2023 IEEE Symposium Series on Computational Intelligence (SSCI)
Mexico City, Mexico. December 5-8, 2023

Online Soft Sensing of Dioxin Emission Based on Fast Tree BLS and Robust PCA

Heng Xia
Faculty of Information Technology
Beijing University of Technology
Beijing, China
xiaheng@emails.bjut.edu.cn

Jian Tang
Faculty of Information Technology
Beijing University of Technology
Beijing, China
freeflytang@bjut.edu.cn

Runyu Zhang
Faculty of Information Technology
Beijing University of Technology
Beijing, China
zhangrunyu@emails.bjut.edu.cn

***Abstract*—Municipal solid waste incineration (MSWI) is a crucial technology for waste treatment in densely populated cities. It plays a vital role in contributing to the hot concept of waste-to-energy. However, the effectively measuring of dioxins (DXN) emission from MSWI plants presents a complex challenge duo to its high economical cost and large lag time. To address the challenge, we propose a soft sensing method of DXN emission concentration based on the fast tree broad learning system (FTBLS) and the robust principal component analysis (RPCA). FTBLS can swiftly construct the DXN emission model with increment learning for obtaining accurate measuring results. RPCA is capable decomposing high-dimensional small sample data into low-rank and noise matrices, achieving robust operation condition drift detection in the presence of noise process data. The similarity estimation is used to aid the soft measuring value's obtainment for concept drift sample. The experiment and application results demonstrate the effectiveness of our proposed online soft sensing approach.**

Keywords—Municipal solid waste incineration (MSWI), Dioxins, online soft sensing, fast tree broad learning system, robust concept detection, similarity estimation

I. INTRODUCTION

Municipal solid waste incineration (MSWI) technology not only addresses the crisis of the "MSW Siege" [1], but also generates energy, contributing to the electricity supply[2]. Up to date, MSWI has become one of the primary options for municipal solid waste treatment in most cities [3]. Currently, over 800 MSWI plants have been built and put into operation in China, with grate furnaces accounting for more than 90% of them [4]. The type grate-based MSWI process is shown in Fig. 1. The operation of the MSWI process involves six stages: reception and storage, feeding, combustion, heat exchange, gas cleaning, and emission [5]. Each stage is continuous and cannot be stopped, and any fault in the sub-stage can impact the stability of the entire production process and result in excessive pollution emissions [6]. Moreover, the combustion process involves various physical and chemical reactions, which makes the process mechanism more complex [7].

The MSWI industry play a significant role as a supporting industry in China. However, there is a general lack of public acceptance of the MSWI technology regarding its environmental and health impact [8]. One particular concern is the dioxins (DXN) emission, which have irreversible negative effects [9]. However, there is currently no real-time online detection technology available to detect DXN emission concentration. Real-time estimation of DXN concentrations emitted from MSWI plants remains a challenging issue [10]. These concerns have led to public opposition to the establishment of MSWI plants, as evident in the "Not-In-My-Back-Yard" phenomenon [11]. To make public trust and support MSWI technology, it is crucial to develop DXN models that can accurately online measuring DXN emission concentration [12].

In recent decades, modeling research on DXN (referring to DXN emission concentration located in stack) can be summarized into two phases. The first phase involved studying the relationship between DXN emission concentration and some flue gas concentration out of the stack and some key process variables. The second phase focused on constructing the soft sensing model between DXN emission concentrations and most of process variables. In the first phase, researcher conducted model using data from multiple MSWI plants located in different regions. They examined the relationship between CO and DXN emission concentration [13], the relationship between CO concentration, oxygen concentration, moisture, and furnace temperature and DXN emission [14], and also explored the relationship between combustion temperature, steam flow, CO concentration, HCl concentration and DXN emission [15]. Subsequently, machine learning and deep learning emerged as the primary tools for constructing soft sensing model. Researchers collected data from a single MSWI plant and attempt to learn a mapping relation between the process variable and DXN concentration [16]. Various methods employed, including genetic programming and back-propagation neural network (BPNN) [17], principal component analysis (PCA) and BPNN [18], deep belief network [19], and random weight neural network [20] et al. Statistical learning based methods involved support vector regression (SVR) [21] and adaptive selected ensemble SVR [22]. Multivariate statistical methods included selective ensemble kernel partial least square [23]. Decision tree (DT) techniques encompassed deep forest regression (DFR) [24], DFR based on cross-layer full connection [25], improved DFR [26], broad hybrid forest regression [27], simplified DFR [2], and fuzzy broad learning [28].

Currently, deep and broad learning methods on the basis of the DT have become the research focus. However, most of the above-mentioned works can only make accurate measuring results under stable operating conditions. However, MSWI is a dynamic industrial process. As a result, DXN emission concentrations fluctuate over time due to these drifting conditions. To address these challenges, this article presents a strategy that combines fast tree broad learning system (FTBLS) and robust PCA (RPCA) for online soft sensing of DXN emission concentration.

II. DESCRIPITON OF MSWI IN VIEW OF DXN

The typical MSWI process is shown in Fig. 1.

Funded by National Key Research and Development Program of China (2021ZD0112302).

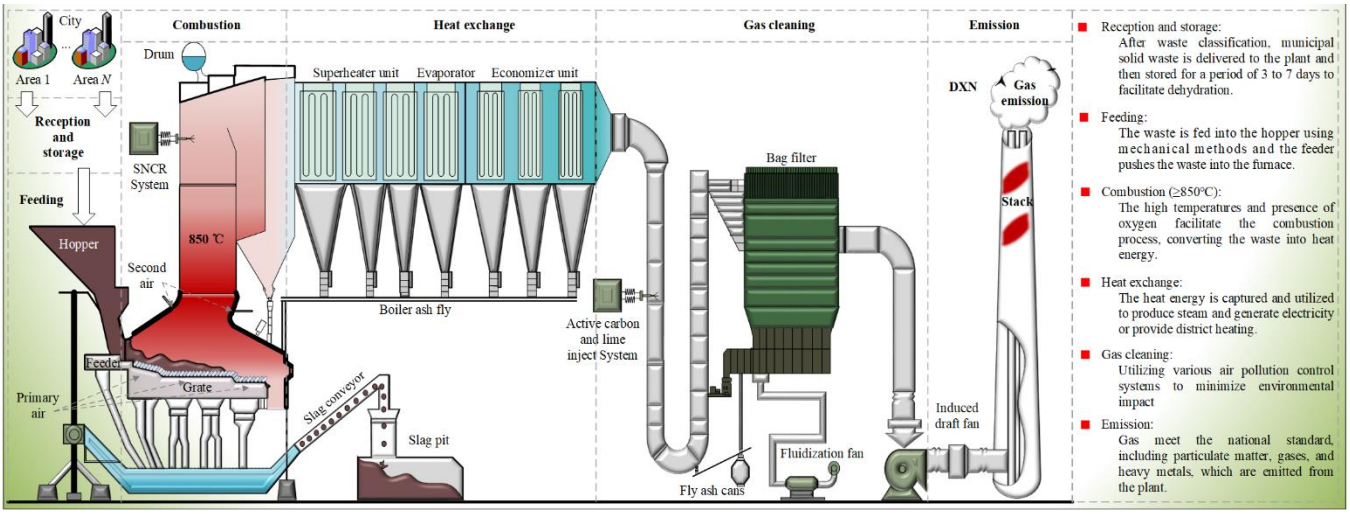

Fig. 1. Process flow of the grate-based MSWI plant.

Currently, the offline measurement of DXN emission concentration involved two stages: manual on-site sampling, laboratory cleaning and extract. It is followed by high-resolution gas chromatography-high-resolution mass spectrometry (HRGC/HRMS) analysis. In the first stage, it takes at least two hours to collect real-time flue gas samples, while the second stage typically takes more than one day for analysis. Thus, it is necessary to construct a soft sensing model by using the measuring value obtained with HRGC/HRMS as the true value of building model. Moreover, the modeling method should suitable to the data with characteristics of high dimensional and small sample.

As shown in Fig.1, the MSWI process consists of multiple stages and operates in a high temperature and high pressure environment. The measured values of the process variables may deviate from or contain errors compared to the true values, resulting in sparse noise data. The existence of such noise data hinders the correctness and validity of the online soft sensing results. Consequently, in order to detect the concept drift effectively, it is crucial to eliminate the noise data by using a noise removal technique. Moreover, it is not reasonable to use the offline constructed model to measure the DXN emission concentration for the samples represented concept drift. How to make such soft measuring in the actual MSWI plant is still a challenging issue.

III. METHOD IMPLEMEATOIN

The proposed method is divided into four parts, which are described as follows.

A. Offline Model Constriuction Based on FTBLS

A new proposed FTBLS approach is used for offline model construction. The key modules involve the mapping, enhancement, and increment layer. In the former two layer, the used fundamental unit is the random DT (RDT). To achieve the fast increment learning performance, the neuron group is considered as the basic unit of the third layer.

Given the history training data $D^{\mathrm{Htd}}=\{\boldsymbol{x}_i^{\mathrm{Htd}},y_i^{\mathrm{Htd}}\}_i^N$, the mapping layer consists of multiple RDT units. Each RDT unit is composed of non-leaf nodes (tree grow) and leaf nodes (tree prediction). During the tree growth process, $(T-1)/2$ nonleaf nodes are determined by random split, which is expressed as,

$$\begin{aligned}&\varphi^t(x_{i,m}^{\mathrm{Htd}})=f^{\mathrm{RS}}(N,M)=\{rand(1,N),rand(1,M)\}\\&s.t.\ \ \min(\boldsymbol{x}_i^{\mathrm{Htd}})<x_{i,m}^{\mathrm{Htd}}<\max(\boldsymbol{x}_i^{\mathrm{Htd}})\end{aligned}\tag{1}$$

where $\varphi^t(x_{i,m}^{\mathrm{Htd}})$ means the tth $(\mathrm{t}\le(T-1)/2)$ nonleaf node.

Based on the tth nonleaf node, the tree structure of the RDT growth by splitting strategy, which can be described as,

$$\begin{cases}D^{\mathrm{Htd}}\xrightarrow{x_i^{\mathrm{Htd}}\ge x_{i,m}^{\mathrm{Htd}}}D_{t+1,1}^{\mathrm{Htd}}=\{\boldsymbol{x}_i^{\mathrm{Htd}},y_i^{\mathrm{Htd}}\}_i^{N_{t+1,1}^{\mathrm{NLN}}}\in\mathbb{R}^{N_{t+1,1}^{\mathrm{NLN}}\times(M+1)}\\D^{\mathrm{Htd}}\xrightarrow{x_i^{\mathrm{Htd}}<x_{i,m}^{\mathrm{Htd}}}D_{t+1,2}^{\mathrm{Htd}}=\{\boldsymbol{x}_i^{\mathrm{Htd}},y_i^{\mathrm{Htd}}\}_i^{N_{t+1,2}^{\mathrm{NLN}}}\in\mathbb{R}^{N_{t+1,2}^{\mathrm{NLN}}\times(M+1)}\end{cases}\tag{2}$$

where $D_{t+1,1}^{\mathrm{Htd}}$ and $D_{t+1,2}^{\mathrm{Htd}}$ denote the data of two sub nodes.

A recursive program is utilized for (1) and (2) to obtain the $(T-1)/2$ nonleaf nodes. These $(T+1)/2$ leaf nodes, on the other hand, use arithmetic average to calculation the final prediction. Taking the tth $(\mathrm{t}\le(T-1)/2)$ leaf node as an example, it can be denoted as follows:

$$\hat{y}_i=\sum_{i=1}^{N_t^{\mathrm{LN}}}y_i^{\mathrm{Htd}}\tag{3}$$

where N_t^{LN} is the number of samples in tth leaf node and y_i^{Htd} is the true value in tth leaf node.

Therefore, the RDT model is represented as follows:

$$f_{n^{\mathrm{ML}}}^{\mathrm{ML}}(\boldsymbol{x}^{\mathrm{Htd}})=\sum_{i=1}^{N_t^{\mathrm{LN}}}y_i^{\mathrm{Htd}}\boldsymbol{\varphi}^t(\boldsymbol{x}^{\mathrm{Htd}})\tag{4}$$

where $f_{n^{\mathrm{ML}}}^{\mathrm{ML}}(\cdot)$ is the n^{ML}th RDT function in mapping layer, $\boldsymbol{\varphi}^t(\cdot)$ denotes the tth route in the n^{ML}th RDT.

Suppose the mapping layer have N^{ML} RDT function, the output of mapping layer can be expressed as follows,

$$\boldsymbol{Z}^{\mathrm{ML}}=[\boldsymbol{z}_1^{\mathrm{ML}},...,\boldsymbol{z}_{n^{\mathrm{ML}}}^{\mathrm{ML}},...,\boldsymbol{z}_{N^{\mathrm{ML}}}^{\mathrm{ML}}]\in\mathbb{R}^{N\times N^{\mathrm{ML}}}\tag{5}$$

$$\boldsymbol{z}_{n^{\mathrm{ML}}}^{\mathrm{ML}}=f_{n^{\mathrm{ML}}}^{\mathrm{ML}}(\boldsymbol{X}^{\mathrm{Htd}})=\sum_{i=1}^{N_t^{\mathrm{LN}}}y_i^{\mathrm{Htd}}\boldsymbol{\varphi}^t(\boldsymbol{X}^{\mathrm{Htd}})\in\mathbb{R}^{N\times 1}\tag{6}$$

The enhancement layer has a similar structure and unit to the mapping layer. Similarly, we assume the enhancement

layer consists of N^{EL} RDT functions, denoted as $\left\{f_{n^{\mathrm{ML}}}^{\mathrm{EL}}(\cdot),\ldots,f_{n^{\mathrm{ML}}}^{\mathrm{EL}}(\cdot),\ldots,f_{N^{\mathrm{ML}}}^{\mathrm{EL}}(\cdot)\right\}$. The output can be denoted as,

$$\boldsymbol{Z}^{\mathrm{EL}}=[\boldsymbol{z}_1^{\mathrm{EL}},\ldots,\boldsymbol{z}_{n^{\mathrm{EL}}}^{\mathrm{EL}},\ldots,\boldsymbol{z}_{N^{\mathrm{EL}}}^{\mathrm{EL}}]\in\mathbb{R}^{N\times N^{\mathrm{EL}}} \tag{7}$$

$$\boldsymbol{z}_{n^{\mathrm{EL}}}^{\mathrm{EL}}=f_{n^{\mathrm{EL}}}^{\mathrm{EL}}(\boldsymbol{X}^{\mathrm{Htd}})=\sum_{i=1}^{N_i^{\mathrm{LN}}}y_i^{\mathrm{Htd}}\boldsymbol{\varphi}^{i}(\boldsymbol{X}^{\mathrm{Htd}})\in\mathbb{R}^{N\times 1} \tag{8}$$

By Integrating the mapping and enhancement layer, the output can be expressed as $\boldsymbol{Z}^k=[\boldsymbol{Z}^{\mathrm{ML}},\boldsymbol{Z}^{\mathrm{EL}}]\in\mathbb{R}^{N\times(N^{\mathrm{EL}}+N^{\mathrm{EL}})}$. Hence, the output of the output layer can be written as follows:

$$\hat{\boldsymbol{y}}=\boldsymbol{Z}^{\mathrm{ML}}\boldsymbol{\omega}_{\mathrm{ML}}^k+\boldsymbol{Z}^{\mathrm{EL}}\boldsymbol{\omega}_{\mathrm{ML}}^k=[\boldsymbol{Z}^{\mathrm{ML}},\boldsymbol{Z}^{\mathrm{EL}}]\begin{bmatrix}\boldsymbol{\omega}_{\mathrm{ML}}^k\\ \boldsymbol{\omega}_{\mathrm{EL}}^k\end{bmatrix}=\boldsymbol{Z}^k\boldsymbol{W}^k \tag{9}$$

where $\boldsymbol{W}^k$ is the weight of each RDT function in mapping layer and enhancement layer, it is further evaluated as follows:

$$W^k=([\boldsymbol{Z}^k]^{\mathrm{T}}\boldsymbol{Z}^k+\lambda_2\boldsymbol{I})^{-1}[\boldsymbol{Z}^k]^{\mathrm{T}}\boldsymbol{y}^{\mathrm{Htd}}=[\boldsymbol{Z}^k]^{*}\boldsymbol{y}^{\mathrm{Htd}} \tag{10}$$

where $W^k\in\mathbb{R}^{(N^{\mathrm{EL}}+N^{\mathrm{EL}})\times 1}$, $[\boldsymbol{Z}^k]^{*}$ represents the pseudoinverse matrix of $\boldsymbol{Z}^k$, $\lambda_2>0$ is a regular term ratio.

When both the mapping layer and enhancement layer show insufficient performance, the model undergoes incremental broadening by using a group of neurons to realize rapid learning and obtain even better performance. In increment layer, each group consists of multiple neurons. When the first neuron group is added, the mapping process is represented as follows:

$$\boldsymbol{Z}_1^{\mathrm{OffIL}}=[\boldsymbol{z}_1^{\mathrm{OffIL}},\ldots,\boldsymbol{z}_{N_1^{\mathrm{OffIL}}}^{\mathrm{OffIL}}]\in\mathbb{R}^{N\times N_1^{\mathrm{OffIL}}} \tag{11}$$

$$\boldsymbol{z}_{n_1^{\mathrm{OffIL}}}^{\mathrm{OffIL}}=f_{n_1^{\mathrm{OffIL}}}^{\mathrm{OffIL}}(\boldsymbol{Z}^k\boldsymbol{\omega}_{n_1^{\mathrm{OffIL}}}^{\mathrm{OffIL}}+b_{n_1^{\mathrm{OffIL}}}^{\mathrm{OffIL}})\in\mathbb{R}^{N\times 1} \tag{12}$$

where $\boldsymbol{Z}_1^{\mathrm{OffIL}}$ means the output of the first neuron group, N_1^{OffIL} denotes the number of neuron, $\boldsymbol{z}_{n_1^{\mathrm{OffIL}}}^{\mathrm{OffIL}}$ indicates the output of the n_1^{OffIL}th $(n_1^{\mathrm{OffIL}}<N_1^{\mathrm{OffIL}})$ neuron, $f_{n_1^{\mathrm{OffIL}}}^{\mathrm{OffIL}}(\cdot)$ is the active function, and $\omega_{n_1^{\mathrm{OffIL}}}^{\mathrm{OffIL}}$ and $b_{n_1^{\mathrm{OffIL}}}^{\mathrm{OffIL}}$ express the weight and bias.

Consequently, the output of the mapping, enhancement, and increment layer is denoted as $\boldsymbol{Z}^{k+1}=[\boldsymbol{Z}^{\mathrm{ML}},\boldsymbol{Z}^{\mathrm{EL}},\boldsymbol{Z}_1^{\mathrm{OffIL}}]\in\mathbb{R}^{N\times(N^{\mathrm{EL}}+N^{\mathrm{EL}}+N_1^{\mathrm{OffIL}})}$. Then, the column recursion algorithm is employed to solve the Moore-Penrose inverse matrix. The new inverse matrix can be written as,

$$[\boldsymbol{Z}^{k+1}]^{*}=\begin{bmatrix}[\boldsymbol{Z}^k]^{*}-\boldsymbol{A}\boldsymbol{B}^{\mathrm{T}}\\ \boldsymbol{B}^{\mathrm{T}}\end{bmatrix} \tag{13}$$

where

$$\begin{cases}\boldsymbol{A}=[\boldsymbol{Z}^{k+1}]^{*}\boldsymbol{H}_{k+1}\\ \boldsymbol{B}^{\mathrm{T}}=\begin{cases}[\boldsymbol{C}]^{*}, & \text{if } \boldsymbol{C}\neq 0\\ [1+\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A}]^{-1}\boldsymbol{A}^{\mathrm{T}}[\boldsymbol{Z}^{k+1}]^{*}, & \text{if } \boldsymbol{C}=0\end{cases}\\ \boldsymbol{C}=\boldsymbol{H}_{k+1}-\boldsymbol{Z}^{k+1}\boldsymbol{A}\end{cases} \tag{14}$$

After that, the weight matrix of the mapping, enhancement, and increment layer is updated as follows:

$$\boldsymbol{W}^{k+1}=\begin{bmatrix}\boldsymbol{W}^k-\boldsymbol{A}\boldsymbol{B}^{\mathrm{T}}\boldsymbol{y}^{\mathrm{Htd}}\\ \boldsymbol{B}^{\mathrm{T}}\boldsymbol{y}^{\mathrm{Htd}}\end{bmatrix} \tag{15}$$

Finally, after adding the first neuron group, the prediction of the offline model is obtained as follows:

$$\begin{aligned}\hat{\boldsymbol{y}}&=\boldsymbol{Z}^{k+1}\boldsymbol{W}^{k+1}=[\boldsymbol{Z}^{\mathrm{ML}},\boldsymbol{Z}^{\mathrm{EL}},\boldsymbol{Z}_1^{\mathrm{OffIL}}]\begin{bmatrix}\boldsymbol{\omega}_{\mathrm{ML}}^{k+1}\\ \boldsymbol{\omega}_{\mathrm{EL}}^{k+1}\\ \boldsymbol{\omega}_{\mathrm{OffIL}}^{k+1}\end{bmatrix}\\ &=\boldsymbol{Z}^{\mathrm{ML}}\boldsymbol{\omega}_{\mathrm{ML}}^{k+1}+\boldsymbol{Z}^{\mathrm{EL}}\boldsymbol{\omega}_{\mathrm{ML}}^{k+1}+\boldsymbol{Z}_1^{\mathrm{OffIL}}\boldsymbol{\omega}_{\mathrm{OffIL}}^{k+1}\end{aligned} \tag{16}$$

B. Offline Drift Indicators' Control Limit Calculation Based on RPCA

The offline concept indicators' control limit is performed using RPCA. The history training data D^{Htd} consists of high-dimensional process variables $\boldsymbol{X}^{\mathrm{Htd}}=[\boldsymbol{x}_1^{\mathrm{Htd}},\boldsymbol{x}_2^{\mathrm{Htd}},\ldots,\boldsymbol{x}_N^{\mathrm{Htd}}]^{\mathrm{T}}$, and the true values $\boldsymbol{y}^{\mathrm{Htd}}=[y_1^{\mathrm{Htd}},y_2^{\mathrm{Htd}},\ldots,y_N^{\mathrm{Htd}}]^{\mathrm{T}}$. We suppose the noise data matrix is represented by $\boldsymbol{E}^{\mathrm{Htd}}$ and the actual process data matrix is denoted by $\bar{\boldsymbol{X}}^{\mathrm{Htd}}$ ($rank(\bar{\boldsymbol{X}}^{\mathrm{Htd}})=r, and\ r<M$). The original process data matrix $\boldsymbol{X}^{\mathrm{Htd}}$ can be decomposed as,

$$\boldsymbol{X}^{\mathrm{Htd}}=\bar{\boldsymbol{X}}^{\mathrm{Htd}}+\boldsymbol{E}^{\mathrm{Htd}} \tag{17}$$

where $\bar{\boldsymbol{X}}^{\mathrm{Htd}}$ and $\boldsymbol{E}^{\mathrm{Htd}}$ are low-rank and sparse matrix.

To address the matrix separation problem, we can transform (17) into a convex optimization problem as follows,

$$\begin{aligned}&\min\left\|\bar{\boldsymbol{X}}^{\mathrm{Htd}}\right\|_{*}+\lambda_1\left\|\boldsymbol{E}^{\mathrm{Htd}}\right\|_{1}\\ &s.t.\ \ \bar{\boldsymbol{X}}^{\mathrm{Htd}}+\boldsymbol{E}^{\mathrm{Htd}}=\boldsymbol{X}^{\mathrm{Htd}}\end{aligned} \tag{18}$$

where $\left\|\bar{\boldsymbol{X}}^{\mathrm{Htd}}\right\|_{*}=\sum\sigma_i(\bar{\boldsymbol{X}}^{\mathrm{Htd}})$ is the norm of the matrix $\bar{\boldsymbol{X}}^{\mathrm{Htd}}$, and $\left\|\boldsymbol{E}^{\mathrm{Htd}}\right\|_{1}=\sum\left|\boldsymbol{E}^{\mathrm{Htd}}\right|$ represents the sum of the magnitudes vector of the matrix $\boldsymbol{E}^{\mathrm{Htd}}$.

According to (18), an iterative strategy is employed to minimize both $\bar{\boldsymbol{X}}^{\mathrm{Htd}}$ and $\boldsymbol{E}^{\mathrm{Htd}}$, which can be expressed as,

$$\boldsymbol{E}_{k+1}^{\mathrm{Htd}}=\arg\min_{\boldsymbol{E}^{\mathrm{Htd}}}\left\{L(\bar{\boldsymbol{X}}_k^{\mathrm{Htd}},\boldsymbol{E}^{\mathrm{Htd}},Q_k^{\mathrm{Htd}},\beta)\right\} \tag{19}$$

$$\bar{\boldsymbol{X}}_{k+1}^{\mathrm{Htd}}=\arg\min_{\bar{\boldsymbol{X}}^{\mathrm{Htd}}}\left\{L(\bar{\boldsymbol{X}}^{\mathrm{Htd}},\boldsymbol{E}_k^{\mathrm{Htd}},Q_k^{\mathrm{Htd}},\beta)\right\} \tag{20}$$

$$Q_{k+1}^{\mathrm{Htd}}=Q_k^{\mathrm{Htd}}-\beta(\boldsymbol{X}_{k+1}^{\mathrm{Htd}}+\boldsymbol{E}_{k+1}^{\mathrm{Htd}}-\bar{\boldsymbol{X}}^{\mathrm{Htd}},\beta) \tag{21}$$

The correlation matrix R^{Htd} of the low-rank matrix $\bar{\boldsymbol{X}}^{\mathrm{Htd}}\in\mathbb{R}^{N\times M}$ can be written as follows:

$$R^{\mathrm{Htd}}=\frac{1}{N-1}[\bar{\boldsymbol{X}}^{\mathrm{Htd}}]^{\mathrm{T}}\bar{\boldsymbol{X}}^{\mathrm{Htd}} \tag{22}$$

By using SVD method to compute the eigenvalue of the correlation matrix R^{Htd}, it can be denoted as follows:

$$R^{\mathrm{Htd}}=\boldsymbol{U}^{\mathrm{Htd}}\boldsymbol{\Sigma}^{\mathrm{Htd}}[\boldsymbol{V}^{\mathrm{Htd}}]^{\mathrm{T}} \tag{23}$$

where $\boldsymbol{U}^{\mathrm{Htd}}$ and $\boldsymbol{V}^{\mathrm{Htd}}$ express the orthogonal matrix ($\boldsymbol{V}^{\mathrm{Htd}}=[\boldsymbol{U}^{\mathrm{Htd}}]^{\mathrm{T}}$), $\boldsymbol{\Sigma}^{\mathrm{Htd}}=diag(\sigma_1,\sigma_2,\ldots,\sigma_M)$ is a diagonal matrix and σ_m is the mth eigenvalue of R^{Htd}.

To reduce redundant information, the main principal components are selected based on the following formula:

$$\eta^{\mathrm{Htd}}=\left(\sum_{m^{\mathrm{PC}}}^{M^{\mathrm{PC}}}\sigma_i \Big/ \sum_{m}^{M}\sigma_i\right)\le\delta^{\mathrm{PC}} \tag{24}$$

where M^{PC} is the number of selected principal components and δ^{PC} indicates the selected threshold by experience setting.

Based on the selected M^{PC} value, the principal component loading matrix $\boldsymbol{U}^{\mathrm{Htd}}$ is expressed by $\left[[\boldsymbol{U}_{\mathrm{PC}}^{\mathrm{Htd}}]^{M\times M^{\mathrm{PC}}}\ [\boldsymbol{U}_{\mathrm{UC}}^{\mathrm{Htd}}]^{M\times(M-M^{\mathrm{PC}})}\right]$. Consequently, the low-rank matrix $\bar{\boldsymbol{X}}^{\mathrm{Htd}}\in\mathbb{R}^{N\times M}$ is written as,

$$\begin{aligned}\bar{\boldsymbol{X}}^{\mathrm{Htd}}&=\bar{\boldsymbol{X}}^{\mathrm{Htd}}\boldsymbol{I}=\bar{\boldsymbol{X}}^{\mathrm{Htd}}(\boldsymbol{U}_{\mathrm{PC}}^{\mathrm{Htd}}[\boldsymbol{U}_{\mathrm{PC}}^{\mathrm{Htd}}]^{\mathrm{T}}+\boldsymbol{I}-\boldsymbol{U}_{\mathrm{PC}}^{\mathrm{Htd}}[\boldsymbol{U}_{\mathrm{PC}}^{\mathrm{Htd}}]^{\mathrm{T}})\\&=\bar{\boldsymbol{X}}^{\mathrm{Htd}}\boldsymbol{U}_{\mathrm{PC}}^{\mathrm{Htd}}[\boldsymbol{U}_{\mathrm{PC}}^{\mathrm{Htd}}]^{\mathrm{T}}+\bar{\boldsymbol{X}}^{\mathrm{Htd}}(\boldsymbol{I}-\boldsymbol{U}_{\mathrm{PC}}^{\mathrm{Htd}}[\boldsymbol{U}_{\mathrm{PC}}^{\mathrm{Htd}}]^{\mathrm{T}})\\&=\hat{\boldsymbol{X}}_{\mathrm{PC}}^{\mathrm{Htd}}+\tilde{\boldsymbol{X}}_{\mathrm{UC}}^{\mathrm{Htd}}\end{aligned} \tag{25}$$

where $\hat{\boldsymbol{X}}_{\mathrm{PC}}^{\mathrm{Htd}}$ and $\tilde{\boldsymbol{X}}_{\mathrm{UC}}^{\mathrm{Htd}}$ are the principal and residual space.

The T^2 and SPE statistical indicators are utilized for monitoring the drift of operation condition. Given the low-rank matrix $\bar{\boldsymbol{X}}^{\mathrm{Htd}}\in\mathbb{R}^{N\times M}$, the control limit of the T^2 (θ_0^{T2}) and SPE (θ_0^{SPE})indicators can be denoted as,

$$\theta_0^{\mathrm{T2}}=\frac{M_{\mathrm{PCA}}(N-1)}{(N-M_{\mathrm{PCA}})}f_\alpha\left(M_{\mathrm{PCA}},N-M_{\mathrm{PCA}}\right) \tag{26}$$

$$\theta_0^{\mathrm{SPE}}=\Theta_1\left(\frac{c_\alpha\sqrt{2\Theta_2h_0^2}}{\Theta_1}+1+\frac{\Theta_2h_0(h_0-1)}{\Theta_1^2}\right)^{1/h_0} \tag{27}$$

where $f_\alpha\left(M_{\mathrm{PCA}},N-M_{\mathrm{PCA}}\right)$ indicates the F distribution with M_{PCA} and $(N-M_{\mathrm{PCA}})$ degrees of freedom, c_α indicates the normal bias regard to the upper $(1-\alpha)$ percentile.

C. Online Drift Indicator Calculation

A moving sampling window of fixed size is utilized. For online data within the window W1, the new data matrix $\boldsymbol{X}_{k+1}=[\boldsymbol{X}^{\mathrm{Htd}};\boldsymbol{X}_{\mathrm{W1}}^{\mathrm{Otd}}]$ and correlation matrix R^{Htd} are recursively calculated as follows:

$$\begin{cases}\boldsymbol{q}_{k+1}=\dfrac{N_k}{N_{k+1}}\boldsymbol{q}_{k+1}+\dfrac{1}{N_{k+1}}(\boldsymbol{X}_k)^{\mathrm{T}}\boldsymbol{I}_{k+1}^{\mathrm{T}}\\ \boldsymbol{X}_{k+1}=\begin{bmatrix}\boldsymbol{X}_k\boldsymbol{\Sigma}_k\boldsymbol{\Sigma}_{k+1}^{-1}-\boldsymbol{1}_k[\boldsymbol{q}_{k+1}-\boldsymbol{q}_k]^{\mathrm{T}}\boldsymbol{\Sigma}_{k+1}^{-1}\\ \boldsymbol{X}_{k+1}^{\mathrm{W1}}\end{bmatrix}\\ R_{k+1}=\dfrac{N_k-1}{N_{k+1}-1}\boldsymbol{\Sigma}_{k+1}^{-1}\boldsymbol{\Sigma}_kR_k\boldsymbol{\Sigma}_k\boldsymbol{\Sigma}_{k+1}^{-1}\\ \qquad\qquad+\dfrac{N_k}{N_{k+1}-1}\boldsymbol{\Sigma}_{k+1}^{-1}[\boldsymbol{q}_{k+1}-\boldsymbol{q}_k][\boldsymbol{q}_{k+1}-\boldsymbol{q}_k]^{\mathrm{T}}\boldsymbol{\Sigma}_{k+1}^{-1}\\ \qquad\qquad+\dfrac{1}{N_{k+1}-1}[\boldsymbol{X}_{k+1}^{\mathrm{W1}}]^{\mathrm{T}}\boldsymbol{X}_{k+1}^{\mathrm{W1}}\end{cases} \tag{28}$$

where

$$\begin{cases}N_k=N\\ \boldsymbol{X}_k=\boldsymbol{X}^{\mathrm{Htd}}\\ \boldsymbol{\Sigma}_i=diag(\sigma_{i,1},\sigma_{i,2},...,\sigma_{i,M}),\ i=k,k+1\\ \boldsymbol{X}_{k+1}^{\mathrm{W1}}=(\boldsymbol{X}_{\mathrm{W1}}^{\mathrm{Otd}}-\boldsymbol{I}_{k+1}[\boldsymbol{q}_{k+1}]^{\mathrm{T}})\boldsymbol{\Sigma}_{k+1}^{-1}\end{cases} \tag{29}$$

After the low-rank matrix $\bar{\boldsymbol{X}}^{\mathrm{W1}}$ is obtained, the principal component loading matrix $\boldsymbol{U}_{\mathrm{PC}}^{\mathrm{W1}}$ is calculated based on M^{PC} principal components. The T^2 and SPE of samples within the window W1 are denoted as follows:

$$T_{k+1}^2=[\boldsymbol{x}_{k+1}^{\mathrm{W1}}]^{\boldsymbol{T}}\boldsymbol{U}_{\mathrm{PC}}^{\mathrm{W1}}\Lambda[\boldsymbol{U}_{\mathrm{PC}}^{\mathrm{W1}}]^{\mathrm{T}}\boldsymbol{x}_{k+1}^{\mathrm{W1}} \tag{30}$$

$$SPE_{k+1}=[\boldsymbol{x}_{k+1}^{\mathrm{W1}}]^{\mathrm{T}}(\boldsymbol{I}-\boldsymbol{U}_{\mathrm{PC}}^{\mathrm{W1}}[\boldsymbol{U}_{\mathrm{PC}}^{\mathrm{W1}}]^{\mathrm{T}})\boldsymbol{x}_{k+1}^{\mathrm{Otd}} \tag{31}$$

D. Online Soft Measuring on Actual MSWI Plant

Therefore, the defined concept drift detection strategy and online soft measuring method are as follows.

1) The condition is considered as normal (no drift) when the T_{k+1}^2 value is less than a certain control limit θ_0^{T2} and the SPE_{k+1} value is less than a certain control limit θ_0^{SPE}. The offline constructed FTBLS model is used in this case.

2) The condition is deem abnormal (drift) when the T_{k+1}^2 value exceed the control limit θ_0^{T2} or the SPE_{k+1} value surpasses the control limit θ_0^{SPE}. It is important to note that no DXN truth data is available for offline constructed model to make updating. To address this issue, we provides the following steps.

- Step 1: Calculate the similarity between an abnormal sample and historical sample using Euclidean distance (as shown in (40)). Select the most similarity historical sample $\underbrace{[...,\boldsymbol{x}_n^{\mathrm{History}},...]}_{similar\ samples\ set}\in\mathbb{R}^{N^{\mathrm{Sim}}\times M}(N^{\mathrm{Sim}}\ll N)$ by artificial setting threshold δ^{Sim};

$$\begin{aligned}\boldsymbol{d}_i^{\mathrm{MSWI}}&=[d_1^i,...,d_N^i]\in\mathbb{R}^{1\times N}\\&=\left[\sqrt{\sum_{m=1}^{M}(x_{i,m}^{\mathrm{MSWI}}-x_{1,m}^{\mathrm{History}})^2},...,\sqrt{\sum_{m=1}^{M}(x_{i,m}^{\mathrm{MSWI}}-x_{N,m}^{\mathrm{History}})^2}\right]\end{aligned} \tag{32}$$

$$\underbrace{[...,\boldsymbol{x}_n^{\mathrm{History}},...]}_{similar\ samples\ set}\in\mathbb{R}^{N^{\mathrm{Sim}}\times M}\leftarrow\bar{\boldsymbol{d}}_i^{\mathrm{MSWI}}\xleftarrow{d_1^i\ge\delta^{\mathrm{Sim}}}[d_1^i,...,d_N^i] \tag{33}$$

- Step 2: Obtain a prediction vector $[...,y_n,...]\in\mathbb{R}^{1\times N^{\mathrm{Sim}}}$ by inputting the most similarity historical sample $\underbrace{[...,\boldsymbol{x}_n^{\mathrm{History}},...]}_{similar\ samples\ set}$ into the model;
- Step 3: Arithmetic averaging is performed on prediction vectors $[...,y_n,...]\in\mathbb{R}^{1\times N^{\mathrm{Sim}}}$ to calculate the final measured values of the abnormal samples.

IV. Expriment and Application

A. Experiment on the DXN dataset

The DXN data used for experiment is obtained from an actual MSWI plant located in Beijing, China. The data spans from 2009 to 2020. The details are presented in Table I.

TABLE I. The Description of the DXN Data

	Data	Training data	Testing data
Sample	141	71	70
Feature	116	116	116
Time	2009s~2020s	2009s~2016s	2016s~2020s

The historical data consists of true value samples collected from 2009 to 2016, while the online testing data contains samples collected from 2016 to 2020. The true value of DXN concentration is determined by averaging the emission concentration over a two-hour period in the MSWI processes. To obtain a sample corresponding to the DXN true value, the process data are processed by using simply averaged algorithm over the sampling time.

In the comparison of offline modeling methods, the following techniques are employed, i.e., RF, BPNN, SVR, DFR, and BLS, fuzzy tree BLS (FuzzyTBLS), and FTBLS. The experimental results are presented in Table II.

TABLE II. COMPARISON RESULTS OF DIFFERENT METHODS

Method	History training data (2009s~2016s)		Testing data (2016s~2020s)	
	RMSE	Explained variance	RMSE	Explained variance
RF	0.0132	0.7913	0.0274	-0.1980
BP	0.0009	0.9990	0.0573	-16.6480
SVR	0.0067	0.7749	0.0202	-6.8958
DFR	0.0154	0.7176	0.0286	-0.1468
BLS	0.0101	0.8795	0.0270	-2.4269
FuzzyTBLS [28]	0.0071	0.9398	0.0236	-1.9953
FTBLS	0.0000	1.0000	0.0198	-1.8706

As shown in Table II, all the methods demonstrate effective modeling results for the history training data. Among them, FTBLS (with an RMSE of 0 and explained variance of 1) and BP (with an RMSE of 0.0009 and explained variance of 0.999) achieve optimal learning performance compared to the other methods. Furthermore, the methods based on DT (such as RF and DFR) exhibit weaker performance compared to the neural network-based methods (BP and BLS). This suggest that tree-based methods have lower learning performance on training dataset compared to neuron-based methods. In terms of the testing data, all methods show a negative explained variance, indicating that they struggle to fit the original dataset. However, RF and DFR perform relatively better than other methods. This suggest that the tree-based methods (RF and DFR) show more stability and less over-fitting compared to the neuron-based methods (BP and BLS).

In the testing process, a moving window of 5 samples is employed. Table III demonstrates the number abnormal samples, the online increment neurons group, the RMSE of updated FTBLS, and update time of online increment neurons group.

TABLE III. RESULTS OF TESTING PROCESS

Windows	Abnormal	Increment neurons group	RMSE	Update time (second)
1	5	4	1.0350e-05	0.04
2	5	3	6.6022e-06	0.03
3	5	3	7.9588e-06	0.03
4	5	3	6.9205e-06	0.02
5	5	3	1.0293e-05	0.02
6	5	3	6.6765e-06	0.02
7	5	5	8.0492e-06	0.04
8	5	3	1.2056e-05	0.02
9	5	3	7.1671e-06	0.03
10	5	3	9.5905e-06	0.03
11	5	4	7.8853e-06	0.04
12	5	3	8.6091e-06	0.03
13	5	3	7.6234e-06	0.02
14	5	3	6.2579e-06	0.02

The results of SPE for each sample in moving window, as well as the control limit are shown in Fig. 8.

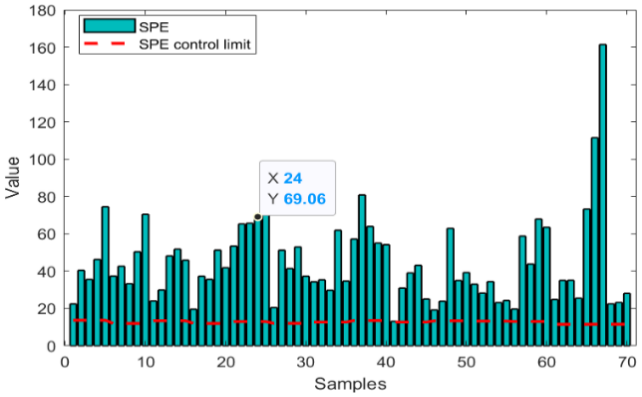

Fig. 2. Results of online operation condition detection.

B. *Appllication on the MSWI plant*

The developed software (Upside of Fig. 3) is deployed on a secure isolation experimental system (middle of Fig. 3) located on the edge side of the MSWI plant. It is connected to the DCS (distribution control system) system of the actual MSWI plant for implementation purposes. After running for 1 hours, the obtained results are illustrated in Fig. 3. The proposed method has been validated in the normal operation condition of an actual MSWI plant, which indicates successful implementation. Therefore, the industrial application proved the effectiveness of the proposed method in online detection of DXN emission concentration.

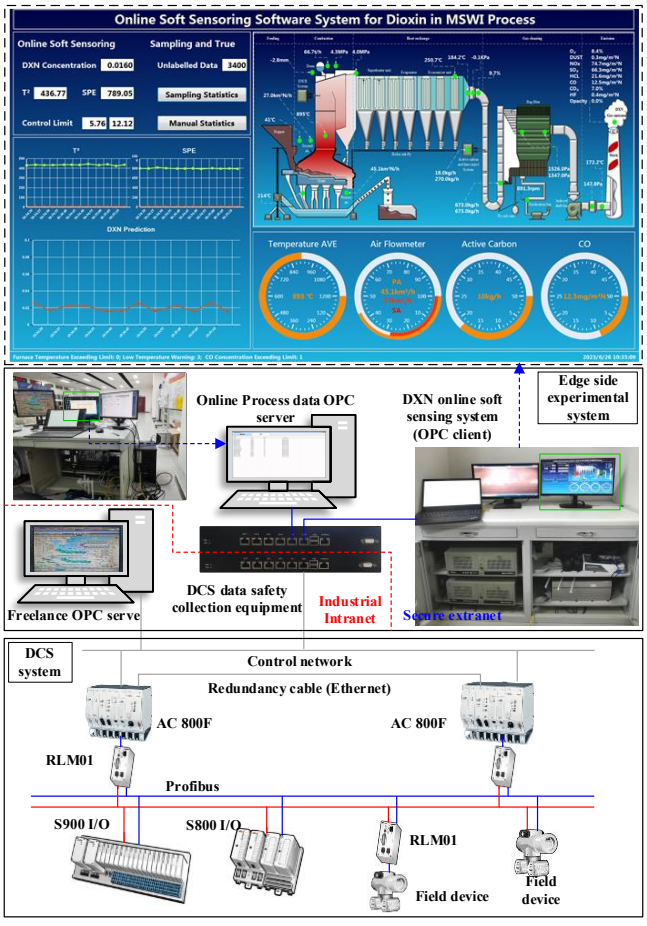

Fig. 3. Online soft sensing system application on the actual MSWI plant

V. CONCLUSIONS

To enable online soft sensing of DXN emission concentration, this article introduces a method based on RPCA and FTBLS algorithms. The main contributions of this study are as follows. (1) A novel approach called FTBLS is presented for modeling historical data, which involved the mapping and enhancement layer based on DT and increment layer based on neurons. (2) The RPCA technique is applied to eliminate noise from process data for effective detection of operating condition drift. This approach enhances the robustness of condition detection and improves the accuracy of identifying variations in operating conditions. (3) The similarity between an abnormal sample and historical sample is used to aid the soft measuring value's obtainment for concept drift sample. A software system for soft sensing DXN emission concentration is developed in an actual MSWI plant.

REFERENCES

[1] L. Wang, D. Y. Yan, Y. Xiong, and L. H., Zhou, "A review of the challenges and application of public-private partnership model in Chinese garbage disposal industry," Journal of Cleaner Production, vol. 230, pp. 219-229, 2019.

[2] H. Xia, J. Tang, L. Aljerf, C. L. Cui, B. Y. Gao, and P. O. Ukaogo, "Dioxin emission modeling using feature selection and simplified DFR with residual error fitting for the grate-based MSWI process," Waste Management, vol. 168, pp. 256-271, 2023.

[3] G. F. Meng, Z. Guo, and J. L. Li, "The dynamic linkage among urbanisation, industrialisation and carbon emissions in China: Insights from spatiotemporal effect," Science of The Total Environment, vol. 760, pp 1-11, 2021.

[4] J. Tang, H. Xia, W. Yu, and J. F. Qiao, "Research status and prospects of intelligent optimization control for municapal solid waste incineration processes," Acta Automatica Sinica, pp. 1-41, 2023, DOI:10.16383/j.aas.c220810.

[5] H. Xia, J. Tang, and L. Aljerf, "Dioxin emission prediction based on improved deep forest regression for municipal solid waste incineration process," Chemosphere, vol. 294, 133716, 2022.

[6] T. Yang, X. L. Yi, S. W. Lu, K. H. Johansson, and T. Y. Chai, "Intelligent manufacturing for the process industry driven by industrial artificial intelligence," Engineering, vol. 7, no. 9, pp. 1224-1230, 2021.

[7] G. Vilardi, and N. Verdone, "Exergy analysis of municipal solid waste incineration processes: The use of O2-enriched air and the oxy-combustion process," Energy, vol. 239, 122147, 2022.

[8] R. Caferra, I. D'Adamo, and P. Morone, "Wasting energy or energizing waste? The public acceptance of waste-to-energy technology," Energy, vol. 263, Part E, 2023.

[9] H. Xia, J. Tang, L. Aljerf, T. Z. Wang, B. Y. Gao, Q. D. Xu, and P. Ukaogo, "Assessment of PCDD/Fs formation and emission characteristics at a municipal solid waste incinerator for one year," Science of The Total Environment, vol. 883, 163705, 2023.

[10] Y. Peng, S. Lu, X. Li, J. Yan, and K. Cen, "Formation, measurement, and control of dioxins from the incineration of municipal solid wastes: Recent advances and perspectives," Energy & Fuels, vol. 34, no. 11, pp. 13247-13267, 2020.

[11] J. X. Wei, H. Li, and J. G. Liu, "Fate of dioxins in a municipal solid waste incinerator with state-of-the-art air pollution control devices in China," Environmental Pollution, vol. 289, pp. 1-10, 2021.

[12] J. F. Qiao, Z. H. Guo, and J. Tang, "Dioxin emission concentration measurement approaches for municipal solid wastes incineration process: a survey," Acta Automatica Sinica, vol. 46, no. 6, pp. 1063-1089, 2020.

[13] F. Hasselriis, "Minimizing trace organic emissions from combustion of municipal solid waste by the use of carbon monoxide monitors," National Waste Processing Conference, Denver. 1986.

[14] F. Hasselriis, "Optimization of combustion conditions to minimize dioxin emissions," Waste Management & Research, vol. 5, no. 3, pp. 311-326, 1987.

[15] N. B. Chang, and S. H.Huang, "Statistical modelling for the prediction and control of PCDDs and PCDFs emissions from municipal solid waste incinerators," Waste Management & Research, vol. 13, no. 4, pp. 379-400, 1995.

[16] Y. Zhuo and Z. Ge, "Data guardian: a data protection scheme for industrial monitoring systems," IEEE Transactions on Industrial Informatics, vol. 18, no. 4, pp. 2550-2559, April 2022.

[17] N. B. Chang, and W. C. Chen, "Prediction of PCDDs/PCDFs emissions from municipal incinerators by genetic programming and neural network modeling," Waste Management & Research, vol. 18, no. 4, pp. 341-351, 2000.

[18] S. Bunsan, W. Y. Chen, H. W. Chen, Y. H. Chuang, and N. Grisdanurak, "Modeling the dioxin emission of a municipal solid waste incinerator using neural networks," Chemosphere, vol. 92, no. 3, pp. 258-264, 2013.

[19] Z. Guo, J. Tang, J. F. Qiao and H. He, "Dioxin emission concentration soft measurement model of MSWI process based on unmarked samples and improved deep belief network," the 39th Chinese Control Conference (CCC), Shenyang, China, 2020, pp. 5784-5789.

[20] J. Tang, H. Xia, L. Aljerf, D. D. Wang, and P. O. Ukaogo, "Prediction of dioxin emission from municipal solid waste incineration based on expansion, interpolation, and selection for small samples," Journal of Environmental Chemical Engineering, vol. 10, no. 5, 108314, 2022.

[21] X. D. Xiao, J. W. Lu, J. Hai, and L. Liao, "Prediction of dioxin emissions in flue gas from waste incineration based on support vector regression," Renewable Energy Resources, vol. 35, no. 8, pp. 1107-1114, 2017.

[22] J. Tang, J. F. Qiao and Z. H. Guo, "Dioxin emission concentration soft measurement based on multi-source latent feature selective ensemble modeling for municipal solid waste incineration process," Acta Automatica Sinica, vol. 48, no. 1, pp. 223-238, 2022.

[23] J. Tang, J. F. Qiao X. Zhe, and Z. H. Guo, "Soft measuring approach of dioxin emission concentration in municipal solid waste incineration process based on feature reduction and selective ensemble algorithm," Control Theory & Applications, vol. 38, no. 1, pp. 110-120, 2021.

[24] J. Tang, H. Xia, J. F. Qiao, and Z. H. Guo, "Modeling method of deep ensemble forest regression with its application," Journal of Beijing University of Technology, vol. 47, no. 11, pp. 1219-1229, 2021.

[25] J. Tang, H. Xia, J. Zhang, J. F. Qiao, and W. Yu, "Deep forest regression based on cross-layer full connection," Neural Computing and Applications, vol. 33, no. 15, pp. 9307-9328, 2021.

[26] H. Xia, J. Tang, and L. Aljerf, "Dioxin emission prediction based on improved deep forest regression for municipal solid waste incineration process," Chemosphere, vol. 294, 133716, 2022.

[27] H. Xia, J. Tang, C. L. Cui, and J. F. Qiao, "Soft sensing method of dioxin emission in municipal solid waste incineration process based on broad hybrid forest regression," Acta Automatica Sinica, vol. 49, no. 2, pp. 343-365, 2023.

[28] H. Xia, J. Tang, W. Yu and J. Qiao, "Online Measurement of Dioxin Emission in Solid Waste Incineration Using Fuzzy Broad Learning," IEEE Transactions on Industrial Informatics, doi: 10.1109/TII.2023.3259419.