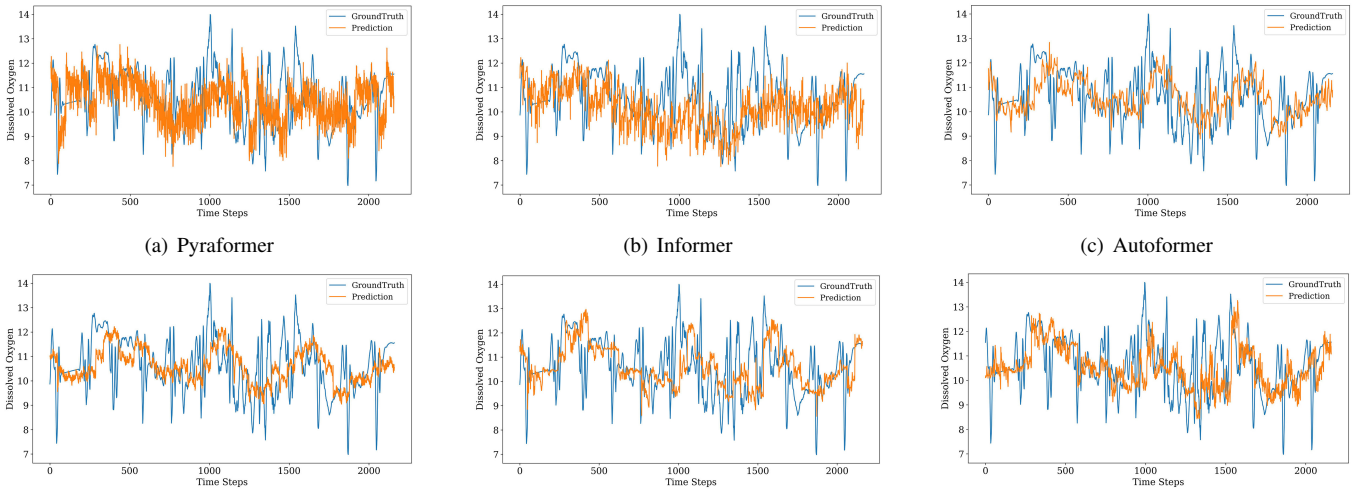
Model	MSE			
	48steps	96steps	192steps	336steps
Autoformer	1.260	1.339	1.463	1.730
Pyraformer	1.522	1.491	1.506	1.538
Informer	1.598	1.545	1.523	1.522
FEDformer	1.351	1.539	1.709	1.897
LightTS	1.491	1.479	1.663	1.681
TimesNet	1.161	1.200	1.319	1.515
LMF-TimesNet	**1.012**	**1.066**	**1.165**	**1.377**

prediction accuracy of the Pyraformer and Informer is significantly lower than that of the other models. The predicted values show high volatility and lack a smooth output sequence, with the Informer showing particularly poor performance around time step 1000. It is shown in Figs. 5(c)-5(e) that the long-term prediction results of the Autoformer, FEDformer, and LightTS enhance performance compared to Pyraformer and Informer. Their predicted values show a smoother sequence and align more closely with the overall trend of the predicted values. However, the prediction results for the peak are relatively conservative, resulting in a shifted and lagging curve compared to the true values. It is shown in Fig. 5(f) that TimesNet achieves relatively accurate prediction of sequence trends and peaks compared to the aforementioned models. However, there is still some lags compared to the true values. Fig. 6 shows that LMF-TimesNet performs well in predicting peaks, stationary sequences, overall trends, and prediction lags, outperforming the benchmark models.

IV. CONCLUSIONS

Water quality prediction is an essential task of water environment management, and it is of great significance in preventing and controlling water pollution. Many monitoring devices are deployed for comprehensive water environment management, and water quality data shows multimodal characteristics. However, current water quality prediction models overlook the mutual influence of changes in multimodal data and fail to achieve long-term water quality predictions. To solve the above problems, a novel water quality prediction model named Low-rank Multimodal Fusion TimesNet (LMF-TimesNet) is proposed in this work. Feature-level fusion of hydrological time series and remote sensing images is performed by low-rank multimodal fusion. Then, the fused feature is utilized by TimesNet to achieve long-term water quality prediction. Experimental results with real-life water quality datasets reveal that the prediction accuracy of LMF-TimesNet outperforms the existing state-of-the-art models.

Our future work intends to incorporate intelligent optimization algorithms [19] to tune the model parameters and improve the prediction accuracy. In addition, we intend to design a better computer vision model to replace the inception block module in TimesNet to extract multimodal water environment features, further enhancing the model's performance.



(a) Pyraformer (b) Informer (c) Autoformer

(d) FEDformer (e) LightTS (f) TimesNet

Fig. 5. Prediction results for different models on the BTH datasets.

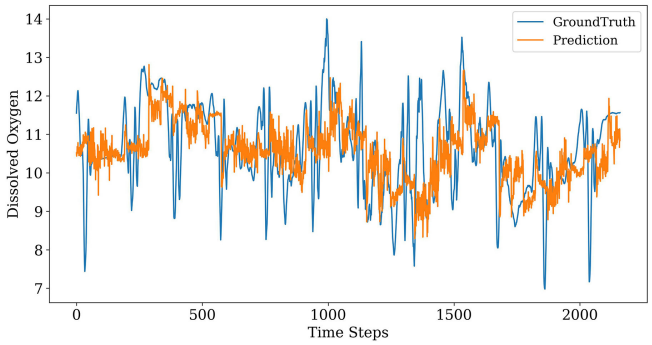


Fig. 6. Comparison between predicted and true values of LMF-TimesNet.

REFERENCES

[1] M. Salucci, L. Tenuti, G. Oliveri and A. Massa, "Efficient Prediction of the EM Response of Reflectarray Antenna Elements by an Advanced Statistical Learning Method," *IEEE Transactions on Antennas and Propagation*, vol. 66, no. 8, pp. 3995–4007, Aug. 2018.

[2] S. Yu and S. Shen, "Compaction Prediction for Asphalt Mixtures Using Wireless Sensor and Machine Learning Algorithms," *IEEE Transactions on Intelligent Transportation Systems*, vol. 24, no. 1, pp. 778–786, Jan. 2023.

[3] X. Kong and Z. Ge, "Deep PLS: A Lightweight Deep Learning Model for Interpretable and Efficient Data Analytics," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 34, no. 11, pp. 8923-8937, Nov. 2023.

[4] Z. Li, F. Liu, W. Yang, S. Peng and J. Zhou, "A Survey of Convolutional Neural Networks: Analysis, Applications, and Prospects," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 33, no. 12, pp. 6999–7019, Dec. 2022.

[5] N. Jin, Y. Zeng, K. Yan and Z. Ji, "Multivariate Air Quality Forecasting With Nested Long Short Term Memory Neural Network," *IEEE Transactions on Industrial Informatics*, vol. 17, no. 12, pp. 8514–8522, Dec. 2021.

[6] Z. Qu, X. Liu and M. Zheng, "Temporal-Spatial Quantum Graph Convolutional Neural Network Based on Schrödinger Approach for Traffic Congestion Prediction," *IEEE Transactions on Intelligent Transportation Systems*, vol. 24, no. 8, pp. 8677–8686, Aug. 2023.

[7] B. Xiao, J. Hu, W. Li, C. M. Pun and X. Bi, "CTNet: Contrastive Transformer Network for Polyp Segmentation," *IEEE Transactions on Cybernetics*, vol. 52, no. 9, pp. 8456–8467, Sep. 2023.

[8] J. Wang, J. Li, Y. Shi, J. Lai and X. Tan, "AM³Net: Adaptive Mutual-Learning-Based Multimodal Data Fusion Network," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 32, no. 8, pp. 5411–5426, Aug. 2022.

[9] Z. He, H. Shi, Y. Wu and Z. Tu, "Low-rank Fusion Network for Multi-modality Person Re-identification," *2023 8th International Conference on Intelligent Computing and Signal Processing (ICSP)*, Xian, China, 2023, pp. 1578–1581.

[10] D. Zhang, R. Wang, Y. Fan, C. Wang and X. Chen, "Short-term Metro Station Power Lighting Load Prediction Based on TimesNet," *2023 IEEE 4th China International Youth Conference On Electrical Engineering (CIYCEE)*, Chengdu, China, 2023, pp. 1–6.

[11] A. Becoulet and A. Verguet, "A Depth-First Iterative Algorithm for the Conjugate Pair Fast Fourier Transform," *IEEE Transactions on Signal Processing*, vol. 69, pp. 1537–1547, Feb. 2021.

[12] C. He, X. Li, Y. Xia, J. Tang, J. Yang and Z. Ye, "Addressing the Overfitting in Partial Domain Adaptation With Self-Training and Contrastive Learning," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 34, no. 3, pp. 1532–1545, Mar. 2024.

[13] J. Bi, Z. Wang, H. Yuan, J. Zhang and M. Zhou, "Self-adaptive Teaching-learning-based Optimizer with Improved RBF and Sparse Autoencoder for High-dimensional Problems," *Information Sciences*, vol. 630, pp. 463–481, Jun. 2023.

[14] Szegedy C, Liu W, Jia Y, Sermanet P, Reed S, Anguelov D, Erhan D, Vanhoucke V and Rabinovich A. "Going deeper with convolutions." *Proceedings of the IEEE conference on computer vision and pattern recognition*, Boston, USA, 2015, pp. 1–9.

[15] B. Wang, R. Luo and W. Zhu, "Application Research of Facial Expression Recognition Based on Improved ResNet101," *2023 IEEE 6th International Conference on Automation, Electronics and Electrical Engineering (AUTEEE)*, Shenyang, China, 2023, pp. 770–776.

[16] X. Hu, W. Wang, J. Tang and M. Liu, "Multi-Step Ahead Prediction of Main Steam Flow Using PCA and Pyraformer in MSWI Process," *2023 35th Chinese Control and Decision Conference (CCDC)*, Yichang, China, 2023, pp. 372–376.

[17] T. Zhou, Z. Ma, Q. Wen, X. Wang, L. Sun and R. Lin, "Fedformer: Frequency enhanced decomposed transformer for long-term series forecasting." *International conference on machine learning*, Baltimore, USA, 2022, pp. 27268–27286.

[18] C. David, M. Zhang, B. Yang, T. Kieu, C. Guo and Christian S. Jensen, "LightTS: Lightweight time series classification with adaptive ensemble distillation." *Proceedings of the ACM on Management of Data*, Seattle, USA, 2023, pp. 1–27.

[19] J. Bi, Z. Wang, H. Yuan, J. Zhang and M. Zhou, "Cost-Minimized Computation Offloading and User Association in Hybrid Cloud and Edge Computing," *IEEE Internet of Things Journal*, vol. 11, no. 9, pp. 16672–16683, May 2024.