

# Simultaneous Identification of Process Structure, Parameters and Time-delay Based on Non-negative Garrote

Jian-Guo Wang<sup>1</sup> Qian-Ping Xiao<sup>1</sup> Tiao Shen<sup>1</sup> Shi-Wei Ma<sup>1</sup> Wen-Tao Rao<sup>2</sup> Yong-Jie Zhang<sup>2</sup>

<sup>1</sup>School of Mechatronic Engineering and Automation, Shanghai University, Shanghai Key Lab of Power Station Automation Technology, Shanghai 200072, China

<sup>2</sup>Institute of Environment and Resources, Baosteel Research Institute, Shanghai 201900, China

**Abstract:** In practice, the model structure, parameters and time-delay of the actual process may vary simultaneously. However, the general identification methods of the 3 items are performed with separate procedures which is very inconvenient in practical application. In view of the fact that variable selection procedure can ensure a compact model with robust input-output relation and in order to explore the feasibility of variable selection algorithm for the simultaneous identification of process structure, parameters and time-delay, non-negative garrote (NNG) algorithm is introduced and applied to system identification and the corresponding procedures are presented. The application of NNG variable selection algorithm to the identification of single input single output (SISO) system, multiple input multiple output (MIMO) system and Wood-Berry tower industry are investigated. The identification accuracy and the time-series variable selection results are analyzed and compared between NNG and ordinary least square (OLS) algorithms. The derived excellent results show that the proposed NNG-based modeling algorithm can be utilized for simultaneous identification of the model structure, parameters and time-delay with high precision.

**Keywords:** Model structure, model parameter, time-delay, non-negative garrote, variable selection.

## 1 Introduction

System identification is the foundation of model predictive control and control performance assessment. Many critical real-time process variables with time-series are contained in many industry processes. It is better to use identified model to predict the key response variables that are difficult to be measured with some existed sensor variables. This model is named inferential sensor model or soft-sensor model<sup>[1]</sup>. The application and many kinds of soft-sensor models or identification methods in different fields were studied in [2, 3]. The development of soft sensor faces many challenges. For example, the relationship between input and output may be nonlinear or vary from time to time when process operating conditions are changing<sup>[4]</sup>. So, non-parametric model, for example, neural network or multiple local model network, is used to deal with nonlinear problems generally<sup>[5]</sup>. The problems of over-fitting and vagueness of interpretation occur in the black box model inevitably. In order to deal with the time-variant characteristics in the industry process, adaptive model and real-time model are used usually. Many sensors are used in the modern industry process. But, the availability of the data does not mean that one can get useful information success-

fully. The method of linear multivariate statistical analysis, such as, principle component regression (PCR), partial least squares (PLS)<sup>[6]</sup> and canonical variables analysis (CVA), are used to overcome collinearity of high-dimensional and candidate predictor variable problems in the development of soft-sensor process<sup>[7]</sup>. Those methods try to reduce some latent variables of sensors. In the PCR algorithm, the latent variables represent the main direction of variation information in the sensor variables. However, the quality of prediction result has nothing to do with the latent variables. The prediction will be successful only if the eigenvalues of latent variables are very small. The aim of PLS is to find the direction of max covariance between sensor variables and response variables. The recursive PLS algorithm was proposed by Qin<sup>[8]</sup>. The study proved that recursive PLS algorithm is a useful process monitoring tool. CVA algorithm maximizes the correlation between input and output, however it cannot ensure the coefficient of irrelevant variable to be zero. Although the PCR/PLS/CVA algorithms can shrink sensor variables into several key latent variables, many contributing information from predictor variables are contained. This is not an ideal situation, because the operator cannot understand the physical meaning between input and output and focus on the critical factors of controllable variables. An important characteristic of model identification is the appropriate variable selection. It increases the transparency and robustness of the model. The process variables are subjected to simple statistical analyses to identify a subset of measurements to be used in the inferential

Research Article  
Manuscript received September 29, 2014; accepted March 30, 2015;  
published online June 20, 2016  
Recommended by Guest Editor Yang Song  
© Institute of Automation, Chinese Academy of Sciences and  
Springer-Verlag Berlin Heidelberg 2016

scheme. This can help us understand the physical meaning of input and output and focus on analyzing the main factor. It is important to analyze and decide what makes the system performance worse<sup>[9]</sup>. Recently, the tightening method based on shrinking or setting coefficients of a greedy model to has drawn a lot of attention. The two popular methods are non-negative garrote (NNG)<sup>[10]</sup> and least absolute shrinkage and selection operator (LASSO)<sup>[11]</sup>. The NNG method has two stages which are shrinkage amplitude and the parameter estimation of regression model with ordinary least square (OLS). LASSO and its mutation algorithm can calculate the parameter and shrink amplitude at the same time, but one must make sure that the objective function is a piecewise continuous function. The advantage of these methods is that the irrelevant variable will be shrunk to zero. Meanwhile, this method and other similar methods are very popular in the field of biological data exploration and machine learning<sup>[12]</sup>. However, they are not widely used in the modeling of industry process. Pan et al.<sup>[5]</sup> put forward an improved partial least square NNG algorithm, in order to establish a static input and output model of final quality prediction during the semiconductor manufacturing process. The classical variable selection methods are based on Cp criteria (CP), Akaike information criterion (AIC), the Bayes information criterion (BIC), etc. However, these methods are computationally infeasible for even moderate numbers of predictors. This type of method is implemented in stepwise fashion through forward selection or backward elimination<sup>[13]</sup>. The implementations are known to be sub-optimal in many applications because of the myopic nature of the stepwise. However, NNG algorithm can select the variable via shrinking or setting greedy model parameter to be zero. NNG method can make up for the deficiency of the traditional variable selection. It is proved that NNG method is path consistent in the sense of probability and the coefficient estimation or variable selection has a consistency with actual process model at least. Many existing process identification methods suppose that the model structure is known. In fact, this is not true in most cases. The model order is determined by system input and output data, when there is no prior knowledge of model structure. The accuracy of model cannot be met if the model order is too low. However, the model will be too complicated with a high model order. There existed some specific methods to determine model order, however, its separated from parameter identification<sup>[14, 15]</sup>. Furthermore, the time delay or interaction matrix identification plays an important role in determination of performance benchmarks for control system performance evaluation/monitoring. The NNG variable selection algorithm is used as a new method to identify the open loop process parameters. The identification result and the model order can be obtained simultaneously. The delay time of the model or interactor matrix can be achieved as a bonus. The details are as follows: we can construct a high order (higher than the real order) model during the identification, that is construct enough time series as alternative variables, then select the vital variables and the correspond-

ing model order with the usage of NNG algorithm. This identification method can help us understand the physical meaning of the input and output by analyzing the key factors. Furthermore, it is helpful to figure out the causes of affecting the performance. For the variable selection problem in the linear regression model, the requirements are as follows: the independent variables contained in the regression equation are as little as possible; small variance and less mean square error of regression coefficient and small mean square error of prediction value<sup>[10]</sup>. Therefore, the accuracy of prediction and interpretability are two important indicators of the estimation of regression model. The final goal of variable selection is to determine a simpler variable among all the variables. There are two steps for NNG modeling<sup>[16]</sup>. The first step is to achieve a set of regression coefficient with conventional least square (LS) method. The second step is to solve a series of constrained quadratic programming problem to shrink the regression coefficient with the goal of variable selection. In practice, the model structure, parameter and time-delay of the actual process may vary simultaneously. However, the general identification methods of the three items are performed with separate procedures and it is very inconvenient in practical application. As we know, proper variable selection procedure can ensure a compact model and results in robust input-output relation, so the variable selection algorithm is introduced to investigate the feasibility of simultaneous identification of process structure, parameter and time-delay. A novel method of system identification is proposed based on NNG variable selection in this paper. This method can obtain the model structure, parameter and time-delay at the same time rather than in three separate procedures. The derived excellent results show that the proposed NNG-based modeling algorithm can be utilized for simultaneous identification of the model structure, parameter and time-delay with high precision.

## 2 Problem formulation

An Auto-regressive exogenous (ARX) model is considered as follows:

$$A(z^{-1})y(k) = B(z^{-1})u(k-d) + \xi(k) \quad (1)$$

where  $\xi(k)$  is a white noise with zero mean and unit variance, and  $A(z^{-1}) = 1 + a_1z^{-1} + a_2z^{-2} + \dots + a_{n_a}z^{-n_a}$ ,  $B(z^{-1}) = b_1z^{-1} + b_2z^{-2} + \dots + b_{n_b}z^{-n_b}$ ,  $\{a_1, a_2, \dots, a_{n_a}\}$  are the parameters to be identified using process input and output data  $\{u_1, u_2, \dots, u_N\}$  and  $\{y_1, y_2, \dots, y_N\}$  respectively.

Equation (1) can be written in another way as below:

$$y(k) = -a_1y(k-1) - \dots - a_{n_a}y(k-n_a) + b_1u(k-d-1) + \dots + b_{n_b}u(k-d-n_b) + \xi(k). \quad (2)$$

The structure parameters  $n_a, n_b$  and  $d$  must be known as priori knowledge for the most system identification algorithms. However, it is difficult to know the structure pa-

parameter in some cases. So, the regular system identification algorithms do not work in this situation.

We assign a value to the structure parameters  $n_a, n_b$ , for example,  $n_a = N_a, n_b = N_b$ . We can ensure that  $N_a$  and  $N_b$  are bigger enough than the actual value of structure parameter. Then, (2) can be re-written as

$$y(k) = -a_1y(k-1) - \dots - a_{N_a}y(k-N_a) + b_1u(k-1) + \dots + b_{N_b}u(k-N_b) + \xi(k). \tag{3}$$

Then NNG variable selection algorithm can be utilized to calculate the values and frequencies of each time-series variables. When the tightening method based on shrinking is utilized, many setting coefficients of the model will become 0. Suppose that the result is as follows:

$$\begin{aligned} \{a_1, a_2, \dots, a_{N_a}\} &= \{a_{r_1}, a_{r_2}, \dots, a_{r_a}, 0, \dots, 0\} \\ \{b_1, b_2, \dots, b_{N_b}\} &= \{0 \dots 0, b_{r_1}, b_{r_2}, \dots, b_{r_b}, 0, \dots, 0\}. \end{aligned} \tag{4}$$

Then, from (4), we know that there are  $N_a - ra$  parameters of  $\{y(k-1), \dots, y(k-N_a)\}$  being 0 and  $N_b - rb$  parameters of  $\{u(k-1), \dots, u(k-N_b)\}$  being 0. The identified system can be written as

$$y(k) = -a_{r_1}y(k-1) - \dots - a_{r_a}y(k-ra) + b_{r_1}u(k-d-1) + \dots + b_{r_b}u(k-d-rb) + \xi(k). \tag{5}$$

Thus, it can be concluded that  $ra$  is the order of the system and  $d$  is the delay time of the process.

### 3 The conventional identification algorithm

#### 3.1 Least squares identification

As described in (1) and (2), (2) can be transformed into

$$y(k) = \varphi^T(k)\theta + \xi(k) \tag{6}$$

where  $\varphi^T(k) \in \mathbf{R}^{(n_a+n_b+1) \times 1}$ ,  $\varphi^T(k) = [-y(k-1), \dots, -y(k-n_a), u(k-d), \dots, u(k-d-n_b)]^T$  and  $\theta = [a_1, \dots, a_{n_a}, b_0, \dots, b_{n_b}]^T \in \mathbf{R}^{(n_a+n_b+1) \times 1}$ .

It is supposed that the input and output data  $\{y(k), u(k), k = 1, 2, \dots, L\}$  can be rewritten, where  $L$  is the length of the data.

$$\begin{aligned} y(1) &= \varphi^T(1)\theta + \xi(1) \\ y(2) &= \varphi^T(2)\theta + \xi(2) \\ &\vdots \\ y(L) &= \varphi^T(L)\theta + \xi(L). \end{aligned} \tag{7}$$

Assume that

$$Y = \begin{bmatrix} y(1) \\ y(2) \\ \vdots \\ y(L) \end{bmatrix} \in \mathbf{R}^{L \times 1}, \Phi = \begin{bmatrix} \varphi^T(1) \\ \varphi^T(2) \\ \vdots \\ \varphi^T(L) \end{bmatrix} \in \mathbf{R}^{L \times (n_a+n_b+1)}. \tag{8}$$

From (8), the least squares equation can be obtained as below:

$$Y = \Phi\theta + \zeta \tag{9}$$

where  $Y$  and  $\Phi$  are formed by input and output data,  $\theta$  is the parameters that need to be identified.

$$\theta = (\Phi^T\Phi)^{-1}\Phi^TY. \tag{10}$$

However, the structure parameters  $n_a, n_b$  and  $d$  must be known as priori knowledge for the least square identification algorithm. Furthermore, it is difficult to know the structure parameter in some cases. So, the LS identification algorithm does not work in the circumstance of identifying model structure, parameter and time-delay simultaneously.

#### 3.2 Subspace identification

A state space model of linear invariant system is considered as follows:

$$\begin{aligned} x(k+1) &= Ax(k) + Bu(k) + Ke(k) \\ y(k) &= Cx(k) + Du(k) + e(k) \end{aligned} \tag{11}$$

where  $x(k) \in \mathbf{R}^n$ ,  $u(k) \in \mathbf{R}^m$ ,  $y(k) \in \mathbf{R}^p$  and  $e(k) \in \mathbf{R}^p$  is white noise sequence with covariance  $K_e$ .  $(A, B, C, D)$  are the system matrices of controller with appropriate dimension.

We should construct the Hankel matrix of the input and output data to identify the plant system parameters. The standard subspace notation can be achieved through the iteration of (11).

$$\begin{aligned} Y_f &= \Gamma_N X_f + H_N^d U_f + H_N^s E_f \\ Y_p &= \Gamma_N X_p + H_N^d U_p + H_N^s E_p \\ X_f &= A^N X_p + \Delta_N^d U_p + \Delta_N^s E_p \end{aligned} \tag{12}$$

where subscript  $p$  denotes past horizon and  $f$  denotes future horizon. Respectively, the past and future input block Hankel matrices are defined as

$$U_p = \begin{pmatrix} u_0 & u_1 & \dots & u_{j-1} \\ u_1 & u_2 & \dots & u_j \\ \vdots & \vdots & \ddots & \vdots \\ u_{N-1} & u_N & \dots & u_{N+j-2} \end{pmatrix}$$

$$U_f = \begin{pmatrix} u_N & u_{N+1} & \dots & u_{N+j-1} \\ u_{N+1} & u_{N+2} & \dots & u_{N+j} \\ \vdots & \vdots & \ddots & \vdots \\ u_{2N-1} & u_{2N} & \dots & u_{2N+j-2} \end{pmatrix}.$$

Meanwhile, the past and future output block Hankel matrices  $Y_p$  and  $Y_f$  have the same structure. The state matrices  $X_p, X_f$  are defined as follows:

$$\begin{aligned} X_p &= X_0 = \begin{pmatrix} x_0 & x_1 & \cdots & x_{j-1} \end{pmatrix} \in \mathbf{R}^{n \times j} \\ X_f &= X_N = \begin{pmatrix} x_N & x_{N+1} & \cdots & x_{N+j-1} \end{pmatrix} \in \mathbf{R}^{n \times j}. \end{aligned} \tag{13}$$

The lower triangular Toeplitz matrices  $H_N^d, H_N^s$  are defined as

$$H_N^d = \begin{pmatrix} D & 0 & 0 & \cdots & 0 \\ CB & D & 0 & \cdots & 0 \\ CAB & CB & D & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ CA^{N-2}B & CA^{N-2}B & CA^{N-2}B & \cdots & D \end{pmatrix} \in \mathbf{R}^{lN \times mN}$$

$$H_N^s = \begin{pmatrix} I & 0 & 0 & \cdots & 0 \\ CK & I & 0 & \cdots & 0 \\ CAK & CK & I & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ CA^{N-2}K & CA^{N-2}K & CA^{N-2}K & \cdots & I \end{pmatrix} \in \mathbf{R}^{lN \times lN}.$$

The extended observability matrix is

$$\Gamma_N = \begin{pmatrix} C^T & (CA)^T & (CA^2)^T & \cdots & (CA^{N-1})^T \end{pmatrix}^T.$$

The future input  $U_f$  and  $E_f$  can be eliminated by conducting an oblique projection  $\frac{Y_f}{U_f W_p}$  of the subspace matrix (14) for the case of open loop system.

$$Y_f = \Gamma_N X_f + H_N^d U_f + H_N^s E_f \tag{14}$$

$$O_N = \frac{Y_f}{U_f W_p} = \Gamma_N \hat{X}_N \tag{15}$$

where  $W_p = \begin{pmatrix} U_p & Y_p \end{pmatrix}^T$ ,  $\hat{X}_f$  is the estimation state sequence. The column space of  $O_N$  and  $\Gamma_N$  are overlapped, when  $\Gamma_N$  is full rank. The system order is equal to the dimension of  $\Gamma_N$ , that is  $rank(O_N) = n$ .

$O_N$  can be obtained by conducting singular value decomposition (SVD) of  $\Gamma_N$ .

$$O_N = \begin{pmatrix} U_1 & U_2 \end{pmatrix} \begin{pmatrix} S_1 & 0 \\ 0 & S_2 \end{pmatrix} \begin{pmatrix} V_1^T \\ V_2^T \end{pmatrix} \tag{16}$$

where  $S_1 \in \mathbf{R}^{n \times n}$ . It is supposed that  $\Gamma_N = U_1 S_1^{\frac{1}{2}}$ . Then  $\hat{X}_f = \Gamma_N^\dagger O_N$ .  $S_2 = 0$ , when the system is without noise. The elements of  $S_2$  are much closed to zero, when there is existence of noise, while the diagonal of  $S_1$  are bigger than zero. So the numerical value that is greater than zero stands for the order of the system.

There are two steps of subspace identification: 1) Construct the Hankel matrices with the past and future input

and output data to obtain the predicted subspace. Then the system observation matrix or state sequence and the system order can be achieved by the SV decomposition of the predicted subspace. 2) Obtain the system state matrices  $\{A, B, C, D\}$  by solving the least squares problems. There are two kinds of identification algorithms, that is observation matrix based and state sequence estimation based.

From this section, we know that either least squares identification algorithm or subspace identification algorithm can identify the model structure, parameter and time-delay in separate procedures. However, the model structure, parameter and time-delay of the actual process may vary at the same time. So the conventional identification algorithms are very inconvenient in practical application. Thus, its so urgent and valuable to have a methodology to identify the model structure, parameter and time-delay simultaneously.

### 4 System identification based on non-negative garrote

Assuming that a set of observed data  $\{X, y\}$ ,  $X \in \mathbf{R}^{n \times p}$  is the input matrix.  $p$  represents the number of input variables.  $y \in \mathbf{R}^{n \times 1}$  is the quality variables (response variables). Suppose that  $S$  and  $y$  have normalized zero mean and unit standard deviation.  $\hat{\beta}^0 \in \mathbf{R}^{p \times 1}$  is the estimation of linear regression model with OLS.

$$y = X\beta + \varepsilon. \tag{17}$$

Equation (6) can be also established, when  $\varepsilon$  is a color noise rather than white noise. There is no noise correlation between  $y$  and  $X$  in the open-loop system. The second step is to shrink the coefficient by solving the optimization problem:

$$\begin{aligned} J &= \min_{c_j} \left\| y - X\hat{\beta}^0 \cdot *c \right\|^2 \\ \text{s.t.} \quad c_j &\geq 0, \quad \sum_{j=1}^{p_x} c_j \leq s. \end{aligned} \tag{18}$$

With  $s$  decreasing, more  $c_j$  is equal to 0, the other non-zero coefficients are shrunk. Given  $0 \leq s \leq p$ , we can get a set of solutions. Solution trajectory will appear when  $s$  is varying from 0 to  $p$ . Prediction error is estimated by the method of v-fold cross validation<sup>[17]</sup> to find the optimal  $s$  to minimize prediction error, then we can get the optimal regression coefficient. The detail of steps for determining optimal regression coefficient with the method of v-fold cross validation are as below:

1) The data set  $L = \{X, y\}$  can be separated into subset  $L_1 \cdots L_v$ . The initial prediction coefficient  $\hat{\beta}$  of every subset is calculated with ordinary least square, then, the optimization problem of (18) is solved as

$$\begin{aligned} c^{(v)}(s) &= \arg \min_{c_j} \sum_{i=1}^{n_{L_v}} (y_i - X_i c \cdot * \hat{\beta})^2 \\ \text{s.t.} \quad c_j &\geq 0, \quad \sum_{j=1}^{p_x} c_j \leq s. \end{aligned} \tag{19}$$

Suppose that  $\bar{L}_v = L - L_v$ , then, the cross validation error is

$$CVE^{(v)}(s) = \sum_{i=1}^{n\bar{L}_v} (y_i - X_i c^{(v)}(s) \cdot * \hat{\beta})^2. \tag{20}$$

Optimal  $s$  can be selected by minimizing v-fold cross validation error, that is

$$s^* = \arg \min_{0 \leq s \leq p_x} \frac{1}{v} \sum_{v=1}^v CVE^{(v)}(s) \tag{21}$$

2) Once again the following equation is solved to achieve the overall optimal regression coefficient:

$$c^* = \arg \min_{c_j} \sum_{i=1}^{n_L} (y_i - X_i c \cdot * \hat{\beta})^2$$

$$\text{s.t.} \quad c_j \geq 0, \sum_{j=1}^{p_x} c_j \leq s^*. \tag{22}$$

3) The regression coefficient derived from NNG algorithm can be expressed as follows:

$$\hat{\beta}_{NNG} = c^* \cdot * \hat{\beta}. \tag{23}$$

The advantages of NNG algorithm are easiness and convenience in practice. The differences between NNG and other subset selection methods are as follows:

In the method of subset selection, the trade-off between the effectiveness of model fitting and the penalty strength of the number of selected variables is presented via different criteria, while the approach of NNG is implemented by directly selecting the parameter  $s$ . Different values of  $s$  correspond to different penalty strength<sup>[10]</sup>.

## 5 Simulation application

### 5.1 SISO system

In order to verify the validity of process model and time delay identification NNG-based method, considering a second order model described by (24)

$$A(q)y(t) = B(q)u(t) + \varepsilon(t) \tag{24a}$$

where  $\varepsilon(k)$  is a white noise with mean 0 and unit variance. And where

$$\begin{cases} A(q) = 1 - 1.5q^{-1} + 0.7q^{-2} \\ B(q) = q^{-d}(1 + 0.5q^{-1}) \end{cases} \tag{24b}$$

$d$  is delay time of second order system, and  $d \geq 1$ .

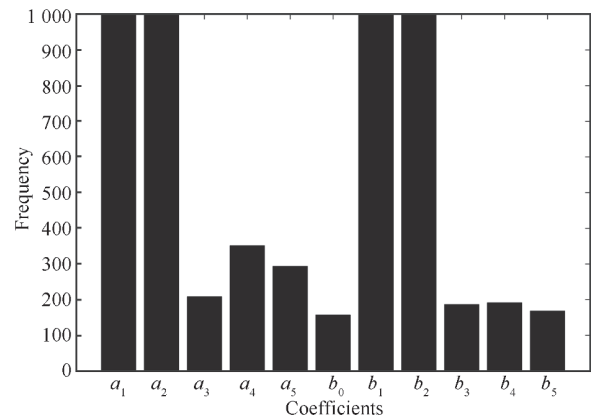
Zero mean and unit variance white noise is used to motivate the system in the experience. Suppose that the orders of polynomial  $A(q)$  and  $B(q)$  of ARX model is (5, 5) when generating open-loop input and output data. So, the parameters that need to be identified are  $a_1, \dots, a_5, b_0, \dots, b_5$  respectively. Set sample size as  $n_L = 1500$ . The value of  $d$  can be selected as 1, 2, 3, 4. Regression coefficient of 1000 sets of open-loop input and output data can be estimated

by NNG algorithm. The frequency of every non-zero coefficient in 1000 experiments is shown as in Fig. 1.

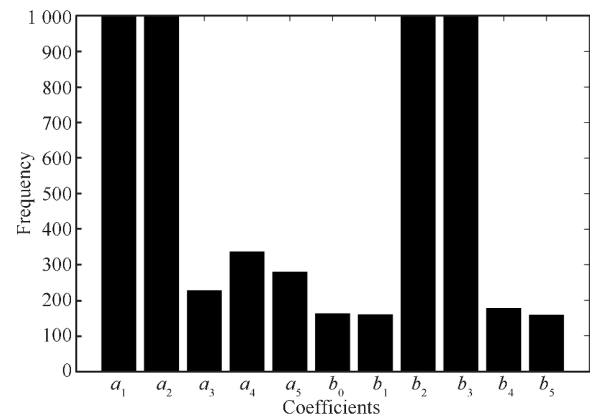
From Fig.1(a), we notice that the frequency of  $a_1, a_2, b_1, b_2$  whose coefficients are not equal to zero is 1000 respectively, whereas the others are less than 300 when  $d = 1$ . So, we know that  $a_1, a_2, b_1, b_2$  are not zero, while the other coefficients are equal to zero, with the method of NNG algorithm in the ARX model. The order of polynomial  $A(q)$  is 2. Because  $a_1 a_2$  are not equal to zero, while  $a_3 a_4 a_5$  are equal to zero. In the same way, we know that the order of  $B(q)$  is 2. The system delay time  $d$  is 1, because  $b_0$  is zero. We can derive the system order and system delay time from the other figures in the same way. So, NNG algorithm is useful for the identification of the system order and system delay time.

When  $d = 3$ , the parameter trajectory of 1000 experiments with the identification method of NNG algorithm is shown in Fig. 2.

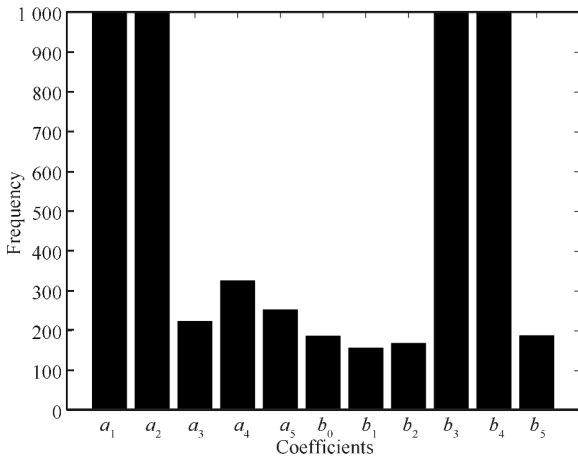
From Fig.2, the identified parameter fluctuates up and down in the vicinity of the actual value, which illustrates the effectiveness of the identification result with NNG algorithm. When  $d = 3$ , the compared result of the mean value and standard deviation of parameters with NNG and OLS algorithms using the same data is shown in Table 1. We can see that the resulting value identified with OLS is larger than that with NNG for the zero actual parameter and so is the standard deviation of sample data. As all the



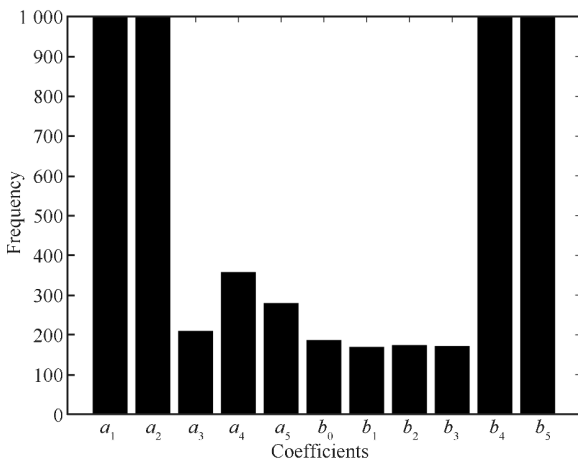
(a)  $d = 1$



(b)  $d = 2$



(c)  $d = 3$

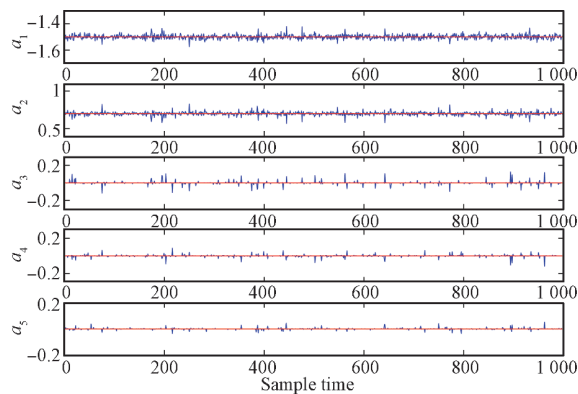


(d)  $d = 4$

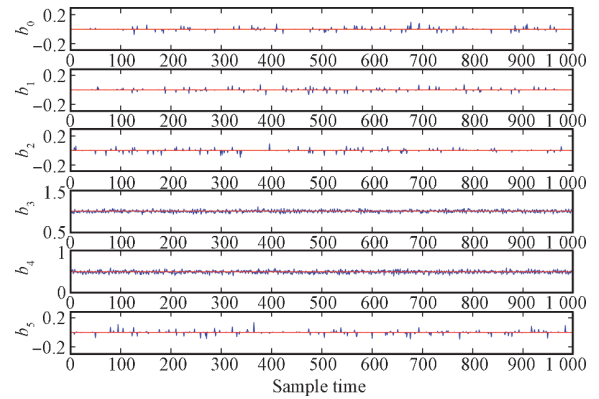
Fig. 1 The frequency of non-zero regression coefficients in 1000 experiments

regression coefficient identified by OLS are not equal to 0 during the 1000 experiments, so model structure and delay time cannot be obtained via OLS identification method.

We can conclude that the regression model derived from NNG method are fewer parameter, less complexity and good accuracy.



(a) The identified coefficient  $A(q)$



(b) The identified coefficient  $B(q)$

Fig. 2 The identified coefficient trajectory with NNG algorithm in 1000 experiments

### 5.2 MIMO system

In order to verify the effectiveness of NNG algorithm for MIMO system, consider a MIMO system (25)

$$\begin{cases} A(q)y(k) = B(q)u(k) + \varepsilon(k) \\ A(q) = 1 - 0.8q^{-1} - 0.2q^{-2} + 0.6q^{-3} \\ B(q) = \begin{bmatrix} B_{11} & B_{12} \\ B_{21} & B_{22} \end{bmatrix} \\ B_{11} = 3q^{-1} - 3.5q^{-2} - 1.5q^{-3} \\ B_{12} = -4q^{-2} - 2q^{-3} - q^{-4} \\ B_{21} = q^{-1} - 0.2q^{-2} - 0.5q^{-3} \\ B_{22} = q^{-1} - 1.5q^{-2} + 0.5q^{-3} + 0.2q^{-4}. \end{cases} \quad (25)$$

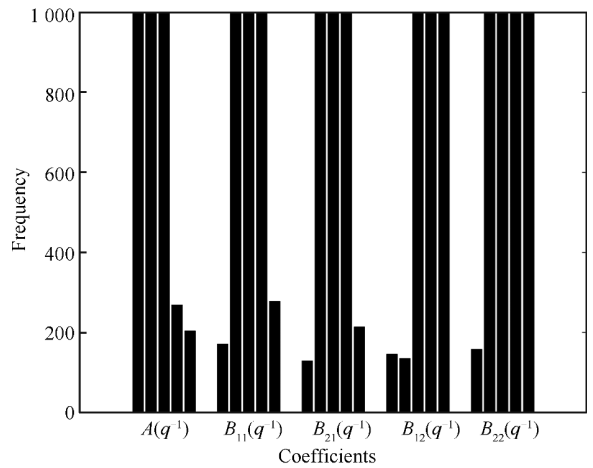
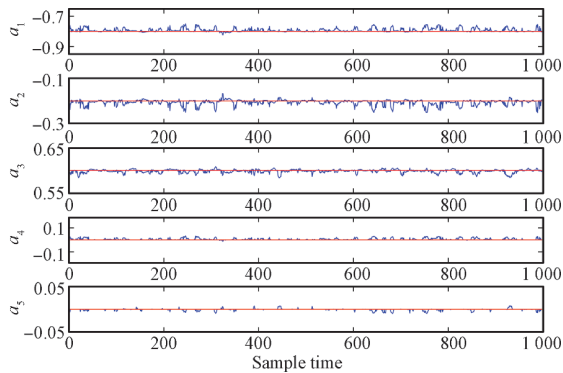


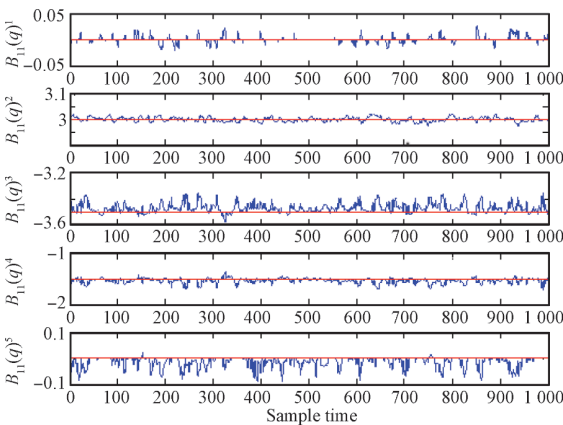
Fig. 3 The frequency of non-zero regression coefficient in 1000 experiments

Here,  $\varepsilon(k)$  is a white noise with zero mean and unit variance. In the experiment, a white noise with zero mean and unit variance is used to excite the system. Suppose that the orders of  $A(q)$  and  $B(q)$  of ARX model are (5, 5, 5, 5, 5) when generating open loop input and output data, where  $A(q) = [a_1, a_2, a_3, a_4, a_5]$ ,  $B(q) = [B_{11}(q), B_{12}(q), B_{21}(q), B_{22}(q)]$ . The parameters need to

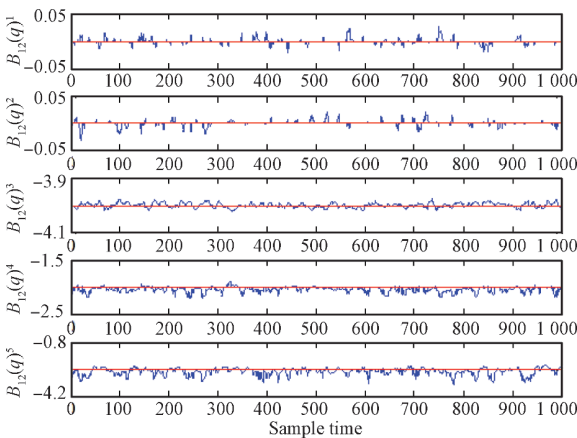
be identified are  $a_1, a_2, a_3, a_4, a_5$  and the coefficient of  $B_{11}(q), B_{12}(q), B_{21}(q), B_{22}(q)$ . The data length is  $n_L = 1500$ . The regression coefficient of 1000 sets of input and output data are estimated with NNG algorithm. The frequencies of non-zero coefficient in the 1000 experiments are shown as bar graph in Fig. 3. We can obtain that the order of  $A(q), B_{11}(q), B_{21}(q), B_{12}(q), B_{22}(q)$  are 3, 3, 3, 4, 4 respectively. The coefficient trajectories of  $A(q), B_{22}(q)$  with the NNG identification method in 1000 experiments are shown in Fig. 4. The identified parameters fluctuate around the actual value. This proves the effectiveness of the identification result with NNG algorithm.



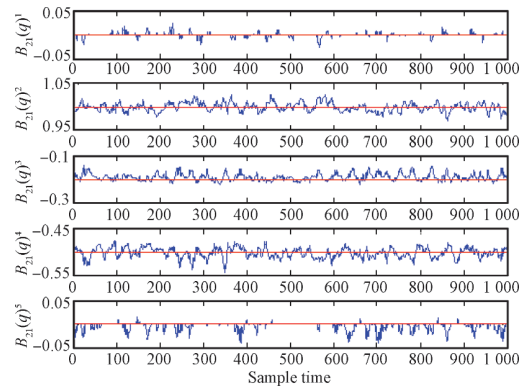
(a) The identified coefficient  $A(q)$



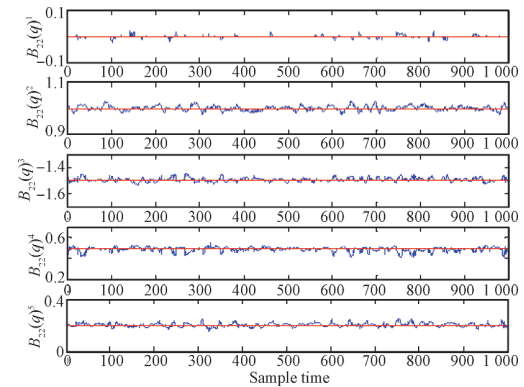
(b) The identified coefficient  $B_{11}(q)$



(c) The identified coefficient  $B_{12}(q)$



(d) The identified coefficient  $B_{21}(q)$



(e) The identified coefficient  $B_{22}(q)$

Fig. 4 The identified coefficients trajectory of  $A(q), B_{11}(q), B_{12}(q), B_{21}(q), B_{22}(q)$  with NNG algorithm in 1000 experiments

### 5.3 Industrial example

A two input two output Wood-Berry tower model is considered in this application. This methane/water rectifying tower is a classic MIMO system with large time delay. The tower has strong temperature influence between top and bottom part. The outputs of this model are the concentration  $y_1$  of distillate from the top of tower and the concentration  $y_2$  of liquid from the bottom of tower. The outputs are controlled by the back flow  $u_1$  from the top and steam reboiler  $u_2$  from the bottom. Feed flow  $w$  is unmeasurable disturbance variable.  $G(s)$  is the transfer function and  $H(s)$  is the disturbance transfer function.

$$y = G(s)u + H(s)w \tag{26a}$$

$$G(s) = \begin{bmatrix} \frac{12.8e^{-s}}{16.7s + 1} & \frac{-18.9e^{-3s}}{21.0s + 1} \\ \frac{6.6e^{-7s}}{10.9s + 1} & \frac{-19.4e^{-3s}}{14.4s + 1} \end{bmatrix} \tag{26b}$$

$$H(s) = \begin{bmatrix} \frac{3.8e^{-8s}}{14.9s + 1} \\ \frac{4.9e^{-3s}}{13.2s + 1} \end{bmatrix}$$

Sample time  $T_s = 1$  s. The discretized process model is  $G(q)$ :

$$G(q) = \frac{1}{A(q)} \begin{bmatrix} B_{11}(q) & B_{12}(q) \\ B_{21}(q) & B_{22}(q) \end{bmatrix} \quad (27a)$$

$$\begin{cases} A(q) = 1.0000 - 3.7406q^{-1} + 5.2467q^{-2} - 3.2704q^{-3} + 0.7644q^{-4} \\ B_{11}(q) = 0.7440q^{-1} - 2.0822q^{-2} + 1.9422q^{-3} - 0.6038q^{-4} \\ B_{12}(q) = -0.8789q^{-3} + 2.4496q^{-4} - 2.2756q^{-5} + 0.7046q^{-6} \\ B_{21}(q) = 0.5786q^{-7} - 1.6363q^{-8} + 1.5426q^{-9} - 0.4847q^{-10} \\ B_{22}(q) = -1.3015q^{-3} + 3.6543q^{-4} - 3.4195q^{-5} + 1.0664q^{-6} \end{cases} \quad (27b)$$

The disturbance  $w$  is a white noise with 0 mean and unit variance. The excitation input has also a 0 mean and unit variance white noise. Suppose that the order of ARX model is (5, 11, 11, 11, 11) when generating the input and output data. Here,  $A(q) = [a_1 a_2 \dots a_{12}]$ . There are twelve parameters in each polynomial  $B_{11}(q)$ ,  $B_{12}(q)$ ,  $B_{21}(q)$ ,  $B_{22}(q)$ . Given  $n_L = 1000$ . The regression coefficients of 1000 sets of input and output data are estimated with NNG algorithm. The frequency of non-zero coefficient in the 1000 experiments are shown as bar graph in Fig. 5. We can obtain that the order of  $A(q)$ ,  $B_{11}(q)$ ,  $B_{12}(q)$ ,  $B_{21}(q)$ ,  $B_{22}(q)$  as 4, 4, 6, 10, 6, respectively.

It can be seen from Fig. 5 that the selected frequency of variables with zero coefficients is less than 200 and most of them is about 100. This results illustrate that the variables with zero coefficient are not been selected in about 90% of the experiments. This figure can prove the effectiveness of variable selection with NNG algorithm. The tendency chart of the identified  $A(q)$  which fluctuates around the true value is shown in Fig. 6.

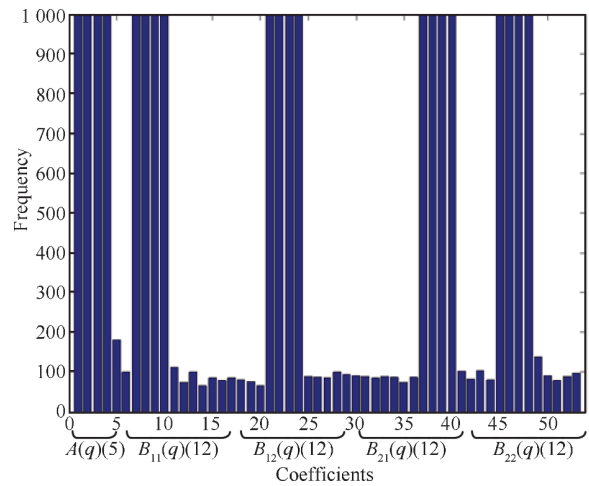


Fig. 5 The frequency of non-zero regression coefficients in 1000 experiments

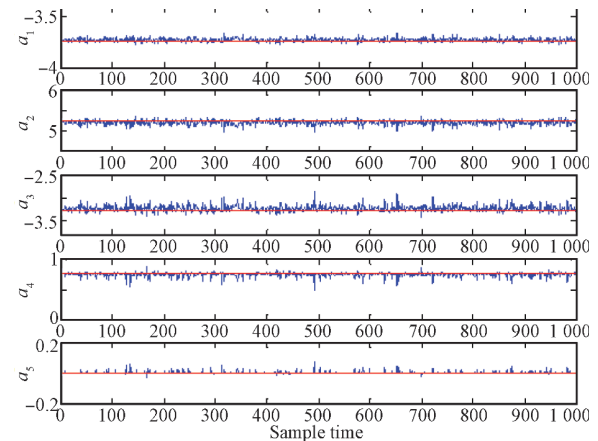


Fig. 6 The identified coefficients trajectory with NNG algorithm in 1000 experiments

Table 1 The identified coefficients with different identification methods

Parameter	True value	Mean of NNG	Standard variance of NNG	Mean of OLS	Standard variance of OLS
$a_1$	-1.5	-1.499 8	0.017 4	-1.498 8	0.024 5
$a_2$	0.7	0.699 8	0.024	0.697 7	0.046 1
$a_3$	0	0.000 1	0.016 2	0.001 8	0.044 4
$a_4$	0	0.000 1	0.012	-0.001 1	0.035 2
$a_5$	0	0.000	0.005 5	0.000 4	0.017 2
$b_0$	0	0.000 5	0.012 2	0.000 8	0.026 1
$b_1$	0	0.000 2	0.010 6	-0.000 2	0.025 4
$b_2$	0	-0.000 1	0.010 3	0.001	0.026
$b_3$	1	0.998 7	0.025 7	0.999 7	0.026 3
$b_4$	0.5	0.495 2	0.030 6	0.502 3	0.035 3
$b_5$	0	0.000 1	0.014 7	0.000 6	0.025 6



## 6 Conclusions

Many existing process identification methods treat the structure, parameter and time-delay identification with separate procedures. In practice, the model structure is known as priori knowledge. In fact, this is unrealistic in most cases. The model accuracy cannot be met if the model order is supposed smaller than the actual value. In contrast, the model will become too complicated with a high model order. The usual way to determine the model order is to increase gradually the order from low to high and to meet the accuracy requirements during the process identification. In other words, the model structure identification and the parameter identification are treated as two independent tasks. What is more, the time delay, which plays an important role in determination of control performance benchmarks, has to be determined with other special methods. Thus, the existing identification methods for the three items are performed with separate procedures which is very inconvenient in practical application. In order to explore the feasibility of variable selection algorithm for the simultaneous identification of process structure, parameter and time-delay, NNG algorithm is introduced and applied to system identification. The application of NNG variable selection algorithm in simulation examples and Wood-Berry tower model are investigated. The identification accuracy and the time-series variable selection results are compared and analyzed between NNG and OLS algorithm. The derived excellent results show that the proposed NNG-based modeling algorithm can be utilized for simultaneous identification of the model structure, parameters and time-delay with high precision. But as the NNG algorithm can be only used to open-loop system, so how the NNG algorithm can be used for closed-loop system can be further studied.

## Acknowledgments

This work was supported by National Natural Science Foundation of China (No. 61171145).

## References

- [1] L. Fortuna, S. Graziani, A. Rizzo. *Soft Sensors for Monitoring and Control of Industrial Processes*, Berlin, Germany: Springer-Verlag, 2007.
- [2] Y. P. Badhe, J. Lonari, S. S. Tambe, B. D. Kulkarni, N. K. Valecha, S. V. Deshmukh, S. Ravichandran. Improve polyethylene process control and product quality. *Hydrocarbon Processing*, vol. 86, no. 3, pp. 53–60, 2007.
- [3] K. Desai, Y. Badhe, S. S. Tambe, B. D. Kulkarni. Soft-sensor development for fed-batch bioreactors using support vector regression. *Biochemical Engineering Journal*, vol. 27, no. 3, pp. 225–239, 2006.
- [4] K. Sun, J. L. Liu, J. L. Kang, S. S. Jang, D. S. H. Wong, D. S. Chen. Development of a variable selection method for soft sensor using artificial neural network and nonnegative garrote. *Journal of Process Control*, vol. 24, no. 7, pp. 1068–1075, 2014.
- [5] T. H. Pan, D. S. H. Wong, S. S. Jang. Development of a novel soft sensor using a local model network with an adaptive subtractive clustering approach. *Industrial and Engineering Chemistry Research*, vol. 49, no. 10, pp. 4738–4747, 2010.
- [6] C. Y. Li, W. G. Li. Partial least squares method based on least absolute shrinkage and selection operator. In *Proceedings of the 3rd International Conference on Advanced Computer Theory and Engineering (ICACTE)*, IEEE, Chengdu, China, vol. 4, pp. 591–593, 2010.
- [7] R. J. Shi, J. F. MacGregor. Modeling of dynamic systems using latent variable and subspace methods. *Journal of Chemometrics*, vol. 14, no. 5–6, pp. 423–439, 2000.
- [8] S. J. Qin. Recursive PLS algorithms for adaptive data modeling. *Computers & Chemical Engineering*, vol. 22, no. 4–5, pp. 503–514, 1998.
- [9] S. Bhartiya, J. R. Whiteley. Development of inferential measurements using neural networks. *ISA Transactions*, vol. 40, no. 4, pp. 307–323, 2001.
- [10] L. Breiman. Better subset regression using the nonnegative garrote. *Technometrics*, vol. 37, no. 4, pp. 373–384, 1995.
- [11] R. Tibshirani. Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society: Series B (Statistical Methodological)*, vol. 58, no. 1, pp. 267–288, 1996.
- [12] H. Chun, S. Keles. Sparse partial least squares regression for simultaneous dimension reduction and variable selection. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, vol. 72, no. 1, pp. 3–25, 2010.
- [13] M. Yuan, Y. Lin. On the non-negative garrote estimator. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, vol. 69, no. 2, pp. 143–161, 2007.
- [14] J. Mohieddine. An overview of control performance assessment technology and industrial application. *Control Engineering Practice*, vol. 14, no. 5, pp. 441–466, 2006.
- [15] A. Alenany, H. Shang, M. Soliman, I. Ziedan. Improved subspace identification with prior information using constrained least squares. *Control Theory & Applications*, vol. 5, no. 13, pp. 1568–1576, 2010.
- [16] C. C. Pan, J. Bai, G. G. Yang, D. S. H. Wong, S. S. Jang. An inferential modeling method using enumerative PLS based nonnegative garrote regression. *Journal of Process Control*, vol. 22, no. 9, pp. 1637–1646, 2012.
- [17] H. Chun, S. Keles. Sparse partial least squares regression for simultaneous dimension reduction and variable selection. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, vol. 72, no. 1, pp. 3–25, 2010.

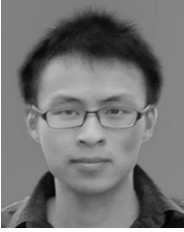


**Jian-Guo Wang** received the Ph. D. degree in control science and engineering from Shanghai Jiao Tong University, China in 2007. Now he is a faculty member in the Department of Automation, Shanghai University, China.

His research interests include the statistical modeling and control of multivariate process, performance monitoring and energy saving optimization of power system, steel rolling and heating furnace.

E-mail: jgwang@shu.edu.cn

ORCID iD: 0000-0003-0232-030X

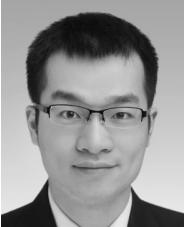


**Qian-Ping Xiao** received the B. Sc. degree in electronic science and technology from Jingdezhen Ceramic Institute, China in 2012, and now is a master student in pattern recognition and intelligent system at Shanghai University, China.

His research interests include subspace identification and performance evaluation of model predictive control systems.

E-mail: xqp.jier.shu@gmail.com (Corresponding author)

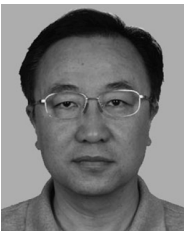
ORCID iD: 0000-0001-5880-2741



**Tiao Shen** received the B. Sc. degree in electrical engineering and automation from Shanghai University, China in 2014, and now is a master student in pattern recognition and intelligent system at Shanghai University, China.

His research interests include machine learning and data mining.

E-mail: 564306302@qq.com



**Shi-Wei Ma** received the B. Sc. and the M. Sc. degrees in electronics from Lanzhou University, China in 1986 and 1991, respectively, and received the Ph. D. degree in control theory and engineering from Shanghai University, China in 2000. From 2001 to 2003, he was a Japan science and technology research fellow at the National Institute of Industrial Safety of Japan. From 2003 to

2008, he was an associate professor, and since 2008, he has been

a professor, in the Department of Automation at Shanghai University, China.

His research interests include signal processing, pattern recognition and intelligent system.

E-mail: swma@mail.shu.edu.cn



**Wen-Tao Rao** received the B. Sc. and M. Sc. degrees in thermal energy power engineering from Kunming University of Science and Technology, China in 1990 and 1993, respectively, and received the Ph. D. degree in thermal energy power engineering from Tongji University in 2002. He is a postdoctoral of Kunming University of Science and Technology in 2006.

His research interests include furnaces technology and control technology, energy saving technology and equipment, new energy, energy management contract.

E-mail: 13501785249@163.com



**Yong-Jie Zhang** received the B. Sc. degree in non-ferrous metallurgy from Northeastern University, China in 1992, and received the M. Sc. and Ph. D. degrees in thermal power engineering from Northeastern University in 1992 and 2005, respectively.

His research interests include the basic study of new generation of steel materials and the development of electromagnetic continuous casting technology and equipment.

E-mail: zyj@baosteel.com