Journal of Water Process Engineering 78 (2025) 108747

Contents lists available at ScienceDirect

Journal of Water Process Engineering

journal homepage: www.elsevier.com/locate/jwpe

Hybrid water quality prediction based on attention combined with frequency enhancement and multi-seasonal decomposition

Yibo Li [a], Ziqi Wang [b], Haitao Yuan [c], Jing Bi [a,*]

[a] College of Computer Science, Beijing University of Technology Beijing, China
[b] School of Software Technology, Zhejiang University Ningbo, China
[c] School of Automation Science and Electrical Engineering, Beihang University Beijing, China

ARTICLE INFO

Editor: Zhugen Yang

Keywords:
Water quality forecasting
Seasonal trend decomposition
Discrete Fourier transform
Savitsky-Golay filter
Frequency enhancement

ABSTRACT

Water quality forecasting methods can reflect the water quality situation and development trend in the short or long-term future and provide important support for water environment management. Due to the influence of other factors fluctuating in the water environment and errors in collection equipment, water quality time series data are characterized by instability and high nonlinearity, and a nonlinear regression problem of non-smooth series data is difficult in the prediction field. This work proposes a hybrid model for water quality prediction called SMDF2 to improve the prediction accuracy. SMDF2 integrates the Savitsky-Golay (SG) filter, Multi-seasonal trend decomposition using loss (MSTL), Discrete Fourier Transform (DFT), Frequency Enhanced Block (FEB) and Frequency Domain Enhanced Attention (FEA) in an encoder-decoder architecture. The SG filter is employed to smooth out the noise to diminish the instability in time series. MSTL is used to extract periodic and trend components for the nonlinear sequences, and DFT is utilized to achieve the conversion between the time domain and the frequency domain. FEB and FEA are employed for the frequency-domain feature extraction and the frequency-domain feature correlation learning. Experimental results with real domestic and international water environment datasets for both long-term and short-term predictions demonstrate that SMDF2 surpasses various advanced models in accuracy for both single-element and multi-element predictions. Specifically, it improves prediction accuracy by an average of 21.73 % for single-element tasks and 18.14 % for multi-element tasks.

1. Introduction

Water quality prediction, which forecasts future changes in water quality by analyzing historical data, has become a central objective of water quality monitoring. This technology provides critical decision-making support for water environment management and pollution control, playing an indispensable role in preventing water pollution and safeguarding ecological health. Therefore, developing stable and highly efficient predictive models for water quality is crucial [1–3].

Water quality prediction methods fall into two broad categories: mechanistic and data-driven models. Data-driven approaches include traditional statistical methods and recent deep learning models. Traditional techniques extract linear relationships: autoregressive integrated moving average models [4–6] provide recursive sequential predictions, while support vector machines [7] flexibly handle low and high-dimensional data. Advances in deep learning [8] are enabling more sophisticated data-driven models for water quality time series forecasting.

Recurrent neural networks (RNNs) [9,10] demonstrate strong potential for time series forecasting, where gated architectures like long short-term memory (LSTM) networks [11–13] and gated recurrent units [14] effectively address gradient issues through gated mechanisms. Autoregressive RNNs enable probabilistic forecasting [15] using extensive temporal data. Recently, Transformers [16] have gained prominence for capturing long-range dependencies *via* attention mechanisms, with significant enhancements including: Informer [17], which employs ProbSparse self-attention and distillation for linear-complexity long-sequence processing; Autoformer [18] utilizing seasonal-trend decomposition and autocorrelation mechanisms for multi-resolution analysis; and Pyraformer [19] implementing hierarchical pyramidal attention through tree structures to capture multi-scale dependencies efficiently.

Due to the influence of environmental factors and data acquisition

* Corresponding author.
E-mail address: bijing@bjut.edu.cn (J. Bi).

https://doi.org/10.1016/j.jwpe.2025.108747
Received 12 July 2025; Received in revised form 25 August 2025; Accepted 11 September 2025
Available online 22 September 2025

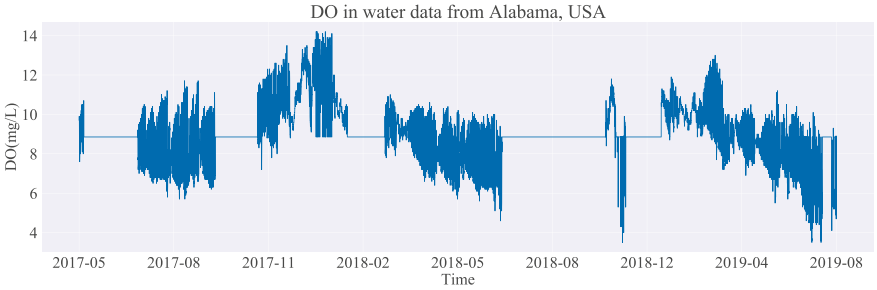

Fig. 1. Time series of dissolved oxygen (DO, mg/L) in surface waters of Alabama, USA.

equipment errors, water quality time series data typically exhibit instability and high nonlinearity. Addressing the nonlinear regression problems associated with processing such non-stationary sequential data represents a significant challenge in the current field of prediction. Specifically, water quality data often contains substantial noise and outliers, which interfere with subsequent analysis and may lead to model overfitting. Simultaneously, influenced by regular environmental factors such as seasons and weather, water quality data demonstrates complex periodic and trend components, requiring models capable of capturing intricate patterns. However, conventional forecasting methods struggle to effectively extract and utilize these features to enhance model performance. Furthermore, when processing complex sequences, relying solely on time-domain analysis may inadequately reveal all intrinsic patterns within the data, particularly exhibiting limitations in capturing dynamic variations at different frequencies. Finally, efficiently identifying key features and their inter-correlations in processed data is crucial for improving model performance.

To address the aforementioned challenges and enhance water quality prediction accuracy, we propose a hybrid forecasting model called SMDF2. This model integrates the Savitsky-Golay (SG) filter [20], Multi-seasonal trend decomposition using loss (MSTL) [21], Discrete Fourier Transform (DFT) [22], Frequency-enhanced block (FEB), and Frequency-enhanced attention (FEA) [23]. Key contributions of this work are summarized as follows.

- To capture unsteady and nonlinear features of water quality data, SMDF2 adopts the SG filter to mitigate outliers and noise. Meanwhile, SMDF2 innovatively uses MSTL to decompose water quality data, which extracts multi-seasonal components and trend ones of multiple indicators in the water quality data to capture global patterns of the series.
- SMDF2 combines DFT for time-frequency domain interconversion and uses FEB and FEA for frequency-domain feature extraction and capturing frequency-domain feature correlations.
- Experimental results from both long-term and short-term predictions with actual domestic and international water quality data demonstrate that SMDF2 outperforms other models with respect to accuracy for both single-factor and multi-factor predictions.

The experiment utilized two internationally representative river monitoring datasets:

- Wu Village Station, Langfang, China (39.81444°N, 116.942629°E), a critical node in the North China Plain's water quality monitoring network.
- USGS 02423130, Alabama River at Lees County, USA (32.379000°N,-86.308000°W), a federally managed station in the southeastern U.S. agricultural belt.

To clarify, we highlight key differences between the current work and our previous study [24] as follows.

- The study in [24] exclusively employs attention for feature extraction and learning, whereas this work introduces convolution layers in the FEB module to capture features. Integrating a convolutional neural network (CNN) with self-attention enables simultaneous consideration of local and global information in handling time series data, enhancing both the model's prediction accuracy and its generalization ability.
- The study in [24] only uses a dataset spanning from August 2018 to December 2021 from a river water quality automatic station in Hebei Province. Unlike that, this work additionally uses a water quality dataset obtained from a river segment of the Alabama River spanning from May 2017 to August 2019 for validating the prediction accuracy of SMDF2.
- The study in [24] exclusively conducts multi-element water quality prediction for the potential of hydrogen (pH). In contrast, this work introduces single-element water quality prediction. It extends multi-element experiments to include total nitrogen (TN) prediction, offering a more thorough evaluation of SMDF2's prediction performance.

2. Material and methods

2.1. Dataset description

Experiments adopt two data sets for single-element and multi-element prediction, respectively. The data for the single-element experiment includes dissolved oxygen (DO) detected by USGS Station 02423130 on the Alabama River at Lees County, USA (32.379000°N,

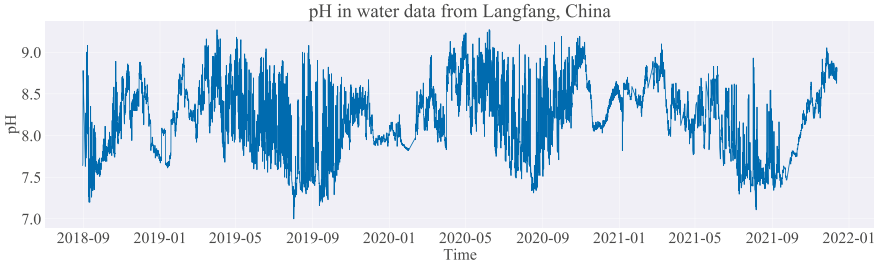

Fig. 2. Time series of potential of hydrogen (pH) in surface waters of Langfang, Hebei, China.

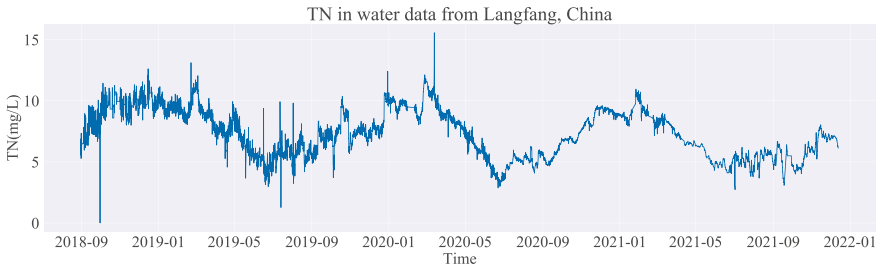

Fig. 3. Time series of total nitrogen (TN, mg/L) in surface waters of Langfang, Hebei, China

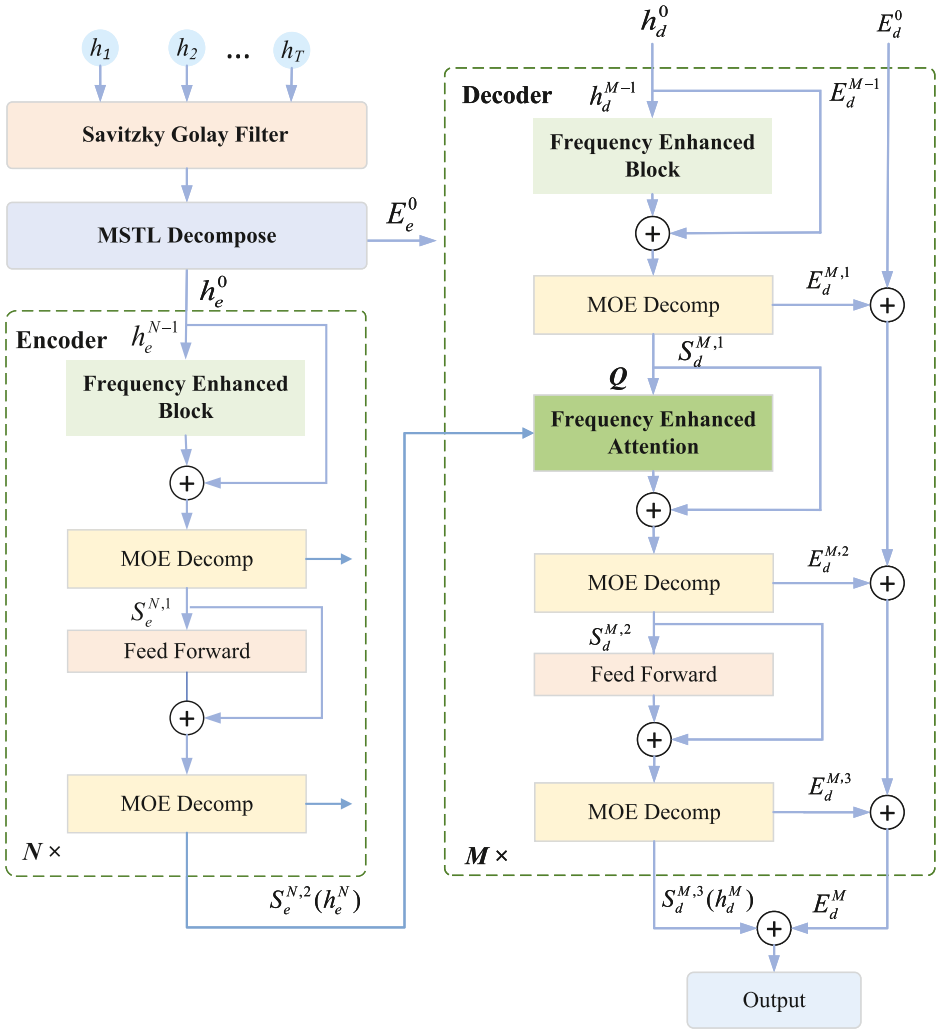

Fig. 4. Structure of SMDF2. SMDF2 is composed of N encoders and M decoders. h represents the time sequence node, S represents the seasonal component, and E represents the trend component. The lower corner marker represents the encoder or decoder. The upper corner marker represents the number of layers passed through, *e.g.*, $S^{M,3}$ denotes the seasonal component output by the third decomposition block after the decoder in the M th layer.

-86.308000°W). Exploration Service from May 2017 to Aug. 2019, with a time point interval of one hour and more than 19,800 samples in total. The multi-element experiments utilize real-time data collected from Wu Village Automatic Monitoring Station in Langfang, Hebei Province, China (39.81444°N, 116.94263°E). The data collection period spans from Aug. 2018 to Dec. 2021. This high-resolution regional monitoring exemplifies the increasing emphasis on quality data in hydrology. As underscored by [25] in their global isotope dataset analysis, robust observational foundations are essential for advancing predictive accuracy. The dataset comprises approximately 6500 samples, and the sampling interval is 4 h. Water quality indicators include pH, TN, DO, and electrical conductivity (EC). The DO time series in the Alabama dataset and pH and TN in the Hebei dataset are presented in Figs. 1–3, respectively.

2.2. SMDF2

This section outlines the overall architecture of SMDF2, as shown in Fig. 4. The design of SMDF2 integrates the SG filter, MSTL, DFT, FEB, and FEA modules to address the challenges of instability, high nonlinearity, and multi-scale variability in water quality time series. The integration is motivated by the complementary strengths of these

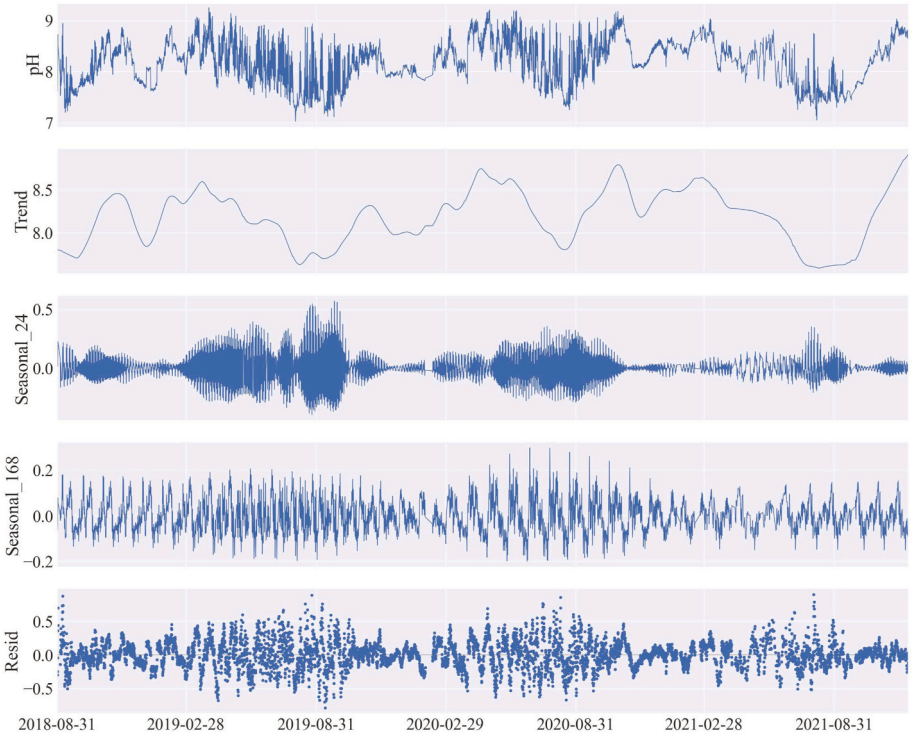

Fig. 5. Example MSTL decomposition of pH data into trend, seasonal components (periods 24 and 168), and residual.

components. The SG filter effectively reduces high-frequency noise and outliers without distorting underlying trends, which is crucial for non-stationary environmental data. The MSTL algorithm enables decomposition into multiple seasonal, trend, and residual components, thereby capturing complex periodicity and improving robustness under environmental variability. The DFT facilitates time–frequency domain transformation, revealing periodicity, frequency distribution, and other informative spectral features for prediction. Building on evidence that sophisticated feature integration frameworks, such as the trustworthy multi-focus fusion strategy of [26] can significantly enhance classification and prediction in complex environmental systems, SMDF2 incorporates attention mechanisms in FEB and FEA to strengthen nonlinear feature extraction and correlation learning across nodes. These modules extend the capabilities of conventional self- and cross-attention by enhancing frequency-domain feature learning, enabling the model to capture both local and global nonlinear correlations between nodes in multi-element water quality series.

2.2.1. Savitsky-Golay filter

The SG filter is applied to smooth the water quality data. A polynomial curve is fitted with least squares within a given data window, and the smoothed data points are obtained by calculating the derivatives of polynomials. The SG filter is computationally fast. It has low storage requirements and can extract valuable information from the underlying trends. It does not alter the underlying shape and length of the time series. In addition, the SG filter can remove noise and unwanted signals from the data samples, reducing interference with the prediction results. The selection of the window size g and polynomial orders p was determined empirically. A comparative analysis of different parameter settings is presented in Section 3.1. The formula of the SG filter is shown as follows:

$$\overline{h}[\epsilon] = \sum_{k=0}^{2m} c[k] \cdot h[\epsilon + k - m] \tag{1}$$

where the length of the filter window is $g = 2m + 1$, and each measurement point is $h = (-m, -m+1, \ldots, 0, 1, \ldots, m-1, m)$. $\overline{h}[\epsilon]$ is the ϵth output data point. The data points from $h[\epsilon - m]$ to $h[\epsilon + m]$ are neighboring to the ϵth data point. c is a vector of coefficients of the SG filter, which depends on the window length (g) and the polynomial order (p). g determines the number of neighboring data points used for p, while p determines the order of the fitted polynomial.

2.2.2. Multi-seasonal trend decomposition using loss

Water quality indicators are affected by seasons, climate, human factors, and usually exhibit complex seasonal patterns that may vary across time scales. As demonstrated in studies of environmental variability [27], such complexity necessitates advanced decomposition methods. Existing sequence decomposition techniques can only extract a single seasonal component, limiting their accuracy. In contrast, MSTL allows for multiple seasonal parameters to be specified, enabling separation of complex seasonal patterns. To fully capture the periodicity of water quality series and enhance prediction robustness under environmental variability, SMDF2 employs the MSTL algorithm to decompose the series into multiple seasonal, trend, and residual components. MSTL obtains extremely accurate seasonal components with low computational cost. The formula is given as:

$$Y_t = E_t + S_1 t + \ldots + S_n t + R_t \quad t = 1, \ldots, T \tag{2}$$

$$F_E = \max\left(0, 1 - \frac{\mathrm{Var}(R_t)}{\mathrm{Var}(E_t + R_t)}\right) \tag{3}$$

$$F_S = \max\left(0, 1 - \frac{\mathrm{Var}(R_t)}{\mathrm{Var}(S_t + R_t)}\right) \tag{4}$$

where t represents a specific time point, Y is a time series containing T time points, E_t signifies the trend value obtained by decomposing Y at time point t, Sn_t indicates the nth seasonal value acquired through decomposition at time point t, and R_t represents the residual value derived from decomposing Y at time point t. F_E and F_S represent trend intensity and seasonal intensity, respectively.

The trend component captures lower-frequency or long-lasting

changes in the time series, while the seasonal component represents changes that occur at regular intervals. The residual component represents the remaining value of the original time series after removing both the trend and seasonal components. The predicted pH values are decomposed using MSTL, as shown in Fig. 5. In implementation, the MSTL decomposition was performed using the stats models library [28]. The seasonal periods were set to (24, 168), which corresponds to daily and weekly cycles when the data are sampled at hourly resolution. For the dataset with four-hour intervals, the same configuration equivalently captures fluctuations at approximately four-day and four-week scales. The trend window was set to 101 to ensure sufficient smoothing of long-term variations, while seasonal smoothing was constrained to a constant level. This configuration allows MSTL to effectively separate high-frequency daily variations from longer-term weekly fluctuations, while preserving interpretable trend and residual components.

2.2.3. Discrete Fourier transform

Frequency-domain signals can better reveal features such as periodicity, frequency distribution, and energy distribution of the data, which provide more informativeness during the model training. Consequently, the time-domain signals are transformed into frequency-domain ones for feature learning. DFT is applied to FEB and FEA in SMDF2. In implementation, instead of utilizing the entire frequency spectrum, a fixed number of 64 frequency modes are retained. This strategy reduces computational complexity to a linear scale with respect to sequence length, while ensuring that the main periodic information of the series is preserved. The time domain is converted to the frequency domain with the discrete Fourier variation $\mathscr{D}$ in (5). The inverse discrete Fourier variation $\mathscr{D}^{-1}$ realizes the conversion of the frequency domain back to the time domain in (6).

$$h_l = \sum_{n=0}^{T-1} h_n e^{-j2\pi l\frac{n}{T}}, \quad l = 1, 2, \ldots, L \tag{5}$$

$$h_n = \sum_{L=0}^{L-1} h_l e^{-j2\pi l\frac{n}{T}}, \quad n = 1, 2, \ldots, T \tag{6}$$

where h_n is the nth time point in the time series. j represents an imaginary unit, and h_l is the lth complex number in the frequency domain. $e^{-j2\pi l\frac{n}{T}}$is a complex exponential function representing the complex amplitude value and phase information in h_l.

2.2.4. Frequency enhanced block

FEB is similar to the traditional Transformer's self-attention, which realizes the extraction of local features from the input time series. However, SMDF2 uses a convolutional layer in FEB, which is more useful for capturing local features in the time series. FEB is employed in both the encoder and decoder. Its functions are frequency domain projection, random sampling, feature learning, and frequency domain complementation.

- Frequency domain projection: First the input $\boldsymbol{h} \in \mathbb{R}^{N\times D}$is linearly projected with$\boldsymbol{w} \in \mathbb{R}^{D\times D}$to obtain $\boldsymbol{x}$, *i.e.*,$\boldsymbol{x} = \boldsymbol{h}\cdot\boldsymbol{w}$, adjusting the length and increasing the nonlinearity. Then convert $\boldsymbol{x}$ from the time domain to the frequency domain $\boldsymbol{X}$ by DFT.
- Random sampling: This is done in the frequency domain with a very small amount of Q randomly selected for retention. Typically, this type of sampling is lossy to the input. However, this loss has minimal impact on the accuracy of the final prediction. Due to the sparse nature of frequency domain information and the prevalence of random noise in the high-frequency components, much of this information can be discarded in time series prediction tasks. Random sampling significantly reduces the length of the input vectors, thus decreasing computational complexity and information redundancy, so a random selection is employed:

$$\widetilde{\boldsymbol{X}} = \text{Choose}(\boldsymbol{X}) = \text{Choose}(\mathscr{D}(x)) \tag{7}$$

- Feature learning: Randomly initialize the matrix $\boldsymbol{I}$ and multiply it with randomly chosen frequency components$\widetilde{\boldsymbol{X}}$.
- Frequency domain complementation: This process is the reverse of frequency-domain sampling. To ensure that the projection back into the frequency domain matches the dimensions of the original input signal, the result of $\widetilde{\boldsymbol{X}}\odot\boldsymbol{I}$ must be zero-padded. Finally, it is transformed back to the time domain using an inverse Fourier transform. In practice, the number of retained frequency modes is fixed at 64 to control complexity, and zero-padding is applied to restore the original sequence length before the inverse transform.

FEB is defined as:

$$\text{FEB}(x) = \mathscr{D}^{-1}(\text{Padding}(\widetilde{\boldsymbol{X}}\odot\boldsymbol{I})) \tag{8}$$

where $\boldsymbol{A}$ is temporary matrix, *i.e.*, $\boldsymbol{A} = \widetilde{\boldsymbol{X}}\odot\boldsymbol{I}and\boldsymbol{A} \in \mathbb{C}^{M\times D}$. Its element $A_{m,do}$, $m\in(1, 2, \ldots, M)$ and $d_o\in(1, 2, \ldots, D)$, is given as:

$$A_{m,d_o} = \sum_{d_i=1}^{D} \widetilde{X}_{m,d_i} I_{d_i,d_o,m}, \quad d_i \in (1, 2, \ldots, D) \tag{9}$$

2.2.5. Frequency domain enhanced attention

FEA is similar to cross-attention in a traditional Transformer, where the signals from the encoder and decoder are subjected to cross-attention operation to capture the intrinsic relationships between these two sets of features. The implementation process of FEA is similar to that of FEB, which is frequency domain projection, random sampling, feature learning, and frequency domain complementation. Initially, the FEA inputs include the keys $(\boldsymbol{x} \in \mathbb{R}^{L\times D})$ and values $(\boldsymbol{v} \in \mathbb{R}^{L\times D})$ from the encoder, as well as the queries $(\boldsymbol{u} \in \mathbb{R}^{L\times D})$ from the decoder. FEA is defined as:

$$\begin{aligned}\widetilde{\boldsymbol{X}} &= \text{Choose}(\mathscr{D}(x))\\ \widetilde{\boldsymbol{V}} &= \text{Choose}(\mathscr{D}(v))\\ \widetilde{\boldsymbol{U}} &= \text{Choose}(\mathscr{D}(u))\end{aligned} \tag{10}$$

$$\text{Atten}(\widetilde{\boldsymbol{X}}, \widetilde{\boldsymbol{V}}, \widetilde{\boldsymbol{U}}) = \text{Softmax}\left(\frac{\widetilde{\boldsymbol{X}}\widetilde{\boldsymbol{V}}^{\top}}{\sqrt{d_{\tilde{\boldsymbol{V}}}}}\right)\widetilde{\boldsymbol{U}} \tag{11}$$

$$\text{FEA}(\boldsymbol{x}, \boldsymbol{v}, \boldsymbol{u}) = \mathscr{D}^{-1}(\text{Padding}(\text{Atten}(\widetilde{\boldsymbol{X}}, \widetilde{\boldsymbol{V}}, \widetilde{\boldsymbol{U}}))) \tag{12}$$

where $\widetilde{\boldsymbol{X}}$, $\widetilde{\boldsymbol{V}}$ and $\widetilde{\boldsymbol{U}}$ are the frequency domain signals obtained by DFT of $\boldsymbol{x}$, $\boldsymbol{v}$and$\boldsymbol{u}$.

2.2.6. Mixture of experts decomposition blocks

SMDF2 employs MOE for progressive decomposition in both encoder and decoder in addition to global MSTL multi-seasonal decomposition of water quality time series. MOE contains a set of averaging filters of different sizes that progressively separate the trend and the period from the inputs to achieve optimization of the prediction results alternating with the decomposition of the series. MOE is denoted as:

$$\mathbf{H}_{\text{trend}} = \text{Softmax}(L(x))^{*}(P(x)) \tag{13}$$

where the weights of the mixed trends are obtained by applying the Softmax to $L(x)$, and $P(\cdot)$ is the average filter with various sizes.

3. Results and discussion

3.1. Experimental setup

During the training stage, hyperparameters are selected through a

Table 1
MSE and MAE of SG filters with different g and p.[a]

g	p	Alabama dataset		Hebei dataset	
		MSE	MAE	MSE	MAE
5	**3**	**0.158**	**0.232**	**0.186**	**0.276**
7	3	0.172	0.276	0.209	0.313
9	3	0.273	0.315	0.302	0.397
7	5	0.281	0.323	0.335	0.426
9	5	0.169	0.247	0.204	0.311
11	5	0.183	0.251	0.217	0.329
9	7	0.189	0.263	0.225	0.341
11	7	0.162	0.235	0.193	0.279
		0.238	0.303	0.265	0.354

[a] Note: Values in bold indicate the best results.

Table 2
SMDF2's Performance under different temporal split ratios.

Setting	Split ratio	MSE	MAE
single-element	70/10/20	$0.159 \pm 2.5 \times 10^{-5}$	$0.232 \pm 2.8 \times 10^{-5}$
	65/10/25	$0.157 \pm 5.3 \times 10^{-6}$	$0.241 \pm 1.9 \times 10^{-6}$
	75/10/15	$0.161 \pm 3.3 \times 10^{-6}$	$0.239 \pm 2.8 \times 10^{-6}$
multi-element	70/10/20	$0.181 \pm 1.4 \times 10^{-7}$	$0.269 \pm 5.4 \times 10^{-7}$
	65/10/25	$0.179 \pm 2.4 \times 10^{-7}$	$0.243 \pm 1.5 \times 10^{-7}$
	75/10/15	$0.181 \pm 2.4 \times 10^{-5}$	$0.272 \pm 1.1 \times 10^{-5}$

systematic process combining literature-based initial ranges, preliminary sensitivity analysis, and iterative comparisons of prediction results, and final values are set as follows. The number of attention heads is 8, the batch size is 64, the number of input features in the encoder or decoder is 512, the learning rate is 0.0001, the number of encoder layers is 3, and the number of decoder layers is 2. The input step size is selected from the set of {12, 24, 36, 48, 72, 96}. For the SG filter, its window size is 5, and its polynomial order is 3. These values are chosen based on their impact on MSE and MAE over multiple validation runs, balancing accuracy and computational efficiency. Table 1 presents the MSE and MAE results for SMDF2 with different SG filter window lengths (g) and polynomial orders (p) with the prediction step of 6.

To prevent time series data leakage, the dataset is split into training, validation, and testing sets strictly according to chronological order, with a ratio of 7:1:2. No shuffling is applied before splitting, ensuring that validation and testing data points are always temporally ahead of the training set. All preprocessing steps are fitted using only the training set statistics and then applied to the validation and test sets, thereby avoiding any inadvertent exposure of future information during training.

To assess robustness with respect to the temporal split, we evaluate three alternative cut-offs for both single-element and multi-element settings: 70/10/20 (baseline), 65/10/25, and 75/10/15 for train/val/test. Each configuration is independently trained and evaluated five times, and the mean ± standard deviation of MSE and MAE for prediction step 6 are reported. As shown in Table 2, results are consistent across cut-offs and across single-element and multi-element settings, indicating that conclusions are not artifacts of a particular split.

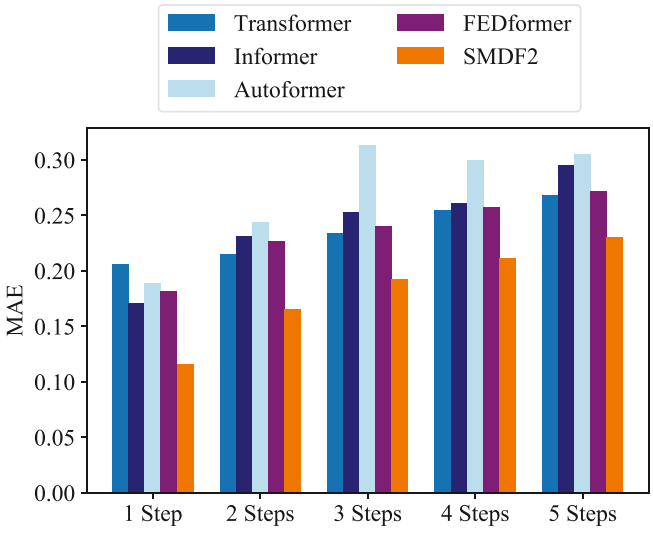

Fig. 7. MAEs of different models for DO prediction at horizons from 1 to 5 steps.

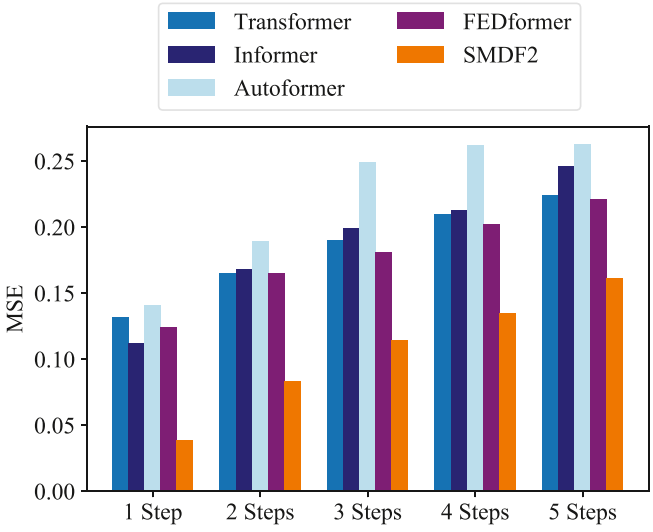

Fig. 8. MSEs of different models for DO prediction at horizons from 1 to 5 steps.

3.2. Model prediction and comparison

To evaluate SMDF2's performance in water quality time series prediction, three advanced deep learning models, including Transformer, Autoformer, and Informer, are selected for comparison. Additionally, recent studies on fuzzy neural networks [29] have demonstrated the efficacy of integrating fuzzy logic and neural networks for handling

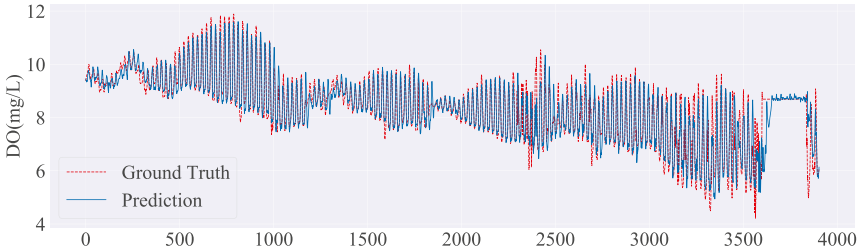

Fig. 6. Comparison of ground-truth DO (mg/L) values and SMDF2-predicted ones for the Alabama dataset.

Table 3
Comparison of prediction results for DO across different models.[a]

Step	Metric	Transformer	Informer	Autoformer	SMDF2
6	MAE	0.265 ± 0.00	$0.283 \pm 3.2 \times 10^{-4}$	$0.324 \pm 4.1 \times 10^{-4}$	$\mathbf{0.232 \pm 2.8 \times 10^{-5}}$
	MSE	$0.219 \pm 7.6 \times 10^{-8}$	$0.230 \pm 7.2 \times 10^{-4}$	$0.272 \pm 9.7 \times 10^{-4}$	$\mathbf{0.158 \pm 2.5 \times 10^{-5}}$
12	MAE	$0.342 \pm 4.0 \times 10^{-8}$	$0.340 \pm 6.8 \times 10^{-4}$	$0.342 \pm 5.7 \times 10^{-3}$	$\mathbf{0.283 \pm 1.6 \times 10^{-5}}$
	MSE	$0.321 \pm 1.7 \times 10^{-8}$	$0.299 \pm 7.1 \times 10^{-4}$	$0.304 \pm 3.6 \times 10^{-3}$	$\mathbf{0.220 \pm 2.9 \times 10^{-5}}$
18	MAE	0.338 ± 0.00	$0.357 \pm 8.9 \times 10^{-4}$	$0.385 \pm 3.8 \times 10^{-3}$	$\mathbf{0.298 \pm 2.6 \times 10^{-5}}$
	MSE	$0.305 \pm 1.2 \times 10^{-8}$	$0.320 \pm 1.0 \times 10^{-3}$	$0.342 \pm 2.1 \times 10^{-3}$	$\mathbf{0.244 \pm 2.9 \times 10^{-5}}$
24	MAE	$0.363 \pm 1.7 \times 10^{-8}$	$0.384 \pm 1.1 \times 10^{-3}$	$0.368 \pm 7.2 \times 10^{-3}$	$\mathbf{0.313 \pm 1.8 \times 10^{-6}}$
	MSE	0.323 ± 0.00	$0.355 \pm 1.6 \times 10^{-3}$	$0.345 \pm 6.9 \times 10^{-3}$	$\mathbf{0.264 \pm 3.2 \times 10^{-6}}$
30	MAE	$0.367 \pm 1.9 \times 10^{-8}$	$0.388 \pm 3.5 \times 10^{-4}$	$0.391 \pm 3.4 \times 10^{-3}$	$\mathbf{0.321 \pm 3.8 \times 10^{-6}}$
	MSE	$0.372 \pm 3.1 \times 10^{-8}$	$0.359 \pm 5.9 \times 10^{-4}$	$0.386 \pm 2.5 \times 10^{-3}$	$\mathbf{0.281 \pm 2.5 \times 10^{-6}}$

[a] Note: Values in bold indicate the best results.

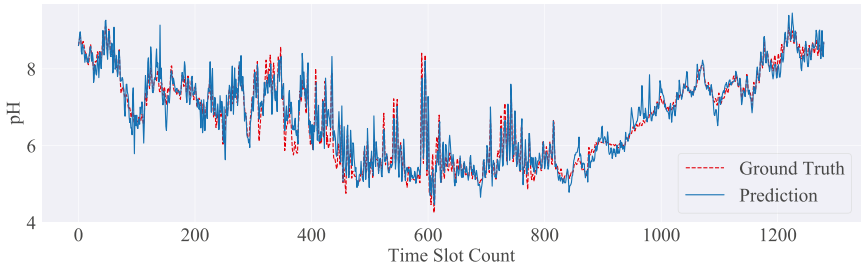

Fig. 9. Comparison of ground-truth pH values and SMDF2-predicted ones for the Hebei dataset.

nonlinear water quality dynamics, providing a relevant parallel to SMDF2's design. Mean absolute error (MAE) [30] and mean squared error (MSE) [31] are used as evaluation metrics.

3.2.1. *Single-element prediction*

The Alabama dataset is used to implement single-element water quality prediction with SMDF2, which predicts future values of DO from its past values. Fig. 6 displays the ground truth and predicted values for DO with the prediction step size of 6. The red line represents the ground-truth result, while the blue line shows the predicted result.

SMDF2 is further validated in the short-term prediction of single-element water quality. Figs. 7 and 8 illustrate the MAE and MSE values for different models across prediction step lengths ranging from 1 to 5. The results demonstrate that SMDF2 outperforms the other models in single-element prediction when the prediction step length is short.

The performance of SMDF2 for single-element long-term water quality prediction is also investigated. Table 3 presents the MAE and MSE results as the mean ± standard deviation over five independent runs for various models across prediction step sizes of 6, 12, 18, 24, and 30. The results indicate that SMDF2 surpasses the other three models, with lower MAE and MSE values and an average improvement in prediction accuracy of 21.73 %. These results suggest that SMDF2 can effectively capture the temporal dependencies and seasonal patterns in single-element water quality series. The notable advantage in short-term forecasting indicates that the model architecture is highly responsive to recent variations in DO, while the consistent superiority in long-term

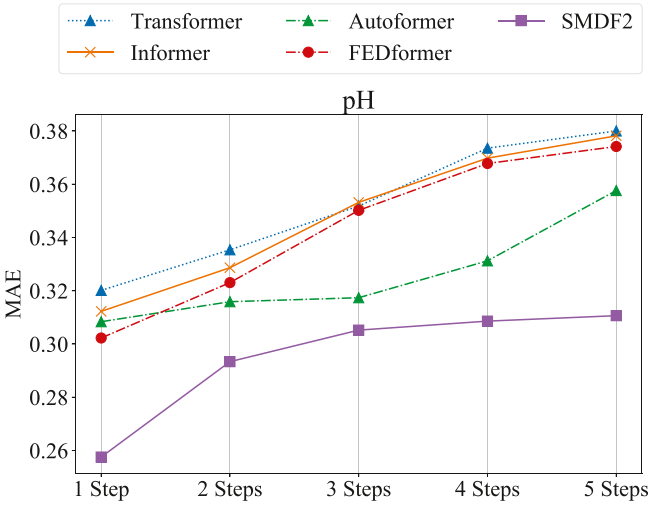

Fig. 11. MAEs of different models for pH prediction at horizons from 1 to 5 steps.

horizons reflects its ability to preserve relevant trend information over extended periods. Beyond the quantitative results, it is worth noting that existing approaches have inherent limitations. CNN have been applied in water quality forecasting, usually perform well in capturing local

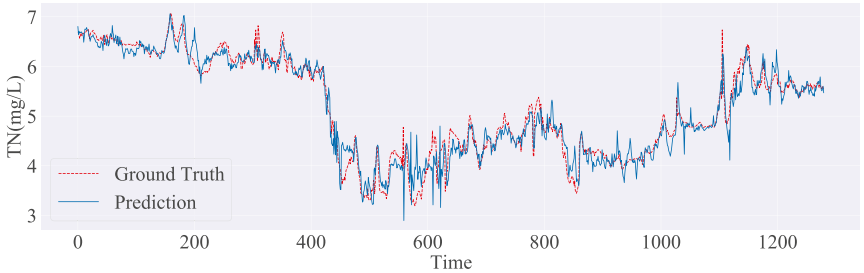

Fig. 10. Comparison of ground-truth TN (mg/L) values and SMDF2-predicted ones for the Hebei dataset.

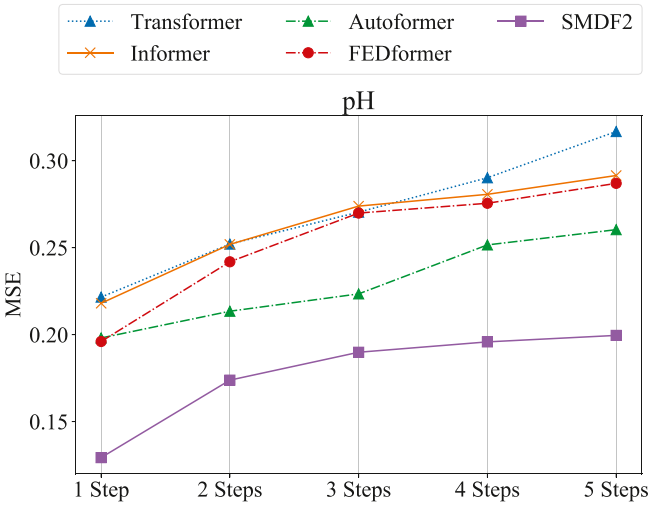

Fig. 12. MSEs of different models for pH prediction at horizons from 1 to 5 steps.

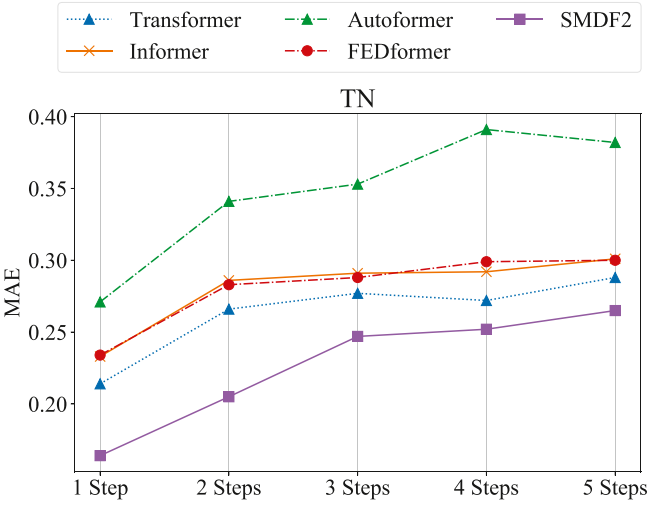

Fig. 13. MAEs of different models for TN prediction at horizons from 1 to 5 steps.

dependencies but struggle with long-term temporal patterns. Transformer-based models, including the baselines tested here, are effective at modeling global dependencies but may suffer from overfitting on small-scale datasets and incur high computational costs. Our results align with these observations, although Transformer variants achieve competitive accuracy, SMDF2 consistently attains lower MAE and MSE, especially for short-term horizons, highlighting its ability to balance accuracy and efficiency in practical applications.

3.2.2. *Multi-element prediction*

SMDF2 is also applied to predict water quality in a multi-element manner. Specifically, pH is predicted with TN, DO, and EC, while TN is predicted with pH, DO, and EC. Figs. 9 and 10 display the ground truth and predicted values for pH and TN, respectively, with a prediction step of 6.

SMDF2 is subsequently validated for multi-element water quality prediction. The MAE values for various models with forecasting step lengths ranging from 1 to 5 for pH are illustrated in Fig. 11, while MSE values are presented in Fig. 12. Similarly, Fig. 13 shows the MAE values for TN prediction across different forecasting step lengths, and Fig. 14 displays the MSE values. These results indicate that SMDF2 outperforms

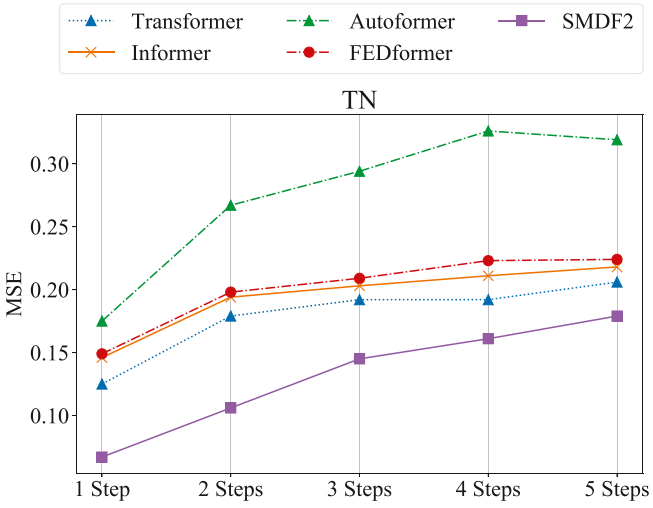

Fig. 14. MSEs of different models for TN prediction at horizons from 1 to 5 steps.

the other three models in multi-element forecasting for shorter prediction lengths.

SMDF2's performance is evaluated in multi-element long-term prediction of water quality series. MAE and MSE are reported as the mean ± standard deviation over five independent runs for various models across prediction step lengths in 6, 12, 18, 24, and 30. The pH and TN prediction results are summarized in Table 4. SMDF2 demonstrates superior performance, achieving the lowest MAE and MSE values, with an average accuracy improvement of 17.95 % for pH prediction and 18.33 % for TN prediction.

The improved performance in multi-element forecasting highlights SMDF2's ability to leverage inter-variable relationships in water quality data. By jointly modeling multiple elements, the model benefits from complementary information, such as the correlation between DO and pH or TN, which can enhance predictive accuracy. The relatively larger gains in short-term horizons suggest that the cross-variable dependencies are most influential in capturing immediate fluctuations, while the sustained advantage in long-term horizons indicates that the decomposition-based architecture is capable of maintaining informative features across extended time spans.

3.2.3. *Robustness under anomalous events*

To further evaluate the robustness of the proposed model under rare or sudden conditions, we identified anomalous events in the test set using thresholds calculated from the training set to avoid data leakage. A point was considered anomalous when its standardized value was greater than three standard deviations from the training set mean, or when the standardized first-order difference was greater than three standard deviations. Consecutive anomalous points were merged into events, and each event was extended with a small context window. For the Alabama dataset with 1-hour sampling, the window length was plus or minus 6 h. For the Hebei dataset with 4-hour sampling, the window length was plus or minus two time steps. Events without enough historical input or a forecast horizon were excluded. Table 5 shows the number of events and the percentage of the test set they cover. We compared the MSE and MAE of each model on three subsets: the full test set, the anomalous subset, and the normal subset. Table 6 presents the ratio between the error on the anomalous subset and the error on the normal subset for both metrics. A smaller ratio means better robustness.

The proposed SMDF2 model consistently has the smallest increase in error across all datasets and prediction horizons.

Table 4
Comparison of prediction results for different models.[a]

Target	Step	Metric	Transformer	Informer	Autoformer	SMDF2
pH	6	MAE	$0.387 \pm 6.3 \times 10^{-8}$	$0.372 \pm 5.4 \times 10^{-3}$	$0.351 \pm 3.2 \times 10^{-3}$	$\mathbf{0.321 \pm 1.4 \times 10^{-7}}$
		MSE	$0.328 \pm 5.1 \times 10^{-8}$	$0.299 \pm 5.6 \times 10^{-3}$	$0.269 \pm 4.5 \times 10^{-3}$	$\mathbf{0.211 \pm 5.4 \times 10^{-7}}$
	12	MAE	$0.455 \pm 1.8 \times 10^{-8}$	$0.384 \pm 2.0 \times 10^{-3}$	$0.391 \pm 1.2 \times 10^{-3}$	$\mathbf{0.349 \pm 3.1 \times 10^{-5}}$
		MSE	$0.382 \pm 2.2 \times 10^{-8}$	$0.325 \pm 1.9 \times 10^{-3}$	$0.321 \pm 2.5 \times 10^{-3}$	$\mathbf{0.248 \pm 2.4 \times 10^{-5}}$
	18	MAE	$0.468 \pm 2.3 \times 10^{-7}$	$0.439 \pm 8.4 \times 10^{-4}$	$0.397 \pm 5.6 \times 10^{-3}$	$\mathbf{0.387 \pm 1.1 \times 10^{-6}}$
		MSE	$0.421 \pm 1.7 \times 10^{-7}$	$0.381 \pm 7.9 \times 10^{-4}$	$0.351 \pm 7.7 \times 10^{-3}$	$\mathbf{0.296 \pm 5.7 \times 10^{-7}}$
	24	MAE	$0.551 \pm 5.5 \times 10^{-8}$	$0.475 \pm 6.8 \times 10^{-4}$	$0.436 \pm 3.9 \times 10^{-3}$	$\mathbf{0.438 \pm 5.6 \times 10^{-7}}$
		MSE	$0.568 \pm 4.1 \times 10^{-8}$	$0.442 \pm 3.7 \times 10^{-4}$	$0.391 \pm 2.8 \times 10^{-3}$	$\mathbf{0.381 \pm 3.9 \times 10^{-7}}$
	30	MAE	$0.575 \pm 7.3 \times 10^{-8}$	$0.567 \pm 1.8 \times 10^{-3}$	$0.466 \pm 6.4 \times 10^{-3}$	$\mathbf{0.447 \pm 8.1 \times 10^{-8}}$
		MSE	$0.597 \pm 8.5 \times 10^{-8}$	$0.525 \pm 1.3 \times 10^{-3}$	$0.438 \pm 7.3 \times 10^{-3}$	$\mathbf{0.396 \pm 3.5 \times 10^{-8}}$
TN	6	MAE	$0.276 \pm 2.9 \times 10^{-8}$	$0.310 \pm 1.0 \times 10^{-3}$	$0.377 \pm 2.2 \times 10^{-4}$	$\mathbf{0.266 \pm 4.6 \times 10^{-8}}$
		MSE	$0.189 \pm 1.7 \times 10^{-8}$	$0.226 \pm 1.0 \times 10^{-3}$	$0.311 \pm 3.5 \times 10^{-4}$	$\mathbf{0.179 \pm 6.8 \times 10^{-8}}$
	12	MAE	$0.352 \pm 4.1 \times 10^{-8}$	$0.379 \pm 2.2 \times 10^{-3}$	$0.419 \pm 5.3 \times 10^{-4}$	$\mathbf{0.317 \pm 4.4 \times 10^{-5}}$
		MSE	$0.273 \pm 5.7 \times 10^{-8}$	$0.296 \pm 1.9 \times 10^{-3}$	$0.380 \pm 4.1 \times 10^{-4}$	$\mathbf{0.242 \pm 2.0 \times 10^{-5}}$
	18	MAE	$0.421 \pm 8.3 \times 10^{-8}$	$0.413 \pm 4.9 \times 10^{-3}$	$0.405 \pm 1.5 \times 10^{-3}$	$\mathbf{0.348 \pm 5.1 \times 10^{-6}}$
		MSE	$0.356 \pm 7.6 \times 10^{-8}$	$0.337 \pm 3.1 \times 10^{-3}$	$0.346 \pm 2.4 \times 10^{-3}$	$\mathbf{0.273 \pm 8.3 \times 10^{-7}}$
	24	MAE	$0.453 \pm 3.1 \times 10^{-8}$	$0.446 \pm 1.7 \times 10^{-4}$	$0.419 \pm 8.7 \times 10^{-4}$	$\mathbf{0.378 \pm 4.6 \times 10^{-8}}$
		MSE	$0.391 \pm 2.7 \times 10^{-8}$	$0.369 \pm 3.7 \times 10^{-4}$	$0.368 \pm 7.4 \times 10^{-4}$	$\mathbf{0.309 \pm 6.9 \times 10^{-8}}$
	30	MAE	$0.515 \pm 5.7 \times 10^{-8}$	$0.464 \pm 4.1 \times 10^{-4}$	$0.416 \pm 6.9 \times 10^{-4}$	$\mathbf{0.403 \pm 4.8 \times 10^{-7}}$
		MSE	$0.492 \pm 6.3 \times 10^{-8}$	$0.400 \pm 2.0 \times 10^{-4}$	$0.360 \pm 7.3 \times 10^{-4}$	$\mathbf{0.332 \pm 2.9 \times 10^{-7}}$

[a] Note: Values in bold indicate the best results.

Table 5
Number and coverage of anomalous events in the test sets.

Dataset	Variable(s)	Number of events	Coverage %
Alabama	DO	45	6.2
Hebei	pH, TN, DO, EC	38	7.5

Table 6
Error ratios between anomalous and normal subsets.[a]

Dataset and Step	Transformer	Informer	Autoformer	SMDF2
Alabama 1 step	1.38	1.34	1.32	**1.16**
Alabama 6 step	1.41	1.36	1.33	**1.18**
Hebei pH 6 step	1.19	1.22	1.20	**1.16**
Hebei TN 6 step	1.23	1.26	1.22	**1.15**

[a] Note: Values in bold indicate the best results.

Table 7
Computational efficiency comparison across models.[a]

Metric	Transformer	Informer	Autoformer	SMDF2
FLOPs	8.89×10^{10}	6.58×10^{10}	$\mathbf{4.44 \times 10^{10}}$	5.51×10^{10}
Number of Parameters	2.21×10^{7}	2.37×10^{7}	2.21×10^{7}	$\mathbf{2.21 \times 10^{7}}$
Single-element Duration	2.317 ± 0.006	$\mathbf{1.998 \pm 0.003}$	6.538 ± 0.205	2.045 ± 0.033
Multi-element Duration	0.863 ± 0.088	0.649 ± 0.025	2.484 ± 0.105	$\mathbf{0.554 \pm 0.005}$

[a] Note: Values in bold indicate the best results.

3.3. Method validation

To ensure that SMDF2 is accurate, practical, and robust, this section presents a comprehensive validation of the proposed method from three perspectives: computational efficiency, component contribution, and training convergence.

3.3.1. Computational efficiency analysis

To further evaluate the practical applicability of SMDF2, its computational complexity, model size, and runtime were compared with those of Transformer, Informer, and Autoformer under a prediction step of 6. Table 7 summarizes the results in terms of floating-point operations (FLOPs), number of parameters, single-element prediction time, and multi-element prediction time.

The results show that SMDF2 requires fewer computations than Transformer and Informer, while maintaining a comparable model size to Transformer and a smaller size than Informer. In single-element prediction, SMDF2 runs significantly faster than Autoformer and achieves performance close to the fastest baseline. In multi-element prediction, SMDF2 demonstrates the shortest runtime among all compared models, highlighting its superior inference efficiency.

Overall, SMDF2 maintains high prediction accuracy while reducing computational cost and improving inference speed. These advantages make it suitable for real-world water quality monitoring applications, especially in scenarios where real-time performance is essential.

3.3.2. Ablation experiment

The effect of SG filters and MSTL on SMDF2 both in single-element experiments and multi-elements is also tested. Specifically, the structure of.

SMDF2 is decomposed step by step, and their effect through experiments is explored. Table 8 shows the result of single-element prediction with the step lengths of 1 to 9 for DO. Table 9 demonstrates the results of multi-element prediction experiments with step lengths of 1 to 9 for pH and TN, respectively. Comparing results with and without the SG filter or MSTL shows that both are essential for improving forecasting performance.

3.3.3. Training convergence analysis

The convergence behavior of SMDF2 is evaluated by comparing its training loss curves with those of baseline models. SMDF2 reaches stable convergence in fewer epochs and maintains a lower loss throughout training, as shown in Fig. 15. The relatively small gap between its training and validation losses indicates good generalization without clear signs of overfitting. This faster and more stable convergence can be linked to the model's design, where the SG filters and MSTL module facilitate efficient feature extraction and optimization.

Overall, the proposed SMDF2 demonstrates stable and strong predictive performance across diverse scenarios, including different geographical regions such as Alabama in the United States and Hebei in China, single-element and multi-element prediction tasks, varied dataset partitions, and anomalous water quality events. These results suggest that the SMDF2 has the potential to generalize beyond the specific conditions seen during training.

Table 8
Ablation result of SG filter and MSTL in SMDF2 for single-element prediction.[a]

Step	Original		SG		MSTL		SG + MSTL	
	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE
1	0.124	0.182	0.052	0.127	0.073	0.146	**0.039**	**0.116**
2	0.165	0.227	0.095	0.187	0.114	0.199	**0.083**	**0.165**
3	0.181	0.240	0.135	0.203	0.156	0.227	**0.114**	**0.192**
4	0.202	0.257	0.150	0.225	0.169	0.239	**0.135**	**0.211**
5	0.221	0.272	0.173	0.244	0.195	0.249	**0.161**	**0.230**
6	0.217	0.279	0.169	0.243	0.190	0.254	**0.158**	**0.232**
7	0.225	0.283	0.186	0.259	0.212	0.273	**0.176**	**0.248**
8	0.231	0.283	0.197	0.268	0.219	0.281	**0.183**	**0.257**
9	0.243	0.296	0.218	0.281	0.233	0.291	**0.197**	**0.265**

[a] Note: Values in bold indicate the best results.

Table 9
Ablation result of SG filter and MSTL in SMDF2 for multi-element prediction.[a]

Target	Step	Original		SG		MSTL		SG + MSTL	
		MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE
pH	1	0.196	0.302	0.133	0.257	0.164	0.278	**0.129**	**0.257**
	2	0.241	0.323	0.204	0.319	0.214	0.199	**0.174**	**0.293**
	3	0.269	0.350	0.241	0.347	0.252	0.358	**0.189**	**0.305**
	4	0.275	0.367	0.251	0.354	0.257	0.359	**0.196**	**0.309**
	5	0.286	0.374	0.261	0.362	0.269	0.364	**0.199**	**0.310**
	6	0.299	0.385	0.263	0.364	0.281	0.372	**0.211**	**0.321**
	7	0.311	0.392	0.277	0.374	0.293	0.380	**0.214**	**0.323**
	8	0.320	0.399	0.285	0.379	0.304	0.391	**0.223**	**0.332**
	9	0.328	0.404	0.296	0.387	0.315	0.398	**0.226**	**0.334**
TN	1	0.149	0.234	0.080	0.187	0.108	0.205	**0.067**	**0.164**
	2	0.198	0.283	0.149	0.257	0.169	0.247	**0.106**	**0.205**
	3	0.209	0.288	0.163	0.262	0.180	0.269	**0.145**	**0.247**
	4	0.223	0.299	0.181	0.275	0.192	0.269	**0.161**	**0.252**
	5	0.224	0.300	0.190	0.283	0.197	0.275	**0.179**	**0.265**
	6	0.237	0.311	0.202	0.293	0.208	0.182	**0.186**	**0.276**
	7	0.241	0.315	0.211	0.301	0.215	0.308	**0.205**	**0.291**
	8	0.254	0.324	0.223	0.309	0.234	0.314	**0.214**	**0.298**
	9	0.260	0.328	0.232	0.316	0.241	0.319	**0.227**	**0.308**

[a] Note: Values in bold indicate the best results.

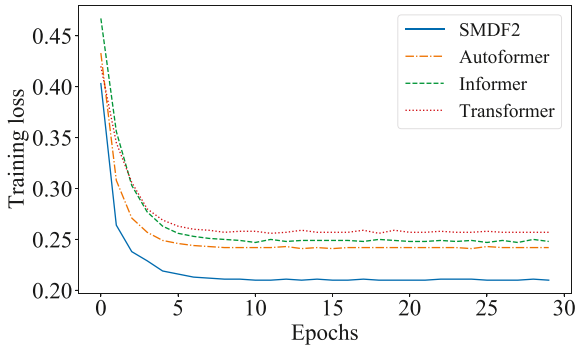

Fig. 15. Training loss curves of SMDF2, Autoformer, Informer, and Transformer models over 30 epochs.

4. Conclusions

The essence of water quality forecasting is time series forecasting, which refers to analyzing past data on water quality and reflecting future trends of indicators in water. Water quality may vary due to seasonal changes, weather conditions, and different aquatic environments, leading to nonlinearity in water quality data. Traditional prediction models are unable to capture the nonlinear features. This work introduces a hybrid water quality prediction model called SMDF2, which integrates the Savitsky-Golay filter, Multi-seasonal trend decomposition using loss, Discrete Fourier Transform, Frequency-enhanced block, and Frequency-enhanced attention, serving for eliminating noise and outliers, extracting multi-seasonal components and trend components for time series, frequency domain time domain interconversion, and capturing serial correlation, respectively. Experimental results with real-world water quality datasets show that SMDF2 achieves higher accuracy in single-element and multi-element prediction than the other three state-of-the-art forecasting models.

Future research will fully consider the broader factors influencing water environments, including meteorological conditions, hydrological data, and pollutant emission data. [32,33] have clearly demonstrated the impact of atmospheric moisture on precipitation, which is also a key factor influencing water quality indicators. Therefore, by integrating these data with water quality time-series data using advanced multimodal data fusion techniques [34], complementary information can be extracted from multi-source data, thereby enhancing the accuracy of predictive models. Additionally, to optimize accuracy and computational efficiency in complex models, we will explore intelligent parameter tuning based on efficient optimization frameworks [35], ultimately supporting more scientifically informed environmental monitoring and evidence-based decision-making.

CRediT authorship contribution statement

Yibo Li: Project administration, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **Ziqi Wang:** Formal analysis, Data curation, Conceptualization. **Haitao Yuan:** Methodology, Investigation, Funding acquisition, Formal analysis. **Jing Bi:** Funding acquisition, Data curation, Conceptualization.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgement

This work was supported by the National Natural Science Foundation of China under Grants 62473014 and 6217301, the Beijing Natural Science Foundation under Grants L233005 and 4232049, and Beihang World TOP University Cooperation Program.

Data availability

Data will be made available on request.

References

[1] G. Wang, H. Chen, H. Han, J. Bi, J. Qiao, E.B. Tirkolaee, Predicting water quality with nonstationarity: event-triggered deep fuzzy neural network, IEEE Trans. Fuzzy Syst. 32 (5) (2024) 2690–2699.
[2] G. Barletta, S. Moitra, S. Derrible, A. Mathew, A.M. Nair, C.M. Megaridis, Exploring machine learning models to predict atmospheric water harvesting with an ion deposition membrane, J. Water Process Eng. 72 (2025) 107476.
[3] D. Wu, D. Sotnikov, M. Haajari, T. Salmi, Minimizing conductor consumption in high-field hts solenoid design using adaptive ann-based optimization algorithm, IEEE Trans. Appl. Supercond. 35 (5) (2025) 1–5.
[4] L. Kong, G. Li, W. Rafique, S. Shen, Q. He, M.R. Khosravi, R. Wang, L. Qi, Time-aware missing healthcare data prediction based on arima model, IEEE/ACM Trans. Comput. Biol. Bioinform. 21 (4) (2024) 1042–1050.
[5] L. Wang, L. Chen, S. Jin, C. Li, Forecasting the green behaviour level of chinese enterprises: A conjoined application of the autoregressive integrated moving average (arima) model and multi-scenario simulation, Technol. Soc. 81 (2025) 102825.
[6] J. Bi, H. Yuan, S. Li, K. Zhang, J. Zhang, M. Zhou, Arima-based and multiapplication workload prediction with wavelet decomposition and savitzky–golay filter in clouds, IEEE Trans. Syst. Man. Cybern. Syst. 54 (4) (2024) 2495–2506.
[7] L.T. Zhou, G.L. Long, C.C. Hu, K. Zhang, Reservoir water level prediction using combined CEEMDAN-FE and Run-SVM-RBFNN machine learning algorithms, Water Sci. Eng. 18 (2) (2025) 177–186.
[8] J. Bi, Z. Chen, H. Yuan, J. Zhang, Accurate water quality prediction with attention-based bidirectional lstm and encoder–decoder, Expert Syst. Appl. 238 (2024) 121807.
[9] Z. Ma, H. Zhang, J. Liu, Db-rnn: an rnn for precipitation nowcasting deblurring, IEEE J. Select. Top. Appl. Earth Observ. Remot. Sens. 17 (2024) 5026–5041.
[10] R. Chai, D. Liu, T. Liu, A. Tsourdos, Y. Xia, S. Chai, Deep learning-based trajectory planning and control for autonomous ground vehicle parking maneuver, IEEE Trans. Autom. Sci. Eng. 20 (3) (2023) 1633–1647.
[11] J. Bi, Z. Guan, H. Yuan, J. Zhang, Improved network intrusion classification with attention-assisted bidirectional lstm and optimized sparse contractive autoencoders, Expert Syst. Appl. 244 (2024) 122966.
[12] J. Xiao, X. Cui, X. Liu, H. Fang, P. Li, Improved 3-d lstm: a video prediction approach to long sequence load forecasting, IEEE Trans. Smart Grid 16 (2) (2025) 1885–1896.
[13] S. Mao, H. Li, Y. Zhang, Y. Shi, Prediction of ionospheric electron density distribution based on cnn-lstm model, IEEE Geosci. Remote Sens. Lett. 21 (2024) 1–5.
[14] M.S. Pandianchery, V. Sowmya, E.A. Gopalakrishnan, V. Ravi, K.P. Soman, Centralized cnn–gru model by federated learning for covid-19 prediction in india, IEEE Trans. Comput. Soc. Syst. 11 (1) (2024) 1362–1371.
[15] P. Arora, S.M.J. Jalali, S. Ahmadian, B.K. Panigrahi, P.N. Suganthan, A. Khosravi, Probabilistic wind power forecasting using optimized deep auto-regressive recurrent neural networks, IEEE Trans. Industr. Inform. 19 (3) (2023) 2814–2825.
[16] N.R. Monteiro, J.L. Oliveira, J.P. Arrais, Tag-dta: binding-region-guided strategy to predict drug-target affinity using transformers, Expert Syst. Appl. 238 (2024) 122334.
[17] S. Ma, J. He, J. He, Q. Feng, Y. Bi, Forecasting air quality index in yan'an using temporal encoded informer, Expert Syst. Appl. 255 (2024) 124868.
[18] J. Ouyang, Z. Zuo, Q. Wang, Q. Duan, X. Zhu, Y. Zhang, Seasonal distribution analysis and short-term pv power prediction method based on decomposition optimization deep-autoformer, Renew. Energy 246 (2025) 122903.
[19] S. Li, Y. Sun, Y. Han, Z. Zhang, M. Yao, J. Chen, Q. Tian, Y. Lin, Cl-mfgcn: Graph structure contrastive learning and multiscale feature fusion graph convolutional network for spectrum prediction, IEEE Internet Things J. 12 (4) (2025) 3600–3612.
[20] L. Zhao, H. Yuan, K. Xu, J. Bi, B.H. Li, Hybrid network attack prediction with Savitzky–Golay filter-assisted informer, Expert Syst. Appl. 235 (2024) 121126.
[21] H. Nan, X. Zhu, J. Ma, Mstl-gltp: a global–local decomposition and prediction framework for wireless traffic, IEEE Internet Things J. 10 (6) (2023) 5024–5034.
[22] C.R. Pochimireddy, A. Siripuram, B. Osgood, Fast dft computation for signals with structured support, IEEE Trans. Inf. Theory 70 (2) (2024) 1498–1524.
[23] J. Hong, Y. Bai, Y. Huang, Z. Chen, Hybrid carbon price forecasting using a deep augmented fedformer model and multimodel optimization piecewise error correction, Expert Syst. Appl. 247 (2024) 123325.
[24] J. Bi, Y. Li, X. Chang, H. Yuan, J. Qiao, Hybrid water quality prediction with frequency domain conversion enhancement and seasonal decomposition, in: 2023 IEEE International Conference on Systems, Man, and Cybernetics (SMC), 2023, pp. 5200–5205.
[25] R. Li, G. Zhu, L. Chen, X. Qi, S. Lu, G. Meng, Y. Wang, W. Li, Z. Zheng, J. Yang, Global stable isotope dataset for surface water, Earth Syst. Sci. Data 17 (5) (2025) 2135–2145.
[26] C. Hu, C. Zhao, H. Shao, J. Deng, Y. Wang, Tmff: trustworthy multifocus fusion framework for multi-label sewer defect classification in sewer inspection videos, IEEE Trans. Circuits Syst. Video Technol. 34 (12) (2024) 12274–12287.
[27] L. Chen, G. Zhu, X. Lin, R. Li, S. Lu, Y. Jiao, D. Qiu, G. Meng, Q. Wang, The complexity of moisture sources affects the altitude effect of stable isotopes of precipitation in inland mountainous regions, Water Resour. Res. 60 (6) (2024).
[28] B.C. Bandara K, R.J. Hyndman, Mstl: a seasonal-trend decomposition algorithm for time series with multiple seasonal patterns, Int. J. Oper. Res. 52 (1) (2025) 79–98.
[29] C. Wei, T. Zhao, J. Cao, P. Li, Water quality prediction model based on interval type-2 fuzzy neural network with adaptive membership function, Int. J. Fuzzy Syst. (2025) 1–15.
[30] J. Bi, S. Li, H. Yuan, M. Zhou, Integrated deep learning method for workload and resource prediction in cloud systems, Neurocomputing 424 (2021) 35–48.
[31] Z. He, H. Shen, W. Xu, Y.C. Eldar, X. You, Mse-based training and transmission optimization for mimo isac systems, IEEE Trans. Signal Process. 72 (3) (2024) 3104–3121.
[32] Q. Wang, Y. Liu, G. Zhu, S. Lu, L. Chen, Y. Jiao, W. Li, W. Li, Y. Wang, Regional differences in the effects of atmospheric moisture residence time on precipitation isotopes over eurasia, Atmos. Res. 314 (2025) 107813.
[33] F. Li, H. Lu, G. Wang, J. Qiu, Long-term capturability of atmospheric water on a global scale, Water Resour. Res. 60 (12) (2024).
[34] N. Song, J. Nie, Q. Wen, Y. Yuan, X. Liu, J. Ma, Z. Wei, Gl-st: a data-driven prediction model for sea surface temperature in the coastal waters of china based on interactive fusion of global and local spatiotemporal information, IEEE J. Select. Top. Appl. Earth Observ. Remot. Sens. 18 (2025) 2959–2974.
[35] K. Du, S. Yang, W. Xu, F. Zheng, H. Duan, A novel optimization framework for efficiently identifying high-quality pareto-optimal solutions: maximizing resilience of water distribution systems under cost constraints, Reliab. Eng. Syst. Saf. 261 (2025) 111136.