Electronics and Electrical Engineering

AIMS Electronics and Electrical Engineering, 10(2): 240–264.
DOI: 10.3934/electreng.2026010
Received: 27 October 2025
Revised: 30 January 2026
Accepted: 24 February 2026
Published: 16 March 2026

https://www.aimspress.com/journal/ElectrEng

Research article

A data-driven predictive modeling approach for converter inlet temperature in the copper smelting acid-making process

Chunbo Wang[1,*], Xiaoli Li[1,2] and Kang Wang[1]

[1] School of Information Science and Technology, Beijing University of Technology. Beijing 100124, China
[2] Beijing Key Laboratory of Computational Intelligence and Intelligent System, Beijing 100124, China

* **Correspondence:** Email: bjut_wcb@emails.bjut.edu.cn.

Abstract: In the copper smelting Flue Gas Sulfuric Acid, accurate prediction of the converter inlet temperature is of great engineering significance for ensuring safe operation, reducing energy consumption, and improving system stability. To address the strong coupling, long time delay, and pronounced nonlinearity inherent in this process, we proposed a data-driven soft-sensing method designed for real-world industrial applications. The proposed approach was validated in an actual "two-conversion–two-absorption" industrial system. First, the Partial Autocorrelation Function was employed to analyze Distributed Control System data, identifying the most relevant input variables and time windows to provide a high-correlation feature foundation for model construction. Then, a Gated Recurrent Unit neural network was utilized to model temporal dependencies, effectively capturing complex dynamic relationships while suppressing noise interference. Furthermore, an Attention Mechanism was incorporated to enhance the model's focus on key temporal features, improving adaptability to process fluctuations and enhancing the interpretability of prediction results. Finally, the proposed model was implemented and verified in an industrial Flue Gas Sulfuric Acid process. The experimental results demonstrated that the method maintains stable predictive performance and accurately captures the dynamic variation of the first-layer converter inlet temperature. It effectively assists operators in adjusting process parameters in real time, thereby improving system safety, continuity, and operational reliability. The findings verify the engineering feasibility and practical applicability of the proposed method, providing strong technical support for the intelligent operation of copper smelting sulfuric acid production systems.

Keywords: copper smelting; flue gas sulfuric acid; data-driven modeling; GRU; attention mechanism; soft sensing

1. Introduction

In recent years, air pollution caused by sulfur dioxide (SO_2) emissions has attracted increasing global attention [1]. As a typical atmospheric pollutant, SO_2 can form acid rain in the atmosphere, leading to ecosystem acidification, soil degradation, and serious threats to human health through respiratory system damage [2]. Environmental monitoring results indicate that the metallurgy, chemical, and coal-fired power generation industries remain the major sources of SO_2 emissions [3]. Among them, smelters are considered one of the major regional contributors to SO_2 pollution due to their long-term use of high-sulfur raw materials and continuous emission characteristics [4]. To mitigate these impacts, smelting enterprises commonly adopt the Flue Gas Sulfuric Acid (FGSA) process as a crucial technology for SO_2 reduction and resource recovery. In this process, SO_2-rich flue gas is oxidized to sulfur trioxide (SO_3) and subsequently converted into sulfuric acid (H_2SO_4), thereby achieving the dual objectives of pollutant abatement and byproduct utilization [5]. Owing to its high desulfurization efficiency, low operating cost, and excellent process stability, the FGSA technology has been widely applied in China's metallurgical industry.

In the FGSA process, the conversion stage is a critical step that determines the deep removal efficiency of SO_2, with the primary objective of promoting the oxidation of SO_2 into SO_3 under catalytic conditions [6]. The conversion rate not only affects the purity and yield of sulfuric acid products but also serves as a key indicator of the system's economic performance and operational stability [7]. Since this reaction is a reversible exothermic process, temperature plays a decisive role in maintaining reaction equilibrium. Excessive temperature can shift the equilibrium backward, reducing the conversion rate, whereas insufficient temperature slows the reaction rate and decreases system efficiency. Therefore, precise temperature control of the converter is essential for sustaining thermal balance and improving overall conversion efficiency. However, most smelting plants exhibit a low level of process informatization in the acid-making section. Operators typically rely on empirical adjustments to control temperature, lacking data-based predictive and control models, which limits process optimization and intelligent operation [8]. To address this issue, it is crucial to develop a data-driven prediction model for the converter inlet temperature, enabling accurate characterization of the dynamic thermal behavior of the reaction and providing a scientific foundation for temperature control strategies.

Modeling research serves as the theoretical foundation for achieving control and optimization in complex industrial processes [9,10]. In recent years, studies on modeling approaches for copper smelting FGSA and related desulfurization processes have become a focal point in academic and industrial fields. Generally, modeling methods can be categorized into two major types: mechanism-based modeling and data-driven modeling approaches [11]. Mechanism-based models rely on a deep understanding of the system's reaction kinetics, fluid dynamics, and thermodynamic characteristics. They describe the conversion mechanism of SO_2 to SO_3 by establishing mass and energy conservation equations. For instance, in Flue Gas Desulfurization (FGD) research, scholars have proposed hybrid models that integrate fluid dynamics and gas absorption mechanisms to characterize SO_2 removal processes [12]. Other researchers have developed comprehensive models

based on multiple subprocesses to evaluate desulfurization tower performance [13]. Moreover, researchers have developed and validated an oxidation model for wet flue gas desulfurization to describe oxygen mass transfer and sulfite oxidation behavior, and to support the optimization of the oxidation air flow rate [14]. Although mechanism-based models offer strong physical interpretability, they often involve highly complex parameters and are computationally demanding. In high-temperature, multiphase, and multi-reaction systems such as smelting FGSA processes, it is particularly difficult to obtain accurate kinetic parameters, which severely limits their applicability for real-time prediction and process control.

In contrast, data-driven modeling approaches rely solely on process measurement data to learn system characteristics, thereby avoiding the need for explicit modeling of complex reaction mechanisms [15]. This approach has been widely applied in industrial soft sensing and predictive control. Studies have demonstrated that soft-sensing models based on neural networks and deep learning frameworks exhibit outstanding dynamic prediction performance in chemical and metallurgical processes. For instance, Ren et al. integrated the Particle Swarm Optimization (PSO) algorithm with a Back-Propagation (BP) neural network to construct a soft-sensing model, achieving improved prediction accuracy [16]. Yuan et al. proposed a Layer-Wise Data Augmentation (LWDA) strategy and developed a Stacked Autoencoder (LWDA-SAE) model for deep feature extraction [17]. Lui et al. introduced a Supervised Bidirectional Long Short-Term Memory (SBiLSTM) network that leverages bidirectional temporal information to capture dynamic features [18]. Yuan et al. further incorporated a spatiotemporal attention mechanism into an LSTM architecture to dynamically extract key features [19]. Additionally, Yang et al. proposed a variable attention-based gated Gated Recurrent Unit (VAGGRU) network, which enhances nonlinear system modeling accuracy through adaptive weighting [20]. These studies collectively demonstrate that data-driven modeling can effectively capture the nonlinear and dynamic characteristics of complex systems and achieve high-precision prediction of key process variables, thereby providing reliable support for intelligent control in industrial applications.

However, most researchers adopt static or simplified dynamic structures, which fail to adequately capture the strong time-varying characteristics and multivariable coupling inherent in the copper smelting FGSA process. To address this limitation, we propose a novel dynamic data-driven modeling approach that integrates the Partial Autocorrelation Function (PACF), Gated Recurrent Unit (GRU) neural network, and Attention Mechanism (AM). This hybrid framework is designed to characterize the dynamic variation of the converter temperature, enabling high-precision soft sensing and prediction of the conversion process. In the proposed model, the PACF is employed to identify significant time-lag features among variables and determine the optimal input time window. The GRU is utilized to learn nonlinear dynamic relationships within multivariate time series, while the attention mechanism adaptively assigns weights to emphasize critical time steps and key feature variables. Through this synergistic integration, the model enhances prediction accuracy and stability.

Our objective of this study is to construct a converter temperature prediction framework that combines the strengths of statistical analysis and deep learning, thereby providing data-driven support for the intelligent control of copper smelting FGSA systems. The remainder of this paper is organized as follows: In Section II, we present the fundamental mechanism of the FGSA process, including SO_2 oxidation kinetics, temperature effects, and major equipment characteristics. In Section III, we introduce the proposed PACF–GRU–AM dynamic modeling method and explain its structure and training strategy. In Section IV, we provide experimental studies and result analysis,

validating the proposed method by comparing its performance with several benchmark models. Finally, in Section V, we conclude the paper by summarizing the major findings and innovations, and discuss potential applications of the proposed model in industrial intelligent control and energy efficiency optimization.

2. Copper smelting FGSA process

2.1. Process overview

During the copper smelting process, the combustion of high-sulfur raw materials generates a large volume of SO_2-containing flue gas. If emitted directly without proper treatment, it can cause serious environmental issues, such as acid rain, smog, and ecosystem acidification, as well as pose significant threats to human health. To achieve pollutant resource utilization and cleaner production, the smelting industry widely adopts the FGSA process, which converts SO_2 in the flue gas into sulfuric acid, thereby achieving a balance between environmental protection and economic benefits. After purification and drying, the sulfur-containing flue gas produced from smelting is effectively freed from particulates and moisture, resulting in a dry and clean SO_2 gas stream. The gas then enters the core stage of the acid-making system: the conversion process. In this stage, SO_2 is oxidized to SO_3 under the action of a catalyst, and the generated SO_3 is subsequently absorbed by dilute sulfuric acid to produce concentrated H_2SO_4. The overall reactions can be expressed as follows:

$$2SO_2 + O_2 \underset{\text{high temperature}}{\overset{\text{catalyst}}{\rightleftharpoons}} 2SO_3 + Q \tag{1}$$

$$SO_3 + H_2O = H_2SO_4 + Q \tag{2}$$

The conversion rate of SO_2 to SO_3 is a key indicator for evaluating the performance of the FGSA process. A higher conversion rate indicates a more complete reaction, higher acid concentration, and cleaner tail gas emissions. To achieve this objective, it is essential to maintain coordinated control of temperature, gas flow rate, and oxygen concentration under complex multistage catalytic reaction conditions. In particular, the inlet temperature of the first converter layer plays a decisive role in determining the reaction rate, thermal balance, and overall energy efficiency of the system. Therefore, it serves as the primary focus of modeling and prediction in this study.

2.2. Conversion system process and operating principles

The FGSA process generally consists of a blower, heat exchangers, converters, and absorption towers. A typical configuration of the “two-conversion–two-absorption” process in copper smelting acid production is illustrated in Figure 1. This configuration represents the standard process design adopted in most copper smelting FGSA systems. The process can be summarized as follows: After the SO_2-containing flue gas undergoes drying and absorption, the cleaned gas is delivered into the system by the blower and preheated to a specified temperature through a heat exchanger before entering the first layer of the converter. The inlet temperature at this stage directly determines the activation degree of the catalyst and the initial reaction rate, making it a critical parameter for

achieving high conversion efficiency. After the first catalytic reaction, the gas passes through the heat exchanger for heat recovery and then enters the second and third catalyst layers for continued oxidation, further reducing the SO_2 concentration. Once the gas has undergone multistage catalytic conversion and its temperature decreases to a level suitable for absorption, it is directed into the first absorption tower, where SO_3 reacts with the water content in concentrated sulfuric acid to produce H_2SO_4, completing the first absorption stage. The partially converted tail gas is reheated through another heat exchanger and introduced into the fourth catalyst layer for deep conversion. It is then cooled and fed into the second absorption tower for the final absorption reaction, thereby completing the two-conversion–two-absorption process.

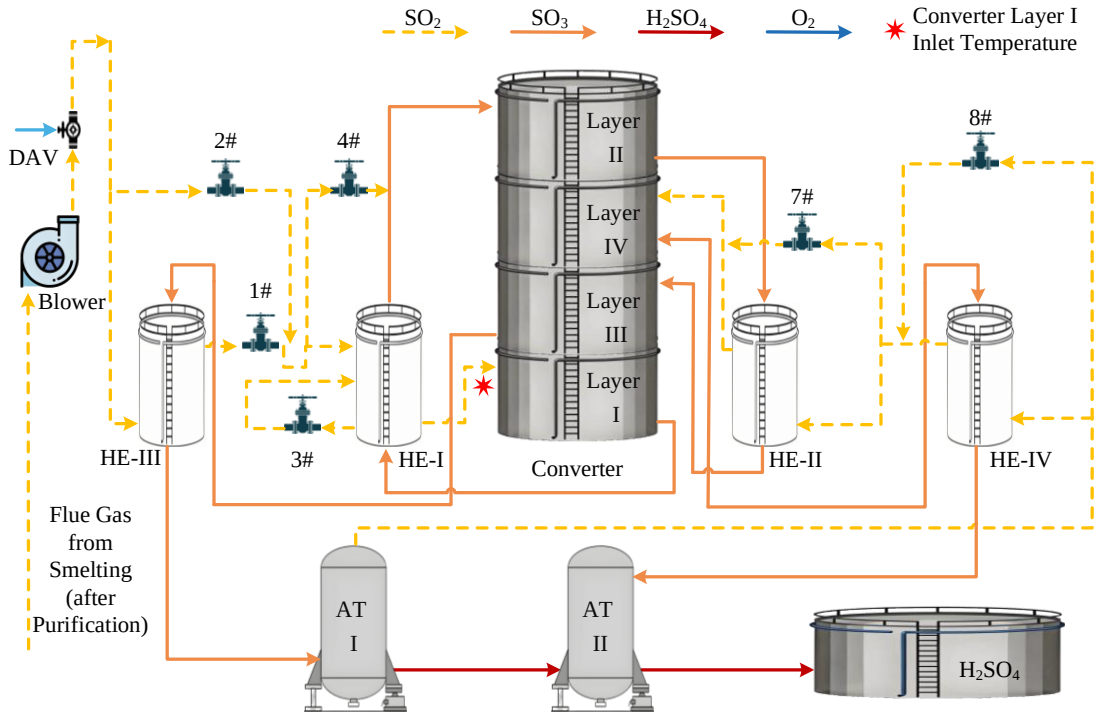

Figure 1. Process flow of the FGSA conversion section.

The major functions of each device in the FGSA conversion system are described as follows:

Blower: The blower delivers the purified SO_2 flue gas into the conversion system and serves as the power core of the acid-making process. By adjusting the guide vane opening, the flow rate and pressure of the gas can be controlled, thereby influencing the temperature distribution in the downstream reaction section. This control is essential for maintaining stable system operation.

Dilution Air Valve (DAV): Located between the blower outlet and the converter inlet, the DAV is a key component for regulating gas composition and reaction conditions. Its primary function is to control the amount of dilution air introduced into the system, thereby adjusting the SO_2-to-O_2 ratio in the flue gas. This ensures that the reactant gas concentration remains within an optimal range for stable catalytic conversion.

Heat Exchanger (HE): The heat exchangers perform dual functions of preheating the flue gas and cooling the reaction products, making them critical to maintaining the system's thermal balance. The SO_2 gas must be heated to an appropriate temperature before entering the converter to activate the

catalytic reaction, while the generated SO_3 gas must be cooled before entering the absorption tower. This process enables energy recovery and redistribution, providing the necessary conditions for stabilizing the converter inlet temperature.

Converter: The converter is the core equipment of the FGSA system, responsible for the catalytic oxidation of SO_2 into SO_3. It typically consists of four catalyst beds arranged in a series to balance the reaction rate with thermodynamic constraints. Among them, the inlet temperature of the first layer is the most critical parameter, as it determines the initial kinetic conditions of the reaction system. Excessive temperature can accelerate the reaction and cause excessive local heat release, leading to catalyst sintering and equilibrium reversal. Conversely, an insufficient temperature reduces the reaction rate and conversion efficiency. Therefore, the first-layer inlet temperature of the converter is the key parameter governing system thermal balance and conversion efficiency, and it constitutes the main focus of modeling and prediction in this study.

Absorption Tower (AT): The absorption towers facilitate the absorption and conversion of SO_3 into H_2SO_4. In industrial practice, the two-conversion–two-absorption configuration is typically adopted, enabling the flue gas to undergo two catalytic conversion stages followed by two absorption stages. Concentrated sulfuric acid is used as the absorbent instead of water to prevent acid mist formation and to improve absorption efficiency.

This multistage reaction–absorption system achieves a high overall conversion efficiency. However, during actual operation, temperature fluctuations in the earlier reaction stages are often propagated to subsequent sections, thereby disrupting the system's thermal balance. Among all process parameters, the inlet temperature of the first converter layer is the most sensitive. Its variations directly affect the catalytic reaction rate and temperature gradient distribution across the layers, the thermal load and energy recovery efficiency of the heat exchangers, as well as the residual SO_2 concentration in the tail gas and the overall conversion rate. Therefore, developing a high-precision prediction model for the inlet temperature of the first converter layer is of great importance for maintaining system thermal balance, optimizing energy consumption, and improving product quality.

2.3. SO_2 conversion reaction mechanism and temperature effects

The oxidation of SO_2 to SO_3 is a typical exothermic reversible reaction, whose kinetic and thermodynamic characteristics determine the complexity of temperature control. According to Le Chatelier's principle, an increase in temperature shifts the equilibrium toward the reactants, thereby decreasing the SO_2 conversion rate. Conversely, a lower temperature favors equilibrium conversion but slows the reaction rate. In practical operation, the relationship between the reaction rate constant k_r and temperature can be expressed by the Arrhenius equation:

$$k_r = Ae^{-E_a/RT^{\text{abs}}} \tag{3}$$

where k_r represents the reaction rate constant, which characterizes the reaction rate per unit concentration. A denotes the pre-exponential factor, indicating the probability of effective molecular collisions. E_a is the apparent activation energy, corresponding to the minimum energy threshold required for reactant molecules to undergo chemical reactions. R is the universal gas constant, with a

value of 8.314 $J \cdot mol^{-1} \cdot K^{-1}$, and T^{abs} represents the absolute temperature, expressed in kelvins (K). This equation indicates that as temperature increases, the absolute value of the exponential term $-E_a/RT^{abs}$ decreases, thereby increasing the value of k_r and accelerating the overall reaction rate.

It can thus be concluded that temperature exerts a promoting and inhibiting effect on the oxidation reaction. To balance the reaction rate and equilibrium conversion efficiency, the inlet temperature of the first converter layer is typically maintained within the range of 375–395 ℃ in industrial operations. This temperature range enables rapid reaction initiation while preventing excessive exothermic heat release that could trigger the reverse reaction.

If the inlet temperature is not properly controlled, several issues may arise: Excessive temperature: Intense heat release at the early stage leads to catalyst deactivation and difficulty in maintaining the temperature of subsequent layers. Insufficient temperature: The SO_2 oxidation rate becomes too slow, reducing the reaction driving force and lowering the overall conversion efficiency. Large temperature fluctuations: Instability in heat exchanger load occurs, adversely affecting downstream absorption efficiency. Therefore, the inlet temperature of the first converter layer not only determines the initial reaction conditions but also influences the system's overall energy distribution and thermal balance, making it the key control parameter for achieving efficient and stable conversion performance.

In this chapter, we provide a systematic analysis of the overall process, key equipment, and reaction mechanisms of the copper smelting FGSA system. The results indicate that the converter temperature is the critical factor influencing the $SO_2 \rightarrow SO_3$ reaction rate and equilibrium conversion efficiency. In particular, the inlet temperature of the first converter layer plays a decisive role in determining the initial reaction conditions, catalyst activation, and overall thermal balance of the system. Dynamic variations in the inlet temperature directly affect the conversion rate, heat distribution, and residual SO_2 concentration in the tail gas. However, most acid-making units rely on operator experience for temperature adjustment, lacking systematic predictive models and intelligent control strategies. As a result, they struggle to manage process fluctuations and nonlinear characteristics effectively. To address these challenges, we propose a data-driven dynamic modeling approach that integrates the PACF, GRU, and AM. The proposed method aims to achieve accurate prediction and soft sensing of the first-layer converter inlet temperature, providing a solid technical foundation for the intelligent and digital transformation of copper smelting FGSA systems.

3. Predictive modeling of the first-layer converter inlet temperature

3.1. Standardization and optimization strategies for industrial process data

In the copper smelting FGSA conversion process, the Distributed Control System (DCS) collects real-time field data containing multiple process variables, including temperature, flow rate, SO_2/O_2 concentrations, and valve signals. These industrial signals typically exhibit nonlinear characteristics, strong noise, missing values, and time delays. To ensure the effectiveness of model training and the reliability of the prediction results for the first-layer converter inlet temperature, a systematic data preprocessing procedure is conducted prior to modeling. The preprocessing workflow primarily includes data selection, outlier detection and correction, missing-value imputation, signal smoothing, and normalization.

Overall, the preprocessing is organized as a sequential pipeline: (1) Select and filter

high-relevance variables to form a consistent input set; (2) detect and correct abnormal points to remove non-physical disturbances; (3) compensate missing values and align timestamps to ensure multivariate synchronization; (4) smooth signals to suppress high-frequency noise while preserving major trends; and (5) normalize all variables to a unified scale for stable and efficient model training. This ordered workflow ensures that each subsequent step is performed on cleaner and better-aligned data, thereby forming a coherent and reliable dataset for dynamic modeling.

Feature selection and data filtering: Based on the mechanistic understanding of the acid-making process, combined with expert knowledge and correlation analysis, a set of key variables closely related to the first-layer converter inlet temperature is identified. These include: flue gas temperature after drying absorption T_{gas}, SO_2 concentration C_{SO_2}, O_2 concentration C_{O_2}, system flue gas flow rate F_{gas}, air supply rate F_{air}, blower guide vane angle θ_{fan}, inlet and outlet temperatures of each heat exchanger T_{HX}^{in}, T_{HX}^{out}, and valve opening degrees V_i of key control units.

Furthermore, the Pearson correlation coefficient method is applied to preliminarily evaluate the linear correlation between each variable and the target temperature.

$$r_{xy} = \frac{\sum_{i=1}^{N}(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^{N}(x_i - \bar{x})^2 \sum_{i=1}^{N}(y_i - \bar{y})^2}}, \tag{4}$$

where r_{xy} denotes the correlation coefficient between variables x and y, with a value range of $[-1, 1]$. When $|r_{xy}|>0.5$, the variable is considered to have a strong linear correlation with the target temperature. After feature screening, the selected variables are retained to form the model input matrix: $\mathbf{X} = [x_1, x_2, \ldots, x_n]$. With the input variables determined, the raw signals are then cleaned to eliminate abnormal points and missing segments before any smoothing and scaling operations were performed.

Due to instrument measurement errors, data delays, or transient disturbances in the field, anomalies may exist in the raw industrial data. In this study, the Pauta criterion is applied for outlier detection, which can be expressed as follows:

$$|x_i - \bar{x}| > 3\sigma \tag{5}$$

where $\bar{x}$ denotes the sample mean, and σ represents the standard deviation.

If outliers are detected, the correction method is determined based on their duration and positional distribution. For isolated outliers, a linear interpolation method is applied to replace the abnormal data points. For short-term local disturbances, a moving-average strategy is further used to weaken abrupt fluctuations while maintaining the overall physical continuity of the process signals.

$$x_i = x_{i-1} + \frac{(x_{i+1} - x_{i-1})}{2} \tag{6}$$

This method effectively mitigates the influence of persistent anomalies and stabilizes the overall signal by averaging adjacent data points within a defined window.

$$x_i^{\text{new}} = \frac{1}{m} \sum_{k=i-\frac{m-1}{2}}^{i+\frac{m-1}{2}} x_k, \tag{7}$$

where m represents the window width of the moving average.

The application of the above methods effectively suppresses abnormal fluctuations and preserves the physical continuity of the process signal.

Missing-Value Compensation and Time Alignment: To ensure the synchronization of multivariable time series, all feature data are aligned to a unified timestamp and subjected to missing-value compensation. When the number of missing points is less than or equal to three sampling intervals, a linear interpolation method is applied to restore the missing values.

If the missing gap is large, Cubic Spline Interpolation is employed. Continuity of the first and second derivatives at the interval endpoints is ensured, thereby enabling the time series trend to be smoothly reconstructed. Finally, all variables are resampled at 1-minute intervals to guarantee temporal consistency across the dataset. After completing alignment and imputation, the signals become temporally consistent, which provides a suitable basis for subsequent noise suppression.

Signal Smoothing and Noise Suppression: To address the high-frequency noise present in the DCS measurement signals, a Weighted Moving Average method is employed for signal smoothing, as expressed by the following equation:

$$x_t^{\text{smooth}} = \frac{\sum_{i=-p}^{p} w_i x_{t+i}}{\sum_{i=-p}^{p} w_i}, \tag{8}$$

where w_i denotes the smoothing weight, which decreases with increasing distance from the central point. The smoothing process effectively eliminates transient spikes and random disturbances while preserving the dominant trend characteristics of temperature variation.

Normalization and Data Partitioning: To eliminate dimensional differences among variables and accelerate model training, a Min–Max linear normalization method is applied to map all data values into the range of [−1, 1], as expressed by:

$$x_i^{'} = 2 \times \frac{x_i - x_{\min}}{x_{\max} - x_{\min}} - 1, \tag{9}$$

where $x_{\min}$ and $x_{\max}$ represent the minimum and maximum values of each variable, respectively. The normalized data preserve the relative variation among variables, preventing those with larger magnitudes from dominating the model optimization process.

Through the above preprocessing procedures, the raw process data from the acid-making system are significantly improved in terms of completeness, smoothness, and consistency. The preprocessed dataset exhibits good stationarity and continuity, accurately reflecting the dynamic evolution characteristics of the copper smelting acid-making process. This ensures the availability of high-quality input data for subsequent model training and prediction. In summary, the preprocessing

strategy follows an end-to-end flow from “variable screening - anomaly handling - missing-value compensation and alignment - smoothing – normalization”, which ensures that the final dataset is physically reasonable and numerically stable for the subsequent PACF analysis and recurrent neural network training.

3.2. Time-delay feature identification

In the sulfuric acid conversion process of copper smelting off-gas, the inlet temperature of the first catalyst bed in the converter is influenced by the combined effects of historical operating conditions, upstream heat-exchange delays, and the cumulative exothermic reactions. Such multivariable and multi-time-delay dynamic characteristics pose considerable challenges for direct modeling. To reasonably determine the temporal dependency length of input sequences and to identify the significant time-delay relationships between temperature and its influencing factors, we introduce the PACF for time-delay analysis and feature selection. In this work, the PACF is used to reveal how many minutes of historical information significantly contribute to the current temperature, thereby characterizing the time-lag behavior of temperature variations and providing a practical basis for selecting the input time steps of the model.

The PACF is an important statistical measure in time series analysis, used to evaluate the true correlation between lag orders of a variable after eliminating the influence of intermediate lags. Unlike the conventional Autocorrelation Function (ACF), the PACF can effectively remove the indirect effects transmitted through intermediate variables, thereby more accurately reflecting the direct dependency between a specific lag and the target variable. Consequently, the PACF can efficiently identify the dominant memory depth of temperature dynamics and provide a quantitative basis for determining the time-window length L in the subsequent GRU network. More importantly, the identified memory depth corresponds to the effective historical horizon required for prediction: An overly short window may miss delayed thermal effects, while an excessively long window may introduce redundant information and accumulate noise, reducing training efficiency.

Assuming that the time series $\{x_t\}$ is a stationary process, the partial autocorrelation coefficient at lag k can be obtained using the following recursive relationship:

$$\phi_{kk} = \frac{r_k - \sum_{j=1}^{k-1} \phi_{k-1,j} r_{k-j}}{1 - \sum_{j=1}^{k-1} \phi_{k-1,j} r_j} \tag{10}$$

$$\phi_{kj} = \phi_{k-1,j} - \phi_{kk} \phi_{k-1,k-j}, \quad j = 1, 2, \ldots, k-1 \tag{11}$$

$$r_k = \frac{\mathrm{Cov}(x_t, x_{t-k})}{\mathrm{Var}(x_t)}, \tag{12}$$

where ϕ_k denotes the partial autocorrelation coefficient at lag k; r_k represents the sample autocorrelation coefficient; and $\mathrm{Cov}(x_t, x_{t-k})$ and $\mathrm{Var}(x_t)$ denote the sample covariance and variance, respectively.

The PACF results are typically visualized as a plot of partial autocorrelation coefficients versus lag order. When the partial autocorrelation coefficient at a certain lag exceeds the confidence interval, it indicates that the lag has a significant influence on the target variable. By analyzing the statistical significance distribution of the PACF, the temporal dependency range between the converter inlet temperature and the major process variables can be identified. Therefore, the PACF provides an objective criterion to determine whether historical temperature (or an input variable) at time $t-k$ has a direct contribution to the current temperature at time t.

In the modeling process, the PACF distributions between each input variable and the first-bed inlet temperature sequence of the converter are computed sequentially. The significant lag orders are then used to determine the time-window length for the GRU model inputs. If the PACF rapidly decays to zero after lag k, it implies that the system mainly depends on the most recent k time steps. Conversely, if the PACF remains significant over a wider lag range, it suggests that the system exhibits strong long-memory characteristics and requires a longer modeling window to capture the dynamic information. Based on the selected window length L, the model input at time t is constructed by stacking the historical observations within $[t-L+1, \ldots, t]$ for all selected variables, so that delayed effects from heat exchange and reaction heat release can be embedded into the temporal input structure.

The PACF analysis results reveal that, in the sulfuric acid conversion process of copper smelting off-gas, the temperature sequence typically exhibits significant lags within 6-10 orders, corresponding to an approximate response delay of 6-10 minutes. This observation is consistent with the practical durations of gas flow, heat exchange, and catalytic reactions in the process. Therefore, the upper bound of the significant lag range obtained from PACF analysis is selected as the sliding-window length L for GRU inputs, ensuring that the model effectively captures the major dynamic features of temperature variations while avoiding computational redundancy and noise accumulation caused by excessively long windows. This setting also makes the chosen time steps interpretable from a process perspective, linking the data-driven window selection to the physical delay range of the industrial system.

In summary, the Partial Autocorrelation Function provides a data-driven approach for identifying time-delay features in this study. It enables an objective quantification of the lag effects of various process variables on the converter inlet temperature, thereby guiding the design of the GRU model's input structure along the temporal dimension. As a result, the predictive model is endowed with sufficient historical information while maintaining high training efficiency and strong generalization capability. As a result, the input sequence length is determined in a principled and interpretable manner, balancing information sufficiency and computational efficiency for subsequent temporal modeling.

3.3. Converter temperature dynamic feature extraction mechanism

To model the temporal dependencies of the industrial process variables, we adopt a GRU network as the core sequence-learning module. The major reasons are as follows: The GRU uses a relatively simple gating structure and therefore requires fewer parameters than more complex recurrent architectures, which helps reduce the risk of overfitting when the amount of labeled industrial data is limited; it also improves training efficiency and convergence stability, making it suitable for real-time or near-real-time deployment requirements in industrial systems. In addition,

the gating mechanism can effectively retain long-range temporal information while suppressing the influence of noise and time-delay-induced fluctuations, which matches the characteristics of the flue gas sulfuric acid conversion process data.

The GRU is an improved variant of the RNN. Compared with conventional RNNs, the GRU introduces gating mechanisms into its architecture to mitigate the vanishing gradient problem and effectively learn long-term dependencies. The structure of the GRU is illustrated in Figure 2. Unlike the LSTM network, the GRU removes the explicit memory cell state and retains only two gating units: the update gate and the reset gate. This simplification reduces the number of parameters and computational complexity while maintaining comparable or superior performance, which is particularly beneficial for ensuring model stability in industrial process modeling.

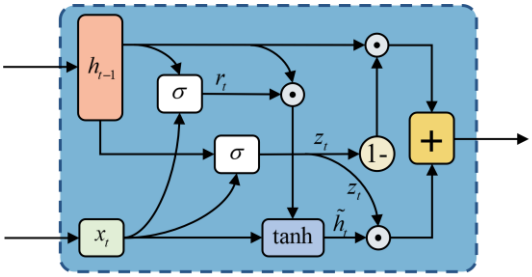

Figure 2. Schematic diagram of the GRU network architecture.

In the modeling framework of this study, the GRU is employed to learn the dynamic evolution patterns of the converter temperature sequence over time. The core concept of the GRU lies in regulating the information flow along the temporal dimension through its gating units, thereby achieving a balance between long-term dependency modeling and short-term disturbance responsiveness. The structure of the GRU is illustrated in Figure 2, where σ denotes the Sigmoid activation function, and tanh represents the hyperbolic tangent function; both are applied in an element-wise manner.

The mathematical formulation of the GRU can be expressed as follows:

$$\begin{cases} z_t = \sigma(W_z[u_t, h_{t-1}] + b_z) \\ r_t = \sigma(W_r[u_t, h_{t-1}] + b_r) \\ \tilde{h}_t = \tanh(W_h[u_t, (r_t \circ h_{t-1})] + b_h) \\ h_t = (1 - z_t) \circ h_{t-1} + z_t \circ \tilde{h}_t \end{cases}, \tag{13}$$

where u_t denotes the input vector at time t, corresponding to process variables such as flue gas temperature, SO_2/O_2 concentration, heat exchanger outlet temperature, flow rate, and valve opening. h_t represents the hidden state vector at time t, which encapsulates the historical information of the input sequence along the temporal dimension. W_z, W_r, and W_h are the weight matrices associated with the update gate, reset gate, and candidate state, respectively, while b_z, b_r, and b_h denote the corresponding bias terms. z_t is the update gate vector, which controls the proportion of the previous hidden state h_{t-1} and the new candidate information $\tilde{h}_t$ contributing to the current hidden state. r_t is the reset gate vector that determines how much of the past information should be forgotten when generating the candidate state $\tilde{h}_t$. The symbol "$\circ$" denotes the element-wise (Hadamard) product

operation.

According to Eq. (13), the GRU network can automatically determine whether to retain or discard historical information based on sequential data, thereby enabling adaptive learning of the dynamic variation trends of the converter inlet temperature. Unlike conventional feed-forward neural networks, the GRU propagates its hidden states continuously over time, enabling it to capture multi-scale temperature dynamics induced by heat-exchange delays, flue gas disturbances, and the exothermic heat of catalytic reactions.

In the sulfuric acid conversion process, the stability of the inlet temperature is directly related to the reaction rate and equilibrium conversion of $SO_2 \rightarrow SO_3$. By learning the implicit mapping relationships between historical temperature sequences and process variables, the GRU model can predict the time-delay response of temperature following changes in input signals, thereby modeling and forecasting the dynamic evolution of the thermal balance.

Furthermore, to enhance the model's nonlinear representation capacity and training stability, dropout regularization is introduced within the GRU layers to prevent overfitting. In addition, a stacked multi-layer GRU architecture is adopted to extract higher-level dynamic features. The final hidden state h_t produced by the GRU serves as the input to the subsequent Attention Mechanism Layer, which extracts temporal features most relevant to the target temperature.

In summary, the GRU network serves not only as the core component for time-series modeling in this study but also as a critical bridge connecting the dynamic characteristics of process data with the temperature prediction task. Through its compact gating architecture, the GRU efficiently represents the complex, nonlinear, and time-varying thermodynamic behavior inherent in the copper smelting off-gas sulfuric acid process, thereby providing a stable temporal representation foundation for subsequent attention-based weighting and soft-measurement prediction.

3.4. Key timing feature focusing mechanism

In the sulfuric acid conversion process of copper smelting off-gas, the variation in the first-bed inlet temperature of the converter is influenced not only by the coupling of multiple variables but also exhibits pronounced time-varying and nonlinear characteristics. The contribution of input variables to temperature fluctuations differs across operating periods. For instance, during smelting condition transitions, the flue gas flow rate and heat exchange temperature act as dominant factors, whereas during steady-state operation, the SO_2 concentration and oxygen ratio play more significant roles.

To enable the model to adaptively recognize and emphasize the varying importance of different temporal segments and feature dimensions, an AM is incorporated into the GRU network.

3.4.1. The principle of attention mechanics

The attention mechanism was originally developed in the field of Natural Language Processing (NLP). Its core concept is that, among a large amount of input information, the model should not treat all time steps equally; instead, it should assign different weights according to the task objective, thereby focusing on the information most relevant to the output.

In time series modeling, the attention mechanism dynamically adjusts the contribution of different temporal segments by computing the correlation between the hidden states of each input

time step and the target output. This enables the model to enhance key temporal features while suppressing irrelevant information.

Let the sequence of hidden states encoded by the GRU be denoted as:

$$H = [h_1, h_2, \ldots, h_T], \quad h_t \in \mathbb{R}^d , \tag{14}$$

where h_t denotes the hidden state at time t, and T represents the length of the time window. The attention mechanism computes a relevance score e_t between each hidden state and the current output target, which is then normalized to obtain the attention weight α_t. The calculation is expressed as follows:

$$e_t = v_a^T \tanh(W_a h_t + b_a) \tag{15}$$

$$\alpha_t = \frac{\exp(e_t)}{\sum_{i=1}^{T} \exp(e_i)} \tag{16}$$

$$c_t = \sum_{i=1}^{T} \alpha_i h_i , \tag{17}$$

where W_a denotes the trainable weight matrix, v_a represents the attention vector, and b_a is the bias term. α_t denotes the attention distribution, which reflects the attention intensity of the model at time step t. c_t is the context vector, obtained through a weighted summation, representing the information within the historical sequence that is most relevant to the current output.

From Eq. (15)-(17), it can be observed that the attention mechanism automatically assigns a weight to each time step during the modeling process. When the features at a particular moment are more critical for predicting the converter temperature, the corresponding attention coefficient α_t increases significantly, thereby contributing more strongly to the final prediction. Conversely, time steps containing less relevant or redundant information receive lower weights, which helps suppress noise interference and improve the robustness of the prediction.

3.4.2. Role in converter temperature prediction

Applying the attention mechanism to temperature prediction in the copper smelting off-gas sulfuric acid process helps reveal the dominant temporal segments and key driving features underlying temperature variations. In this system, the dynamic behavior of the flue gas temperature is typically governed by the interaction of the following factors:

Short-term disturbance characteristics: Fluctuations in flow rate caused by changes in blower guide vane positions or valve operations lead to instantaneous temperature variations.

Medium-term lag characteristics: The heat accumulation effect in the heat exchanger introduces delayed responses of the upstream gas temperature to the converter inlet temperature.

Long-term equilibrium characteristics: Variations in SO_2 and O_2 concentrations adjust the intensity of exothermic reactions, thereby influencing the overall thermal balance.

Traditional GRU models tend to process all historical information uniformly across time, making it difficult to emphasize critical moments that dominate temperature evolution. By incorporating the attention mechanism, the model can automatically learn to identify and emphasize historical periods that contribute most to temperature fluctuations at different operational stages. For instance, during sudden load changes, the model strengthens the feature weights of recent time steps, whereas under steady-state conditions, it allocates greater attention to slowly varying variables related to thermal equilibrium, thus achieving adaptive focusing on dynamic features.

Another advantage of the attention mechanism lies in its interpretability. By analyzing the temporal distribution of the learned attention weights α_t, one can intuitively observe which time segments and process variables the model focuses on during temperature prediction. For example, when a sudden increase in flue gas flow occurs, the attention weights tend to concentrate on the previous 3–5 times steps of the heat exchanger outlet temperature and flow signals, consistent with the operational insight that "flow variations lead to rapid temperature rise." In contrast, during catalyst activity transitions, the model's attention may shift toward SO_2 concentration signals over longer intervals, reflecting the delayed effect of reaction heat accumulation. This attention visualization not only validates the model's reliability but also provides interpretable insights for operators, supporting a transition from black-box prediction to explainable intelligent forecasting.

3.4.3. Synergy with GRU

In the model architecture, the attention mechanism is embedded after the GRU layer. The GRU is responsible for extracting the dynamic temporal features of the input sequence, while the attention mechanism performs weighted integration of these hidden states, thereby achieving a fusion of local and global features. This hybrid configuration preserves the GRU's capability for temporal dependency modeling and, at the same time, enables dynamic and flexible feature selection through the attention weight allocation mechanism. The resulting context vector c_t is then fed into a fully connected regression layer to generate the predicted value of the converter's first-bed inlet temperature.

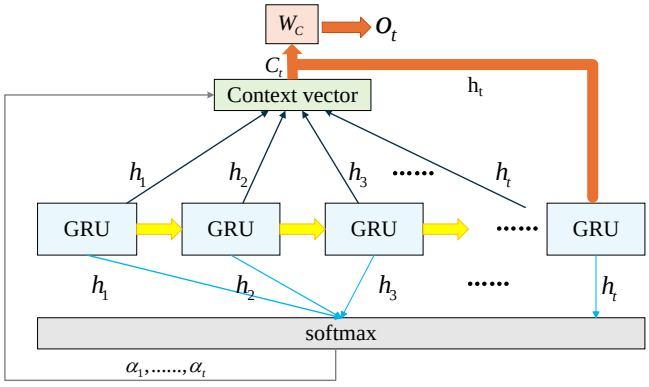

Figure 3. Workflow of the global attention GRU network.

In summary, the attention mechanism in this study not only improves prediction accuracy but also significantly enhances the model's interpretability for complex smelting processes. From an engineering perspective, the attention weights enable the predictor to automatically emphasize the

most informative historical moments (such as periods immediately following smelting disturbances, converter blowing-period changes, or valve switching), so the model relies less on noisy or low-relevance time steps and produces more stable temperature forecasts.

3.5. Comprehensive modeling framework for dynamic temperature prediction

To achieve accurate prediction of the first-bed inlet temperature in the sulfuric acid conversion process of copper smelting off-gas, we propose an integrated dynamic modeling approach based on the Partial Autocorrelation Function, a Gated Recurrent Unit network, and an attention mechanism. The proposed framework combines statistical time-delay analysis and sequence learning, enabling effective representation of the nonlinear and strong time-delay characteristics inherent in the acid production process. As illustrated in Figure 4, the model is organized as a clear end-to-end pipeline, and each block corresponds to a practical function in the temperature forecasting task.

(1) Temporal input construction guided by time-delay analysis. First, the Partial Autocorrelation Function is employed to analyze the lag dependence between the target temperature and candidate process variables. The statistically significant lag range is used to determine the input window length L, which defines how many historical minutes are included for prediction. With the selected variables and window length, the multivariate sequence is organized into a temporal input matrix at time t by stacking historical observations within $[t-L+1,\ldots,t]$. This step ensures that the model input explicitly contains the dominant delayed effects induced by upstream heat exchange and reaction heat accumulation while avoiding redundant history that may introduce noise and unnecessary computation.

(2) Dynamic feature extraction by the stacked Gated Recurrent Unit network. The constructed input matrix is then fed into a stacked Gated Recurrent Unit network. The gated recurrent structure propagates state information along the temporal dimension and extracts nonlinear dynamic coupling relationships among variables. In practice, this module converts raw multivariate signals into a sequence of hidden representations that summarize both short-term disturbances (e.g., rapid fluctuations) and slower thermal-evolution patterns (e.g., heat-balance transitions). Therefore, it provides a compact temporal feature basis for subsequent temperature regression.

(3) Key-temporal information aggregation by the attention layer and regression output. On top of the Gated Recurrent Unit hidden states, an attention mechanism is introduced to perform weighted aggregation. By adaptively assigning attention weights, the model emphasizes the most informative time steps for the current prediction and suppresses less relevant or noisy segments. From an engineering perspective, this means the predictor can automatically focus on critical periods that dominate temperature evolution (for example, around disturbances or switching events), thereby improving stability of the temperature forecast. The attention-weighted context representation is finally passed to a fully connected regression layer to output the predicted first-bed inlet temperature.

In summary, Figure 4 demonstrates the complete data flow of the proposed model: The Partial Autocorrelation Function determines an interpretable and data-driven time window for input construction, the stacked Gated Recurrent Unit network extracts multiscale temporal dynamics, and the attention mechanism highlights the most useful historical information before regression. This modular design improves the readability of the modeling logic and the practicality of deployment in industrial temperature prediction scenarios.

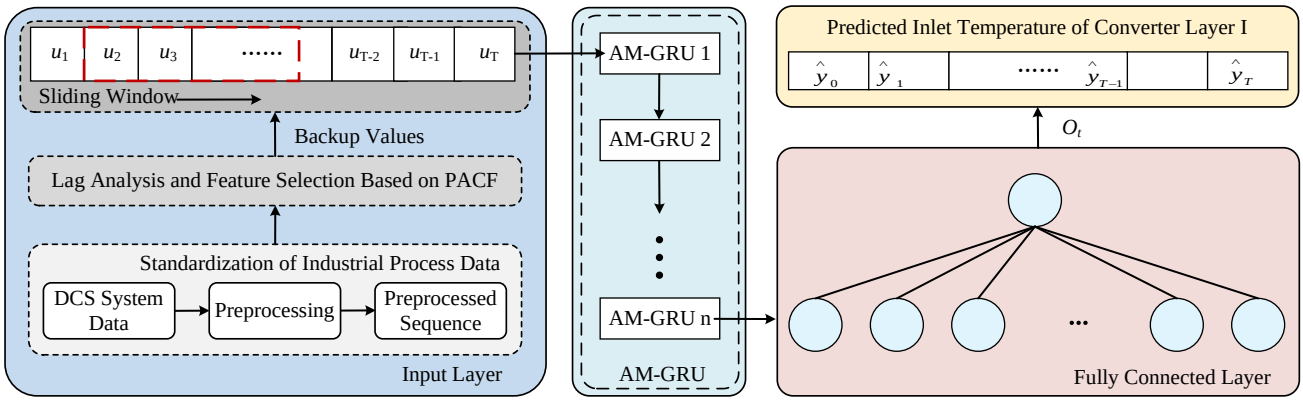

Figure 4. Schematic diagram of the converter layer I inlet temperature prediction model.

4. Experimental verification

4.1. Data source and preprocessing

In this section, we present a case study using the flue gas sulfuric acid system of a smelter in China as the research object. Figure 5 shows a photograph of the converter section in this system. The experimental object is the conversion process of the flue gas sulfuric acid plant, and the data are collected from the on-site distributed control system. The flue gas sulfuric acid process is affected by the intermittent air supply of the smelting operation, and each smelting cycle lasts approximately eight hours. Therefore, the sampling interval is uniformly set to one minute, and a total of 24,000 samples of process parameters recorded by various sensors in the conversion section are obtained. To avoid information leakage, the data are divided chronologically into training, validation, and test sets with a ratio of 7:1:2, ensuring that the model evaluation is consistent with the real-time online prediction scenario. Since the complete dataset spans multiple smelting cycles and includes steady-state segments and fluctuating operating conditions, such as smelting-induced disturbances, changes in converter blowing periods, valve switching, flash smelting furnace inspection, and unexpected events, all three subsets naturally cover stable and dynamic operating conditions.

Figure 5. Copper smelter flue gas acid system.

To comprehensively evaluate the predictive performance of the proposed model, three commonly used metrics are adopted in this study: The Root Mean Square Error (RMSE), the Mean Absolute Percentage Error (MAPE), and the Coefficient of Determination (R^2), as shown in Eq. (18) ~ (20).

Root Mean Square Error (RMSE): Reflects the average magnitude of deviation between the predicted and actual values.

Mean Absolute Percentage Error (MAPE): Measures the relative error level in percentage terms.

Coefficient of Determination (R^2): Indicates the goodness of fit of the model, where a value closer to 1 represents better predictive accuracy.

$$\mathrm{RMSE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(\tilde{y}_i - y_i)^2} \tag{18}$$

$$\mathrm{MAPE} = \frac{100\%}{n}\sum_{i=1}^{n}\left|\frac{\tilde{y}_i - y_i}{y_i}\right| \tag{19}$$

$$\mathrm{R}^2 = 1 - \frac{\sum_{i=1}^{n}(\tilde{y}_i - y_i)^2}{\sum_{i=1}^{n}(\tilde{y} - y_i)^2} \tag{20}$$

4.2. Time-delay characteristics and step-size determination

In the dynamic prediction of the converter's first-bed inlet temperature, temperature variations are influenced not only by the current operating conditions but also closely related to the flue gas composition and flow fluctuations generated during the smelting process. To quantify this temporal dependency, the results are illustrated in Figure 6. It can be observed that the PACF values for the first seven lag orders significantly exceed the 95% confidence interval, indicating a strong linear dependence between these historical time steps and the current converter inlet temperature. Beyond the seventh lag, the PACF coefficients gradually decay within the confidence bounds, suggesting that historical information over longer time spans contributes little to the prediction.

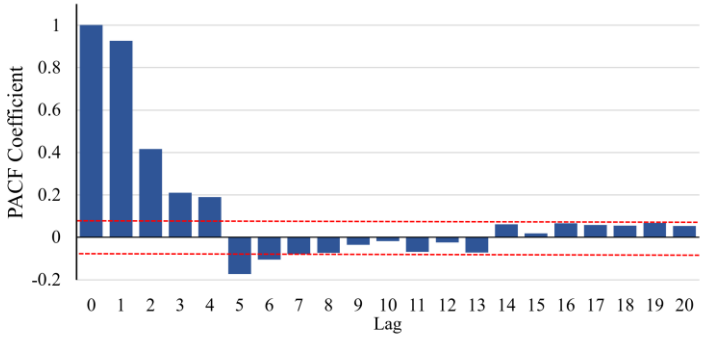

Figure 6. PACF of the converter layer inlet temperature.

Considering the model's ability to capture historical dependencies and the complexity of training, the time-window length (look-back value) is set to 7 in this study. In other words, the model uses the multivariate features from the previous seven sampling points as inputs for predicting the current temperature. This configuration effectively reflects the system's dynamic inertia while avoiding overfitting caused by redundant information, thereby improving the model's stability and generalization performance.

4.3. Hyperparameter optimization strategy and optimal results

To further enhance the predictive performance of the model, a systematic hyperparameter optimization is conducted for the GRU-AM-based network architecture. Given the high structural and parametric complexity of such nonlinear models, a hybrid optimization strategy combining coarse-grained Grid Search and Bayesian Optimization is adopted to efficiently explore the parameter space under limited computational resources. Considering the architectural characteristics of the GRU-AM model, the key hyperparameters optimized in this study include the number of network layers, number of hidden units, learning rate, dropout ratio, and batch size. The search ranges are summarized in Table 1.

Table 1. Hyperparameters.

Hyper-parameter	Range	Best
Dropout rate	{0.1, 0.15, 0.2, 0.25,0.3}	0.15
Stacked units number	{1, 2, 3, 4}	3
Hidden state number	{2, 4, 8, 16, 32, 64}	16, 64, 32

4.4. Model performance comparison and result analysis

To verify the effectiveness and superiority of the proposed model for predicting the first-bed inlet temperature of the converter, three representative benchmark models are selected for comparison.

RNN: Equipped with a basic recurrent structure, capable of capturing short-term dependencies in time series, but prone to gradient vanishing or explosion problems when dealing with long sequences.

LSTM: Introduces forget and memory gates to improve the modeling of long-term dependencies, representing a classical model for traditional time-series prediction.

Multilayer Perceptron (MLP): A multilayer feedforward neural network used to evaluate the static prediction performance without temporal feature modeling.

All four models are trained using the same input features, training samples, and data partitioning strategy, with the time-window length $L = 7$ determined in Section 4.2 as a unified input structure. For fair comparison, the initial learning rate is set to 0.002, the batch size to 64, and the number of training epochs is kept identical across all models.

Figure 7 presents the comparison between the predicted and measured values of the converter's first-bed inlet temperature over one smelting cycle. From the time-series comparison curves, it can be observed that the proposed model accurately tracks the temperature variation trends of the

converter inlet under different fluctuation amplitudes and disturbance conditions. The predicted outputs closely fit the measured values, with concentrated residual distributions and relatively small oscillation ranges.

In contrast, the MLP model exhibits noticeable lag and underfitting when handling dynamic temporal features, showing limited sensitivity to rapid temperature fluctuations. The RNN model possesses basic temporal modeling capability, but suffers from gradient attenuation in capturing long-term dependencies, resulting in an excessively smoothed prediction curve. Although the LSTM model outperforms RNN and MLP overall, its response to high-frequency disturbances shows a slight time delay.

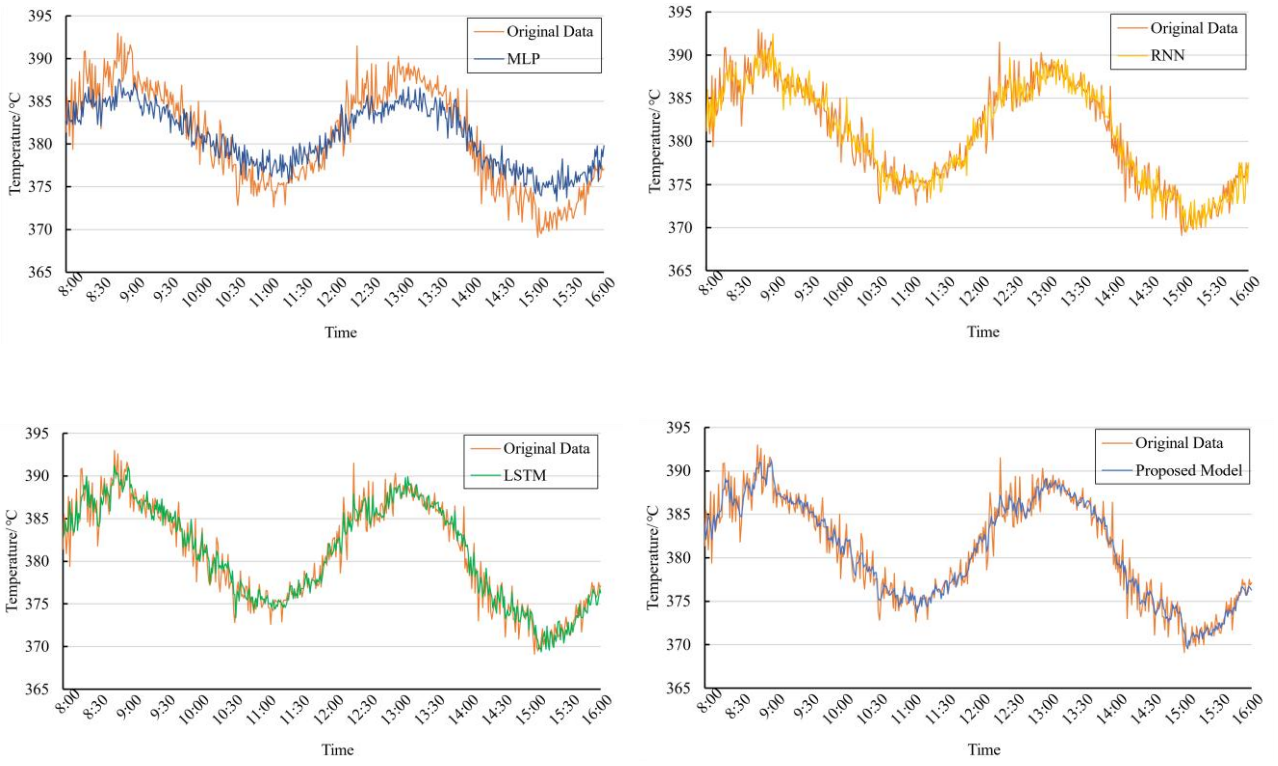

Figure 7. Comparison of prediction results of each model.

To provide a more comprehensive evaluation beyond the test-set example shown in Figure 7, we additionally visualize the prediction performance on the training, validation, and test subsets in Figure 8. In this figure, the Original Data and the corresponding model outputs are plotted along the same time axis, while the split boundaries are indicated by vertical dashed lines. Moreover, we report the prediction accuracy (R^2) for each subset to summarize the overall goodness-of-fit on the training set and the generalization performance on the validation and test sets.

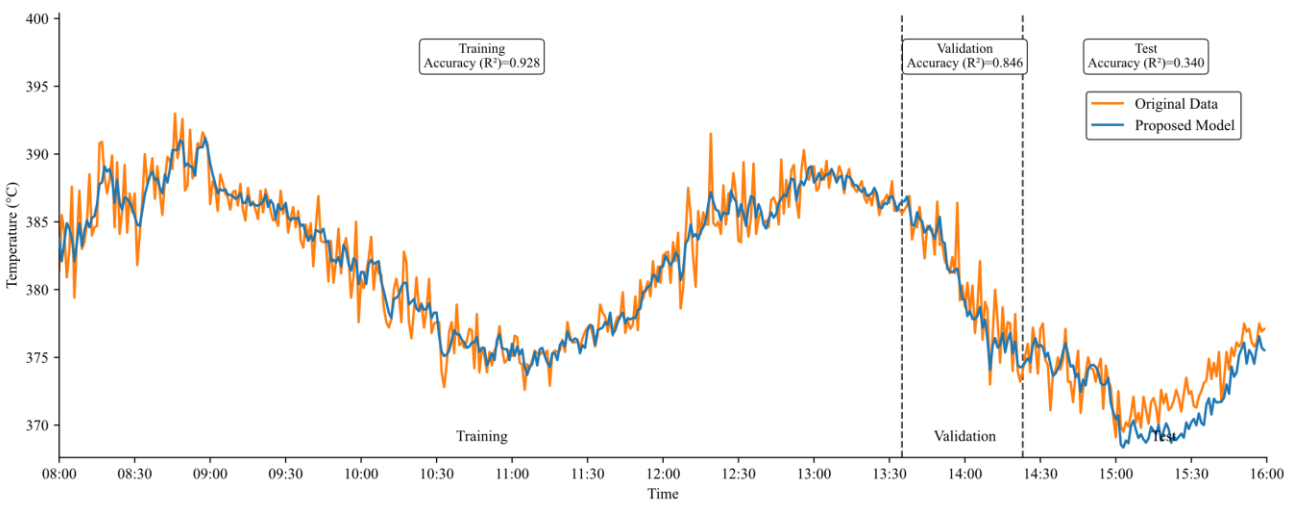

Figure 8. Prediction performance of the proposed model on the training, validation, and test sets.

Table 2 presents a comparison of the prediction results of different models on the test dataset. As shown in the results, the proposed model achieves the best performance across all three error metrics, with the lowest RMSE and MAPE values and the highest R^2. This indicates that the predicted temperatures from the proposed model exhibit the highest degree of agreement with the actual measurements, demonstrating superior accuracy and stability.

In comparison, the LSTM model performs relatively well in capturing temporal dependencies but shows slight lag during rapid fluctuation periods. The RNN model suffers from gradient attenuation due to long-sequence dependencies, leading to a smoother overall prediction curve that fails to fully capture nonlinear variations. The MLP model, lacking the ability to model temporal dependencies, produces larger errors and exhibits the widest fluctuation range among all models.

Table 2. Statistical results of prediction errors for different models.

Model Name	RMSE	MAPE	R^2
MLP	6.241	1.931	0.942
RNN	5.312	1.675	0.952
LSTM	4.276	1.509	0.959
Proposed Model	2.913	1.47	0.962

Figure 9 illustrates the residual distributions of the four prediction models for the converter inlet temperature forecasting task. As shown in the figure, the residuals of the proposed PACF-GRU-Attention model are the most concentrated, exhibiting an approximately symmetric unimodal distribution with most residuals clustered near zero. This indicates that the proposed model achieves small prediction errors, limited fluctuation ranges, and stable accuracy under different operating conditions.

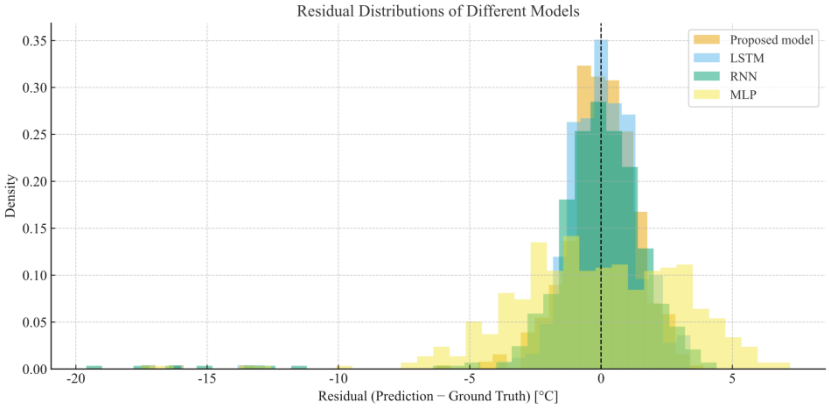

Figure 9. Prediction model residuals.

In contrast, the LSTM model shows a slightly wider residual distribution with mild positive and negative shifts, suggesting a certain degree of response delay during high-fluctuation periods. The residuals of the RNN model are more dispersed, revealing its limited ability to model long-term dependencies. The MLP model exhibits the widest and flattest residual distribution, indicating poor temporal feature extraction and an obvious underfitting tendency. Overall, the proposed model demonstrates a residual distribution closest to the ideal state, reflecting higher predictive accuracy and stability, and providing reliable technical support for dynamic temperature prediction in industrial smelting processes.

In summary, under the current experimental setting and a consistent training configuration, the proposed soft-sensing model integrating PACF-based feature selection, GRU-based temporal modeling, and an attention mechanism achieves improved predictive performance compared with the considered benchmark models. The PACF helps determine an effective input time window and relevant variables, the GRU captures multivariate temporal dependencies with moderate complexity, and the attention mechanism enhances sensitivity to key temporal segments during fluctuating phases. Nevertheless, it should be noted that the comparative results are obtained with a limited set of baseline models and a fixed hyperparameter setting, and other methods may achieve stronger performance after more extensive optimization. Moreover, this generalization evaluation is mainly based on a chronological split of data collected from the same industrial system; therefore, additional validation across plants, operating regimes, and longer time horizons is required.

5. Conclusions

In this study, a data-driven prediction model integrating the PACF, a GRU network, and an AM is proposed for forecasting the first-bed inlet temperature of the converter in the copper smelting off-gas sulfuric acid process. The proposed framework combines statistical time-delay analysis with temporal sequence learning to capture nonlinear dynamics and operating fluctuations. Industrial data from a “two-conversion, two-absorption” system demonstrate that, under a consistent experimental setup, the proposed model provides lower prediction errors and more concentrated residual distributions than the selected baseline models. These results suggest that the proposed framework is a practical and interpretable solution for on-site temperature prediction and has potential value for operational monitoring and process adjustment.

Limitations and future work: First, the benchmark comparison in this work includes a limited number of models and does not involve an exhaustive hyperparameter search for each baseline; therefore, the reported superiority should be interpreted within the current tuning effort and experimental scope. Second, the validation is mainly conducted using data from a single plant with a chronological split; broader generalization across plants, equipment conditions, and long-term distribution shifts remains to be verified. Third, we do not explicitly quantify prediction uncertainty or investigate robustness under severe sensor faults and communication delays. In future work, we will focus on expanding cross-site datasets, incorporating more competitive and carefully tuned baselines, developing online adaptation strategies to handle drift, and introducing uncertainty estimation to improve decision support in real industrial applications.

Author contributions

Chunbo Wang was responsible for investigation, validation, and drafting the original manuscript. Xiaoli Li contributed to conceptualization, methodology, investigation, and validation. Kang Wang contributed to formal analysis, project administration, and supervision. All authors reviewed and approved the final version of the manuscript.

Use of Generative-AI tools declaration

The authors declare they have not used Artificial Intelligence (AI) tools in the creation of this article.

Acknowledgments

This study is supported by the National Natural Science Foundation of China (61873006).

Conflict of interest

The authors declare no conflicts of interest in this paper.

References

1. Cui H, Cao Y (2023) How can market-oriented environmental regulation improve urban energy efficiency? Evidence from quasi-experiment in China's SO2 trading emissions system. *Energy* 278: 127660. https://doi.org/10.1016/j.energy.2023.127660
2. Zhang T, Peng H, Wu C, Guo Y, Wang J, Chen X, et al. (2023) Process compatible desulfurization of NSP cement production: A novel strategy for efficient capture of trace SO2 and the industrial trial. *J Clean Prod* 411: 137344. https://doi.org/10.1016/j.jclepro.2023.137344
3. Córdoba P, Rojas S (2024). Carbon sequestration through mineral carbonation: Using commercial FGD-gypsum from a copper smelter for sustainable waste management and environmental impact mitigation. *J Environ Chem Eng* 12: 112510. https://doi.org/10.1016/j.jece.2024.112510

4. Li L, Zhang X, Lian X, Zhang L, Zhang Z, Liu X, et al. (2025) Flue gas desulfurization and SO2 recovery within a flexible hydrogen-bonded organic framework. *Nature Chemistry* 17: 727–733. https://doi.org/10.1038/s41557-025-01744-9
5. Zhou Z, Lu J, Ma X (2024) SO3/sulfuric acid mist removal in simulated flue gas: Multi-factor study based on two-film theory mass transfer process. *Fuel* 357: 129698. https://doi.org/10.1016/j.fuel.2023.129698Get rights and content
6. Li X, Han J, Liu Y, Dou Z, Zhang T (2022) Summary of research progress on industrial flue gas desulfurization technology. *Sep Purif Technol* 281: 119849. https://doi.org/10.1016/j.seppur.2021.119849
7. Wang X, Li H (2018) Soft sensor based on stacked auto-encoder deep neural network for air preheater rotor deformation prediction. *Adv eng inform* 36: 112–119. https://doi.org/10.1016/j.aei.2018.03.003
8. Weng Q, Tian X, Wang H, Wu X, Wang S, Zhuo Y, et al. (2022) Comprehensive effect of oxidant addition in an FGD slurry on the removal and distribution of selenium: a field study. *Environ Sci Technol* 56: 3544–3551. https://pubs.acs.org/doi/10.1021/acs.est.1c07708
9. Qiao Z, Wang X, Gu H, Tang Y, Si F, Romero C, et al. (2019) An investigation on data mining and operating optimization for wet flue gas desulfurization systems. *Fuel* 258: 116178. https://doi.org/10.1016/j.fuel.2019.116178
10. Hou P, Bai J, Yin J (2013) On-line monitoring and optimization of performance indexes for limestone wet desulfurization technology. *Applied Mechanics and Materials* 295: 1020–1028. https://doi.org/10.4028/www.scientific.net/AMM.295-298.1020
11. Li X, Dong J, Wang K (2023) Constrained nonlinear model predictive control of pH value in wet flue gas desulfurization process. *Optim Contr Appl Meth* 44: 1523–1539. https://doi.org/10.1002/oca.2763
12. Darake S, Hatamipour M, Rahimi A, Hamzeloui P (2016) SO2 removal by seawater in a spray tower: Experimental study and mathematical modeling. *Chem Eng Res Des* 109: 180–189. https://doi.org/10.1016/j.cherd.2015.11.027
13. Zhao Z, Zhang Y, Gao W, Baleta J, Liu C, Li W, et al. (2021) Simulation of SO2 absorption and performance enhancement of wet flue gas desulfurization system. *Process Saf Environ Prot* 150: 453–463. https://doi.org/10.1016/j.psep.2021.04.032
14. Liu C, Zhao Z, Gao W, Baleta J, Li W, Li Q, et al. (2021). Process optimization of S (IV) oxidation in flue gas desulfurization scrubbers. *Process saf environ prot* 149: 610–618. https://doi.org/10.1016/j.psep.2021.03.020
15. Xin J, Zhou C, Jiang Y, Tang Q, Yang X, Zhou J (2023) A signal recovery method for bridge monitoring system using TVFEMD and encoder-decoder aided LSTM. *Measurement* 214: 112797. https://doi.org/10.1016/j.measurement.2023.112797
16. Ren C, An N, Wang J, Li L, Hu B, Shang D (2014) Optimal parameters selection for BP neural network based on particle swarm optimization: A case study of wind speed forecasting. *Knowledge-based systems* 56: 226–239. https://doi.org/10.1016/j.knosys.2013.11.015
17. Yuan X, Li L, Shardt Y, Wang Y, Yang C (2020) Deep learning with spatiotemporal attention-based LSTM for industrial soft sensor model development. *IEEE T Ind Electron* 68: 4404–4414. https://doi.org/10.1109/TIE.2020.2984443

18. Lui C, Liu Y, Xie M (2022) A supervised bidirectional long short-term memory network for data-driven dynamic soft sensor modeling. *IEEE T Instrum Meas* 71: 1–13. https://doi.org/10.1109/TIM.2022.3152856
19. Yuan X, Ou C, Wang Y, Yang C, Gui W (2019) A layer-wise data augmentation strategy for deep learning networks and its soft sensor application in an industrial hydrocracking process. *IEEE T Neur Net Lear Syst* 32: 3296–3305. https://doi.org/10.1109/TNNLS.2019.2951708
20. Yang Y, Yu H, Liu X (2022) A Variable Attention-Based Gated GRU Approach for Soft Sensor Models. In *2022 34th Chinese Control and Decision Conference (CCDC)*, 92–96. https://doi.org/10.1109/CCDC55256.2022.10033933

AIMS Press