A Spatial–Temporal Structural Estimation Model Based on GATE-PCGRU for Multirate Industrial Process

Bei Sun, Shengyu Liu, and Yonggang Li

***Abstract*— Real-time estimation of key performance indicators (KPIs) is crucial for the safe and cost-effective operation of an industrial process. Due to multiple sampling rates, spatial correlation of industrial topologies, and temporal correlation of distinct time intervals, the online estimation of KPIs is challenging. This study proposes a spatial–temporal structural multirate KPI estimation framework based on graph attention autoencoder and gated recurrent unit network with period control unit (GATE-PCGRU), which is capable of simultaneously capturing the spatial and temporal dependence. First, graph attention autoencoder (GATE) is employed to comprehensively extract the spatial topology correlation among process variables by reconstructing the data feature attributes. Second, a gated recurrent unit network with period control unit (PCGRU) is employed to extract the temporal correlation and learn how the process variables vary over time, achieving KPI estimation for different periods. Next, a combined spatial–temporal mechanism is presented to incorporate the extracted spatial and temporal correlations into the framework. The output of the GATE encoder is used as input to the PCGRU, and a loss function that combines the spatial and temporal losses is designed to fuse the two models for training. The proposed method is compared with existing methods in a multirate industrial process, obtaining R^2 and mean absolute percentage error (MAPE) of 0.731 and 0.132, respectively, which is improved compared with the existing methods.**

***Index Terms*— Deep learning, key performance indicator (KPI) estimation, multirate industrial process, spatial–temporal structural.**

I. Introduction

A. Background

THE key indicators of a production process reflect its stability and safety. They are crucial for the monitoring and decision-making processes of industrial operations [1], [2]. To maintain a steady rate of production in practice, personnel focus on monitoring key performance indicators (KPIs) and making modest adjustments based on prior knowledge and experience. However, some data are typically acquired by low-rate sampling (several hours apart) and chemical analysis using expensive equipment [3], [4]. This results in considerable delays and high costs, which impedes the improvement of product quality and operational safety. However, process operating variables, such as temperature and flow rates, of a process may be gathered via high-rate sampling sensors (a few minutes apart). The abovementioned data are typically obtained from different steps of the production process using different sampling intervals. The complexity of the industrial process, nonlinear relationships between key indicators and other process variables, and data information asymmetry [5], [6] pose as challenges for the real-time measurement of key indicators during production [7], [8], [9], [10], [11], [12]. We summarize the challenges facing the industry.

1) *Multiple Sampling Rates:* The majority of modeling techniques [13], [14], [15], [16] assume the same sample size for input and output variables. In practice, however, production lines contain multiple control and measurement points for complex industrial processes. Thus, the sampling rates are distinct, and the sampling intervals may vary due to workers' control. Fig. 1 shows an industry's multirate characteristics.
2) *Complex Industrial Structure:* Machine learning and deep learning, the prevalent estimate modeling methodologies, perform estimation tasks primarily by modeling the relationship between input–output data [17], [18], such as partial least squares (PLS) [19], slow feature analysis [20], artificial neural network (ANN) [21], and stacked autoencoder (SAE) [22], The majority of these methods assume a common cascade of readings or observations from all sensors as multivariate inputs of the model, while ignoring the underlying system that generates the data or explicit industrial structure. Nonetheless, in the majority of industrial installations, a complex industrial process has a well-defined and explicitly fixed structure with various interconnected units, which mutually influence each other.
3) *Complex Spatial–Temporal Correlations:* The spatial dependence of the industrial structure and the temporal dependence of multivariate data are key to addressing the complex industrial process problem. When designing the model, considering only the temporal or spatial dependence might lead to the omission of crucial information that influences estimation outcomes.

Manuscript received 8 January 2023; revised 14 May 2023; accepted 20 June 2023. Date of publication 3 August 2023; date of current version 8 August 2023. This work was supported in part by the National Key Research and Development Program of China under Grant 2022YFE0125000 and Grant 2020YFB1713701, in part by the Projects of International Cooperation and Exchanges of the National Natural Science Foundation of China under Grant 61860206014, and in part by the National Natural Science Foundation of China under Grant 61973321. The Associate Editor coordinating the review process was Dr. David Ayllon. *(Corresponding author: Yonggang Li.)*

The authors are with the School of Automation, Central South University, Changsha 410083, China (e-mail: sunbei@csu.edu.cn; 214611078@csu.edu.cn; liyonggang@csu.edu.cn).

Digital Object Identifier 10.1109/TIM.2023.3291796

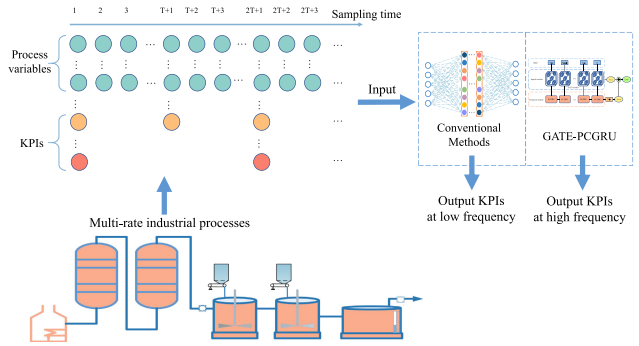


Fig. 1. Characteristics of variables in a multirate industrial process.

B. Related Work

First, for multiple sampling rates, a feasible option in this regard is using interpolative methods to replace missing variables with suitable values [23], [24], [25]. However, this strategy is overly dependent on the quality of the interpolation method used. Inaccurate interpolation can negatively affect the final outcome. In addition, multirate issues may be solved via semisupervised learning. Zhou et al. [26] presented a semisupervised probabilistic latent variable regression model with unequal sample sizes for process monitoring variables and quality variables. However, in some multirate industrial processes with scarce labels, the performance of the semisupervised model degraded considerably. Second, for complex industrial structures, graph neural networks (GNNs) have been demonstrated to be effective at extracting features from such node data [27], [28], [29]. Narwariya et al. [30] recommended capturing the structure of complex equipment as a graph and modeling it with GNNs. Wu et al. [31] employed GNNs to detect industrial anomalies. Third, for complex spatial–temporal correlations, Yang et al. [32] introduced a model combined with graph convolutional neural network and gated recurrent unit network (GCN-GRU) for traffic prediction that performs well. Consequently, this modeling approach may be applied to the KPI estimation for industrial processes. However, simply fusing spatial and temporal modules is difficult to ensure that spatial–temporal features are fully captured simultaneously. Therefore, it is challenging to establish a reasonable spatial–temporal fusion mechanism.

C. Contribution

Considering the abovementioned challenges, this study proposed a graph attention autoencoder and gated recurrent unit network with period control unit (GATE-PCGRU)-based spatial–temporal structural estimation model by combining the spatial and temporal features for the estimation of key quality indicators in multirate industrial processes. The spatial- and temporal-dependent modules were modeled separately; subsequently, a spatial–temporal fusion mechanism was used to construct the network. GATE-PCGRU addresses the three challenges above with the following improvements.

1) Instead of using interpolation or semisupervised methods for multirate industrial processes, a gated recurrent unit network with period control unit (PCGRU) is proposed. It adds a period control unit to the gated recurrent unit network (GRU), which learns the trend of the process variables over time and captures the relationship between the process variables and the KPIs. If the dynamics are learned, then the estimation period can be controlled by adjusting the parameter ΔT.
2) In a complex industrial structure, PCGRU can extract the temporal features while losing the spatial features. As a result, the graph attention network (GAT) is introduced to facilitate the model simultaneously capturing the spatial and temporal dependences.
3) The combined model with direct splicing of GAT and PCGRU may not adequately capture the spatial features. This is because the characteristics during neural network training may lead to GAT in GAT-PCGRU overreliant on labeled KPI data. Therefore, it is hard to guarantee whether the update of GAT model parameters is beneficial. To ensure that topological information is learned sufficiently, GATE-PCGRU is proposed, which adds the error of GATE's reconstructed data to the overall loss calculation of the model. Moreover, the appropriate weights are obtained by an experimental analysis to balance the error when reconstructing the input data and the error when estimating the results.

The remainder of this study is organized as follows. Section II highlights the backbone methods, including GAT, autoencoder (AE), and GRU. Section III describes the GATE-PCGRU framework in detail. Section IV presents a case study of a wet zinc decoppering process using on-site industrial data, as well as thorough model analysis experiments. Section V presents the conclusions of the study.

II. Preliminary Knowledge

A. Graph Attention Networks

Research has demonstrated that the GAT [33] model is preferable for the topology analysis of graph data in multiple fields [34], [35], [36]. GAT is a GNN based on attention mechanism, and it calculates weights by aggregating features of the current node and its neighboring nodes. The model considers the topology relationships between neighboring nodes and integrates the correlation between the features of each node in the graph, thus allowing it to learn topology information. In addition, the mutual influence of devices is directional in industrial systems, and GAT outperforms graph convolutional neural network (GCN) [37] in the application of directed graphs. As shown in Fig. 2, the primary structure of GAT consists of two steps: 1) attention coefficient calculation and 2) feature aggregation.

1) The attention coefficient calculation is given as follows.

$$e_{ij} = a\left(W\overrightarrow{x_i}, W\overrightarrow{x_j}\right) \tag{1}$$

where e_{ij} denotes the importance of node j for i, W represents the weight matrix of input-to-output feature dimensions transformation, $\overrightarrow{x_i}$ represents a feature vector of node i, and $\overrightarrow{x_j}$ denotes a feature vector of node j. The attention mechanism was introduced into the

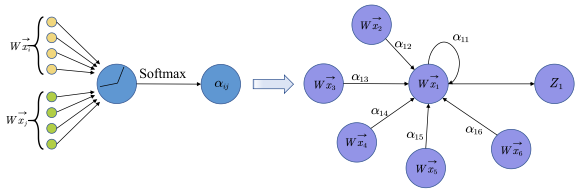


Fig. 2. GAT network structure.

directed graph by calculating the attention coefficients of the neighboring nodes of node. Subsequently, the softmax function was used to process e_{ij}, which was computed by all neighboring nodes to get the final attention coefficient α_{ij}, as per the following equation:

$$\alpha_{ij} = \frac{\exp\left(\text{LeakyReLU}\left(a\left(W\overrightarrow{x_i}, W\overrightarrow{x_j}\right)\right)\right)}{\sum_{k\in N(i)}\exp\left(\text{LeakyReLU}\left(a\left(W\overrightarrow{x_i}, W\overrightarrow{x_k}\right)\right)\right)} \tag{2}$$

where $k \in N(i)$ and k is the neighboring node of node i, a is a learnable weight vector, and LeakyReLU is a nonlinear activation function.

2) *Feature Aggregation:* After obtaining the normalized attention coefficients α_{ij}, feature aggregation was performed according to the following equation as the output features of each node:

$$\overrightarrow{x_i'} = \text{ELU}\left(\sum_{j\in N_i}\alpha_{ij}W\overrightarrow{x_j}\right) \tag{3}$$

where $\overrightarrow{x_i'}$ represents a new feature of each output node i of the GAT layer after incorporating neighborhood information and ELU is an activation function. In the GAT, ELU is usually used as the activation function and its expression is

$$\text{ELU}(x) = \begin{cases} x, & x \geq 0 \\ \beta(e^x - 1), & x < 0. \end{cases} \tag{4}$$

To make the model more stable, a multiheaded attention mechanism was used to calculate K sets of attention coefficients and to integrate multiple attention heads using an average, calculated as follows:

$$\overrightarrow{x'}_i = \begin{bmatrix} \text{ELU}\left(\sum_{j\in N_i}\alpha_{ij}^1 W^1\overrightarrow{x_j}\right) \\ \cdots \\ \text{ELU}\left(\sum_{j\in N_i}\alpha_{ij}^k W^k\overrightarrow{x_j}\right) \\ \cdots \\ \text{ELU}\left(\sum_{j\in N_i}\alpha_{ij}^K W^K\overrightarrow{x_j}\right) \end{bmatrix} \tag{5}$$

where k represents concatenation, α_{ij}^k are normalized attention coefficients of nodes i and j calculated by the kth attention mechanism, and W^k denotes the corresponding linear transformation matrix of the kth attention mechanism.

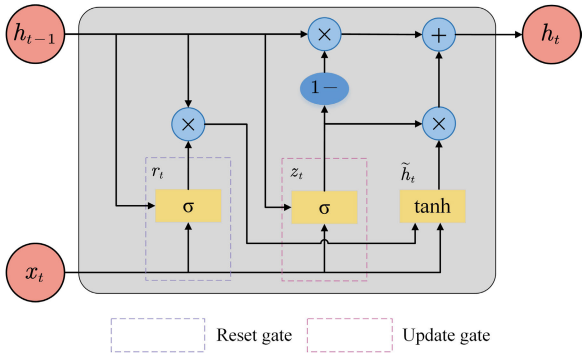


Fig. 3. GRU network structure.

B. Autoencoder

AE is a two-part unsupervised learning model comprised of an encoder and a decoder. The parameter updating process backpropagates the error between the reconstructed input data and the original input data as loss and then iteratively trains and optimizes the model parameters to obtain accurate feature information. The encoding and decoding processes of the AE are

$$f_{\text{encoder}} : X \longrightarrow Z \tag{6}$$

$$f_{\text{decoder}} : Z \longrightarrow \hat{X} \tag{7}$$

where f_{encoder} denotes the encoding, f_{decoder} is the decoding, X is the input data, Z represents the feature vector, and $\hat{X}$ is the reconstructed input data.

C. Gated Recurrent Unit Network

GRU is an improvement from the long short-term memory (LSTM) network, whose complex internal structure is optimized by reducing and merging gate structure units, thus increasing the training speed of the network while ensuring accuracy [38], [39]. In addition, the reduction of parameters reduces the risk of overfitting. The structure is shown in Fig. 3. The functions and calculations of the two gates involved in the storage unit are shown in the following.

1) The update gate is given as

$$z_t = \sigma(W_z x_t + U_z h_{t-1} + b_z) \tag{8}$$

where z_t is the updated gate state at moment t to determine how much information should be brought to the next state unit, x_t is the input vector at moment t, and h_{t-1} is the hidden state passed down from the previous moment.

2) The reset gate is given as

$$r_t = \sigma(W_r x_t + U_r h_{t-1} + b_r) \tag{9}$$

where r_t denotes the reset gate that determines how much previous information should be ignored.

3) The hidden layer output is given as

$$\tilde{h}_t = \tanh(W_h x_t + U_h(r_t \odot h_{t-1})) \tag{10}$$

$$h_t = (1 - z_t) \odot h_{t-1} + z_t \odot \tilde{h}_t \tag{11}$$

where $\tilde{h}_t$ is the candidate hidden layer state information at moment t and h_t is the hidden layer state information at moment t.

4) Thus, the final output obtained is

$$y_t = \sigma\left(W_y h_t + b_y\right) \tag{12}$$

where y_t is the output of the GRU unit at moment t.

The GRU model has a straightforward structure, and it has the advantages of fast convergence and a limited number of parameters; moreover, in the GRU, each unit cycle captures the dependencies at different time scales in an adaptive manner. Consequently, in the proposed method, the GRU was utilized to extract the intrinsic information included in the data sequence.

III. Proposed Model

We presented a novel framework for estimating key metrics and applied it to a multirate industrial process. Fig. 4 shows the modeling framework, which primarily consists of spatial topological information, temporal features, a spatial-dependent module, a temporal-dependent module, and the GATE-PCGRU spatial–temporal structural estimation network. First, the industrial reactors were analyzed based on a priori knowledge and expert experience. Subsequently, the node variables of the industrial network and the relationship between them were determined to constitute the industrial topology as the spatial structure information. Second, the multivariate process variables collected by the corresponding high-rate sensors of each variable node, the key indicators obtained from the assay, and the fixed parameters constituted the temporal characteristics of the multirate data of the industry. Next, in the spatial-dependent module, the spatial dependence was modeled based on the graph attention autoencoder (GATE). With GATE, AE was used for unsupervised representative learning based on GAT, and the stability of the features extracted by GAT was ensured. In the temporal-dependent module, GRU with period control unit was proposed to process the features of multirate industry samples. In conclusion, a spatial–temporal feature fusion mechanism was employed to extract the spatial–temporal correlation of industrial processes for accurate key indicator estimates.

In addition, the framework had two primary components: offline training and online estimation. For the offline training, the proposed GATE-PCGRU was trained using a sample of historical data with a sampling interval of 2 h. For the online estimation, the model was subject to relevant optimizations and adjustments. In particular, the period control part of the time module was added to ensure that the model could estimate the key metrics in real time using the process variable data collected by the high-rate sensors.

The rest of this section is organized as follows. Section III-A covers the spatial-dependent modeling for multirate industrial processes. Section III-B covers the temporal-dependent modeling for multirate industrial processes, and Section III-C covers the spatial–temporal fusion mechanism for multirate industrial processes.

A. Spatial-Dependent Modeling

This study aimed to obtain an online estimation of key metrics for multirate industrial processes, which is essentially a GNN regression task. Extensive studies have been conducted [40], [41], [42], [43] to use node attributes for graph representation learning. Notably, GATs, which can focus only on task-relevant parts rather than the entire graph structure owing to the attention-based feature aggregation mechanism of GAT, are widely used in various domains. However, the above approaches still have the following two limitations.

1) Industrial structures are complex and variable, and their performance is limited facing networks with diverse industry structures [44], [45].
2) These are supervised graph embedding methods that are unsuitable for scenarios where the industrial data are rich, but the labeled data are scarce. This is because in real-world industrial production, numerous data labels tend to have a high access threshold.

AEs have recently become popular in unsupervised learning due to their ability to capture complex relationships between input data attributes via stacked nonlinear layers. However, traditional AEs are unable to exploit explicit relationships in graph-structured data. Therefore, to exploit the relationships in graph-structured data and to reduce the dependence on labeled data, the proposed method was constructed based on the architecture of GAEs using a GAT encoder and decoder scheme that considers the error of the feature attributes of reconstructed data as the final loss. The method is supported by a specially designed variant of GNNs: GATE. This is the first time that GAEs were used to improve the performance of online estimation of key industrial metrics. It consists of two main modules: an encoder and a decoder, as shown in Fig. 5.

1) *Encoder:* We believed that a single-layer GAT would not represent the characteristics of the nodes well, which was confirmed later in the experiment. Therefore, we employed a multilayer GAT stacking to obtain the graphical structure of the industrial process with the representation Z of the nodes, as shown in Fig. 5. Each node of the current encoder layer utilized the representations of its neighbors according to their correlation to generate a new representation. The output of the kth graph attention layer in the encoding process was

$$Z_i^{(k)} = \mathrm{ELU}\left(\sum_{j \in N_i} \alpha_{ij}^{(k)} W^{(k)} Z_j^{(k-1)}\right). \tag{13}$$

2) *Decoder:* Instead of computing the loss by reconstructing the topology of the graph through the inner product of Z and Z^T, the GAT whose layers were symmetric to the GATE encoder was used to reconstruct the features of the node data $\hat{X}$. Subsequently, the following loss function was defined to optimize the model:

$$\sum_{i=1}^{N} \left\| X_i - \hat{X}_i \right\|_2. \tag{14}$$

Thus, a model of the spatial module and an optimized loss function were constructed using GATE.

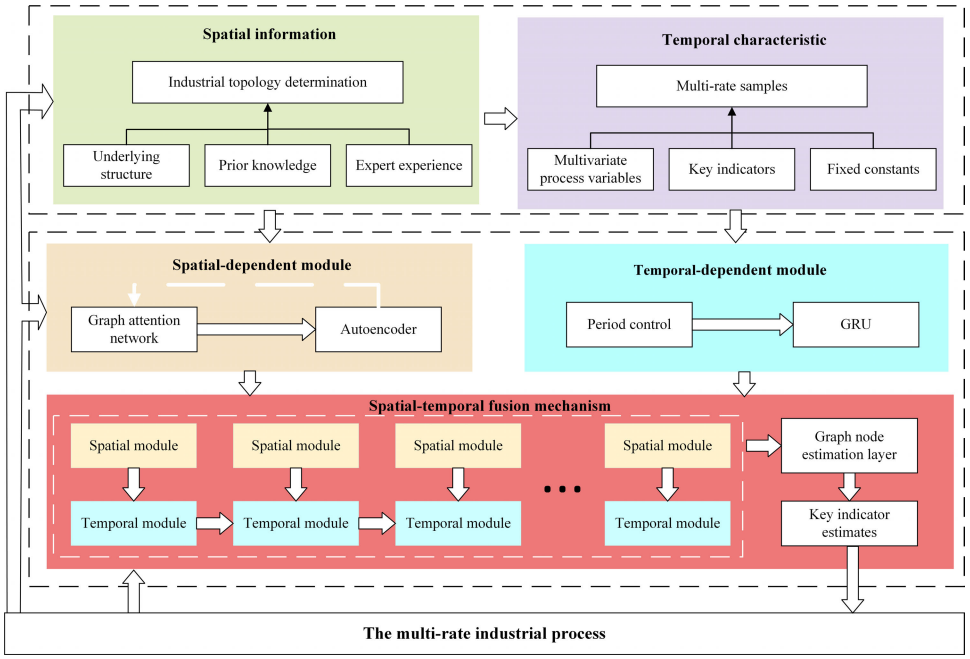


Fig. 4. Overall framework of GATE-PCGRU modeling.

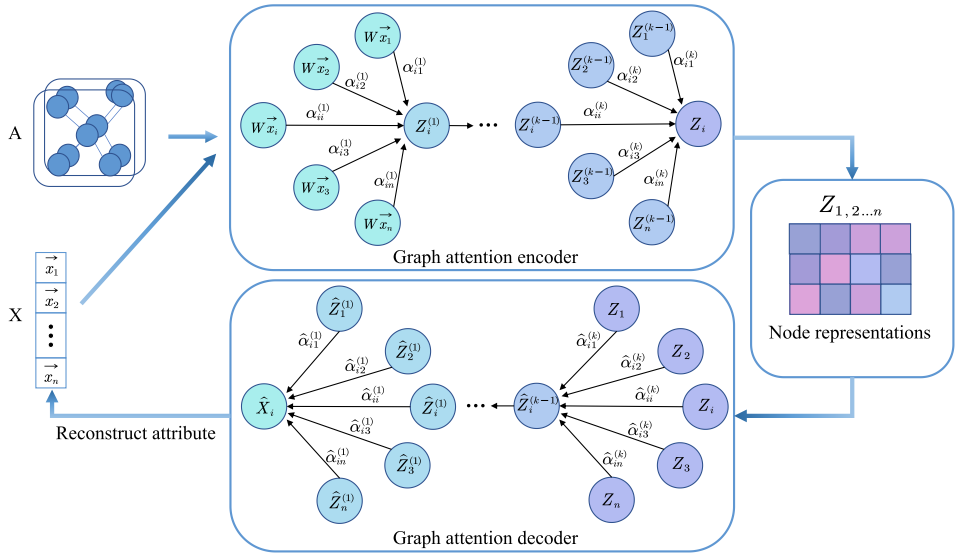


Fig. 5. Calculation process of GATE's encoder and decoder.

B. Temporal-Dependent Modeling

To apply GRU to the multirate state estimation problem, a PCGRU was proposed. The concept of our model's design was to alter the internal structure of the GRU and apply what was learned using data with a low sampling rate to the online estimation at high rates by adding a period control unit. In particular, as shown in Fig. 1, some variables with long manual test intervals measuring hours were observed. Consequently, downsampling was necessary to ensure that all data of the training set retained the same frequency as the variables with low sampling rates. Therefore, only long-period estimation could be performed in the model, and thus, achieving online real-time estimation was difficult. A period control unit was added to learn the changing pattern of high sampling rate variables switching within different estimation periods. Thus, the period-related parameters could be adjusted for a short-period estimation. Hence, the internal structure of the GRU was modified, as shown in Fig. 6. Compared to the GRU structural diagram, a period control unit was added and the hidden layer output became

$$h_t' = h_t + \Delta T v_{[t-1,t]} \tag{15}$$

$$v_{[t-1,t]} = \sigma(W_v h_{t-1} + b_v) \tag{16}$$

where h_t' is the new hidden layer output, ΔT is the period-related parameters of the period control unit, and $v_{[t-1,t]}$ is expressed as the changing rates of h_t in $[t-1, t]$. The other model parameters were consistent with the GRU, and

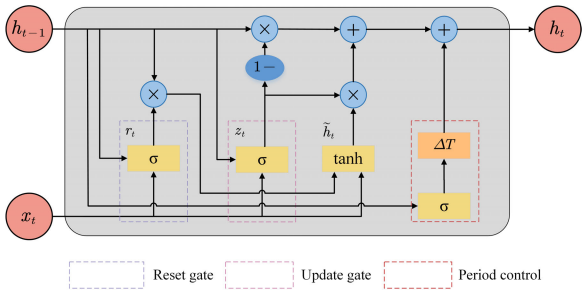


Fig. 6. PCGRU network structure.

the input x_t was GATE encoder layer output, Z. The Eqs. (15) and (16) shows the physical meaning of the modification: to use the period control unit to learn the change rate $v_{[t-1,t]}$ among blocks of the input sequence data during the period. By learning this mapping relationship, an estimation with different periods could be achieved by varying the length of the estimation period ΔT.

C. Spatial–Temporal Fusion Mechanism

Timely and accurate estimation of key indicators remains a challenge due to the high nonlinearity and dynamic spatial–temporal dependence among industrial process variables, which are given as follows. The first is the spatial correlation between variables of different steps, including correlation between near and distant variables. The second is the temporal correlation between variables of a process at different time intervals, including temporal proximity, periodicity, and trend properties. Third, incoming and outgoing flows are highly correlated with intermediate flows and contribute to each other. Therefore, to estimate key indicator flows, whether they are input–output or transition variables, the dependencies between them in the overall process must be considered.

To fully consider the spatial and temporal influences of process variables, a spatial–temporal fusion mechanism, which integrates the spatial–temporal correlations while ensuring computational accuracy and efficiency, was proposed. In particular, the GATE can compensate for the shortcomings that exist in the complex existing graph networks while modeling, but without large improvements in accuracy. It solves the problem of insufficient learning of spatial feature information due to the implication of the temporal correlation modeling method. Simultaneously, it uses the unique control period structure of PCGRU in temporal order to extract multirate features, which can describe the correlation of time series data more comprehensively. Thus, the two models were fused and optimized using a spatial–temporal fusion mechanism, which enabled the entire framework to fully extract the spatial–temporal correlations to describe the dynamic characteristics of the multirate industrial process more comprehensively. The spatial–temporal fusion mechanism is shown in Fig. 7, which consists primarily of two parts: 1) offline training of the model and 2) online deployment of the model.

1) *Offline Training of the Model:* To train the entire GATE-PCGRU model, in addition to calculating the loss L_{PCGRU} based on the estimation results and the target value, GATE was introduced to reconstruct the loss of the input data L_{GATE}; subsequently, the two losses were combined to obtain the final loss function L

$$L = L_{\text{PCGRU}} + \lambda L_{\text{GATE}} \tag{17}$$

where λ is a tradeoff parameter. By minimizing L, the spatial and temporal models avoided overconstraining each other for better spatial–temporal integration. Essentially, this is a joint optimization process that mutually benefits both GATE and PCGRU, in which the GATE provides a better potential representation of nodes for the PCGRU multirate temporal-dependent modeling, and the PCGRU module effectively enhances the GATE's capacity to encode the input spatial topology information, thereby enhancing the potential and low-dimensional representation.

2) *Online Deployment of the Model:* The parameters ΔT of the period control unit of the PCGRU need to be adjusted during the deployment, which is used to estimate the short period of the key indicators. This adjustment is based on the relationship between samples with different sampling frequencies, which can be expressed as

$$\Delta T_{\text{test}} = \Delta T_{\text{train}} \frac{T_s}{T_l} \tag{18}$$

where ΔT_{train} and ΔT_{test} are different ΔT inputs for model training and testing, respectively; T_s denotes the high-frequency sampling interval; and T_l denotes the low-frequency sampling interval.

IV. Case Study

A. Process Description

Copper removal is an important part of the wet zinc purification process, as shown in Fig. 8. It involves two open, continuously stirred reactors and a thickener for reactors 1# and 2#. External chutes connect the reactors in sequence. After leaching, zinc sulfate solution is pumped through the overflow tank to reactor 1# to begin removing copper by the addition of zinc powder. Zinc powder is added separately to both reactors to ensure that copper ions in the solution are separated and removed by the precipitation of predominantly monomeric copper and cuprous oxide. In addition, following the solid–liquid separation in the thickener, a small portion of the bottom stream containing solid particles is reintroduced to reactor 1#, where the solid particles act as crystal seeds, whereas the remainder of the solid particles is pressed and dried to produce copper slag for recycling. The clear liquor from the thickener's upper section is delivered to the cobalt–nickel removal section for further purification. This copper removal method with a bottom stream return is more cost-effective and eco-friendly than the conventional copper removal method. However, the employment of various reactions complicates the mechanism of the copper removal process and necessitates precise control. To improve control performance and copper removal efficiency, the copper ion concentration at the outlet must be estimated in real time.

The goal of the copper removal process is to stabilize the export copper ions within a certain range as a small amount of

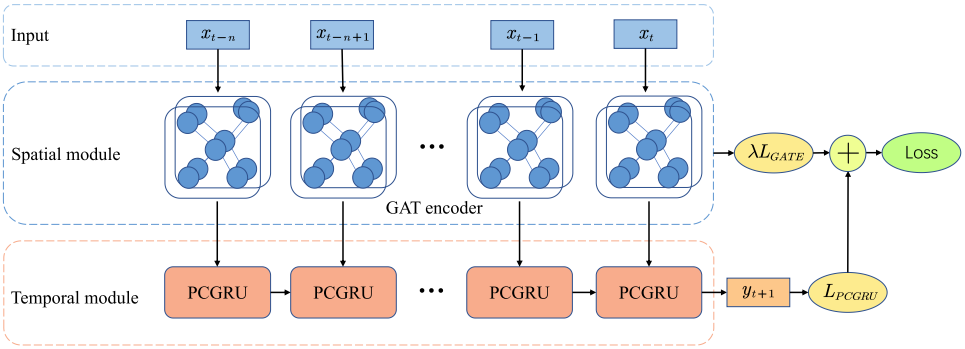


Fig. 7. GATE-PCGRU framework: spatial–temporal fusion mechanism.

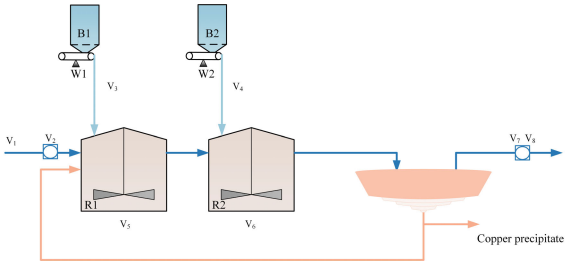


Fig. 8. Flowchart of the copper removal process in wet zinc refining.

TABLE I
PROCESS VARIABLES FOR COPPER REMOVAL

Variables	Description	Unit	Sampling frequency
V1	Total flow rate	m^3/h	1min
V2	Concentration of imported copper ions	mg/L	2h
V3	Amount of zinc powder added for 1#	kg/h	1min
V4	Amount of zinc powder added for 2#	kg/h	1min
V5	1# Potential detection value	mv	1min
V6	2# Potential detection value	mv	1min
V7	Concentration of exported cobalt ions	mg/L	4h
V8	Concentration of exported copper ions	mg/L	2h

copper ions is required as a catalyst for purification during the cobalt–nickel removal process. The process variables for copper removal and their sampling frequencies are summarized in Table I. The different sampling rates between the offline assay variables and real-time online detection variables make the entire process a multirate system. The low sampling frequency of KPIs results in delayed feedback of key information and difficulties in meeting the control accuracy requirements of the copper removal process. Therefore, an effective model must be developed to estimate KPI in real time for such a copper removal process that belongs to a multirate industry.

B. Experimental Settings

In this study, experimental data were collected during actual production at a wet process zinc refinery in China over a sample collection period of 8 m, starting in January 2020, and the total number of samples collected was 25 921. (The data are desensitized for confidentiality reasons. In the result figures, the vertical axis values are scaled.) Each data sample contained the values of output states, inlet conditions, and reaction conditions of the industrial process data presented in Table I. The computational runtime environment used an Intel Core i7-8700 CPU, 42-GB RAM, and Nvidia GeForce RTX 3090 GPU and was implemented based on the deep learning framework Pytorch 1.7.0. Four accuracy metrics, namely, root-mean-square error (RMSE), correlation coefficient (R^2), mean absolute error (MAE), and mean absolute percentage error (MAPE) [46], were used to evaluate the model and validate the proposed GATE-PCGRU method.

As listed in Table I, all online process variables were measured every minute, whereas offline copper ion concentration and cobalt ion concentration were assayed every 2 and 4 h, respectively. Numerous studies in the literature [47], [48], [49] provide valid information for determining the reaction type, reaction steps, and the influence of process variables on the copper removal reaction. This study determined the relationships between the variables listed in Table I based on prior knowledge and expert experience and subsequently constructed the industrial topology of the copper removal process based on graphs. The collected data corresponding to the graph nodes and the data from the assays were normalized before being input into the model to eliminate magnitude differences. During offline training, to investigate the KPI estimation task for the multirate process, the data were sampled by taking the assay frequency of the exported copper ion concentration as one sampling interval to obtain the training set, and the parameter ΔT of the PCGRU model was set to 120. Next, for online estimation, the port copper ion value was obtained at a frequency of 10 min, which implied that the parameter ΔT of the PCGRU was set to 10.

C. Results and Analysis

During the online estimation test, PCGRU contributes to the increment of the frequency of KPI estimation from every 2 h to every 10 min. In addition, To fully evaluate the performance of the GATE-PCGRU, two sets of ablation experiments were conducted.

1) *PCGRU:* Removing the spatial-dependent module and using only the GRU with a period control unit. The

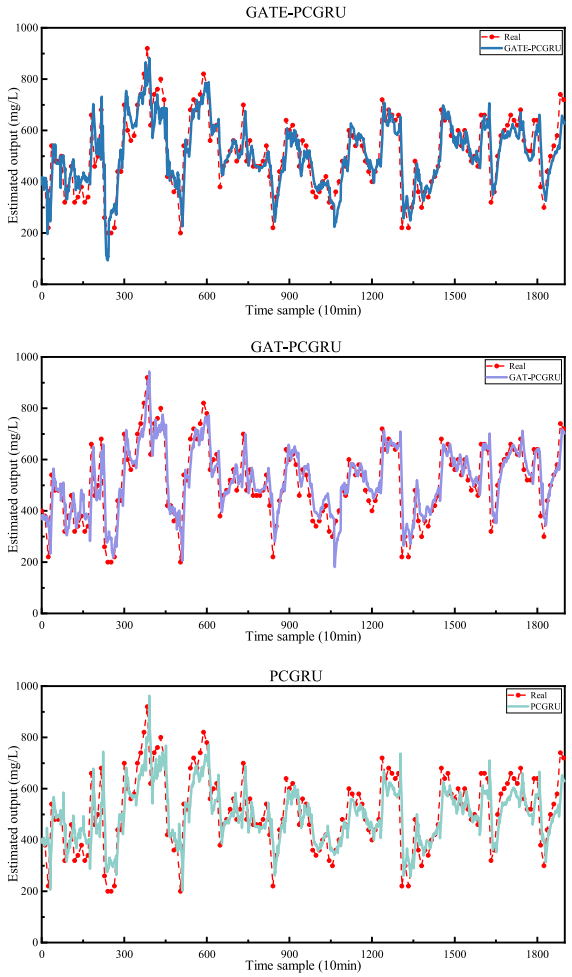


Fig. 9. Estimation results of GATE-PCGRU, GAT-PCGRU, and PCGRU.

comparative results of the estimation values are shown in Fig. 9.

2) *GAT-PCGRU:* Replacing GATE with GAT in our proposed model, thus demonstrating that GATE is more reliable and stable for capturing spatial features in the GATE-PCGRU.

Notably, the performances of GRU and GATE were not shown in the ablation comparison because they cannot estimate the KPI at a 10-min frequency due to the multirate problem. However, the experimental results of GRU and GATE estimate the KPI at 2-h frequency are described later in the baseline comparison experiment. What is mentioned above also illustrates the essential role of PCGRU in achieving real-time estimation in the multirate industrial process.

Fig. 9 shows that all three models are capable of describing the varying KPI over time, which implies that the inclusion of a period control unit makes the entire framework better able to handle the industrial multirate problem. However, considering only the temporal-dependent problem is insufficient because the PCGRU deviates the most and the effect is worse when the KPI fluctuates frequently. For example, as seen from samples No.1425 to No.1710, the actual value of KPI fluctuates frequently, and the estimated value of PCGRU exhibits a high

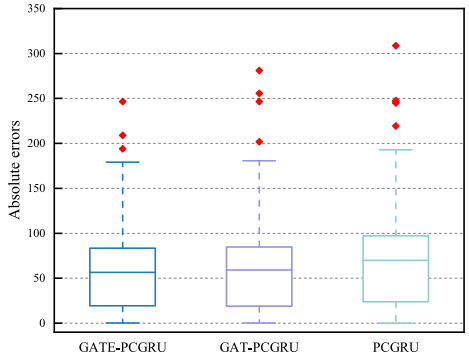


Fig. 10. Box plot of absolute estimation error among GATE-PCGRU, GAT-PCGRU, and PCGRU.

TABLE II
PERFORMANCE INDICATORS OF GATE-PCGRU, GAT-PCGRU, AND PCGRU

Methods	MAE	MAPE	RMSE	R^2
PCGRU	69.82	0.160	91.94	0.606
GAT-PCGRU	59.10	0.142	79.44	0.706
Proposed	**56.58**	**0.132**	**75.92**	**0.731**

deviation. In contrast, the GATE-PCGRU and GAT-PCGRU are able to maintain a satisfactory estimation accuracy with low deviation. This indicates that the model incorporates a spatial-dependent extraction module, which can characterize the spatial–temporal features of the data and increase the accuracy. Moreover, GAT-PCGRU is unstable during particularly large fluctuations. For instance, from samples No.238 to No.258, copper ion concentration first decreases and then increases sharply, but the trend estimated by GAT-PCGRU is diametrically opposite. In addition, GATE-PCGRU maintains excellent stability, and the estimated trend remains consistent with the actual value. Furthermore, Fig. 10 shows the box plots of the absolute errors in the estimation results for the three models; here, the GATE-PCGRU exhibited better performance in terms of the mean and outliers. In the estimation results of GAT and GRU, the mean of the errors is higher, and the outliers occur more often and deviate further from their mean. The aforementioned conclusions indicate that GATE-PCGRU is more robust and can maintain the stability of the estimation. The accuracy metrics of the three models are listed in Table II.

The specific indicators listed in Table II reveal that GATE-PCGRU was optimal in all indicators, with an R^2 improvement of 20.72% and 3.65% over PCGRU and GAT-PCGRU, respectively. This indicates that the exported copper ion content estimated based on the GATE-PCGRU and the actual assayed copper ion concentration curves are more consistent in terms of value and trends.

The parameters of the data-driven model contain all the information that the model obtains from the sample; moreover, they determine the effectiveness of the model estimation. This section analyzes the number of layers of the GATE encoder and λ in the loss function, which may have an impact on GATE-PCGRU's estimation performance.

1) *GATE Encoder Layer Number:* It is used for extracting spatial multidimensional features. Layers 1–4 (encoder

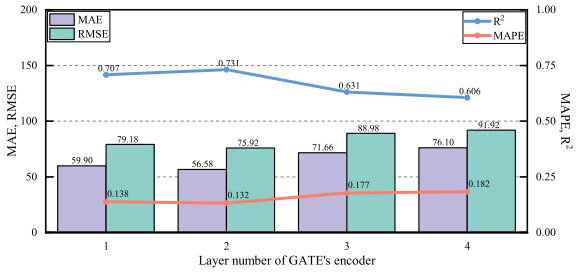


Fig. 11. Performance indicators of GATE-PCGRU with different numbers of layers of the GATE encoder.

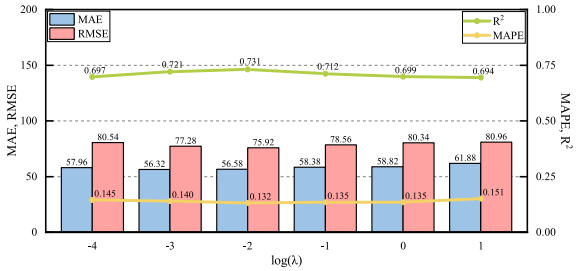


Fig. 12. Performance indicators of GATE-PCGRU with different λ's.

and decoder are symmetric, and the total number of layers is 2, 4, 6, and 8) were analyzed to evaluate the model performance at different layers. As observed in Fig. 11, although the perceptual field of output features of the node expands with the addition of a deeper layer, the MAE, MAPE, RMSE, and R^2 do not improve. The model achieves a better performance when the encoder layer is set to 2, and the model performance gets worse as the layer increases. This indicates that if there are limited layers, the feature aggregation range will be too small, whereas when the layers are excessive, feature degradation might occur.

2) λ *in the Loss Function:* It is used to weigh L_{PCGRU} and L_{GATE}, which can be defined as the degree of spatial and temporal influences. In this study, the values of λ were varied in the range of $\{0.0001, 0.001, 0.01, 0.1, 1, 10\}$, and the results are shown in Fig. 12. The results show that starting from 0.0001, as λ increased, R^2 increased, but the rest of the indicators did not improve considerably. Moreover, the model worked best when λ was 0.01. When λ exceeded 0.01, a slight degradation on the metrics' performance was noted, but the changes were flat and no further consequences were noted, which indicates that the model performs better in the range of $\{0.0001, 0.001, 0.01, 0.1, 1, 10\}$. This demonstrates that the node optimization of the GATE and PCGRU can achieve better performance. Therefore, λ was set as 0.01, which obtains sufficient feature information without adding additional time and space complexities.

In addition, the proposed GATE-PCGRU was compared with the current mainstream and conventional industrial KPI estimation models through experiments. The compared methods include deep learning networks (LSTM, GRU, and CNN-GRU) for solving industrial temporal problems, graph networks (ChebNet, GCN, GAT, GAE, and GATE) for extracting industrial topology features, and spatial–temporal network GCN-GRU. The results are listed in Table III.

TABLE III
PERFORMANCE INDICATORS OF BASELINE EXPERIMENTS

Methods	MAE	MAPE	RMSE	R^2	Average test time
LSTM	72.22	0.177	96.10	0.569	12.80 ms
GRU	73.46	0.168	96.00	0.570	7.18 ms
CNN-GRU	67.16	0.153	87.39	0.644	16.25 ms
ChebNet	74.18	0.182	94.07	0.587	5.48 ms
GCN	74.69	0.183	93.60	0.591	6.87 ms
GAT	68.99	0.165	90.74	0.616	8.84 ms
GAE	70.50	0.178	88.96	0.631	8.10 ms
GATE	70.02	0.157	88.43	0.635	10.93 ms
GCN-GRU	64.32	0.145	82.46	0.683	13.16 ms
Proposed	**56.58**	**0.132**	**75.92**	**0.731**	**11.65 ms**

The GATE-PCGRU proposed in this study exhibited the best estimation performance, as observed in Table III. It is noteworthy that these baseline models can only output KPIs at low frequencies. Among the single models GRU, LSTM, ChebNet, GCN, GAT, GAE, and GATE, the single time series models GRU and LSTM disregard the spatially correlated features of industrial topologies, which results in unsatisfactory estimate outcomes. According to the estimation results, GNNs, such as ChebNet, GCN, GAT, GAE, and GATE, which consider the industrial topology, are more effective in terms of R^2 than single temporal models but perform less consistently in terms of MAPE. Furthermore, ChebNet and GCN are less effective than GAT due to the fact that they can only assign the same importance coefficients to their different process variable nodes. This indicates that the estimation effect can be improved by considering the influence of each variable on KPI. The performance of GAE and GATE is more effective than that of GAT and GCN because adaptive spatial correlation feature extraction improves the stability.

The combined models CNN-GRU, GCN-GRU, and GATE-GRU achieved better estimation results than the single model, which indicates that the multimodel approach is better than the single model in estimating the exported copper ion concentration during the copper removal. This is because the abstract feature representation learned by the CNN in the CNN-GRU helps to explore the implied relationships in the data, which in turn improves the performance of the GRU. Therefore, they are better than the single model. Moreover, the GCN-GRU considers the spatial and temporal dependence simultaneously, and thus, it also obtains desirable results. Compared with these two methods, the GATE-PCGRU performs better due to the consideration of adaptive spatial correlation between different process variables and the ability to ensure that spatial features are captured. Furthermore, the absolute error statistics of the baseline experiment are shown in Fig. 13; evidently, the GATE-PCGRU outperforms the other models as it exhibits the least number of samples for which the absolute value of the error is more than 0.1. Therefore, GATE-PCGRU has better reliability and stability. In addition, as a combined model, the proposed method is more time-consuming than other models in the model training process because it has more parameters, and the

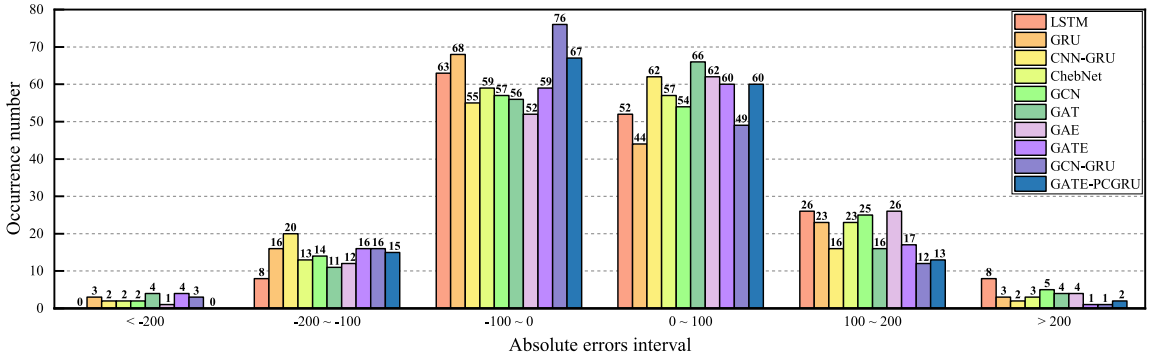


Fig. 13. Absolute error statistics of baseline experiments.

loss calculation is more complicated. However, we are more concerned about the efficiency of real-time estimation. The time consumed in estimation is related to the running state of the computer, so we performed multiple estimations and recorded the average test time in Table III, indicating a small difference in running speed between them. This is acceptable in the real-time estimation of KPI in industrial operations.

V. Conclusion

In this study, a novel spatial–temporal multirate industrial process KPI estimation framework was proposed to improve the multirate industrial process estimation. It could efficiently extract the spatial dependence among variables in a variety of industrial processes and capture the temporal features of multirate variables. A study of a real industrial case demonstrated the superior performance of the proposed model for KPI estimation, indicating that the proposed GATE-PCGRU learns the complex relationship between the quality variables with low sampling frequency and the process variables with high sampling frequency.

Specifically, the GAT effectively captured the complex topology of industrial processes on a directed graph constructed from expert experience. The proposed PCGRU could control the estimation period to realize the real-time estimation of KPI in the multirate industrial process. Furthermore, based on a combined spatial–temporal mechanism, GATE-PCGRU simultaneously captured the spatial–temporal features, and the estimation accuracy was improved compared to the direct combined model GAT-PCGRU. The experimental results showed that GATE-PCGRU was optimal in all indicators, with an R^2 improvement of 20.72% and 3.65% over PCGRU and GAT-PCGRU, respectively. Furthermore, an analysis of experimental results compared to several prevalent estimation models demonstrated that the GATE-PCGRU is capable of achieving more accurate estimations.

References

[1] X. Chen, W. Zhong, C. Jiang, Z. Li, X. Peng, and H. Cheng, "Key performance index estimation based on ensemble locally weighted partial least squares and its application on industrial nonlinear processes," *Chemometric Intell. Lab. Syst.*, vol. 203, Aug. 2020, Art. no. 104031.

[2] C.-F. Lindberg, S. Tan, J. Yan, and F. Starfelt, "Key performance indicators improve industrial performance," *Energy Proc.*, vol. 75, pp. 1785–1790, Aug. 2015.

[3] C. F. Lui, Y. Liu, and M. Xie, "A supervised bidirectional long short-term memory network for data-driven dynamic soft sensor modeling," *IEEE Trans. Instrum. Meas.*, vol. 71, pp. 1–13, 2022.

[4] M. Mou and X. Zhao, "Gated broad learning system based on deep cascaded for soft sensor modeling of industrial process," *IEEE Trans. Instrum. Meas.*, vol. 71, pp. 1–11, 2022.

[5] X. Yuan, Y. Wang, C. Yang, Z. Ge, Z. Song, and W. Gui, "Weighted linear dynamic system for feature representation and soft sensor application in nonlinear dynamic industrial processes," *IEEE Trans. Ind. Electron.*, vol. 65, no. 2, pp. 1508–1517, Feb. 2018.

[6] H. Luo, S. Yin, T. Liu, and A. Q. Khan, "A data-driven realization of the control-performance-oriented process monitoring system," *IEEE Trans. Ind. Electron.*, vol. 67, no. 1, pp. 521–530, Jan. 2020.

[7] L. Zhou, Y. Wang, Z. Ge, and Z. Song, "Multirate factor analysis models for fault detection in multirate processes," *IEEE Trans. Ind. Informat.*, vol. 15, no. 7, pp. 4076–4085, Jul. 2019.

[8] Y. Cong, L. Zhou, Z. Song, and Z. Ge, "Multirate dynamic process monitoring based on multirate linear Gaussian state-space model," *IEEE Trans. Autom. Sci. Eng.*, vol. 16, no. 4, pp. 1708–1719, Oct. 2019.

[9] J. Li, B. Yang, H. Li, Y. Wang, C. Qi, and Y. Liu, "DTDR–ALSTM: Extracting dynamic time-delays to reconstruct multivariate data for improving attention-based LSTM industrial time series prediction models," *Knowl.-Based Syst.*, vol. 211, Jan. 2021, Art. no. 106508.

[10] X. Yuan, L. Feng, K. Wang, Y. Wang, and L. Ye, "Deep learning for data modeling of multirate quality variables in industrial processes," *IEEE Trans. Instrum. Meas.*, vol. 70, pp. 1–11, 2021.

[11] Z. Chai, C. Zhao, and B. Huang, "Variational progressive-transfer network for soft sensing of multirate industrial processes," *IEEE Trans. Cybern.*, vol. 52, no. 12, pp. 12882–12892, Dec. 2022.

[12] B. Sun, J. Dai, K. Huang, C. Yang, and W. Gui, "Smart manufacturing of nonferrous metallurgical processes: Review and perspectives," *Int. J. Minerals, Metall. Mater.*, vol. 29, no. 4, pp. 611–625, Apr. 2022.

[13] Y. Duan, M. Liu, and M. Dong, "A metric-learning-based nonlinear modeling algorithm and its application in key-performance-indicator prediction," *IEEE Trans. Ind. Electron.*, vol. 67, no. 8, pp. 7073–7082, Aug. 2020.

[14] X. Yuan, C. Ou, Y. Wang, C. Yang, and W. Gui, "A layer-wise data augmentation strategy for deep learning networks and its soft sensor application in an industrial hydrocracking process," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 32, no. 8, pp. 3296–3305, Aug. 2021.

[15] B. Shen and Z. Ge, "Weighted nonlinear dynamic system for deep extraction of nonlinear dynamic latent variables and industrial application," *IEEE Trans. Ind. Informat.*, vol. 17, no. 5, pp. 3090–3098, May 2021.

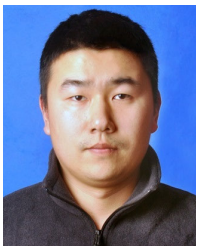
[16] Y. He, Z. Ying, Y. Wang, and J. Wang, "Enhanced dynamic dual-latent variable model for multirate process monitoring and its industrial application," *IEEE Trans. Instrum. Meas.*, vol. 72, pp. 1–8, 2023.

[17] Q. Sun and Z. Ge, "Probabilistic sequential network for deep learning of complex process data and soft sensor application," *IEEE Trans. Ind. Informat.*, vol. 15, no. 5, pp. 2700–2709, May 2019.

[18] K. Liu, M. Zheng, Y. Liu, J. Yang, and Y. Yao, "Deep autoencoder thermography for defect detection of carbon fiber composites," *IEEE Trans. Ind. Informat.*, vol. 19, no. 5, pp. 6429–6438, May 2023.

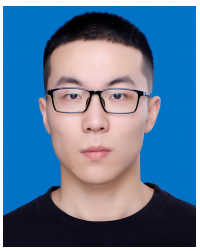
[19] C. Zhao, F. Wang, Z. Mao, N. Lu, and M. Jia, "Quality prediction based on phase-specific average trajectory for batch processes," *AIChE J.*, vol. 54, no. 3, pp. 693–705, 2008.

[20] J. Corrigan and J. Zhang, "Integrating dynamic slow feature analysis with neural networks for enhancing soft sensor performance," *Comput. Chem. Eng.*, vol. 139, Aug. 2020, Art. no. 106842.
[21] J. C. B. Gonzaga, L. A. C. Meleiro, C. Kiang, and R. M. Filho, "ANN-based soft-sensor for real-time process monitoring and control of an industrial polymerization process," *Comput. Chem. Eng.*, vol. 33, no. 1, pp. 43–49, Jan. 2009.
[22] D. P. Kingma and M. Welling, "Auto-encoding variational Bayes," 2013, *arXiv:1312.6114*.
[23] R. Amirthalingam and J. H. Lee, "Subspace identification based inferential control applied to a continuous pulp digester," *J. Process Control*, vol. 9, no. 5, pp. 397–406, Oct. 1999.
[24] Y. Liang, Z. Zhao, and L. Sun, "Memory-augmented dynamic graph convolution networks for traffic data imputation with diverse missing patterns," *Transp. Res. C, Emerg. Technol.*, vol. 143, Oct. 2022, Art. no. 103826.
[25] H. Qin, X. Zhan, Y. Li, X. Yang, and Y. Zheng, "Network-wide traffic states imputation using self-interested coalitional learning," in *Proc. 27th ACM SIGKDD Conf. Knowl. Discovery Data Mining*, Aug. 2021, pp. 1370–1378.
[26] L. Zhou, J. Chen, Z. Song, and Z. Ge, "Semi-supervised PLVR models for process monitoring with unequal sample sizes of process variables and quality variables," *J. Process Control*, vol. 26, pp. 1–16, Feb. 2015.
[27] K. Xu, W. Hu, J. Leskovec, and S. Jegelka, "How powerful are graph neural networks?" 2018, *arXiv:1810.00826*.
[28] J. Xue et al., "Quantifying the spatial homogeneity of urban road networks via graph neural networks," *Nature Mach. Intell.*, vol. 4, no. 3, pp. 246–257, Mar. 2022.
[29] M. Jia, D. Xu, T. Yang, Y. Liu, and Y. Yao, "Graph convolutional network soft sensor for process quality prediction," *J. Process Control*, vol. 123, pp. 12–25, Mar. 2023.
[30] J. Narwariya, P. Malhotra, T. V. Vishnu, L. Vig, and G. Shroff, "Graph neural networks for leveraging industrial equipment structure: An application to remaining useful life estimation," 2020, *arXiv:2006.16556*.
[31] Y. Wu, H. Dai, and H. Tang, "Graph neural networks for anomaly detection in industrial Internet of Things," *IEEE Internet Things J.*, vol. 9, no. 12, pp. 9214–9231, Jun. 2022.
[32] L. Yang, X. Gu, and H. Shi, "A noval satellite network traffic prediction method based on GCN-GRU," in *Proc. Int. Conf. Wireless Commun. Signal Process. (WCSP)*, Oct. 2020, pp. 718–723.
[33] P. Veličković, G. Cucurull, A. Casanova, A. Romero, P. Liò, and Y. Bengio, "Graph attention networks," 2017, *arXiv:1710.10903*.
[34] H. Zhao et al., "Multivariate time-series anomaly detection via graph attention network," in *Proc. IEEE Int. Conf. Data Mining (ICDM)*, Nov. 2020, pp. 841–850.
[35] Y. Zhu, Z.-J. Zha, T. Zhang, J. Liu, and J. Luo, "A structured graph attention network for vehicle re-identification," in *Proc. 28th ACM Int. Conf. Multimedia*, Oct. 2020, pp. 646–654.
[36] M. Song, W. Zhao, and E. Haihong, "KGAnet: A knowledge graph attention network for enhancing natural language inference," *Neural Comput. Appl.*, vol. 32, no. 18, pp. 14963–14973, Sep. 2020.
[37] T. N. Kipf and M. Welling, "Semi-supervised classification with graph convolutional networks," 2016, *arXiv:1609.02907*.
[38] Y. Tian, Y. Xu, Q.-X. Zhu, and Y. He, "Novel stacked input-enhanced supervised autoencoder integrated with gated recurrent unit for soft sensing," *IEEE Trans. Instrum. Meas.*, vol. 71, pp. 1–9, 2022.
[39] Z. Que, X. Jin, and Z. Xu, "Remaining useful life prediction for bearings based on a gated recurrent unit," *IEEE Trans. Instrum. Meas.*, vol. 70, pp. 1–11, 2021.
[40] B. Perozzi, R. Al-Rfou, and S. Skiena, "DeepWalk: Online learning of social representations," in *Proc. 20th ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining*, Aug. 2014, pp. 701–710.
[41] A. Grover and J. Leskovec, "node2vec: Scalable feature learning for networks," in *Proc. 22nd ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining*, Aug. 2016, pp. 855–864.
[42] M. Defferrard, X. Bresson, and P. Vandergheynst, "Convolutional neural networks on graphs with fast localized spectral filtering," in *Proc. 30th Conf. Adv. Neural Inf. Process. Syst. (NIPS)*, Barcelona, Spain, 2016, pp. 3844–3852.
[43] W. Hamilton, Z. Ying, and J. Leskovec, "Inductive representation learning on large graphs," in *Proc. 31st Conf. Adv. Neural Inf. Process. Syst. (NIPS)*, Long Beach, CA, USA, 2017, pp. 1024–1034.
[44] D. Wu and J. Zhao, "Process topology convolutional network model for chemical process fault diagnosis," *Process Saf. Environ. Protection*, vol. 150, pp. 93–109, Jun. 2021.
[45] D. Zhang, E. Stewart, M. Entezami, C. Roberts, and D. Yu, "Intelligent acoustic-based fault diagnosis of roller bearings using a deep graph convolutional network," *Measurement*, vol. 156, May 2020, Art. no. 107585.
[46] S. P. Tiwari, S. Adhikary, and S. Banerjee, "Estimation of chlorophyll-A from oceanographic properties—An indirect approach," in *Proc. IEEE Int. Geosci. Remote Sens. Symp. (IGARSS)*, Jul. 2022, pp. 6872–6875.
[47] B. Zhang, C. Yang, H. Zhu, Y. Li, and W. Gui, "Evaluation strategy for the control of the copper removal process based on oxidation–reduction potential," *Chem. Eng. J.*, vol. 284, pp. 294–304, Jan. 2016.
[48] M. Huang, X. Zhou, T. Huang, C. Yang, and W. Gui, "Dynamic optimization based on state transition algorithm for copper removal process," *Neural Comput. Appl.*, vol. 31, no. 7, pp. 2827–2839, Jul. 2019.
[49] X. Zhou, M. Huang, T. Huang, C. Yang, and W. Gui, "Dynamic optimization for copper removal process with continuous production constraints," *IEEE Trans. Ind. Informat.*, vol. 16, no. 12, pp. 7255–7263, Dec. 2020.



Bei Sun received the Ph.D. degree in control science and engineering from Central South University, Changsha, China, in 2015.

From 2012 to 2014, he was with the Department of Electrical and Computer Engineering, Polytechnic School of Engineering, New York University, New York, NY, USA. From 2016 to 2018, he was a Post-Doctoral Researcher with the School of Chemical Engineering, Aalto University, Espoo, Finland. He is currently an Associate Professor with the School of Automation, Central South University. His research interests include data-driven modeling, optimization, and control of complex industrial processes.



Shengyu Liu received the B.A. degree in automation from the Central South University of Forestry and Technology, Changsha, China, in 2021, where he is currently pursuing the M.A. degree in control science and engineering.

His research interests include soft sensor modeling and deep learning.

Yonggang Li received the M.S. degree in control science and engineering and the Ph.D. degree in control science and engineering from Central South University, Changsha, China, in 2000 and 2004, respectively.

From 2011 to 2012, he was a Visiting Scholar with Curtin University, Perth, WA, Australia. Since 2013, he has been a Full Professor with the School of Information Science and Engineering, Central South University. His current research interests include modeling and optimal control of complex industrial processes, process control, intelligent control system, and knowledge-driven automation.