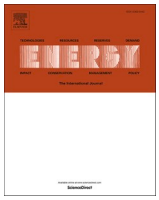
Energy 305 (2024) 132222

Contents lists available at ScienceDirect

Energy

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



NOx concentration prediction based on multi-channel fused spectral temporal graph neural network in coal-fired power plants

Zheng Wu [a,b], Yue Zhang [a,b], Ze Dong [a,b,*]

[a] School of Control and Computer Engineering, North China Electric Power University, Beijing, 102206, China
[b] Hebei Technology Innovation Center of Simulation & Optimized Control for Power Generation, Baoding, 071003, China

ARTICLE INFO

Handling editor: Isabel Soares

Keywords:
NOx concentration prediction
Coal-fired power plant
Pollutant emission
Graph neural network
Modal decomposition
Feature selection

ABSTRACT

Establishing an accurate and stable NOx concentration prediction model is the foundation for achieving environmental protection in coal-fired power plant denitrification. In this research, a data-driven hybrid prediction model (BMFE-MFST-GNN) is proposed to improve the prediction accuracy of inlet NOx concentrations. First, we develop the adaptive variational modal decomposition method (FEVMD) to decompose historical NOx concentration data into simple and smooth subsequences and extract time-frequency features. Second, we propose the boosting mutual information feature selection algorithm (BMIFS) to determine the best set of auxiliary variables. The delay time is calculated based on the maximal information coefficient (MIC) to reconstruct the datasets. Then, the multi-channel fused spectral temporal graph neural network (MFST-GNN) is created to build the graph feature information of decomposed subsequences and reconstructed auxiliary variables to predict the concentration subsequences. Finally, we integrate the subsequence prediction results to obtain the future NOx concentrations. The experimental results demonstrate that the proposed method outperforms several comparative models in predicting NOx concentrations.

1. Introduction

Coal-fired power plants emit large amounts of nitrogen oxides (NOx), which are a major source of air pollution. Rain and snow absorb NOx and other pollutants in the air during the creation process, resulting in acid rain that devastates the natural environment [1]. Simultaneously, NOx generates photochemical smog with other pollutants when exposed to ultraviolet light, endangering human health [2].

Selective catalyst reduction (SCR) is a popular denitrification method for thermal power units [3]. To reduce NOx emissions, NOx in the flue gas and the reducing agent (i.e. generally ammonia) undergo selective catalytic reduction reactions to produce nitrogen (N_2) and water (H_2O) under the action of a catalyst [4]. Excessive or insufficient ammonia spraying will result in ammonia or NOx escape and secondary pollution. However, the continuous emission monitoring system (CEMS) [5] operates in a harsh environment and frequently fails. This results in inaccurate NOx concentration measurements [6]. Therefore, accurate prediction of SCR system inlet NOx concentration is critical for achieving denitrification and environmental protection in coal-fired power plants.

Based on the principles, the NOx concentration prediction methods primarily include mechanistic models [7–9] and data-driven models [10–14]. The mechanistic models mainly construct the corresponding differential equations in the boiler based on chemical reactions and energy transformations [15]. However, the mechanistic model is computationally inefficient because it necessitates making a number of assumptions that are challenging to verify in actual production. As a result, it is difficult to apply the mechanistic models to actual production for quick and accurate concentration prediction [16].

With the introduction of various intelligent systems into power plants, data-driven models such as machine learning [10] and deep learning [12–14] have produced promising results in predicting NOx concentrations. Tuttle et al. [11] reviewed ten data-driven NOx prediction models and conducted a comparative experimental analysis. The results revealed that the GRU model has the best prediction performance. Wang et al. [12] proposed a lightweight convolutional neural network (C3–CNN) to predict inlet NOx concentration. The model utilized a cross-channel communication block to achieve mutual communication between the same convolutional layers, which improved the extraction of hidden features in boiler data. Wang [13] and Tang [14] proposed two recent hybrid models (CEEMDAN-AM-LSTM and AE-ELM). Both models exhibited high accuracy in NOx concentration

* Corresponding author. School of Control and Computer Engineering, North China Electric Power University, Beijing, 102206, China
E-mail addresses: wuzhengncepu@ncepu.edu.cn (Z. Wu), zhyhot_zhyhot@sina.com (Y. Zhang), dongze@ncepu.edu.cn (Z. Dong).

https://doi.org/10.1016/j.energy.2024.132222
Received 19 March 2023; Received in revised form 20 February 2024; Accepted 24 June 2024
Available online 6 July 2024

prediction. However, thermal process data in coal-fired power plants exhibits nonlinearity, coupling, and time lag, and the operating trend at future points is highly unstable [17]. This leads to the fact that the generalization performance and accuracy of the NOx prediction models developed at this stage remain deficient [18]. Therefore, the key to improving the model's prediction performance is to address the problem of volatility in the operating trend of thermal process data and establish effective correlation information between feature sequences.

The NOx concentration sequences can be divided into multiple band-limited intrinsic mode functions (IMFs) with limited bandwidth using variational mode decomposition (VMD) [19]. The decomposed.

Nomenclature

SCR	selective catalyst reduction	CEMS	continuous emission monitoring system
NOx	nitrogen oxides	MIC	maximum information coefficient
GRU	gated recurrent unit	GNN	graph neural network
MIFS	mutual information feature selection	StemGNN	spectral temporal graph neural network
BIMF	band-limited intrinsic mode function	BMIFS	boosting mutual information feature selection
FC	fully connected layer	ARIMA	autoregressive integrated moving average
GFT	graph fourier transform	DFT	discrete fourier transform
IGFT	inverse graph fourier transform	IDFT	inverse discrete fourier transform
BP	back propagation	HVMD	hybrid gray wolf variational modal decomposition
EMD	empirical mode decomposition	FEVMD	adaptive variational modal decomposition
VMD	variational modal decomposition	MVMD	multivariate variational modal decomposition
DELM	deep extreme learning machine	MST-GNN	multidimensional spatio-temporal graph neural network
FFT	fast fourier transform	MFST-GNN	multi-channel fused spectral temporal graph neural network
AM	attention mechanism	AE-ELM	automatic encoder extreme learning machine
LSTM	long short term memory	CEEMDAN	complete ensemble empirical mode decomposition adaptive noise
MAE	mean absolute error	RMSE	root mean square error
MAPE	mean absolute percentage error	R^2	determination coefficient

subsequences with stable time-frequency trends are straightforward to forecast. The VMD outperforms empirical mode decomposition (EMD) [20] in terms of nonlinearity and noise handling. Meng et al. [21] created a hybrid model by combining multivariate variational modal decomposition (MVMD) to improve and optimized extreme learning machine (DELM)'s wind power prediction performance. Korkmaz D et al. [22] used VMD to decompose PV power, and the model's prediction accuracy improved. But right now, one of the main problems with the VMD is how to determine the best value for the decomposition layer and penalty factor.

After the modal decomposition, the effective selection and fusion of correlation information among the thermal process feature data also play an important role in improving the NOx prediction accuracy [23]. Graph neural network (GNN) can use adjacency matrices to establish the connection relationship of topology between feature data, which has obvious advantages for predicting complex industrial data [24,25]. Multiple prediction tasks in power systems have currently been successfully completed utilizing GNN [26–29]. Liu et al. [30] built a NOx emission prediction model using mutual information and GNN, achieving great prediction results. Wu et al. [31] created a multidimensional spatio-temporal graph neural network (MST-GNN) to predict wind speed with high accuracy. Cao et al. [32] proposed a spectral temporal graph neural network model (StemGNN). StemGNN can extract both inter-series correlation and time dependence from spectral data. On all ten datasets, it achieves excellent prediction performance. Therefore, the GNN offers a fresh approach to developing a high-accuracy NOx concentration prediction model.

Table 1
Main parameters of the boiler.

No.	Parameter description	Unit	BMCR	ECR
1	Superheated steam flow	t/h	2008	1775
2	Superheated steam pressure	Mpa	17.47	17.27
3	Superheated steam temperature	°C	540	540
4	Reheat steam flow	t/h	1662.2	1482
5	Reheater inlet pressure	MPa	3.81	3.39
6	Reheater outlet pressure	MPa	3.61	3.21
7	Reheater inlet temperature	°C	320	309
8	Reheater outlet temperature	°C	540	540
9	Economizer inlet feedwater temperature	°C	278	270
10	Economizer inlet feedwater pressure	MPa	19.23	18.70
11	Tank pressure	MPa	18.84	18.36

This paper proposes a hybrid model (BMFE-MFST-GNN) for predicting inlet NOx concentrations. First, we implement the adaptive decomposition of VMD (FEVMD) utilizing discrete fourier transform (DFT) and sample entropy. FEVMD is employed to decompose NOx concentration into multiple modal subsequences. Next, the auxiliary variables are chosen using the boosting mutual information feature selection algorithm (BMIFS). The maximum information coefficient (MIC) is applied to calculate latency and reconstruct the dataset. Finally, the subsequences and auxiliary variables are expanded into a graph structure in non-euclidean space by a multi-channel fused spectral temporal graph neural network (MFST-GNN). Meanwhile, the deep graph neural network effectively fuses feature information among graph nodes to achieve accurate prediction of NOx concentration.

The contributions of this paper are as follows.

- We propose a hybrid prediction model (BMFE-MFST-GNN) for inlet NOx concentration at a coal-fired power plant. A signal decomposition method is innovatively applied to decompose NOx into multiple trend-stable subsequences. The correlation information between thermal process sequences is then established and fused in the form of graph structure by the graph neural network to mitigate the effects of data nonlinearity and coupling.
- An adaptive modal decomposition method (FEVMD) is proposed for decomposing historical NOx concentrations. The discrete fourier transform and sample entropy are utilized to determine the optimal number of decomposition layers and penalty factor.
- The multi-channel fused spectral temporal graph neural network (MFST-GNN) is created to establish the graph structure of decomposed subsequences and feature parameters through a dual-channel form. The mapping relationship between historical and future NOx concentrations is constructed by the deep graph network.
- The boosting mutual information feature selection method (BMIFS) is proposed to determine the best set of auxiliary variables affecting NOx generation. It solves the drawback of the MIFS, which weakens mutual information terms in the later stages of the search.

The rest of this paper is organized as follows: Section 2 describes the boiler in this research and NOx affecting factors. Section 3 provides a discussion of the problem formulation. Section 4 presents the proposed models in detail. Section 5 conducts a comparison experiment. The final section concludes this paper.

2. The boiler and mechanistic knowledge

2.1. Description of the boiler object

This paper focuses on a 600 MW subcritical pressure primary

Table 2
Candidate input variables.

No.	Parameter description	Unit	Notation	Scope
1	Unit load	MW	UL	[178.22–603.13]
2	Main steam temperature	°C	ST	[510.87–546.11]
3	Main steam pressure	MPa	SP	[8.60–16.05]
4	Total coal	t/h	TC	[84.94–259.70]
5	Total air volume	t/h	TA	[1115.1–2214.2]
6	Furnace pressure	kPa	P_f	[-316.85–181.32]
7	Furnace flue gas temperature	°C	T_f	[85.82–175.85]
8	Furnace flue gas oxygen content	%	O_{2f}	[3.69–9.50]
9	Denitrification inlet flue gas flow	t/h	G_d	[532.14–1253.57]
10	Denitrification inlet flue pressure	kPa	P_d	[-939.56–82.42]
11	Denitrification inlet flue temperature	°C	T_d	[284.96–360.32]
12	Denitrification inlet flue oxygen content	%	O_{2d}	[3.82–8.76]
13	Feedwater flow	t/h	W_f	[289.59–1799.11]
14	Ammonia flow rate	m^3/h	A_f	[1.09–79.98]
15	Excess air coefficient	–	E_a	[1.24–1.92]
16–21	A-F coal feeder instantaneous coal feed rate	t/h	$F_{a\text{-}f}$	[0.02–47.88]
22–27	A-F coal mill mixing regulator feedback	%	$M_{a\text{-}f}$	[0.82–99.66]
28–31	1-4 corner AA layer dampers position feedback	%	$AA_{1\text{-}4}$	[19.32–90.93]
32–35	A layer	%	$A_{1\text{-}4}$	[5.22–68.65]
36–39	AB layer	%	$AB_{1\text{-}4}$	[23.90–87.03]
39–42	B layer	%	$B_{1\text{-}4}$	[8.64–64.62]
43–46	BC layer	%	$BC_{1\text{-}4}$	[21.95–87.64]
47–50	C layer	%	$C_{1\text{-}4}$	[7.36–64.87]
51–54	CD layer	%	$CD_{1\text{-}4}$	[4.37–57.23]
55–58	D layer	%	$D_{1\text{-}4}$	[8.94–56.38]
59–62	DE layer	%	$DE_{1\text{-}4}$	[0.52–55.95]
63–66	E layer	%	$E_{1\text{-}4}$	[9.11–51.98]
67–70	EF layer	%	$EF_{1\text{-}4}$	[2.84–32.20]
71–74	F layer	%	$F_{1\text{-}4}$	[8.52–33.06]
75–78	FF layer	%	$FF_{1\text{-}4}$	[0.46–27.99]
79–82	OFA dampers	%	$OFA_{1\text{-}4}$	[3.27–74.15]
83–106	1-4 corner of the secondary air door flap feedback 15-20	%	$S_{1a\text{-}4f}$	[2.84–73.53]
107	Superheat	°C	S_h	[6.51–67.66]
108	Dilution fan flow	m^3/h	D_f	[2421.0–3347.0]
109	Dilution fan current	A	D_c	[0.09–55.10]

intermediate reheat control cycle steam packet boiler from a power plant in Hebei, China. The boiler adopts swing burner for temperature control, positive pressure direct blowing pulverizing system, single hearth, Π-type open-air arrangement. The boiler is closed below the operating floor with tight body, solid slag discharge, all-steel frame structure, and balanced ventilation. The burners are arranged on the front and rear walls, forming a reverse tangent circle. Ammonia is used as the reducing agent to control NOx concentration. Fig. 1 presents the overall structure and production flow of the SCR denitrification plant. Table 1 depicts the boiler's main parameters.

2.2. Factors affecting inlet NOx generation

The combustion conditions in the boiler will directly affect the SCR system's inlet NOx concentration. According to the generation mechanism, NOx generated from boilers can be categorized into thermal NOx, fast NOx, and fuel NOx.

Thermal NOx is primarily defined as nitrogen oxides produced by oxidizing N2 in the air at high temperatures during the combustion process. Temperature is a major factor in the production of thermal NOx. When the combustion temperature exceeds 1600 °C, the production of thermodynamic NOx increases dramatically with increasing temperature.

Table 3
Comparison of operating combustion and design coal properties.

Project	Notation	Unit	Design	Operation
Carbon	C_{ar}	%	62.58	58.41
Oxygen	O_{ar}	%	10.05	9.45
Hydrogen	H_{ar}	%	3.70	3.79
Nitrogen	N_{ar}	%	1.07	0.97
Sulfur	$S_{t.ar}$	%	0.40	0.34
Ash	A_{ar}	%	7.70	11.04
Moisture	M_t	%	14.50	16.00
Volatile matter	V_{daf}	%	37.89	30.83
High level heat generation	$Q_{gr.ar}$	MJ/kg	25.20	24.12
Low level heat generation	$Q_{net.ar}$	MJ/kg	24.00	22.86
Hardgrove grindability index	HGI	–	61.00	50.00

Fast NOx is the rapid formation of nitrogen oxides from hydrocarbons near the reaction zone when fuel is burned too densely. Unlike thermal NOx, fast NOx production has a stronger relationship with pressure than temperature. For large boilers, fast NOx production during coal combustion is negligible.

Fuel NOx are nitrogen oxides generated by the thermal decomposition and oxidation of nitrogen compounds in fuel during the combustion process. Fuel NOx is the main NOx generated by coal-fired power plants, accounting for 60–80 % of total NOx. Its generation is not only affected by coal properties, such as ash and volatile matter, but also by oxygen, air volume, excess air coefficient, temperature, and other boiler combustion conditions, as well as burner structure.

It is known from the analysis that the overall three NOx produced by the fuel combustion process in the power plant are influenced by a number of factors, including temperature, pressure, oxygen, airflow, coal quantity, coal properties, excess air coefficient, and boiler burner arrangement structure. Based on the above mechanism analysis and field engineering experience, we initially identified 109 operating parameters related to inlet NOx concentration as candidate inputs. Table 2 summarizes the candidates' input variables.

The combustion coal for the boiler in this research is Shenfu Dongsheng coal. Table 3 shows the results of a comparison of actual combustion coal properties and design coal properties under the collected experimental dataset. It can be seen that the coal properties is stable during the operation of the experimental data. Meanwhile, parameters such as flue gas oxygen content and air volume can also reflect coal properties information to a certain extent. Therefore, we did not consider coal properties in the candidate variables.

3. Problem formulation

The historical NOx concentration sequence of the SCR system is set to $\mathbf{x}_h = [x_1, x_2, \ldots, x_i]$, and the predicted future NOx concentration sequence is set to $\mathbf{x}_f = [x_i, x_{i+1}, \ldots, x_{i+t}]$. The length of the time series is i. The length of the predicted sequence is t. The quantitative relationship between the historical and future sequences can be described as:

$$x_f = \mathscr{F}(x_h) \tag{1}$$

where $\mathscr{F}(\cdot)$ is the correlation function between $\mathbf{x}_h$ and $\mathbf{X}_f$. Note that, **Eq.** (1) does not consider the effect of other operating parameters of the boiler. In this research, we incorporate other operating parameters as auxiliary variables to improve the accuracy of prediction.

The denitrification process has a certain delay characteristic due to the delay in the denitrification reaction and the system's ammonia injection regulation [33]. Therefore, to eliminate the effect of delayedness, we introduce a delay-related parameter t to reconstruct the unit operating parameters. **Eq.** (1) is rewritten as:

$$\mathbf{x}_f = \mathscr{F}[\mathbf{x}_h, AS(v_h)] \tag{2}$$

$$AS(v_h) = [v_{i-\tau}, v_{i-\tau+1}, \ldots, x_{i-\tau+i}]^T \tag{3}$$

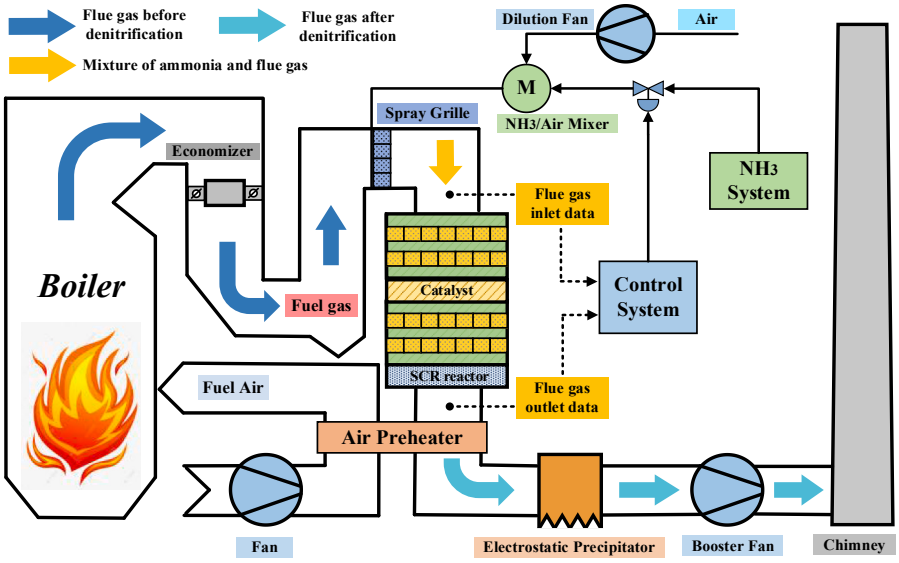


Fig. 1. The overall structure and production flow of the SCR denitrification plant.

where v_h is the set of auxiliary variables. $AS(\cdot)$ denotes the reconstruction function.

The inlet NOx concentration sequence is a broadband time series with unstable future trends. This makes it more difficult to construct the function. It is generally recognized that narrowband time series demonstrate more stable future trends.

Therefore, we adopt the signal decomposition method to decompose the NOx concentration into multiple narrowband time series. The decomposed subsequences have a stable time series trend, and the model can extract time-frequency information from the sequences more effectively. The decomposition process is expressed as:

$$\mathbf{x}_f^k = \mathscr{F}_k\left[\mathbf{x}_h^k, AS(v_h)\right] \tag{4}$$

$$\mathbf{x}_f = \sum_K \mathbf{x}_f^k \tag{5}$$

where $\mathbf{x}_h^k$ is the historical NOx data segment for the mode k. $\mathbf{x}_f^k$ is the future prediction segment of the mode k. $\mathscr{F}_k(\cdot)$ is the correlation function between $\mathbf{x}_h^k$ and $\mathbf{x}_f^k$. K is the number of modal decompositions.

Based on Eqs. (4) and (5), we propose three steps to predict NOx concentrations.

Step 1 . Decompose historical NOx concentrations into trend-stable subsequences using the signal decomposition method.

Step 2 . Based on the decomposed subsequences $\mathbf{x}_h^k$ and the reconstructed auxiliary variables $AS(v_h)$, the correlation function $\mathscr{F}_k(\cdot)$ is fitted utilizing the deep learning to calculate $\mathbf{x}_f^k$.

Step 3 . Sum $\mathbf{x}_f^k$ to establish the predicted NOx concentration $\mathbf{x}_f$.

Combined with the above analysis, we propose FEVMD, BMIFS, and MFST-GNN, and then combine these methods for the NOx concentration prediction task. MIC is also used to calculate latency and reconstruct the dataset. These methods will be presented one at a time in the section that follows.

4. Methodology

4.1. FEVMD

In this research, an adaptive variational modal decomposition method (FEVMD) is proposed to decompose the inlet NOx concentration sequences into simple and smooth subsequences. The decomposed subsequences have different local time-frequency feature information, which allows the hybrid model to extract time series information more easily.

4.1.1. Basic theory

The original time series is set to $x(t)$, and the mode $u_k(t)$ is a finite bandwidth of k with central frequency decomposed by the input signal. It is assumed that all of the components are narrowband signals with peaks near their respective center frequencies. Dragomiretskiy K et al. [19] defined the constrained variational problem in three steps: resolving the signal with the Hilbert transform, flattening the signal to baseband, and estimating the bandwidth using H_1 gaussian smoothness. The above problem can be described by **Eq.** (6):

$$\begin{cases} \min_{\{u_k\},\{\omega_k\}} \left\{ \sum_k \left\| \partial_t \left[\left(\delta(t) + \dfrac{\mathrm{j}}{\pi t} \right) * u_k(t) \right] \mathrm{e}^{-\mathrm{j}\omega_k t} \right\|_2^2 \right\} \\ \text{s.t.} \quad \sum_k u_k = x(t) \end{cases} \tag{6}$$

where $\delta(t)$ is the dirac distribution at time t. ∂_t is the partial derivative. * is the convolution operator. $\{u_k\} = \{u_1, \ldots, u_k\}$ is the decomposition of each modal component. $\{\omega_k\} = \{\omega_1, \ldots, \omega_k\}$ is the center frequency of each modal component.

The penalty factor term and the lagrangian operator term are introduced to reconstruct the variational bounding problem, and the reconstructed unboundedness model is:

$$L(\{u_k\},\{\omega_k\},\lambda) = \alpha \sum_k \left\| \partial_t \left[\left(\delta(t) + \frac{\mathrm{j}}{\pi t} \right) \times u_k(t) \right] \mathrm{e}^{-\mathrm{j}\omega_k t} \right\|_2^2 + \left\| x(t) - \sum_k u_k(t) \right\|_2^2 + \left[\lambda(t), x(t) - \sum_k u_k(t) \right] \tag{7}$$

where $L(\cdot)$ is the lagrangian function. α is the penalty factor, and it ensures the reconstruction accuracy of the signal in the presence of gaussian noise. $\lambda(t)$ is the lagrangian multiplication operator.

The modal components $u_k(t)$, center frequency ω_k, and lagrangian operator $\lambda(t)$ are updated alternately by frequency iterations to find the optimal solution of the modal. The update equation is given by:

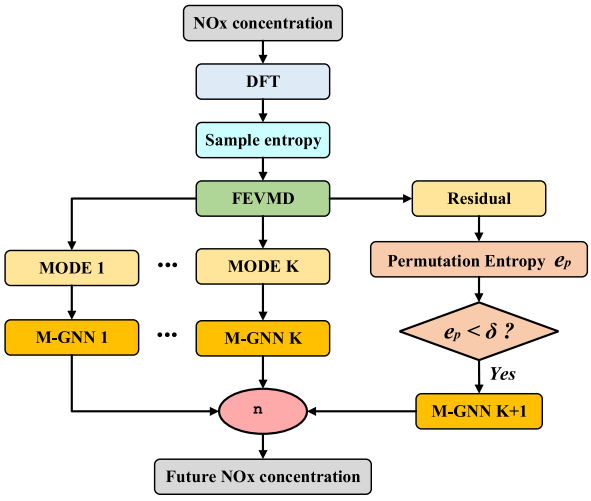


Fig. 2. The structure of FEVMD method .[11]

$$\widehat{u}_k^{n+1}(\omega)=\frac{\widehat{x}(\omega)-\sum_{i\neq k}\widehat{u}_i(\omega)+\frac{\widehat{\lambda}(\omega)}{2}}{1+2\alpha(\omega-\omega_k)^2} \tag{8}$$

$$w_k^{n+1}=\frac{\int_0^\infty \omega\left|\widehat{u}_k^{n+1}(\omega)\right|^2 d\omega}{\int_0^\infty \left|\widehat{u}_k^{n+1}(\omega)\right|^2 d\omega} \tag{9}$$

$$\widehat{\lambda}^{n+1}(\omega)=\widehat{\lambda}^{n}(\omega)+\tau\left[\widehat{x}(\omega)-\sum_k \widehat{u}_k^{n+1}\right] \tag{10}$$

where $\widehat{u}_k^{n+1}(\omega)$, $\widehat{x}(\omega)$, and $\widehat{\lambda}^{n+1}(\omega)$ represent the Fourier transforms of u_k^{n+1}, $x(t)$, and λ_k^{n+1}, respectively. τ is the step size of the Lagrangian operator. Given a discriminant accuracy $\varepsilon > 0$, the iteration is stopped when $\sum_k \|u_k^{n+1}-u_k^n\|_2^2/\|u_k^n\|_2^2 < \varepsilon$ is satisfied.

4.1.2. Ways of improvement

The parameter setting in VMD is critical, and incorrect parameters have a serious influence on decomposition performance [34]. FEVMD adaptively determines the number of decomposition layers and penalty factors using the discrete fourier transform (DFT) [35] and sample entropy [36]. The structure of FEVMD is depicted in Fig. 2.

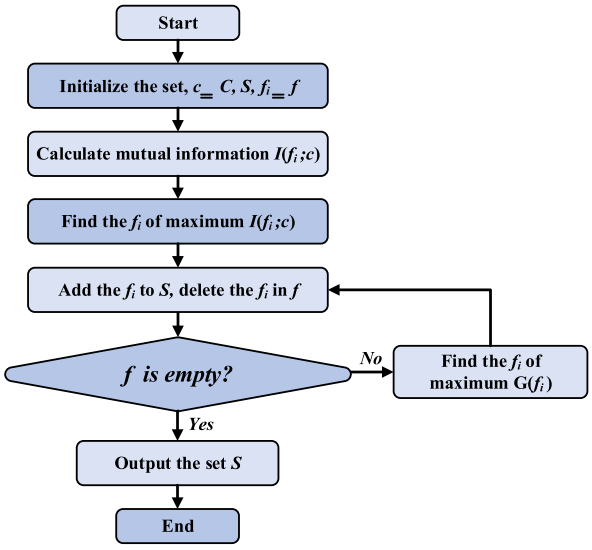


Fig. 3. The operational flow of the BMIFS method.

An incorrect number of decomposition layers setting easily results in over- or under-decomposition. The DFT can better reflect the number of frequencies in the original signal [35]. The goal of VMD is to separate the modal components in the signal that have different center frequencies. The decomposed modal components are closely distributed around the center frequency [19]. Therefore, the optimal number of decomposition layers can be reflected by the number of peaks in the DFT spectrum of the decomposed signal.

The penalty factor also has a significant impact on the decomposition results. A time series' complexity can be measured using sample entropy. The theory demonstrates that a sequence's self-similarity has an inverse relationship with its sample entropy value [36]. Thermal process data has a high degree of self-similarity. Therefore, the optimal value of the penalty factor can be determined by combining the entropy values of modal components. The operational steps of FEVMD are.

Step 1 . Define the signal $u(t)$ with sequence length N. $u(t)$ consists of $N-m+1$ vectors $u_m(i) = [u(i),u(i+1),\ldots,u(i+m-1)]$. m is the embedding dimension. $i = 1,\cdots,N-m+1$.

Step 2 . Calculate the mean matching probability between sequences.

$$B_m(r)=\frac{1}{N-m}\sum_{i=1}^{N-m}\frac{1}{N-m+1}v_m(i) \quad i=1,\cdots,N-m+1 \tag{11}$$

$$d_m[u_m(i),u_m(j)]=\max[u_m(i+k)-u_m(j+k)],0\leqslant k\leqslant m-1 \tag{12}$$

where $B_m(r)$ is the mean matching probability. d_m is the vector spacing. r is the tolerance for accepting matrices. v_m is the number of $d_m \leq r$.

Step 3 . Calculating the sample entropy $SampleEn(m,r)$. The step size is set to 100, and the traversal is performed between the intervals [1000, 8000]. The penalty factor with the lowest sum of entropy values among the modal components is selected as the optimal value.

$$SampleEn(m,r) = -\ln\frac{B_{m+1}(r)}{B_m(r)} \tag{13}$$

Step 4 . The signal to be decomposed is analyzed by an DFT spectrogram. The number of peaks in the spectrum is set to the optimal number of decomposition layers.

4.2. BMIFS

Mutual information feature selection (MIFS) is a feature selection method based on mutual information [37]. MIFS augments the evaluation function with penalty terms, successfully reducing information redundancy among variables. The calculation formula is:

$$J(f_i)=I(f_i;c)-\beta\sum_{S_j\in S}I(f_i;S_j) \tag{14}$$

where $J(\cdot)$ is the evaluation function. $I(\cdot)$ is the mutual information between variables. $f_i \in F$ is the candidate variable. c is the dominant variable. $S_j \in S$ is the selected variable. β is the penalty factor used to measure the ratio relationship between subsets.

However, as the search progresses, the number of variables chosen will gradually increase. The weight of the right-hand penalty term in Eq. (14) will continue to rise, while the role of the left-hand mutual information term will diminish [38]. At the later stage of screening, certain feature variables that are more related to the dominant variables will be missed.

To address the shortcomings of MIFS, we propose a boosting mutual information feature selection algorithm (BMIFS). We consider the influence of the number of selected variables $|S|$ in the search process and add a penalty weight term of $1/|S|$. Meanwhile, we reflect the correlation

Table 4
The results of feature selection for each candidate variable.

NOx			NOx		
Variable name	BIMFS coef.	Correlation	Variable name	BIMFS coef.	Correlation
UL	0.9409	High	$AA_{1\text{-}4}$	-0.1716	Low
ST	2.0305	Perfect	$A_{1\text{-}4}$	-0.2536	Low
SP	1.1159	High	$AB_{1\text{-}4}$	-0.0929	Low
P_f	-0.7299	Low	$B_{1\text{-}4}$	-0.0798	Low
T_f	2.0759	Perfect	$BC_{1\text{-}4}$	-0.3619	Low
O_{2f}	0.9178	High	C_1	0.7172	High
G_d	0.4423	Moderate	C_2	0.6901	High
P_d	0.0611	Moderate	C_3	0.7240	High
T_d	1.6998	High	C_4	0.7772	High
O_{2d}	1.4583	High	CD_1	-0.2725	Low
W_f	2.7955	Perfect	D_1	-0.2720	Low
A_f	0.5696	Moderate	DE_1	-0.6634	Low
E_a	1.4796	High	E_1	-0.8193	Irrelevant
F_a	-0.2821	Low	EF_1	-0.3027	Low
F_b	0.3031	Moderate	F_1	-0.6523	Low
F_c	0.2992	Moderate	FF_1	-0.7305	Low
F_d	0.0725	Moderate	OFA_1	-0.1716	Irrelevant
F_e	0.1363	Moderate	S_{1b}	-0.9208	Irrelevant
F_f	-0.5115	Low	S_{2d}	-0.5332	Low
M_a	-0.3909	Low	S_{3f}	0.1191	Moderate
M_b	-0.7894	Low	S_h	0.1363	Moderate
M_c	-0.8260	Irrelevant	D_f	0.3446	Moderate
M_d	-0.9906	Irrelevant	D_e	-1.2985	Irrelevant
M_e	-0.8294	Irrelevant	TC	1.4801	Perfect
M_f	-0.9450	Irrelevant	TA	2.9717	Perfect

between the variables to be selected and those selected in the evaluation function. The expression of the evaluation function $G(\cdot)$ is:

$$G(f_i) = I(f_i; c) - \frac{\beta}{|S|}\sum_{S_j \in S} MR \tag{15}$$

where $|S|$ indicates the number of selected feature variables. MR is the minimum redundancy in the candidate variable f_i with respect to the selected variable S_j. MR and $I(\cdot)$ are calculated as follows:

$$MR = \frac{I(f_i; S_j)}{I(f_i; c)} \tag{16}$$

$$I(X, Y) = -\int_y \int_x p(x,y) \log \frac{p(x,y)}{p(x)p(y)} dxdy \tag{17}$$

where $p(x)$ is the marginal probability distribution of X. $p(y)$ is the marginal probability distribution of Y. $p(x,y)$ is the joint probability distribution between X and Y.

Fig. 3 presents the operational flow of the BMIFS method. The specific steps of BMIFS are as follows.

Step 1 . Initialize the set of candidate variables f, selected variables S and dominant variables c.

Step 2 . Calculate the mutual information $I(f_i; c)$ between the candidate variables and the selected variables in turn, and determine the candidate variable f_i with the maximum mutual information.

Step 3 . Add f_i to the set of selected variables S. Meanwhile, remove f_i from the set of candidate variables f.

Step 4 . Determine whether the set of selected variables S reaches the target number.

Step 5 . If not. Calculate the value of the evaluation function $G(f_i)$ for each candidate variable and the selected variable in turn, and determine the maximum value of the function f_i. Return to **Step 3.**

Step 6 . If it is reached. Output the set of selected variables S and finish the feature selection.

This research finally identifies fifteen auxiliary variables. Table 4 presents the results of feature selection for each candidate variable.

4.3. Elimination of time delay

The boiler is a substantial and time-delayed system. There is a time delay between the feature variables and NOx generation concentrations during the combustion process [39]. The current values of the feature variables do not accurately reflect NOx concentrations.

The maximal information coefficient (MIC) is a statistical indicator that measures the strength of correlation between two variables [40]. Higher correlation between variables is indicated by higher MIC values. The calculation formula is:

$$M_{ic}(x,y) = \max_{xy<B(n)} \frac{I(x,y)}{\log_2(\min(x,y))} \tag{18}$$

where $I(x,y)$ is the mutual information between variables. $B(n)$ is usually set to the 0.6 power of the total data volume.

We calculate the MIC values between the auxiliary variables and

[1] M-GNN denotes the modal prediction result of MFST-GNN.

Table 5
MIC and delay time of each auxiliary variable.

No.	Notation	Delay time (s)
1	*UL*	120
2	*ST*	80
3	*SP*	10
4	*TC*	100
5	*TA*	50
7	T_f	200
8	O_{2f}	0
11	T_d	170
12	O_{2d}	10
13	W_f	10
15	E_a	10
47	C_1	20
48	C_2	20
49	C_3	20
50	C_4	20

NOx concentrations at each moment in the 0–5 min, with a time interval of 10 s. The moment with the largest MIC value is selected as the optimal delay time. Then, all auxiliary variables are calculated sequentially, and the feature dataset is reconstructed based on the delay time. Table 5 depicts the results of the time delay calculation for each auxiliary variable.

4.4. MFST-GNN

Spectral temporal graph neural network (StemGNN) [32] is a multivariate time series prediction method that does not require a predefined topology. It has obvious advantages for predicting multivariate NOx concentration series. Based on the StemGNN, we propose a multi-channel fused spectral temporal graph neural network (MFST-GNN).

MFST-GNN introduces a multi-channel structure to capture the graph feature information of NOx concentration subsequences and auxiliary variables. It adaptively generates the feature graph structure by using two latent correlation layer modules. The adjacency matrix is employed to establish topological information between operated data. The MFST-GNN block and output layer module are utilized to combine and iterate the neighboring node information to complete the concentration prediction. The overall framework of MFST-GNN is depicted in Fig. 4. Fig. 5 shows the model's specific components.

4.4.1. Latent correlation layer

The Latent Correlation Layer module employs a self-attention mechanism to learn the adjacency relationships between input variables in an adaptive manner and construct the feature graph structure [32]. It computes the implicit correlations in multivariate time series using the gated recurrent unit (GRU). The sequence representation is the final hidden state, and the adjacency matrix is computed using the attention mechanism. The formula for calculation is:

$$Q = RW^Q, K = RW^K, W = \mathrm{Softmax}\left(\frac{QK^T}{\sqrt{d}}\right) \tag{19}$$

where Q and K are the query matrix and key matrix. d is the input dimension. W^Q and W^K are learnable parameters. W is the output matrix. R is the hidden state obtained from the input after GRU calculation.

4.4.2. MFST-GNN block

MFST-GNN Block inherits the excellent performance of StemGNN Block in learning the time series implicit relations in the spectral domain. First, we employ two latent correlation layer modules to create a multi-channel. The graph topology of NOx concentration subsequences and auxiliary variables is adaptively generated and summed.

Next, we use the graph fourier transform (GFT) and discrete fourier transform (DFT) to calculate the inter-feature adjacency relationships and frequency domain representations. The spectral filtering is then

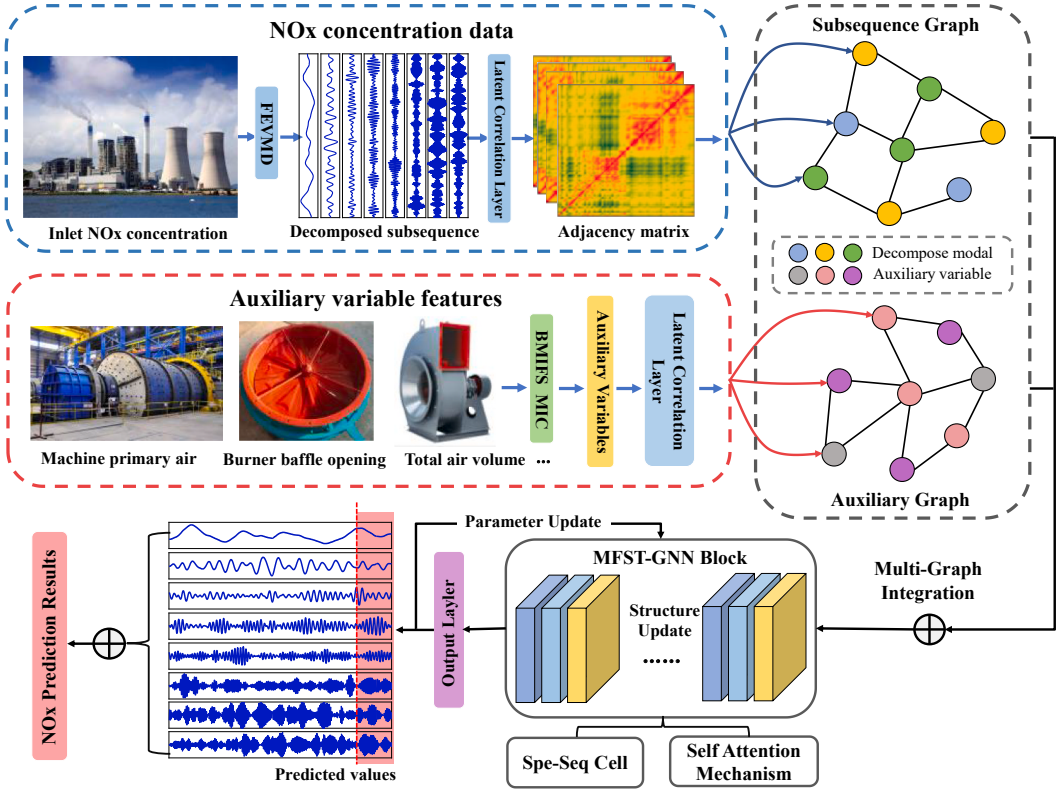


Fig. 4. The overall framework of MFST-GNN.

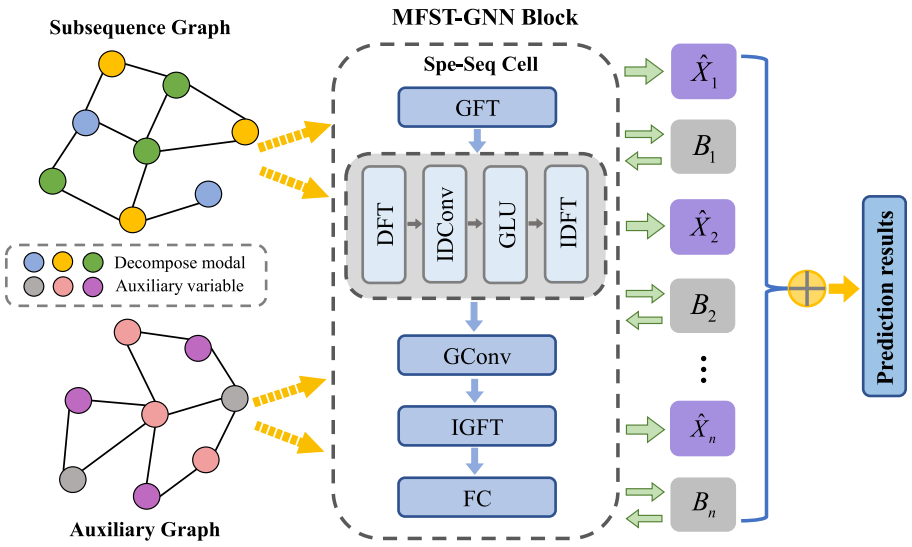


Fig. 5. Specific components of MFST-GNN .[21]

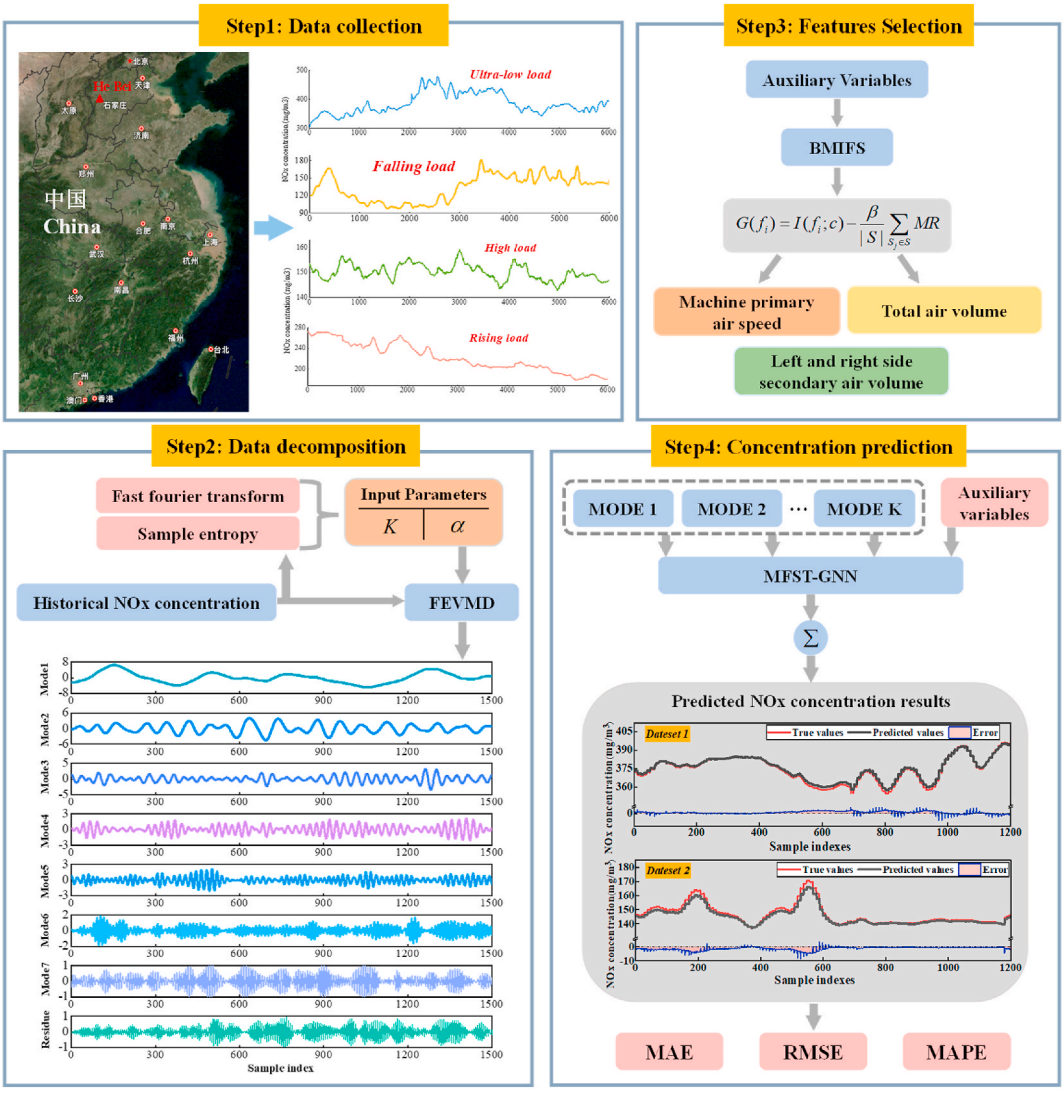


Fig. 6. The overall steps of the BM-GNN method.

Table 6
Description of the datasets.

Dataset	Data volume	Load scopes	NOx scopes
D1	6000	[178.22–242.69]	[305.60–477.03]
D2	6000	[239.27–380.42]	[97.71–181.86]
D3	6000	[593.85–603.13]	[142.77–159.05]
D4	6000	[368.21–551.85]	[178.89–272.75]

Table 7
The best number of decomposition layers and penalty factors.

Dataset	Decomposition number	Penalty factor
D1	8	3100
D2	9	4200
D3	8	2200
D4	9	3600

calculated utilizing the graph convolution operator, and the output is obtained using the inverse graph fourier transform (IGFT). The calculation formula is:

$$Z_j = \mathscr{GF}^{-1}\left(\sum_i g_{\Theta_{ij}}(\Lambda_i)\mathscr{S}(\mathscr{GF}(X_i))\right) \tag{20}$$

where $\mathscr{GF}$, $\mathscr{GF}^{-1}$, and $\mathscr{S}$ represent the computation process of GFT, IGFT, and Spe-Seq Cell, respectively. Λ_i is the laplace eigenvalue matrix. Θ_{ij} is the graph convolution kernel. Z_j is the network output.

The Spe-Seq cell learns the frequency-based features of multivariate time series. It mainly consists of DFT, 1D convolution, GLU, and the inverse discrete fourier transform (IDFT) [32]. DFT and IDFT are applied for interconversion between the time and frequency domains of data. The computational equation is:

$$M^*\left(\widehat{X}_u^*\right) = \mathrm{GLU}\left(\theta_\tau^*\left(\widehat{X}_u^*\right), \theta_\tau^*\left(\widehat{X}_u^*\right)\right) = \theta_\tau^*\left(\widehat{X}_u^*\right) \odot \sigma^*\left(\theta_\tau^*\left(\widehat{X}_u^*\right)\right), * \in \{r, i\} \tag{21}$$

where $\widehat{X}_u^r$ is the real output information. $\widehat{X}_u^i$ is the imaginary output information. θ_τ^* is the depth map convolution kernel. $\odot$ is the hadamard product. σ^* is the nonlinear sigmoid gate. $M^r\left(\widehat{X}_u^r\right) + iM^i\left(\widehat{X}_u^i\right)$ is the output to IDFT.

Finally, we obtain the basis expansion coefficients ϑ for the total output $Z = \left[Z_1, Z_2, \ldots, Z_j\right]$ through the fully connected layer. Combined with the learnable basis vectors ω, the outputs are calculated $y = \omega \cdot \vartheta$.

4.4.3. Output layer

We stack the MFST-GNN blocks to construct the deep graph prediction model. The output layer of the model keeps the original, consisting of GLU and the fully connected layer (FC). The output results include predicted values and historical callback values. The historical callback values help to regulate the temporal representation of the block function space [32]. The loss functions include forecasting loss and backcasting loss.

Set the timing data $X = [X_{t-K}, \ldots, X_{t-1}]$ for the first K timestamps and predict the timing data $\widehat{X} = [\widehat{X}_t, \widehat{X}_{t+1} \ldots, \widehat{X}_{t+H-1}]$ for the next H timestamps. $X \in R^{N \times T}$. N is the total number of timing nodes. T is the number of timestamps. The loss function is calculated as:

$$\mathscr{L}(\widehat{X}, X; \Delta_\theta) = \sum_{t=0}^{T} \left\|\widehat{X}_t - X_t\right\|_2^2 + \sum_{t=K}^{T} \sum_{i=1}^{K} \left\|B_{t-i}(X) - X_{t-i}\right\|_2^2 \tag{22}$$

where the former term is the forecasting loss. The latter term is the backcasting loss. $\widehat{X}$ is the predicted value. $B(\cdot)$ is the network structure for the backcasting loss. Δ_θ is the set of parameters in the network.

4.5. Overall prediction model

We propose a hybrid prediction model (BMFE-MFST-GNN) for NOx concentration in coal-fired power plants. For convenience, it is abbreviated as BM-GNN in the following sections. Fig. 6 presents the overall steps of BM-GNN. The specific processes are as follows.

- Data preprocessing. Collect the raw dataset. Detect outliers in the dataset. Fill in outlier and null data, and then normalize the data. Data with a difference between thermal process data greater than triple the mean value of the series is defined as outlier data points, and outlier and null data are filled with the mean values of the 15 points before and after.
- Determine the auxiliary variables and time delays. Candidate feature variables are initially chosen based on mechanistic knowledge. The correlation analysis and ranking are performed using BMIFS to determine the auxiliary variables. Meanwhile, MIC is applied to calculate the correlation among the features with different delays to derive the delay time of the set of auxiliary variables.
- Concentration modal decomposition. The FEVMD is utilized to divide the inlet NOx concentration into simple and smooth IMF and residual pairs. The subsequences extract the deep time-frequency

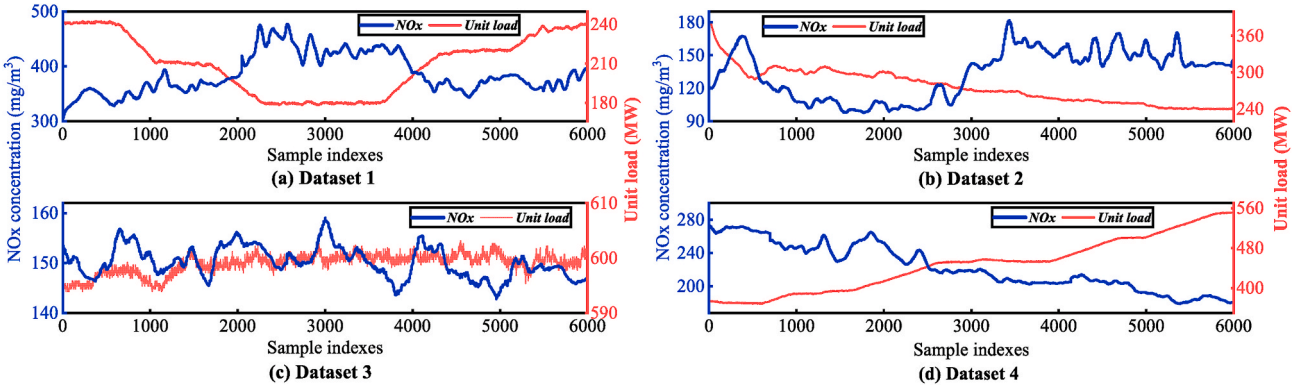


Fig. 7. The images of the experimental datasets.

[2] $\widehat{X}_i$ is the predicted concentration value. B_i indicates the network that generates backcasting output.

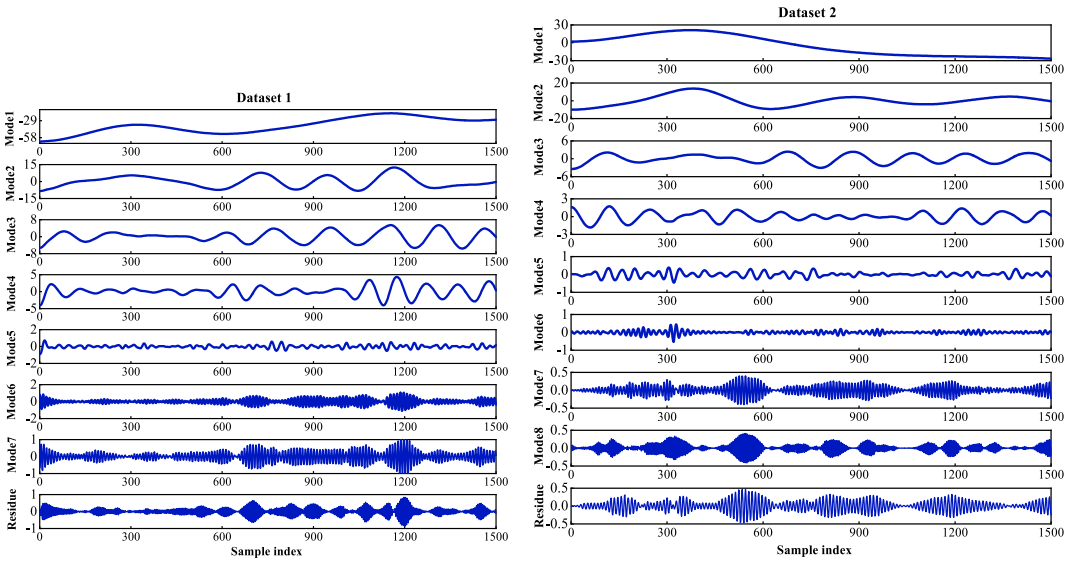


Fig. 8. Results of NOx concentration decomposition for D1 and D2 datasets .[31]

Table 8
The central frequency and permutation entropy of each mode [a].

Mode	D1		D2	
	u_k	e_p	u_k	e_p
Mode1	0.0002	0.6694	0.0002	0.7001
Mode2	0.0036	0.9370	0.0022	0.8347
Mode3	0.0063	1.1243	0.0046	0.9530
Mode4	0.0102	1.2121	0.0086	1.1348
Mode5	0.3538	2.1416	0.0233	1.8147
Mode6	0.0920	2.9913	0.0585	2.5937
Mode7	0.1071	3.1467	0.1052	3.0360
Mode8	–	–	0.2000	3.8188
Residue	0.1988	3.7592	0.2947	3.9513

[a] The results of D3 and D4 are given in the supplementary materials (Table 13).

information. The randomness of the residuals generated by the decomposition is filtered according to the permutation entropy.

- Model training. Input NOx concentration subsequences and reconstructed auxiliary variables into the model. Initialize the model parameters. Set the parameters for model training, such as learning rate and number of iterations.
- Model testing. Input test dataset. Output the prediction results of NOx concentration by the BM-GNN model.

5. Prediction experiment and analysis

5.1. Experimental datasets

In this research, we conduct prediction comparison experiments utilizing historical operating data from a 600 MW coal-fired generating unit in a power plant in Hebei, China. To validate the model's generalizability, we selected the measured data in different unit operation states to form the experimental datasets. Because the load coverage of the four datasets is extensive, including four operating states such as rising and falling load, high load, and ultra-low load under deep peaking, they can fully reflect the power plant's various operating characteristics.

For convenience, the four datasets are denoted as D1, D2, D3, and D4. The datasets contain a total of 24,000 sample points. The data sampling interval is 10s. The division ratio of the training set, validation set, and test set is 7:1:2. Table 6 provides the load coverage of four datasets. The load variation range contains the vast majority of the unit's operating states. Fig. 7 presents images of the four experimental datasets.

Table 10
Model parameter setting.

Parameter	Numerical setting
Window size	20
Multi layer	10
Train length	7
Valid length	1
Test length	2
Dropout rate	0.1
Decay rate	0.5
Leakyrelu rate	0.2
Learning rate	0.0001
Batch size	32
Norm method	Z_Score
Optimizer	RMSProp

Table 9
Reconstructed signal performance comparison [a].

Mode	D1					D2				
	R^2	MAE	RMSE	MAPE	TIME(s)	R^2	MAE	RMSE	MAPE	TIME(s)
EMD	0.8889	1.2275	1.3180	1.5909	**1.5534**	0.8750	0.6079	0.7552	1.0328	**1.5602**
VMD	0.9013	0.9582	1.1255	1.1740	2.0896	0.9166	0.5130	0.6697	0.9271	2.0701
CEEMD	0.9439	0.5372	0.9815	1.3126	2.0731	0.9283	0.4977	0.5916	0.7943	2.0755
HVMD	0.9850	0.3673	0.7391	0.9945	2.5721	0.9523	0.2712	0.3560	0.6723	2.5031
FEVMD	**0.9899**	**0.3140**	**0.5631**	**0.7539**	2.7606	**0.9794**	**0.1916**	**0.2886**	**0.4933**	2.9270

[a] The results of D3 and D4 are given in the supplementary materials (Table 14).

Table 11
Comparison of NOx concentration prediction results [a].

Model	D1			D2			D3			D4		
	MAE	RMSE	MAPE	MAE	RMSE	MAPE	MAE	RMSE	MAPE	MAE	RMSE	MAPE
GRU	1.61	1.96	0.44	2.45	3.23	1.65	1.80	2.08	1.21	2.30	2.41	1.22
BP	1.76	2.11	0.48	2.94	3.78	2.00	1.87	2.17	1.26	2.69	2.82	1.43
AE-ELM	1.43	1.78	0.39	1.69	2.28	1.13	1.27	1.48	0.85	1.75	1.85	0.94
CA-LSTM	1.38	1.71	0.37	1.99	2.63	1.34	1.07	1.26	0.72	1.83	1.94	0.98
StemGNN†	1.59	1.94	0.43	2.03	2.64	1.37	1.45	1.68	0.98	2.07	2.18	1.10
StemGNN	1.47	1.81	0.40	1.87	2.41	1.25	1.39	1.62	0.94	1.98	2.09	1.06
BM-GNN†	1.12	1.46	0.30	1.44	1.94	0.96	0.97	1.15	0.66	1.74	1.85	0.93
BM-GNN	**1.00**	**1.34**	**0.27**	**1.38**	**1.79**	**0.92**	**0.85**	**1.01**	**0.58**	**1.40**	**1.50**	**0.75**

[a] .† *denotes the prediction model without MIC time delay reconstruction.*

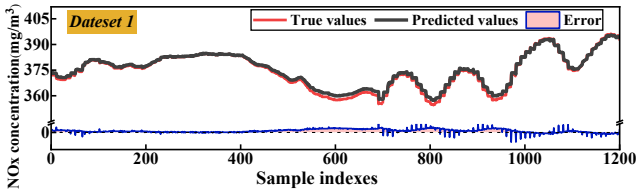


Fig. 9. Predicted concentration results for D1 dataset.

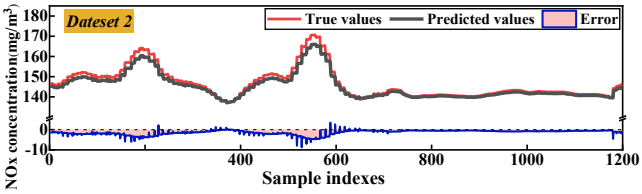


Fig. 10. Predicted concentration results for D2 dataset.

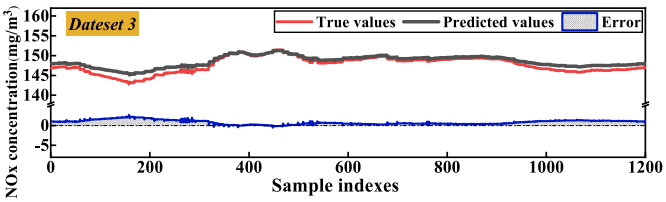


Fig. 11. Predicted concentration results for D3 dataset.

It should be noted that the NOx concentration data will produce significant distortions during the CEMS blowback [41]. We mark this part of the data when the blowdown signal is issued and do not use it.

5.2. Assessment indicators

We choose the assessment indicators of mean absolute error (MAE), root mean square error (RMSE), mean absolute percentage error (MAPE), and determination coefficient (R^2) to reflect the prediction performance of the model. The calculation formula is:

$$MAE = \frac{1}{N}\sum_{t=1}^{N}|x(t) - \hat{x}(t)| \tag{23}$$

$$RMSE = \sqrt{\frac{1}{N}\sum_{t=1}^{N}(x(t) - \hat{x}(t))^2} \tag{24}$$

$$MAPE = \frac{1}{N}\sum_{t=1}^{N}\left|\frac{x(t) - \hat{x}(t)}{x(t)}\right| \tag{25}$$

$$R^2 = 1 - \frac{\sum_{t=1}^{n}(x(t) - \hat{x}(t))^2}{\sum_{t=1}^{n}(x(t) - \bar{x}(t))^2} \tag{26}$$

where $x(t)$ is the real NOx concentration value. $\hat{x}(t)$ is the forecasted NOx concentration value, and N is the total number of data.

5.3. NOx concentration decomposition

The inlet NOx concentration is adaptively decomposed using FEVMD in this research. The time-series data are decomposed into simple and smooth subsequences and residual pairs. Table 7 provides the best number of decomposition layers and penalty factors for four datasets.

Fig. 8 depicts the NOx concentration decomposition results for D1 and D2. The central frequency (u_k) and permutation entropy (e_p) of each mode are shown in Table 8. According to the results, the permutation entropy exhibits a positive correlation with the central frequency. The residual term has a relatively large value of permutation entropy. Therefore, we use the permutation entropy [42] as a random metric reference and set a threshold value of 3.80 to filter the invalid residuals.

We conduct signal reconstruction experiments to further validate the effectiveness of FEVMD. The comparison methods include VMD [19], EMD [20], CEEMD [12], and HVMD [43]. Table 9 compares the signal reconstruction results of D1 and D2.

According to the results, the reconstructed signal after FEVMD has the highest correlation with the original signal. In terms of R^2, MAE, RMSE, and MAPE, FEVMD's reconstruction accuracy is superior to the other compared methods. The effectiveness of the proposed method improvement is demonstrated.

5.4. Comparative experiments

We compare the prediction performance of the proposed BM-GNN with a variety of models to experimentally validate its effectiveness. The single models GRU [14], BP [44], and StemGNN [32], as well as the recently proposed hybrid models CA-LSTM[4] [12] and AE-ELM [13], are included in the comparison models. The experimental datasets are D1, D2, D3, and D4.

The auxiliary variables are entered based on the BMIFS correlation analysis results. The datasets are rebuilt utilizing the time delays shown in Table 5. Table 10 displays the parameter settings for the BM-GNN model. The remaining comparison model parameters are kept

[3] To shorten the text, the results of the decomposition of D1 and D2 are shown in the supplemental materials (Fig. 15).

[4] We abbreviate CEEMDAN-AM-LSTM as CA-LSTM in Ref. [12].

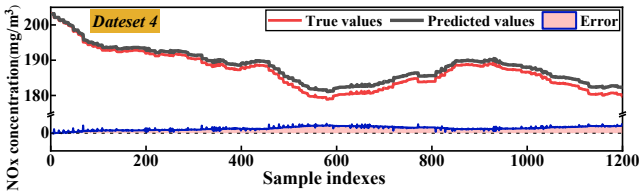


Fig. 12. Predicted concentration results for D4 dataset.

unchanged from the original articles. The experimental training and testing procedures are implemented in Python version 3.7.

Table 11 summarizes the prediction results for each model. Figs. 9–12 present the prediction curves for the four datasets. We can obtain that: 1) Due to the complex engineering characteristics of the thermal process data, such as nonlinearity and non-smoothness, it is difficult to achieve better prediction results with a single prediction model. Single models such as GRU, BP, and SVM have lower prediction accuracy than hybrid models such as CA-LSTM and AE-ELM. 2) After calculating the time delay with MIC and reconstructing the dataset, the influence of the boiler combustion process delay on the prediction accuracy of NOx concentration is reduced. The BM-GNN prediction accuracy is enhanced. 3) On all four datasets, BM-GNN outperforms the Sother comparison methods in prediction. Meanwhile, the proposed model's prediction accuracy and generalization performance are significantly improved over StemGNN, and the performance is nearly identical across the four datasets. This is because the addition of signal decomposition and auxiliary variables allows the model to capture and fuse inter-feature correlations and time dependencies more effectively, improving overall prediction results.

5.5. *The effectiveness of the StemGNN method*

To validate the prediction effect of StemGNN in the hybrid model

Table 12
Prediction results for each mode of the D1 dataset [a].

Mode	ARIAM			LSTM			StemGNN		
	MAE	RMSE	MAPE	MAE	RMSE	MAPE	MAE	RMSE	MAPE
Mode1	**0.49**	**0.70**	**1.56**	0.50	0.72	1.57	0.55	0.82	1.67
Mode2	0.44	0.61	2.57	**0.37**	**0.55**	**2.01**	0.40	0.61	2.26
Mode3	0.30	0.57	3.62	**0.29**	**0.42**	**2.90**	0.37	0.50	3.35
Mode4	0.25	0.40	5.62	**0.20**	**0.36**	**5.60**	0.22	0.39	5.62
Mode5	0.20	0.37	8.01	0.19	0.34	6.58	**0.17**	**0.32**	**6.43**
Mode6	0.19	0.29	9.01	0.14	0.27	8.19	**0.13**	**0.21**	**7.01**
Mode7	0.15	0.23	10.07	0.13	0.22	10.19	**0.10**	**0.15**	**9.07**
Residue	0.12	0.18	11.78	0.10	0.17	12.10	**0.09**	**0.12**	**11.73**

[a] The results of D2, D3, and D4 are presented in the supplementary materials (Tables 15–17).

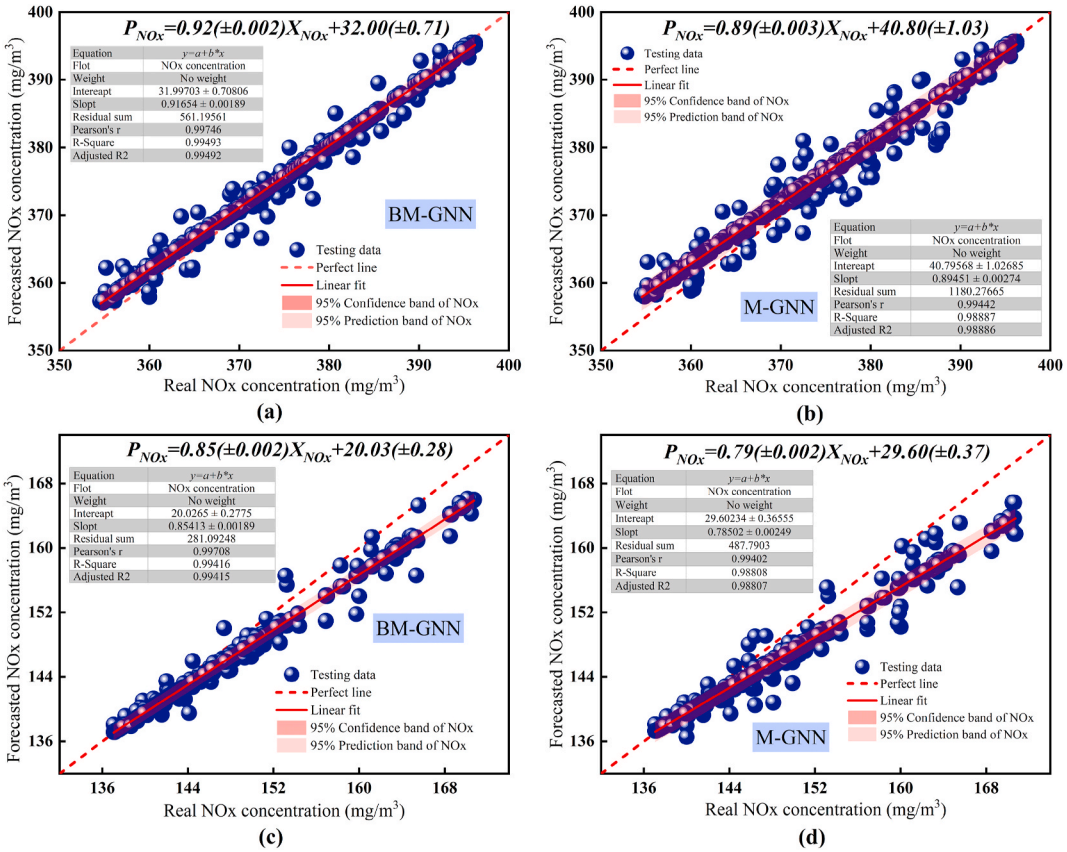


Fig. 13. Fitting results of forecasted and real values of different models: (a) Results of BM-GNN on D1; (b) Results of M-GNN on D1; (c) Results of BM-GNN on D2; (d) Results of M-GNN on D2.

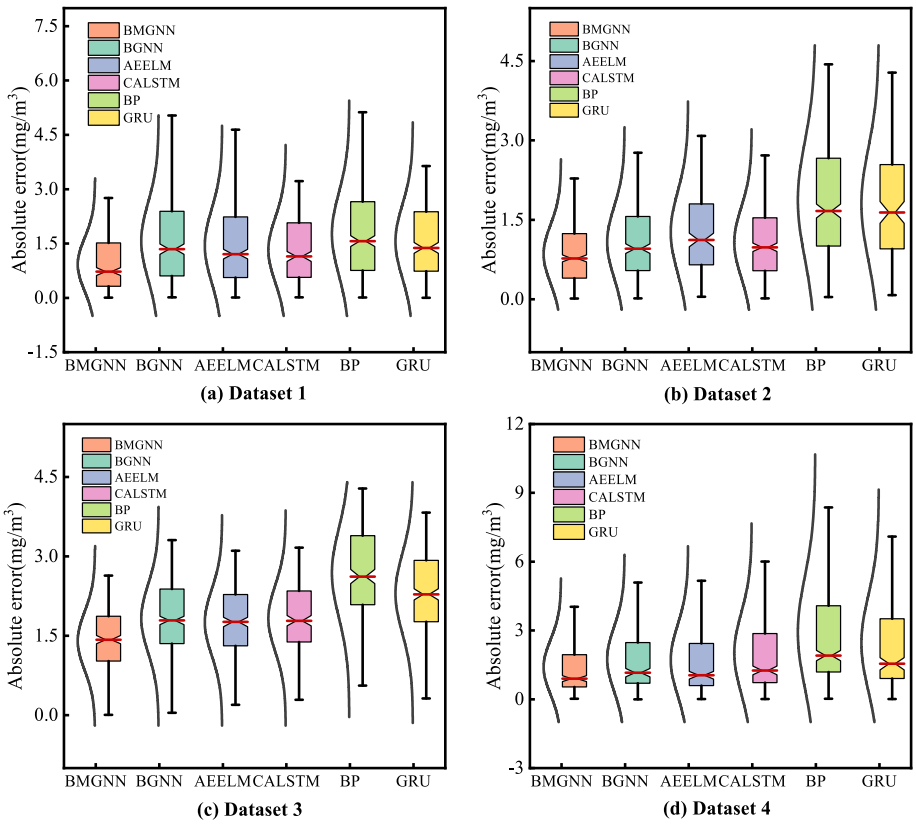


Fig. 14. Comparison of prediction error results of the models.

based on signal decomposition, we perform a prediction effect comparison experiment in different modes. The comparison method employs the widely used timing prediction methods ARIAM and LSTM. Table 12 presents the comparison results for D1.

According to the results, StemGNN performs well in both low-frequency and high-frequency prediction modes. For each mode, the prediction accuracy of LSTM is higher than that of ARIAM. StemGNN's prediction performance is comparable to LSTM in low-frequency modes and superior to LSTM in high-frequency modes. Therefore, StemGNN can show its own excellent prediction performance in multivariate hybrid prediction models based on signal decomposition.

5.6. *The effectiveness of the BMIFS method*

Fig. 13 depicts a comparison of the model's prediction results with and without the addition of BMIFS extracted auxiliary variables. D1 and D2 are taken as examples. According to Fig. 13(b)(d), the fitted line of the model without auxiliary variables (M-GNN) deviates from the ideal straight line. The area of both the 95 % confidence interval and the prediction interval is larger. The slopes of the two fitted curves are 0.89 $\pm$ 0.003 and 0.79 $\pm$ 0.002. There is a large error between the predicted and true values.

From Fig. 13(a)(c), the prediction results of the BMIFS extracted auxiliary variable model (BM-GNN) are essentially distributed around the ideal straight line. The slopes of the two fitted curves are 0.92 $\pm$ 0.002 and 0.85 $\pm$ 0.002. The area of the prediction interval and the 95 % confidence interval are both small. Therefore, the prediction performance of the model improves significantly after incorporating the auxiliary variables extracted by BMIFS. The proposed BMIFS method's effectiveness is demonstrated.

5.7. *The effectiveness of the FEVMD method*

The prediction errors for the model with modal decomposition using FEVMD (BM-GNN), the model without modal decomposition (B-GNN), and the other models are compared in Fig. 14. The normal distribution curve of the absolute error is depicted on the left side of the box shape. The red line in the box shape represents the median error. The lower the median error and the more concentrated the absolute error distribution range, the better the model's prediction accuracy.

According to Fig. 14, FEVMD reduces the median and absolute error of prediction results after modal decomposition of NOx concentration. Its prediction performance in the four datasets tends to be consistent when compared to the comparison method. The prediction accuracy and generalization performance of the model have greatly improved. The effectiveness of the proposed FEVMD method is demonstrated.

6. Conclusions

Accurate NOx concentration prediction is critical for controlling pollutant emissions from coal-fired power plants and protecting the environment. Due to the nonlinearity, coupling, and hysteresis of the SCR system, NOx concentration prediction models frequently perform poorly in terms of accuracy and generalization. We propose a data-driven hybrid prediction model (BMFE-MFST-GNN) to address this problem. First, FEVMD is proposed to decompose NOx concentration into multiple subsequences. Second, BMIFS is developed to determine the auxiliary variables to improve the model prediction accuracy. Finally, MFST-GNN constructs and iterates the graph structure information between features to achieve NOx prediction. We conduct comparative experiments utilizing actual historical operating data from a 600 MW coal-fired generating unit at a power plant in Hebei, China. The results demonstrate that the proposed method outperforms several comparable methods in predicting NOx concentrations under various

operating conditions, including rising and falling loads, high loads, and ultra-low loads.

The hybrid prediction model provides a new solution to the complex NOx concentration prediction problem. However, there are also some weaknesses in the proposed model. At this stage, the problems of coal property variation, operation condition division, model failure, and blowback data distortion still impede the practical application of many prediction models in the NOx prediction task. In our future research, we will further consider and attempt to solve these problems. Meanwhile, we will continue to deepen our collaboration with power plants to discuss the possibility of applying the models to actual operating systems.

CRediT authorship contribution statement

Zheng Wu: Conceptualization, Methodology, Writing – original draft, Writing – review & editing, Supervision. **Yue Zhang:** Investigation, Validation. **Ze Dong:** Resources, Conceptualization, Data curation.

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.

Data availability

The data that has been used is confidential.

Acknowledgements

This work was supported by the S&T Program of Hebei (No. 22567643H) and the National Natural Science Foundation of China (No. 71471060) and the Fundamental Research Funds for Central Universities (No. 2018QN096) and the Natural Science Foundation of Hebei Province (No. E2018502111).

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.energy.2024.132222.

References

[1] Jin Y, Scherer L, Sutanudjaja E, et al. Climate change and CCS increase the water vulnerability of China's thermoelectric power fleet. Energy 2022;245:123339.
[2] Jiang Y, Lee B, Oh D, et al. Influence of various air-staging on combustion and NO_X emission characteristics in a tangentially fired boiler under the 50% load condition. Energy 2022;244:123167.
[3] Yuan Z, Meng L, Gu X, et al. Prediction of NOx emissions for coal-fired power plants with stacked-generalization ensemble method. Fuel 2021;289:119748.
[4] Du X, Yang G, Chen Y, et al. The different poisoning behaviors of various alkali metal containing compounds on SCR catalyst. Appl Surf Sci 2017;392:162–8.
[5] Gu X, Li B, Sun C, et al. An improved hourly-resolved NOx emission inventory for power plants based on continuous emission monitoring system (CEMS) database: a case in Jiangsu, China. J Clean Prod 2022;369:133176.
[6] Yang T, Ma K, Lv Y, et al. Real-time dynamic prediction model of NOx emission of coal-fired boilers under variable load conditions. Fuel 2020;274:117811.
[7] Schluckner C, Gaber C, Landfahrer M, et al. Fast and accurate CFD-model for NOx emission prediction during oxy-fuel combustion of natural gas using detailed chemical kinetics. Fuel 2020;264:116841.
[8] Usberti N, Jablonska M, Blasi MD, et al. Design of a "high-efficiency" NH3-SCR reactor for stationary applications. A kinetic study of NH3 oxidation and NH3-SCR over V-based catalysts. Appl Catal B Environ 2015;179:185–95.
[9] Devarakonda M, Tonkyn R, Tran D, et al. Modeling Species inhibition of NO oxidation in urea-SCR catalysts for diesel engine NOx control. J Eng Gas Turbines Power 2011;133:491–8.
[10] Yuan Z, Meng L, Gu X, et al. Prediction of NOx emissions for coal-fired power plants with stacked-generalization ensemble method. Fuel 2021;289:119748. wang.
[11] Tuttle JF, Blackburn LD, Andersson K, Powell KM. A systematic comparison of machine learning methods for modeling of dynamic processes applied to combustion emission rate modeling. Appl Energy 2021;292:116886.
[12] Wang Z, Pen X, Cao S, et al. NOx emission prediction using a lightweight convolutional neural network for cleaner production in a down-fired boiler. J Clean Prod 2023;389:136060.
[13] Wang X, Liu W, Wang Y, Yang G. A hybrid NOx emission prediction model based on CEEMDAN and AM-LSTM. Fuel 2022;310:122486.
[14] Tang Z, Wang S, Chai X, et al. Auto-encoder-extreme learning machine model for boiler NOx emission concentration prediction. Energy 2022;256:124552.
[15] Tang Z, Wang S, Cao s, et al. Dynamic prediction model for NOx emission of SCR system based on hybrid data-driven algorithms. Proceedings of the CSEE 2022;42:3295–307.
[16] Li F, Wang Q, Hu J, Tang Y. Combined data-driven and knowledge-driven methodology research advances and its applied prospect in power systems. Proceedings of the CSEE 2021;41:4377–90.
[17] Qiao J. A novel online modeling for NOx generation prediction in coal-fired boiler. Sci Total Environ 2022;847:157542.
[18] Zhang L, Lin D, Wang Y, et al. Review of applications of machine learning in nitrogen oxides reduction in thermal power plants. Therm Power Gener 2023;52:7–17.
[19] Dragomiretskiy K, Zosso D. Variational mode decomposition. IEEE Trans Signal Process 2014;62:531–44.
[20] Huang N, Shen Z, Long S, et al. The empirical mode decomposition and the Hilbert spectrum for nonlinear and non-stationary time series analysis. Proceedings of the Royal Society of London Series A 1998;454.
[21] Meng A, Zhu Z, Deng W, et al. A novel wind power prediction approach using multivariate variational mode decomposition and multi-objective crisscross optimization based deep extreme learning machine. Energy 2022;260:124957.
[22] Korkmaz D. SolarNet: a hybrid reliable model based on convolutional neural network and variational mode decomposition for hourly photovoltaic power forecasting. Appl Energy 2021;300:117410.
[23] Liu H, Wang Y, Li X, Yang G. Prediction of NOx emissions of coal-fired power plants based on mutual information-graph convolutional neural network. Proceedings of the CSEE 2021;42:1052–60.
[24] Ma C, Sun L, Zhong Z, et al. ReLaText: exploiting visual relationships for arbitrary-shaped scene text detection with graph convolutional networks. Pattern Recogn 2021;111:107684.
[25] Kipf T, Welling M. Semi-supervised classification with graph convolutional networks[C]. Proceedings of the international conference on learning representations. Toulon: iclr. 2017.
[26] Hu J, Song Y. Piezoelectric modulus prediction using machine learning and graph neural networks. Chem Phys Lett 2022;791:139359.
[27] Liu J, Wang X, Xie F, et al. Condition monitoring of wind turbines with the implementation of spatio-temporal graph neural network. Eng Appl Artif Intell 2023;121:106000.
[28] Wu Q, Zheng H, Guo X, et al. Promoting wind energy for sustainable development by precise wind speed prediction based on graph neural networks. Renew Energy 2022;199:977–92.
[29] Sun L, Liu T, Wang D, et al. Deep learning method based on graph neural network for performance prediction of supercritical CO2 power systems. Appl Energy 2022;324:119739.
[30] Liu H, Wang Y, Li X, et al. Prediction of NOx emissions of coal-fired power plants based on mutual information-graph convolutional neural network. Renew Energy 2022;42:1052–60.
[31] Wu Q, Zheng H, Guo X, et al. Promoting wind energy for sustainable development by precise wind speed prediction based on graph neural networks. Renew Energy 2022;199:977–92.
[32] Cao D, Wang Y, Duan J, et al. Spectral temporal graph neural network for multivariate time-series forecasting[C]//Conference on Neural Information Processing Systems. Canada: NeurIPS; 2021.
[33] Dong Z, Yan L. Modelling and simulation for NOx emission concentration of SCR denitrification system. J Syst Simul 2020;32:172–81.
[34] Yin H, Liu Y, et al. Variational mode decomposition denoising combined the detrended fluctuation analysis. Signal Process 2016;125:349–64.
[35] Behrendt M, Angelis M, Comerford L, et al. Projecting interval uncertainty through the discrete Fourier transform: an application to time signals with poor precision. Mech Syst Signal Process 2022;172:108920.
[36] Widodo A, Shim M, Caesarendra W, Yang B. Intelligent prognostics for battery health monitoring based on sample entropy. Expert Syst Appl 2011;38:11763–9.
[37] Battiti R. Using mutual information for selecting features in supervised neural net learning. IEEE Trans Neural Network 1994;5:537–50.
[38] Macedo F, Oliveira M, Pacheco A, et al. Theoretical foundations of forward feature selection methods based on mutual information. Neurocomputing 2019;325:67–89.

[39] Yan L, Dong Z. Dynamic prediction model of SCR denitrification system based on k-nearest neighbor mutual information and WKOPLS. Proceedings of the CSEE 2019; 39:2970–80.
[40] Reshef D, Reshef Y, Finucane H, et al. Detecting novel associations in large data sets. Science 2011;334:1518–24.
[41] Yan L. Research on data-driven modeling and optimal control of SCR flue gas denitration system[D]. Beijing: North China Electric Power University; 2020.
[42] Cao H, Tung W, Gao J, et al. Detecting dynamical changes in time series using the permutation entropy. Phys Rev 2004;70:046217.
[43] Gai J, Shen J, Hu Y, Wang H. An integrated method based on hybrid grey wolf optimizer improved variational mode decomposition and deep neural network for fault diagnosis of rolling bearing. Measurement 2020;162:107901.
[44] Yu T, Geng P, Huo E, Cao M. Combustion optimization of a coal-fired boiler based on intelligent algorithm. Journal of Chinese Society of Power Engineering 2016;36: 594–9.