A Data-Driven Approach to Furnace Temperature Prediction: Decoupling Static and Dynamic Feature

Yusen Gang and Chen Peng

***Abstract*—Boilers are significant contributors to carbon emissions and pollutants across various industrial sectors. Accurately modeling furnace temperature is critical for optimizing combustion and enhancing operational efficiency. However, modeling poses significant challenges due to the interplay between rapidly changing dynamic processes and slowly varying static data. To address the coupling and redundancy inherent in these heterogeneous features, a hybrid framework for boiler temperature modeling (HFBTM) is proposed in this paper. The framework utilizes a multi-layer dense network to extract static features and a selective state space model (S3M) to capture dynamic features. These features are effectively combined through a hybrid feature fusion module using weighted integration, generating accurate temperature predictions across multiple future time steps. Compared with traditional single dynamic models, HFBTM mitigates information redundancy, reduces error propagation, and serves as an end-to-end furnace temperature prediction model that integrates both static and dynamic features. Experimental results demonstrate that the HFBTM framework delivers superior prediction performance across forecasting tasks of varying lengths. Compared with the existing methods, the proposed framework achieves higher accuracy and meets the requirements for precise modeling of boiler systems.**

***Index Terms*—Artificial intelligence, boiler dynamics modeling, combustion system, energy intelligence, boiler**

I. Introduction

Boilers are essential components in various industrial sectors, including manufacturing and power generation. Accurate monitoring and estimation of furnace temperature facilitates the timely detection of anomalies, thereby enhancing the overall stability of system operations [1]. Furthermore, precise modeling of the internal temperature of boiler serves as the foundation for effective combustion control, which improves thermal efficiency and reduces both energy consumption and environmental impact [2, 3]. However, achieving accurate temperature estimation is hindered by several challenges, including uncertainties in fuel calorific values, the complex and nonlinear nature of heat transfer processes, and the delays in system responses.

Manuscript received: 25 August 2024; revised: 29 October 2024; accepted: 5 November 2024. *(Corresponding author: Chen Peng.)*

Citation: Y. Gang and C. Peng, A data-driven approach to furnace temperature prediction: Decoupling static and dynamic feature, *IJICS*, 2024, 29(4), 171–176.

Yusen Gang and Chen Peng are with School of Mechanical Engineering and Automation, Shanghai University, Shanghai 200444, China (e-mail: yusengang@shu.edu.cn; c.peng@i.shu.edu.cn).

Digital Object Identifier 10.62678/IJICS202412.10139

Traditional boiler temperature modeling approaches are grounded in physical principles, including thermodynamics, mass conservation, and energy conservation [4, 5]. These methods utilize numerical analysis to develop mathematical representations of the boiler combustion system. However, they often rely on numerous empirical formulas, complicating the construction of accurate quantitative models [6, 7]. In the control field, state-space models [8] are commonly employed, providing a structured framework for describing energy conversion, material flow, and heat transfer in complex systems. Despite their utility, state-space models face certain limitations, such as parameter uncertainty, system nonlinearity, and sensitivity to disturbances, which can hinder their generalization in practical applications [9, 10].

Recent advancements in artificial intelligence have introduced data-driven approaches, which show significant potential for improving boiler temperature modeling [11, 12]. Among these, recurrent neural networks (RNNs) capture temporal dependencies by transferring latent features across time steps, enabling the prediction of boiler temperature [13]. Furthermore, while maintaining the original RNN structure, variants such as long short-term memory (LSTM) [14, 15] and gated recurrent unit (GRU) [16, 17] employ gating mechanisms to control the amount of latent feature transfer, thereby addressing the vanishing and exploding gradient problems encountered during temperature prediction. When facing the long-period prediction tasks, it is necessary for the model to capture dependencies over extended time ranges, which shares similarities with the objectives in natural language processing (NLP). As a result, the transformer model [18], originally designed for NLP tasks, has become increasingly popular in time-series forecasting. Its self-attention mechanism allows it to capture dependencies across distant time points, making it highly effective in modeling long-range dependencies and complex dynamics [19–21]. However, the computational cost of the transformer model increases as the sequence length grows, leading to decreased efficiency with larger inputs. To address these limitations, the selective state space model (S3M) has been designed for Mamba [22, 23], which introduces a dynamic state matrix that selectively filters and retains essential information, thereby improving computational efficiency and model performance.

Despite these advancements, most artificial intelligence based boiler temperature modeling methods utilize a single model to simultaneously process inputs. Specifically, all inputs are combined into a one-dimensional column vector

and fed into the model for training. This approach relies on the model to capture the relationships between various variables and temperatures. However, the combustion process encompasses both rapidly changing dynamic processes and slowly changing static processes. When a single model is employed for training, this heterogeneity can lead to information redundancy and ineffective feature selection, thereby hindering the stability and convergence of model.

In order to resolve these problems, a hybrid framework for boiler temperature modeling (HFBTM) that integrates both static and dynamic features is proposed in this paper. The primary components of the research are structured as follows:

(1) The HFBTM employs two distinct networks to extract different types of features: a multi-layer dense network for static characteristics and the S3M for dynamic temperature variations.

(2) The HFBTM decouples and integrates static and dynamic features, enabling a more comprehensive feature representation to capture the complex behaviors of the boiler system.

(3) Detailed experiments and comparisons are conducted, with experimental results validating the effectiveness of our proposed method.

II. Methodology

A. Problem Description

In general, the furnace temperature prediction problem can be described as

$$Y_t^{t+p} = f\left(X_{t-h}^{t-1}, Y_{t-h}^{t-1}\right) \tag{1}$$

where $X_{t-h}^{t-1} = (x_{t-1}, x_{t-2}, \ldots, x_{t-h})$ denotes the historical furnace input from time step $t-h$ to time step $t-1$, $Y_{t-h}^{t-1} = (y_{t-1}, y_{t-2}, \ldots, y_{t-h})$ denotes the historical furnace temperature from time step $t-h$ to time step $t-1$, and $Y_t^{t+p} = (y_t, y_{t+1}, \ldots, y_{t+p})$ denotes the furnace temperature to be predicted in the future time step p. In order to distinguish different types of input features, we extend this problem as

$$Y_t^{t+p} = f\left(S_{t-h}^{t-1}, D_{t-h}^{t-1}, Y_{t-h}^{t-1}\right) \tag{2}$$

where S_{t-h}^{t-1} denotes the static input parameter and D_{t-h}^{t-1} denotes the dynamic input parameter, $X_{t-h}^{t-1} = [S_{t-h}^{t-1}, D_{t-h}^{t-1}]$. Two different types of input data are shown in Fig. 1.

B. Hybrid Framework for Boiler Temperature Modeling

The HFBTM is structured around two core components: the static feature encoding module and the dynamic feature encoding module. The static feature encoding module employs a multi-layer dense network, whereas the dynamic feature encoding module is built on the S3M model. Static features, such as the physical structure of boiler and environmental conditions, remain relatively stable or change gradually over short periods. In contrast, dynamic features, such as power regulation and temperature variations during combustion, demonstrate significant temporal fluctuations and exhibit complex, nonlinear behaviors. Both static and dynamic features collaboratively influence the temperature dynamics of the boiler system.

Using a single model to process both types of features may cause feature entanglement, introducing redundant information that disrupts the learning process and impairs convergence, especially when handling static features. To mitigate this issue, the HFBTM framework decouples static and dynamic features, capitalizing on the distinct strengths of both modules to effectively capture feature-specific patterns and enhance predictive performance. The overall architecture of the HFBTM network is depicted in Fig. 2. T denotes the final predicted temperature, w_s and w_d denote the learnable attention coefficients.

a. Static Feature Encoding Module

The static feature encoding module utilizes a multi-layer dense network to extract complex and meaningful features from the static input data of the boiler system. These static features, such as the physical structure of the equipment, environmental conditions, and operational parameters, remain relatively stable or change gradually over short periods, but play a crucial role in shaping the operational state of system and temperature control. The multi-layer dense network combines neurons across multiple layers to model latent

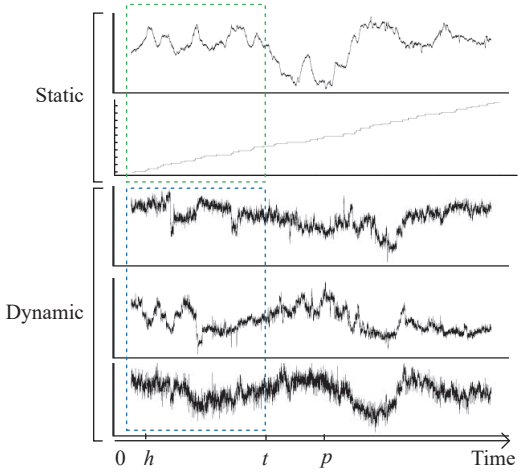

Figure 1 Diagram illustration of the process of prediction.

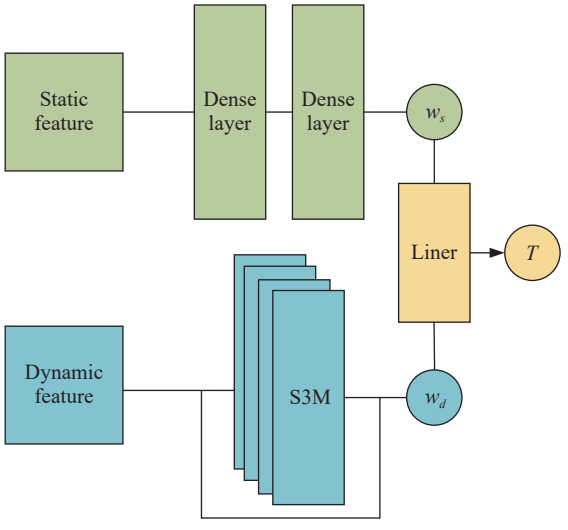

Figure 2 Structure of HFBTM.

interactions among static features. Nonlinear activation functions, such as rectified linear unit (ReLU), are applied between layers to enhance the ability of network to capture complex patterns.

This architecture efficiently processes the high-dimensional static inputs of the boiler system, reflecting the influence of long-term stable states on temperature dynamics and offering more accurate static feature representations for predictive tasks. The mathematical formulation of the multi-layer dense network is provided as

$$f_s^{(l)} = f_{\text{Activation}}\left(w^{(l)} f_s^{(l-1)} + b^{(l)}\right) \tag{3}$$

where $f_s^{(0)} = S_{t-h}$ denotes the input of the multi-layer dense network, $l \in [0, 1, 2, \ldots, L]$ denotes the layer index, L denotes the total number of layers in the network, $w^{(l)}$ denotes the weight matrix, $b^{(l)}$ denotes the bias vector of the L-th layer, and $f_{\text{Activation}}$ denotes the nonlinear activation function applied at each layer. The final output of the network $f_s^{(L)}$ denotes the extracted static feature vector, which is used for the subsequent fusion and prediction of dynamic features.

b. Dynamic Feature Encoding Module

Mamba is a novel neural network architecture designed to efficiently filter and propagate the dynamic states of input data through the S3M. Unlike traditional models with fixed parameters, the state matrix parameters in Mamba adapt dynamically to the input sequence, which is essential for capturing the rapidly changing characteristics of boiler systems. Given that boiler temperatures can fluctuate sharply in response to external disturbances, the Mamba model selectively filters these dynamic features, retaining critical information during state transitions while discarding redundancies. The structure of the S3M is illustrated in Fig. 3.

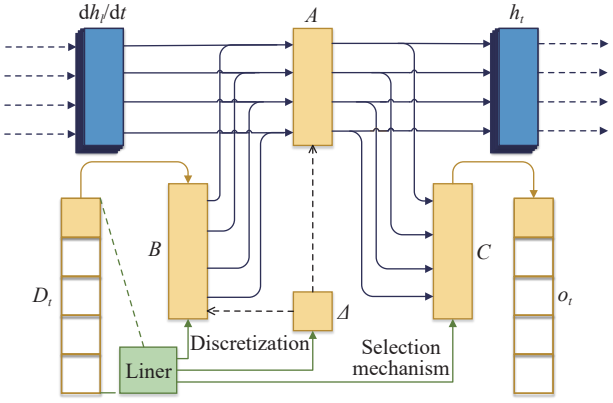

Figure 3 Structure of S3M.

This module incorporates latent states to describe the mapping between input and output sequences using linear state space models. It can be expressed as

$$\frac{\mathrm{d}h_t}{\mathrm{d}t} = Ah_t + BD_t \tag{4}$$

$$o_t = Ch_t \tag{5}$$

where h_t denotes the hidden state, D_t denotes the dynamic input parameter, and A, B, and C denote the parameter matrices of the system state space models, respectively. Specifically, A is the high-order polynomial projection operator (HIPPO) matrix [24], and each element in A is defined as

$$a_{ij} = \begin{cases} (2i+1)^{1/2}(2j+1)^{1/2}, & i > j; \\ i+1, & i = j; \\ 0, & i < j \end{cases} \tag{6}$$

The state matrix A is upper triangular, which means that the hidden state at each time step is influenced solely by previous time steps. This structure ensures that the weights in the upper triangular portion are zero, contributing nothing to the current hidden state. As a result, the model preserves the causality of the time series by preventing future values from influencing the current computation.

Additionally, the parameter matrices B and C are derived from input variables through linear transformations, allowing the model parameters to adapt dynamically based on the input. These modifications allow the S3M to discard unnecessary information and emphasize significant dynamic features.

The data processed by the computer are discrete sequences. Using the discretization techniques of continuous state space models, the resulting linear state space model is constructed as

$$\begin{cases} h(k+1) = \bar{A}h(k) + \bar{B}D(k), \\ o(k) = Ch(k) \end{cases} \tag{7}$$

where $h(k)$ denotes the internal state of the model at step k, $h(k+1)$ denotes the hidden state at the next step $k+1$, $D(k)$ denotes the model input at step k, and $\bar{A}$ and $\bar{B}$ are the discretized system parameter matrices. The specific discretization process can be expressed as

$$\bar{A} = \mathrm{e}^{\Delta A} \tag{8}$$

$$\bar{B} = (\Delta A)^{-1}(\mathrm{e}^{\Delta A} - I) \cdot \Delta B \tag{9}$$

where $\Delta \in (0, 1)$ denotes the discrete step length and I denotes the identity matrix.

c. Feature Fusion Module

To integrate static and dynamic features effectively, the HFBTM framework incorporates a hybrid feature fusion module that seamlessly combines the outputs from the static encoding module and the dynamic encoding module. The temperature of boiler system is influenced by both types of features. Static features represent the physical properties of the equipment and gradually changing factors, and dynamic features capture real-time fluctuations and external disturbances. The fusion module enhances the accuracy and stability of multi-step temperature predictions by combining these features through concatenation and weighted integration. This module can be expressed as

$$T = w_T(w_s \cdot f_s^{(L)} + w_d o(k)) + w_T \tag{10}$$

where T denotes the final predicted temperature, w_s and w_d denote the learnable attention coefficients, and w_T and d_T construct the linear layer for the final predicted temperature.

III. EXPERIMENT

A. Dataset

To validate the effectiveness of HFBTM, operational data are derived from the distributed control system (DCS) in a waste-to-energy plant from 1 April 2019 to 30 May 2019. The height of the incinerator in this power plant is 63.20 m, with a cross-sectional area of 20.53 × 12.56 m^2, utilizing a two-stage reciprocating grate to push the fuel. The primary air fans are located below the grate, with a total of 4 units. The structure of the boiler is shown in Fig. 4.

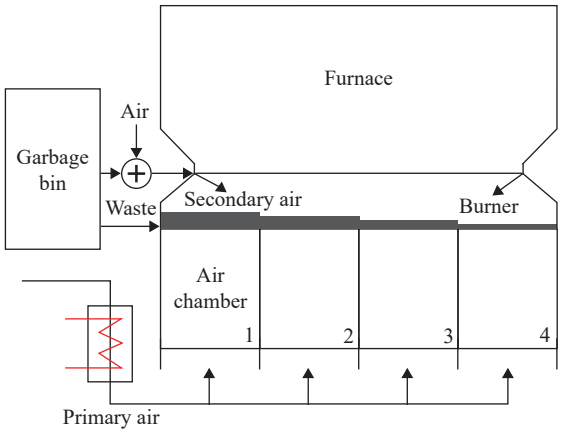

Figure 4 Structure of furnace.

This study identifies a set of 22 variables that influence furnace temperature based on expert insights [25]. The model predicts the average furnace temperature, with the variables and their statistical properties summarized in Tables 1 and 2, respectively.

For each input and output variable, 12,000 samples are collected at a 60 s sampling interval. A sliding window cross-validation strategy is employed to effectively handle the temporal dependencies. In each iteration, a time-ordered data window is used, with the first 80% serving as the training set and the remaining 20% as the testing set. The window is shifted forward across multiple rounds to ensure comprehensive evaluation and mitigate overfitting to specific time segments. To reduce noise, the Savitzky-Golay filter [26] is applied. All data are normalized using Z-score normalization to a range of [0, 1], eliminating the influence of outliers and accelerating network convergence.

B. Experiment Setup

Four time-series prediction methods are selected as comparison, including:

(1) CNN-Seq2Seq [27]: an LSTM-based encoder-decoder architecture with attention and CNN;

(2) Transformer [19]: a transformer architecture with CNN;

(3) Informer [28]: a transformer architecture with sparse self-attention mechanism;

(4) Mamba [22]: a Mamba model.

In the comparative experiments, we select similar parameters for networks with analogous structures to ensure fair and accurate comparisons. Specifically, the CNN convolution kernel is set to 3, and a residual structure is used.

Table 1 Description of variable in dataset.

Classification	Variable	Description	Unit
Static	x_1	Input waste weight	t
	x_2	Air chamber pressure 1	Pa
	x_3	Air chamber pressure 2	Pa
	x_4	Air chamber pressure 3	Pa
	x_5	Air chamber pressure 4	Pa
	x_6	Blower outlet pressure 1	kPa
	x_7	Blower outlet pressure 2	kPa
	x_8	Blower outlet pressure 3	kPa
	x_9	Blower outlet pressure 4	kPa
	x_{10}	Blower air output volume 1	m^3/h
	x_{11}	Blower air output volume 2	m^3/h
	x_{12}	Blower air output volume 3	m^3/h
	x_{13}	Blower air output volume 4	m^3/h
	x_{14}	Waste thickness on unit 1	mm
	x_{15}	Waste thickness on unit 2	mm
	x_{16}	Waste thickness on unit 3	mm
	x_{17}	Waste thickness on unit 4	mm
Dynamic	x_{18}	Oxygen content	%
	x_{19}	Primary air outlet temperature	°C
	x_{20}	Moisture content	%
	x_{21}	Ash content	%
Output	y	Average furnace temperature	°C

The LSTM architecture comprises two layers, each with 128 hidden units and a dropout rate of 0.2. Both the transformer and informer models are configured with the same hyperparameters, each model containing 6 layers, a hidden dimension of 128, 8 attention heads, and a feedforward network dimension of 2048. The Mamba and HFBTM are also configured with identical settings, using 6 layers and a hidden layer dimension of 128. All models use ReLU as the activation function, and the Adam optimizer is employed for optimization. The learning rate is configured at 2.0×10^{-4}, accompanied by a weight decay of 1.0×10^{-4}. The initial learning rate is set to 1.0×10^{-4}, with a dropout rate of 0.2. The training process spans 1000 epochs, incorporating early stopping to mitigate overfitting. The evaluation indicators used in this paper are root-mean-square deviation (RMSE) and mean absolute error (MAE), which can be expressed as

$$\mathrm{RMSE} = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(y_i - \hat{y}_i)^2} \tag{11}$$

$$\mathrm{MAE} = \frac{1}{N}\sum_{i=1}^{N}|y_i - \hat{y}_i| \tag{12}$$

where N denotes the total number of time steps across all sequences, $\hat{y}_i$ denotes the output of the network and y_i denotes the ground-truth value.

The construction and training are conducted using PyTorch 2.3.0 in a Python 3.11.0 environment. The experimental hardware setup consists of an Intel (R) Core i7–117000 CPU

Table 2 Summary statistic of the dataset.

Variable	Description	Mean	Maximum	Minimum	Unit
x_1	Input waste weight	158.484	259.863	53.327	t
x_2-x_5	Air chamber pressure	284.576	634.982	−105.251	Pa
x_6-x_9	Blower outlet pressure	0.345	0.866	−0.058	kPa
$x_{10}-x_{13}$	Blower air output volume	10,654.572	21,568.545	3861.199	m^3/h
$x_{14}-x_{17}$	Waste thickness per unit	60.911	94.998	28.283	mm
x_{18}	Oxygen content	4.599	9.044	2.099	%
x_{19}	Primary air outlet temperature	210.489	230.143	48.720	°C
x_{20}	Moisture content	18.167	22.919	16.342	%
x_{21}	Ash content	17.143	22.047	14.110	%
y	Average furnace temperature	1023.307	1119.185	886.396	°C

running at 2.50 GHz, a 24 GB NVIDIA GTX 3090 GPU, and 64 GB of RAM.

C. Result Analysis

The experimental results, as shown in Table 3, indicate that the CNN-LSTM hybrid model performs the worst across all forecasting tasks. In 10-step forecasting, both transformer and informer demonstrate advantages, achieving RMSE improvements of at least 29.1% and 16.9%, and MAE improvements of 32.7% and 31.1%, respectively, compared with CNN-LSTM. The Mamba model delivers the best performance in 10-step forecasting due to its selective mechanism, which allows the model parameters to adapt quickly to short-term data variations.

Table 3 Evaluation of prediction accuracy for different models. Bold indicates the optimal results.

Method	Step 10		Step 20		Step 30	
	RMSE	MAE	RMSE	MAE	RMSE	MAE
CNN-LSTM	6.864	40.811	8.774	66.629	10.309	91.912
Transformer	4.862	27.455	7.287	45.887	9.413	76.704
Informer	4.531	27.112	6.864	44.811	9.086	71.415
Mamba	**4.137**	**22.221**	6.829	41.795	8.774	66.629
HFBTM	4.216	23.419	**6.797**	**40.811**	**8.425**	**61.430**

As the forecasting horizon extends to 20 steps, both performances of transformer and informer decline, though both continue to outperform CNN-LSTM. This highlights the effectiveness of self-attention mechanisms in capturing long-range dependencies. Under these conditions, the proposed HFBTM framework slightly outperforms Mamba, with RMSE improving by 4.6% and MAE by 2.4%.

In 30-step forecasting, informer achieves higher accuracy than transformer, with RMSE and MAE improvements of 3.5% and 6.9%, respectively. Mamba also surpasses transformer, with RMSE improving by 6.7% and MAE by 13.1%. At this horizon, the proposed HFBTM framework achieves the highest predictive accuracy among all models.

IV. Conclusion

An HFBTM is proposed in this paper, which integrates a multi-layer dense network with a selective state space model. By decoupling static and dynamic features and processing them separately, the framework addresses the issue of feature redundancy in boiler time series data, enhancing the robustness and accuracy of furnace temperature predictions and enabling precise modeling of the boiler combustion system. Compared with the existing models, such as CNN-LSTM, transformer, and informer, HFBTM demonstrates superior adaptability and generalization in multi-step prediction tasks.

This study introduces new methods and insights for improving combustion efficiency and enhancing the operational stability of boiler systems. To further optimize energy utilization, future research will explore control strategies tailored to HFBTM to achieve precise regulation of boiler control parameters. Additionally, we aim to investigate faster modeling techniques to enable real-time modeling of boiler systems.

Acknowledgment

This work was supported by the National Natural Science Foundation of China (No. 61833011).

References

[1] A. B. M. A. Malek, M. Hasanuzzaman, and N. A. Rahim, Prospects, progress, challenges and policies for clean power generation from biomass resources, *Clean Technol. Environ. Policy*, 2020, 22(6), 1229–1253.

[2] M. A. Nemitallah, M. A. Nabhan, M. Alowaifeer, A. Haeruman, F. Alzahrani, M. A. Habib, M. Elshafei, M. I. Abouheaf, M. Aliyu, and M. Alfarraj, Artificial intelligence for control and optimization of boilers' performance and emissions: A review, *J. Cleaner Prod.*, 2023, 417, 138109.

[3] J. He, J. Zhang, L. Wang, X. Hu, J. Xue, Y. Zhao, X. Wang, and C. Dong, Optimizing the controlling parameters of a biomass boiler based on big data, *Energies*, 2023, 16(23), 7783.

[4] J. Porteiro, J. Collazo, D. Patiño, E. Granada, J. C. Moran Gonzalez, and J. L. Míguez, Numerical modeling of a biomass pellet domestic boiler, *Energy Fuels*, 2009, 23(2), 1067–1075.

[5] C. Yin, L. Rosendahl, S. K. Kær, S. Clausen, S. L. Hvid, and T. Hille, Mathematical modeling and experimental study of biomass combustion in a thermal 108 MW grate-fired boiler, *Energy Fuels*, 2008, 22(2), 1380–1390.

[6] X. Su, X. Chen, Q. Fang, L. Ma, P. Tan, C. Zhang, G. Chen, and C. Yin, An integrated model for flexible simulation of biomass combustion in a travelling grate-fired boiler, *Energy*, 2024, 307, 132605.

[7] J. Chang, X. Ma, X. Wang, and X. Li, CPFD modeling of hydrodynamics, combustion and NO_x emissions in an industrial CFB boiler, *Particuology*, 2023, 81, 174–188.

[8] M. Gölles, S. Reiter, T. Brunner, N. Dourdoumas, and I. Obernberger, Model based control of a small-scale biomass boiler, *Control Eng. Pract*., 2014, 22, 94–102.

[9] X. Wu, J. Shen, Y. Li, M. Wang, and A. Lawal, Flexible operation of post-combustion solvent-based carbon capture for coal-fired power plants using multi-model predictive control: A simulation study, *Fuel*, 2018, 220, 931–941.

[10] X. Wu, J. Shen, Y. Li, M. Wang, A. Lawal, and K. Y. Lee, Dynamic behavior investigations and disturbance rejection predictive control of solvent-based post-combustion CO_2 capture process, *Fuel*, 2019, 242, 624–637.

[11] R. P. Nikula, M. Ruusunen, and K. Leiviskä, Data-driven framework for boiler performance monitoring, *Appl. Energy*, 2016, 183, 1374–1388.

[12] S. K. Jha, J. Bilalovic, A. Jha, N. Patel, and H. Zhang, Renewable energy: Present research and future scope of artificial intelligence, *Renewable Sustain. Energy Rev*., 2017, 77, 297–317.

[13] Y. Zhao, Y. Cai, and H. Jiang, Recurrent neural network-based hybrid modeling method for digital twin of boiler system in coal-fired power plant, *Appl. Sci*., 2023, 13(8), 4905.

[14] Z. Tong, X. Chen, S. Tong, and Q. Yang, Dense residual LSTM-attention network for boiler steam temperature prediction with uncertainty analysis, *ACS Omega*, 2022, 7(13), 11422–11429.

[15] S. Pan, X. Shi, B. Dong, J. Skvaril, H. Zhang, Y. Liang, and H. Li, Multivariate time series prediction for CO_2 concentration and flowrate of flue gas from biomass-fired power plants, *Fuel*, 2024, 359, 130344.

[16] J. F. Tuttle, L. D. Blackburn, K. Andersson, and K. M. Powell, A systematic comparison of machine learning methods for modeling of dynamic processes applied to combustion emission rate modeling, *Appl. Energy*, 2021, 292, 116886.

[17] T. Ye, M. Dong, J. Long, Y. Zheng, Y. Liang, and J. Lu, Multi-objective modeling of boiler combustion based on feature fusion and Bayesian optimization, *Comput. Chem. Eng*., 2022, 165, 107913.

[18] K. Han, A. Xiao, E. Wu, J. Guo, C. Xu, and Y. Wang, Transformer in transformer, in *Proc. 35th International Conference on Neural Information Processing Systems*, Montreal, Canada, 2021, 1217.

[19] Y. Ye, H. Lin, and H. Zhou, Modeling and prediction of key parameters of circulating fluidized bed boiler based on transformer, *J. Taiwan Inst. Chem. Eng*., 2024, 162, 105622.

[20] R. Li, D. Zeng, T. Li, B. Ti, and Y. Hu, Real-time prediction of SO_2 emission concentration under wide range of variable loads by convolution-LSTM VE-transformer, *Energy*, 2023, 269, 126781.

[21] Z. Li, X. Zhang, and Z. Dong, TSF-transformer: A time series forecasting model for exhaust gas emission using transformer, *Appl. Intell*., 2023, 53(13), 17211–17225.

[22] A. Gu and T. Dao, Mamba: Linear-time sequence modeling with selective state spaces, arXiv preprint arXiv: 2312.00752, 2023.

[23] M. A. Ahamed and Q. Cheng, Timemachine: A time series is worth 4 Mambas for long-term forecasting, in *Proc. 27th European Conference on Artificial Intelligence*, Santiago de Compostela, Spain, 2024, 1688–1695.

[24] A. Gu, T. Dao, S. Ermon, A. Rudra, and C. Ré, HiPPO: Recurrent memory with optimal polynomial projections, in *Proc. 34th International Conference on Neural Information Processing Systems*, Vancouver, Canada, 2020, 125,

[25] X. Wu, H. Zhang, H. Chen, S. Wang, and L. Gong, Combustion optimization study of pulverized coal boiler based on proximal policy optimization algorithm, *Appl. Therm. Eng*., 2024, 254, 123857.

[26] R. W. Schafer, What is a Savitzky-Golay filter? [lecture notes], *IEEE Signal Process. Mag*., 2011, 28(4), 111–117.

[27] G. Zhang, X. Bai, and Y. Wang, Short-time multi-energy load forecasting method based on CNN-Seq2Seq model with attention mechanism, *Mach. Learn. Appl*., 2021, 5, 100064.

[28] H. Zhou, S. Zhang, J. Peng, S. Zhang, J. Li, H. Xiong, and W. Zhang, Informer: Beyond efficient transformer for long sequence time-series forecasting, in *Proc. 35th AAAI Conference on Artificial Intelligence*, online, 2021, 11106–11115.

Yusen Gang received the ME degree in electrical engineering from China University of Mining and Technology, China, in 2019. He is currently pursuing the PhD degree at School of Mechanical Engineering and Automation, Shanghai University, China. His research interests include time series prediction and system modeling.

Chen Peng received the PhD degree in control theory and control engineering from China University of Mining and Technology, China, in 2002. He is currently a professor at School of Mechanical Engineering and Automation, Shanghai University, China. His research interests include networked control, safety control, and time-delay system analysis.