

Article

Dynamic Soft Sensing of Stack NO_x Concentration in Cement Kiln SNCR–SCR Denitrification Using a DAC-IVY-Optimized TCN-SE-LSTM Model

Zheng Zhao ^{1,2,*}, Si-Yuan Liu ¹, Yu-Xin Zhang ¹, Jia-Le Quan ¹ and Xin-Yu Tang ³ ¹ Department of Automation, North China Electric Power University, Baoding 071003, China² Hebei Technology Innovation Center of Simulation & Optimized Control for Power Generation, North China Electric Power University, Baoding 071003, China³ Tianjin Sinoma Engineering Research Center Co., Ltd., Tianjin 300131, China

* Correspondence: zheng_zhao@ncepu.edu.cn

Abstract

Accurate single-step prediction of stack NO_x concentration is essential for emission monitoring and ammonia-injection control in cement kiln SNCR–SCR hybrid denitrification systems. However, this task is challenging because industrial kiln data are affected by nonstationary emission fluctuations, nonlinear multivariable coupling, process-dependent time delays, and online deployment constraints. To address these process-specific challenges, this study develops a leakage-free dynamic soft-sensing framework for stack NO_x concentration prediction. In the proposed framework, variational mode decomposition (VMD) is used to characterize the multi-scale nonstationarity of the stack NO_x sequence under a sliding-window protocol. Trend-guided maximal information coefficient (MIC) analysis is then applied for nonlinear feature selection and delay compensation using only the training data, and the identified feature subset and delay parameters are fixed for validation and testing. A TCN-SE-LSTM model is constructed to extract temporal dependencies, recalibrate informative feature channels, and capture long-lag dynamic behavior. In addition, the Dual Adaptive Constrained Ivy Algorithm (DAC-IVY) is used only for offline hyperparameter optimization, so that the online stage requires only the trained prediction model. Experiments using 21,600 raw samples collected from an actual cement kiln Distributed Control System (DCS) show that the proposed framework achieves an RMSE of 0.2084 mg/Nm³ and an R^2 of 0.9844 on the test set, outperforming conventional baseline models. These results indicate that the proposed framework can provide an effective soft-sensing basis for subsequent denitrification control and operational optimization.

Keywords: cement kiln; SNCR–SCR hybrid denitrification; stack NO_x concentration; soft sensing; variational mode decomposition; delay compensation; TCN-SE-LSTM; hyperparameter optimization



Academic Editor: Alpaslan Atmanli

Received: 2 June 2026

Revised: 29 June 2026

Accepted: 1 July 2026

Published: 3 July 2026

Copyright: © 2026 by the authors.

Licensee MDPI, Basel, Switzerland.

This article is an open access article distributed under the terms and conditions of the [Creative Commons Attribution \(CC BY\)](https://creativecommons.org/licenses/by/4.0/) license.

1. Introduction

The cement industry is one of the major industrial sources of atmospheric pollutants, following sectors such as thermal power generation and transportation [1,2]. Among these pollutants, nitrogen oxides (NO_x) are of particular concern because they contribute to acid deposition, photochemical smog, secondary particulate formation, and human health risks [3,4]. In cement production, NO_x is mainly generated during high-temperature fuel

combustion in the rotary kiln and precalciner [5]. Therefore, effective NO_x reduction is essential for meeting increasingly stringent emission regulations.

NO_x control in cement kilns is generally achieved through combustion control and flue gas denitrification. Combustion control methods, such as low- NO_x combustion and staged combustion, suppress NO_x formation by adjusting the combustion atmosphere and temperature distribution in the kiln and precalciner [6,7]. However, when combustion control alone is insufficient, flue gas denitrification technologies are required. Selective Non-Catalytic Reduction (SNCR) injects ammonia or urea into a high-temperature reaction zone without a catalyst, offering relatively low operating cost but moderate denitrification efficiency [8]. Selective Catalytic Reduction (SCR), in contrast, uses a catalyst in a medium-temperature flue gas region and can achieve higher denitrification efficiency, although with higher investment and maintenance costs [9].

To balance denitrification efficiency and operating cost, SNCR–SCR hybrid denitrification systems have been increasingly adopted in cement kilns [10,11]. In such systems, SNCR provides primary NO_x reduction under high-temperature conditions, while SCR further removes residual NO_x and helps reduce ammonia slip [12]. However, the stack NO_x concentration in this hybrid process is affected by multiple coupled variables, including ammonia flow, combustion state, raw meal feed, flue gas temperature, and upstream NO_x concentration. These variables exhibit strong nonlinearity, nonstationarity, and process-dependent time delays due to material transport, thermal inertia, and multistage gas-solid reactions [13]. As a result, direct and timely measurement alone is often insufficient for high-quality ammonia-injection control. An accurate dynamic soft-sensing model for stack NO_x concentration is therefore important for real-time monitoring, control support, and operational optimization [14].

Existing NO_x prediction methods can be broadly divided into mechanistic and data-driven approaches [15]. Mechanistic models are built from reaction kinetics, mass transfer, and thermodynamic principles, and they can provide physical interpretability [16]. Mechanistic and process-analysis studies have provided useful insights into NO_x formation and removal mechanisms in cement kiln denitrification and SNCR–SCR-related processes [11,13,17]. However, the internal reactions and operating conditions of cement kiln denitrification systems are highly complex, making it difficult to construct a sufficiently accurate first-principles model for stack NO_x prediction. Data-driven methods, including statistical learning and deep neural networks, have therefore been widely used for industrial soft sensing and time-series prediction [18]. Recent cement-related studies have further demonstrated the feasibility of data-driven NO_x prediction. For example, Liu et al. [19] proposed an EEMD–MImRMR–BASBP-based NO_x concentration prediction model for cement denitrification processes, and Xue et al. [20] developed a hybrid data-driven model for NO_x concentration prediction in a cement clinker calcination system.

Recent studies have demonstrated the potential of data-driven models in complex industrial prediction tasks. For example, hybrid frameworks integrating signal decomposition and Transformer–LSTM structures have been applied to explainable photovoltaic power prediction [21]. Bidirectional LSTM models combined with multimodal information fusion have been used for carbon emission prediction [22]. Compositional generalization models, such as PSRONet and SMNet, have also shown strong performance in industrial robot multi-joint fault diagnosis [23,24]. In addition, state prediction and fault diagnosis frameworks for scenarios with scarce labeled data, such as the Time-Frequency Dual Domain Prediction (TFDDP) model [25] and the Time-Frequency Dual-Domain Contrast and Fusion (TFDDCF) approach [26], have improved feature extraction and generalization by exploiting unlabeled signals. These studies indicate that hybrid data-driven architectures are effective for complex industrial systems. Nevertheless, cement kiln SNCR–SCR denitri-

fication has distinct characteristics: the target signal is strongly nonstationary, the input variables are coupled through combustion and denitrification mechanisms, and different variables influence the stack NO_x concentration with different time delays. These characteristics require a prediction framework that can simultaneously address signal decomposition, feature relevance, delay compensation, and long-term temporal dependence.

Several studies have investigated NO_x prediction in cement kilns and related combustion and denitrification systems. For boiler and power-plant applications, Sun et al. [27] developed an artificial-neural-network-based boiler NO_x emission prediction model, Yang et al. [28] constructed a real-time dynamic NO_x prediction model for coal-fired boilers under variable-load conditions, and Kang et al. [29] introduced a Bi-LSTM-based dynamic model for SCR denitration systems in coal-fired power plants. For other industrial denitrification scenarios, Xie et al. [30] modeled SCR outlet NO_x emission sequences using a sequence-to-sequence LSTM network, He et al. [31] applied a CNN-LSTM hybrid scheme to NO_x emission prediction in an FCC unit, and Zhong et al. [32] developed a data-driven model to reduce outlet NO_x concentration of an SCR system in an FCC unit. Although these studies provide useful references, most existing models do not fully integrate nonstationary signal decomposition, mechanism-guided feature selection, delay compensation, channel-wise feature recalibration, and offline hyperparameter optimization within a unified leakage-free soft-sensing framework for cement kiln SNCR-SCR hybrid denitrification.

It should be emphasized that the objective of this study is not to claim that each individual module is newly proposed at the level of algorithmic principles. Instead, the main novelty lies in constructing a process-specific, leakage-free, and online-feasible dynamic soft-sensing framework for stack NO_x prediction in a cement kiln SNCR-SCR hybrid denitrification system. The investigated industrial process simultaneously exhibits nonstationary stack NO_x fluctuations, nonlinear multivariable coupling, variable-specific time delays, and online deployment constraints. Therefore, a reliable soft-sensing framework should not only improve network prediction accuracy, but also address data leakage prevention, nonlinear feature relevance, delay alignment, channel-wise information recalibration, and online inference feasibility in a unified workflow.

To address these limitations, this study proposes a dynamic soft-sensing framework for single-step stack NO_x concentration prediction in a cement kiln SNCR-SCR hybrid denitrification system. The design of the proposed framework is guided by the above process-specific modeling requirements rather than by a simple mechanical stacking of existing modules. Variational Mode Decomposition (VMD) is used to handle the nonstationarity of the stack NO_x sequence. Trend-guided Maximal Information Coefficient (MIC) analysis is used for nonlinear feature selection and delay compensation. A Temporal Convolutional Network integrated with a Squeeze-and-Excitation block and a Long Short-Term Memory network (TCN-SE-LSTM) is constructed to extract multivariate temporal dependencies, recalibrate informative channels, and capture long-lag dynamic behavior. Finally, the Dual Adaptive Constrained Ivy Algorithm (DAC-IVY) is introduced for offline hyperparameter optimization. During online application, the optimized model parameters are fixed, and DAC-IVY is not executed in the online prediction stage, which helps maintain the real-time feasibility of the proposed soft sensor.

The main contributions of this study are summarized as follows:

- A process-specific leakage-free dynamic soft-sensing framework is developed for stack NO_x concentration prediction in a cement kiln SNCR-SCR hybrid denitrification system. The preprocessing pipeline explicitly separates training, validation, and testing stages, and the feature subset and delay parameters identified from the training data are fixed for validation and testing to reduce the risk of future-information leakage.

- A trend-guided MIC-based feature selection and delay compensation strategy is used to identify nonlinear relevance and process-dependent time lags among key operating variables, improving the temporal alignment of multivariable inputs.
- A hybrid TCN-SE-LSTM architecture is designed for industrial multivariable time-series soft sensing. In this architecture, TCN extracts local temporal features, the SE block recalibrates channel-wise feature importance, and LSTM captures long-lag dynamic behavior.
- A DAC-IVY-based offline hyperparameter optimization strategy is introduced to improve model configuration while keeping the online prediction stage lightweight. Experiments using 21,600 raw samples from an actual cement kiln Distributed Control System demonstrate the effectiveness and engineering feasibility of the proposed framework.

2. SNCR–SCR Hybrid Denitrification System for Cement Kilns

The generation and removal of NO_x in a cement kiln involve several coupled stages, including rotary kiln calcination, precalciner combustion, staged-combustion NO_x reduction, SNCR denitrification, and SCR denitrification. The process flow of the investigated SNCR–SCR hybrid denitrification system is shown in Figure 1.

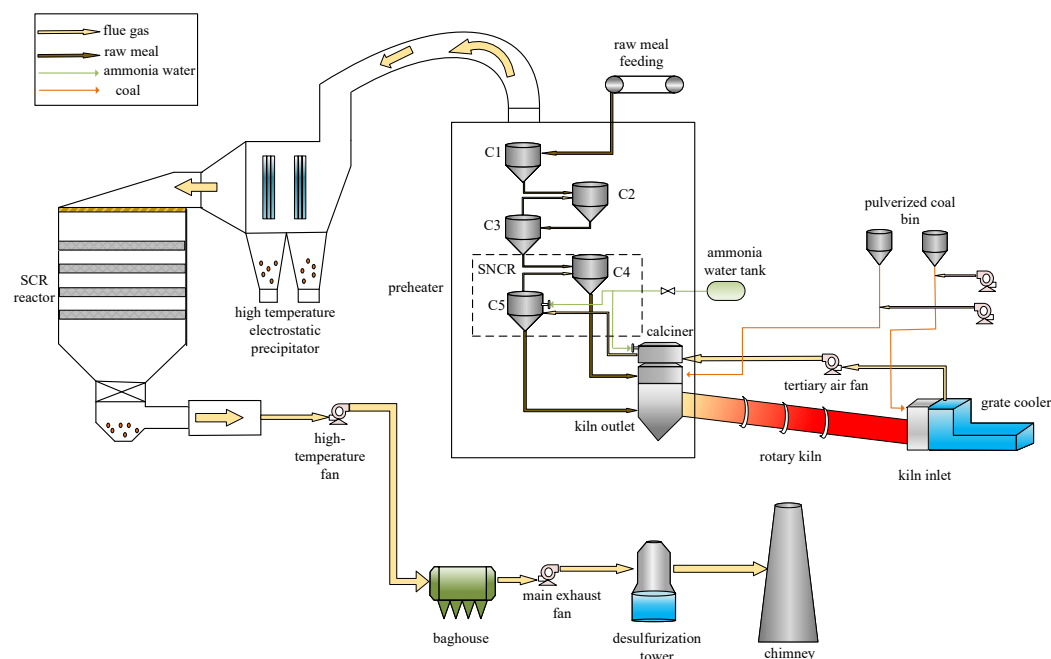


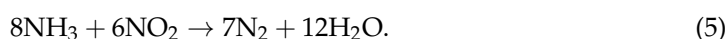
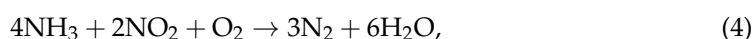
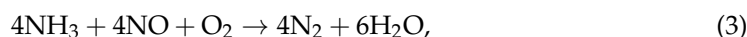
Figure 1. Schematic diagram of the SNCR–SCR hybrid denitrification system.

Raw meal first passes through a five-stage cyclone preheater (C1–C5), where it is heated by high-temperature flue gas before entering the precalciner. After further decomposition in the precalciner, the material enters the rotary kiln and undergoes high-temperature calcination to form cement clinker. The clinker is then cooled in the grate cooler. Pulverized coal is supplied through two streams. One stream is fed to the kiln head for intensive combustion, and the generated NO_x flows toward the kiln inlet with the flue gas. The other stream is injected into the precalciner through the kiln-inlet coal feeding system, which supports staged-combustion denitrification. By adjusting the kiln-inlet coal feed rate, a reducing atmosphere can be formed in the lower part of the precalciner. In this region, reducing gases generated by incomplete fuel combustion react with NO_x carried by

the rotary-kiln flue gas, thereby reducing part of the NO_x to N_2 . The main reactions are expressed as follows:



After staged combustion, the flue gas rises to the middle and upper regions of the precalciner. In the SNCR reaction region, ammonia solution is injected to achieve primary NO_x reduction under high-temperature conditions. The flue gas then leaves the C1 outlet, passes through the high-temperature electrostatic precipitator for dust removal, and enters the SCR reactor. In the investigated system, the SCR reactor does not use an independent ammonia injection unit. Instead, it utilizes the unreacted ammonia from the upstream SNCR process as the reducing agent. Using the catalyst, residual NO_x is further removed through selective catalytic reduction. The main SCR reactions are as follows:



It should be noted that the effective SNCR reaction temperature window refers to the local reaction region where ammonia is injected, whereas the DCS temperature variables used in this study are process measurement points that reflect the overall thermal state of the kiln and denitrification system. Therefore, these temperature variables are used as operating-condition indicators for stack NO_x prediction rather than as direct measurements of the local SNCR reaction temperature.

Based on process-mechanism analysis and available DCS measurements, ten candidate auxiliary variables are initially selected for modeling: ammonia flow rate, precalciner cone-section temperature, kiln-head coal feed rate, kiln-tail coal feed rate, raw meal feed rate, SCR inlet NO_x concentration, SCR inlet temperature, first-layer pressure difference, rotary kiln current, and C1 outlet NO_x concentration. The prediction target is the stack NO_x concentration. Therefore, the initial dataset contains ten candidate input variables and one target variable.

3. Stack NO_x Concentration Prediction Model Based on DAC-IVY-Optimized TCN-SE-LSTM Network

As illustrated in Figure 2, this study establishes a stack NO_x concentration prediction model based on a DAC-IVY-optimized TCN-SE-LSTM network. The modeling procedure consists of six main steps: (1) process-mechanism analysis for initial variable selection, (2) data preprocessing, (3) feature selection, (4) delay compensation, (5) construction of the TCN-SE-LSTM prediction model, and (6) offline hyperparameter optimization using DAC-IVY.

3.1. Data Preprocessing

The raw DCS dataset contains 21,600 consecutive samples collected from an actual cement kiln SNCR–SCR hybrid denitrification system. Industrial DCS data are often affected by sensor noise, missing records, communication interruptions, and abnormal fluctuations. Therefore, data preprocessing is required before model construction.

First, abnormal samples are screened using a box-plot-based outlier detection method combined with process knowledge. Missing values caused by short-term sensor or transmission failures are filled using interpolation or a moving-average method to preserve the

continuity of the time series. The comparison of the stack NO_x sequence before and after preprocessing is shown in Figure 3.

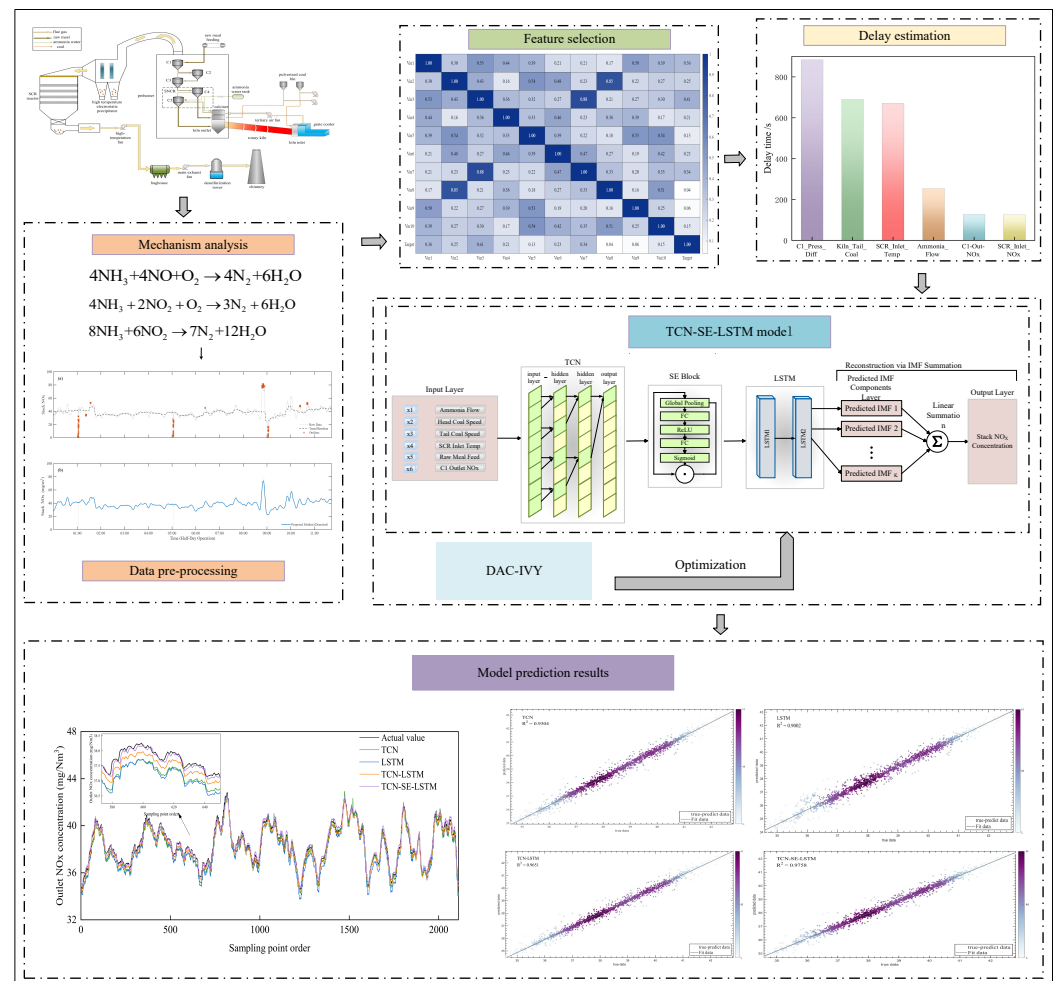


Figure 2. Development procedure of the stack NO_x concentration prediction model.

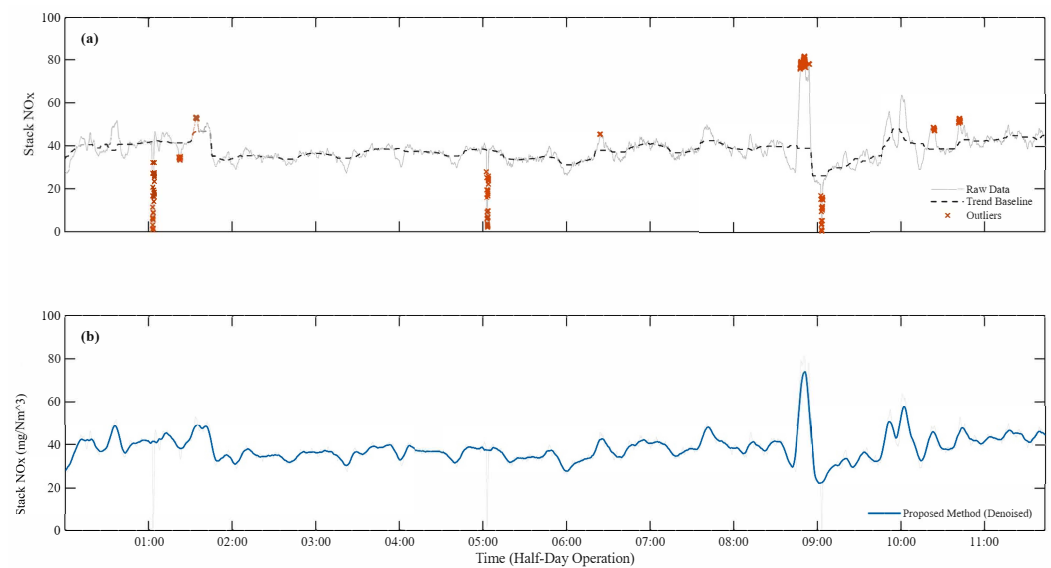


Figure 3. Comparison of the stack NO_x sequence before and after preprocessing. (a) shows the original signal with detected outliers and the trend baseline, whereas (b) shows the preprocessed stack NO_x sequence.

Second, VMD is used to handle the nonstationarity of the stack NO_x sequence. Compared with traditional denoising methods such as wavelet transform, VMD is a non-recursive adaptive signal decomposition method that decomposes a complex signal into several band-limited intrinsic mode functions (IMFs) with specific center frequencies [33]. In this study, VMD is applied only to the stack NO_x sequence rather than to all auxiliary input variables. Each IMF component is predicted using the proposed TCN-SE-LSTM model, and the final stack NO_x prediction is obtained by summing the predicted IMF components.

To avoid future-information leakage, the dataset is chronologically divided into training, validation, and testing sets before decomposition-based target construction, feature selection, delay compensation, model training, and performance evaluation. MIC-based feature selection and delay compensation are performed only on the training set, and the obtained feature subset and delay parameters are then fixed for validation and testing. For testing and online deployment, VMD is implemented using a sliding-window strategy. At each prediction instant, only historical samples available up to the current time are used for decomposition, and future stack NO_x values are never used. This protocol ensures that the reconstructed prediction is obtained only from predicted IMF components rather than from future measured target values.

The variation ranges of the initial candidate variables and the target variable after data screening are listed in Table 1. The variables are presented using standardized engineering names and units. During delay compensation, 500 initial samples are discarded according to the maximum identified delay, and 21,100 temporally aligned samples are finally used for model training, validation, and testing.

Table 1. Variation ranges of the candidate variables and target variable after data screening.

Variable	Parameter	Unit	Minimum	Maximum
Var1	Ammonia flow	L/h	2.85	14.26
Var2	Calcliner temperature	°C	800.00	1100.00
Var3	C1 outlet NO _x	mg/Nm ³	8.81	229.18
Var4	Kiln-head coal feed	t/h	434.77	1436.02
Var5	Kiln-tail coal feed	t/h	1015.03	1964.73
Var6	Raw meal feed	t/h	298.68	380.38
Var7	SCR inlet NO _x	mg/Nm ³	7.10	190.27
Var8	SCR inlet temperature	°C	218.80	233.60
Var9	C1 pressure difference	kPa	12.68	64.27
Var10	Kiln current	A	540.94	876.46
Var11	Stack NO _x target	mg/Nm ³	0.50	81.64

3.2. Feature Selection Based on MIC Algorithm

The cement kiln denitrification process is characterized by strong multivariable coupling and nonlinear interactions. Traditional linear correlation measures, such as the Pearson correlation coefficient [32], may fail to identify nonlinear dependencies among operating variables. Therefore, the Maximal Information Coefficient (MIC) [34] is adopted in this study to evaluate nonlinear associations between candidate variables and the target stack NO_x concentration.

For two random variables X and Y , their mutual information $I(X; Y)$ is defined as

$$I(X; Y) = \sum_{x \in X} \sum_{y \in Y} p(x, y) \log \frac{p(x, y)}{p(x)p(y)}. \quad (6)$$

The MIC value is then calculated as

$$\text{MIC}(X; Y) = \max_{n_x n_y < B} \frac{I(X; Y)}{\log_2(\min(n_x, n_y))}, \quad (7)$$

where $p(x, y)$ denotes the joint probability distribution, $p(x)$ and $p(y)$ denote the marginal probability distributions, n_x and n_y are the grid partition numbers along the X and Y directions, respectively, and B is the upper bound of the total grid number. In this study, B is set to $N^{0.6}$, where N is the number of samples used for MIC calculation. The MIC value ranges from 0 to 1, and a larger value indicates a stronger nonlinear association.

To avoid future-information leakage, MIC-based feature selection is performed only on the training dataset. The selected feature subset is then fixed and directly applied to the validation and testing datasets. Based on the ten candidate auxiliary variables and the target stack NO_x concentration, an 11×11 MIC matrix is calculated using the training data. The corresponding heat map is shown in Figure 4.



Figure 4. MIC heat map of the candidate variables and the target stack NO_x concentration.

To determine the final model inputs, a three-step screening strategy is adopted.

First, threshold-based relevance screening is performed. A baseline MIC threshold of 0.20 is used as the initial relevance criterion, and its influence on prediction performance is further evaluated in the Section 4. According to the MIC values between each candidate variable and the target variable, Var9 and Var10 show weak relevance and are removed from the input set.

Second, mechanism-guided retention is conducted. Although Var5 and Var8 have relatively low MIC values with the target variable, they are retained based on process-mechanism knowledge. Var5 represents the kiln-tail coal feed rate, which affects the combustion state and reducing atmosphere in the precalciner. Var8 represents the SCR inlet temperature, which influences catalyst activity and SCR reaction efficiency. Removing these variables solely according to statistical correlation may weaken the physical completeness of the input set.

Third, redundancy screening is applied to reduce duplicate information among candidate variables. Although Var2 and Var7 satisfy the initial relevance threshold, they show strong nonlinear redundancy with retained variables. Specifically, Var7 is strongly associated with Var3 because both variables describe NO_x concentrations at consecutive process

locations. Similarly, Var2 is strongly associated with Var8 because of thermal propagation in the kiln system. Therefore, Var2 and Var7 are removed to reduce input redundancy and model complexity.

Finally, six variables are retained as model inputs: Var1, Var3, Var4, Var5, Var6, and Var8. These variables provide complementary information about ammonia injection, upstream NO_x concentration, fuel supply, raw meal feeding, and SCR inlet thermal conditions.

3.3. Trend-Guided MIC-Based Delay Compensation

The cement kiln SNCR–SCR hybrid denitrification process has obvious process-dependent time delays caused by material transport, flue gas residence time, thermal inertia, and multistage denitrification reactions. Therefore, delay compensation is required before constructing the prediction model.

In this study, a trend-guided MIC-based delay compensation method is used. First, process knowledge and trend comparison between each selected input variable and the target stack NO_x concentration are used to determine the delay search interval $[\tau_{\min}, \tau_{\max}]$. Then, within the training dataset, each selected input variable is shifted by different candidate delay steps, and the MIC value between the shifted input variable and the target variable is calculated. The delay step corresponding to the maximum MIC value is selected as the optimal delay for that variable.

For each selected input variable x_i , the optimal delay step τ_i^* is obtained as

$$\tau_i^* = \arg \max_{\tau_i \in [\tau_{\min}, \tau_{\max}]} \{MIC(x_i(t - \tau_i), y(t))\}, \quad (8)$$

where $y(t)$ is the stack NO_x concentration at time t , and τ_i is an integer delay step measured in sampling intervals. The DCS sampling period in this study is 2 s; therefore, the corresponding delay time is calculated as $2\tau_i$ seconds.

All delay parameters are estimated only from the training dataset. After the optimal delays are determined, they are fixed and applied to the validation and testing datasets without recalculation. The calculated delay results are listed in Table 2. The maximum identified delay is 884 s, corresponding to 442 sampling steps. Therefore, a conservative alignment length of 500 samples is used to cover the maximum identified delay, and 21,100 aligned samples are retained for subsequent model training, validation, and testing.

Table 2. Delay compensation results of selected input variables.

Variable	Delay Step	Delay Time (s)
Var1	127	254
Var3	64	128
Var4	345	690
Var5	336	672
Var6	63	126
Var8	442	884

3.4. Input–Output Construction for the TCN-SE-LSTM Model

After feature selection and delay compensation, the model input is constructed using the six retained auxiliary variables. For each prediction instant t , a look-back window of length L_w is used to form the input tensor:

$$\mathbf{X}_t \in \mathbb{R}^{L_w \times M}, \quad (9)$$

where $M = 6$ denotes the number of selected input variables. Since VMD is applied only to the stack NO_x sequence, the IMF components are treated as supervised prediction

targets rather than additional input variables. The output layer predicts the next-step IMF components,

$$\hat{\mathbf{u}}_{t+1} = [\hat{u}_{1,t+1}, \hat{u}_{2,t+1}, \dots, \hat{u}_{K,t+1}], \quad (10)$$

where K is the number of VMD modes. The final stack NO_x concentration prediction is reconstructed by summing the predicted IMF components:

$$\hat{y}_{t+1} = \sum_{k=1}^K \hat{u}_{k,t+1}. \quad (11)$$

This design avoids expanding the input dimension by decomposing all auxiliary variables and ensures that the final prediction is reconstructed only from predicted IMF components.

3.4.1. TCN Network

The Temporal Convolutional Network (TCN) is a convolution-based sequence modeling architecture with strong parallel computation capability and stable gradient propagation [20]. Its core mechanisms include causal convolution and dilated convolution.

Causal convolution ensures that the output at time step t depends only on the current and previous input samples within the network. This structure is consistent with the requirement of real-time forecasting because future samples are not used in the convolution operation.

Dilated convolution introduces a dilation factor d to expand the receptive field without substantially increasing the number of trainable parameters. This enables the TCN to capture longer temporal dependencies in industrial time-series data. For an input sequence X , the dilated convolution operation with filter f , kernel size k , and dilation factor d is defined as

$$F(s) = (X *_d f)(s) = \sum_{i=0}^{k-1} f(i) \cdot X_{s-d \cdot i}, \quad (12)$$

where s is the time-step index. When $d = 1$, the operation reduces to standard convolution. By stacking multiple dilated causal convolutional layers with residual connections, the TCN can extract local temporal patterns and short-term fluctuations from delay-aligned multivariable sequences [35,36].

3.4.2. SE Block

After TCN-based feature extraction, the model obtains a multichannel temporal feature map. However, different feature channels do not contribute equally to stack NO_x prediction under varying operating conditions. Although MIC-based feature selection removes weakly relevant and redundant variables before model training, the dynamic contribution of the retained variables may still vary with combustion state, ammonia injection, raw meal feeding, and SCR inlet thermal conditions. Therefore, relying only on TCN feature extraction may not sufficiently distinguish the relative importance of different channels at different operating moments. Therefore, a Squeeze-and-Excitation (SE) block is introduced to learn channel-wise importance weights and recalibrate the TCN feature channels [37,38], as shown in Figure 5.

Let the TCN output feature map be denoted as

$$\mathbf{U} = [u_1, u_2, \dots, u_C], \quad (13)$$

where C is the number of channels and each u_c is a temporal feature sequence. The SE operation consists of three steps.

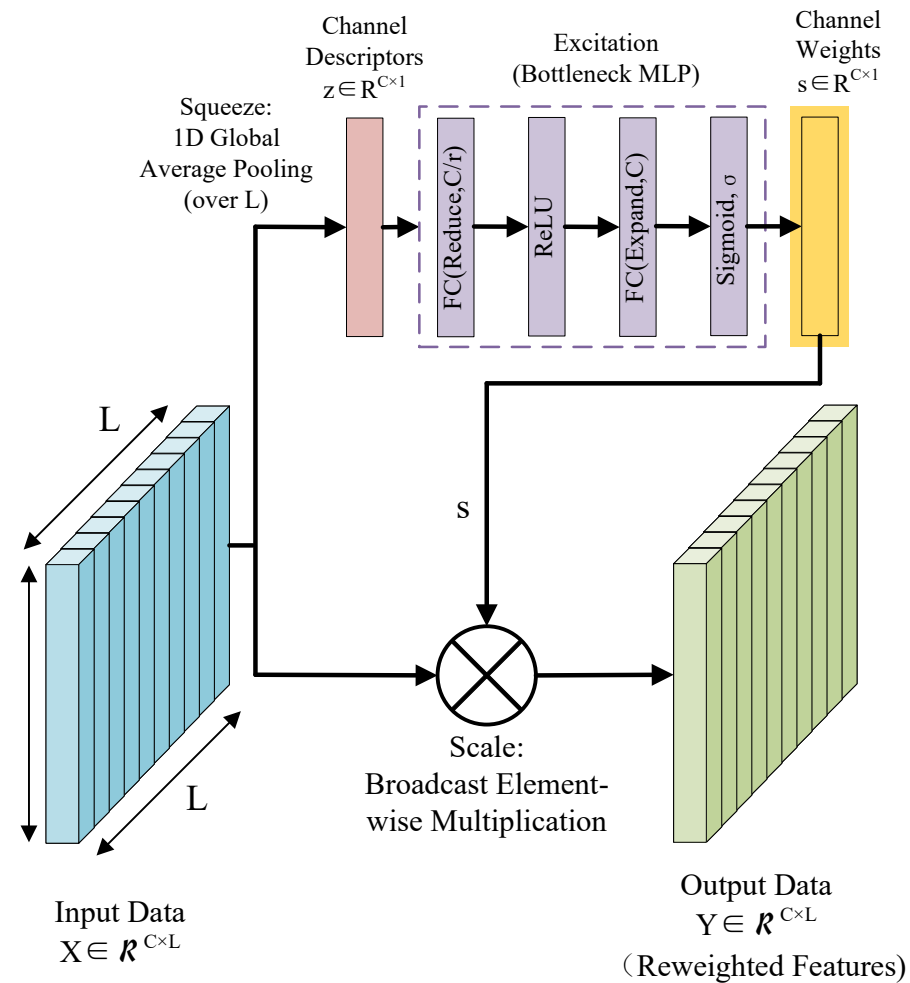


Figure 5. Structure of the SE block.

First, in the squeeze operation, global average pooling is used to aggregate the temporal information of each channel:

$$z_c = F_{sq}(u_c) = \frac{1}{L_w} \sum_{i=1}^{L_w} u_c(i), \quad (14)$$

where z_c is the channel descriptor of the c -th channel.

Second, in the excitation operation, two fully connected layers are used to model nonlinear channel dependencies and generate normalized channel weights:

$$\mathbf{s} = F_{ex}(\mathbf{z}, W) = \sigma(W_2 \delta(W_1 \mathbf{z})), \quad (15)$$

where $W_1 \in \mathbb{R}^{C/r \times C}$, $W_2 \in \mathbb{R}^{C \times C/r}$, r denotes the reduction ratio, $\delta(\cdot)$ is the ReLU activation function, and $\sigma(\cdot)$ is the sigmoid activation function.

Third, in the recalibration operation, the learned channel weight s_c is multiplied by the corresponding feature channel:

$$\tilde{u}_c = s_c \cdot u_c. \quad (16)$$

Through this process, informative channels are enhanced, whereas less relevant channels are suppressed. This mechanism helps the model adaptively emphasize the most influential process features after TCN-based temporal extraction, which is particularly

useful for multivariable industrial soft sensing where variable contributions change with operating conditions.

3.4.3. LSTM Network

The Long Short-Term Memory (LSTM) network is suitable for modeling long-range dependencies in nonlinear time-series regression tasks [39]. In this study, the LSTM module receives the SE-recalibrated feature sequence and further captures long-lag dynamic behavior in the cement kiln denitrification process. The structure of the LSTM unit is shown in Figure 6.

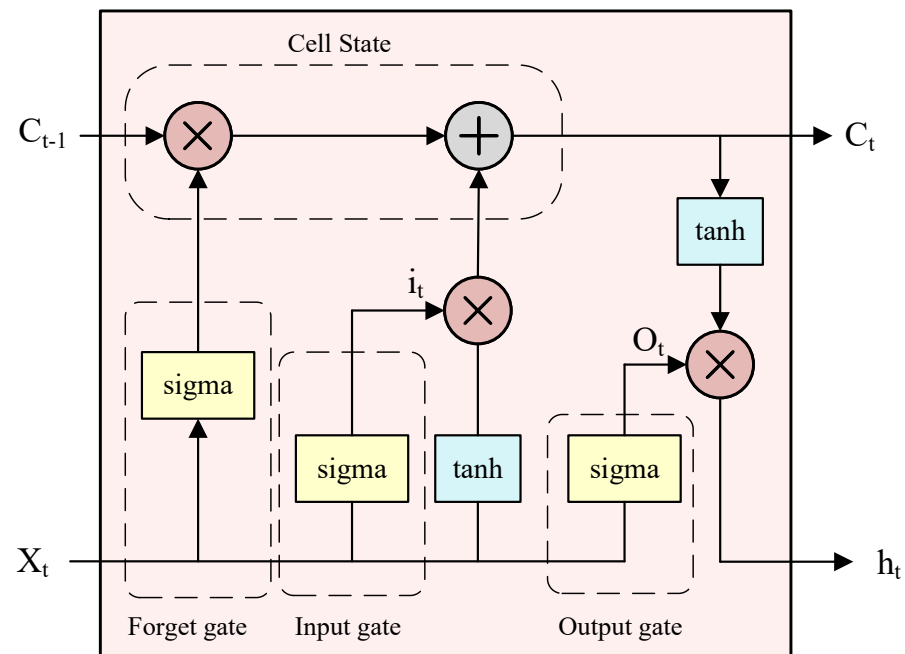


Figure 6. Structure of the LSTM network.

The LSTM state update is controlled by the forget gate, input gate, candidate cell state, and output gate:

$$f_t = \sigma(W_f[h_{t-1}, x_t] + b_f), \quad (17)$$

$$i_t = \sigma(W_i[h_{t-1}, x_t] + b_i), \quad (18)$$

$$\tilde{C}_t = \tanh(W_c[h_{t-1}, x_t] + b_c), \quad (19)$$

$$C_t = f_t \cdot C_{t-1} + i_t \cdot \tilde{C}_t, \quad (20)$$

$$o_t = \sigma(W_o[h_{t-1}, x_t] + b_o), \quad (21)$$

$$h_t = o_t \cdot \tanh(C_t), \quad (22)$$

where x_t denotes the input feature vector at time step t , h_t is the hidden state, C_t is the cell state, and W and b denote trainable weights and biases, respectively.

3.4.4. TCN-SE-LSTM Network

Based on the complementary characteristics of the above modules, a serial TCN-SE-LSTM architecture is constructed for stack NO_x concentration prediction, as shown in Figure 7.

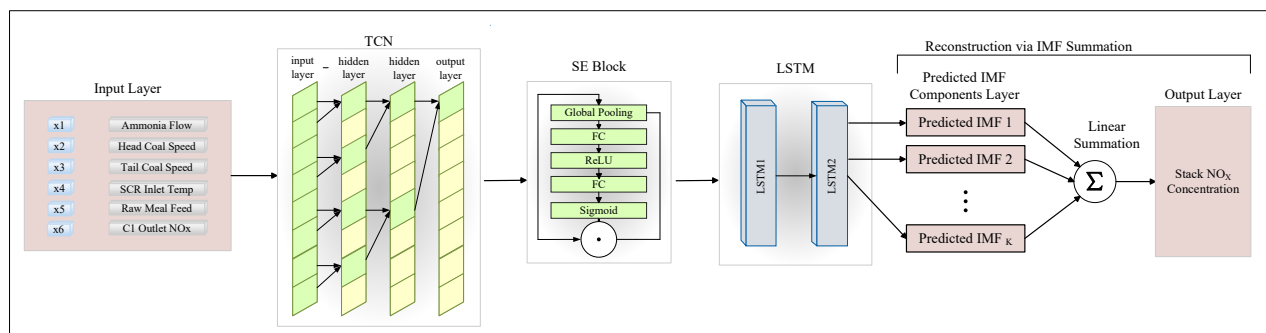


Figure 7. Structure of the TCN-SE-LSTM network.

The TCN module first processes the delay-aligned multivariable input sequence and extracts local temporal patterns through dilated causal convolutions. The SE block then recalibrates the extracted feature channels by assigning adaptive weights according to their relevance to the prediction task. This allows the network to emphasize informative process variables and suppress redundant feature responses under varying operating conditions. Finally, the LSTM module models long-term dependencies in the recalibrated feature sequence and outputs the predicted IMF components for the next time step. The predicted stack NO_x concentration is then obtained by summing the predicted IMF components. In this serial architecture, TCN, SE, and LSTM play different roles rather than being simply stacked: TCN focuses on local temporal feature extraction, SE performs channel-wise feature recalibration, and LSTM captures long-lag dynamic behavior. This complementary design explains why the proposed TCN-SE-LSTM model can outperform the simpler TCN-LSTM structure in the network-comparison experiment.

3.5. DAC-IVY Algorithm

3.5.1. Ivy Algorithm

The Ivy Algorithm (IVY) is a population-based metaheuristic optimization algorithm inspired by the growth and spreading behavior of ivy plants [40]. In IVY, each candidate solution is regarded as an ivy individual, and the population evolves through growth-rate updating, position updating, and fitness-based selection. This mechanism enables the algorithm to search the solution space by combining exploration and exploitation.

Let $X_i(t)$ denote the position of the i -th candidate solution at iteration t , and let $\Delta Gv_i(t)$ denote its growth-rate vector. The growth-rate vector is updated by a stochastic rule:

$$\Delta Gv_i(t+1) = r^2 \circ \mathcal{N}(0,1) \circ \Delta Gv_i(t), \quad (23)$$

where r is a random number in $[0,1]$, $\mathcal{N}(0,1)$ denotes a standard normal random vector, and $\circ$ represents element-wise multiplication.

The position of a new candidate solution is then updated according to the current best solution and the growth-rate vector:

$$X_i^{\text{new}}(t+1) = X_{\text{best}}(t) + \beta \Delta Gv_i(t+1), \quad (24)$$

where $X_{\text{best}}(t)$ is the best solution found so far and β is the growth factor controlling the search step. After the new solution is generated, a fitness-based selection strategy is applied. If the new solution achieves a better fitness value than the original one, it replaces the original solution. Through iterative updating and selection, the population gradually moves toward regions with better fitness.

3.5.2. Dual Adaptive Constrained Ivy Algorithm

To improve the suitability of IVY for hyperparameter optimization in the proposed soft-sensing model, this study develops a Dual Adaptive Constrained Ivy Algorithm (DAC-IVY). The improvement focuses on two aspects: nonlinear adaptive step-size control and constraint handling based on the valid hyperparameter domain.

1. Nonlinear adaptive growth factor: In the standard IVY algorithm, a fixed growth factor may lead to insufficient global exploration in the early stage or unstable local search in the later stage. Therefore, a nonlinear adaptive growth factor is introduced:

$$\tilde{\beta}(t) = \beta_{\min} + (\beta_{\max} - \beta_{\min}) \left(1 - \frac{t}{T_{\max}}\right)^3, \quad (25)$$

where $\beta_{\max}$ and $\beta_{\min}$ are the upper and lower bounds of the growth factor, respectively, and $T_{\max}$ is the maximum number of iterations. In the early stage, a larger $\tilde{\beta}(t)$ enhances exploration over the hyperparameter search space. As the iteration proceeds, $\tilde{\beta}(t)$ decreases nonlinearly, allowing the algorithm to perform a more refined local search around promising candidate solutions.

2. Constraint handling based on the valid hyperparameter domain: The candidate solutions generated by metaheuristic algorithms may exceed physically or structurally meaningful hyperparameter ranges. To avoid invalid model configurations, a constrained fitness function is introduced:

$$F_{\text{fitness}}(X) = \text{RMSE}_{\text{val}}(X) + P(X), \quad (26)$$

where $\text{RMSE}_{\text{val}}(X)$ denotes the validation-set RMSE obtained by the model configured with candidate solution X , and $P(X)$ is the penalty term:

$$P(X) = \begin{cases} 0, & X \in \Omega_{\text{valid}}, \\ \gamma \|X - X_{\text{bound}}\|^2, & X \notin \Omega_{\text{valid}}, \end{cases} \quad (27)$$

where Ω_{valid} is the predefined valid hyperparameter domain, X_{bound} is the nearest boundary point of the valid domain, and γ is the penalty coefficient. For example, the VMD mode number is constrained within $K \in [3, 10]$, and the LSTM hidden units are constrained within a reasonable range to avoid underfitting or excessive model complexity.

It should be emphasized that DAC-IVY is used only for offline hyperparameter optimization. After the optimized hyperparameter combination is determined, the online prediction stage uses the fixed trained TCN-SE-LSTM model and does not execute the DAC-IVY optimization process.

3.6. Hyperparameter Optimization of the Hybrid TCN-SE-LSTM Network Using DAC-IVY

The proposed DAC-IVY algorithm is used to optimize the key hyperparameters of the VMD-assisted TCN-SE-LSTM model. The overall modeling and optimization workflow is shown in Figure 8. First, raw DCS data are collected from the cement kiln SNCR-SCR hybrid denitrification system, and candidate variables are determined based on process-mechanism analysis. Then, data preprocessing, VMD decomposition of the stack NO_x sequence, MIC-based feature selection, and delay compensation are conducted to construct temporally aligned input-output samples. Finally, DAC-IVY searches for an optimized hyperparameter combination using the validation-set RMSE as the fitness criterion.

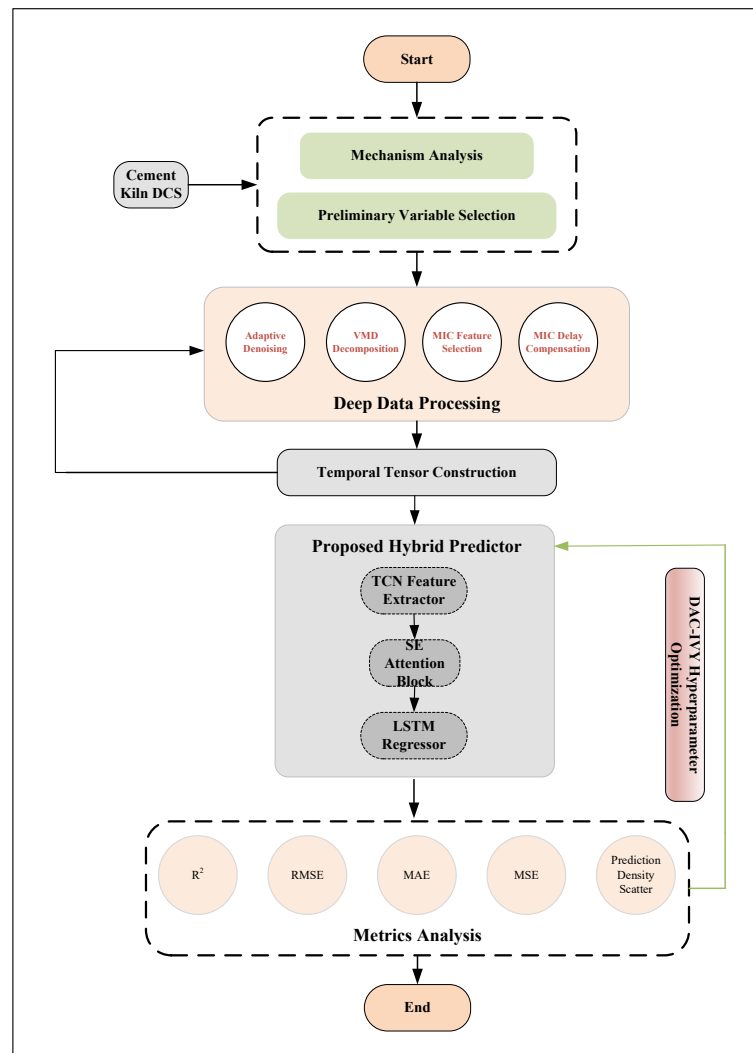


Figure 8. Overall workflow of the proposed dynamic soft-sensing model.

In this study, seven hyperparameters are optimized: the number of convolutional kernels, convolutional kernel size, number of TCN blocks, dropout rate, learning rate, number of LSTM hidden units, and VMD mode number K . Their search ranges and optimized values are listed in Table 3. Continuous parameters are directly updated during the optimization process, whereas integer-valued parameters are rounded to the nearest valid integer before model training and evaluation.

Table 3. Optimized hyperparameters of the VMD-assisted TCN-SE-LSTM model.

Parameter	Search Range	Optimized Value
Number of convolutional kernels	16–128	47
Convolutional kernel size	2–7	6
Number of TCN blocks	1–4	3
Dropout rate	0.01–0.50	0.01
Learning rate	0.0001–0.0100	0.0047578
LSTM hidden units	32–256	113
VMD mode number K	3–10	8

The optimized model uses the delay-aligned auxiliary variables as inputs and predicts the next-step IMF components of the stack NO_x sequence. The predicted IMF components are then summed to obtain the final stack NO_x concentration prediction. In this framework,

the TCN module extracts local temporal patterns from the multivariable input sequence, the SE block recalibrates informative feature channels, and the LSTM module captures long-lag dynamic dependencies. DAC-IVY improves the model configuration through offline hyperparameter optimization, while the online stage only requires sliding-window VMD preprocessing and forward inference using the trained model.

4. Experiments and Analysis

4.1. Evaluation Metrics and Experimental Setup

To evaluate the prediction accuracy and stability of the proposed model, four commonly used metrics are adopted: mean absolute error (MAE), mean squared error (MSE), root mean squared error (RMSE), and coefficient of determination (R^2) [41]. They are defined as follows:

$$e_{\text{MAE}} = \frac{1}{N} \sum_{i=1}^N |y_i - \hat{y}_i|, \quad (28)$$

$$e_{\text{MSE}} = \frac{1}{N} \sum_{i=1}^N (y_i - \hat{y}_i)^2, \quad (29)$$

$$e_{\text{RMSE}} = \sqrt{\frac{1}{N} \sum_{i=1}^N (y_i - \hat{y}_i)^2}, \quad (30)$$

$$R^2 = 1 - \frac{\sum_{i=1}^N (y_i - \hat{y}_i)^2}{\sum_{i=1}^N (y_i - \bar{y})^2}, \quad (31)$$

where y_i is the measured stack NO_x concentration, $\hat{y}_i$ is the predicted value, $\bar{y}$ is the mean value of the measured samples, and N is the number of testing samples.

The raw DCS dataset contains 21,600 consecutive samples collected from the investigated cement kiln SNCR–SCR hybrid denitrification system. After delay compensation, 500 initial samples are discarded according to the maximum identified delay, and 21,100 temporally aligned samples are retained. The aligned samples are chronologically divided into training, validation, and testing sets with a ratio of 8:1:1. Feature selection and delay compensation are performed only on the training set, and the obtained feature subset and delay parameters are fixed for the validation and testing sets. Hyperparameter optimization is conducted using the validation set, while the testing set is used only for final performance evaluation. Unless otherwise specified, all metrics reported in this section are calculated on the testing set. All reported prediction errors are calculated after inverse normalization. Therefore, MAE and RMSE are reported in the original physical unit of mg/Nm^3 , while MSE is reported in $(\text{mg}/\text{Nm}^3)^2$.

4.2. Impact of Feature Selection on Modeling Performance

To evaluate the effect of MIC-based feature selection, the model performance before and after feature selection is compared. Before feature selection, the ten candidate auxiliary variables determined by process-mechanism analysis are used as model inputs. After feature selection, the six retained variables, namely Var1, Var3, Var4, Var5, Var6, and Var8, are used as model inputs. To isolate the effect of feature selection, VMD preprocessing and delay compensation are not applied in this comparison. Other model settings remain unchanged. The prediction curves are shown in Figure 9, and the evaluation metrics are listed in Table 4.

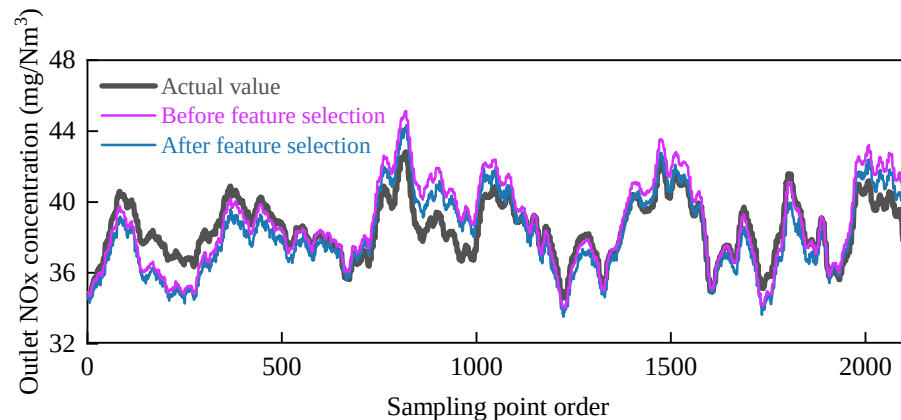


Figure 9. Comparison of stack NO_x concentration prediction performance before and after feature selection.

Table 4. Evaluation metrics before and after feature selection.

Model	MAE	MSE	RMSE	R ²
Before selection	0.6295	0.4123	0.6402	0.9055
After selection	0.3750	0.1937	0.4402	0.9486

Table 4 shows that the RMSE decreases from 0.6402 mg/Nm³ to 0.4402 mg/Nm³ after feature selection, and R² increases from 0.9055 to 0.9486. This indicates that removing weakly relevant and redundant variables reduces input noise and improves model accuracy. However, the remaining prediction error is still relatively large because the process-dependent time delays between the input variables and stack NO_x concentration have not yet been compensated.

4.3. Sensitivity Analysis of the MIC Threshold

In the MIC-based feature selection procedure, the threshold directly affects the number of retained variables and the prediction performance. To justify the selection of 0.20 as the baseline threshold, a sensitivity analysis was conducted by testing three representative thresholds: 0.15, 0.20, and 0.25. For each threshold, the same feature screening and mechanism-guided retention strategy was applied, while VMD preprocessing and delay compensation were not used in this comparison, so as to isolate the influence of the MIC threshold on feature selection. The results are shown in Table 5.

Table 5. Sensitivity analysis of the effects of MIC thresholds on model performance.

Threshold	Retained Variables	Number	MAE	MSE	RMSE	R ²
0.15	Var1, Var3, Var4, Var5, Var6, Var8, Var10	7	0.3920	0.2130	0.4615	0.9380
0.20	Var1, Var3, Var4, Var5, Var6, Var8	6	0.3750	0.1937	0.4402	0.9486
0.25	Var1, Var3, Var5, Var8	4	0.4450	0.2621	0.5120	0.9125

Table 5 shows that the threshold of 0.20 achieves the best prediction performance among the tested thresholds, with the lowest RMSE of 0.4402 and the highest R² of 0.9486. When the threshold is reduced to 0.15, Var10 is additionally retained; however, the RMSE increases to 0.4615, indicating that weakly relevant variables may introduce redundant information and noise. When the threshold is increased to 0.25, only four variables are retained, and the RMSE further increases to 0.5120, suggesting that useful process information is lost when the threshold is too strict. Therefore, 0.20 is selected as the baseline

MIC threshold because it provides the best balance between prediction accuracy, input compactness, and process interpretability.

4.4. Impact of Delay Compensation on Modeling Performance

After feature selection, the effect of trend-guided MIC-based delay compensation is evaluated. The selected variables are first used without delay compensation and then aligned according to the identified delay parameters. Except for the delay compensation operation, all other model settings are kept the same. The prediction curves are shown in Figure 10, and the evaluation metrics are listed in Table 6.

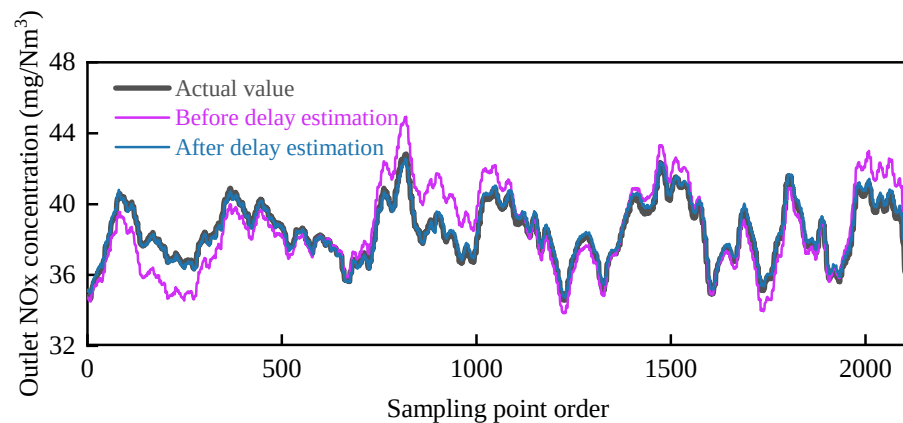


Figure 10. Comparison of prediction performance before and after delay compensation.

Table 6. Evaluation metrics before and after delay compensation.

Model	MAE	MSE	RMSE	R^2
Before delay compensation	0.3750	0.1937	0.4402	0.9486
After delay compensation	0.3120	0.1444	0.3800	0.9615

Table 6 shows that the RMSE decreases from 0.4402 to 0.3800 after delay compensation, while R^2 increases from 0.9486 to 0.9615. This improvement confirms that temporal alignment of the auxiliary variables is necessary for modeling the delayed response of stack NO_x concentration. Nevertheless, high-frequency fluctuations and nonstationary components remain difficult to capture, indicating the need for further signal decomposition.

4.5. Impact of VMD Preprocessing on Modeling Performance

To evaluate the effect of VMD preprocessing, the model performance with and without VMD is compared after feature selection and delay compensation. In the VMD-assisted version, VMD is applied only to the stack NO_x sequence. The proposed model predicts the next-step IMF components, and the final stack NO_x prediction is obtained by summing the predicted IMF components. The testing stage follows the sliding-window VMD protocol described above to avoid future-information leakage. All VMD-assisted testing results were obtained using this sliding-window protocol, and no future stack NO_x measurements were used for test-set decomposition or reconstruction. The prediction curves are shown in Figure 11, and the quantitative results are listed in Table 7.

As shown in Table 7, VMD preprocessing further reduces the RMSE from 0.3800 to 0.2084 and increases R^2 from 0.9615 to 0.9844. The comparison indicates that decomposing the nonstationary stack NO_x sequence into multiple intrinsic mode functions helps the model capture multi-scale temporal patterns and improves prediction stability.

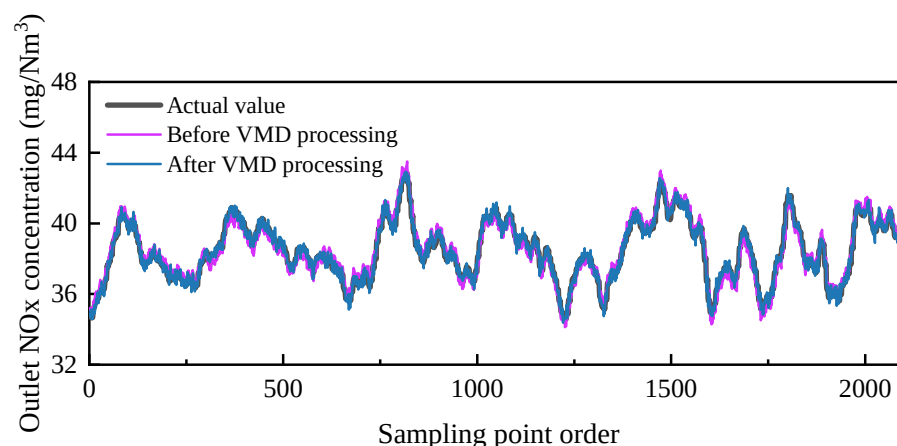


Figure 11. Impact of VMD preprocessing on prediction performance.

Table 7. Evaluation metrics before and after VMD preprocessing.

Model	MAE	MSE	RMSE	R^2
Without VMD	0.3120	0.1444	0.3800	0.9615
With VMD	0.1650	0.0434	0.2084	0.9844

4.6. Impact of Different Network Architectures on Modeling Performance

After feature selection, delay compensation, and VMD preprocessing, different network architectures are compared to evaluate the effectiveness of the proposed TCN-SE-LSTM model. The compared models include TCN, LSTM, TCN-LSTM, Transformer Encoder, and TCN-SE-LSTM. All models use the same training, validation, and testing datasets as well as the same preprocessed input-output samples. The quantitative evaluation metrics of all compared models, including the Transformer Encoder, are listed in Table 8.

Table 8. Evaluation metrics of different network architectures.

Model	MAE	MSE	RMSE	MAPE (%)	R^2
TCN	0.3750	0.1937	0.4402	0.9763	0.9304
LSTM	0.4743	0.2776	0.5269	1.2491	0.9002
TCN-LSTM	0.2701	0.0970	0.3115	0.6994	0.9651
Transformer Encoder	0.3301	0.1745	0.4177	0.8458	0.9373
TCN-SE-LSTM	0.1650 0.0434	0.2084	0.4301	0.9844	

Table 8 shows that the proposed TCN-SE-LSTM model achieves the lowest prediction error and the highest R^2 among the compared network architectures. Specifically, the proposed model obtains an MAE of 0.1650, an MSE of 0.0434, an RMSE of 0.2084, a MAPE of 0.4301%, and an R^2 of 0.9844. Compared with LSTM, the TCN-LSTM model reduces the RMSE from 0.5269 to 0.3115, indicating that the TCN module improves the extraction of local temporal patterns. Compared with TCN-LSTM, the proposed TCN-SE-LSTM further reduces the RMSE from 0.3115 to 0.2084 and reduces the MAPE from 0.6994% to 0.4301%, which verifies the effectiveness of the SE block in recalibrating informative feature channels under varying operating conditions.

To further address the comparison with Transformer-based sequence modeling architectures, a lightweight Transformer Encoder model is evaluated under the same processed dataset, chronological data partition, and evaluation metrics. The key hyperparameters of the Transformer Encoder are selected using the validation set, and the testing set is

used only for final evaluation. As shown in Table 8, the Transformer Encoder achieves an RMSE of 0.4177, a MAPE of 0.8458%, and an R^2 of 0.9373 on the test set, which is better than the single LSTM model but still inferior to the proposed TCN-SE-LSTM model. This result indicates that, for the present single-step stack NO_x soft-sensing task, the proposed architecture is more suitable for capturing local temporal dependencies, channel-wise variable importance, and long-lag dynamic behavior. Although Transformer-based models are powerful for general time-series forecasting, the limited selected input variables, process-dependent delays, and online soft-sensing requirement in this industrial case make the TCN-SE-LSTM structure more effective and compact.

4.7. Superiority of the DAC-IVY Optimization Algorithm

To verify the optimization capability of the proposed DAC-IVY algorithm, it is compared with standard IVY and particle swarm optimization (PSO). The same hyperparameter search ranges, training-validation split, and fitness function are used for all optimization algorithms. The convergence curves are shown in Figure 12. The final fitness values obtained from 30 independent runs are summarized in Figure 13 and Table 9.

As shown in Figure 12, DAC-IVY converges to a lower final fitness value than PSO and standard IVY. The statistical results in Table 9 further show that DAC-IVY achieves the lowest mean fitness and the smallest standard deviation over 30 independent runs. This indicates that the nonlinear adaptive growth factor and constrained search strategy improve both optimization accuracy and stability. These results support the use of DAC-IVY for offline hyperparameter optimization of the proposed prediction model.

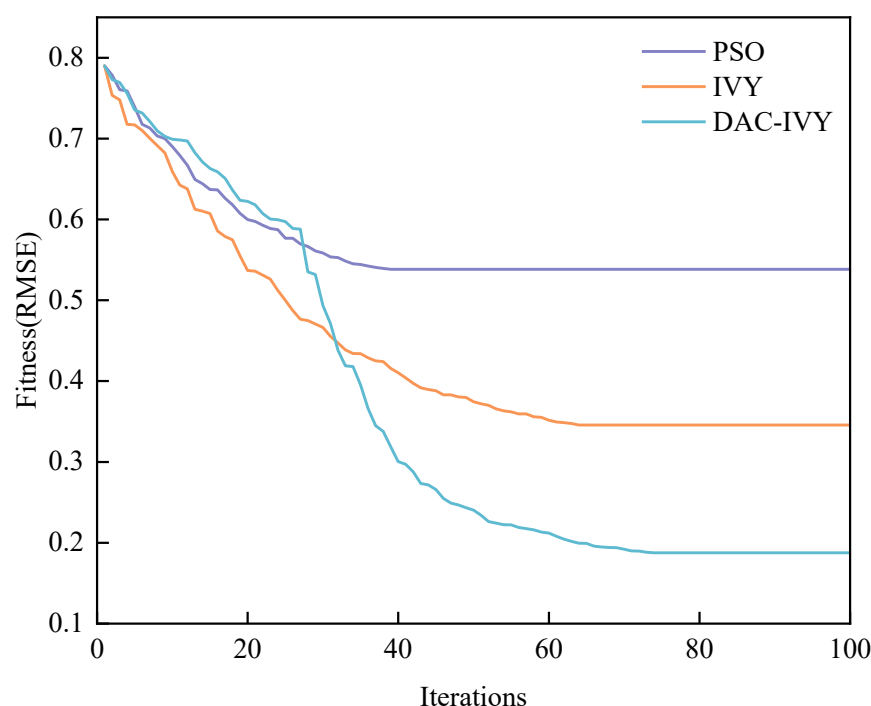


Figure 12. Fitness convergence curves of different optimization algorithms.

Table 9. Statistical comparison of optimization algorithms over 30 independent runs.

Algorithm	Best	Mean	Std.	Worst
PSO	0.3557	0.5393	0.0858	0.6743
IVY	0.2432	0.3537	0.0461	0.4491
DAC-IVY	0.1705	0.1896	0.0128	0.2177

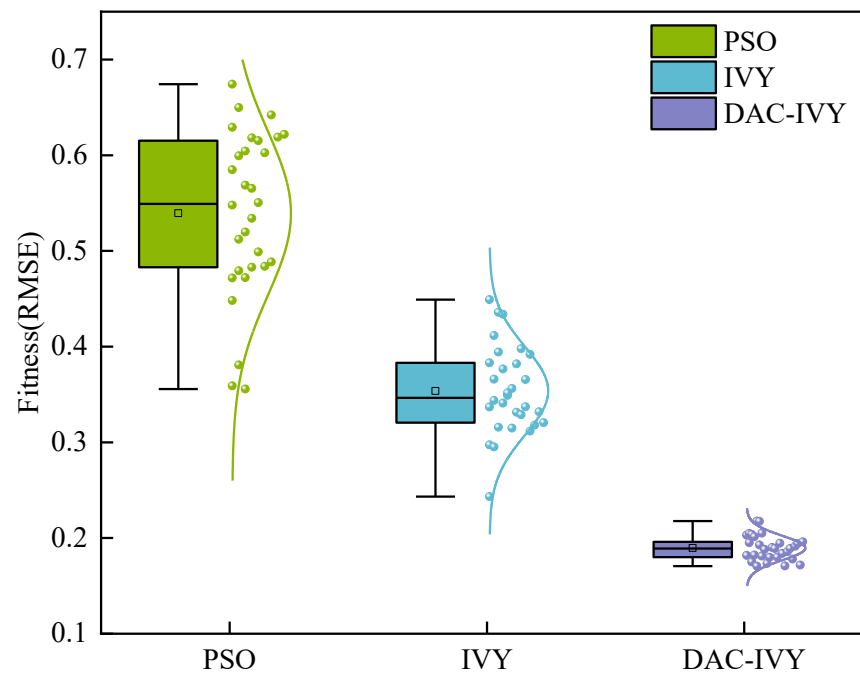


Figure 13. Boxplot of final fitness values over 30 independent runs.

4.8. Statistical Significance and Confidence Interval Analysis

To further evaluate whether the performance improvement of the proposed TCN-SE-LSTM model is statistically significant, a paired Wilcoxon signed-rank test is conducted using the absolute prediction errors of the testing samples. The proposed model is compared with TCN, LSTM, and TCN-LSTM on the same testing set. Since the Transformer Encoder is added as an architecture-level comparison model with summary evaluation metrics, and its sample-wise prediction errors are not included in this statistical test, the Wilcoxon analysis focuses on the four models with available paired testing errors.

As shown in Table 10, all p -values are far below 0.001, indicating that the reduction in absolute prediction error achieved by the proposed TCN-SE-LSTM model is statistically significant compared with TCN, LSTM, and TCN-LSTM.

Table 10. Paired Wilcoxon signed-rank test based on absolute prediction errors.

Comparison	Wilcoxon W	p -Value
TCN-SE-LSTM vs. TCN	333,735.0	7.8006×10^{-171}
TCN-SE-LSTM vs. LSTM	140,362.0	6.8016×10^{-265}
TCN-SE-LSTM vs. TCN-LSTM	479,670.0	1.4746×10^{-113}

In addition, nonparametric bootstrap resampling is used to estimate the confidence intervals of the main evaluation metrics of the proposed model. A total of 5000 bootstrap resamples are generated from the testing samples with replacement, and the 2.5th and 97.5th percentiles are used as the 95% confidence interval.

The bootstrap results in Table 11 show that the 95% confidence intervals of MAE, RMSE, MAPE, and R^2 are narrow. This indicates that the prediction performance of the proposed model is stable on the testing set and that the reported accuracy metrics are not caused by a small number of outlier samples.

Table 11. Bootstrap 95% confidence intervals of the proposed TCN-SE-LSTM model.

Metric	Value	95% CI Lower	95% CI Upper
MAE	0.1650	0.1595	0.1704
RMSE	0.2084	0.1996	0.2173
MAPE (%)	0.4301	0.4156	0.4445
R^2	0.9844	0.9828	0.9858

4.9. Computational Cost and Online Feasibility

The DAC-IVY algorithm is used only for offline hyperparameter optimization. After the optimized hyperparameters are determined, the online prediction stage does not execute DAC-IVY again. Therefore, the online computational burden mainly comes from sliding-window VMD preprocessing, forward inference of the trained TCN-SE-LSTM model, and IMF summation. This offline–online separation is important for industrial deployment because the iterative population-based search of DAC-IVY is not involved in real-time DCS calculations.

From the perspective of online computational complexity, the trained neural network only performs fixed forward propagation. The TCN module extracts temporal features through causal dilated convolutions, the SE block performs lightweight channel-wise recalibration through global pooling and two fully connected layers, and the LSTM module captures long-lag dynamic behavior using the fixed hidden units determined in the offline stage. Therefore, after offline optimization, the online complexity is mainly determined by the look-back window length, the number of selected input variables, the number of convolution filters, and the LSTM hidden units, rather than by the DAC-IVY population size or iteration number.

To evaluate the online feasibility of the proposed model, the average inference time of the trained TCN-SE-LSTM model was measured. As shown in Table 12, the average neural-network inference time per sample is 14.5 ms, which is much lower than the DCS sampling period of 2000 ms. This indicates that the neural-network inference stage can satisfy the real-time calculation requirement of the cement kiln DCS. Although DAC-IVY requires iterative model evaluations during offline optimization, this cost does not affect online deployment because the optimized hyperparameters are fixed before real-time prediction.

Table 12. Online feasibility analysis of the proposed model.

Item	Value
Optimization mode of DAC-IVY	Offline only
Online execution of DAC-IVY	No
Optimized hyperparameters	7
Average TCN-SE-LSTM inference time per sample	14.5 ms
DCS sampling period	2000 ms

4.10. Robustness Evaluation Under Industrial Noise

To evaluate the robustness of the proposed model under measurement noise, Gaussian white noise is added to the selected auxiliary input variables in the testing set at signal-to-noise ratios (SNRs) of 30 dB and 20 dB, while the measured stack NO_x values used for evaluation remain unchanged. The proposed TCN-SE-LSTM model is compared with the baseline LSTM model under the same noise settings. The results are shown in Figure 14 and Table 13.

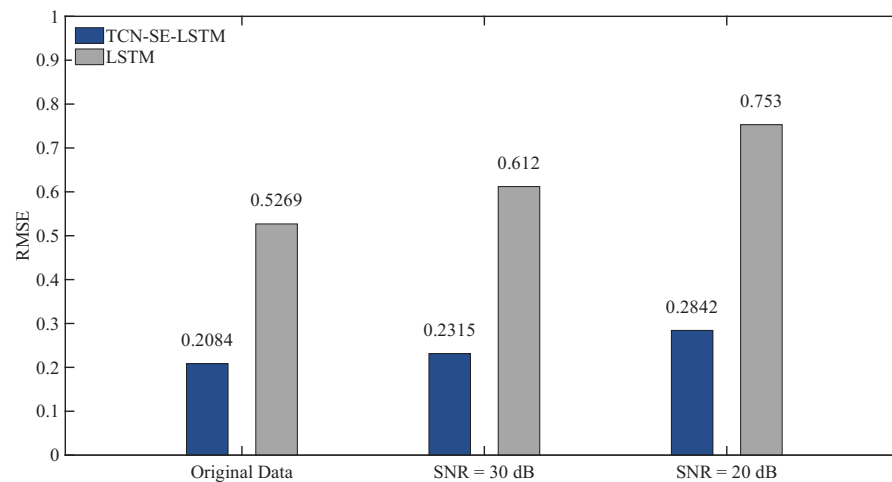


Figure 14. Robustness test under different noise levels.

Table 13. RMSE comparison under different noise levels.

Model	Original Data	SNR = 30 dB	SNR = 20 dB
LSTM	0.5269	0.6120	0.7530
TCN-SE-LSTM	0.2084	0.2315	0.2842

As shown in Table 13, the prediction error of both models increases as the noise level becomes stronger. However, the proposed TCN-SE-LSTM model maintains a lower RMSE than the LSTM model under all tested conditions. Even under the 20 dB noise condition, the RMSE of the proposed model is 0.2842, which is still lower than that of LSTM on the original clean testing data. This result indicates that the proposed framework has better noise robustness in complex industrial environments. It should be noted that this experiment mainly evaluates robustness against measurement noise in the selected auxiliary variables. It does not fully represent severe process disturbances such as start-up, shutdown, fault conditions, or large load-transition scenarios. These more complex transient operating conditions require additional long-term industrial data and will be further investigated in future work.

4.11. Practical Implications and Limitations

The proposed framework is developed for single-step stack NO_x soft sensing in an industrial cement kiln SNCR–SCR hybrid denitrification system. The raw DCS dataset contains 21,600 consecutive samples, and 21,100 temporally aligned samples are retained after delay compensation. Although the scale of these data is moderate compared with large-scale general deep learning datasets, it is practically meaningful for a single-line industrial soft-sensing case. In addition, the prediction task involves six selected auxiliary variables and single-step NO_x concentration estimation, rather than high-dimensional image recognition or long-horizon sequence generation. The chronological training-validation-testing partition and validation-set-based model configuration are adopted to reduce the risk of overfitting and future-information leakage.

The DCS sampling period is 2 s, which is much shorter than the identified delay times of the selected variables, ranging from 126 s to 884 s. Therefore, the sampling interval is sufficiently fine to describe the main transport, thermal-inertia, and denitrification-reaction dynamics considered in this study. The fixed delay parameters should be regarded as engineering approximations of the dominant time-lag relationships within the investigated operating period. For more complex operating conditions, dynamic delay estimation or online delay adaptation may further improve the adaptability of the soft sensor.

Although this study focuses on single-step prediction, the predicted stack NO_x concentration can provide useful support for subsequent ammonia-injection optimization. It can serve as a real-time soft-sensing signal for emission monitoring, auxiliary feedback correction, constraint supervision, and future model predictive control or feedforward-feedback coordinated ammonia-injection optimization. However, the present work does not directly design a closed-loop ammonia-injection controller. Multi-step forecasting and control-oriented soft sensing will be further investigated in future work.

The present study is based on data from one cement kiln production line. Therefore, the generalization capability across different kiln types, raw-material conditions, fuel properties, sensor configurations, and control strategies still requires further validation. The proposed framework's structure can be transferred, but plant-specific recalibration is necessary. In particular, MIC-based feature selection, delay compensation, normalization parameters, and model fine-tuning should be updated using data from the target production line. In addition, the current dataset does not fully cover seasonal variations, long-term operational drift, complete start-up and shutdown periods, or fault-disturbance scenarios. The robustness test mainly evaluates measurement-noise robustness and does not fully represent severe process disturbances or large load-transition conditions. Future work will extend the dataset to longer operating periods and multiple production lines, and will investigate transfer learning, incremental updating, dynamic delay adaptation, multi-step forecasting, and control-oriented optimization.

5. Conclusions

This study proposes a dynamic soft-sensing framework for single-step stack NO_x concentration prediction in a cement kiln SNCR-SCR hybrid denitrification system. The proposed method is designed to address the nonstationarity, multivariable coupling, and process-dependent time delays of industrial denitrification data. First, VMD is applied to the stack NO_x sequence to decompose the nonstationary target signal into several intrinsic mode functions (IMFs). To avoid future-information leakage, VMD is implemented under the chronological data-partitioning and sliding-window protocol. Second, MIC-based feature selection and delay compensation are used to identify nonlinear variable relevance and align the delayed effects of auxiliary variables. Third, a TCN-SE-LSTM prediction model is constructed, in which the TCN module extracts local temporal patterns, the SE block recalibrates informative feature channels, and the LSTM module captures long-lag dynamic dependencies. In addition, DAC-IVY is introduced for offline hyperparameter optimization of the VMD-assisted TCN-SE-LSTM model. During online application, DAC-IVY is not executed again, and the optimized model is used only for forward prediction.

Experimental results based on actual cement kiln DCS data show that the proposed framework achieves the best overall prediction performance among the compared models. The final model obtains an RMSE of 0.2084 mg/Nm³ and an R^2 of 0.9844 on the testing set, indicating good prediction accuracy and robustness under complex industrial operating conditions. The ablation studies further verify the effectiveness of feature selection, delay compensation, VMD preprocessing, the SE block, and DAC-IVY-based hyperparameter optimization. The comparison with the Transformer Encoder model further shows that the proposed TCN-SE-LSTM architecture provides a more suitable balance between local temporal feature extraction, channel-wise feature recalibration, long-lag dynamic modeling, and online feasibility for the investigated single-step industrial soft-sensing task. Moreover, the average inference time of the trained TCN-SE-LSTM model is approximately 14.5 ms, which is much shorter than the DCS sampling period of 2 s. This suggests that the neural-network inference stage has the potential to support real-time stack NO_x soft sensing and subsequent ammonia-injection optimization.

Despite these promising results, this study still has several limitations. First, the model was validated using data from a single cement kiln production line, and its generalization ability across kilns with different capacities, raw materials, fuel properties, sensor configurations, and control strategies requires further investigation. Second, the current work focuses on offline model validation and single-step stack NO_x soft sensing, which is consistent with the short sampling interval of the DCS and the requirement of real-time state monitoring. The extension to multi-step forecasting and control-oriented prediction can be further investigated according to the needs of advanced ammonia-injection optimization. Third, the fixed delay parameters used in this study are engineering approximations of the dominant time-lag relationships, and dynamic delay adaptation may further improve model adaptability under changing operating conditions. Fourth, uncertainty quantification and model reliability under seasonal variations, long-term operational drift, severe load transitions, start-up/shutdown, and fault-disturbance conditions require further study. Our future work will focus on online deployment, cross-kiln validation, transfer learning, incremental updating, dynamic delay estimation, and integration of the proposed soft sensor with denitrification control and multi-objective operational optimization.

Author Contributions: Conceptualization, Z.Z. and S.-Y.L.; methodology, S.-Y.L.; software, S.-Y.L.; validation, S.-Y.L., Y.-X.Z. and J.-L.Q.; formal analysis, S.-Y.L.; investigation, S.-Y.L. and X.-Y.T.; resources, Z.Z. and X.-Y.T.; data curation, S.-Y.L.; writing—original draft preparation, S.-Y.L.; writing—review and editing, Z.Z., Y.-X.Z., J.-L.Q. and X.-Y.T.; visualization, S.-Y.L.; supervision, Z.Z.; project administration, Z.Z.; funding acquisition, Z.Z. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the Baoding Key Laboratory of State Detection and Optimization Regulation for Integrated Energy System and the Key Research and Development Program of Tianjin, grant number 25YFKFYS00190.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The data used in this study are not publicly available due to industrial confidentiality and restrictions from the data-providing enterprise.

Acknowledgments: The authors would like to thank the data-providing enterprise for supporting the collection of industrial DCS data used in this study.

Conflicts of Interest: Author Xin-Yu Tang was employed by the Tianjin Sinoma Engineering Research Center Co., Ltd. The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

References

1. Andrew, R.M. Global CO₂ emissions from cement production, 1928–2018. *Earth Syst. Sci. Data* **2019**, *11*, 1675–1710. [[CrossRef](#)]
2. Poggianti, B.; Palazzolo, R.; Moliner, C. Rotary kiln simulation for energy recovery: The precalciner cement kiln case. *Therm. Sci. Eng. Prog.* **2024**, *54*, 102806. [[CrossRef](#)]
3. Jia, F. Air pollutants and prevention technology in cement industry. *Green Environ. Prot. Build. Mater.* **2021**, 36–37. [[CrossRef](#)]
4. Skalska, K.; Miller, J.S.; Ledakowicz, S. Trends in NO_x abatement: A review. *Sci. Total Environ.* **2010**, *408*, 3976–3989. [[CrossRef](#)] [[PubMed](#)]
5. Fei, M.; Cui, L.; Su, C.; Yin, H. Present Situation and Prospect of Exhaust Gas Denitrification Technology in Cement Kiln. *Cem. Technol.* **2019**, 89–94. [[CrossRef](#)]
6. Chen, D.; Wu, L.; Yuan, M.; Zhang, Q. Summary of Research on Denitrification Engineering Technology under Low Nitrogen Staged Combustion in Cement Kilns. *Hans J. Chem. Eng. Technol.* **2023**, *13*, 252–262.
7. Tang, X. Cement industry flue gas nitrogen oxide control technology and its application. *Cem. Eng.* **2020**, 76–77+83.
8. Liu, J.; Chen, S.; Wang, P.; Jiang, Y.; Zhang, Q.; Han, Y.; Long, J. Multi-parameter co-optimization of SNCR denitration system in a large MSW incinerator. *Chem. Eng. Process.-Process Intensif.* **2025**, *216*, 110458. [[CrossRef](#)]

9. Liu, J.; Xie, Z.; Xu, Y.; Han, Y.; Zhang, Q.; Guo, B.; Wang, Q.; Guo, X.; Du, H. Operating parameters optimization and efficient NO reduction of SNCR process in a municipal solid waste incinerator. *Waste Manag.* **2025**, *205*, 114996. [[CrossRef](#)] [[PubMed](#)]
10. Elkaee, S.; Phule, A.D.; Yang, J.H. Advancements in (SCR) technologies for NOx reduction: A comprehensive review of reducing agents. *Process Saf. Environ. Prot.* **2024**, *184*, 854–880. [[CrossRef](#)]
11. He, D.; Zhan, W.; Kang, Y.; Li, L. Gaussian Process-Based Economic MPC with Delay Compensation for SNCR-SCR Combined Denitration Systems. *Ind. Eng. Chem. Res.* **2024**, *63*, 19628–19639. [[CrossRef](#)]
12. Bose, D.; Ghanta, K.C.; Hens, A. CFD-based investigation of NOx removal from industrial waste gas by selective catalytic reduction (SCR) and selective non-catalytic reduction (SNCR) process using NH₃. *Int. J. Chem. React. Eng.* **2025**, *23*, 701–713. [[CrossRef](#)]
13. Fang, X.; Li, Y.; Qin, C.; Lu, X. Application of High-Temperature Medium-Dust SNCR+SCR Combined Denitrification Technology in Cement Kiln Systems. *Environ. Eng.* **2024**, 50–53.
14. Yan, A.; Cao, S. A NOx Concentration Prediction Model Based on a Sparse Regularization Stochastic Configuration Network. *Instrumentation* **2024**, *3*, 13–22. [[CrossRef](#)]
15. Duan, Q. NOx Emission Concentration Prediction Model for Coal-Fired Boilers Based on Wavenet. *Adv. Appl. Math.* **2023**, *12*, 5072–5082.
16. Ayadi, R.; Abd El-Aziz, R.M.; Taloba, A.I.; Aljuaid, H.; Hamed, N.O.; Khder, M.A. Deep Learning-Based Soft Sensors for Improving the Flexibility for Automation of Industry. *Wireless Commun. Mobile Comput.* **2022**, *2022*, 5450473. [[CrossRef](#)]
17. Fu, S.L.; Song, Q.; Yao, Q. Study on the catalysis of CaCO₃ in the SNCR deNOx process for cement kilns. *Chem. Eng. J.* **2015**, *262*, 9–17. [[CrossRef](#)]
18. Zhao, Z.; Liang, L.; Liu, S. Dynamic Soft Measurement Model of NOx Concentration at the Outlet of SNCR Denitrification System Based on Variable Selection and POA-NARX. *J. Chin. Soc. Power Eng.* **2025**, *45*, 592–601.
19. Liu, X.; Chen, Y.; He, N.; Yao, Y. NOx Concentration Prediction in Cement Denitrification Process Based on EEMD-MImRMR-BASBP. *Processes* **2023**, *11*, 317. [[CrossRef](#)]
20. Xue, S.; Li, F.; Cheng, S.; Feng, J.; Shen, X. A Hybrid Data-Driven Model for Predicting NOx Concentration in a Cement Clinker Calcination System. *IEEE Trans. Instrum. Meas.* **2025**, *74*, 2505312. [[CrossRef](#)]
21. Zou, J.; Wang, Z.; Zhu, Z.; Tan, Z. Explainable and physics-constrained PV power prediction via a hybrid framework Integrating secondary decomposition and improved Transformer-LSTM. *Energy* **2026**, *347*, 140314. [[CrossRef](#)]
22. Li, Y.; Wang, Z.; Liu, S. Enhance carbon emission prediction using bidirectional long short-term memory model based on text-based and data-driven multimodal information fusion. *J. Clean. Prod.* **2024**, *471*, 143301. [[CrossRef](#)]
23. Qiu, X.; Yu, Z.; Hu, Y. Zero-shot compound fault diagnosis method driven by semantic adaptive fusion. *Meas. Sci. Technol.* **2026**, *37*, 086105. [[CrossRef](#)]
24. Hu, X.; Jiang, C.; Huang, Y.; Peng, D.; Su, H.; He, Y.; Chen, Z. SMNet: A Novel Compositional Generalization Model for Industrial Robot Multi-Joint Fault Diagnosis. *IEEE Internet Things J.* **2026**, *13*, 13005–13018. [[CrossRef](#)]
25. Xu, Y.; He, D.; Sun, H.; Jin, Z.; Zhao, M. Self-supervised learning for train bearing fault diagnosis based on time-frequency dual domain prediction. *Struct. Health Monit.* **2026**, 1–24. [[CrossRef](#)]
26. He, D.; Xu, Y.; Sun, H.; Jin, Z.; Zhao, M. Self-supervised learning for vehicle bearing fault diagnosis based on time-frequency dual-domain contrast and fusion. *Nonlinear Dyn.* **2025**, *113*, 17385–17412. [[CrossRef](#)]
27. Sun, B.; Wang, D.; Yang, B.; Zhang, S.H.; Kong, L.Y. Prediction Model for the Boiler NOx Emission with Material Properties Based on the Artificial Neural Network. *Adv. Mater. Res.* **2013**, *2342*, 40–45. [[CrossRef](#)]
28. Yang, T.; Ma, K.; Lv, Y.; Bai, Y. Real-time dynamic prediction model of NOx emission of coal-fired boilers under variable load conditions. *Fuel* **2020**, *274*, 117811. [[CrossRef](#)]
29. Kang, J.; Niu, Y.; Hu, B.; Li, H.; Zhou, Z. Dynamic modeling of SCR denitration systems in coal-fired power plants based on a bi-directional long short-term memory method. *Process Saf. Environ. Prot.* **2021**, *148*, 867–878. [[CrossRef](#)]
30. Xie, P.; Gao, M.; Zhang, H.; Niu, Y.; Wang, X. Dynamic modeling for NOx emission sequence prediction of SCR system outlet based on sequence to sequence long short-term memory network. *Energy* **2020**, *190*, 116482–116482. [[CrossRef](#)]
31. He, W.; Li, J.; Tang, Z.; Wu, B.; Luan, H.; Chen, C.; Liang, H. A Novel Hybrid CNN-LSTM Scheme for Nitrogen Oxide Emission Prediction in FCC Unit. *Math. Probl. Eng.* **2020**, *2020*, 8071810. [[CrossRef](#)]
32. Zhong, C.; Dai, N.; Ouyang, F. A Data-driven Model to Reduce Outlet NOx Concentration of SCR System in FCC Unit. *China Pet. Process. Petrochem. Technol.* **2024**, *26*, 136–146.
33. Zhang, Y.; Chen, B.; Pan, G.; Zhao, Y. A novel hybrid model based on VMD-WT and PCA-BP-RBF neural network for short-term wind speed forecasting. *Energy Convers. Manag.* **2019**, *195*, 180–197. [[CrossRef](#)]
34. Xiao, H.; Huang, G.; Xiong, G.; Jiang, W.; Dai, H. A NOx emission prediction hybrid method based on boiler data feature subset selection. *World Wide Web* **2023**, *26*, 1811–1825. [[CrossRef](#)]
35. Hewamalage, H.; Bergmeir, C.; Bandara, K. Recurrent Neural Networks for Time Series Forecasting: Current status and future directions. *Int. J. Forecast.* **2021**, *37*, 388–427. [[CrossRef](#)]

36. Yan, J.; Mu, L.; Wang, L.; Ranjan, R.; Zomaya, A.Y. Temporal Convolutional Networks for the Advance Prediction of ENSO. *Sci. Rep.* **2020**, *10*, 8055. [[CrossRef](#)] [[PubMed](#)]
37. Hu, J.; Shen, L.; Albanie, S.; Sun, G.; Wu, E. Squeeze-and-Excitation Networks. *IEEE Trans. Pattern Anal. Mach. Intell.* **2020**, *42*, 2011–2023. [[CrossRef](#)] [[PubMed](#)]
38. Jin, X.; Xie, Y.; Wei, X.S.; Zhao, B.R.; Chen, Z.M.; Tan, X. Delving deep into spatial pooling for squeeze-and-excitation networks. *Pattern Recognit.* **2022**, *121*, 108159. [[CrossRef](#)]
39. Greff, K.; Srivastava, R.K.; Koutnik, J.; Steunebrink, B.R.; Schmidhuber, J. LSTM: A Search Space Odyssey. *IEEE Trans. Neural Netw. Learn. Syst.* **2017**, *28*, 2222–2232. [[CrossRef](#)] [[PubMed](#)]
40. Ghasemi, M.; Zare, M.; Trojovský, P.; Rao, R.V.; Trojovská, E.; Kandasamy, V. Optimization based on the smart behavior of plants with its engineering applications: Ivy algorithm. *Knowl.-Based Syst.* **2024**, *295*, 111850. [[CrossRef](#)]
41. Chai, T.; Draxler, R.R. Root mean square error (RMSE) or mean absolute error (MAE)?-Arguments against avoiding RMSE in the literature. *Geosci. Model Dev.* **2014**, *7*, 1247–1250. [[CrossRef](#)]

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.