IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received March 27, 2019, accepted April 8, 2019, date of publication April 17, 2019, date of current version May 20, 2019.

Digital Object Identifier 10.1109/ACCESS.2019.2911741

A Novel Nonlinear Partial Least Square Integrated With Error-Based Extreme Learning Machine

ZE DONG AND NING MA

School of Control and Computer Engineering, North China Electric Power University, Beijing 102206, China
Hebei Engineering Research Center of Simulation and Optimized Control for Power Generation, North China Electric Power University, Baoding 071003, China

Corresponding author: Ning Ma (maningncepu@163.com)

This work was supported in part by the National Natural Science Foundation of China under Grant 71471060, in part by the Natural Science Foundation of Hebei Province under Grant E2018502111, and in part by the Fundamental Research Funds for the Central Universities under Grant 2018QN097.

ABSTRACT This paper proposes a novel nonlinear partial least square (PLS) approach for dealing with the modeling problem of industrial processes with input variables in collinearity. The new method combines the external linear PLS framework with the internal extreme learning machine (ELM) function. First, PLS is used as the outer framework to extract the input and output latent variables as well as eliminating the collinearity of the original variables, and then ELM is employed to describe the nonlinear relation between pairs of latent variables. Besides, the weight updating strategy based on errors minimization is also involved to improve the prediction accuracy. Then, the pH-neutralization process is taken as a benchmark to verify the validity of the new model. Finally, this method is applied to model the NO_x emission of a 1000-MW coal-fired boiler, and root-mean-square error (RMSE) values are 5.9541 for the training dataset and 6.8323 for the testing dataset. Compared with linear PLS and another two nonlinear PLS methods, smaller prediction errors is obtained. The results indicate new nonlinear PLS model can be a better choice for establishing the model of NO_x emission for coal-fired boilers.

INDEX TERMS Nonlinear partial least square, extreme learn machine, weight update, NO_x emission, coal-fired boilers.

I. INTRODUCTION

In many chemical engineering processes, accurate and reliable real-time measurements of key variables such as emission components and product quality are highly desirable, which play an important role in production safety and process control [1]. However, on some occasions, these critical process variables are difficult to measure because of some reasons such as extreme environment, the particularly huge expense of analytical instruments, and measurement time delay [2]. Therefore, setting up a suitable model to obtain critical process variables through variables easy to measure becomes a popular method in industrial processes. Besides, some variables such as emission concentration and product quality always can be optimized by tuning adjustable operational parameters such as temperature, valve opening, and fuel quantity using a variety of intelligent optimization algorithms [3], [4]. In other word, establishing accurate models to predict these variables is a prerequisite for the type of optimization systems.

A mechanism model can be derived from the exactly physical formula and chemical equations. However, due to the complexity of real industrial processes, fundamental models are usually too simplistic to be used accurately and effectively in practice [5]. On the other hand, data-driven models have become an efficient modeling method and have been widely applied to various industrial fields. Unlike the mechanism modeling technique, data-driven models supply accurate information for a particular operation region with less knowledge details of the industrial processes. A lot of data-driven modeling methods have been proposed recently, and Kadlec *et al.* [6] extensively reviewed some of the methods and discussed of some open issues in the data-driven modeling developments. Statistical regression techniques and machine learning techniques such as artificial neural networks (ANNs) and support vector machine (SVM) have found an increasingly wide utilization in chemical and thermal fields for building data-driven models. In order to set

The associate editor coordinating the review of this manuscript and approving it for publication was Ruqiang Yan.

up an accurate data-driven model, a large number of data samples used for training models are needed. However, there is always the problem that the size of data samples obtained is too small to satisfy the demand of modeling. Additionally, the other challenge in large-scale nonlinear system modeling is the correlation among inputs in a model, which can decrease the model prediction performance, increase the model complexity, and cause the over fitting problem [7], [8].

Recently, the partial least squares (PLS), as an effective way to solve correlated inputs and limited sample data problems, has been widely used in industrial process modeling. PLS is a multivariate statistical technique that can obtain the orthogonal characteristic vectors of the independent variables and the dependent variables, with projecting the high dimensional data space between the independent variable and the dependent variable to the corresponding low dimensional feature space [9], [10]. However, PLS method is essentially a linear regression approach, which is suitable for linear processes or local ranges near the equilibrium point of nonlinear processes. For the complex nonlinear processes, nonlinear PLS method is needed. Many nonlinear PLS methods proposed can be generally fall into two categories. The first category is kernel partial least squares (KPLS), which maps nonlinear dependent variable sample data to the high dimensional feature space by the kernel function and then sets up a linear model in the high dimensional space [11]. The other category of nonlinear PLS is to integrate non-linear features within the linear PLS framework using a nonlinear mapping while retaining the outer linear PLS framework unchanged. The latter nonlinear PLS method is the main content in this work. In recent decades, several nonlinear methods have been used as internal mapping functions of nonlinear PLS, such as least squares support vector machine PLS and neural network PLS. Wold *et al.* [10] introduced a nonlinear PLS algorithm that the linear mapping was replaced by the quadratic function. Qin and McAvoy [12] employed the neural network to establish the inner relationship between the latent variables. Li *et al.* [13] described the fuzzy PLS algorithm in which the fuzzy inference system was applied to map the inner model. Lv *et al.* [14] proposed the least squares support vector machine nonlinear PLS algorithm with the weight-updating procedure. What's more, other nonlinear PLS methods have also been found in previous researches [15]–[17].

Although scholars have conducted extensive research into the nonlinear PLS methods, there are still some drawbacks in existing methods. For instance, the spline PLS and artificial neural networks (ANNs) PLS are expected to be appropriate internal mapping function for fitting complex nonlinearity, but the extra flexibility can lead to over fitting and prediction errors [18]. Extreme learning machine (ELM) is a novel single hidden layer feed-forward network with good nonlinear mapping capability [19]. Compared with ANNs and other neural works methods, ELM transforms the learning training problem into solving the least squares norm problem of output weight matrix, which gives it the advantages of avoiding falling into local extremism, fast learning speed, few adjustment parameters and powerful generalization capability [20].

In this article, the ELM was used as the inner function to map nonlinear relationship between the inputs and outputs and PLS was applied as framework. Moreover, the error-based input weight updating was also embedded into the algorithm to form a novel nonlinear PLS method (defined as EBELMPLS algorithm). The pH-neutralization process was taken as a benchmark to verify the validity of new method. The new method and other PLS models were used to set up the relationship among NO_x emissions and operational parameters of a 1000MW coal-fired boiler. Results and comparisons of those models were also discussed.

II. PRELIMINARIES

A. OUTER PLS FRAME

Assume the input set and output set are $\mathbf{X}\in\mathrm{R}^N\times S$ and $\mathbf{Y}\in\mathrm{R}^N\times M$, where the indexes N represents the numbers of samples, S represents the number of dimensions of the input (predictor) variables, and M is the number of dimensions of the output (response) variables. The outer PLS decomposes $\mathbf{X}$ and $\mathbf{Y}$ into a bilinear form,

$$\mathbf{X}=\sum_{i=1}^{A}\mathbf{t}_i\mathbf{p}_i^T+\mathbf{E}\quad \mathbf{Y}=\sum_{i=1}^{A}\mathbf{u}_i\mathbf{q}_i^T+\mathbf{F} \tag{1}$$

where A represents the number of latent variables, $\mathbf{t}$ and $\mathbf{u}$ are the input and output score vectors, respectively. $\mathbf{p}$ and $\mathbf{q}$ represent input and output loading vectors, and $\mathbf{E}$, $\mathbf{F}$ are the residual error matrices of PLS model. The relationship between each vector is as follows

$$\mathbf{p}_i=\mathbf{t}_i^T\mathbf{X}/\mathbf{t}_i^T\mathbf{t}_i \tag{2}$$

$$\mathbf{t}_i=\mathbf{X}w_i \tag{3}$$

where w_i is the weight vector calculated by input and output vectors.

After extracting the input and output score matrices $\mathbf{T}=[\mathbf{t}_1,\mathbf{t}_2,\ldots,\mathbf{t}_A]$ and $\mathbf{U}=[\mathbf{u}_1,\mathbf{u}_2,\ldots\mathbf{u}_A]$, the inner relationship between each pair of scoring vectors is established

$$\mathbf{u}_i=f(\mathbf{t}_i)+\mathbf{e}_i \tag{4}$$

where$f(\cdot)$ is the inner mapping function, and $\mathbf{e}$ represents the residual. In linear PLS, the inner function $f(\cdot)$ is linear, that is, $f(\mathbf{t}_i)=c_\mathrm{i}\mathbf{t}_\mathrm{i}$, where c_i is a linear coefficient. If $f(\cdot)$ is a nonlinear function, the corresponding nonlinear PLS model is established.

B. INNER ELM MAPPING FUNCTION

Due to the nonlinear characteristics of lots of chemical and thermal processes, it is absolutely essential to apply nonlinear functions to map complex and strongly nonlinear inner relationship between latent variables extracted by outer PLS frame. In this study, in order to set up a nonlinear PLS model, ELM is used as the inner function. Different from other network models, which always cost a lot of computation

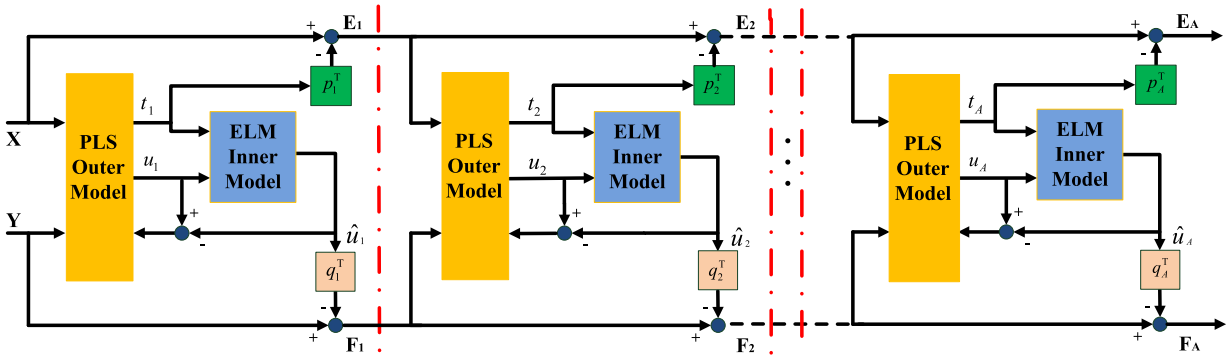

FIGURE 1. The graphic structure of the EBELMPLS method.

time to optimize a large number of parameters, ELM transforms complex network training process into a least squares problem to solve. As a result, this transformation reduces the computational complexity of the network.

Suppose a certain extracted pair of latent variables **t** and **u**, in which $\mathbf{t}=[t_1, t_2,\ldots,t_N]^{\mathrm{T}}$, $\mathbf{u}=[u_1, u_2,\ldots,u_N]^{\mathrm{T}}$, N represents the number of samples. Here, the number of hidden layer nodes of the model is l and activation function is $g(\cdot)$, then the output of the ELM model could be calculated by following form:

$$\sum_{i=1}^{l}\beta_i g_i(t_j)=\sum_{i=1}^{l}\beta_i g(w_i\cdot t_j+b_i)=\mathbf{o}_j \quad j=1,2,\ldots,N \tag{5}$$

where β_i denotes the output weights, which connects the ith hidden node with the output nodes. Simultaneously, w_i is the input weights vector, which connects the ith hidden node with the input nodes. b_i represents the bias of the ith hidden node. Previous studies have shown that the output value of ELM model can be fitted to samples with zero error [21]. So a derivation equation can be obtained as follows:

$$\sum_{j=1}^{l}\left\|\mathbf{o}_j-\mathbf{u}_j\right\|=0 \quad j=1,2,\ldots,N \tag{6}$$

Eq (5) can be transformed as follows:

$$\sum_{i=1}^{l}\beta_i g(w_i\cdot t_j+b_i)=\mathbf{u}_j \quad j=1,2,\ldots,N \tag{7}$$

Eq (7) can be simply written as:

$$\mathbf{H}\boldsymbol{\beta}=\mathbf{u} \tag{8}$$

where

$$\mathbf{H}=\begin{bmatrix} g(w_1\cdot t_1+b_1) & \cdots & g(w_l\cdot t_1+b_l) \\ \vdots & \ddots & \vdots \\ g(w_1\cdot t_N+b_1) & \cdots & g(w_l\cdot t_N+b_l) \end{bmatrix}_{N\times l} \tag{9}$$

$$\boldsymbol{\beta}=[\boldsymbol{\beta}_1,\ldots,\boldsymbol{\beta}_l]^T \quad \mathbf{u}=[u_1,\ldots,u_N]^T \tag{10}$$

In the form, **H** is called hidden layer output matrix. In the training process of ELM, the input weights w_i and bias values b_i of ELM are generated randomly. Then the output matrix H can be obtained, so that the ELM learning training problem is transformed into the least squares norm problem for solving the output weight, and that is:

$$\hat{\boldsymbol{\beta}}=\mathbf{H}^{+}\mathbf{u} \tag{11}$$

where $\mathbf{H}^{+}$ is the Moore-Penrose generalized inverse of hidden layer output matrix **H**. and the calculation results are

$$\hat{\boldsymbol{\beta}}=[\hat{\beta}_1,\ldots,\hat{\beta}_l] \tag{12}$$

Finally, the expression of inner ELM function can be written as

$$f(t)=\sum_{i=1}^{l}\hat{\beta}_i g(w_i\cdot t+b_i) \tag{13}$$

In this work, sigmoid function $g(x)=1/(1+\exp(-x))$ was chosen as activation function due to its good applications as activation in different kinds of ANNs. When substituting the activation function into eq 13, the detail expression of the inner ELM can be showed as follow

$$f(t)=\sum_{i=1}^{l}\hat{\beta}_i\frac{1}{(1+e^{-(w_i\cdot t+b_i)})} \tag{14}$$

III. THE PROPOSED EBELMPLS ALGORITHM

The novel nonlinear PLS is obtained based on combination of the external PLS framework and the internal ELM function. The external PLS is employed to extract feature information in input variables, reducing dimensions and collinearity of the input vectors, and internal ELM model is used to establish the nonlinear relationship in latent space. The overall structure of EBELMPLS is depicted in Figure 1.

As shown in figure 1, score vectors **t**and **u** are obtained by PLS outer model, and then each pair of score vectors are implemented by a SISO (single in single out) ELM as

$$\mathbf{u}=f(\mathbf{t})+\mathbf{e} \tag{15}$$

where $f(\cdot)$ can be calculated by the ELM function (eq 14). In this approach, the sample data is preprocessed through the external PLS model instead of being applied directly to train ELM model, which transforms the original multiple regression issue into a series of single variable regression tasks, and makes ELM model simplified.

The error-based input weight updating procedure can be performed when nonlinear function used to fit the inner relationship between the input and the output latent variables is continuous and differentiable with respect to the input weights $\mathbf{w}$. The sigmoid function, the activation function of ELM, satisfies the given condition. So the weight updating procedure can be implemented at each of iteration within the NIPALS method [22].

For eq 15, the relationship between output and input latent variables can be approximated by Newton-Raphson linearization approach,

$$\mathbf{u} = \mathbf{f}_{00} + \partial \mathbf{f}(\mathbf{t})/\partial \mathbf{w}^T \mid_{\mathbf{w}_{00}} \cdot \Delta \mathbf{w} \tag{16}$$

where

$$\mathbf{f}_{00} = \hat{\mathbf{u}} = \mathbf{f}(\mathbf{t}) \tag{17}$$

Here, $\mathbf{f}_{00}$ is the estimated value of output latent score and can be obtained using ELM function. And $\mathbf{w}_{00}$ donates the current value of $\mathbf{w}$. In point of fact, the first derivative term in the above formula can be written as

$$\begin{aligned}\frac{\partial \mathbf{f}(\mathbf{t})}{\partial \mathbf{w}^T} &= \frac{\partial \mathbf{f}(\mathbf{t})}{\partial \mathbf{t}^T} \cdot \frac{\partial \mathbf{t}}{\partial \mathbf{w}^T} = \frac{\partial \mathbf{f}(\mathbf{t})}{\partial \mathbf{t}^T} \cdot \frac{\partial (\mathbf{X}\mathbf{w})}{\partial \mathbf{w}^T} = \frac{\partial \mathbf{f}(\mathbf{t})}{\partial \mathbf{t}^T} \cdot \mathbf{X} \\ &= \begin{bmatrix} \frac{\partial f(t_1)}{\partial t_1} & \frac{\partial f(t_1)}{\partial t_2} & \cdots & \frac{\partial f(t_1)}{\partial t_N} \\ \frac{\partial f(t_2)}{\partial t_1} & \frac{\partial f(t_2)}{\partial t_2} & \cdots & \frac{\partial f(t_2)}{\partial t_N} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial f(t_N)}{\partial t_1} & \frac{\partial f(t_N)}{\partial t_2} & \cdots & \frac{\partial f(t_N)}{\partial t_N} \end{bmatrix} \cdot \mathbf{X} \\ &= \begin{bmatrix} \frac{\partial f(t_1)}{\partial t_1} & 0 & \cdots & 0 \\ 0 & \frac{\partial f(t_2)}{\partial t_2} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \frac{\partial f(t_N)}{\partial t_N} \end{bmatrix} \cdot \mathbf{X} \end{aligned} \tag{18}$$

where $f'(t)$ is derivative of ELM activation function (eq 14) as follow

$$f'(t) = \sum_{i=1}^{l} \hat{\boldsymbol{\beta}}_i w_i \frac{e^{-(w_i \cdot t + b_i)}}{(1 + e^{-(w_i \cdot t + b_i)})^2} \tag{19}$$

when we define $\partial \mathbf{f}(\mathbf{t})/\partial \mathbf{w}^{\mathrm{T}}$ as matrix $\mathbf{D}$,error matrix can be written as

$$\mathbf{e} = \mathbf{u} - \hat{\mathbf{u}} = \mathbf{D} \cdot \Delta \mathbf{w} \tag{20}$$

The increments $\Delta \mathbf{w}$ can be regressed directly by least-squares method as follows

$$\Delta \mathbf{w} = (\mathbf{D}^T \mathbf{D})^{-1} \mathbf{D}^T \mathbf{e} \tag{21}$$

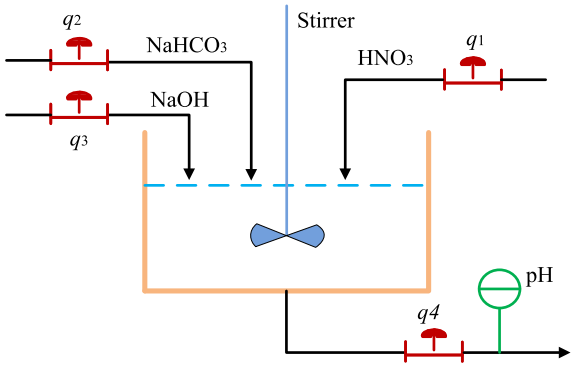

FIGURE 2. The pH-neutralization process.

A detailed description of the modified NIPALS algorithm steps for EBMELMPLS training procedures is shown in Table 1.

After the EMELMPLS training procedures, the vectors $\mathbf{p}$ and $\mathbf{q}$, input weight $\mathbf{w}$, and estimated value $\hat{\mathbf{u}}$ can be obtained in matrices called by $\mathbf{P}$, $\mathbf{Q}$, $\mathbf{W}$ and $\hat{\mathbf{U}}$. Besides, ELM inner mapping function trained through above training procedures can be directly used to predict new data.

For the new input data matrix $\mathbf{X}_t$, the new input score matrix $\mathbf{T}$ can be calculated as follow

$$\mathbf{T} = \mathbf{X}_t \mathbf{W}(\mathbf{P}\mathbf{W})^{-1} \tag{22}$$

where $\mathbf{W}$=[$\mathbf{w}_1 \ldots \mathbf{w}_A$], $\mathbf{P}$=[$\mathbf{p}_1 \ldots \mathbf{p}_A$], and A donates the number of required latent variables. So the new estimated output score vector $\hat{\mathbf{U}}$ ($\hat{\mathbf{U}} = [\hat{\mathbf{u}}_i \ldots \hat{\mathbf{u}}_A]$) of $\mathbf{X}_t$ can be calculated using trained ELM function: $\hat{\mathbf{u}}_i = \mathbf{f}_i(\mathbf{t}_i)$. Finally, the estimated output $\hat{\mathbf{Y}}_t$ is calculated as follow formulation

$$\hat{\mathbf{Y}}_t = \hat{\mathbf{U}} \hat{Q}^T \tag{23}$$

IV. SIMULATIONS

A. BENCHMARK DATA SET TEST

PH-neutralization system, described by Henson and Seborg, is well known to be highly non-linear and has been widely used as a benchmark to test modeling or control algorithms. Figure 2 shows a flow sheet of the pH-neutralization systems process; the neutralization reaction of a strong acid (HNO_3, q_1), a buffer stream ($NaHCO_3$, q_2) and a strong base (NaOH, q_3) in the reaction tank. The pH value of the output stream is measurable and determined by the flow rate of inlet ($q_1 \sim q_3$) and outlet (q_4). In order to keep a constant level in the tank, the flow rate of outlet stream varies with the inlets streams flow rate. A data set of 81 samples points was obtained under the conditions of different inlet flow rate to verify the relationship between the pH and the liquid flow $q_1 \sim q_4$. The sample data was divided into two subsets, a testing data set containing 16 observations randomly selected and a training data comprising remaining observations. The "pH-neutralization" data can be found at https://figshare.com/s/f267027476f6d54ffd11.

In order to get a creditable conclusion, the sample data set was used to set up model by four different PLS methods:

TABLE 1. Modified nipals algorithm steps for EBMELMPLS training procedures.

step	Summary of steps	
1	Mean centre and scale **X** and **Y**	
2	Select the output score **u** from a column of **Y**	
3	Calculate input weights w using **u** and **X**	$\mathbf{w}^T = \mathbf{u}^T/(\mathbf{u}^T\mathbf{u})$
4	Normalize **w** to unit length	$\mathbf{w}=\mathbf{w}/\|\|\mathbf{w}\|\|$
5	Compute the input score **t**	$\mathbf{t}=\mathbf{X}\mathbf{w}$
6	Train the inner ELM model , which predict s **u** with **w**, and obtain the estimated value $\hat{\mathbf{u}}$	$\hat{\mathbf{u}} = \mathbf{f}(\mathbf{t})$
7	Compute output loading **q**	$\mathbf{q}^T = \hat{\mathbf{u}}^T\mathbf{Y} / (\hat{\mathbf{u}}^T\hat{\mathbf{u}})$
8	Normalize **q** to unit length	$\mathbf{q}=\mathbf{q}/\|\|\mathbf{q}\|\|$
9	Calculate the new output score **u**	$\mathbf{u}=\mathbf{Y}\mathbf{q}$
10	Compute the input weights updating parameters **w**	$\Delta\mathbf{w} = (\mathbf{D}^T\mathbf{D})^{-1}\mathbf{D}^T\mathbf{e}$
11	Compute new input weights **w** and normalize w to unit length	$\mathbf{w}= \mathbf{w}+\Delta\mathbf{w}$ $\mathbf{w}=\mathbf{w}/\|\|\mathbf{w}\|\|$
12	Calculate the new input score **t**	$\mathbf{t}=\mathbf{X}\mathbf{w}$
13	Check convergence on **t;** if **t** is convergent, go to 14, otherwise go to 6	
14	Compute input loading **p**	$\mathbf{p}^T = \mathbf{t}^T\mathbf{X}/(\mathbf{t}^T\mathbf{t})$
15	Calculate new nonlinear prediction **u** using new input score **t**	$\hat{\mathbf{u}} = \mathbf{f}(\mathbf{t})$
16	Calculate input and output residual matrices **E** and **F**	$\mathbf{E}=\mathbf{X}-\mathbf{t}\mathbf{p}^T$ $\mathbf{F}=\mathbf{Y}-\hat{\mathbf{u}}\mathbf{q}^T$
17	Replace **X** and **Y** by **E** and **F**, and repeat steps 1-16 until the accuracy of result meets the requirement	

linear PLS (LPLS), error-based RBF neural network PLS (EBRBFPLS), ELMPLS and EBELMPLS. The same external PLS framework was employed for four methods, while four kinds of inner mapping function were implemented for each methods. EBRBFPLS applied a three-layer RBF neural network, which was established with Gaussian function and could be trained by neural network toolbox in MATLAB 2016. ELMPLS and EBELMPLS both used the same ELM model as inner function, but the difference was that EBELMPLS method has the error-based weight updating procedure. Considering that the initial weights and bias values of ELM model can affect the performance of the EBRBFPLS, ELMPLS and EBELMPLS, 30 experiments were performed.

B. RESULTS AND DISCUSSION

In order to accurately evaluate the performance of four PLS methods, root mean square error (RMSE) and mean relative error (MRE) are used as evaluation criteria.

$$\mathrm{RMSE} = \sqrt{[\sum_{i=1}^{n}(\hat{y}_i - y_i)^2]/n} \tag{24}$$

$$\mathrm{MRE} = \frac{1}{n}\sum_{i=1}^{n}\left|\frac{\hat{y}_i - y_i}{y_i}\right| \tag{25}$$

where n represents the number of samples, y_i and $\hat{y}_i$ are real and the corresponding predicted values, respectively. $\mathrm{RMSE_T}$, $\mathrm{MRE_T}$, $\mathrm{RMSE_P}$ and $\mathrm{MRE_P}$ are the criteria values of training set and testing set, respectively. In addition, the variance explanation rate of input variables and output ($R^2\mathbf{X}$ and $R^2\mathbf{Y}$) by extracted latent variables is an important criterion, and that is written as [23]

$$\mathrm{R^2X} = 1 - \mathrm{SS}(\mathbf{E})/\mathrm{SS}(\mathbf{X}) \quad \mathrm{R^2Y} = 1 - \mathrm{SS}(\mathbf{F})/\mathrm{SS}(\mathbf{Y}) \tag{26}$$

where **E** is the residual matrix of input matrix **X**, **F** is the residual matrix of output matrix **Y**, and $\mathrm{SS}(\cdot)$ donates the sum of squares.

In this work, q_1, q_2, q_3 and q_4 were employed as inputs of the pH-neutralization model, and the only output was the pH value. The results of four models are displayed in Table 2. Furthermore, the scatter fitting diagram for inner first latent variable pair of each approach is represented in Figure 3. Note the criteria of the last three methods in Table 2 are the average of the 30 experimental results.

As we can see from Table 2, only 56.35% of variance is interpreted through LPLS approach, and Figure 3a also reveals that LPLS is a linear regression method and it is unable to describe the internal nonlinear relationship between latent variable pair $\mathbf{t}_1$ and $\mathbf{u}_1$. As shown in Figure 3b,d, the performance of the EBRBFPLS and EBELMPLS methods is obviously superior to the LPLS method. The reason for such a result is that the neural network and ELM are capable of fitting nonlinear relations. Two methods can interpret 93.42% and 94.86% of the variance, respectively. What's more, because the error-based weight updating procedure is added, the distribution of the real variables in Figure 3b,d are more compact than that of in Figure 3a,c. However, the ELMPLS method, without weight update, explains 77.14% of variance. This result indicates that weight updating operation

TABLE 2. Comparison of four kinds of methods for PH-neutralization process.

method	A	$RMSE_T$	$MRE_T(\%)$	$RMSE_P$	$MRE_P(\%)$	$R^2X(\%)$	$R^2Y(\%)$
LPLS	3	1.0768	12.67	1.1994	13.17	74.81	56.35
EBRBFPLS	2	0.5738	6.57	0.5940	7.06	66.52	93.42
ELMPLS	3	0.6143	7.59	0.6835	7.97	75.76	77.14
EBELMPLS	2	0.4096	4.22	0.4681	4.89	63.94	94.86

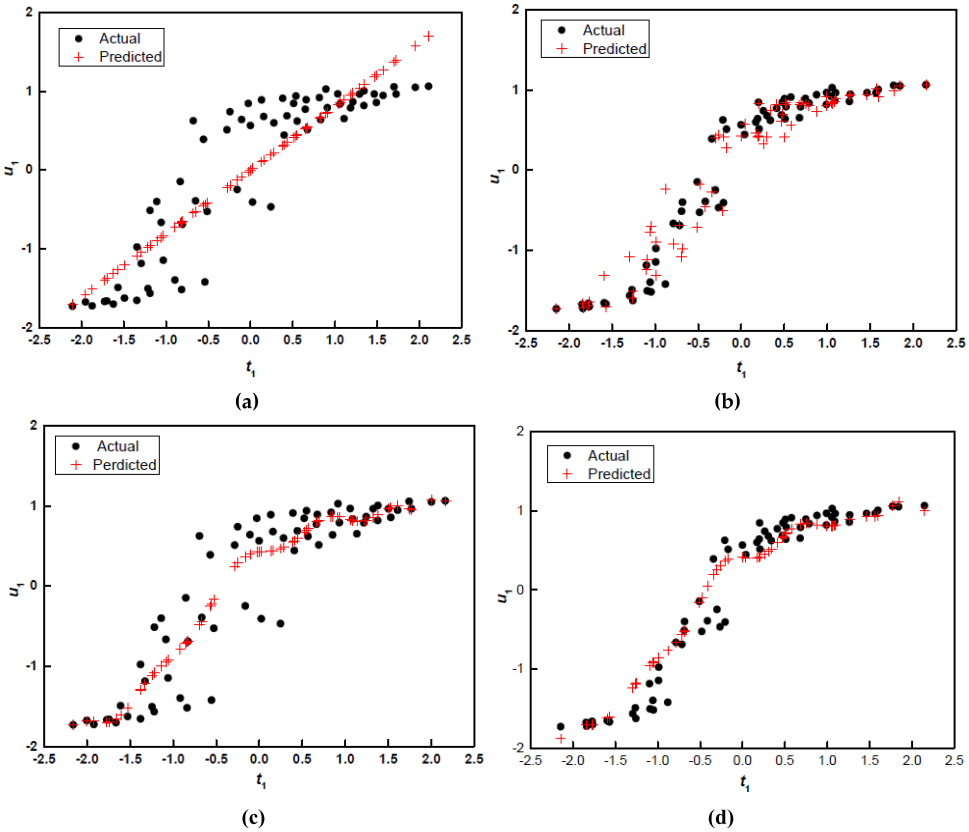

FIGURE 3. The scatter fitting plot for inner first latent variable pair. (a) LPLS. (b) EBRBFPLS. (c) ELMPLS. (d) EBELMPLS.

is advantageous to improve the prediction accuracy of the model.

Moreover, instead of the numerical value of variance captured, RMSE and MRE values of training and testing dataset of four methods are also listed in Table 2. For the linear PLS approach, the RMSE value of LPLS model is 1.0768 for the training dataset and 1.1994 for the testing dataset. The EBRBFPLS method displays an obvious betterment, with $RMSE_T = 0.5738$ and $RMSE_P = 0.5940$. Additionally, the EBELMPLS approach performs the best fitting and predicting ability among the four methods, with $RMSE_T = 0.4096$, $MRE_T = 4.22\%$ and $RMSE_P = 0.4681$, $MRE_P = 4.89\%$. The ELMPLS method, by contrast, shows much poor prediction accuracy, with $RMSE_P = 0.6835$, $MRE_P = 7.97\%$. This validates again that weight update operation is benefit to improving the accuracy. The modeling time of the algorithm can well reflect the complexity of the algorithm. In this simulation, the average modeling time of LPLS, ELMPLS and EBELMPLS is not more than 2s and the average modeling time of EBRBFPLS is more than 40s. This is because the parameters of EBRBFPLS need more time to be adjusted in the training process of the model. The results show that EBELMPLS algorithm has high modeling efficiency.

V. APPLICATION TO NO_x EMISSIONS

In recent years, reducing Nitrogen oxide (NO_x) emissions has become a global concern. NO_x is one of the main pollutants from coal-fired power plants, and is also a contributor to global warming [24]. Combustion optimization technology has been demonstrated as an effective and economic method to the reduction of NO_x emissions from coal-fired boilers by tuning adjustable parameters of a boiler, such as coal feed quantity, excess air ratio, and secondary air etc [25]. Such optimization of operational parameters relies heavily on an

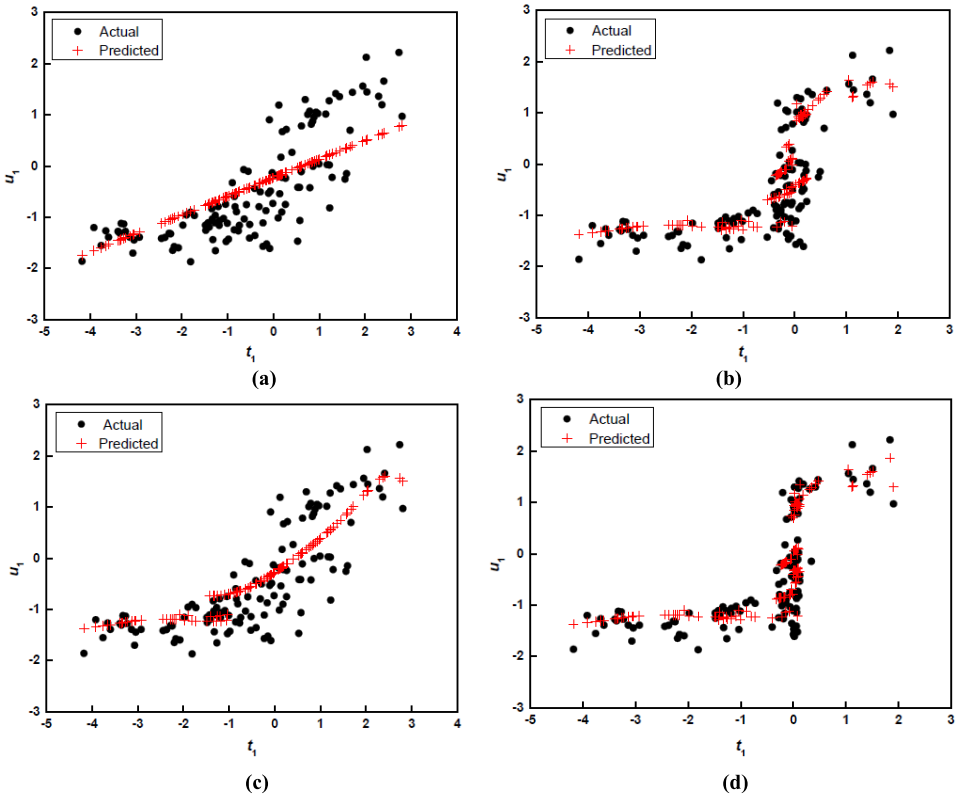

FIGURE 4. Score scatter diagram plot for inner first latent variable pair. (a) LPLS. (b) EBRBFPLS. (c) ELMPLS. (d) EBELMPLS.

accurate correlation between NO_x emissions and operational parameters. However, the generation mechanism of NO_x is complex and there is a serious coupling in each variable of a boiler in power plants, so it is difficult to set up an accurate mechanism model for such a complex system.

In this section, the presented EBELMPLS approach is used to establish the model of NO_x emission for a 1,000 MW ultra-supercritical variable pressure once-through boiler. This boiler is 65.5m in height and has a cross section 32.08m×15.67m. The boiler adopts the ∏-shaped arrangement, the improved low NO_x PM (pollution minimum) main burner, the MACT (Mitsubishi advanced combustion technology) type low NO_x grading air supply combustion system and the reverse double tangent combustion mode. There are forty-eight pulverized coal burners in the corners and around middle of the walls at six different elevations. The boiler is configured with six straight blow medium-speed coal mills and the maximum continuous evaporation of the boiler is 2980 t/h. In this work, 150 patterns were obtained from SIS (supervisory information system) of the power plant at typical steady loads of 700, 850 and 1000MW. The "data of modeling NO_x emissions" can be found at https://figshare.com/s/95eed79776755261a566. The sampling interval was 1 min [26]. Based on the basic knowledge of boilers and the engineers' experience, 23 parameters affecting the NO_x emissions were employed as inputs of prediction model: the boiler load, total secondary air flow rate, total air flow rate, two opening values of the OFA (over fire air) air damper, six pulverized coal flow rates, six primary air flow rates of coal pulverized, and six opening values of the secondary air dampers. In addition, the NO_x emissions data obtained in this study is represented on dry gas basis at 6% O_2.

TABLE 3. Comparison of four kinds of methods for no_x emission.

method	$RMSE_T$	MRE_T(%)	$RMSE_P$	MRE_P(%)
LPLS	16.9825	3.83	24.9043	5.01
EBRBFPLS	7.4129	1.47	10.6751	2.00
ELMPLS	11.9296	2.30	18.7661	3.56
EBELMPLS	5.9541	1.23	6.8323	1.32

The total of 150 cases was divided into two parts: one group of 120 samples for training model and the other part of 30 samples for testing model. Similar to the preceding pH-neutralization experiment, the LPLS, EBRBFPLS, ELMPLS and EBELMPLS were implemented to set up the NO_x emission model of the boiler with these samples. Table 3 represents the simulation results for four methods. Note the criteria of the last three methods in Table 3 are the average of the 30 experimental results.

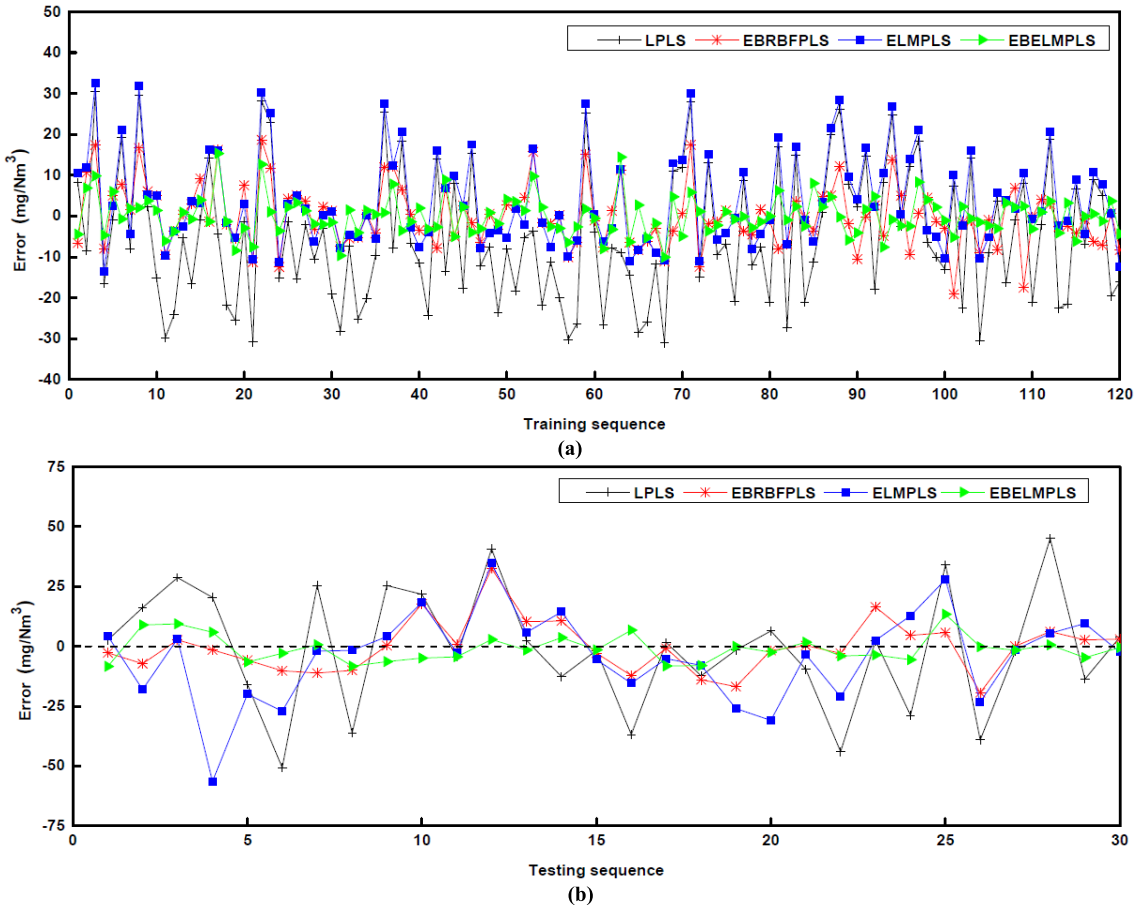

FIGURE 5. Prediction errors of NO_x emissions for different models: (a) prediction errors of training dataset and (b) prediction errors of testing dataset.

The conclusions obtained from Table 3 are similar to those of the previous pH-neutralization case. As expected, the LPLS approach exhibits the worst approximating and predicted abilities, with $RMSE_T = 16.9825$, $MRE_T = 3.83\%$, $RMSE_P = 24.9034$ and $MRE_P = 5.01\%$. ELMPLS model is observed to perform better than LPLS model, but it still seems to have inadequate ability to handle the nonlinearity problem. The EBRBFPLS method results in a better RMSE and MRE for both training set and testing set than LPLS and ELMPLS methods. Simultaneously, the EBELMPLS performs the best among the four methods, and it produces a RMSE of 5.9541 for the training dataset and 6.8323 for the testing set; in addition, the corresponding MRE value are 1.23% and 1.32%, respectively.

Figure 4 gives the fitting diagram of inner first latent variable pair of each approaches and Figure 5 shows the errors between the actual value and predicted value of NO_x emissions with four modeling methods. Note that the data of the EBRBFPLS model, the ELMPLS model, and the EBELMPLS model in Figures 4 and 5 are calculated with one of the 30 experiments in which the criteria is near the average criteria. From the Figure 4 one can see that the EBELMPLS exhibits the best fitting accuracy, with more compact distribution of the real variable.

The prediction results of four models NO_x emissions prediction model for training dataset are shown in Figure 5a, and it can be seen that LPLS method's approximation capability is poor, with large range from -30 to 30. The ELMPLS prediction error is smaller than LPLS, but its prediction error is still relatively large. So it can be concluded that LPLS and ELMPLS approaches do not have enough ability to fit the training dataset adequately. Besides, EBRBFPLS and EBELMPLS display strong fitting ability with small predicting errors.

Figure 5b shows the estimation error of the testing dataset for each of the four models. It can be clearly seen that the prediction error curve of EBELMPLS model is the most stable and the minimum amplitude of fluctuation is the smallest. Because the four models use the same training and testing datasets, the predicted results on testing dataset are reliable in proving the models' generalization and prediction accuracy capability. The LPLS model is observed to have the worst performance among the four models. The ELMPLS and EBRBFPLS perform better than LPLS, but are not as good as EBELMPLS model. Therefore, it can be concluded that EBELMPLS model is more precise and robust than three other models for modeling NO_x emission of the coal-fired boiler.

VI. CONCLUSION

In this work, a novel nonlinear partial least square method was proposed for establishing industrial process models. The new algorithm is composed of external feature extraction and internal nonlinear mapping. The PLS framework is used to extract the principal element of input and output variables and ELM function is applied to map the inner nonlinear relationship. An error-based input weight updating was also proposed to reduce the modeling prediction errors. Then pH-neutralization process is taken as a benchmark to verify the performance of the model and comparing it with other commonly used PLS methods. The results showed that EBELMPLS model provided the best prediction performance among four methods. Then, the method proposed in this article was used to model the NO_x emission of a 1000MW coal-fired boiler. The RMSEs of EBELMPLS were 5.9541 for training dataset and 6.8323 for the testing dataset. Compared with the LPLS, EBRBFPLS and ELMPLS models, the prediction errors were reduced by approximately 72.5%, 35.9% and 63.5%, respectively. The comparison results demonstrate that EBELMPLS model has relatively high precision prediction performance and it can be an alternative when designing a software package to control NO_x emissions from utility boilers. In the future research work, the other kinds internal functions will be studied.

APPENDIX

See Table 4.

TABLE 4. Key parameters list.

$\mathbf{X}$: input data matrix	$\mathbf{Y}$: output data matrix
$\mathbf{E}$: input residual error matrix	$\mathbf{F}$: output residual error matrix
$\mathbf{t}$: input score vector	$\mathbf{u}$: output score vector
$\mathbf{T}$: input score matrix	$\mathbf{U}$: output score matrix
$\mathbf{p}$: input loading vector	$\mathbf{q}$: output loading vector
$\mathbf{P}$: input loading matrix	$\mathbf{Q}$: output loading matrix
$\hat{\mathbf{u}}$: estimated output score vector	$\hat{\mathbf{U}}$: estimated output score matrix
$\mathbf{w}$: weight vector	$\mathbf{W}$: weight matrix
$\mathbf{e}$: residual error vector	A : number of latent variables
l : number of hidden layer nodes of ELM	
w: input weight of ELM	
b: the bias of the hidden node of ELM	
$\mathbf{H}$: hidden layer output matrix of ELM	
$\boldsymbol{\beta}$: output weight matrix of ELM	

REFERENCES

[1] L. Wang, H. Jin, X. Chen, J. Dai, K. Yang, and D. Zhang, "Soft sensor development based on the hierarchical ensemble of Gaussian process regression models for nonlinear and non-Gaussian chemical processes," *Ind. Eng. Chem. Res.*, vol. 55, no. 8, pp. 7704–7719, 2016.

[2] X. Yuan, Z. Ge, and Z. Song, "Locally weighted kernel principal component regression model for soft sensing of nonlinear time-variant processes," *Ind. Eng. Chem. Res.*, vol. 53, no. 35, pp. 13736–13749, 2014.

[3] M. Kano and Y. Nakagawa, "Data-based process monitoring, process control, and quality improvement: Recent developments and applications in steel industry," *Comput. Chem. Eng.*, vol. 32, nos. 1–2, pp. 12–24, 2008.

[4] G.-Q. Li, X.-B. Qi, K. C. C. Chan, and B. Chen, "Deep bidirectional learning machine for predicting NO_x emissions and boiler efficiency from a coal-fired boiler," *Energy Fuels*, vol. 31, no. 10, pp. 11471–11480, 2017.

[5] J. Liu, D.-S. Chen, and J.-F. Shen, "Development of self-validating soft sensors using fast moving window partial least squares," *Ind. Eng. Chem. Res.*, vol. 49, no. 22, pp. 11530–11546, 2010.

[6] P. Kadlec, B. Gabrys, and S. Strandt, "Data-driven soft sensors in the process industry," *Comput. Chem. Eng.*, vol. 33, no. 4, pp. 795–814, Apr. 2009.

[7] J. Liu, "On-line soft sensor for polyethylene process with multiple production grades," *Control Eng. Pract.*, vol. 15, no. 7, pp. 769–778, 2007.

[8] K. Kim, J.-M. Lee, and I.-B. Lee, "A novel multivariate regression approach based on kernel partial least squares with orthogonal signal correction," *Chemometrics Intell. Lab. Syst.*, vol. 79, nos. 1–2, pp. 22–30, 2005.

[9] D. C. Ronen, F. W. Sanders, H. S. Tan, P. R. Mort, and F. J. Doyle, III, "Predictive dynamic modeling of key process variables in granulation processes using partial least squares approach," *Ind. Eng. Chem. Res.*, vol. 50, no. 3, pp. 1419–1426, 2011.

[10] S. Wold, N. Kettaneh-Wold, and B. Skagerberg, "Nonlinear PLS modeling," *Chemometrics Intell. Lab. Syst.*, vol. 7, nos. 1–2, pp. 53–65, 1989.

[11] Q. Jia and Y. Zhang, "Quality-related fault detection approach based on dynamic kernel partial least squares," *Chem. Eng. Res. Des.*, vol. 106, pp. 242–252, Feb. 2016.

[12] S. J. Qin and T. J. McAvoy, "Nonlinear PLS modeling using neural networks," *Comput. Chem. Eng.*, vol. 16, no. 4, pp. 379–391, 1992.

[13] B. Li, P. A. Hassel, A. J. Morris, and E. B. Martin, "A non-linear nested partial least-squares algorithm," *Comput. Stat. Data Anal.*, vol. 48, no. 1, pp. 87–101, 2005.

[14] Y. Lv, J. Liu, and T. Yang, "Nonlinear PLS integrated with error-based LSSVM and its application to NO_x modeling," *Ind. Eng. Chem. Res.*, vol. 51, no. 49, pp. 16092–16100, 2012.

[15] H. Cao, X. Yan, S. Yang, H. Ren, and S. S. Ge, "Low-cost pyrometry system with nonlinear multisense partial least squares," *IEEE Trans. Syst., Man, Cybern. Syst.*, vol. 48, no. 7, pp. 1029–1038, Jul. 2018.

[16] U. Kruger, Y. X. Zhou Wang, D. Rooney, and J. Thompson, "Robust partial least squares regression: Part I, algorithmic developments," *J. Chemometrics*, vol. 22, no. 1, pp. 1–13, 2008.

[17] Y.-P. Zhou *et al.*, "Artificial neural network-based transformation for nonlinear partial least-square regression with application to QSAR studies," *Talanta*, vol. 71, no. 2, pp. 848–853, 2007.

[18] A. I. Abdel-Rahman and G. J. Lim, "A nonlinear partial least squares algorithm using quadratic fuzzy inference system," *J. Chemometrics*, vol. 23, no. 10, pp. 530–537, 2009.

[19] G. B. Huang, Q. Y. Zhu, and C. K. Siew, "Extreme learning machine: Theory and applications," *Neurocomputing*, vol. 70, no. 1, pp. 489–501, 2006.

[20] S. Li, L. Goel, and P. Wang, "An ensemble approach for short-term load forecasting by extreme learning machine," *Appl. Energy*, vol. 170, pp. 22–29, May 2016.

[21] G.-B. Huang, "Learning capability and storage capacity of two-hidden-layer feedforward networks," *IEEE Trans. Neural Netw.*, vol. 14, no. 2, pp. 274–281, Mar. 2003.

[22] G. Baffi, E. B. Martin, and A. J. Morris, "Non-linear projection to latent structures revisited: The quadratic PLS algorithm," *Comput. Chem. Eng.*, vol. 23, no. 9, pp. 395–411, 1999.

[23] S. Wold, J. Trygg, A. Berglund, and H. Antti, "Some recent developments in PLS modeling," *Chemometrics Intell. Lab. Syst.*, vol. 58, no. 2, pp. 131–150, 2001.

[24] A. Molina, E. G. Eddings, D. W. Pershing, and A. F. Sarofim, "Char nitrogen conversion: Implications to emissions from coal-fired utility boilers," *Prog. Energy Combustion Sci.*, vol. 26, nos. 4–6, pp. 507–531, 2000.

[25] Z. Luo, F. Wang, H. Zhou, R. Liu, W. Li, and G. Chang, "Principles of optimization of combustion by radiant energy signal and its application in a 660 MWe down- and coal-fired boiler," *Korean J. Chem. Eng.*, vol. 28, no. 12, pp. 2336–2343, 2011.

[26] J. Smrekar, P. Potočnik, and A. Sengacnik, "Multi-step-ahead prediction of NO_x emissions for a coal-based boiler," *Appl. Energy.*, vol. 106, no. 11, pp. 89–99, 2013.

ZE DONG was born in Baoding, Hebei, China, in 1970. He received the B.S., M.S., and Ph.D. degrees in thermal engineering from North China Electric Power University, in 1992, 1995, and 2001, respectively.

From 1995 to 2003, he was a Lecturer with the Department of Dynamics, North China Electric Power University. Since 2009, he has been a Professor with the Automation Department, North China Electric Power University. He is currently a Ph.D. Supervisor. He is also currently the Director of the Hebei Power Generation Process Simulation and Optimal Control Engineering Research Center and the Director of the Simulation Application Branch of China Computer Users Association. He has authored four books, more than 80 articles. His research interests include the modeling and control of complex systems.

NING MA received the B.S. degree in engineering from Xiangtan University, China, in 2015, and the master's and Ph.D. degrees in advance from North China Electric Power University, China, where he is currently pursuing the Ph.D. degree in control engineering. His main research interest includes the application of machine learning in thermal process modeling and control.