Revista Internacional de Métodos Numéricos para Cálculo y Diseño en Ingeniería

Tech Science Press

https://doi.org/10.32604/rimni.2026.087788

ARTICLE

Leakage-Controlled CEEMDAN RawFusion Forecasting for SCR Inlet NO_x in Municipal Solid Waste Incineration

Weixin Yu[1], Yongyi Xu[1], Lei Hu[1], Jianzhong Zhou[1], Weiwei Shen[1], Peng Zou[1], Ting'an Liu[1,*], Baocheng Zou[1] and Lu Dong[2,*]

[1]China Power Huachuang (Suzhou) Electric Power Technology Research Co., Ltd., Suzhou, China

[2]State Key Laboratory of Low Carbon Catalysis and Carbon Dioxide Utilization, School of Petroleum Engineering, Yangtze University, Wuhan, China

*Corresponding Authors: Ting'an Liu. Email: franklta@163.com; Lu Dong. Email: ludong@yangtzeu.edu.cn

Received: 23 June 2026; Accepted: 28 July 2026; Published: 06 November 2026

ABSTRACT: Decomposition-assisted forecasting is widely used for non-stationary industrial signals, but its reported accuracy can be inflated when decomposition is fitted to a full sequence or to a validation/test block. In that setting, future observations enter the features assigned to earlier prediction samples. This study examines this leakage problem for selective catalytic reduction (SCR) inlet NO_x forecasting in municipal solid waste incineration and evaluates a sample-wise causal CEEMDAN-RawFusion-Attn-InformerLite predictor. Each CEEMDAN representation is generated only from the historical window available at the prediction time, and five-seed averages are used as the primary evidence. Full-series and block-wise CEEMDAN produced apparently high R^2 values of 0.9798 and 0.9745, whereas the leakage-controlled protocol gave R^2 = 0.9357 and RMSE = 6.1210 for the representative proposed run. Across five seeds, the proposed predictor had a slightly lower mean accuracy than BiLSTM (R^2 lower by 0.0011) while using 70.3% fewer trainable parameters. The results indicate that leakage control is essential when using decomposition-based industrial forecasting, and that compact predictors should be evaluated together with decomposition cost before plant-side deployment.

KEYWORDS: NO_x; forecasting; CEEMDAN; RawFusion; leakage-controlled evaluation; InformerLite; municipal solid waste incineration

1 Introduction

Minute-ahead forecasting of SCR inlet NO_x concentration is directly relevant to ammonia injection in flue-gas denitrification systems [1]. In MSW incineration plants, NO_x generation is affected by refuse heterogeneity, combustion-air distribution, furnace thermal state, and flue-gas transport delay. These factors make the sequence non-stationary and locally irregular, which is why data-driven soft sensors have become a practical alternative to direct measurement in the process industry [2]. For operation, the control system needs a timely estimate of whether inlet NO_x is about to rise or fall.

Short-horizon prediction can support ammonia-flow adjustment before a NO_x rise reaches the catalyst inlet. It can also help avoid unnecessary ammonia injection when the inlet concentration is about to decline. A plant-side model should therefore combine forecasting accuracy, causal validity, and feasible computation.

Deep learning methods have been increasingly adopted for industrial soft sensing and sequential prediction. LSTM and bidirectional recurrent networks remain attractive because they can model nonlinear

Copyright © 2027 The Authors. Published by Tech Science Press. This work is licensed under a Creative Commons Attribution 4.0 International License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

temporal dependence and historical context [3,4]. Attention-enhanced recurrent models further improve the selective use of relevant timesteps [5], while Transformer-family architectures summarize temporal interactions in parallel and have become an important alternative for time-series forecasting [6–8].

For highly non-stationary industrial series, decomposition-assisted modeling is often used to reduce scale mixing before prediction. The empirical mode decomposition family, VMD, and CEEMDAN separate the raw signal into components with different characteristic scales [9–12]. A recent VMD-SE-transformer-BiLSTM load-forecasting study also reflects the broader use of VMD with attention and recurrent architectures in short-term industrial prediction [13]. Related NO_x forecasting studies have used decomposition-assisted LSTM, LSTM-Transformer, and graph-based models [14–16]. For MSW incineration specifically, several model families have been reported. Modular and brain-inspired networks decompose the process into cooperating sub-models [17,18], while a denoising-plus-bidirectional-LSTM soft sensor targets the noisy emission signal directly [19]. Dynamic modular, fuzzy, stochastic-configuration, and multitask formulations have also been proposed to handle process nonlinearity and multiple operating regimes [20–24], building on earlier radial-basis and NARX neural predictors [25,26]. These studies show that data-driven NO_x prediction is feasible, but reported performance depends strongly on sequence representation and evaluation protocol. Crucially, most of these decomposition-assisted models apply the decomposition to the full series or to an entire validation or test block, and report accuracy without verifying that every feature could be generated in true real time. Whether decomposition-stage leakage materially inflates the reported accuracy of industrial NO_x forecasters, and how large that inflation is relative to the choice of neural architecture, has therefore not been established under a strictly causal protocol.

Among decomposition methods, EMD is fully adaptive but is vulnerable to mode mixing; EEMD mitigates mode mixing by adding ensemble noise but can leave reconstruction noise; CEEMDAN further averages adaptively extracted components and keeps a more complete reconstruction; ICEEMDAN can reduce residual noise further but increases implementation complexity; and standard VMD formulates mode extraction as a variational problem that requires preselecting the number of modes and penalty settings [9–12]. This limitation should be understood as applying to the standard VMD formulation used in the present benchmark; adaptive and successive VMD variants have been proposed to relax or automate the choice of mode number [27,28]. CEEMDAN was selected here because it is adaptive for nonlinear univariate NO_x signals, has a complete-reconstruction form, and can be executed independently on each historical window.

Recent forecasting studies continue to combine VMD, CEEMDAN, or related decomposition methods with recurrent, convolutional, Transformer, and attention-based predictors in industrial load, precipitation, air-quality, and power-load applications [13,29–31]. These applications support the practical value of decomposition for non-stationary sequences, but they also make leakage control more important because the decomposition operator itself can use information from outside the forecasting history if the evaluation protocol is not designed causally.

A key methodological risk in decomposition-assisted industrial forecasting is future-information leakage during feature construction. When CEEMDAN, EMD, or VMD is applied to an entire validation or test sequence, the representation of an earlier sample can depend on observations that would not be available at the prediction time. This look-ahead bias was first analysed systematically for wavelet-based hydrological forecasting, where decomposing the full series before splitting was shown to inflate reported skill and to break down in genuine real-time operation [32]. The same mechanism applies to EMD-, VMD-, and CEEMDAN-assisted models. This practice can inflate test accuracy while violating the online forecasting setting, and the problem is especially important for short-horizon NO_x prediction because high autocorrelation can already make simple baselines competitive. Chronological and rolling-origin evaluation schemes are also

recommended for time-series model assessment because random or non-causal validation can overestimate forecasting skill [33,34].

We therefore treat CEEMDAN as a causal feature operator rather than as a full-series preprocessing step. The study focuses on three linked questions: how much decomposition-stage leakage inflates apparent forecasting accuracy, whether raw-enhanced mode fusion and a compact attention predictor remain competitive under a causal protocol, and what accuracy-complexity trade-off is obtained when decomposition cost is reported separately from neural-network inference.

We identify the leakage problem, introduce the sample-wise causal decomposition and raw-fusion predictor, and quantify the effect of decomposition protocol, component design, seed-level variability, parameter sensitivity, and runtime cost.

2 Materials and Methods

2.1 Dataset Construction and Preprocessing

The experiments used a univariate SCR inlet NO_x sequence recorded at one MSW incineration plant over January and February. The raw plant record is already a continuous 1 min series with no missing values and no timestamp gaps, so it was used directly without deletion, imputation, or removal of any samples.

Table 1 summarizes the final dataset used for all reported experiments. The final modelling sequence contains 47,312 one-minute observations from 2026-01-04 00:00:00 to 2026-02-05 20:31:00, with zero duplicated timestamps and zero missing NO_x values. The data were split chronologically into training, validation, and test segments before normalization, model fitting, or metric calculation.

Table 1: Jan-Feb SCR inlet NO_x dataset and chronological data split.

Item	Value
Total rows	47,312
Sampling interval	1 min
Time span	2026-01-04 00:00:00 to 2026-02-05 20:31:00
Mean NO_x	98.1678
Standard deviation	28.1679
Minimum/maximum	0.1267/398.5633
Training rows	33,118
Validation rows	7096
Test rows	7098
Duplicate timestamps	0
Missing NO_x values	0

The only preprocessing applied to the NO_x values was min-max normalization. To respect the causal setting, the normalization was fitted on the training subset only and then reused, without refitting, on the validation and test subsets, so that preprocessing statistics were not contaminated by future information. No outlier removal, clipping, smoothing, or other value-altering operation was applied to the sequence.

Although the dataset is univariate, SCR inlet NO_x is not an isolated sensor trace. It reflects refuse-feed fluctuation, staged-air distribution, combustion intensity, flue-gas transport delay, and control carryover. Short-lived disturbances and slower operating-state drift are therefore mixed in the same signal, motivating the joint use of local raw history and multi-scale decomposition.

Fig. 1 summarizes the sample-wise causal prediction procedure, including the continuous NO_x input, CEEMDAN decomposition on historical windows, raw-enhanced mode fusion, compact attention-based forecasting, and leakage-controlled evaluation. Fig. 2 shows the Jan-Feb NO_x sequence, the chronological split used for model development, and a representative dynamic interval from the test segment.

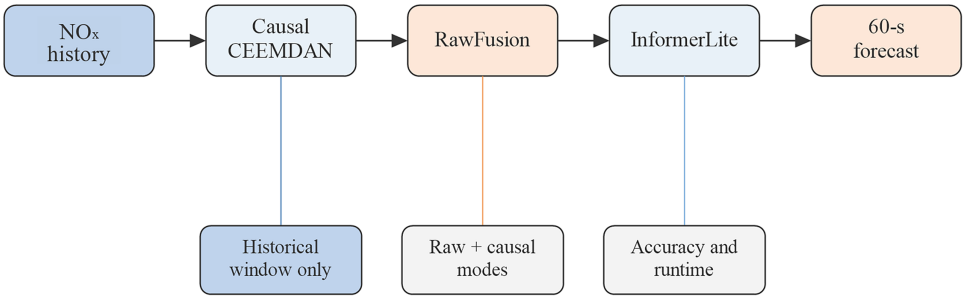


Figure 1: Sample-wise causal CEEMDAN-RawFusion forecasting workflow for 60 s SCR inlet NO_x prediction. The upper row shows the online inference path from the historical NO_x window through causal CEEMDAN, RawFusion and InformerLite to the forecast; the lower boxes summarize the historical-window constraint, raw-mode fusion input and evaluation outputs.

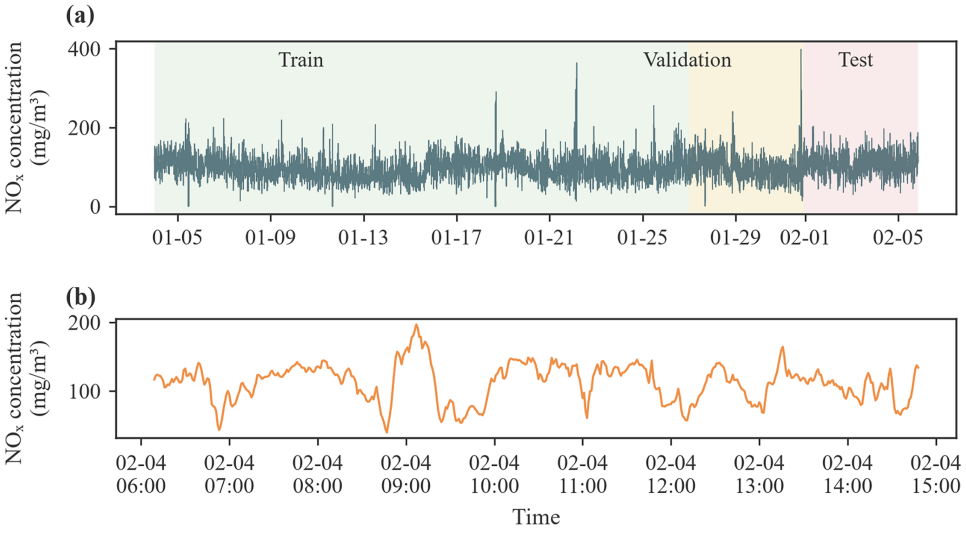


Figure 2: SCR inlet NO_x sequence and representative test-set variation. (**a**) Chronological January–February NO_x sequence with training, validation and test partitions; (**b**) Enlarged high-variation interval selected from the test subset. **NO_x concentration is reported in mg/m^3.**

2.2 Sample-Wise Causal CEEMDAN and Raw-Enhanced Mode Fusion

Let $S = \{y_1, y_2, \ldots, y_T\}$ denote the chronological SCR inlet NO_x sequence and let h be the forecasting horizon. For a prediction issued at time t, the target is y_{t+h}. The admissible information set is $\mathcal{F}_t = \sigma(y_1, \ldots, y_t)$, that is, the σ-field generated by observations no later than time t. A forecasting feature X_t is leakage-free only if it is measurable with respect to $\mathcal{F}_t$, i.e., determined only by observations available up to time *t*.

For each t, a historical decomposition window $H_t = [y_{t-D+1}, \ldots, y_t]$ is extracted. The CEEMDAN operator with parameter vector θ = (N_{trial}, $K_{\max}$, ε) is written as Eq. (1), where $C_t^{(k)}$ is the k-th IMF sequence

and R_t is the residual sequence. In the main experiment we used $N_{\text{trial}} = 8$, $K_{\max} = 6$, $\varepsilon = 0.005$, and a normal CEEMDAN noise scale of 1.0 as reproducible computational settings. These are empirical choices rather than theoretical optima, and their influence is examined through the sensitivity screen in Section 3.7.

In implementation, CEEMDAN was applied independently to each historical window. For each window, multiple noisy replicas were generated with a fixed random seed, empirical-mode extraction was applied to each replica, the corresponding modes were averaged across the ensemble, and the residual was updated after each extracted component. The retained channels therefore consisted of the averaged IMFs and the final residual produced from that historical window only.

$$\mathcal{D}_\theta(H_t) = \{C_t^{(1)}, \ldots, C_t^{(K_t)}, R_t\} \tag{1}$$

The decomposed representation is normalized to a fixed channel number by keeping at most $K_{\max}$ IMFs and the residual. If fewer than $K_{\max}$ IMFs are produced, the missing channels are zero-padded after decomposition; if more are produced, only the first $K_{\max}$ IMFs and the residual are retained. The last L samples of these channels form $M_t \in \mathbb{R}\text{L} \times (K_{\max} + 1)$.

The representative decomposition in Fig. 3b illustrates the resulting scale separation. The high-frequency modes capture rapid minute-level fluctuation, whereas the lower-frequency modes and residual track slower operating drift. This figure is descriptive and is not used as a separate training or validation result.

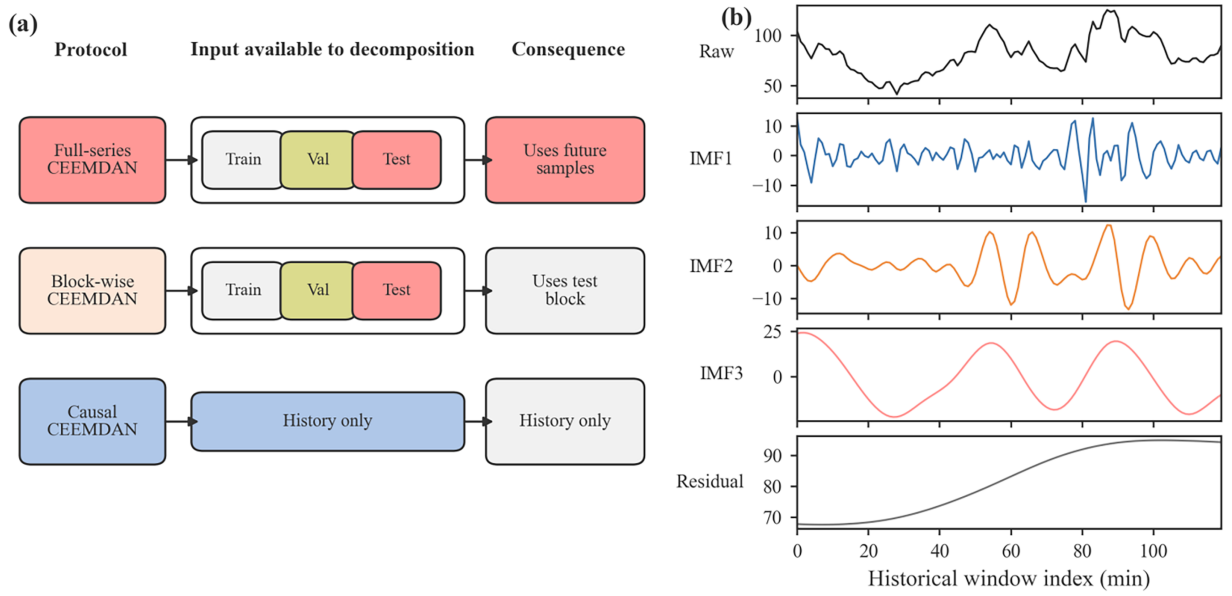


Figure 3: CEEMDAN decomposition protocols and representative sample-wise causal modes. (**a**) Comparison of full-series, block-wise and sample-wise causal CEEMDAN protocols; (**b**) Raw signal, IMFs and residual generated from a representative D = 120 min historical window.

Raw-enhanced mode fusion is then defined as Eq. (2), where $q_t = [y_{t-L+1}, \ldots, y_t]^\top$ is the raw NO_x tail. This differs from standard mode-only fusion because the raw level channel is preserved explicitly instead of relying solely on reconstructed or decomposed modes.

$$X_t = \text{concat}(M_t, q_t), X_t \in \mathbb{R}^{L\times(K_{\max}+2)} \tag{2}$$

The compact attention predictor is written as Eq. (3). The input is first projected to $X_t W_p + P$ to give Z_0. The encoder block produces $Z = \text{Encoder}_\varphi(Z_0)$; a distilling operation then compresses the local sequence, attention pooling computes $\alpha_i = \text{softmax}(v^\top \tanh(Wz_i))$, and $g = \Sigma_i \alpha_i z_i$ is mapped to the forecast by an MLP, $\hat{Y}_{t+h} = \text{MLP}(g)$.

$$\hat{Y}_{t+h} = f_\varphi(X_t) \tag{3}$$

Proposition (Causality). Under fixed random seed and CEEMDAN noise settings, the feature X_t is measurable with respect to $\mathcal{F}_t$, so prediction of y_{t+h} uses no observation later than t.

Proof. The window $H_t = [y_{t-D+1}, \ldots, y_t]$ depends only on observations up to time t. With the seed and noise settings fixed, the CEEMDAN output $\mathcal{D}_\theta(H_t)$ is a deterministic function of H_t, and the raw tail q_t is likewise a function of observations up to t. Hence $X_t = \text{concat}(M_t, q_t)$ is measurable with respect to $\mathcal{F}_t$. At inference the predictor f_φ receives only X_t, so no observation y_τ with $\tau > t$ can enter feature generation or the forecast of y_{t+h}.

The sample-wise computation is represented as the flowchart in Fig. 3a. At each target time, the algorithm reads only the historical window, applies CEEMDAN to that window, retains the last L samples of the selected IMFs and residual, appends the raw NO_x tail, and sends the fused sequence to the predictor. The procedure then advances to the next target time and repeats the decomposition independently, so decomposed features are not shared from future target times.

Because the decomposition is recomputed for every prediction sample, decomposed features are never shared across future target times.

This decomposition maps naturally onto the engineering setting. High-frequency IMFs track rapid combustion disturbances and short-term operational perturbations, while the lower-frequency modes and the residual describe slower drift, and the appended raw channel anchors the current SCR inlet NO_x level. That is why decomposition is paired with raw-level preservation instead of replacing the original signal.

Fig. 3 combines the protocol definition with a representative decomposition result. Full-series and block-wise decomposition create leakage-biased forecasting evaluations because future samples participate in the features assigned to earlier validation or test targets. In the sample-wise causal protocol, only the history available at the target time enters CEEMDAN.

2.3 Attention-Weighted InformerLite Predictor

The predictor receives the fused sequence X_t defined above and maps it to $\hat{Y}_{t+h}$. First, X_t is projected to width d_{model} and combined with a positional encoding P:

$$Z_0 = X_t W_p + P \tag{4}$$

The projected sequence is processed by a compact Transformer encoder block with multi-head self-attention and a feedforward sublayer. A distilling block composed of Conv1D, an exponential linear unit (ELU) activation [35], and MaxPool1D then shortens the token sequence:

$$Z_d = \text{MaxPool}(\text{ELU}(\text{Conv1D}(\text{Encoder}(Z_0)))) \tag{5}$$

Positional encoding in Eq. (4) preserves token order, and the self-attention encoder models position-dependent temporal interactions. After encoding and distillation, content-based additive attention pooling is applied to the distilled tokens z_i. The score in Eq. (6) contains no explicit positional or recency term;

it weights the tokens according to their learned representations and aggregates them into the fixed-length vector g:

$$\alpha_i = \frac{\exp(v^\top \tanh(Wz_i))}{\Sigma_j \exp(v^\top \tanh(Wz_j))} \tag{6}$$

$$g = \Sigma_i \alpha_i z_i \tag{7}$$

$$\hat{Y}_{t+h} = \mathrm{MLP}(g) \tag{8}$$

Eqs. (1) and (2) define the CEEMDAN decomposition and RawFusion construction of X_t. Eqs. (4)–(8) then define the predictor applied to X_t: positional encoding preserves token order, self-attention models position-dependent temporal interactions, the distilling block shortens the encoded sequence, content-based attention pooling forms the fixed-length representation g, and the MLP maps g to the future target.

The architecture is kept compact because the task is short-horizon and univariate. The model does not reproduce the full encoder-decoder Informer; it retains self-attention and distillation in a lower-parameter predictor.

In this architecture, each stage serves a distinct purpose. Sample-wise causal CEEMDAN separates historical disturbances by scale; RawFusion retains the current operating level that may otherwise be distributed across decomposed modes; positional encoding and self-attention model position-dependent temporal interactions; the distilling block shortens the encoded sequence; and content-based attention pooling aggregates the distilled tokens into a fixed representation for single-step forecasting. The contribution of each stage is interpreted through the causal evaluation results reported below.

2.4 Baselines, Hyperparameter Setting, and Component-Isolation Design

The benchmark was organized to compare raw-sequence predictors, decomposition-assisted predictors, and low-complexity autoregressive references. The raw-sequence family contains LSTM, BiLSTM, GRU, BiGRU, TCN, Transformer and InformerLite, together with persistence and Ridge-AR-30 references. The decomposition family uses the same RawFusion-Attn-InformerLite predictor with CEEMDAN, EMD, EEMD and VMD features. This design separates three questions: the value of nonlinear neural prediction at the 60 s horizon, the effect of decomposition relative to raw short-horizon baselines, and the sensitivity of the compact attention predictor to the decomposition method.

All stochastic neural models were evaluated with five seeds under the same chronological split, 60 s forecasting horizon and training-only preprocessing. The deterministic persistence and Ridge-AR-30 baselines were retained as low-complexity references. The decomposition comparisons used EMD, EEMD, CEEMDAN and VMD, which were available in the verified software environment and implemented under the same leakage-controlled feature-generation protocol.

The recurrent baselines were selected from a validation-based candidate pool, with hidden sizes from 128 to 256, one or two recurrent layers, dropout of 0.02–0.05, and learning rates from 5×10^{-4} to 1×10^{-3}. The temporal and CEEMDAN settings of the proposed model were $h = 1$ min, $D = 120$ min, $L = 12$ min, $N_{\mathrm{trial}} = 8$, $K_{\max} = 6$, and $\varepsilon = 0.005$. The neural settings were $d_{\mathrm{model}} = 160$, $n_{\mathrm{head}} = 5$, one encoder layer, a feedforward expansion factor of 3, dropout = 0.02, learning rate = 4×10^{-4}, batch size = 512, up to 48 epochs, patience = 10, and seed = 42 for the representative single run. The five-seed experiments used seeds 40–44 where applicable. The optimizer was Adam [36], and the proposed neural predictor contained 375,201 trainable parameters.

Model complexity was evaluated by trainable parameter count, float 32 weight footprint, predictor-only inference time, CEEMDAN feature-generation time, and estimated end-to-end time, following established practice for reporting the efficiency of deep neural networks [37]. This separation is important because the neural predictor can be compact while the causal CEEMDAN feature-generation step still contributes non-negligible latency.

For stochastic training, component-isolation neural models were evaluated across five random seeds where applicable. Paired seed-level tests were then used to determine whether small metric gaps were larger than seed-level variation. Persistence and Ridge-AR-30 were deterministic or convex linear baselines and therefore have no seed-level standard deviation.

2.5 Evaluation Metrics

Prediction accuracy was evaluated using the coefficient of determination (R^2), root mean square error (RMSE), mean absolute error (MAE), and mean absolute percentage error (MAPE). Higher R^2 indicates better goodness-of-fit, whereas smaller RMSE, MAE, and MAPE indicate lower prediction error. These metrics remain standard in forecasting studies, although MAPE requires careful interpretation near zero values [38]. For numerical stability, MAPE was computed with the absolute denominator clipped at $\varepsilon = 1 \times 10^{-6}$, consistent with the evaluation code used to generate all reported results. Because the NO_x sequence occasionally approaches very low NO_x levels (minimum 0.1267 in the reported NO_x units), MAPE can be disproportionately affected by a small number of near-zero samples; it is therefore reported only as a secondary indicator, and RMSE, MAE, and R^2 are used as the primary basis for model comparison.

The experiments were implemented in Python 3.10.11 using NumPy 2.2.6, pandas 2.3.3, scikit-learn 1.6.1, PyTorch 2.10.0 + cu130, EMD-signal 1.9.0 (imported as PyEMD), vmdpy 0.2 and matplotlib 3.10.8. Neural-network timing was measured on an NVIDIA GeForce RTX 5090 GPU.

3 Results and Discussion

3.1 Data Continuity and Forecasting Task

Figs. 1 and 2 summarize the sample-wise causal prediction procedure and the Jan-Feb NO_x sequence used for model development. The upper panel of Fig. 2 shows that the sequence is continuous over the full Jan-Feb span, with no missing values or timestamp gaps. The lower panel shows a dynamic 180 min interval from the test segment, selected by the largest rolling standard deviation. Several abrupt rises and drops are visible in this interval, which motivates testing whether multi-scale representations remain useful under a strict 1 min-ahead setting.

Sequence continuity matters for both decomposition and evaluation validity: if missing-label intervals or artificial discontinuities were present, both CEEMDAN and sequence modeling would be distorted. For the present record this condition is met by the data themselves, because the Jan-Feb series is already gap-free and evenly sampled at 1 min, so no gap filling or timeline repair was required.

The dynamic interval in Fig. 2 also shows why the task is not a purely smooth autoregressive problem. Abrupt rises and drops play out over a few minutes, yet sit on top of a slower baseline drift. Such behaviour is typical of MSW incineration, where short-lived combustion disturbances and operating adjustments coincide with furnace and flue-gas thermal inertia.

Figs. 1 and 2 therefore define the data setting for the causal evaluation: fast local fluctuation is embedded within slower process evolution, but every feature used for prediction must still be generated without future observations.

3.2 Effect of Decomposition Protocol on Reported Accuracy

The decomposition protocol had a larger effect on the reported performance than the differences among many neural architectures. Full-series CEEMDAN produced R^2 = 0.9798 and RMSE = 3.4332, and block-wise CEEMDAN still produced R^2 = 0.9745 and RMSE = 3.8587. Under the valid sample-wise causal protocol, the same predictor obtained R^2 = 0.9357 and RMSE = 6.1210. This contrast quantifies the accuracy inflation caused by decomposition-stage leakage and supports sample-wise causality as a minimum requirement for decomposition-assisted forecasting; the numerical results are summarized in Table 2 and Fig. 4.

Table 2: Effect of CEEMDAN decomposition protocol on 1 min-ahead forecasting accuracy.

Protocol	R^2	RMSE	MAE	MAPE (%)	Parameters
Full-series CEEMDAN	0.9798	3.4332	2.4732	2.3884	375,201
Block-wise CEEMDAN	0.9745	3.8587	2.7810	2.7235	375,201
Sample-wise causal CEEMDAN	0.9357	6.1210	4.4405	4.2837	375,201

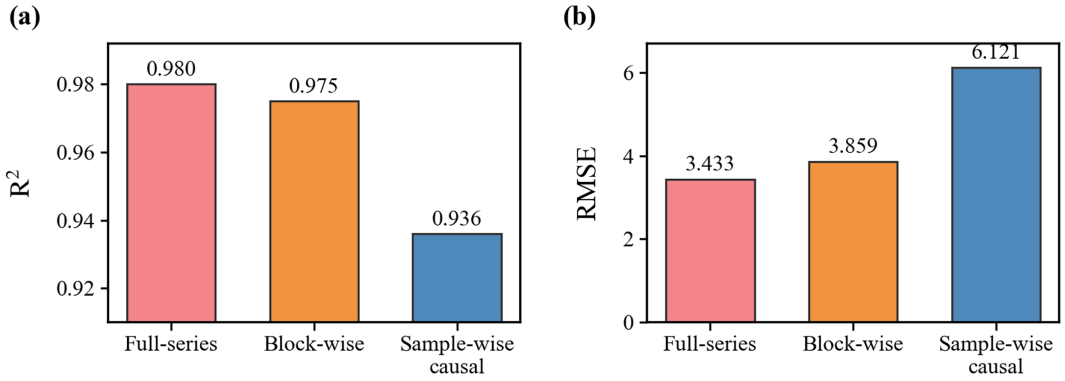


Figure 4: Forecasting accuracy under alternative CEEMDAN decomposition protocols. (**a**) R^2 values for full-series, block-wise and sample-wise causal decomposition; (**b**) Corresponding RMSE values.

These results set the conditions for all later comparisons. The full-series and block-wise scores correspond to diagnostic leakage-biased decompositions rather than deployable online forecasts. All subsequent model comparisons therefore use the sample-wise causal protocol.

3.3 Forecasting Performance under the Causal Protocol

The five-seed averages in Table 3 are treated as the primary evidence for model comparison. The expanded benchmark indicates strong short-horizon autoregression. Persistence was clearly less accurate, whereas Ridge-AR-30 reached R^2 = 0.9351 and RMSE = 6.1501, close to the best deep models. Among the raw neural baselines, BiLSTM gave the lowest mean RMSE (6.1342), followed by InformerLite (6.1389) and GRU (6.1481). The proposed decomposition-fusion predictor was therefore interpreted as competitive and parameter-efficient, not as the most accurate model on every metric.

The added GRU, TCN and Transformer-family baselines further clarify the benchmark pattern. GRU and InformerLite achieved RMSE values close to BiLSTM, TCN was less accurate on average, and the standard Transformer did not improve on the stronger recurrent baselines. This pattern is consistent with a 60 s univariate task in which the recent raw sequence carries most of the next-step predictive information.

Table 3: Expanded benchmark results under sample-wise causal evaluation.

Model	Runs	R^2	RMSE	MAE	MAPE (%)	Parameters	Feature Type
BiLSTM	5	0.9354 ± 0.0003	6.1342 ± 0.0121	4.4462	4.284	1,261,441	Raw
LSTM	5	0.9349 ± 0.0003	6.1608 ± 0.0124	4.4662	4.3027	464,833	Raw
GRU	5	0.9351 ± 0.0002	6.1481 ± 0.0114	4.4647	4.3023	353,281	Raw
BiGRU	5	0.9350 ± 0.0001	6.1564 ± 0.0047	4.467	4.3086	964,609	Raw
TCN	5	0.9314 ± 0.0025	6.3224 ± 0.1132	4.6011	4.4208	64,257	Raw
Transformer	5	0.9345 ± 0.0003	6.1792 ± 0.0149	4.4871	4.3209	271,201	Raw
InformerLite	5	0.9353 ± 0.0006	6.1389 ± 0.0282	4.4479	4.2911	348,161	Raw
CEEMDAN-RawFusion-Attn-InformerLite	5	0.9343 ± 0.0008	6.1873 ± 0.0391	4.4921	4.325	375,201	Decomposition + raw fusion
EMD-RawFusion-Attn-InformerLite	5	0.9348 ± 0.0005	6.1648 ± 0.0248	4.4749	4.3176	375,201	Decomposition + raw fusion
EEMD-RawFusion-Attn-InformerLite	5	0.9346 ± 0.0006	6.1733 ± 0.0265	4.4915	4.3273	375,201	Decomposition + raw fusion
VMD-RawFusion-Attn-InformerLite	5	0.9348 ± 0.0007	6.1640 ± 0.0308	4.4865	4.3291	375,201	Decomposition + raw fusion
Ridge-AR-30	1	0.9351	6.1501	4.4549	4.2889	31	Raw autoregressive
Persistence	1	0.9145	7.0602	5.0873	4.925	0	Raw autoregressive

Note: Neural models are reported as five-seed mean and standard deviation where applicable. Persistence and Ridge-AR-30 are deterministic references and therefore have no seed deviation.

The decomposition comparison gives a similar result. With the same RawFusion-Attn-InformerLite predictor, VMD, EMD, EEMD and CEEMDAN produced RMSE means of 6.1640, 6.1648, 6.1733 and 6.1873, respectively. The decomposition-assisted variants therefore remained close to the leading raw-sequence baselines under sample-wise causal evaluation. The main distinction among the decomposition methods was computational cost: CEEMDAN had the highest feature-generation time in the tested setting, whereas EMD and VMD provided lower-cost alternatives.

Fig. 5 compares the full test segment and a local highly dynamic interval using representative models from the structured benchmark. Across the entire test set, all representative models follow the overall NO_x trend; the differences are most visible near local turning regions.

The turning-point behaviour in Fig. 5 illustrates the local dynamics captured by the representative models. Local differences are visible near rapid reversals, but the five-seed statistics in Table 3 show that the leading models remain within a narrow accuracy range.

3.4 Accuracy-Complexity and Runtime Feasibility

Fig. 6 reports the accuracy-complexity comparison for a representative subset of the neural benchmark in Table 3. The R^2 values of the strongest neural models cluster tightly, whereas their parameter counts differ widely. Together with the Ridge-AR-30 result in Table 3, this result indicates that recent autoregressive information accounts for most of the 1 min-ahead predictability. Accuracy and complexity are therefore interpreted jointly in the following analysis.

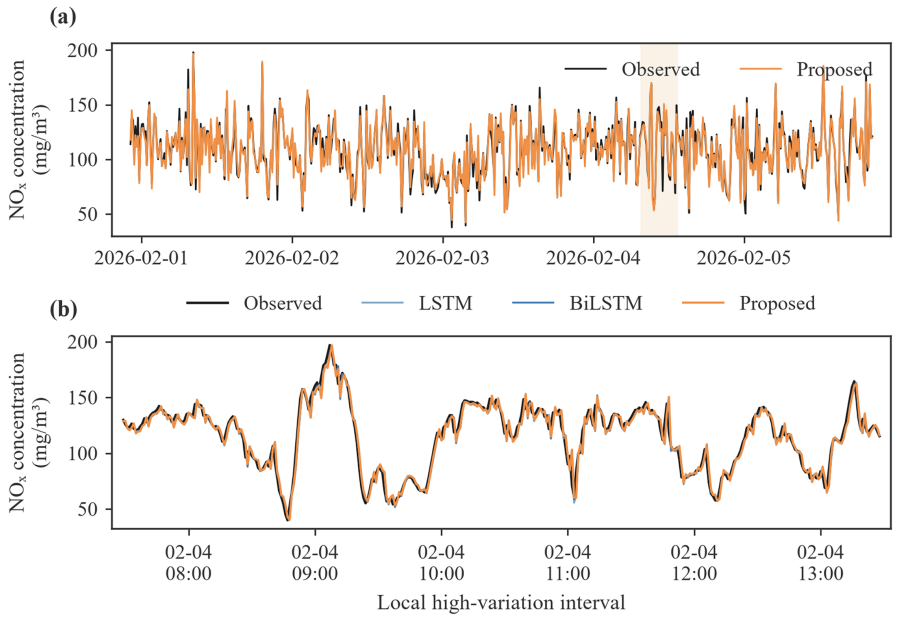


Figure 5: Test-set prediction curves for representative 1-min-ahead models. (**a**) Full test-set trajectories of observed NO_x and model predictions; (**b**) Enlarged high-variation interval showing local differences among representative models. **NO_x concentration is reported in mg/m^3.**

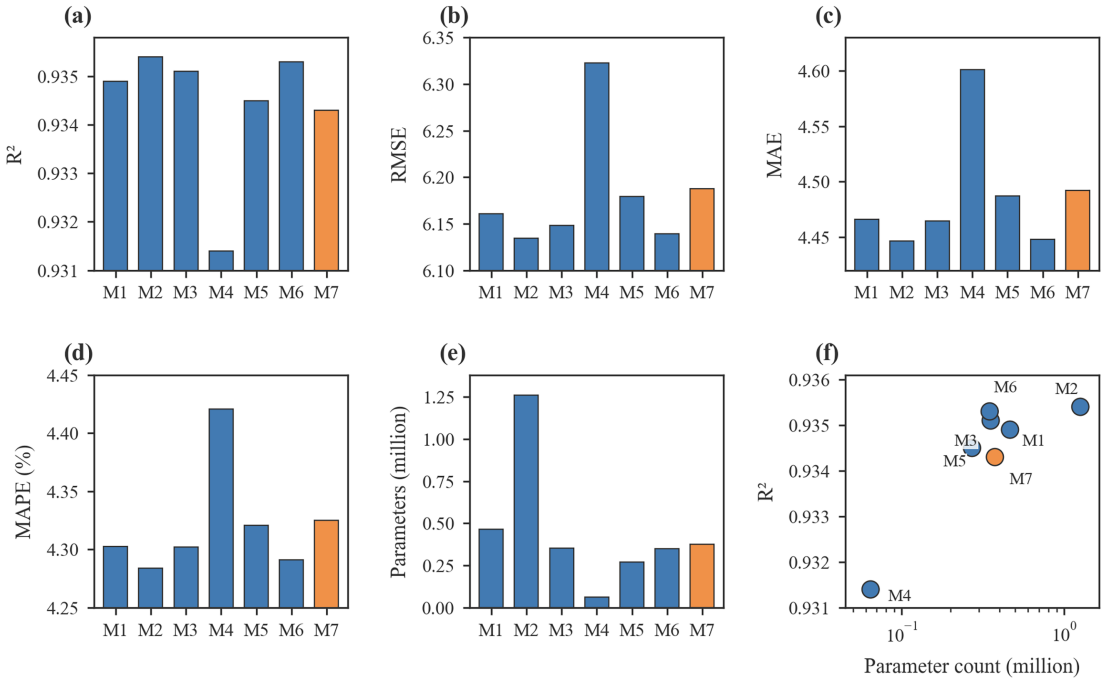


Figure 6: Accuracy and complexity comparison of selected neural forecasting models. (**a**) R^2; (**b**) RMSE; (**c**) MAE; (**d**) MAPE; (**e**) Trainable parameter count; (**f**) R^2-parameter trade-off. M1-M7 denote LSTM, BiLSTM, GRU, TCN, Transformer, InformerLite and CE-RF-Attn-InformerLite, respectively.

This trade-off has two levels. At the neural-predictor level, the proposed Attn-InformerLite predictor uses 375,201 trainable parameters, 70.26% fewer than the 1,261,441 parameters of BiLSTM, but it incurs a small reduction in five-seed mean accuracy (Table 3). Its predictor-only inference time is 0.0082 ms/sample, 63.6% lower than BiLSTM's 0.0225 ms/sample (Table 4). At the full-pipeline level, causal CEEMDAN feature generation raises the end-to-end runtime to 55.9954 ms/sample. The practical advantage is therefore predictor-level compactness and lower predictor-only latency, not lower end-to-end latency; the measured full-pipeline latency nevertheless remains well below the 60 s forecasting interval.

Table 4: Parameter count, model size, and runtime breakdown.

Model	Parameters	Weight (MB)	Predictor (ms/sample)	CEEMDAN (ms/sample)	End-to-End (ms/sample)	Device
BiLSTM	1,261,441	4.8120	0.0225	0.0000	0.0225	Cuda
CE-RF-Attn-InformerLite	375,201	1.4313	0.0082	55.9872	55.9954	Cuda

Note: Predictor-only latency and CEEMDAN feature-generation latency are reported separately to distinguish neural-predictor compactness from full-pipeline runtime. For 14,194 validation-plus-test samples, the estimated total times are 0.32 s for BiLSTM predictor inference, 0.12 s for proposed predictor-only inference, and 794.8 s for the proposed end-to-end CEEMDAN pipeline.

Recent boiler-intelligence work argues that a prediction model earns its place on site through tunability, optimization compatibility, and operational usefulness, not through accuracy alone [39,40]. Read against that benchmark, the proposed model fits cases where leakage-controlled decomposition and mode-level representation are valuable and a feature-generation cost near 56 ms is acceptable. For latency-first deployments, the raw BiLSTM and Ridge-AR-30 baselines provide lower-computation alternatives.

3.5 Control Use and Interpretation Limits

These results carry two practical implications for deployment. First, sample-wise causal decomposition should be treated as a basic requirement for decomposition-assisted industrial forecasting. Whole-segment decomposition can produce attractive but invalid test scores, whereas the causal protocol keeps feature generation within the data available at the forecast issuing time.

Second, compact attention-based encoders can be competitive for short-horizon industrial forecasting, but their advantage is horizon- and data-dependent. In this univariate one-step task, attention did not consistently improve the recurrent baselines, and Ridge-AR-30 was already close to the best neural models. The role of attention here is compact sequence weighting and compression within the causal-fusion predictor.

The present study stops short of that optimization loop. It contributes a forecasting component with a causally valid prediction protocol and moderate neural complexity, leaving the coupling to controller tuning and plant-side validation to later work.

For control use, the proposed method fits as an upper-layer advisory forecasting module. The horizon is 1 min and the measured end-to-end runtime stays well below this interval, so minute-level supervisory use is feasible. CEEMDAN feature generation is the main latency component, whereas raw BiLSTM and Ridge-AR-30 represent lower-computation references. Fig. 7 sketches the deployment: an edge-side model reads recent NO_x measurements, issues a short-horizon inlet-NO_x forecast, and passes it to the supervisory ammonia-injection controller.

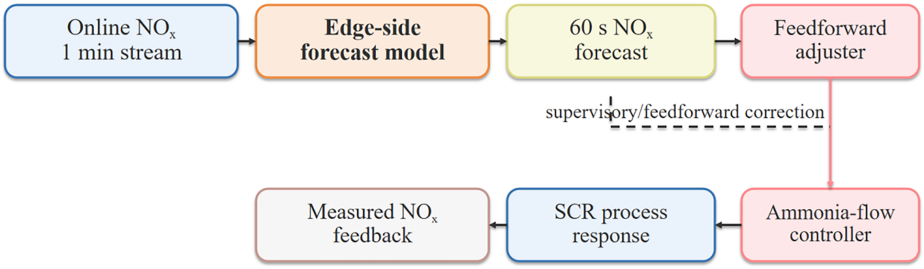


Figure 7: Supervisory edge-deployment concept for SCR ammonia-injection assistance. The upper path shows online NO_x acquisition, edge-side 60 s forecasting and feedforward adjustment; the lower path shows measured NO_x feedback, SCR process response and ammonia-flow controller interaction.

3.6 Multi-Seed Stability and Limitations

Single-run comparisons can be sensitive to random initialization, especially when candidates are close. Fig. 8 shows an accuracy-efficiency trade-off rather than accuracy parity or superiority. Across the five seeds, the proposed model has a lower median R2 (approximately 0.9341 vs. 0.9355) and a higher median RMSE (approximately 6.20 vs. 6.13) than BiLSTM; Table 3 likewise gives a five-seed mean R2 difference of 0.0011 (0.9343 vs. 0.9354). In return, its neural predictor reduces trainable parameters by 70.26% (1.26 million to 0.38 million) and predictor-only inference time by 63.6%, as reported in Table 4. The end-to-end runtime remains governed by causal CEEMDAN feature generation, so the efficiency claim is confined to model size and predictor-only inference.

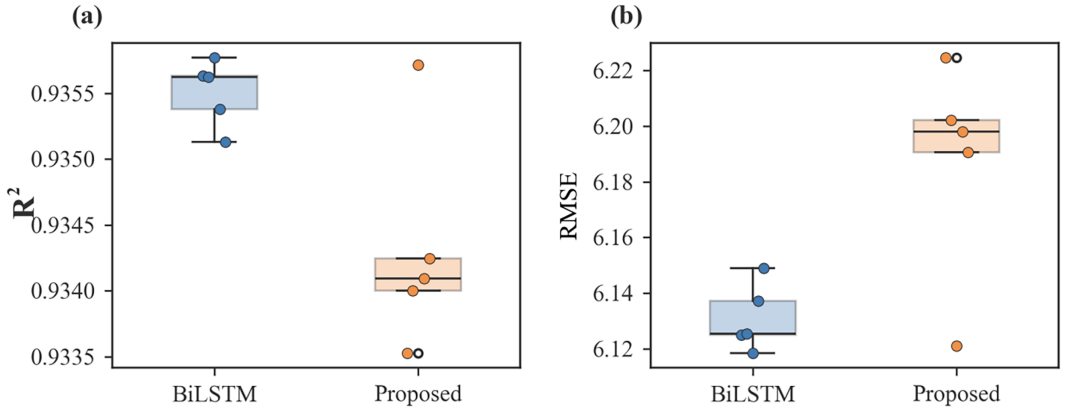


Figure 8: Five-seed prediction stability of BiLSTM and CEEMDAN-RawFusion-Attn-InformerLite. (**a**) R^2 distributions across five random seeds; (**b**) RMSE distributions across five random seeds. Filled markers denote individual seed-level runs; unfilled circular markers denote boxplot outliers.

In Fig. 8, each marker corresponds to one random seed. The box summarizes the interquartile range, the central line gives the median, and unfilled circles indicate observations outside the whisker range under the standard boxplot rule.

The experiments use a single Jan-Feb SCR inlet NO_x dataset from one industrial case. Rolling-origin evaluation improves temporal stability assessment within this sequence, while additional months, plants, pollutants and auxiliary process variables would extend the cross-condition evaluation. This scope is consistent with the main focus on leakage-controlled decomposition and short-horizon forecasting within the available plant record.

These results support a causality-valid forecasting protocol built around a compact predictor and benchmarked against explicit short-horizon references. Model value at the plant level depends on more than isolated accuracy: transparency, reproducibility, runtime feasibility, and compatibility with operational constraints all matter.

The study has four main limitations. First, the current dataset is univariate and does not include auxiliary process variables such as furnace load, oxygen concentration, temperature, ammonia flow or waste composition. Second, the experiments were conducted on one plant and one Jan-Feb segment, so cross-period and cross-plant behaviour requires additional datasets. Third, runtime was measured on a GPU workstation rather than on an industrial controller or edge gateway; field-side CPU load, memory footprint and deployment stability require separate testing. Fourth, the study evaluates predictive performance rather than hardware-in-the-loop or plant closed-loop control performance.

The reported R^2 values apply to this plant NO_x sequence, which was used without deletion, imputation, or clipping and with only training-set-fitted normalization; the accuracy is therefore attributable to the models and the causal protocol rather than to any data-cleaning step. Experiments on additional seasons, other pollutants, and multivariable sensor sets would further quantify robustness.

A natural next step is to couple the forecasting module to an optimization layer and a supervisory control platform, in line with recent combustion-intelligence work [39,40]. That step would test whether the leakage-controlled predictor can support plant-side ammonia-injection decisions under real control constraints.

3.7 CEEMDAN Parameter and Window Sensitivity

A separate sensitivity screen was conducted with a Ridge readout on sampled training and validation windows to limit the cost of CEEMDAN parameter screening. For each trial-history combination, Fig. 9 reports the lowest validation RMSE across the screened maximum-IMF settings and the corresponding feature-generation cost. The lowest sampled Ridge RMSE occurred at D = 240 min with trials = 4 and maximum IMF count = 4, but this setting also required more feature-generation time than D = 120 min. This result belongs only to the Ridge sensitivity screen. The main neural benchmark used D = 120 min throughout, including Fig. 3b, Table 3, and the runtime analysis.

The sensitivity result shows the expected trade-off: longer historical windows can slightly reduce Ridge-screen error, whereas feature-generation cost increases with history length and trial count. The CEEMDAN history length, trial count, and IMF cap should therefore be reported together with the causal protocol and runtime budget rather than treated as universally optimal constants.

3.8 Rolling-Origin Temporal Generalization

Rolling-origin validation, a standard out-of-sample testing scheme for forecasting models [41], was used to probe stability beyond a single fixed chronological split within the Jan-Feb sequence. Across four folds, CEEMDAN-RawFusion-Attn-InformerLite reached mean R^2 = 0.9420 and RMSE = 6.5273, against mean R^2 = 0.9407 and RMSE = 6.6535 for BiLSTM. The two models alternated in local ranking across folds, indicating similar temporal stability on this sequence. Additional seasonal, plant-level and pollutant-level datasets would further test the transferability of this pattern.

3.9 Complexity and Runtime

Fig. 10 and Table 4 summarize this distinction. The complexity analysis separates neural-predictor compactness from end-to-end computational cost. The InformerLite predictor is smaller and faster than BiLSTM in predictor-only inference (0.0082 vs. 0.0225 ms/sample), whereas causal CEEMDAN feature

generation dominates the full pipeline at approximately 55.99 ms/sample. For the combined validation and test sets (14,194 samples), the estimated predictor-only times are 0.32 s for BiLSTM and 0.12 s for the proposed predictor, while the full proposed pipeline including CEEMDAN feature generation is approximately 794.8 s. The present implementation is therefore best described as parameter-efficient at the predictor level, with end-to-end runtime mainly governed by the causal decomposition step.

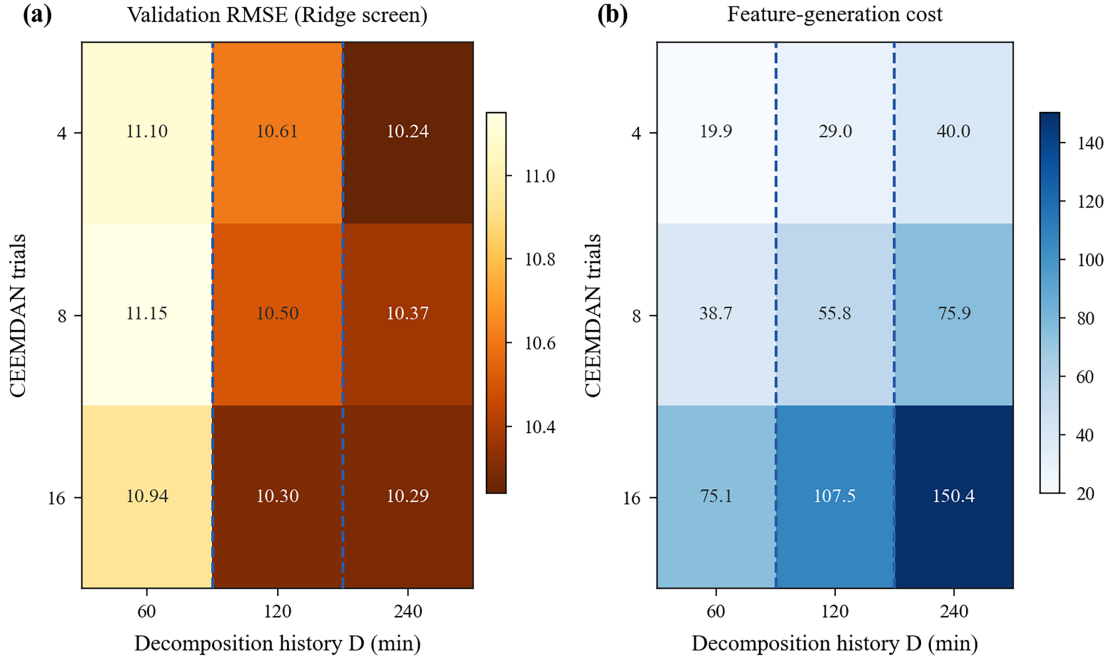


Figure 9: CEEMDAN trial and historical-window sensitivity screened with Ridge readout. (**a**) Validation RMSE across sampled trial-history settings; (**b**) Corresponding feature-generation cost. Dashed lines bracket D = 120 min used in the main neural benchmark; D = 240 min is retained only as a sensitivity condition.

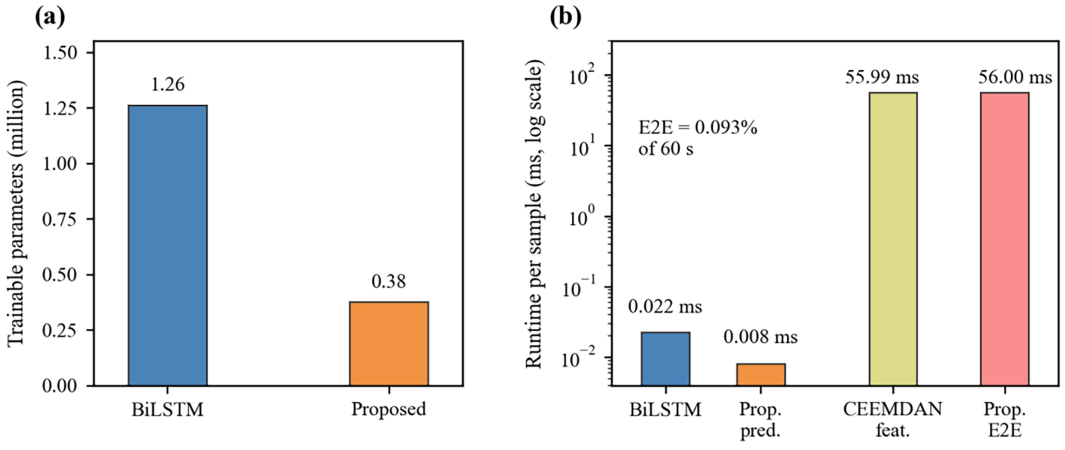


Figure 10: Predictor compactness and end-to-end runtime components. (**a**) Trainable parameter counts of BiLSTM and the proposed predictor; (**b**) Predictor-only inference time, CEEMDAN feature-generation time and end-to-end runtime per sample. Prop. pred. denotes the proposed neural predictor; Prop. E2E denotes the full proposed pipeline.

4 Conclusion

This study presented a leakage-controlled decomposition forecasting protocol for short-horizon SCR inlet NO_x prediction in MSW incineration. The main methodological result is that each decomposition feature must be generated only from information available at the prediction time; otherwise, full-series or block-wise decomposition can substantially inflate reported accuracy.

The expanded benchmark with GRU, TCN, Transformer-family models and EMD/EEMD/VMD/CEEMDAN comparisons shows that the strongest mean accuracy is obtained by raw short-horizon autoregressive models. CEEMDAN-RawFusion-Attn-InformerLite remains close to these leading baselines while using fewer neural-predictor parameters than BiLSTM. The end-to-end cost is mainly controlled by CEEMDAN feature generation, which motivates future work on incremental or cached causal decomposition.

Future work should extend the evaluation to additional operating periods, plants and pollutants. In parallel, incremental or cached causal decomposition could reduce feature-generation cost, widen the margin between end-to-end runtime and the forecasting interval, and ease deployment on plant-side hardware.

Acknowledgement: Not applicable.

Funding Statement: This work was supported by the National Natural Science Foundation of China (grant number 52560021).

Author Contributions: Weixin Yu, Yongyi Xu, Lei Hu, Jianzhong Zhou, Weiwei Shen, Peng Zou, Ting'an Liu, Baocheng Zou and Lu Dong contributed to conceptualisation, methodology, investigation, data curation, formal analysis, visualisation and manuscript drafting. Ting'an Liu and Lu Dong contributed to supervision, review and editing. All authors reviewed and approved the final version of the manuscript.

Availability of Data and Materials: The processed data and analysis scripts supporting the findings of this study are available from the corresponding authors upon reasonable request, subject to plant data-governance restrictions.

Ethics Approval: Not applicable.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Lin X, Yan M, Dai A, Zhan M, Fu J, Li X, et al. Simultaneous suppression of PCDD/F and NO_x during municipal solid waste incineration. Chemosphere. 2015;126:60–6. doi:10.1016/j.chemosphere.2015.02.005.
2. Kadlec P, Gabrys B, Strandt S. Data-driven soft sensors in the process industry. Comput Chem Eng. 2009;33(4):795–814. doi:10.1016/j.compchemeng.2008.12.012.
3. Hochreiter S, Schmidhuber J. Long short-term memory. Neural Comput. 1997;9(8):1735–80. doi:10.1162/neco.1997.9.8.1735.
4. Schuster M, Paliwal KK. Bidirectional recurrent neural networks. IEEE Trans Signal Process. 1997;45(11):2673–81. doi:10.1109/78.650093.
5. Qin Y, Song D, Chen H, Cheng W, Jiang G, Cottrell GW. A dual-stage attention-based recurrent neural network for time series prediction. In: Proceedings of the Twenty-Sixth International Joint Conference on Artificial Intelligence; 2017 Aug 19–26; Melbourne, Australia. p. 2627–33. doi:10.24963/ijcai.2017/366.
6. Vaswani A, Shazeer N, Parmar N, Uszkoreit J, Jones L, Gomez AN, et al. Attention is all you need. In: Proceedings of the 31st Conference on Neural Information Processing Systems (NIPS 2017); 2017 Dec 4–9; Long Beach, CA, USA. p. 5998–6008.

7. Zhou H, Zhang S, Peng J, Zhang S, Li J, Xiong H, et al. Informer: beyond efficient transformer for long sequence time-series forecasting. In: Proceedings of the Thirty-Fifth AAAI Conference on Artificial Intelligence AAAI; 2021 Feb 2–9; Virtual Conference. p. 11106–15. doi:10.1609/aaai.v35i12.17325.
8. Lim B, Arık SÖ, Loeff N, Pfister T. Temporal fusion transformers for interpretable multi-horizon time series forecasting. Int J Forecast. 2021;37(4):1748–64. doi:10.1016/j.ijforecast.2021.03.012.
9. Huang NE, Shen Z, Long SR, Wu MC, Shih HH, Zheng Q, et al. The empirical mode decomposition and the Hilbert spectrum for nonlinear and non-stationary time series analysis. Proc R Soc Lond A. 1998;454(1971):903–95. doi:10.1098/rspa.1998.0193.
10. Dragomiretskiy K, Zosso D. Variational mode decomposition. IEEE Trans Signal Process. 2014;62(3):531–44. doi:10.1109/TSP.2013.2288675.
11. Torres ME, Colominas MA, Schlotthauer G, Flandrin P. A complete ensemble empirical mode decomposition with adaptive noise. In: Proceedings of the 2011 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP); 2011 May 22–27; Prague, Czech Republic. p. 4144–7. doi:10.1109/ICASSP.2011.5947265.
12. Colominas MA, Schlotthauer G, Torres ME. Improved complete ensemble EMD: a suitable tool for biomedical signal processing. Biomed Signal Process Control. 2014;14:19–29. doi:10.1016/j.bspc.2014.06.009.
13. Wei Y, Chen H, Tian Y, Wang C, Hu Y, Xie Z, et al. QPSO-optimized VMD-SE-transformer-BiLSTM for short-term air conditioning load forecasting in industrial buildings. AIP Adv. 2026;16(2):025315. doi:10.1063/5.0319827.
14. Wang X, Liu W, Wang Y, Yang G. A hybrid NO_x emission prediction model based on CEEMDAN and AM-LSTM. Fuel. 2022;310:122486. doi:10.1016/j.fuel.2021.122486.
15. Guo Y, Mao Z. Long-term prediction model for NO_x emission based on LSTM–transformer. Electronics. 2023;12(18):3929. doi:10.3390/electronics12183929.
16. Chen Y, Liang C, Liu D, Niu Q, Miao X, Dong G, et al. Embedding-graph-neural-network for transient NO_x emissions prediction. Energies. 2023;16(1):3. doi:10.3390/en16010003.
17. Duan H, Meng X, Tang J, Qiao J. Prediction of NO_x concentration using modular long short-term memory neural network for municipal solid waste incineration. Chin J Chem Eng. 2023;56:46–57. doi:10.1016/j.cjche.2022.06.028.
18. Meng X, Tang J, Qiao J. NO_x emissions prediction with a brain-inspired modular neural network in municipal solid waste incineration processes. IEEE Trans Ind Inform. 2022;18(7):4622–31. doi:10.1109/TII.2021.3116528.
19. Li Z, Yu Z, Chen D, Li L, Lu Z, Yao S. Soft sensing of NO_x emission from waste incineration process based on data de-noising and bidirectional long short-term memory neural networks. Waste Manage Res. 2025;43(4):602–15. doi:10.1177/0734242X241259643.
20. Duan H, Meng X, Tang J, Qiao J. NO_x emissions prediction for MSWI process based on dynamic modular neural network. Expert Syst Appl. 2024;238:122015. doi:10.1016/j.eswa.2023.122015.
21. Qiao J, Sun Z, Meng X. A comprehensively improved interval type-2 fuzzy neural network for NO_x emissions prediction in MSWI process. IEEE Trans Ind Inform. 2023;19(11):11286–97. doi:10.1109/TII.2023.3245640.
22. Wang R, Li F, Yan A. Modular stochastic configuration network based prediction model for NO_x emissions in municipal solid waste incineration process. Eng Appl Artif Intell. 2024;127:107315. doi:10.1016/j.engappai.2023.107315.
23. Qiao J, Zhou J, Meng X. A multitask learning model for the prediction of NO_x emissions in municipal solid waste incineration processes. IEEE Trans Instrum Meas. 2023;72:2502214. doi:10.1109/TIM.2022.3225056.
24. Wang R, Yan A, Tang J. Prediction of NO_x emissions based on ensemble stochastic configuration network for municipal solid waste incineration process. In: Proceedings of the 2023 35th Chinese Control and Decision Conference (CCDC); 2023 May 20–22; Yichang, China. p. 355–9. doi:10.1109/CCDC58219.2023.10327055.
25. Zhou JL, Duan HS, Meng X, Qiao JF. Prediction of NO_x emissions based on CSO-RBF neural network for municipal solid waste incineration. In: Proceedings of the 2021 China Automation Congress (CAC); 2021 Oct 22–24; Beijing, China. p. 2990–5. doi:10.1109/cac53003.2021.9727631.
26. Tan C, Bai Y, Li X. Dynamic prediction of NO_x emissions based on factor analysis and NARX neural network. In: Proceedings of the 2018 Chinese Automation Congress (CAC); 2018 Nov 30–Dec 2; Xi'an, China. p. 177–81. doi:10.1109/cac.2018.8623690.

27. Lian J, Liu Z, Wang H, Dong X. Adaptive variational mode decomposition method for signal processing based on mode characteristic. Mech Syst Signal Process. 2018;107:53–77. doi:10.1016/j.ymssp.2018.01.019.
28. Nazari M, Sakhaei SM. Successive variational mode decomposition. Signal Process. 2020;174:107610. doi:10.1016/j.sigpro.2020.107610.
29. Xu H, Guo Z, Cao Y, Cheng X, Zhang Q, Chen D. Research on short-term precipitation forecasting method based on CEEMDAN-GRU algorithm. Sci Rep. 2024;14(1):31885. doi:10.1038/s41598-024-83365-9.
30. Wu X, Zhu J, Wen Q. Short-term prediction of $PM_{2.5}$ concentration by hybrid neural network based on sequence decomposition. PLoS One. 2024;19(5):e0299603. doi:10.1371/journal.pone.0299603.
31. Guo X, Gao Y, Song W, Zen Y, Shi X. Short-term power load forecasting based on CEEMDAN-WT-VMD joint denoising and BiTCN-BiGRU-attention. Electronics. 2025;14(9):1871. doi:10.3390/electronics14091871.
32. Quilty J, Adamowski J. Addressing the incorrect usage of wavelet-based hydrological and water resources forecasting models for real-world applications with best practices and a new forecasting framework. J Hydrol. 2018;563:336–53. doi:10.1016/j.jhydrol.2018.05.003.
33. Bergmeir C, Hyndman RJ, Koo B. A note on the validity of cross-validation for evaluating autoregressive time series prediction. Comput Stat Data Anal. 2018;120:70–83. doi:10.1016/j.csda.2017.11.003.
34. Cerqueira V, Torgo L, Mozetič I. Evaluating time series forecasting models: an empirical study on performance estimation methods. Mach Learn. 2020;109(11):1997–2028. doi:10.1007/s10994-020-05910-7.
35. Clevert DA, Unterthiner T, Hochreiter S. Fast and accurate deep network learning by exponential linear units (ELUs). arXiv:1511.07289. 2015.
36. Kingma DP, Ba J. Adam: a method for stochastic optimization. arXiv:1412.6980. 2015.
37. Sze V, Chen YH, Yang TJ, Emer JS. Efficient processing of deep neural networks: a tutorial and survey. Proc IEEE. 2017;105(12):2295–329. doi:10.1109/JPROC.2017.2761740.
38. Hyndman RJ, Koehler AB. Another look at measures of forecast accuracy. Int J Forecast. 2006;22(4):679–88. doi:10.1016/j.ijforecast.2006.03.001.
39. Wu Y, Wang Z, Shi C, Jin X, Xu Z. A novel data-driven approach for coal-fired boiler under deep peak shaving to predict and optimize NO_x emission and heat exchange performance. Energy. 2024;304:132106. doi:10.1016/j.energy.2024.132106.
40. Wang C, Xu J, Xu K, Jiang L, Wang Y, Su S, et al. Real-time prediction and optimization of NO_x emissions using artificial intelligence and online combustion data. Fuel. 2025;391:134836. doi:10.1016/j.fuel.2025.134836.
41. Tashman LJ. Out-of-sample tests of forecasting accuracy: an analysis and review. Int J Forecast. 2000;16(4):437–50. doi:10.1016/S0169-2070(00)00065-0.