

Research Article

Lu Yang, Hui Liu*, and Fugang Chen

Soft sensor method of multimode BOF steelmaking endpoint carbon content and temperature based on vMF-WSAE dynamic deep learning

<https://doi.org/10.1515/htmp-2022-0270>

received September 26, 2022; accepted January 11, 2023

Abstract: The difficulty of endpoint determination in basic oxygen furnace (BOF) steelmaking lies in achieving accurate real-time measurements of carbon content and temperature. For the characteristics of serious nonlinearity between process data, deep learning can perform excellent nonlinear feature representation for complex structural data. However, there is a process drift phenomenon in BOF steelmaking, and the existing deep learning-based soft sensor models cannot adapt to changes in the characteristics of samples, which may lead to their performance degradation. To deal with this problem, considering the characteristics of multimode distribution of process data, an adaptive updating deep learning model based on von-Mises Fisher (vMF) mixture model and weighted stacked autoencoder is proposed. First, the stacked autoencoder (SAE) and vMF mixture model are constructed for complex structural data, which can initially establish nonlinear mapping relationships and division of different distributions. Second, for each query sample, the basic SAE network will perform online adaptive fine-tuning according to its data with the same distribution to achieve dynamic updating. Moreover, each

sample is assigned a weight according to its similarity with the query sample. Through the designed weighted loss function, the updated deep network will better match the working conditions of the query sample. Experimental studies with numerical examples and actual BOF steel-making process data are provided to demonstrate the effectiveness of the proposed method.

Keywords: BOF steelmaking endpoint, stacked autoencoder, vMF mixture model, similarity measure, adaptive weighted updating

1 Introduction

Basic oxygen furnace (BOF) steelmaking, as the main technical means of steelmaking, has the characteristics of high production efficiency and relatively low cost [1]. It is a complex physicochemical reaction process that converts raw materials such as molten iron, scrap steel, and pig iron into steel [2]. Figure 1 illustrates the process of BOF steelmaking. Among them, the endpoint carbon content and temperature of molten steel are important indicators in the steelmaking process, and they are the key to determining whether the steel quality is up to the standard. Therefore, it is of great significance to achieve accurate prediction of the endpoint carbon content and temperature for improving steel production efficiency, reducing production costs, and reducing emissions [3].

Currently, there are two main methods to obtain the endpoint carbon content and temperature of molten steel, one is contact measurement and the other is noncontact measurement [4]. The substance is a typical contact measuring device. It immerses the probe in the high-temperature molten pool for measurement and then takes out the used probe from the substance, which increases the cost of steelmaking. In addition, this measurement method can only be measured intermittently [5]. Judging the endpoint

* **Corresponding author: Hui Liu**, Faculty of Information Engineering and Automation, Kunming University of Science and Technology, Kunming 650500, Yunnan Province, China; Yunnan Key Laboratory of Artificial Intelligence, Kunming University of Science and Technology, Kunming 650500, Yunnan Province, China, e-mail: liuhui621@126.com

Lu Yang: Faculty of Information Engineering and Automation, Kunming University of Science and Technology, Kunming 650500, Yunnan Province, China; Yunnan Key Laboratory of Artificial Intelligence, Kunming University of Science and Technology, Kunming 650500, Yunnan Province, China

Fugang Chen: Yunnan Kungang Electronic and Information Science Ltd, Kunming 650302, Yunnan Province, China

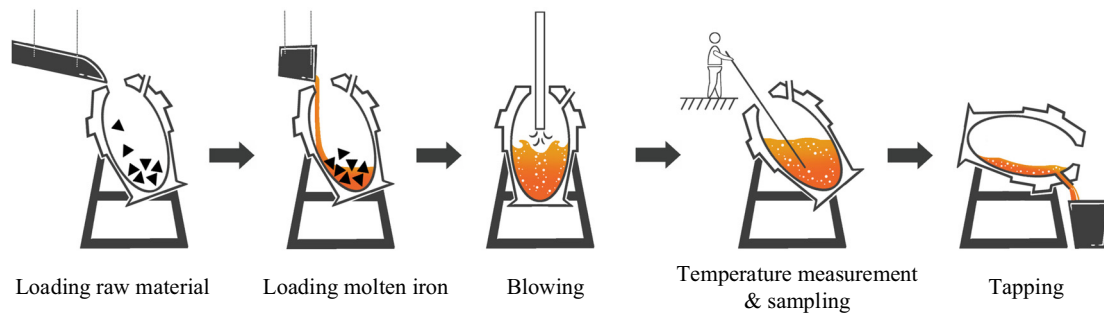


Figure 1: The process step of BOF steelmaking.

by manual experience is one of the most common noncontact methods, and workers judge whether the heat has reached the end point based on experience. However, it will be affected by observers' subjective emotions and proficiency, and it is difficult to achieve accurate judgment of the endpoint [6]. Inspired by this method, some noncontact measurement models based on the flame radiation image of the vessel mouth are proposed to predict the endpoint carbon content and temperature [7–9]. Since the flame radiation image is easily affected by the scene environment, it poses a great challenge to extract the key features of the flame image [10]. With the development of computer technology and data collection technology of BOF steelmaking process, the data-driven endpoint carbon content and temperature soft sensor method in the BOF steelmaking production process have become a current research focus [11]. The researchers use a large amount of steelmaking process data as process variables and employ the endpoint carbon content and temperature as target/output variables to build an intelligent prediction model to predict the carbon content and temperature in molten steel [12–16]. However, these methods are all shallow learning models with a narrow scope of application. For BOF steelmaking process data with highly nonlinear and complex data structures, more reasonable methods are needed to model.

Soft sensor methods of the industrial process are mainly divided into local model modeling and global model modeling. Just-in-time learning (JITL) is the most typical algorithmic framework in the local model modeling. For each query sample, its similar sample set is selected from historical data for local modeling and the output is predicted [17]. In this way, the process time-varying problem can be dealt with refs. [17–19]. For example, Fan *et al.* [18] proposed a similarity criterion based on the Gaussian mixture model-weighted Mahalanobis distance under the JITL framework for soft sensor modeling of non-Gaussian industrial processes. It can be seen that the difficulty of JITL modeling method lies in the use of the similarity criterion [19]. For the complex

industrial process of BOF steelmaking, the distribution of process data varies greatly. If the similarity criterion fails, it is difficult for JITL to obtain a good prediction effect. Global model modeling methods are divided into traditional machine learning modeling [20,21] and soft sensor modeling based on deep learning technology [22–26]. Due to the shallow architecture of traditional machine learning models, their expressive ability is still insufficient to describe complex industrial process data structures, which can easily lead to unstable model performance or even failure [23]. Deep learning technology has powerful feature extraction and nonlinear processing capabilities and can alleviate the restrictive problems such as gradient disappearance in network training through layer-by-layer unsupervised pretraining and supervised fine-tuning [24]. Therefore, deep learning can greatly improve the prediction performance of soft sensor models. For example, Liu *et al.* [22] discussed the unique advantages of deep learning in industrial processes using a modified long short-term memory neural network. Yuan *et al.* [23–25] improved the deep learning network by introducing output information to guide the pretraining process, and achieved good results in industrial systems such as hydrocracking processes. In addition, the introduction of some new theoretical techniques in soft sensor modeling has also achieved good results. For example, Liu *et al.* [27,28], inspired by the idea of transfer learning, introduced the domain adaptive method into soft sensor. Deng *et al.* [29] and Liu *et al.* [30] facilitate the efficient modeling and prediction of soft sensors by introducing active learning.

BOF steelmaking is a complex industrial production process, and there are serious nonlinear characteristics between the data [16]. Currently, the research on the quality prediction of steelmaking is still in the stage of using the shallow model for modeling prediction [31,32], which does not fit well with more complex industrial processes. Deep learning techniques can be used for nonlinear modeling of the complex structured data [23]. Among them, stacked autoencoder (SAE) [33], as a

representative deep learning method, has been widely used in the field of industrial process soft sensing modeling [23–25]. However, most of the aforementioned methods are static models, i.e., they use historical data to train the network offline and then use it online to predict target variables for new samples, with little attention to the dynamic changes of the process.

In addition, in actual steelmaking production, the distribution of the collected process data characteristics varies considerably due to the quality differences of different batches of raw materials, variations in product specifications, and changes in operating conditions [34]. Industrial processes often have multiple operating conditions or cycles [35]. The data of different working conditions or construction periods have different statistical characteristics, and the converter process presents multimode distribution characteristics [36]. Specifically, the production conditions required to smelt steel of different qualities are also different. In the initial stage of blowing, the number of auxiliary materials added and the time period of addition are one of the factors that affect the end of steelmaking. During the steelmaking process, workers have to calculate the amount of steel scrap added according to the composition of molten iron. The amount of scrap steel added will also vary according to different qualities, and the amount of heavy scrap steel added is lower than that of light scrap steel. Workers will also adjust the blowing time of the oxygen lance and control of the amount of oxygen blowing according to the actual situation on site. Overall, data characteristics such as mean, variance, and correlation will differ when a production process is run under different conditions or cycles. Since static models cannot be updated in time according to changes in actual working conditions, these models are difficult to be effective for a long time. Although retraining the deep network can solve the aforementioned problems, it is difficult to meet the real-time requirements in actual production. For the multimode process data of BOF steelmaking with complex data distribution, how to maintain the model performance is the goal of deep learning application in the endpoint soft sensor modeling of the BOF steelmaking. Therefore, an effective deep learning adaptive strategy is needed to avoid the aforementioned problems. Besides, for the multimode distribution phenomenon of industrial process data, researchers often use a mixed model to express multimode behavior in historical databases [16,18,37]. Qi et al. [16] demonstrated the advantages of introducing the von-Mises Fisher (vMF) mixture model [38] for soft sensor modeling in industrial process data with large differences in distribution. This motivates this article to try to combine the vMF mixture model with the SAE deep learning model to solve the problem of the high

nonlinearity of the data in the BOF steelmaking process and the multimode distribution of the data, which leads to the degradation of the model performance.

Based on the aforementioned analysis, in view of the high nonlinearity of the BOF steelmaking process data, this article introduces SAE into the steelmaking process data modeling. For the problem that the performance of the static model degrades due to the process drift, considering that the data characteristics of BOF steelmaking process have the characteristics of multimodal distribution, an adaptive updating deep learning model based on the vMF mixture model and weighted SAE (vMF-WSAE) is proposed. First, through the pretraining and supervised fine-tuning of SAE, features can be extracted from complex structural data and a nonlinear mapping relationship can be initially established to obtain a basic prediction network. At the same time, the vMF mixture model is trained on the historical data, so that the samples in each component have a certain consistency between the process states. When a new query sample is predicted online, the offline-trained SAE is adaptively fine-tuned online with the data within the vMF component to which it belongs, which realizes dynamic updating of the network. Moreover, to pay more attention to the samples that are highly similar to the query samples when the model is updated, the loss function of the model update is improved considering the local characteristics of each vMF component. Different weights are assigned according to the similarity between the query sample and the samples in its belonging component, and the more relevant sample indicates that it has a data structure closer to the query sample of the same working condition. Weighted adaptive updates are performed in this way to improve prediction accuracy. Finally, a numerical simulation example and a simulation experiment and comparative analysis of actual BOF steelmaking data are provided.

In summary, the main contributions of this study are as follows:

- (1) An updating strategy of adaptive dynamic deep learning is proposed, which can dynamically update the deep network according to the distribution characteristics of the query sample, to realize the endpoint soft sensor of multimode BOF steelmaking.
- (2) Through the improved weighted loss function, SAE can quickly adapt to the process running status in the update phase and achieve accurate prediction.
- (3) Experimental studies on numerical examples and practical BOF steelmaking processes are provided. The effectiveness of the proposed method is verified by ablation experiments and compared with other soft sensor methods.

2 Dynamic deep learning model based on vMF-WSAE adaptive updating

In this section, relevant knowledge is introduced in Sections 2.1 and 2.2, including traditional SAE and vMF mixture model algorithms. Then, based on this existing knowledge, the proposed vMF-WSAE-based adaptive deep learning model is introduced. The structure of the theoretical part is shown in Figure 2. First, through the SAE to train the basic deep network on historical data, a nonlinear mapping relationship can be initially established. At the same time, through the improved Bayesian information criterion (iBIC) shown in definition 1, the optimal number of mixed model components is determined, and the complex distributed data are divided into different vMF components. Next, to solve the problem that the static model cannot adapt to the fluctuation of sample characteristics for accurate prediction, an adaptive updating strategy is constructed in Section 2.3. The parameters of the offline model are adaptively updated using the vMF components to which the query samples belong. Moreover, WSAE utilizes an improved network loss function, whose weighting parameters are obtained by the similarity criterion shown in definition 2, so that the model pays more attention to samples that are similar in data structure to the query samples when updating.

2.1 SAE

Autoencoder (AE) is the constituent unit of SAE, which is a three-layer deep learning network including encoder and decoder [33]. Figure 3(a) shows the network structure of AE. AE tries to reconstruct its input data with hidden layer features, $\mathbf{h}$ is usually regarded as a feature representation learned from the input data. The encoder converts the input variable vector $\mathbf{x} \in \mathbb{R}^{d_x}$ into a hidden variable vector $\mathbf{h} \in \mathbb{R}^{d_h}$ through a nonlinear activation function. The decoder maps the hidden layer vector to the output layer to obtain the output variable vector $\tilde{\mathbf{x}} \in \mathbb{R}^{d_x}$ as follows:

$$\mathbf{h} = f(\mathbf{W}\mathbf{x} + \mathbf{b}), \quad (1)$$

$$\tilde{\mathbf{x}} = \tilde{f}(\tilde{\mathbf{W}}\mathbf{h} + \tilde{\mathbf{b}}), \quad (2)$$

where f and $\tilde{f}$ are the nonlinear activation functions of the hidden layer and output layer, respectively; $\mathbf{W}$ and $\mathbf{b}$ are the weight matrix and bias vector of the encoder, respectively; $\tilde{\mathbf{W}}$ and $\tilde{\mathbf{b}}$ are the weight matrix and bias vector of the decoder, respectively. The set of parameters to be optimized by the AE can be expressed as $\boldsymbol{\theta}_{\text{AE}} = \{\mathbf{W}, \tilde{\mathbf{W}}, \mathbf{b}, \tilde{\mathbf{b}}\}$. Assuming there are N training data, the parameters and hidden layer features $\mathbf{H} = \{\mathbf{h}_i\}_{i=1}^N$ of the model are obtained by minimizing the following objective function through multiple iterations:

$$J_{\text{AE}}(\boldsymbol{\theta}_{\text{AE}}) = \sum_{i=1}^N \|\tilde{\mathbf{x}}_i - \mathbf{x}_i\|^2 / 2N. \quad (3)$$

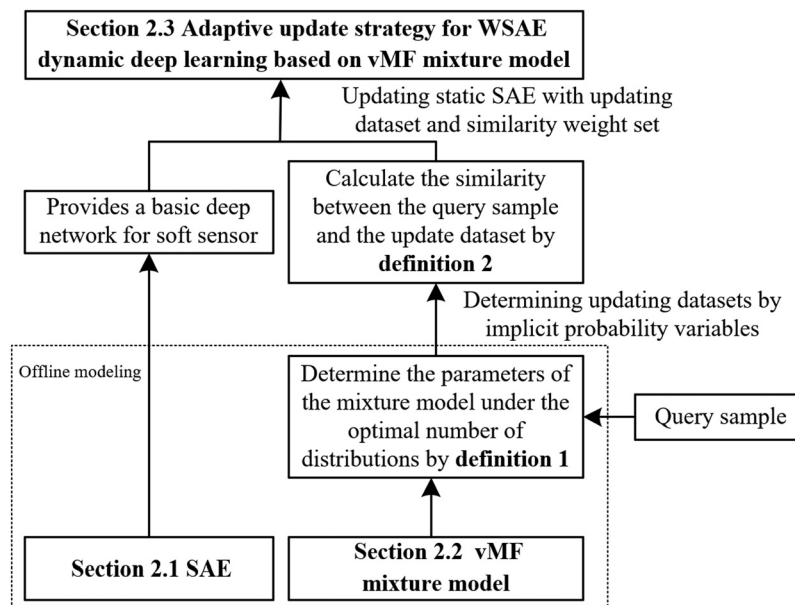


Figure 2: The structure of the proposed method.

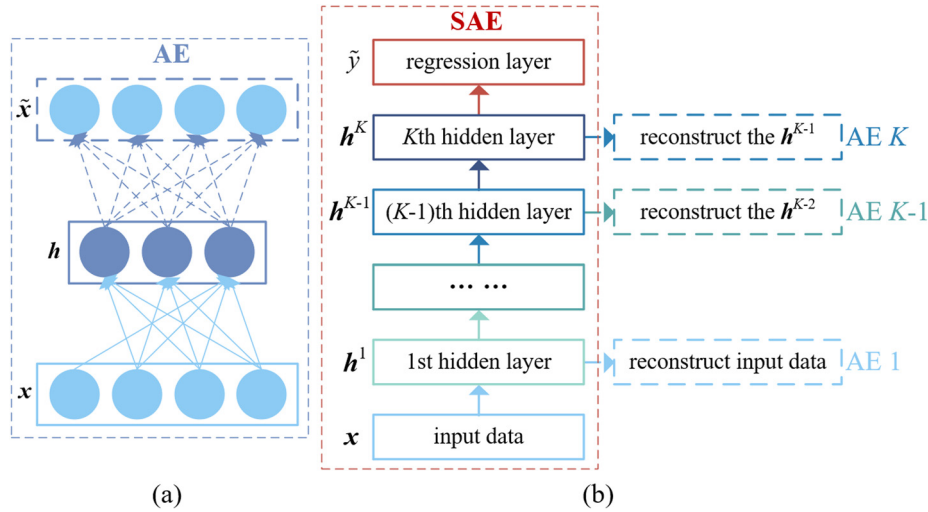


Figure 3: Structure for AE and SAE. (a) AE network. (b) SAE network with K AEs.

To extract deeper abstract features for prediction and improve the nonlinear processing capability of the network, multiple AEs with a single structure are stacked to form a SAE deep neural network, and the basic structure is shown in Figure 3(b). Here, h^k represents the hidden variable vector at the k th hidden layer of the deep network. The training process is divided into an unsupervised pretraining stage and a supervised fine-tuning stage. First, the network parameters $\{\theta_{ij}\}_{i=1}^K$ of each AE are pre-trained greedily in an unsupervised manner. In the fine-tuning stage, the weights obtained from pretraining are regarded as the initial weights of the stacked network. The hidden layer in the last AE is connected to the regression layer for the prediction output. There is an overall fine-tuning of the weights and biases with minimizing the mean square error of the label and prediction output as the objective function, thus optimizing the parameters of all layers.

2.2 vMF mixture model

To make samples within the same distribution component have a certain consistency between process states, this section introduces an algorithm for partitioning the distribution. The vMF distribution is considered a popular direction distribution [39]. For a D -dimensional unit vector $\mathbf{x} = [x_1, x_2, \dots, x_D]^T$, its probability density function can be expressed as follows:

$$v(\mathbf{x}|\boldsymbol{\mu}, \lambda) = \frac{\lambda^{\frac{D}{2}-1}}{(2\pi)^{\frac{D}{2}} I_{\frac{D}{2}-1}(\lambda)} e^{\lambda \boldsymbol{\mu}^T \mathbf{x}}, \quad (4)$$

where $\|\boldsymbol{\mu}\| = 1$, $\lambda \geq 0$, and $I_{(D/2)-1}(\cdot)$ indicates the modified Bessel function of the first kind of order $(D/2) - 1$.

The probability density function $v(\mathbf{x}|\boldsymbol{\mu}, \lambda)$ is described by the mean direction $\boldsymbol{\mu}$ and the concentration parameter λ . The greater value of λ means a more robust concentration around the mean direction.

For a set of data that contains N vectors $\mathbf{X} = \{\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_N\}$, where each vector $\mathbf{x}_i = [x_{i1}, x_{i2}, \dots, x_{iD}]^T$ indicates a D -dimensional data. The probability density of its vMF mixture model is defined as follows:

$$p(\mathbf{X}|\boldsymbol{\mu}, \boldsymbol{\lambda}, \boldsymbol{\tau}) = \prod_{i=1}^N \sum_{m=1}^M \tau_m v(\mathbf{x}_i|\boldsymbol{\mu}_m, \boldsymbol{\lambda}_m), \quad (5)$$

where M is the number of mixture components and τ_m is the weight of the m th vMF component.

For the vMF mixture model, if the traditional expectation maximization algorithm is used for training, it may lead to high computational cost and overfitting due to the complexity of the model or poor initialization. To solve the problem of parameter estimation, the method of variational inference (VI) is used for Bayesian estimation of vMF mixture models [40]. This approach provides a variational treatment for the vMF model by approximating posterior distributions.

First, a set of latent probability variables is defined as $\mathbf{Z} = \{\mathbf{z}_1, \mathbf{z}_2, \dots, \mathbf{z}_N\}$. Each latent probability variable $\mathbf{z}_i = \{z_{i1}, z_{i2}, \dots, z_{iM}\}$ corresponds to a D -dimensional observation vector $\mathbf{x}_i = [x_{i1}, x_{i2}, \dots, x_{iD}]^T$, where $\sum_{m=1}^M z_{im} = 1$ and $0 \leq z_{im} \leq 1$. If $z_{im} = 1$, then the sample $\mathbf{x}_i$ completely belongs to the m th vMF distribution component. The probability density function can be expressed as follows:

$$p(\mathbf{X}|\mathbf{Z}, \boldsymbol{\mu}, \boldsymbol{\lambda}) = \prod_{i=1}^N \sum_{m=1}^M v(\mathbf{x}_i|\boldsymbol{\mu}_m, \boldsymbol{\lambda}_m)^{z_{im}}. \quad (6)$$

Next, in VI methods, a lower bound called the evidence lower bound is provided to approximate the posterior

distribution, whose evidence lower bound is defined as follows:

$$\begin{aligned} L(q) &= \ln \frac{p(\mathbf{X}, \mathbf{Z}, \boldsymbol{\mu}, \boldsymbol{\lambda}, \boldsymbol{\tau})}{q(\mathbf{Z}, \boldsymbol{\mu}, \boldsymbol{\lambda}, \boldsymbol{\tau})} \\ &= \ln \frac{p(\mathbf{X}|\mathbf{Z}, \boldsymbol{\mu}, \boldsymbol{\lambda})p(\boldsymbol{\mu}, \boldsymbol{\lambda})p(\mathbf{Z}|\boldsymbol{\tau})p(\boldsymbol{\tau})}{q(\mathbf{Z})q(\boldsymbol{\mu}, \boldsymbol{\lambda}, \boldsymbol{\tau})}, \end{aligned} \quad (7)$$

where $p(\mathbf{X}, \mathbf{Z}, \boldsymbol{\mu}, \boldsymbol{\lambda}, \boldsymbol{\tau})$ represents the joint distribution of all random variables, and $q(\mathbf{Z}, \boldsymbol{\mu}, \boldsymbol{\lambda}, \boldsymbol{\tau})$ is the approximation for the posterior distributions. The variation factor $q(\mathbf{Z})$ is optimized by maximizing the evidence lower bound $L(q)$. The optimal variational posterior distribution of the latent probability variable $\mathbf{Z}$ is defined as follows:

$$q^*(\mathbf{Z}) = \prod_{i=1}^N \prod_{m=1}^M \xi_{im}^{z_{im}}, \quad (8)$$

where $q^*(\cdot)$ represents the optimized posterior distribution and ξ_{im} is given by

$$\xi_{im} = \frac{e^{\ln \rho_{im}^*}}{\sum_{j=1}^M e^{\ln \rho_{ij}^*}}, \quad (9)$$

where $\sum_{m=1}^M \xi_{im} = 1$ and $0 \leq \xi_{im} \leq 1$. For $q^*(\mathbf{Z})$, we obtain $(z_{im})_{\mathbf{Z}} = \xi_{im}$. $\ln \rho_{im}^*$ is obtained by:

$$\begin{aligned} \ln \rho_{im}^* &= \varphi(\alpha_m) - \varphi(\hat{\alpha}) - \frac{D}{2} \ln 2\pi + \left(\frac{D}{2} - 1\right)(\varphi(\alpha_m) - \ln b_m) \\ &\quad + \frac{a_m}{b_m} \boldsymbol{\omega}_m^T \mathbf{x}_i - \ln I_{\frac{D}{2}-1}(\hat{\lambda}_m) - \left(\frac{\partial}{\partial \lambda_m} \ln I_{\frac{D}{2}-1}(\hat{\lambda}_m)\right) \left(\frac{a_m}{b_m} - \hat{\lambda}_m\right), \end{aligned} \quad (10)$$

where $\varphi(\alpha) = \frac{d}{d\alpha} \ln \alpha$ and $\hat{\alpha} = \sum_{m=1}^M \alpha_m$. Finally, when using VI to estimate the parameters of the vMF mixture model, $\alpha_m, \beta_m, a_m, b_m$ and $\boldsymbol{\omega}_m$ is recorded as the posterior parameters of the m th vMF component, as follows:

$$\begin{aligned} \alpha_m &= \alpha_{0,m} + \sum_{i=1}^N \xi_{im}, \\ \beta_m &= \left\| \beta_{0,m} \boldsymbol{\omega}_{0,m} + \sum_{i=1}^N \xi_{im} \mathbf{x}_i \right\|, \\ \boldsymbol{\omega}_m &= \left(\beta_{0,m} \boldsymbol{\omega}_{0,m} + \sum_{i=1}^N \xi_{im} \mathbf{x}_i \right) \beta_m^{-1}, \\ a_m &= a_{0,m} + \left(\frac{D}{2} - 1\right) \sum_{i=1}^N \xi_{im} \\ &\quad + \left(\frac{\partial}{\partial \beta_m \lambda_m} \ln I_{\frac{D}{2}-1}(\beta_m \hat{\lambda}_m) \right) \beta_m \hat{\lambda}_m, \\ b_m &= b_{0,m} + \sum_{i=1}^N \xi_{im} \left(\frac{\partial}{\partial \lambda_m} \ln I_{\frac{D}{2}-1}(\hat{\lambda}_m) \right) \\ &\quad + \beta_{0,m} \left(\frac{\partial}{\partial \beta_{0,m} \lambda_m} \ln I_{\frac{D}{2}-1}(\beta_{0,m} \hat{\lambda}_m) \right). \end{aligned} \quad (11)$$

For the trained finite vMF mixture model, the total space of BOF historical data is divided into multiple vMF components, each of which represents a distribution with similar parameters. Each process sample has a corresponding latent probability variable $\mathbf{z}_i = \{z_{i1}, z_{i2}, \dots, z_{iM}\}$, and z_{im} can be considered as the probability that the sample belongs to the m th vMF component. For example, if the largest element in $\mathbf{z}_i$ is z_{i1} , the sample is considered to belong to the first vMF component.

2.3 An adaptive dynamic deep learning strategy based on vMF mixture model and WSAE

As mentioned in Section 1, for most deep learning-based soft sensor methods, the model is trained offline and then used online to make predictions, and it has no mechanism for model updating. That is, although SAE can learn the deep features of the data for nonlinear modeling after offline training, when the model is applied online, frequent changes in operating conditions or data characteristics may cause the model performance to degrade. In the actual BOF steelmaking process, affected by the fluctuation of raw material quality of different batches, changes in product specifications and changes in operating conditions, the process data presents the characteristics of multimode distribution. Based on the aforementioned analysis, this section proposes a dynamic deep learning adaptive updating strategy based on the vMF mixture model and weighted SAE (vMF-WSAE) to maintain the performance of the model and track the working conditions in time. The vMF-WSAE modeling framework is divided into two main stages, including offline modeling and online adaptive updating.

2.3.1 Base model constructed based on SAE and vMF mixture model

The BOF steelmaking data has serious nonlinear characteristics. In the offline modeling stage, SAE is used to extract the features of historical data and initially construct nonlinear mapping relationships to obtain the basic deep network. To obtain the weight and bias parameters of each AE, the stacked network is hierarchically pretrained using the limited-memory BFGS (LBFGS) [41] algorithm. Given the input data $\mathbf{X} = \{\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_N\}$, the first-level hidden layer feature $\{\mathbf{h}_1^1, \mathbf{h}_2^1, \dots, \mathbf{h}_N^1\}$ can be learned by the first AE pretraining. The parameter $\boldsymbol{\theta}_1 = \{\mathbf{W}^1, \hat{\mathbf{W}}^1, \mathbf{b}^1, \hat{\mathbf{b}}^1\}$ of AE 1 is obtained by minimizing the loss function shown in

equation (3). After training, its encoder part remains in the first layer of the SAE network structure, and $\mathbf{h}^1$ will be used as the input part of AE 2 to extract the next-level hidden layer feature $\{\mathbf{h}_1^2, \mathbf{h}_2^2, \dots, \mathbf{h}_N^2\}$. Likewise, deep features can be obtained by layer-wise pretraining in a similar manner. Assuming that the feature data extracted in AE k ($k = 2, \dots, K - 1$) is $\{\mathbf{h}_1^k, \mathbf{h}_2^k, \dots, \mathbf{h}_N^k\}$, it will be used to extract the feature data $\{\mathbf{h}_1^{k+1}, \mathbf{h}_2^{k+1}, \dots, \mathbf{h}_N^{k+1}\}$ of level $k + 1$. AE ($k + 1$) is trained by minimizing the following objective function to obtain the network parameters θ_{k+1} :

$$J_{\text{AE}}^{k+1}(\theta_{k+1}) = \sum_{i=1}^N \|\tilde{\mathbf{h}}_i^k - \mathbf{h}_i^k\|^2 / 2N. \quad (12)$$

By pretraining in the aforementioned manner, the initial parameter $\{\mathbf{W}_k, \mathbf{b}_k\}_{k=1,2,\dots,K}$ of the SAE network and the feature vector $\{\mathbf{h}_i^k\}_{k=1,2,\dots,K}$ corresponding to the K hidden layers can be obtained. After that, the output layer of the target variables is added to the top hidden layer feature $\mathbf{h}^K$ to construct the SAE regression prediction network. The weight matrix and the bias vector of the output layer are denoted by $\mathbf{W}_0$ and $\mathbf{b}_0$, respectively, and they are randomly initialized first. The predicted output $\tilde{\mathbf{y}}$ is computed by forward propagation as follows:

$$\tilde{\mathbf{y}} = f(\mathbf{W}_0 \mathbf{h}^K + \mathbf{b}_0). \quad (13)$$

Then, the LBFGS algorithm is used to minimize the error between the predicted value and the true value to fine-tune the parameter $\theta_{\text{SAE}} = \{\mathbf{W}_k, \mathbf{b}_k, \mathbf{W}_0, \mathbf{b}_0\}_{k=1,2,\dots,K}$ of the entire network, the loss function of the network is calculated as follows:

$$J_{\text{SAE}}(\theta_{\text{SAE}}) = \sum_{i=1}^N (y_i - \tilde{y}_i)^2 / 2N. \quad (14)$$

At the same time, considering the characteristics of process data showing multimode distribution, the vMF mixture model is trained to distinguish the differences between the distributions of steelmaking process data. The model parameters are estimated by the method of VI, thereby dividing the BOF steelmaking process data into different distributions, and the samples in the same distribution have the same data characteristics to a certain extent. It will be used for subsequent adaptive updates of the model.

But in mixture modeling, the determination of the number of components is a key issue. For the BOF steelmaking process data, if the mixture model contains too many distributions (components), the model may overfit the observations; if there are too few components, the model may be undertrained. Therefore, a reasonable calculation method is needed to determine the number of components of the mixture model. Inspired by Mehrjou

et al. [42], this article defines the evaluation index of vMF mixture model based on the iBIC to determine the optimal number of distributions.

Definition 1. The definition of the number of components to determine the vMF mixture model based on iBIC is as follows:

Given a training set $\mathbf{X}$ of N samples, the probability density function of its vMF mixture model is $p(\mathbf{X}|\boldsymbol{\mu}, \boldsymbol{\Lambda}, \boldsymbol{\tau})$, and iBIC is calculated as follows:

$$\begin{aligned} \text{iBIC} = & -2 \log p(\mathbf{X}|\boldsymbol{\mu}, \boldsymbol{\Lambda}, \boldsymbol{\tau}) + d \log N - \sum_{m=1}^M \log \tau_m \\ & - d \log 2\pi + \sum_{m=1}^M d_m \log \tau_m + \sum_{m=1}^M \log |J_m|, \end{aligned} \quad (15)$$

where $d_m = D + D(D + 1)/2$ denotes the number of free parameters for the m th component, $d = d_m \times M + M$ is the number of free parameters in the mixture model, and J_m is the Fisher information matrix of the m th component. The smaller the iBIC value is, the better the effect of the component M is.

The defined iBIC can be used to select the optimal number of components for the vMF mixture model offline. First, a vMF mixture model with specific M components is trained and then evaluated using iBIC. Next, after evaluating all possible M values, the optimal number of components M can be determined.

2.3.2 Adaptive updating strategy based on vMF-WSAE

The SAE trained in the offline phase is essentially a static model. To maintain the performance of the model and keep track of the working conditions in time, an adaptive updating strategy is designed. When a query sample arrives, the latent probability of belonging to each vMF component is calculated, and the data belonging to the same distribution as the query sample are used to form its updating dataset. In addition, the update dataset has different degrees of similarity with the query sample, and the trend of the model update should be closer to the working conditions of the query sample. Therefore, this study improves the loss function in the fine-tuning stage of SAE and dynamically assigns different weights according to the similarity between the updating sample set and the query sample.

Among them, the traditional Euclidean distance (ED)-based similarity criterion has defects in measuring the data of the BOF steelmaking process [16], which may

lead to the performance of the model degraded after updating. To more reasonably calculate the similarity between the update sample set and the query sample, considering the local characteristics of different distributions in the mixture model, a similarity criterion based on the weighted Euclidean distance (WED) is defined. In the traditional similarity criterion, the contributions of features to the similarity calculation are treated equally. If the features that are more related to the output variables are given a larger weight when calculating the similarity, the similarity of the calculated similar samples on the output variable is also higher. Therefore, different weights are introduced to different features in the ED calculation. The maximum mutual information coefficient (MIC) has been shown to be effective in mining linear and nonlinear relationships between variables [43], and the weights are determined by the MIC between the input features and the output variables. The defined WED not only retains the stability of ED for measurement but also achieves a comprehensive evaluation of characteristics and output endpoint using a weighting method. In addition, the correlations of the same features with endpoint carbon content and temperature were different within different vMF compositions, and the calculated MICs were also different.

Definition 2. The similarity criterion based on the WED distance is defined as follows:

Step 1: Given a dataset $(\mathbf{X}, \mathbf{Y}) = [\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_D, \mathbf{y}]^T$ of n sample, the mutual information for the d th dimensional feature $\mathbf{x}_d$ of the sample and the target variable vector $\mathbf{y}$ is calculated as follows:

$$MI(\mathbf{x}_d, \mathbf{y}) = \sum p(\mathbf{x}_d, \mathbf{y}) \log \frac{p(\mathbf{x}_d, \mathbf{y})}{p(\mathbf{x}_d)p(\mathbf{y})}, \quad d \in (1, D). \quad (16)$$

Here, $p(\mathbf{x}_d, \mathbf{y})$ represents the joint probability distribution of the two variables, and $p(\mathbf{x}_d)$ and $p(\mathbf{y})$ are the marginal probability distributions of $\mathbf{x}_d$ and $\mathbf{y}$, respectively.

Step 2: Divide the scatter plots formed by $\mathbf{x}_d$ and $\mathbf{y}$ into $a \times b$ grids, and calculate the mutual information of each grid separately. Since there are many grid division methods, the maximum value of MI under different division methods is selected to obtain the MIC. It is calculated as follows:

$$MIC(\mathbf{x}_d, \mathbf{y}) = \max_{a \cdot b < n^{0.6}} \frac{MI(\mathbf{x}_d, \mathbf{y})}{\log_2 \min(a, b)}. \quad (17)$$

Step 3: From Step 2, the MIC value between the input features and the target variable can be obtained, and the weighting coefficient $\{w_d\}_{d=1}^D$ of the measurement can be obtained by calculating the MIC ratio by

$$w_d = \left| \frac{MIC(\mathbf{x}_d, \mathbf{y})}{\sum_{d=1}^D MIC(\mathbf{x}_d, \mathbf{y})} \right|. \quad (18)$$

Then $\mathbf{w} = (w_1, \dots, w_d, \dots, w_D)$ is the weight of the feature measure, and a higher weight indicates a greater degree of association with the target variable and a greater contribution to the calculation of similarity.

Step 4: Assuming two random samples $\mathbf{x}_i, \mathbf{x}_j \in \mathbf{X}$, the WED distance between two samples is calculated as follows:

$$WED(\mathbf{x}_i, \mathbf{x}_j) = \sqrt{\sum_{d=1}^D w_d \cdot (x_{id} - x_{jd})^2}. \quad (19)$$

The smaller the distance is, the more similar the historical sample is to the query sample. To distinguish the similarity of different samples, the designed similarity measure function based on WED distance is as follows:

$$\varepsilon = \exp(-WED(\mathbf{x}_i, \mathbf{x}_j)^2). \quad (20)$$

It can be seen that the larger ε is, the more similar the two samples are, and ε is in the range of 0–1.

The vMF-WSAE algorithm diagram is shown in Figure 4, and the detailed algorithm steps are as follows:

Step 1: After offline modeling, assume that the input part of the new query sample is $\mathbf{x}_q$. Calculate the latent probability variable $\{z_{qm}\}_{m=1,2,\dots,M}$ for the vMF mixture model according to equation (9). Assuming that $\mathbf{x}_q$ belongs to the m th vMF distribution, and the index of the samples in this distribution in the historical data set is $\{u_1, u_2, \dots, u_{N_m^{\text{vMF}}}\}$, the representation of the online updating dataset is $\mathbf{B}_u = \{\mathbf{x}_u, \mathbf{y}_u\}_{u=1}^{N_m^{\text{vMF}}}$.

Step 2: Calculate the similarity weighting coefficient $\mathbf{w}_m = (w_1, w_2, \dots, w_D)$ under the vMF component to which the query sample belongs according to the MIC. Next, calculate the WED similarity between the query sample and all samples in the updating dataset $\mathbf{B}_u$ according to definition 2, denoted as $\varepsilon_m = \{\varepsilon_u\}_{u=1}^{N_m^{\text{vMF}}}$.

Step 3: To enhance the influence of similar samples on network training when calculating the loss, map ε_m to the $[0, 2]$ interval as the weight of the loss function updated by the WSAE model. The mapping method is expressed as follows:

$$\varepsilon_m = 2 \frac{\varepsilon_m - \varepsilon_{\min}}{\varepsilon_{\max} - \varepsilon_{\min}}, \quad (21)$$

where $\varepsilon_{\min}$ and $\varepsilon_{\max}$ are the minimum and maximum values in ε_m , respectively.

Step 4: Implement adaptive variable weight fine-tuning for static SAE. In detail, although these updating samples belong to the same distribution to the query

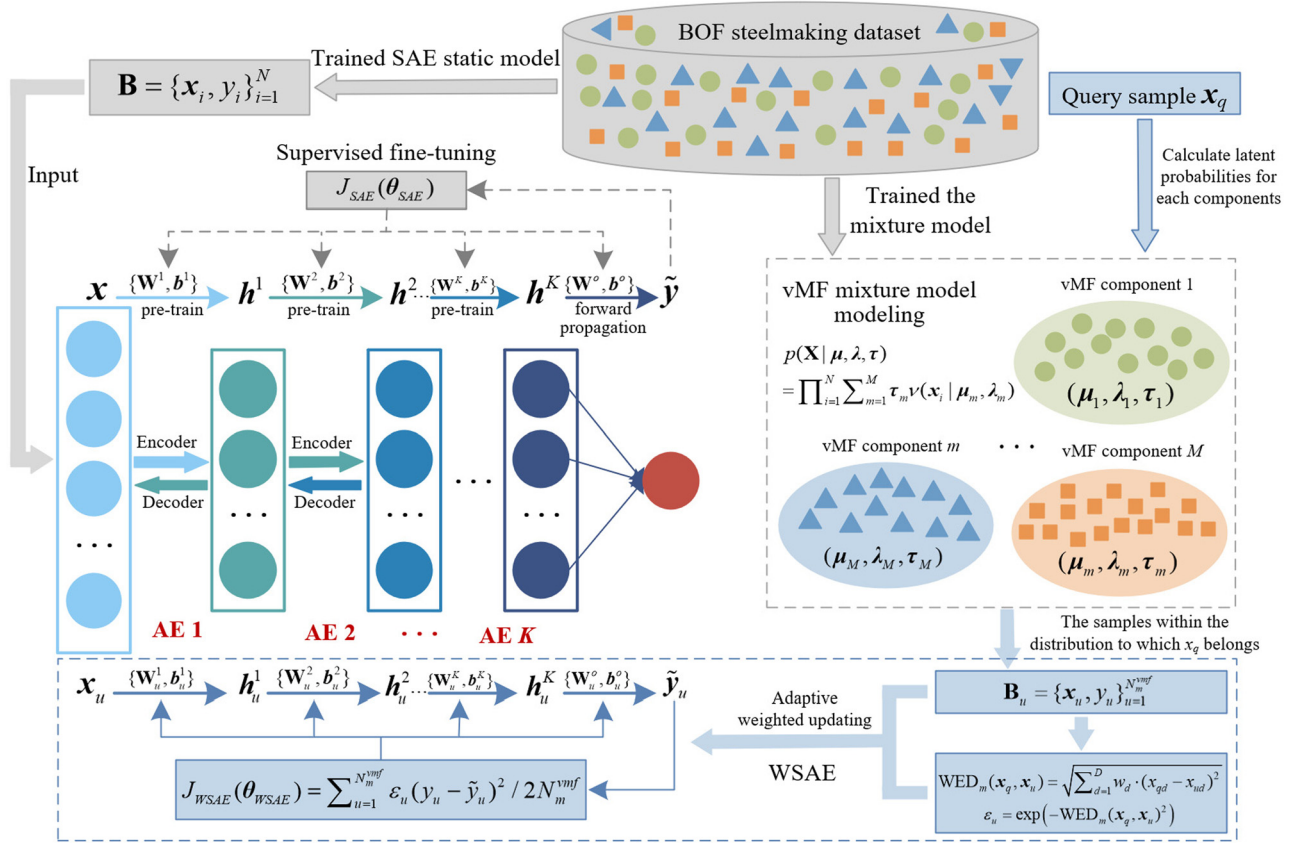


Figure 4: Algorithm diagram of vMF-WSAE.

sample, they also have different similarities. To pay more attention to the samples that are highly similar to the query sample when the model is updating, an updating strategy of variable-weighted SAE is constructed. Updating dataset B_u and similarity weight set $\epsilon_m = \{\epsilon_{u,m}\}_{m=1}^{N_m^{vmf}}$ are used to update the static SAE model. First, through forward propagation, the predicted output variable $\{\tilde{y}_{u,m}\}_{m=1}^{N_m^{vmf}}$ of B_u can be obtained. Then, the entire SAE network is updating with a weighted loss function defined as follows:

$$J_{WSAE}(\theta_{WSAE}) = \sum_{u=1}^{N_m^{vmf}} \epsilon_u (y_u - \tilde{y}_u)^2 / 2N_m^{vmf}. \quad (22)$$

The network is iterated in turn by the LBFGS algorithm to obtain the updated network model parameters $\theta_{WSAE} = \{W_k^*, b_k^*, W_0^*, b_0^*\}_{k=1,2,\dots,K}$.

Step 5: The output of query sample can be quickly predicted by forward propagation based on the updated model parameters θ_{WSAE} . For the next query sample, the aforementioned process of latent probability calculation of the vMF mixture model, similarity calculation, weight calculation, model update, and target variable prediction will be performed again.

Among them, when using the LBFGS algorithm to iteratively update the network parameters to minimize the objective function J_{WSAE} , as with the backpropagation algorithm, the critical step is to compute the partial derivatives of the function. For a single-labeled training sample (x_u, y_u) in WSAE, the calculation process of the partial derivatives of its parameters is as follows:

- (1) For layers 2 and 3 of the network up to the output layer l_n , the activation values of each hidden layer are obtained using forward propagation.
- (2) For the output layer l_n , the following derivative terms are computed:

$$\delta^{(l_n)} = \frac{\partial}{\partial z^{(l_n)}} J_{WSAE}(W, b; x_u, y_u) = \epsilon_u (a^{(l_n)} - y_u) \cdot f'(z^{(l_n)}), \quad (23)$$

where $z^{(l_n)}$ is the weighted sum of the output layer l_n , the calculation methods is $z^{(l_n)} = W^{(l_n-1)}a^{(l_{n-1})} + b^{(l_{n-1})}$, and ϵ_u is the similarity between the sample and the query sample, which is calculated by definition 2.

- (3) For each hidden layer of $l = l_{n-1}, l_{n-2}, \dots, l_2$, compute the following derivatives:

$$\delta^{(l)} = ((W^{(l)})^T \cdot \delta^{(l+1)}) \cdot f'(z^{(l)}). \quad (24)$$

(4) Calculate the final required partial derivatives:

$$\frac{\partial}{\partial \mathbf{W}^{(l)}} J_{\text{WSAE}}(\mathbf{W}, \mathbf{b}; \mathbf{x}_u, y_u) = \delta^{(l+1)} \cdot (a^{(l)})^T, \quad (25)$$

$$\frac{\partial}{\partial \mathbf{b}^{(l)}} J_{\text{WSAE}}(\mathbf{W}, \mathbf{b}; \mathbf{x}_u, y_u) = \delta^{(l+1)}. \quad (26)$$

From the aforementioned equation, the partial derivatives of the total objective function can be obtained and the optimal values can be obtained iteratively updating the parameters.

It can be seen that the updating of the vMF-WSAE is based on the global regression model trained in the offline phase. After the arrival of the sample to be tested, the samples in the vMF distribution to which it belongs are selected from the historical samples as the online update dataset. Through the weighting of the loss function, the model pays more attention to the samples that are highly similar to the data patterns of the query sample when updating. Therefore, only a small amount of iterative calculation is required to update the regression model online adaptively to improve the prediction performance of the model.

3 Endpoint carbon content and temperature soft sensor method based on vMF-WSAE for multimode BOF steelmaking process data

Due to the phenomenon of process drift, the distribution of BOF steelmaking process data will also change, and the characteristics of the data show the characteristics of multimodal distribution, making training an effective deep learning model a challenging problem. In this section, an adaptive update soft sensor modeling method based on vMF-WSAE is introduced, which will be used for endpoint carbon content and temperature prediction in the BOF steelmaking process.

Figure 5 shows the offline training and online updating process of the vMF-WSAE-based soft sensor model. The specific steps of the offline training part are as follows:

- (1) Use SAE to pretrain and fine-tune the historical dataset of BOF steelmaking, and save the basic SAE model for subsequent updating.
- (2) In addition, the vMF mixture model is trained on the historical data of BOF steelmaking by means of VI.

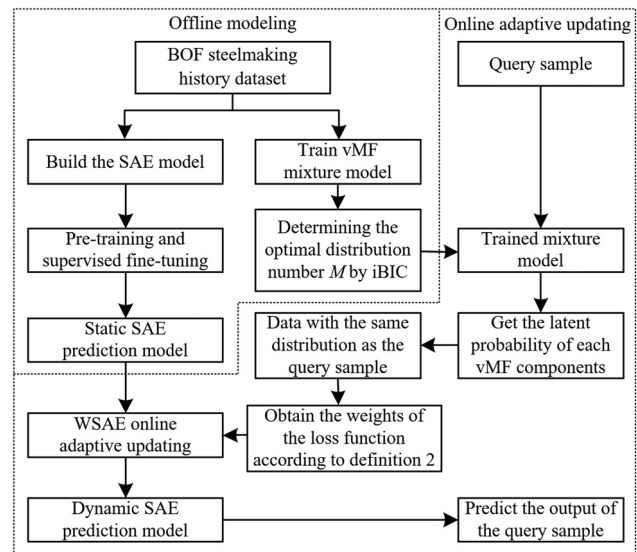


Figure 5: The flowchart of the proposed vMF-WSAE-based soft sensor model.

- (3) The vMF mixture model with different number of components was evaluated by iBIC to determine the optimal number of components.
- (4) According to equations (16)–(18), the correlation degree of different features in each vMF component to the target variable is calculated, respectively, and the weighting coefficient of the WED metric criterion under each component is obtained.

The specific steps for the online updating section are as follows:

- (1) For a query sample, calculate the probability that belongs to each trained vMF component by equation (9). The samples in the vMF component to which the query sample belong are used as the updating dataset.
- (2) Calculate the WED between the query sample and the updating samples by equation (20), and convert it into the updating weight of WSAE by equation (21).
- (3) Adaptive updating of weighted SAE using updating dataset and similarity weight set. Then, the output of the query sample can be quickly predicted based on the updated model parameters.

4 Experimental results and analysis

In this section, a numerical example is first utilized to illustrate the effectiveness of the proposed vMF-WSAE model over the original SAE. Then, a series of experiments

are carried out on BOF steelmaking process data to demonstrate the effectiveness and advantages of the soft sensor modeling method proposed in this article. The simulation configurations of the computer are as follows: RAM: 16GB, CPU: R7-4800H (AMD Ryzen Processor @2.9 GHz), and MATLAB version: 2020a.

The root mean square error (RMSE) and mean absolute percentage error (MAPE) are used to evaluate the prediction performance of the soft sensor model, as shown in equations (33) and (34). The smaller the value, the better the prediction performance of the model.

$$\text{RMSE} = \sqrt{\frac{1}{N_{\text{test}}} \sum_{i=1}^{N_{\text{test}}} (\tilde{y}_i - y_i)^2}, \quad (27)$$

$$\text{MAPE} = \frac{1}{N_{\text{test}}} \sum_{i=1}^{N_{\text{test}}} \left| \frac{\tilde{y}_i - y_i}{y_i} \right|, \quad (28)$$

where N_{test} is the number of samples in the test set, and y_i and $\tilde{y}_i$ are the actual and predicted values, respectively, of the output endpoint of the i th sample.

Also, in the prediction of BOF endpoint, regression accuracy (RA) is an important evaluation criterion, which represents the prediction accuracy of the model on the test set within the allowable error range. The calculation method of RA in this article is as follows:

$$\text{RA} = \frac{1}{N_{\text{test}}} \sum_{i=1}^{N_{\text{test}}} \text{Match}(\tilde{y}_i, y_i) \cdot 100\%, \quad (29)$$

where $\text{Match}(\cdot)$ is a conditional function, and if the difference between the two inputs is within the tolerable error (Te) range, it will output 1, which is a prediction hit. This function formulation can be expressed as follows:

$$\text{Match}(\tilde{y}_i, y_i) = \begin{cases} 1, & |\tilde{y}_i - y_i| \leq \text{Te} \\ 0, & |\tilde{y}_i - y_i| > \text{Te} \end{cases}. \quad (30)$$

4.1 Numerical example

To verify the performance of the proposed method, the numerical case described in reference [44] was used for experiments. The seven input variables are simulated as linear combinations of a total of five source variables as follows:

$$\begin{cases} s_1(k) = 2 \cos(0.08i) \sin(0.06i), \\ s_2(k) = \sin(0.3i) + 3 \cos(0.1i), \\ s_3(k) = \sin(0.4i) + 3 \cos(0.1i), \\ s_4(k) = \cos(0.1i) - \sin(0.05i), \\ s_5(k) = \text{uniformly distributed noise in } [-1, 1]. \end{cases} \quad (31)$$

In total, 1,500 samples belonging to three different modes are generated as follows:

The first 500 samples are generated in mode 1, with

$$z = A^T s + e_1, \quad y = 0.8z_1 + 0.6z_2 + 1.5z_3. \quad (32)$$

The second 500 samples are generated in mode 2 as follows:

$$z = (AB)^T s + e_1, \quad y = 2.4z_2 + 1.6z_3 + 4z_4. \quad (33)$$

The last 500 samples are generated in mode 3 as follows:

$$z = (AB^2)^T s + e_1, \quad y = 1.2z_1 + 0.4z_2 + z_4. \quad (34)$$

Here z is the input variable vector, $e_1 \sim N(0, 0.01)$ is the Gaussian noise, $A \in \mathbb{R}^{5 \times 7}$ is a random coefficient matrix, and $B \in \mathbb{R}^{7 \times 7}$ is a lower triangular matrix of all ones. The output variable is defined as follows:

$$y = y + e_2, \quad e_2 \sim N(0, 0.1). \quad (35)$$

For each mode, 250 samples are selected as the training set, and the remaining 250 samples are used as the testing set for the soft sensor. Thus, there are 750 samples in each training set and the testing set.

It can be seen that there is a complex nonlinear relationship between the input and output variables. In addition, data consists of multiple patterns. To verify the effectiveness of the proposed method, in addition to vMF-WSAE, soft-sensor models of partial least squares (PLS), BP neural network (BPNN), and SAE are also employed for output prediction. In the vMF-WSAE-based method, the probability distribution of the training set is approximated as a mixture of 3 vMF distributions, and the prediction network has the structure of [7-4-1]. The maximum number of iterations for pretraining and offline fine-tuning is 400, and the maximum number of iterations for updating the network online is 100. Furthermore, both BPNN and SAE have the same network structure as [7-4-1]. For NN-based regression network, the network parameters are randomly initialized. The maximum number of iterations for pretraining and fine-tuning of SAE is 400. These hyperparameters are the best results under their respective methods.

Table 1 outlines the results of PLS, BPNN, SAE, and vMF-WSAE in terms of RMSE and MAPE. From Table 1, it is found that the multilayer PLS has the largest prediction RMSE and MAPE. Differently, vMF-WSAE can achieve the best prediction performance under the new model update strategy. In detail, Figure 6 shows the prediction errors on the testing samples for the four methods. It can be seen that BPNN and SAE do not consider the characteristics of query sample, and their prediction effect is poor. The proposed vMF-WSAE can achieve good results because

Table 1: Performance comparison among the proposed vMF-WSAE and other methods in terms of RMSE and MAPE

Indicators	PLS	BPNN	SAE	vMF-WSAE
RMSE	7.6241	5.0840	4.5675	1.7467
MAPE	1.5043	1.0471	0.7016	0.2078

Bold represent the relative optimal value under this indicator in the comparison algorithm.

the model can be adaptively updated in nonlinear data modeling to effectively adapt to changes in data characteristics. As can be seen from this example, the proposed vMF-WSAE model is more effective in nonlinear modeling of multimode data.

4.2 BOF steelmaking process data experiment

4.2.1 Data introduction and parameter setting

Accurate forecasting of the endpoint carbon content and temperature of the molten steel is the key to endpoint

control in BOF steelmaking. The experimental data in this article come from the real BOF steelmaking production data of a steel plant, and the endpoint carbon content and temperature are the target variables, namely, the output variables. The quality of the raw materials changes over time, and the production conditions are constantly changing, and these process samples are collected under different production conditions. After previous research, the feature selection method is employed to select the original features shown in Table 3 as the input variables of the model [10,45]. Table 2 describes the details of the two datasets in terms of the number of input variables, the number of samples, and the output range. The specific meaning of each input variable is shown in Table 3, where the location of oxygen lance 1 refers to the first location of the oxygen lance and the oxygen pressure 1 represents the oxygen pressure value measured at the first time. To remove effects between dimensions, all data are normalized by zero-mean normalization. During the experiment, a total of 2,500 samples under normal working conditions are collected, of which 2,000 process samples are used as the training set and 500 samples are used as the test set to evaluate the performance of the model online.

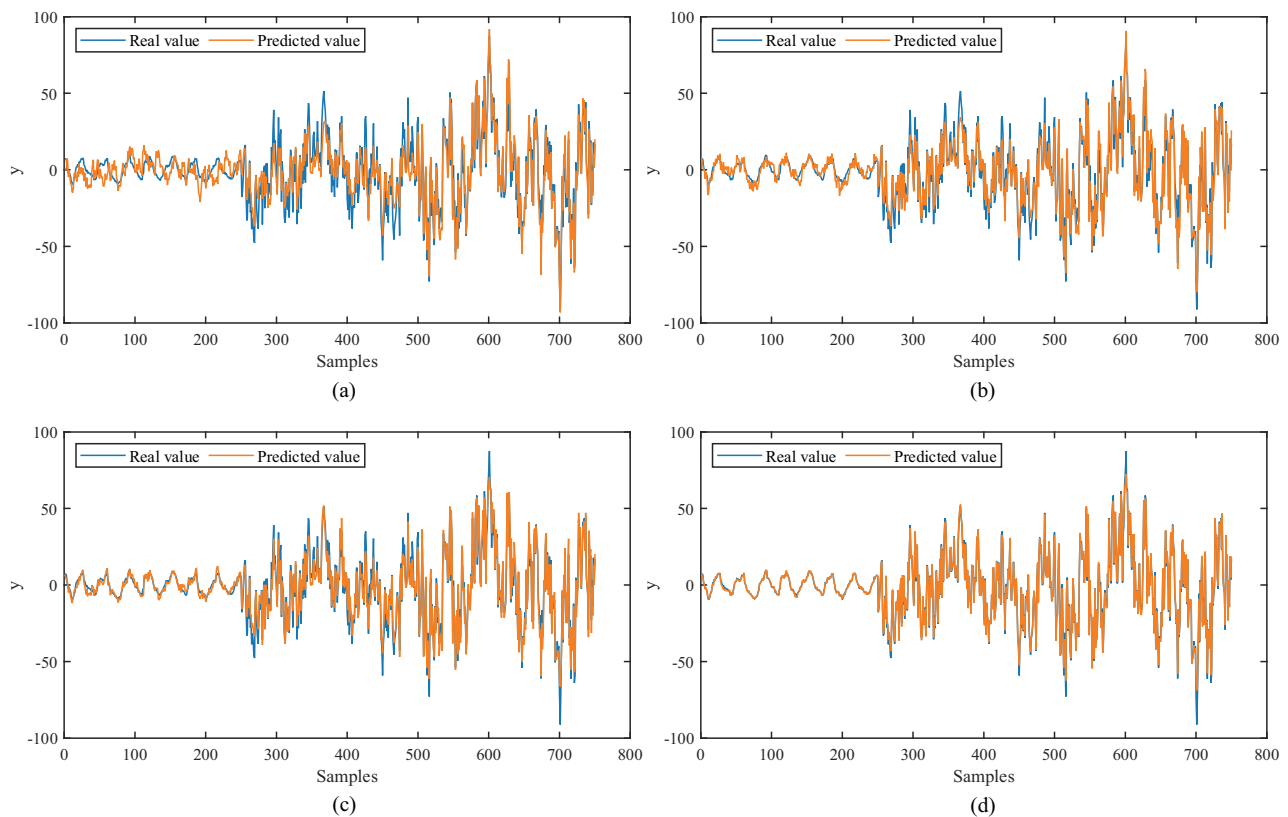
**Figure 6:** Prediction results of (a) PLS, (b) BPNN, (c) SAE, and (d) vMF-WSAE for the numerical simulation data.

Table 2: Details of the datasets of the BOF steelmaking production process

Datasets	No. of input variables	No. of samples	Range of outputs
Carbon content	15	2,500	[0.04, 0.18] (%)
Temperature	11	2,500	[1,600, 1,700] (°C)

In this article, there are many parameters that need to be set reasonably. First, for the network hyperparameters of SAE, different network structure candidates are designed. Second, these candidates are evaluated by a trial-and-error technique based on their predictive performance on the test set. Table 4 shows some network structure candidates and their performance. After comparing the predicted RMSE, [15-13-10-5-1] was selected as the network structure of carbon content and [11-9-7-5-1] as the network structure of temperature. Table 5 shows the settings of specific or range parameter values.

Moreover, for the finite mixture model, the number of mixture components is a significant parameter. In this article, iBIC of definition 1 is used to determine the number of vMF mixture model components, M varies from 2 to 10 with the step size of 1. The changes of iBIC values for vMF mixture models with respect to various choices of M are drawn in Figure 7. As can be seen from the figure, the two curves show an upward trend after reaching the minimum value. From definition 1, the smaller the value of iBIC is, the better the parameter M . Thus, the number of components M in vMF mixture

Table 3: The description of input variables

Target variable	Input variables	Target variable	Input variables
Carbon content	Amount of loaded steel scrap Total of transfer Temperature of iron End of iron mixing to oxygen opening time The time from tapping to starting to mix iron Manganese content in molten iron Phosphorus content in molten iron Arsenic content in molten iron Time to mix iron Average location of oxygen lance Location of oxygen lance 1 Location of oxygen lance 2 Oxygen pressure 1 Oxygen pressure 2 Oxygen pressure 3	Temperature	Total of transfer Duration of first blowing oxygen Arsenic content in molten iron Phosphorus content in molten iron Manganese content in molten iron Location of oxygen lance 3 Location of oxygen lance 4 Location of oxygen lance 5 Oxygen pressure 4 Oxygen pressure 5 Oxygen pressure 6

Table 4: Prediction performance of some candidate schemes of network parameters under the method of this paper

Network structure of carbon content	RMSE	Network structure of temperature	RMSE
[15-12-9-4-1]	1.7552×10^{-2}	[11-8-6-4-1]	8.3925
[15-9-4-1]	2.3244×10^{-2}	[11-6-4-1]	9.4695
[15-4-1]	2.6251×10^{-2}	[11-4-1]	10.3243
[15-13-10-5-1]	1.4185×10^{-2}	[11-9-7-5-1]	6.8536
[15-10-5-1]	1.8742×10^{-2}	[11-7-5-1]	8.4183
[15-5-1]	1.9301×10^{-2}	[11-5-1]	9.9253
[15-14-11-6-1]	1.7289×10^{-2}	[11-10-8-6-1]	9.2093
[15-11-6-1]	2.0978×10^{-2}	[11-8-6-1]	9.3557
[15-6-1]	2.2167×10^{-2}	[11-6-1]	10.3423

Bold represent the relative optimal value under this indicator in the comparison algorithm.

Table 5: The parameter settings

Parameters	Value(s)
Network structure of offline SAE	Carbon content: [15-13-10-5-1], Temperature: [11-9-7-5-1]
Maximum iterations for pretraining	400
Maximum iterations for offline fine-tuning	300
Maximum iterations for online adaptive updating	100
The number of mixture model components M	Min = 2, max = 10
The tolerable error of carbon content (Te)	$\{\pm 0.01, \pm 0.02, \pm 0.03\}$ (%)
The tolerable error of temperature (Te)	$\{\pm 5, \pm 10, \pm 15\}$ ($^{\circ}\text{C}$)

model for carbon content and temperature is set to 5 and 7, respectively.

4.2.2 Evaluation of the proposed method

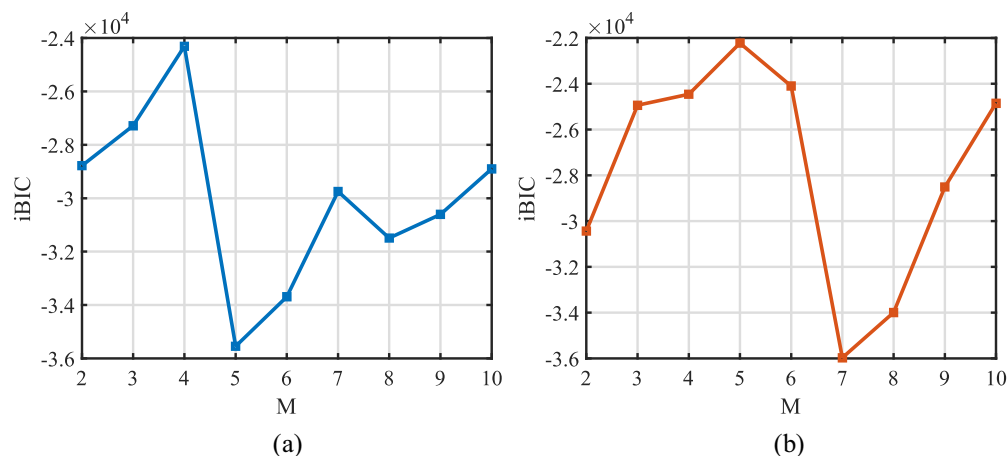
In this section, the performance of the proposed vMF-WSAE is evaluated. First, the influence of the parameter M on the prediction results is evaluated. Then, the prediction performance of the linear model PLS and the traditional nonlinear model BPNN is compared. The ablation experiments of the proposed method are compared, including:

- (1) The original SAE-based soft sensor model.
- (2) The offline stage remains unchanged, and there is no method to improve the loss function in the online update stage, which is expressed as vMF-SAE.
- (3) In the update phase, the weighting parameter of WSAE is determined by the ED between the query sample and the updating dataset, which is expressed as vMF-WSAE (ED).
- (4) The method proposed in this article is denoted as vMF-WSAE (WED).

Various independent experiments are carried out on the production process data of BOF steelmaking, and the model parameters of the compared methods are adjusted to the optimum according to the trial-and-error technique.

First, to evaluate the influence of parameter M on the performance of the proposed method, comparative experiments are carried out by changing the parameter values. The effect of M on the prediction error is shown in Figure 8. It can be noticed that the RMSE varies widely, and the parameter M plays an important role in the variation of prediction accuracy. With the remaining parameters unchanged, for the number of components M of the vMF mixture model, M values of the carbon content and temperature are the smallest when the RSME is taken to be 5 and 7, respectively. Furthermore, the number of components of the mixed model with the best prediction performance agrees with the optimal number of components determined by iBIC, proving the validity of definition 1.

Second, ablation experiments are performed to verify the effectiveness of the proposed method. Table 6 summarizes the predicted evaluation indicators of the compared methods under the carbon content and temperature

**Figure 7:** Demonstration of iBIC for vMF mixture model with respect to various choices of M on (a) carbon content and (b) temperature.

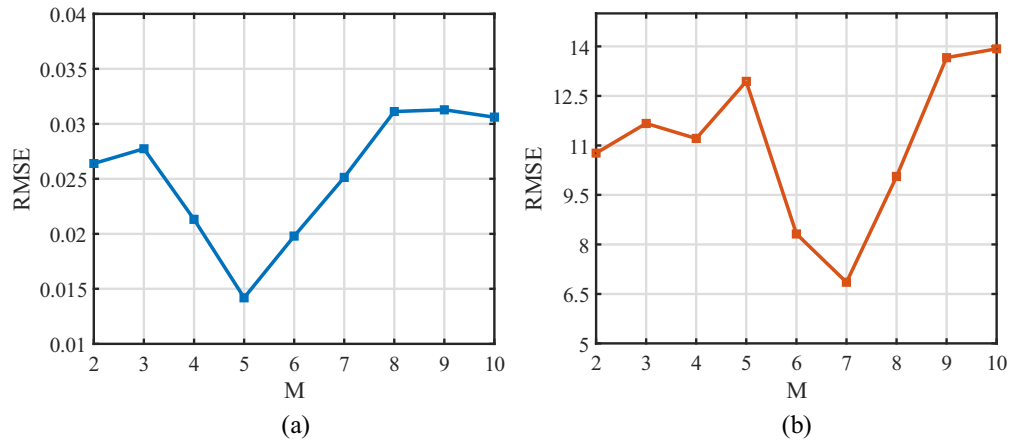


Figure 8: The RMSE values of the test set of vMF-WSAE vary with different choices of M at (a) carbon content and (b) temperature.

Table 6: Prediction performance comparison of the proposed method with other ablation experiments

Indicators	Te	PLS	BPNN	SAE	vMF-SAE	vMF-WSAE (ED)	vMF-WSAE (WED)
Carbon content regression accuracy	0.01%	22.20%	28.20%	35.20%	40.60%	52.80%	70.40%
	0.02%	41.40%	51.00%	60.40%	70.40%	84.00%	90.60%
	0.03%	59.80%	71.00%	75.00%	86.00%	94.00%	95.80%
RMSE(E-02)	—	3.4810	2.9890	2.6967	2.2067	1.6274	1.4185
MAPE	—	0.2973	0.2466	0.2246	0.1772	0.1254	0.1092
Temperature regression accuracy	5°C	22.60%	38.60%	41.00%	47.20%	57.20%	59.80%
	10°C	42.60%	60.80%	67.40%	77.20%	84.20%	92.00%
	15°C	60.40%	78.00%	83.60%	89.60%	96.60%	98.20%
RMSE	—	18.0490	12.4408	11.5903	10.0790	7.6974	6.8536
MAPE(E-03)	—	8.9517	5.8475	5.3840	4.5601	3.6153	3.2788

Bold represent the relative optimal value under this indicator in the comparison algorithm.

datasets. It can be seen that on the two datasets, the RA of the proposed vMF-WSAE (WED) under various Te is much higher than that of other methods, and the statistical results of RMSE and MAPE also show that the performance of the proposed method is also better than other methods.

In detail, the prediction performance of PLS is the worst because the complex nonlinear relationship between the features and the output makes the linear model to not be effectively modeled. Comparing the results of BPNN and SAE, SAE can better express the complex nonlinear relationship between data after pretraining and fine-tuning. While vMF-SAE adaptively updates offline SAE by considering the data pattern characteristics of query samples, its prediction performance is better than traditional SAE modeling. vMF-WSAE (ED) and vMF-WSAE (WED) improve the loss function in the SAE update stage, and use ED and WED to obtain the similarity weight for weighted update respectively. Among them, the traditional ED-based similarity criterion only calculates the input variables. Although the prediction performance has been improved, the evaluation ability of similar samples is still insufficient. For the vMF-WSAE

(WED) method proposed in this article, WED comprehensively considers the information of input and output variables to calculate the similarity, and the evaluation of similar samples is more reasonable. In the adaptive weighted updating, the model pays more attention to the samples in the updating dataset whose data characteristics are more similar to the query sample, thereby obtaining better prediction results.

To more intuitively show the prediction and the tracking effect of the model, Figure 9 further shows the detailed prediction results of these methods for carbon content and temperature, and it displays the actual and predicted outputs sorted by the actual order of the samples. Looking at Figure 9(a), the linear model cannot effectively fit the complex process data of BOF steelmaking. As can be seen from Figure 9(b) and (c), the predicted values of BPNN and SAE have a large deviation from the real values, and the static models cannot adapt well to changing working conditions. Looking at Figure 9(d), although the prediction effect of vMF-SAE is improved, it does not fully consider the data structure characteristics of

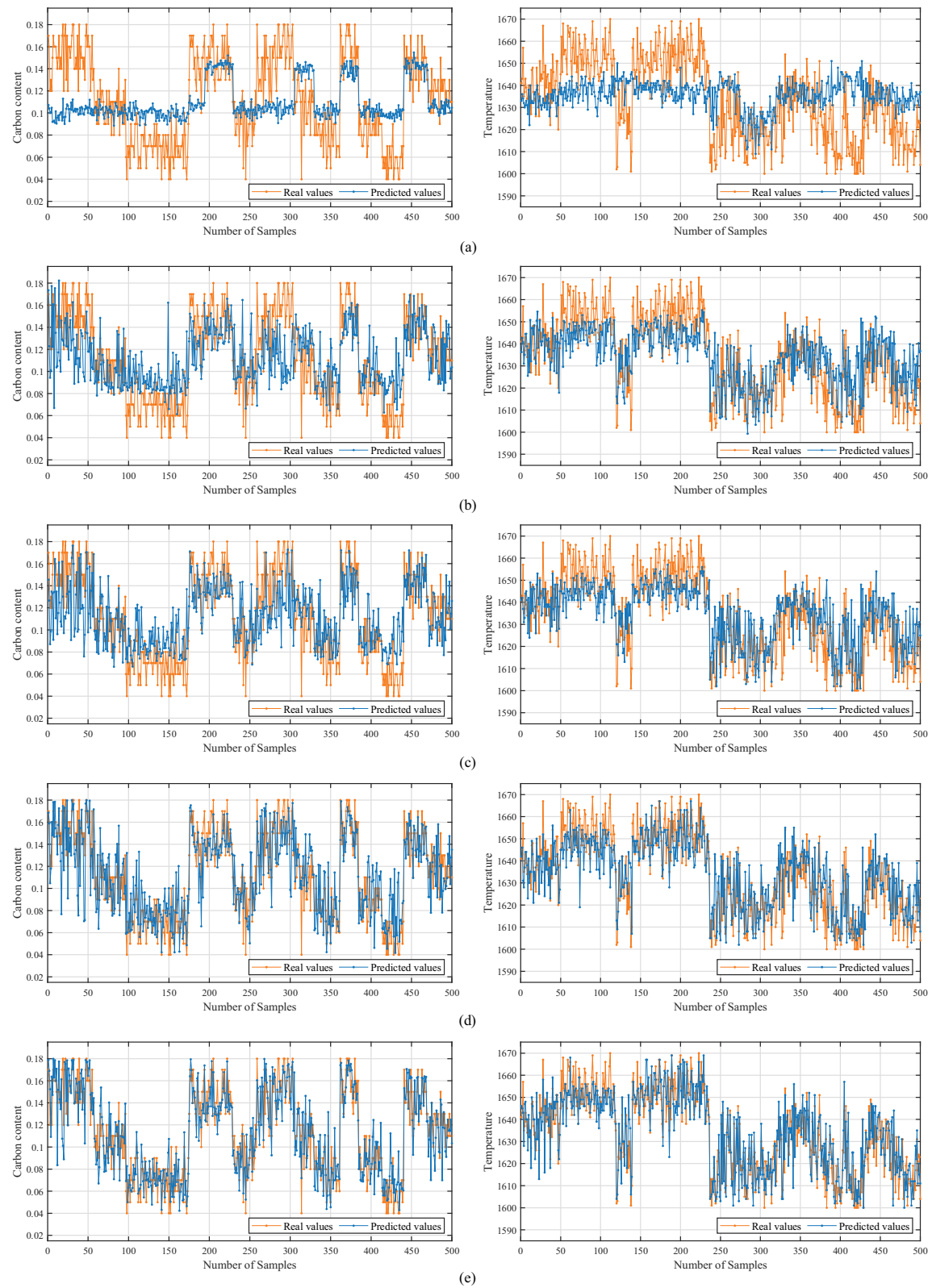


Figure 9: Detailed prediction results for carbon content (left) and temperature (right) test datasets: (a) PLS, (b) BPNN, (c) SAE, (d) vMF-SAE, (e) vMF-WSAE(ED), and (f) vMF-WSAE(WED).

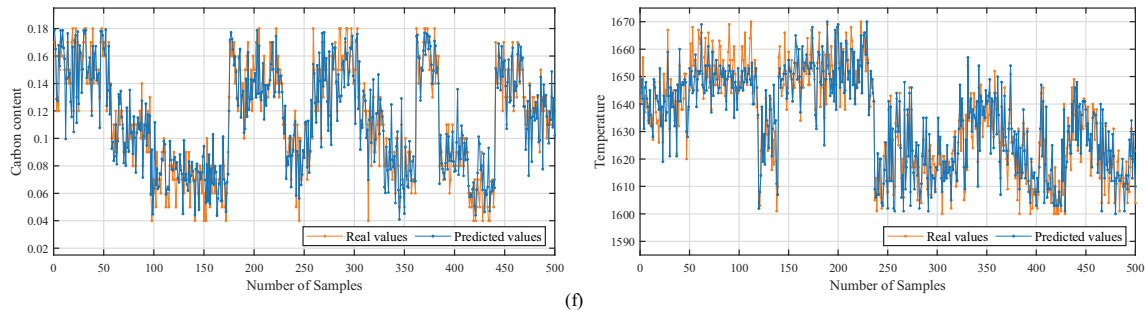


Figure 9: (Continued)

the query samples to effectively update. Observing Figures 9(e) and (f), the predicted values of the proposed vMF-WSAE (WED) fit best with the actual values of carbon content and temperature, and vMF-WSAE(ED) has a certain degree of prediction value that cannot accurately track the true value. These results fully demonstrate that the proposed method has great advantages over the traditional static model.

Table 7 summarizes the computation time of the proposed method and ablation experiments in seconds. Compared with traditional SAE, the offline part of the proposed method increases the time required to fit the vMF mixture model on the training samples. The online test part needs more time to update the model to adapt to the working conditions of the query samples. The time taken to predict the 500 test sets under the two datasets is 46.9063 s and 31.8944 s, respectively (Table 7). The substance detection as

a real-time detection technique, according to the description of the substance inspection of the BOF in reference [46] and the description of the experts in the steelmaking field, the average time consuming to detect the carbon content of the molten steel is about 1 min per time. The time required to predict the endpoint of steelmaking by the proposed method is much less than that of the substance detection, so that it can meet the real-time requirements of the actual BOF steelmaking process.

4.2.3 Comparison with other deep learning soft sensor methods

In this section, a comparison between the proposed vMF-WSAE and other deep learning soft sensor methods, such

Table 7: Performance comparison of the proposed method and ablation experiments in terms of computation time

Datasets	Phase	PLS (s)	BPNN (s)	SAE (s)	vMF-SAE (s)	vMF-WSAE (ED) (s)	vMF-WSAE (WED) (s)
Carbon Content	Offline	0.0229	5.0691	2.8579	20.5336	20.8746	20.4332
	Online	0.0146	0.3066	0.0171	48.0876	44.1774	46.9063
Temperature	Offline	0.0894	1.5921	2.1042	18.7954	19.1849	18.5827
	Online	0.0133	0.0801	0.0138	30.6217	29.7163	31.8944

Table 8: Comparison the proposed method with other methods in terms of prediction evaluation indicators

Indicators	Te	VW-SAE [23]	NVW-SAE [24]	SQAE [25]	vMF-WSAE
Carbon content regression accuracy	0.01%	31.00%	35.60%	36.20%	70.40%
	0.02%	61.80%	62.20%	67.00%	90.60%
	0.03%	77.20%	77.00%	81.60%	95.80%
RMSE (E-02)	—	2.6292	2.6424	2.5279	1.4185
MAPE	—	0.2200	0.2142	0.1929	0.1092
Temperature regression accuracy	5°C	43.20%	46.60%	48.60%	59.80%
	10°C	71.20%	73.80%	76.40%	92.00%
	15°C	84.80%	88.40%	89.20%	98.20%
RMSE	—	10.8009	10.2448	10.4834	6.8536
MAPE (E-03)	—	5.0830	4.7856	4.6657	3.2788

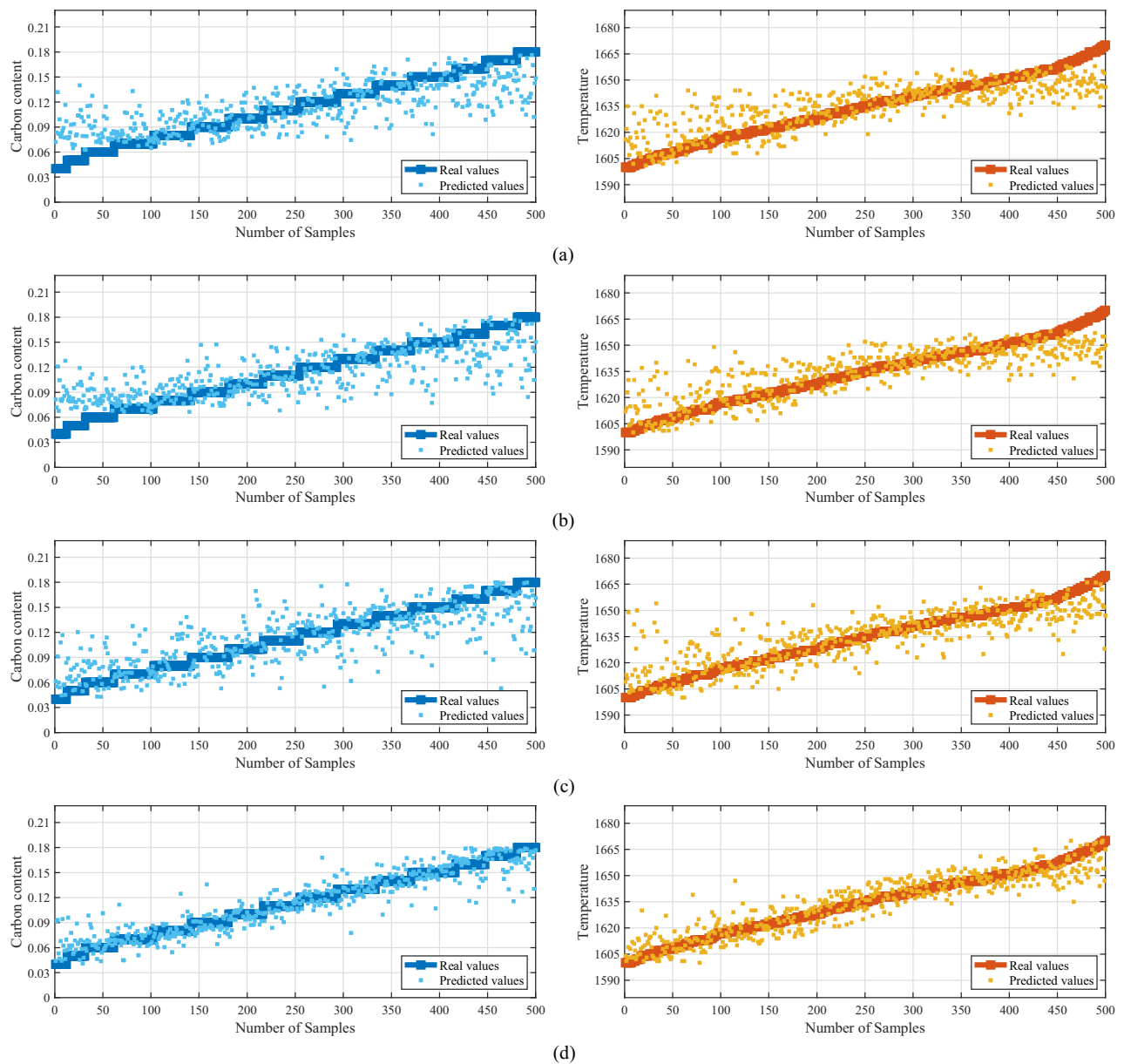
Bold represent the relative optimal value under this indicator in the comparison algorithm.

Table 9: Performance comparison of the proposed method and other experiments in terms of computation time

Datasets	Phase	VW-SAE (s)	NVW-SAE (s)	SQAE (s)	vMF-WSAE (s)
Carbon Content	Offline	1.5459	13.759	2.2398	20.4332
	Online	0.0161	0.0389	0.0133	46.9063
Temperature	Offline	1.7043	18.8453	1.3772	18.5827
	Online	0.008	0.0421	0.0103	31.8944

as variable weighted stacked autoencoder (VW-SAE) [23], nonlinear variable weighted stacked autoencoder (NVW-SAE) [24], and stacked quality-driven autoencoder (SQAE) [25]. The parameters of the compared methods are the best performance under the data of the BOF steelmaking process.

Table 8 presents the prediction performance metrics of the proposed vMF-WSAE and other soft sensor methods. It can be noted that the RA of the proposed vMF-WSAE method under various T_e is much higher than all other methods on both datasets, and observing the RMSE and

**Figure 10:** Scatter plots of actual and predicted values for carbon content (left) and temperature (right): (a) VW-SAE, (b) NVW-SAE, (c) SQAE, and (d) vMF-WSAE.

MAPE statistics, the vMF-WSAE method still outperforms the other methods. For VW-SAE, NVW-SAE, and SQAE, the performance indicators for predicting temperature are better than those for predicting the carbon content, and the model performance of SQAE is better. Compared with traditional machine learning, these deep learning models can solve the highly nonlinear problem of BOF steelmaking process data to a certain extent. Although these methods improve the performance of SAE by improving unsupervised pre-training to supervised, they are essentially global offline static models, which cannot automatically update the parameters of the prediction model according to the characteristics of the query samples.

Table 9 shows the computation time of the proposed method and other methods. Compared with the static deep learning algorithms in these references, the online part of the proposed method requires more time to update the model to adapt to the working conditions of the query sample. In the case of meeting the production requirements, the proposed method achieves better predictive performance.

To more clearly show the fit between the predicted value and the real value of the compared methods, Figure 10 shows the scatter plots of the actual output and predicted output of the proposed method and other methods under the two datasets. In this figure, the sample order is shown sorted by the label size, and each data point represents a test sample. That is to say, the better the prediction performance of the model, the closer the scatter point of the predicted value will be to the curve of the real value. It can be seen from Figure 10(a) and (b) that the predicted values of VW-SAE and NVW-SAE have a large deviation from the real values, and the static deep learning soft sensor model cannot fit the data of the multimode BOF steelmaking process well. Looking at Figure 10(c), although the prediction effect of SQAE is improved compared with the first two methods, there are too many samples with excessive prediction deviation. While in Figure 10(d), vMF-WSAE can always fit the actual output curve well due to the excellent adaptive updating mechanism of vMF-WSAE. Experiments show that the method proposed in this article has certain advantages over other deep learning soft sensor methods in the endpoint carbon content and temperature prediction of BOF steelmaking.

5 Conclusions

For the characteristics of highly nonlinear and multimode distribution of BOF steelmaking process data, the

static model cannot adapt to the changes in sample characteristics leading to the degradation of prediction performance. In this article, an adaptive dynamic deep learning soft sensor model based on the vMF mixture model and weighted SAE (vMF-WSAE) is proposed for predicting the endpoint carbon content and temperature in BOF steelmaking.

For most existing deep networks, they are static feature learning models. This may not be an effective soft sensor modeling method for BOF steelmaking processes with widely varying data distributions. vMF-WSAE adopts an effective adaptive model update strategy, which enables deep learning to quickly adapt to the process running state and achieve accurate prediction. Through numerical examples and modeling simulations of BOF steelmaking process data, the proposed method is compared with ablation experiments. Moreover, other deep learning soft measurement methods are compared. Various evaluation criteria are used to evaluate the performance of each method on the carbon content and temperature datasets. Compared with traditional machine learning, the proposed method can effectively solve the highly nonlinear problem of BOF steelmaking process data. Compared with the offline static model, the mechanism of adaptively updating model parameters can update the model according to the distribution and data structure characteristics of the query sample to improve the prediction performance. This article provides a better research direction for soft sensor modeling of BOF endpoint carbon content and temperature using deep learning, which has a certain reference value in practical applications.

Acknowledgement: The authors are grateful to the National Natural Science Foundation of China (Nos. 62263016 and 61863018) and the Applied Basic Research Foundation of Yunnan Province (No. 202001AT070038) for funding to support this research.

Funding information: This work was supported by the National Natural Science Foundation of China (Nos. 62263016 and No. 61863018); the Applied Basic Research Foundation of Yunnan Province (No. 202001AT070038).

Author contributions: Lu Yang: conceptualization, methodology, software, validation, writing – original draft, writing – review and editing, and visualization; Hui Liu: conceptualization, methodology, supervision, project administration, and funding acquisition; Fugang Chen: data curation, investigation, and formal analysis.

Conflict of interest: The authors state no conflict of interest.

References

- [1] Han, M. and C. Liu. Endpoint prediction model for basic oxygen furnace steel-making based on membrane algorithm evolving extreme learning machine. *Applied Soft Computing*, Vol. 19, 2014, pp. 430–437.
- [2] Xie, S. and T. Chai. Prediction of BOF endpoint temperature and carbon content. *IFAC Proceedings Volumes*, Vol. 32, No. 2, 1999, pp. 7039–7043.
- [3] Han, M. and Y. Zhao. Dynamic control model of BOF steel-making process based on ANFIS and robust relevance vector machine. *Expert Systems with Applications*, Vol. 38, No. 12, 2011, id. 14786–14798.
- [4] Wang, Z., Q. Liu, H. Liu, H. T. Liu and S. Z. Wei. A review of end-point carbon prediction for BOF steelmaking process. *High Temperature Materials and Processes*, Vol. 39, No. 1, 2020, pp. 653–662.
- [5] Zhou, M., Q. Zhao, and Y. Chen. Endpoint prediction of BOF by flame spectrum and furnace mouth image based on fuzzy support vector machine. *Optik*, Vol. 178, 2019, pp. 575–581.
- [6] Gruner, H., H. E. Wiemer, and W. Fix. New metallurgical insight into BOF-steelmaking and improved process control using subplasma technique and bottom gas stirring. *Steelmaking Proceedings*, Vol. 67, 1984, pp. 113–120.
- [7] Liu, H., B. Wang, and X. Xiong. Basic oxygen furnace steel-making end-point prediction based on computer vision and general regression neural network. *Optik*, Vol. 125, No. 18, 2014, pp. 5241–5248.
- [8] Gao, S., Y. Dai, Y. Li, Y. Jiang, and Y. Liu. Augmented flame image soft sensor for combustion oxygen content prediction. *Measurement Science and Technology*, Vol. 34, No. 1, 2022, id. 015401.
- [9] Liu, X. C., H. Liu, F. G. Chen, and C. Li. A real-time prediction method of carbon content in converter steelmaking based on DDMCN flame image feature extraction. *Control and Decision*, vol. 2022, 2022, pp. 1–9. Doi: 10.13195/j.kzyjc.2021.2166.
- [10] Qi, L. and H. Liu. Feature selection of BOF steelmaking process data based on denary salp swarm algorithm. *Arabian Journal for Science and Engineering*, Vol. 45, No. 12, 2020, id. 10401–10416.
- [11] Xie, S. M., T. Y. Chai, and J. Tao. A kind of new method for LD dynamic endpoint prediction. *Acta Automatica Sinica*, Vol. 27, No. 1, 2001, pp. 136–139.
- [12] Xie, S. M., J. Tao, and T. Y. Chai. BOF steelmaking endpoint control based on neural network. *Control Theory & Applications*, Vol. 20, No. 6, 2003, pp. 903–907.
- [13] Wang, X. Z. and M. Han. Prediction model of converter steel-making endpoint based on variable selection. *Control and Decision*, Vol. 25, No. 10, 2010, pp. 1589–1592.
- [14] Cox, I. J., R. W. Lewis, and R. S. Ransing. Application of neural computing in basic oxygen steelmaking. *Journal of Materials Processing Technology*, Vol. 120, No. 1–3, 2002, pp. 310–315.
- [15] Zeng, P. F. and H. Liu. A soft-sensing method for carbon temperature at the end of converter steelmaking based on quadratic similarity measurement. *Computer Integrated Manufacturing Systems*, Vol. 27, No. 5, 2021, pp. 1429–1439.
- [16] Qi, L., H. Liu, Q. Xiong, and Z. X. Chen. Just-in-time-learning based prediction model of BOF endpoint carbon content and temperature via vMF mixture model and weighted extreme learning machine. *Computers & Chemical Engineering*, Vol. 154, 2021, id. 107488.
- [17] Liu, Y. and J. Chen. Integrated soft sensor using just-in-time support vector regression and probabilistic analysis for quality prediction of multi-grade processes. *Journal of Process Control*, Vol. 23, No. 6, 2013, pp. 793–804.
- [18] Fan, M., Z. Ge, and Z. Song. Adaptive gaussian mixture model-based relevant sample selection for JITL soft sensor development. *Industrial & Engineering Chemistry Research*, Vol. 53, No. 51, 2014, id. 19979–19986.
- [19] Liu, Y., T. Chen, and J. Chen. Auto-switch Gaussian process regression-based probabilistic soft sensors for industrial multigrade processes with transitions. *Industrial & Engineering Chemistry Research*, Vol. 54, No. 18, 2015, pp. 5037–5047.
- [20] Wang, Y., D. Wu, and X. Yuan. A two-layer ensemble learning framework for data-driven soft sensor of the diesel attributes in an industrial hydrocracking process. *Journal of Chemometrics*, Vol. 33, No. 12, 2019, id. e3185.
- [21] Gao, C., L. Jian, and S. Luo. Modeling of the thermal state change of blast furnace hearth with support vector machines. *IEEE Transactions on Industrial Electronics*, Vol. 59, No. 2, 2011, pp. 1134–1145.
- [22] Liu, Q., M. Jia, Z. Gao, Z. Gao, L. Xu, and Y. Liu. Correntropy long short term memory soft sensor for quality prediction in industrial polyethylene process. *Chemometrics and Intelligent Laboratory Systems*, Vol. 231, 2022, id. 104678.
- [23] Yuan, X., B. Huang, and Y. Wang. Deep learning-based feature representation and its application for soft sensor modeling with variable-wise weighted SAE. *IEEE Transactions on Industrial Informatics*, Vol. 14, No. 7, 2018, pp. 3235–3243.
- [24] Yuan, X., C. Ou, Y. L. Wang and C. H. Yang. Nonlinear VW-SAE based deep learning for quality-related feature learning and soft sensor modeling. *IECON 2018-44th Annual Conference of the IEEE Industrial Electronics Society*, IEEE, Washington, DC, 2018, pp. 5400–5405.
- [25] Yuan, X., J. Zhou, and B. Huang. Hierarchical quality-relevant feature representation for soft sensor modeling: A novel deep learning strategy. *IEEE Transactions on Industrial Informatics*, Vol. 16, No. 6, 2019, pp. 3721–3730.
- [26] Liu, Y., C. Yang, Z. Gao, Z. Gao, and Y. Yao. Ensemble deep kernel learning with application to quality prediction in industrial polymerization processes. *Chemometrics and Intelligent Laboratory Systems*, Vol. 174, 2018, pp. 15–21.
- [27] Liu, Y., C. Yang, M. Zhang, Y. Dai, and Y. Yao. Development of adversarial transfer learning soft sensor for multigrade processes. *Industrial & Engineering Chemistry Research*, Vol. 59, No. 37, 2020, id. 16330–16345.
- [28] Liu, Y., C. Yang, K. Liu, B. Chen, and Y. Yao. Domain adaptation transfer learning soft sensor for product quality prediction. *Chemometrics and Intelligent Laboratory Systems*, Vol. 192, 2019, id. 103813.
- [29] Deng, H., K. Yang, Y. Liu, S. Zhang, and Y. Yao. Actively exploring informative data for smart modeling of industrial multiphase flow processes. *IEEE Transactions on Industrial Informatics*, Vol. 17, No. 12, 2020, pp. 8357–8366.
- [30] Liu, Y., Q. Wu, and J. Chen. Active selection of informative data for sequential quality enhancement of soft sensor models with

- latent variables. *Industrial & Engineering Chemistry Research*, Vol. 56, No. 16, 2017, pp. 4804–4817.
- [31] Liu, Y. and Z. Gao. Enhanced just-in-time modelling for online quality prediction in BF ironmaking. *Ironmaking & Steelmaking*, Vol. 42, No. 5, 2015, pp. 321–330.
- [32] He, X., J. Ji, K. Liu, Z. Gao, and Y. Liu. Soft sensing of silicon content via bagging local semi-supervised models. *Sensors*, Vol. 19, No. 17, 2019, id. 3814.
- [33] Liou, C. Y., W. C. Cheng, J. W. Liou, and D. R. Liou. Autoencoder for words. *Neurocomputing*, Vol. 139, 2014, pp. 84–96.
- [34] Wang, X., M. Han, and J. Wang. Applying input variables selection technique on input weighted support vector machine modeling for BOF endpoint prediction. *Engineering Applications of Artificial Intelligence*, Vol. 23, No. 6, 2010, pp. 1012–1018.
- [35] Bingyao, C., Z. Hui, and Y. Youjun. Research on the BOF steelmaking endpoint temperature prediction. *2011 International Conference on Mechatronic Science, Electric Engineering and Computer (MEC)*, IEEE, Jilin, China, 2011, pp. 2278–2281.
- [36] Sun, Y. N., Z. L. Zhuang, and H. W. Xu. Data-driven modeling and analysis based on complex network for multimode recognition of industrial processes. *Journal of Manufacturing Systems*, Vol. 62, 2022, pp. 915–924.
- [37] Cui, L. L., B. B. Shen, and Z. Q. Ge. A Mixture Variational Autoencoder Regression Model for Soft Sensor Application. *Acta Automatica Sinica*, Vol. 48, No. 2, 2022, pp. 398–407. Doi: 10.16383/j.aas.c210035.
- [38] Mardia, K. V., P. E. Jupp and K. V. Mardia. *Directional statistics*, Wiley, New York, 2000.
- [39] Watson, G. S. and E. J. Williams. On the construction of significance tests on the circle and the sphere. *Biometrika*, Vol. 43, No. 3–4, 1956, pp. 344–352.
- [40] Taghia, J., Z. Ma, and A. Leijon. Bayesian estimation of the von-Mises Fisher mixture model with variational inference. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 36, No. 9, 2014, pp. 1701–1715.
- [41] Liu, D. C. and J. Nocedal. On the limited memory BFGS method for large scale optimization. *Mathematical Programming*, Vol. 45, No. 1, 1989, pp. 503–528.
- [42] Mehrjou, A., R. Hosseini, and B. N. Araabi. Improved Bayesian information criterion for mixture model selection. *Pattern Recognition Letters*, Vol. 69, 2016, pp. 22–27.
- [43] Wei, G., J. Zhao, and Y. Feng. A novel hybrid feature selection method based on dynamic feature importance. *Applied Soft Computing*, Vol. 93, 2020, id. 106337.
- [44] Zeng, J., L. Xie, C. Gao, J. Sha. Soft Sensor Development Using non-Gaussian Just-In-Time Modeling. In *50th IEEE Conference on Decision and Control and European Control Conference (CDC-ECC)*, IEEE Press, Piscataway, NJ, 2011, pp. 5868–5873.
- [45] Liu, H., P. F. Zeng, S. Q. Wu, and F. G. Chen. Feature selection of converter steelmaking process based on the improved genetic algorithm. *Chinese Journal of Scientific Instrument*, Vol. 40, No. 12, 2019, pp. 185–195.
- [46] IIDA, Y., K. Emoto, M. Ogawa, Y. Masuda, M. Onishi, and H. Yamada. Fully automatic blowing technique for basic oxygen steelmaking furnace. *Transactions of the Iron and Steel Institute of Japan*, Vol. 24, No. 7, 1984, pp. 540–546.