IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received November 17, 2020, accepted December 1, 2020, date of publication December 14, 2020,
date of current version December 29, 2020.

Digital Object Identifier 10.1109/ACCESS.2020.3044272

An Over-Sampling Amplitude-Limited Variational Bayesian Method for the Identification of Hammerstein Model

BAOCHANG XU, LIKUN YUAN, AND YAXIN WANG
Department of Automation, China University of Petroleum, Beijing 102249, China
Corresponding author: Likun Yuan (ylk1221@sina.com)

This work was supported by the National Key Research and Development Project of China under Grant 2019YFA0708304.

ABSTRACT Chemical industrial processes involve numerous multivariable nonlinear systems. Nonlinear Muli-Input Muli-Output (MIMO) models seem more suitable to represent most systems and control problems in industrial processes. Furthermore, the outputs of the real systems might be corrupted with the colored noises, which do not satisfy the assumption of the white noises. In order to solve the impact of the colored noises, an Amplitude-Limiting Variational Bayesian (ALVB) method combined with multivariable nonlinear model (Hammerstein model) working in over-sampling closed-loop structure is proposed in this article. This method is the improvement of the Variational Bayesian (VB) method combining Hammerstein model and over-sampling closed-loop structure. Simulation experiments show that for the nonlinear model (Hammerstein model), the proposed algorithm not only overcomes the unidentifiable disadvantage of the traditional structure but also contributes to a higher identification accuracy. Furthermore, even under situation that the processes output noise is a colored noise, the proposed algorithm still maintains and converges to the achieved accuracy.

INDEX TERMS Over-sampling closed-loop structure, hammerstein model, variational Bayesian (VB) method, amplitude-limited variational Bayesian (ALVB) method, colored noise.

I. INTRODUCTION

In current chemical industrial processes, the performance of the controller is determined by the accuracy of the process models [1]. An industrial process involves a class of multivariable nonlinear systems. The problems of the multivariable nonlinear plants operating in industrial processes have not received a lot of attention so far. According to industrial demands and relative theory, we choose to replace the nonlinear model with a linear model for research, it is just an approximation which is easier for process analysis. This method to describe nonlinear models requires additional numerous restrictive conditions, and the description is still not correct [2]. Therefore, it is necessary to do the further research on multivariable nonlinear model identification to meet the needs of practical industrial processes.

In the field of multivariable nonlinear model identification, a Hammerstein model which consists of a static nonlinear module and a dynamic linear module is the focus of the nonlinear modelling [3]. The Hammerstein model has less computational complexity and represents the characteristics of a process [4], which is always used for industrial nonlinear process analysis, such as continuous reactors [5], PH neutralization processes [6], and pressurized boilers [7]. There are lots of methods of Hammerstein models, such as the traditional iterative method [8], over-parameter identification method [10], subspace identification method [11], blind identification method [13], neural network [15], and particle swarm algorithm [16]. In the identification methods above, the extra excitation should be applied to ensure the informative data for the system parameter estimate [17]-[18]. In the large industrial chemical process, the extra excitation is generally limited to ensure that the identification experiments do not cause the unqualified products and the emergency shutdown [19]. The extra excitation might not only cause a huge cost for identification procedure, but also produce an effect on the industrial processes regularly operating [20]. In order to solve the practical problems above, this article

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

provides a new method to obtain an accurate model at a low cost.

There have been a lot of researches on the model structure identifiability. Sun first proposed the over-sampling closed-loop structure and proved that the over-sampling closed-loop structure identification could ensure the identifiability without extra excitation [21]. Wang proved the identifiability of the linear over-sampling closed-loop structure without input signal in the frequency domain [22], [23]. By analyzing the asymptotic variance expression of linear over-sampling structure, Zhu concluded the high-frequency parts in the output noise can be converted into persisting exciting with the over-sampling structure [24].

With the deep study of the over-sampling structure, a series of traditional identification methods, such as the least square method [25], the prediction-error [26] method and the asymptotic variance method [27], are improved by the combination with the over-sampling structure. The researches above are most based on the univariate linear models, less on the multivariate models. A new identification method, the Variational Bayesian (VB) method can greatly improve the accuracy of the model and widely used in the industrial identification experiments with the fast convergence. The Variational Bayesian (VB) methods based on several model types have been proposed, such as multi-switched model [28], multi-switched model under gamma noise distribution, time-varying model [29], and autoregressive exogenous (ARX) model with random missing output data [30]. In the practical industrial processes, the Variational Bayesian (VB) methods above require extra excitation to ensure the informativity of the identification experiments, which might cause a large cost for the normal plant operation. There is also a problem that the original VB method might not converge with the colored noise.

In order to obtain an accurate dynamic model at a low cost, this article provided a VB method based on the multivariable nonlinear model with over-sampling which combines the Hammerstein models and over-sampling closed-loop structure. An Amplitude-Limited Variational Bayesian (ALVB) method combined with the over-sampling closed-loop structure with colored noise for multivariable nonlinear models is proposed. This algorithm is improved from the traditional VB method and applicable to the condtion of colored noise. Simulations show that the Amplitude-Limited Variational Bayesian (ALVB) method combined with the over-sampling closed-loop structure overcomes the shortcoming of the traditional closed-loop structure identifiability and achieves higher accuracy. Simulations also prove the algorithm convergences in the condtion of colored noise.

II. MULTIVARIABLE NONLINEAR OVER-SAMPLING CLOSED-LOOP STRUCTURE

In the large-scale industrial processes, the continuously operated plants are always nonlinear Multi-Input Multi-Output (MIMO) systems, the frequency in the operating system is much lower than that in the DCS sampling system, which is suitable for the multivariable nonlinear over-sampling structure identification. Therefore, multivariable nonlinear oversampling structure can be used for the large-scale industrial process modelling.

Fig.1 displays how the Hammerstein model over-sampling structure operated in closed-loop identification. The Hammerstein model basically consists of a nonlinear static multivariable block $\boldsymbol{F}(\bullet)$ and a dynamic linear multivariable block $\boldsymbol{G_c}(s)$. There exists a multivariable controller $\boldsymbol{K}(z^{-1})$ of the control period T, and z^{-1} is the backward shift operator that corresponds to T, i.e, $z^{-1}y(t) = y(t-1)$. $\boldsymbol{K}(z^{-1})$ generates piecewise model input $\boldsymbol{U}(m)$ through a zero-order holder. In the over-sampling closed-loop structure, the output is sampled at a period of $\Delta = T/p$ to generate $\boldsymbol{Y_\Delta}(m)$ for identification, while the sampling time for output in conventional identification is T, p is the positive integer indicating the over-sampling rate.

In the Hammerstein model over-sampling structure shown in Fig.1, $\boldsymbol{U}(m) = [\boldsymbol{u}_1(m)\cdots\boldsymbol{u_n}(m)]^T$ is the nonlinear part input and $\boldsymbol{\zeta}(m) = [\boldsymbol{\zeta}_1(m)\cdots\boldsymbol{\zeta_n}(m)]^T$ is the nonlinear part output. $\boldsymbol{R}(m) = [\boldsymbol{r}_1(m)\cdots\boldsymbol{r_n}(m)]^T$ is the system input and $\boldsymbol{Y_\Delta}(k) = [\boldsymbol{y_{\Delta 1}}(k)\cdots\boldsymbol{y_{\Delta n}}(k)]^T$ is the system output. $\boldsymbol{V_\Delta}(k) = [\boldsymbol{v_{\Delta 1}}(k)\cdots\boldsymbol{v_{\Delta n}}(k)]^T$ is a white zero mean noise or colored noise vector.

It is assumed that the nonlinear static multivariable block $\boldsymbol{F}(\bullet)$ can be expressed as

$$\begin{aligned}\zeta_{\boldsymbol{i}}(m) &= F_i\left(\boldsymbol{u_i}(m)\right)\\ &= d_{i1}f_{i1}\left(\boldsymbol{u_i}(m)\right) + d_{i2}f_{i2}\left(\boldsymbol{u_i}(m)\right) + \cdots d_{in_d}f_{in_d}\left(\boldsymbol{u_i}(m)\right)\\ &= \sum_{l=1}^{n_d} d_{il}f_{il}\left(\boldsymbol{u_i}(m)\right).\end{aligned} \tag{1}$$

Denote the plant model $\boldsymbol{G_{c\Delta}}(s)$ with respect to sampling time Δ as

$$\boldsymbol{Y_\Delta}(k) = \boldsymbol{G_{c\Delta}}\left(q^{-1}\right)\boldsymbol{\zeta_\Delta}(k) + \boldsymbol{V_\Delta}(k), \tag{2}$$

then the specific forms of $\boldsymbol{G_{c\Delta}}(q^{-1})$ and $\boldsymbol{V_\Delta}(k)$ are separately as follows

$$\begin{aligned}&\boldsymbol{G_{c\Delta}}\left(q^{-1}\right)\\ &\quad= \frac{\boldsymbol{B_\Delta}\left(q^{-1}\right)}{A_\Delta\left(q^{-1}\right)}\\ &\quad= \frac{1}{A_\Delta\left(q^{-1}\right)}\begin{bmatrix} B_{\Delta 11}\left(q^{-1}\right) & \cdots & B_{\Delta 1n}\left(q^{-1}\right)\\ \vdots & \ddots & \vdots\\ B_{\Delta n1}\left(q^{-1}\right) & \cdots & B_{\Delta nn}\left(q^{-1}\right)\end{bmatrix},\\ &\boldsymbol{V_\Delta}(k) = \boldsymbol{H_\Delta}\left(q^{-1}\right)v_{\Delta *}(k)\\ &\quad= \frac{\boldsymbol{C_\Delta}\left(q^{-1}\right)}{A_\Delta\left(q^{-1}\right)}v_{\Delta *}(k)\\ &\quad= \frac{1}{A_\Delta\left(q^{-1}\right)}\begin{bmatrix} C_{\Delta 1}\\ \vdots\\ C_{\Delta n}\end{bmatrix}\boldsymbol{v}_{\Delta *}(k)\end{aligned} \tag{3}$$

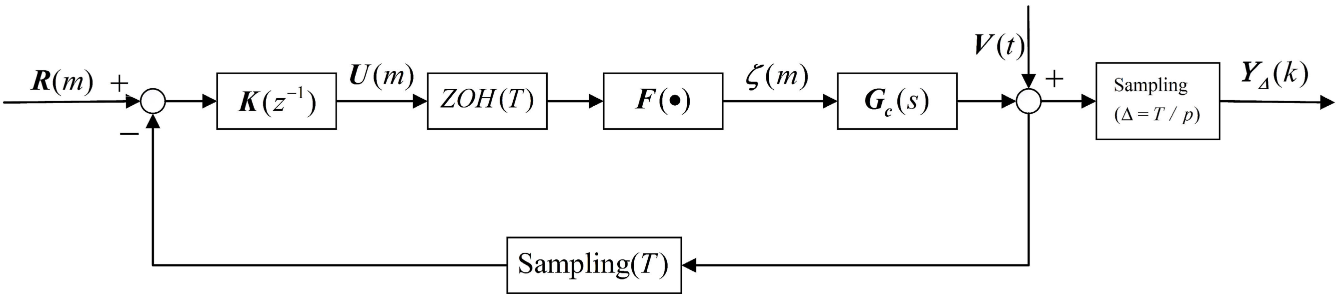

FIGURE 1. Over-sampling structure in closed-loop identification.

where $\boldsymbol{H}_{\boldsymbol{\Delta}}(q^{-1})$ is the stable minimum phase transfer function. $v_{\Delta *}(k)$ is a white noise with zero mean and δ^{-1} variance. $\boldsymbol{B}_{\boldsymbol{\Delta}}(q^{-1})$ is the $n \times n$ numerator of the trancfer function $\boldsymbol{G}_{\boldsymbol{c\Delta}}(q^{-1})$ and $\boldsymbol{C}_{\boldsymbol{\Delta}}(q^{-1})$ is the $n \times 1$ numerator of the transfer function $\boldsymbol{H}_{\boldsymbol{c\Delta}}(q^{-1})$. $A_{\Delta}(q^{-1})$ is the denominator of the transfer function $\boldsymbol{G}_{\boldsymbol{c\Delta}}(q^{-1})$.

In (3), $A_{\Delta}(q^{-1})$, $\boldsymbol{B}_{\boldsymbol{\Delta}}(q^{-1})$, and $\boldsymbol{C}_{\boldsymbol{\Delta}}(q^{-1})$ are separately expressed as

$$
\begin{aligned}
A_{\Delta}\left(q^{-1}\right) &= 1 + a_{\Delta,1}q^{-1} + \cdots + a_{\Delta,n_a}q^{-n_a}, \\
B_{\Delta ij}\left(q^{-1}\right) &= b_{\Delta ij,1}q^{-1-\tau_{b\Delta ij}} + \cdots + b_{\Delta ij,n_b}q^{-n_b-\tau_{b\Delta ij}}, \\
C_{\Delta i}\left(q^{-1}\right) &= 1 + c_{\Delta i,1}q^{-1} + \cdots + c_{\Delta i,n_c}q^{-n_c}
\end{aligned} \tag{4}
$$

where $a_{\Delta,1} \cdots a_{\Delta,n_a}$ are the parameters of the $A_{\Delta}(q^{-1})$ and n_a is the number of $A_{\Delta}(q^{-1})$ parameters. $b_{\Delta ij,1} \cdots b_{\Delta ij,n_b}$ are the parameters of $B_{\Delta ij}(q^{-1})$ and n_b is the number of $B_{\Delta ij}(q^{-1})$ parameters. $\tau_{b\Delta ij}$ is the time delay of $B_{\Delta ij}(q^{-1})$. $c_{\Delta,1} \cdots c_{\Delta,n_c}$ are the parameters of $C_{\Delta i}(q^{-1})$ and n_c is the number of parameters of $C_{\Delta i}(q^{-1})$.

In the over-sampling closed-loop structure, the input $\boldsymbol{U}_{\boldsymbol{\Delta}}(k)$ is also over-sampled. Due to the zero-order holder, the input $\boldsymbol{U}_{\boldsymbol{\Delta}}(k)$ for identification is actually generated as

$$
\boldsymbol{U}_{\boldsymbol{\Delta}}(k\Delta) = \boldsymbol{U}(mT), \;\; k = mp, m\,p+1, \ldots, (m+1)p-1. \tag{5}
$$

Referring to (1) - (5), we can obtain the following

$$
\begin{aligned}
\boldsymbol{Y}_{\boldsymbol{\Delta}}(k) = &-\sum_{q_a=1}^{n_a} a_{\Delta,q_a}\boldsymbol{Y}_{\boldsymbol{\Delta}}(k-q_a) \\
&+ \sum_{q_b=1}^{n_b} \boldsymbol{B}_{\boldsymbol{\Delta},q_b}F\left(\boldsymbol{U}_{\boldsymbol{\Delta}}(k-q_b)\right) \\
&+ \sum_{q_c=1}^{n_c} \boldsymbol{c}_{\boldsymbol{\Delta},q_c}\boldsymbol{e}_{\boldsymbol{\Delta}}(k-q_c) + \boldsymbol{e}_{\boldsymbol{\Delta}}(k),
\end{aligned} \tag{6}
$$

$$
\boldsymbol{B}_{\boldsymbol{\Delta},q_b} = \begin{bmatrix} b_{\Delta 11,q_b} & \cdots & b_{\Delta 1n,q_b} \\ \vdots & \ddots & \vdots \\ b_{\Delta n\,1\,q_b} & \cdots & b_{\Delta nn,q_b} \end{bmatrix},
$$

$$
\boldsymbol{c}_{\boldsymbol{\Delta},q_c} = \begin{bmatrix} c_{\Delta 1,q_c} \\ \vdots \\ c_{\Delta n,q_c} \end{bmatrix}. \tag{7}
$$

(8) can be obtained as

$$
\boldsymbol{Y}_{\boldsymbol{\Delta}}(k) = \boldsymbol{\Phi}_{\boldsymbol{\Delta}}(k)\boldsymbol{\theta}_{\boldsymbol{\Delta}} + \boldsymbol{e}_{\boldsymbol{\Delta}}(k) \tag{8}
$$

where $\boldsymbol{\Phi}_{\boldsymbol{\Delta}}(k)$ is the matrix composed of the model input and output data shown as

$$
\boldsymbol{\Phi}_{\boldsymbol{\Delta}}(k) = \begin{bmatrix} \boldsymbol{y}_{\boldsymbol{\Delta 1}}(k) & F\left(\boldsymbol{U}_{\boldsymbol{\Delta}}(k)\right) & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ \boldsymbol{y}_{\boldsymbol{\Delta n}}(k) & \cdots & 0 & F\left(\boldsymbol{U}_{\boldsymbol{\Delta}}(k)\right) \end{bmatrix}. \tag{9}
$$

In (8) - (9), the specific forms of $\boldsymbol{y}_{\boldsymbol{\Delta}i}(k)$ and $F(\boldsymbol{U}_{\boldsymbol{\Delta}}(k))$ are as follows

$$
\begin{aligned}
\boldsymbol{y}_{\boldsymbol{\Delta}i}(k) &= \left[-y_{\Delta i}(k-1) \quad \cdots \quad -y_{\Delta i}(k-n_a)\right], \\
F\left(\boldsymbol{U}_{\boldsymbol{\Delta}}(k)\right) &= \left[F_1\left(\boldsymbol{u}_{\boldsymbol{\Delta}1}(k)\right) \quad \cdots \quad F_n\left(\boldsymbol{u}_{\boldsymbol{\Delta}n}(k)\right)\right]^T, \\
\boldsymbol{e}_{\boldsymbol{\Delta}}(k) &= A_{\Delta}\left(q^{-1}\right)\boldsymbol{V}_{\boldsymbol{\Delta}}(k) \\
&= \left[e_{\Delta i}(k) \quad \cdots \quad e_{\Delta i}(k-n_c)\right]^T
\end{aligned} \tag{10}
$$

where

$$
\begin{aligned}
F_i\left(\boldsymbol{u}_{\boldsymbol{\Delta}i}(k)\right) &= \left[F_{i1}\left(u_{\Delta i}(k)\right) \quad \cdots \quad F_{in_d}\left(\boldsymbol{u}_{\boldsymbol{\Delta}i}(k)\right)\right]^T, \\
F_{ij}\left(u_i(k)\right) &= \left[f_{ij}\left(u_{\Delta i}\left(k-1-\tau_{b\Delta ij}\right)\right) \quad \cdots \right. \\
&\qquad \left. f_{ij}\left(u_{\Delta i}\left(k-n_b-\tau_{b\Delta ij}\right)\right)\right].
\end{aligned} \tag{11}
$$

$\boldsymbol{\theta}_{\boldsymbol{\Delta}}$ is the parameter vector of the Δ model shown as

$$
\begin{aligned}
\boldsymbol{\theta}_{\boldsymbol{\Delta}} &= \left[\boldsymbol{a}_{\boldsymbol{\Delta}} \quad \boldsymbol{b}_{\boldsymbol{\Delta}i} \quad \cdots \quad \boldsymbol{b}_{\boldsymbol{\Delta}n}\right]^T, \\
\boldsymbol{a}_{\boldsymbol{\Delta}} &= \left[a_{\Delta 1} \quad \cdots \quad a_{\Delta,n_a}\right], \\
\boldsymbol{b}_{\boldsymbol{\Delta}i} &= \left[\boldsymbol{b}_{\boldsymbol{\Delta}i,1} \quad \cdots \quad \boldsymbol{b}_{\boldsymbol{\Delta}i,n_b}\right], \\
\boldsymbol{b}_{\Delta ij} &= \left[b_{\Delta ij,1} \quad \cdots \quad b_{\Delta ij,n_d}\right], \\
b'_{\Delta ij,l} &= \left[b_{\Delta ij,1} \times d_{il} \quad \cdots \quad b_{\Delta ij,n} \times d_{il}\right].
\end{aligned} \tag{12}
$$

$\boldsymbol{\theta}_{\boldsymbol{\Delta}}$ can be obtained from the input data $\boldsymbol{U}_{\boldsymbol{\Delta}}$ and output data $\boldsymbol{Y}_{\boldsymbol{\Delta}}$ of the Δ model by using identification algorithm, such as recursive least squares and prediction error method. We assumed $d_{i1} = 1$ to ensure the identification uniqueness, then the $d_{il} = 1$ and $b_{\Delta ij,l}$ can be separated by the mean value method. $\hat{d}_{il}$ is the mean value of $\hat{d}_{il,q_b}$, calculated and used as the accurate parameters estimate

$$\hat{d}_{il} = \frac{1}{n_b}\sum_{q=1}^{n_b}\hat{d}_{il,q_b} \tag{13}$$

where

$$\begin{aligned}\hat{d}_{il,q_b} &= \frac{\hat{b}_{\Delta ij,l}(q_b)}{\hat{b}_{\Delta ij,q_b}},\\ \hat{b}_{\Delta ij,q_b} &= \hat{b}'_{\Delta ij,1}.\end{aligned} \tag{14}$$

Denote the plant model $\boldsymbol{G_c}(z^{-1})$ with respect to sampling time T as

$$\begin{aligned}\boldsymbol{G_c}\left(z^{-1}\right) &= \frac{\boldsymbol{B}\left(z^{-1}\right)}{A\left(z^{-1}\right)}\\ &= \frac{1}{A\left(z^{-1}\right)}\begin{bmatrix} B_{11}\left(z^{-1}\right) & \cdots & B_{1n}\left(z^{-1}\right)\\ \vdots & \ddots & \vdots\\ B_{n1}\left(z^{-1}\right) & \cdots & B_{nn}\left(z^{-1}\right)\end{bmatrix}\end{aligned} \tag{15}$$

where $A(z^{-1})$ and $B_{ij}(z^{-1})$ are the $n \times n$ denominator and numerator of the transfer function $\boldsymbol{G_c}(z^{-1})$, respectively.

In (15), $A(z^{-1})$ and $B_{ij}(z^{-1})$ are separately expressed as

$$\begin{aligned}A(z^{-1}) &= 1 + a_1 z^{-1} + \cdots + a_{n_a} z^{-n_a},\\ B_{ij}(z^{-1}) &= b_{ij,1} z^{-1-\tau_{bij}} + \cdots + b_{ij,n_b} z^{-n_b-\tau_{bij}}\end{aligned} \tag{16}$$

where parameters $a_1 \cdots a_{n_a}$ are the parameters of the $A(z^{-1})$. $b_{ij,1} \cdots b_{ij,n_b}$ are the parameters of $B_{ij}(z^{-1})$. τ_{bij} is the time delay of the $B_{ij}(z^{-1})$. To achieve the relationship between the T model and the Δ model, the Multi-Input Multi-Output (MIMO) system can be divided into Multiple-Input and Single-Output (MISO) subsystems shown as

$$\begin{aligned}y_{\Delta i}(k) &= -\sum_{q_a=1}^{n_a} a_{\Delta,q_a} y_{\Delta i}(k-q_a)\\ &\quad + \sum_{j=1}^{n}\sum_{q_b=1}^{n_b} b_{\Delta ij,q_b}\zeta_{\Delta j}(k-q_b-\tau_{b\Delta i}).\end{aligned} \tag{17}$$

Referring to (17), each subsystem can be converted into the state-space model structure

$$\begin{aligned}\boldsymbol{X_i}(k) &= \boldsymbol{A}\boldsymbol{X_i}(k-1) + \boldsymbol{B_i}\boldsymbol{\zeta_\Delta}(k-1-\tau_{b\Delta i}),\\ y_{\Delta i}(k) &= \boldsymbol{C}\boldsymbol{X_i}(k)\end{aligned} \tag{18}$$

where the A, B_i, and C are as follows

$$\begin{aligned}\boldsymbol{A} &= \begin{bmatrix} -a_{\Delta,1} & 1 & 0 & 0\\ -a_{\Delta,2} & 0 & \ddots & 0\\ \vdots & \vdots & 0 & 1\\ -a_{\Delta,n_{max}} & 0 & \cdots & 0\end{bmatrix},\\ \boldsymbol{B}_i &= \begin{bmatrix} b_{\Delta i,1} & \cdots & b_{\Delta in,1}\\ \vdots & \ddots & \vdots\\ b_{\Delta i,n_{max}} & \cdots & b_{\Delta in,n_{max}}\end{bmatrix},\\ \boldsymbol{C} &= \begin{bmatrix}1 & 0 & \cdots & 0\end{bmatrix},\end{aligned}$$

$$\boldsymbol{\tau_{b\Delta i}} = \begin{bmatrix}\tau_{b\Delta 1}\\ \vdots\\ \tau_{b\Delta in}\end{bmatrix} \tag{19}$$

where $n_{max} = \max(n_a, n_b)$. When $n_{max} = n_b$, $a_{\Delta,q_a} = 0$. When $n_{max} = n_a$, $b_{\Delta ij,q_b} = 0$.

Referring to (18), we can get the following expression

$$\boldsymbol{X_i}(k) = \boldsymbol{A^p}\boldsymbol{X_i}(k-p) + \sum_{t=0}^{p-1}\boldsymbol{A^t}\boldsymbol{B_i}\boldsymbol{\zeta_\Delta}(k-1-t-\tau_{b\Delta i}). \tag{20}$$

Referring to (1), (6) - (7), $\boldsymbol{\zeta_\Delta}(k-1-t-\tau_{b\Delta i})$ can be obtained as

$$\begin{aligned}\boldsymbol{\zeta_\Delta}(k-1-t-\tau_{b\Delta i}) &= \begin{bmatrix}F_1(\boldsymbol{u}_1(k-1-t))\\ \vdots\\ F_n(\boldsymbol{u}_n(k-1-t))\end{bmatrix}\\ &= \begin{bmatrix}F_1(\boldsymbol{u}_1(k-p))\\ \vdots\\ F_n(\boldsymbol{u}_n(k-p))\end{bmatrix}\\ &= \boldsymbol{\zeta_\Delta}(k-p-\tau_{b\Delta i}).\end{aligned} \tag{21}$$

Referring to (17) - (21), we can obtain that

$$\begin{aligned}\boldsymbol{X_i}(k) &= \boldsymbol{A^p}\boldsymbol{X_i}(k-p) + \sum_{t=0}^{p-1}\boldsymbol{A^t}\boldsymbol{B_i}\boldsymbol{\zeta_\Delta}(k-1-t-\tau_{b\Delta i}),\\ y_{\Delta i}(k) &= \boldsymbol{C}\boldsymbol{X_i}(k).\end{aligned} \tag{22}$$

Due to $q^{-p} = z^{-1}$ and $k\Delta = mT$, the $(k-p)\Delta = (m-1)T$. The model $G_i(z^{-1})$ of the subsystem obtained from (22) is

$$G_i(z^{-1}) = \boldsymbol{C}(\boldsymbol{I} - \boldsymbol{A^p}z^{-1})\sum_{t=0}^{p-1}\boldsymbol{A^t}\boldsymbol{B_i}z^{-\frac{\tau_{b\Delta i}}{p}}. \tag{23}$$

Therefore, the $A(z^{-1})$ and $\boldsymbol{B_i}(z^{-1})$ can be expressed as

$$\begin{cases} A\left(z^{-1}\right) = \det\left(\boldsymbol{I} - \boldsymbol{A}^p z^{-1}\right)\\ \boldsymbol{B}_i\left(z^{-1}\right) = \text{Cadj}\left(\boldsymbol{I} - \boldsymbol{A}^p z^{-1}\right)\sum_{t=0}^{p-1}\boldsymbol{A}^t\boldsymbol{B_i}z^{-\frac{\tau_{bs}}{p}}.\end{cases} \tag{24}$$

III. VARIATIONAL BAYESIAN (VB) METHOD FOR NONLINEAR MULTIVARIABLE OVER-SAMPLING CLOSED-LOOP STRUCTURE

The $\boldsymbol{V_\Delta}(k)$ is a white noise with zero mean and the variance δ^{-1}, $\boldsymbol{\theta_\Delta}$ is the normal distribution with the variance λ, then the probability density of $\boldsymbol{\theta_\Delta}$ can be expressed as

$$P(\boldsymbol{\theta_\Delta} \mid \lambda) = N\left(0, \lambda\boldsymbol{I}_{\dim(\boldsymbol{\theta_\Delta})}\right). \tag{25}$$

The $\boldsymbol{V_\Delta}(k)$ is the normal distribution, assuming the variance λ is the Gamma distribution, then the probability density of λ can be expressed as

$$P(\delta \mid \alpha, \beta) = \text{gamma}(\alpha, \beta\boldsymbol{I}_{n\times n}) \tag{26}$$

where α is the shape parameter and β is the scale parameter.

Referring to (25) - (26), the prior probability distribution of $\boldsymbol{\Theta} = \left\{\boldsymbol{\theta}_{\boldsymbol{\Delta}}, \delta^{-1}\right\}$ can be expressed as

$$P(\boldsymbol{\Theta}) = P(\boldsymbol{\theta}_{\boldsymbol{\Delta}} \mid \lambda)P(\delta \mid \alpha, \beta). \tag{27}$$

Through the Variational Bayesian (VB) method, the posterior probability distribution of $\boldsymbol{\Theta}$ can be expressed as

$$\begin{aligned} &F(Q(\boldsymbol{\Theta})) \\ &= \int Q(\boldsymbol{\Theta}) \log P(\boldsymbol{Y}_{\boldsymbol{\Delta}} \mid \boldsymbol{\Theta}) d\boldsymbol{\Theta} \\ &\quad + \int Q(\boldsymbol{\Theta}) \log \frac{P(\boldsymbol{\Theta})}{Q(\boldsymbol{\Theta})} d\boldsymbol{\Theta} \\ &= \int Q(\boldsymbol{\Theta}) \log P(\boldsymbol{Y}_{\boldsymbol{\Delta}} \mid \boldsymbol{\Theta}) d\boldsymbol{\Theta} \\ &\quad + \int Q(\boldsymbol{\Theta}) \log P(\boldsymbol{\Theta}) d\boldsymbol{\Theta} - \int Q(\boldsymbol{\Theta}) \log Q(\boldsymbol{\Theta}) d\boldsymbol{\Theta}. \end{aligned} \tag{28}$$

By taking the first-order partial derivative of (28) with respect to $\boldsymbol{\theta}_{\boldsymbol{\Delta}}$, referring to (27) the posterior probability distribution $Q(\boldsymbol{\theta}_{\boldsymbol{\Delta}})$ can be achieved as

$$\begin{aligned} &Q(\boldsymbol{\theta}_{\boldsymbol{\Delta}}) \\ &= P(\boldsymbol{Y}_{\boldsymbol{\Delta}} \mid \boldsymbol{\Theta})P(\boldsymbol{\Theta}) \\ &= P(\boldsymbol{Y}_{\boldsymbol{\Delta}} \mid \boldsymbol{\Theta})P(\boldsymbol{\theta}_{\boldsymbol{\Delta}} \mid \lambda)P(\delta \mid \alpha, \beta) \\ &= \frac{1}{C_\theta} \exp\left(-\frac{1}{2\lambda}\boldsymbol{\theta}_{\boldsymbol{\Delta}}^T \boldsymbol{I} \boldsymbol{\theta}_{\boldsymbol{\Delta}}\right) \\ &\quad \exp\left(\sum_{k=1}^{N} -\frac{\delta\left(\boldsymbol{Y}_{\boldsymbol{\Delta}}(k) - \boldsymbol{\Phi}_{\boldsymbol{\Delta}}^T(k)\boldsymbol{\theta}_{\boldsymbol{\Delta}}\right)^2}{2}\right) \\ &= \frac{1}{C_\theta} \exp\left\{ \begin{aligned} &-\tfrac{1}{2}\boldsymbol{\theta}_{\boldsymbol{\Delta}}{}^T \left[\lambda^{-1}\boldsymbol{I} + \sum_{k=1}^{N} \boldsymbol{\Phi}_{\boldsymbol{\Delta}}(k)\delta\boldsymbol{\Phi}_{\boldsymbol{\Delta}}{}^T(k)\right]\boldsymbol{\theta}_{\boldsymbol{\Delta}} \\ &+ \sum_{k=1}^{N} \delta \boldsymbol{Y}_{\boldsymbol{\Delta}}(k)\boldsymbol{\Phi}_{\boldsymbol{\Delta}}{}^T(k)\boldsymbol{\theta}_{\boldsymbol{\Delta}} \end{aligned} \right\} \end{aligned} \tag{29}$$

where $P(\boldsymbol{Y}_{\boldsymbol{\Delta}} \mid \boldsymbol{\Theta})$ is the normal distribution with mean value $\boldsymbol{\Phi}_{\boldsymbol{\Delta}}^T(k)\boldsymbol{\theta}_{\boldsymbol{\Delta}}$ and variance δ^{-1}. C_θ is a constant and

$$\begin{aligned} &C_\theta \\ &= 2\pi(\delta^{-1}\lambda)^{\frac{1}{2}} \left[\frac{\beta^\alpha \delta^{\alpha-1}}{\Gamma(\alpha)} \exp\left(-\beta\delta - \sum_{k=1}^{N} \frac{\delta}{2} \boldsymbol{Y}_{\boldsymbol{\Delta}}(k)\boldsymbol{Y}_{\boldsymbol{\Delta}}{}^T(k)\right)\right]^{-1}. \end{aligned}$$

Based on (29), it is obtained that $\boldsymbol{\theta}_{\boldsymbol{\Delta}}$ is a normal distribution, the mean value $\overline{\boldsymbol{\theta}}_{\boldsymbol{\Delta}}$ and the variance $\mathrm{Var}(\boldsymbol{\theta}_{\boldsymbol{\Delta}})$ are shown as follows

$$\overline{\boldsymbol{\theta}}_{\boldsymbol{\Delta}} = \mathrm{Var}(\boldsymbol{\theta}_{\boldsymbol{\Delta}}) \sum_{k=1}^{N} \delta \boldsymbol{Y}_{\boldsymbol{\Delta}}(k)\boldsymbol{\Phi}_{\boldsymbol{\Delta}}^T(k), \tag{30}$$

$$\mathrm{Var}(\boldsymbol{\theta}_{\boldsymbol{\Delta}}) = \left[\lambda^{-1}\boldsymbol{I} + \sum_{k=1}^{N} \boldsymbol{\Phi}_{\boldsymbol{\Delta}}(k)\delta\boldsymbol{\Phi}_{\boldsymbol{\Delta}}^T(k)\right]^{-1}, \tag{31}$$

$$\overline{\boldsymbol{\theta}}_{\boldsymbol{\Delta}}^2 = \left\langle \boldsymbol{\theta}_{\boldsymbol{\Delta}}^T \boldsymbol{\theta}_{\boldsymbol{\Delta}} \right\rangle_{Q(\boldsymbol{\theta}_{\boldsymbol{\Delta}})} = \mathrm{Var}\left(\boldsymbol{\theta}_{\boldsymbol{\Delta}}\right) + \boldsymbol{\theta}_{\boldsymbol{\Delta}}\boldsymbol{\theta}_{\boldsymbol{\Delta}}^T. \tag{32}$$

By taking the first-order partial derivative of (28) with respect to δ, referring to (27) the posterior probability distribution $Q(\delta)$ can be achieved as

$$\begin{aligned} &Q(\delta) \\ &= P(\boldsymbol{Y}_{\boldsymbol{\Delta}} \mid \boldsymbol{\Theta})P(\boldsymbol{\Theta}) \\ &= P(\boldsymbol{Y}_{\boldsymbol{\Delta}} \mid \boldsymbol{\Theta})P(\boldsymbol{\theta}_{\boldsymbol{\Delta}} \mid \lambda)P(\delta \mid \alpha, \beta) \\ &= \frac{1}{C_\delta} \exp\left[\sum_{k=1}^{N} \frac{1}{2}\ln(\delta) - \frac{1}{2}\delta\left(\boldsymbol{Y}_{\boldsymbol{\Delta}}(k) - \boldsymbol{\Phi}_{\boldsymbol{\Delta}}^T(k)\boldsymbol{\theta}_{\boldsymbol{\Delta}}\right)^2\right] \\ &\delta^{\alpha-1}\exp(-\beta\delta) \\ &= \frac{1}{C_\delta} \exp\left[\sum_{k=1}^{N} \frac{1}{2}\ln(\delta) - \frac{1}{2}\delta \begin{pmatrix} (\boldsymbol{Y}_{\boldsymbol{\Delta}}(k)\boldsymbol{Y}_{\boldsymbol{\Delta}}^T(k) \\ -\boldsymbol{Y}_{\boldsymbol{\Delta}}(k)\boldsymbol{\theta}_{\boldsymbol{\Delta}}^T\boldsymbol{\Phi}_{\boldsymbol{\Delta}}(k) \\ -\boldsymbol{\Phi}_{\boldsymbol{\Delta}}^T(k)\boldsymbol{\theta}_{\boldsymbol{\Delta}}\boldsymbol{Y}_{\boldsymbol{\Delta}}^T(k) \\ +\boldsymbol{\Phi}_{\boldsymbol{\Delta}}(k)^T\overline{\boldsymbol{\theta}}_{\boldsymbol{\Delta}}^2\boldsymbol{\Phi}_{\boldsymbol{\Delta}}^T(k)) \end{pmatrix}\right] \\ &\delta^{\alpha-1}\exp(-\beta\delta) \\ &= \frac{1}{C_\delta}\delta^{\alpha+\frac{1}{2}N-1} \\ &\quad \exp\left\{-\delta\left[\beta\boldsymbol{I} + \frac{1}{2}\sum_{k=1}^{N} \begin{pmatrix} (\boldsymbol{Y}_{\boldsymbol{\Delta}}(k)\boldsymbol{Y}_{\boldsymbol{\Delta}}^T(k) \\ -\boldsymbol{Y}_{\boldsymbol{\Delta}}(k)\boldsymbol{\theta}_{\boldsymbol{\Delta}}^T\boldsymbol{\Phi}_{\boldsymbol{\Delta}}(k) \\ -\boldsymbol{\Phi}_{\boldsymbol{\Delta}}^T(k)\boldsymbol{\theta}_{\boldsymbol{\Delta}}\boldsymbol{Y}_{\boldsymbol{\Delta}}^T(k) \\ +\boldsymbol{\Phi}_{\boldsymbol{\Delta}}(k)^T\overline{\boldsymbol{\theta}}_{\boldsymbol{\Delta}}^2\boldsymbol{\Phi}_{\boldsymbol{\Delta}}^T(k)) \end{pmatrix}\right]\right\} \end{aligned} \tag{33}$$

where C_δ is a constant and $C_\delta = 2\pi(\lambda)^{\frac{1}{2}}[\frac{\beta^\alpha}{\Gamma(\alpha)}\exp(-\frac{1}{2\lambda}\boldsymbol{\theta}_{\boldsymbol{\Delta}}^T \boldsymbol{I}\boldsymbol{\theta}_{\boldsymbol{\Delta}})]^{-1}$.

(33) shows that δ is a Gamma distribution with

$$\bar{\delta} = (2\alpha + N)(2\beta\boldsymbol{I} + \gamma)^{-1} \tag{34}$$

where γ is expressed as

$$\gamma = \sum_{k=1}^{N} \begin{pmatrix} \boldsymbol{Y}_{\boldsymbol{\Delta}}(k)\boldsymbol{Y}_{\boldsymbol{\Delta}}^T(k) - \boldsymbol{Y}_{\boldsymbol{\Delta}}(k)\boldsymbol{\theta}_{\boldsymbol{\Delta}}^T\boldsymbol{\Phi}_{\boldsymbol{\Delta}}(k) \\ -\boldsymbol{\Phi}_{\boldsymbol{\Delta}}^T(k)\boldsymbol{\theta}_{\boldsymbol{\Delta}}\boldsymbol{Y}_{\boldsymbol{\Delta}}^T(k) + \boldsymbol{\Phi}_{\boldsymbol{\Delta}}^T(k)\overline{\boldsymbol{\theta}}_{\boldsymbol{\Delta}}^2\boldsymbol{\Phi}_{\boldsymbol{\Delta}}^T(k) \end{pmatrix}. \tag{35}$$

Referring to (30) - (32), (34), and (35), where k is the lable such as $\boldsymbol{U}_{\boldsymbol{\Delta}}(k)$, h is the iteration number, the recursive steps of the multivariate Variational Bayesian (VB) method are obtained as follows:

1.Initialization: when $k \leq 0$, define $\boldsymbol{\Phi}_{\boldsymbol{\Delta}}(k) = 0$, $\boldsymbol{U}_{\boldsymbol{\Delta}}(k) = 0$, and a non-negative number ε_0.

2.Define initial parameter Θ, the initial iteration number $h = 1$, and non-negative numbers λ, α, β.

3.Update $\overline{\boldsymbol{\theta}}_{\boldsymbol{\Delta}}^{h+1}$, $\mathrm{Var}\left(\boldsymbol{\theta}_{\boldsymbol{\Delta}}\right)^{h+1}$, and $\left\langle \boldsymbol{\theta}_{\boldsymbol{\Delta}}\boldsymbol{\theta}_{\boldsymbol{\Delta}}^T \right\rangle_{Q(\boldsymbol{\theta}_{\boldsymbol{\Delta}})}^{h+1}$ by (30) - (32).

4.Update $\bar{\delta}^{h+1}$ by (34) - (35).

5. If $\left\|\overline{\boldsymbol{\theta}}_{\boldsymbol{\Delta}}^{h+1} - \overline{\boldsymbol{\theta}}_{\boldsymbol{\Delta}}^{h}\right\| \leq \varepsilon_0$, then $\overline{\boldsymbol{\theta}}_{\boldsymbol{\Delta}}^{h+1}$ is the estimated value of $\boldsymbol{\theta}_{\boldsymbol{\Delta}}$. Otherwise, $h = h + 1$, and repeat step 3.

IV. AMPLITUDE-LIMITED VARIATIONAL BAYESIAN (ALVB) METHOD FOR NONLINEAR MULTIVARIABLE OVER-SAMPLING STRUCTURE

A. AMPLITUDE-LIMITED VARIATIONAL BAYESIAN (ALVB) METHOD FOR COLORED NOISE

When $\boldsymbol{V}_{\boldsymbol{\Delta}}(k)$ is a colored noise, the matrix $\boldsymbol{\Phi}_{\boldsymbol{\Delta}}(k)$ contains the unknown noise $\boldsymbol{e}_{\boldsymbol{\Delta}}(k)$. The estimate of the $\boldsymbol{e}_{\boldsymbol{\Delta}}(k)$ can be expressed as

$$\hat{\boldsymbol{e}}_{\boldsymbol{\Delta}}(k) = \boldsymbol{Y}_{\boldsymbol{\Delta}}(k) - \boldsymbol{\Phi}_{\boldsymbol{\Delta}}(k)\hat{\boldsymbol{\theta}}_{\boldsymbol{\Delta}}. \tag{36}$$

In the algorithm iteration, the estimated value $\hat{\boldsymbol{e}}_{\boldsymbol{\Delta}}$ might not converge caused by the huge difference between the initial value $\bar{\boldsymbol{\theta}}_{\boldsymbol{\Delta}}$ and the true value $\boldsymbol{\theta}_{\boldsymbol{\Delta}}$. Therefore, define the $\hat{\boldsymbol{e}}_{\boldsymbol{\Delta}}$ satisfy an amplitude limiting rule

$$\hat{\boldsymbol{e}}_{\Delta}(k) = \begin{bmatrix} \hat{\boldsymbol{e}}_{\Delta 1}(k) \\ \vdots \\ \hat{\boldsymbol{e}}_{\Delta n}(k) \end{bmatrix} \tag{37}$$

where

$$\hat{e}_{\Delta i}(k) = \begin{cases} \kappa, & \hat{e}_{\Delta i}(k) \geq \kappa \\ \hat{e}_{\Delta i}(k), & -\kappa < \hat{e}_{\Delta i}(k) < \kappa \\ -\kappa, & \hat{e}_{\Delta i}(k) \leq -\kappa. \end{cases} \tag{38}$$

In (38), κ is defined as a large positive number.

According to (30) - (38), where k is the lable such as $\boldsymbol{U}_{\boldsymbol{\Delta}}(k)$, h is the iteration number, the recursive steps of the multivariable Amplitude-Limited Variational Bayesian (ALVB) method for colored noise are obtained as follows:

1. When $k \leq 0$, define $\boldsymbol{\Phi}_{\boldsymbol{\Delta}}(k) = 0$, $\boldsymbol{U}_{\boldsymbol{\Delta}}(k) = 0$, and a non-negative number ε_0.
2. Define initial parameter Θ, the initial iteration number $h = 1$, and non-negative numbers λ, α, β.
3. $\hat{\boldsymbol{e}}_{\boldsymbol{\Delta}}$ is estimated by (36) - (38).
4. Update $\overline{\boldsymbol{\theta}}_{\boldsymbol{\Delta}}^{h+1}$, $\mathrm{Var}(\boldsymbol{\theta}_{\boldsymbol{\Delta}})^{h+1}$, and $\langle \boldsymbol{\theta}_{\boldsymbol{\Delta}}\boldsymbol{\theta}_{\boldsymbol{\Delta}}^T\rangle_{Q(\boldsymbol{\theta}_{\boldsymbol{\Delta}})}^{h+1}$ by (30) - (32).
5. Update $\bar{\delta}^{h+1}$ by (34) - (35).
6. If $\left\|\overline{\boldsymbol{\theta}}_{\boldsymbol{\Delta}}^{h+1} - \overline{\boldsymbol{\theta}}_{\boldsymbol{\Delta}}^{h}\right\| \leq \varepsilon_0$, then $\overline{\boldsymbol{\theta}}_{\boldsymbol{\Delta}}^{h+1}$ is the estimated value of $\boldsymbol{\theta}_{\boldsymbol{\Delta}}$. Otherwise, $h = h + 1$, and repeat step 4.

B. ERROR EXPRESSION OF NOISE ESTIMATION

ε is defined as the error between the true value $\boldsymbol{\theta}_{\boldsymbol{\Delta}}$ and the initial value $\overline{\boldsymbol{\theta}}_{\boldsymbol{\Delta}}^1$, meaning $\varepsilon = \boldsymbol{\theta}_{\boldsymbol{\Delta}} - \overline{\boldsymbol{\theta}}_{\boldsymbol{\Delta}}^1$. The Multi-input Multi-Output (MIMO) system can be divided into Multiple-Input and Single-Output (MISO) subsystems, $\hat{e}_{\Delta i}(k)$ can be estimated in each subsystem when $h = 1$. The relationship between $\hat{e}_{\Delta i}$ and $e_{\Delta i}$ satisfies

$$I_{\Delta ij}(k) = \begin{cases} 0, k \neq n_a + (n \times n_b \times n_d + n_c)(i-1) \\ \quad + (n_a + n_b + j) \\ 1, k = n_a + (n \times n_b \times n_d + n_c)(i-1) \\ \quad + (n_a + n_b + j), \end{cases} \tag{39}$$

$$\Delta e_{\Delta i}(k) = \hat{e}_{\Delta i}(k) - e_{\Delta i}(k). \tag{40}$$

The relationship between $\Phi_{\Delta i}(k)$ and $\hat{\Phi}_{\Delta i}(k)$ satisfies

$$\hat{\Phi}_{\Delta i}(k) = \Phi_{\Delta i}(k) + \sum_{q_c=1}^{n_c} \Delta e_{\Delta i}(k-1) I_{\Delta ij}. \tag{41}$$

Referring to (8), (37), and (39) - (41), when $k = 1$, (40) can be expressed as

$$\begin{aligned} \hat{e}_{\Delta i}(1) &= y_{\Delta_i}(1) - \Phi_{\Delta i}(1)\bar{\boldsymbol{\theta}}_{\boldsymbol{\Delta}}^1 \\ &= y_{\Delta_i}(1) - \Phi_{\Delta i}(1)(\boldsymbol{\theta}_{\boldsymbol{\Delta}} - \varepsilon) \\ &= e_{\Delta i}(1) + \Phi_{\Delta i}(1)\varepsilon, \end{aligned} \tag{42}$$

$$\Delta e_{\Delta i}(1) = \Phi_{\Delta i}(1)\varepsilon. \tag{43}$$

Referring to (8), (37), and (42) - (43), when $k = 2$, (40) can be expressed as

$$\begin{aligned} &\hat{e}_{\Delta i}(2) \\ &= y_{\Delta i}(2) - \hat{\Phi}_{\Delta i}(2)\bar{\boldsymbol{\theta}}_{\boldsymbol{\Delta}}^1 \\ &= y_{\Delta i}(2) - \hat{\Phi}_{\Delta i}(2)(\boldsymbol{\theta}_{\boldsymbol{\Delta}} - \varepsilon) \\ &= y_{\Delta i}(2) - (\Phi_{\Delta i}(2) + \Delta e_{\Delta i}(1) I_{\Delta i1})(\boldsymbol{\theta}_{\boldsymbol{\Delta}} - \varepsilon) \\ &= e_{\Delta}(2) + \Phi_{\Delta i}(1)\varepsilon^2 I_{\Delta i1} + (\Phi_{\Delta i}(2) - c_{\Delta i1}\Phi_{\Delta i}(1))\varepsilon, \end{aligned} \tag{44}$$

$$\begin{aligned} &\Delta e_{\Delta i}(2) \\ &= \Phi_{\Delta}(1)\varepsilon^2 I_{\Delta i1} + (\Phi_{\Delta i}(2) - c_{\Delta i1}\Phi_{\Delta i}(1))\varepsilon \\ &= \Phi_{\Delta i}(1)\varepsilon^2 I_{\Delta i1} + R_{\Delta i}(2). \end{aligned} \tag{45}$$

Therefore, based on iteration above, (40) can be expressed as

$$\begin{aligned} \Delta e_{\Delta i}(k) &= \Phi_{\Delta i}(1)\varepsilon\,(I_{\Delta i1}\varepsilon)^{k-1} + R_{\Delta i}(k) \\ &= \Phi_{\Delta i}(1)\varepsilon^k\,(I_{\Delta i1})^{k-1} + R_{\Delta i}(k). \end{aligned} \tag{46}$$

(46) shows that if the initial error $\varepsilon_{\text{in}} > 1$ and when $k \to \infty$, $e_{\Delta i}(k)$ might not converge, shown as

$$\begin{aligned} \lim_{k\to\infty} |\Delta e_{\Delta i}(k)| &\geq |\Phi_{\Delta i}(1)\varepsilon\|\varepsilon_{in}|^{k-1} - |R_{\Delta i}(k)| \\ &\approx |\Phi_{\Delta i}(1)\varepsilon|\,|\varepsilon_{in}|^{k-1} = +\infty. \end{aligned} \tag{47}$$

Referring to (47), $\hat{e}_{\Delta i}(k)$, $\hat{\Phi}_{\Delta i}(k)$, and $\bar{\boldsymbol{\theta}}_{\Delta}$ might not converge to the achieved accuracy during the whole algorithm iteration. To solve the problem above, $\hat{e}_{\Delta i}(k)$ would be defined in a certain range, which ensures the algorithm operate well, the $\bar{\boldsymbol{\theta}}_{\boldsymbol{\Delta}}^h$ and $\hat{e}_{\Delta i}(k)$ converge to the achieved accuracy.

V. SIMULATIONS

The benchtop neutralization system studied by Henson and Seborg is taken as an example for simulation. The base stream Q_1 is *NaOH* of 0.003 mol/L, the buffer stream Q_2 is $NaHNO_3$ of 0.03 mol/L, and the acid stream Q_3 is HNO_3 of 0.003 mol/L. Lakshminarayanan proposed that the process above can be expressed by a 2 × 2 Hammerstein model. The outputs of the model are the liquid level $h(y_1)$ and the PH value $PH(y_2)$, the inputs are the base stream Q_1 (u_1) and the acid stream Q_3 (u_2). The process above operated in the over-sampling closed-loop structure, and the controllers are K_1 and K_2. Experiments for Model-1 and Model-2 are conducted. The specific experimental parameters are as follows:

$$\begin{aligned} \zeta_1(k) &= u_1(k) + 0.2735u_1^2(k) + 2.347u_1^3(k), \\ \zeta_2(k) &= u_1(k) - 2.0381u_2^2(k) + 10.869u_2^3(k). \end{aligned} \tag{48}$$

Model-1 is (49) as shown at the bottom of the next page.

Model-2 is (50) and (51) as shown at the bottom of the next page.

Example of white noise:

$$K_1\left(z^{-1}\right) = \frac{0.1 - 0.06\, z^{-1}}{1 + 0.6\, z^{-1}},$$
$$K_2\left(z^{-1}\right) = \frac{-0.012 - 0.06\, z^{-1}}{1 - 0.9\, z^{-1}},$$
$$H\left(z^{-1}\right) = \begin{bmatrix} \dfrac{1}{1 - 1.8656z^{-1} + 0.8717z^{-2}} \\ \dfrac{1}{1 - 1.8656z^{-1} + 0.8717z^{-2}} \end{bmatrix}. \quad (52)$$

Example of colored noise:

$$K_1\left(z^{-1}\right) = 0.1 + 0.06z^{-1},$$
$$K_2\left(z^{-1}\right) = -0.012 - 0.06z^{-1},$$
$$H\left(z^{-1}\right) = \begin{bmatrix} \dfrac{1 + 1.5z^{-1} + 2z^{-2}}{1 - 1.8656z^{-1} + 0.8717z^{-2}} \\ \dfrac{1 + z^{-1} + 3z^{-2}}{1 - 1.8656z^{-1} + 0.8717z^{-2}} \end{bmatrix}. \quad (53)$$

A. VARIATIONAL BAYESIAN (VB) METHOD IN MULTIVARIABLE NONLINEAR OVER-SAMPLING CLOSED-LOOP STRUCTURE FOR WHITE NOISE

When $\boldsymbol{V}_{\boldsymbol{\Delta}}(k)$ is white noise, Variational Bayesian (VB) method and Recursive least squares (RLS) are separately used to achieve the multivariable nonlinear over-sampling closed-loop structure model parameters. Here first Model-1 is taken as the experimental model. The relative errors of RLS and VB are shown in Fig. 2 and Fig. 3, respectively. The parameter estimates of the two algorithms are shown in Tables 1 and 2. The probability distributions of the parameter estimates are shown in Fig. 4 and Fig. 5. The above identification experiments are repeated for Model-1 and Model-2. The statistical results are shown in Fig. 6.

Fig. 2 and Fig. 3 show that the identification experiments of multivariate nonlinear over-sampling closed-loop structure model can achieve the accuracy of the parameter estimates, without the extra excitations and the controller order is lower than the model order. The structure widens the identifiability of multivariable nonlinear models. Fig. 6 and Tables 1 and 2 shows that the experiment results of the multivariate nonlinear over-sampling closed-loop structure model, the error of VB is smaller than that of RLS under the same conditions. Fig. 4 and Fig. 5 show that the probability density reaches the maximum at the true value $\boldsymbol{\theta}_{\boldsymbol{\Delta}}$.

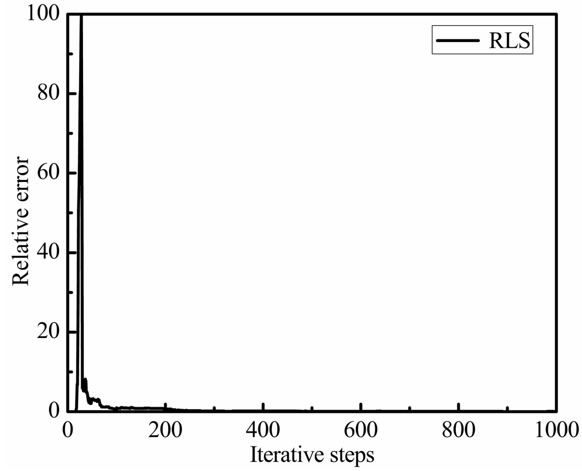

FIGURE 2. Relative error of RLS.

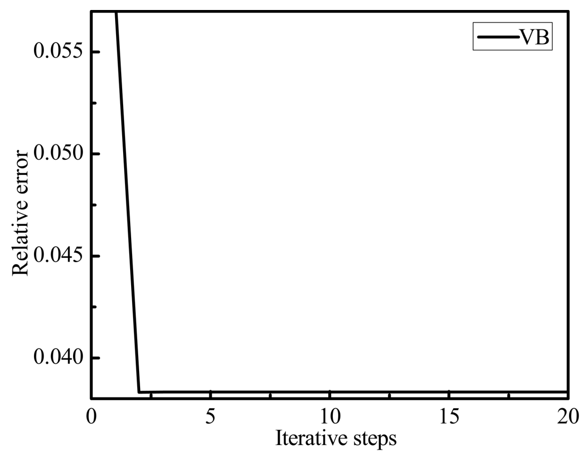

FIGURE 3. Relative error of VB.

B. AMPLITUDE-LIMITED VARIATIONAL BAYESIAN (ALVB) METHOD IN MULTIVARIABLE NONLINEAR OVER-SAMPLING CLOSED-LOOP STRUCTURE FOR COLORED NOISE

When $\boldsymbol{V}_{\boldsymbol{\Delta}}(k)$ is colored noise, Amplitude-Limited Variational Bayesian (ALVB) method, Variational Bayesian (VB)

$$\boldsymbol{G_c}\left(z^{-1}\right) = \begin{bmatrix} \dfrac{0.0699z^{-1} - 0.0632z^{-2}}{1 - 1.8656z^{-1} + 0.8717z^{-2}} & \dfrac{0.0069z^{-2}}{1 - 1.8656z^{-1} + 0.8717z^{-2}} \\ \dfrac{0.0042z^{-2}}{1 - 1.8656z^{-1} + 0.8717z^{-2}} & \dfrac{-0.1748q^{-1} + 0.1679q^{-2}}{1 - 1.8656z^{-1} + 0.8717z^{-2}} \end{bmatrix}. \quad (49)$$

$$\boldsymbol{G_c}\left(z^{-1}\right) = \begin{bmatrix} \dfrac{0.0599z^{-1} - 0.0732z^{-2}}{1 - 1.8656z^{-1} + 0.8717z^{-2}} & \dfrac{0.0069z^{-2}}{1 - 1.8656z^{-1} + 0.8717z^{-2}} \\ \dfrac{0.0042z^{-2}}{1 - 1.8656z^{-1} + 0.8717z^{-2}} & \dfrac{-0.1848z^{-1} + 0.1579z^{-2}}{1 - 1.8656z^{-1} + 0.8717z^{-2}} \end{bmatrix}. \quad (50)$$

$$R(k) = \begin{bmatrix} 0 \\ 0 \end{bmatrix}, p = 4. \quad (51)$$

TABLE 1. $\theta_{\Delta 1}$ **With the White Noise as the Output Noise.**

Algorithm	$a_{\Delta 1}$	$a_{\Delta 2}$	$b_{\Delta 11,1}$	$b_{\Delta 11,2}$	$b_{\Delta 12,1}$	$b_{\Delta 12,2}$	$d_{\Delta 1,1}$	$d_{\Delta 1,2}$	$d_{\Delta 1,3}$	$\sigma(\%)$
RLS	−1.8489	0.8552	−0.0244	−0.0515	−0.0195	0.0286	1	0.2723	2.345	5.33
VB	−1.8496	0.8559	0.0208	−0.0947	−0.0210	0.0294	1	0.2734	2.348	3.83
Truth value	−1.8656	0.8717	0.0069	−0.0632	0	0.0069	1	0.2735	2.347	——

TABLE 2. $\theta_{\Delta 2}$ **With the Colored Noise as the Output Noise.**

Algorithm	$a_{\Delta 1}$	$a_{\Delta 2}$	$b_{\Delta 11,1}$	$b_{\Delta 11,2}$	$b_{\Delta 12,1}$	$b_{\Delta 12,2}$	$d_{\Delta 1,1}$	$d_{\Delta 1,2}$	$d_{\Delta 1,3}$	$\sigma(\%)$
RLS	−1.8489	0.8552	0.0234	−0.0045	−0.1529	0.1415	1	−2.0380	10.871	5.33
VB	−1.8496	0.8559	0.0228	0.0002	−0.1553	0.1443	1	−2.0381	10.868	3.83
Truth value	−1.8656	0.8717	0	0.0042	−0.1748	0.1679	1	−2.0381	10.869	——

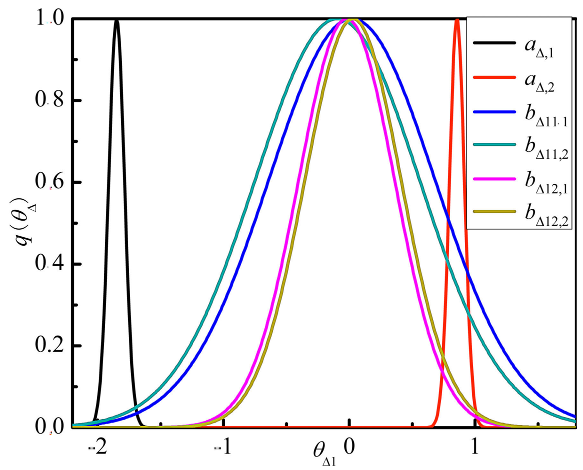

FIGURE 4. Probability distribution of parameter $\theta_{\Delta 1}$.

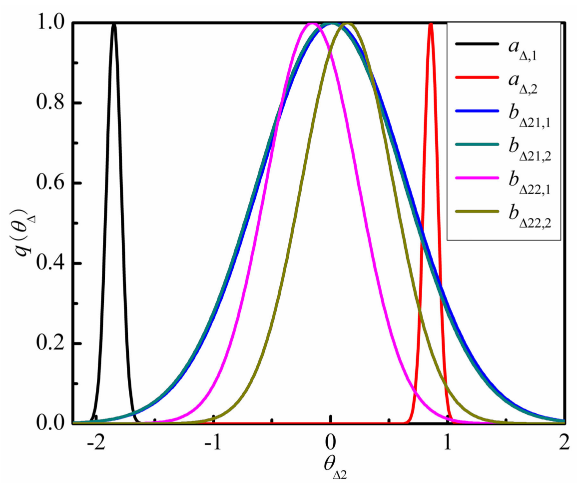

FIGURE 5. Probability distribution of parameter $\theta_{\Delta 2}$.

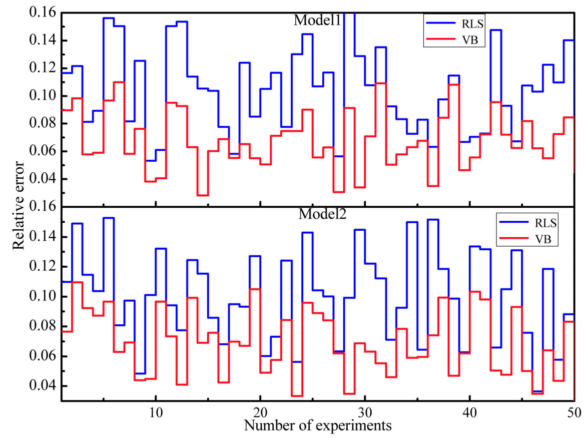

FIGURE 6. Relative errors of repetitive experiments for Model-1 and Model-2.

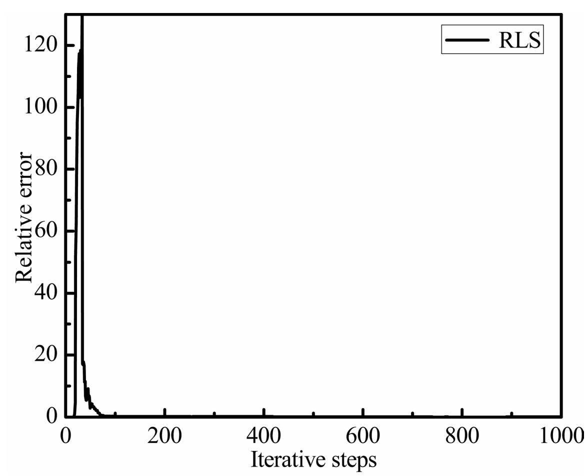

FIGURE 7. Relative error of RLS.

method, and Recursive least squares (RLS) are separately used to achieve the multivariable nonlinear over-sampling closed-loop structure model parameters. Here first Model-1 is taken as the experimental model. The relative errors of RLS and ALVB are shown in Fig. 7 and Fig. 8, respectively. The parameter estimates of the three algorithms are shown in Table 3 and Table 4. The probability distributions of the parameter estimates are shown in Fig. 9 and Fig. 10. The above identification experiments are repeated for Model-1 and Model-2. The statistical results are shown in Fig. 11. Table 3 and Table 4 show the ALVB method overcomes the shortcomings that the traditional VB method might not converge, which satisfies the situation that the output is colored noise. Fig. 7, Fig.8, Table 3, Table 4, and Fig. 11 show

TABLE 3. $\theta_{\Delta 1}$ **With the White Noise as the Output Noise.**

Algorithm	$a_{\Delta 1}$	$a_{\Delta 2}$	$b_{\Delta 11,1}$	$b_{\Delta 11,2}$	$b_{\Delta 12,1}$	$b_{\Delta 12,2}$	$d_{\Delta 1,1}$	$d_{\Delta 1,2}$	$d_{\Delta 1,3}$	$\sigma(\%)$
RLS	-1.9197	0.9306	0.0629	-0.0682	0.0461	-0.0493	1	0.2722	2.345	6.16
ALVB	-1.8700	0.8774	0.0796	-0.0659	0.0236	-0.0058	1	0.2735	2.346	1.18
VB	∞	∞	∞	∞	∞	∞	∞	∞	∞	∞
Truth value	-1.8656	0.8717	0.0069	-0.0632	0	0.0069	1	0.2735	2.347	——

TABLE 4. $\theta_{\Delta 2}$ **With the Colored Noise as the Output Noise.**

Algorithm	$a_{\Delta 1}$	$a_{\Delta 2}$	$b_{\Delta 11,1}$	$b_{\Delta 11,2}$	$b_{\Delta 12,1}$	$b_{\Delta 12,2}$	$d_{\Delta 1,1}$	$d_{\Delta 1,2}$	$d_{\Delta 1,3}$	$\sigma(\%)$
RLS	-1.9179	0.9306	0.0036	0.0141	-0.1654	0.2358	1	-2.0379	10.869	6.16
ALVB	-1.8700	0.8774	-0.0027	0.0099	-0.1662	0.1912	1	-2.0380	10.870	1.18
VB	∞	∞	∞	∞	∞	∞	∞	∞	∞	∞
Truth value	-1.8656	0.8717	0	0.0042	-0.1748	0.1679	1	-2.0381	10.869	——

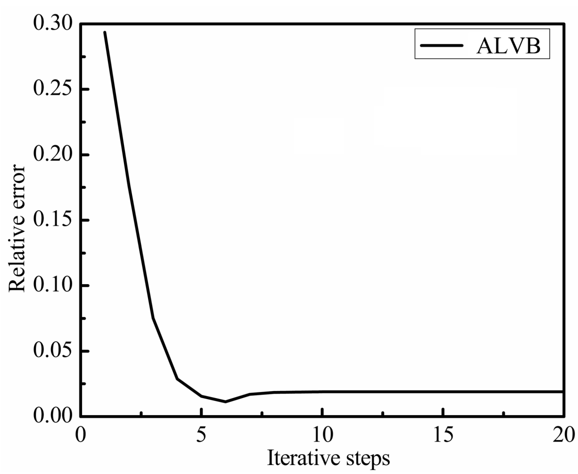

FIGURE 8. **Relative error of ALVB.**

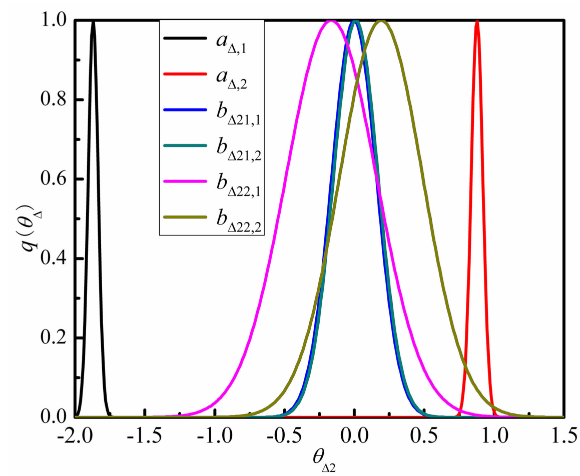

FIGURE 10. **Probability distribution of parameters** $\theta_{\Delta 2}$.

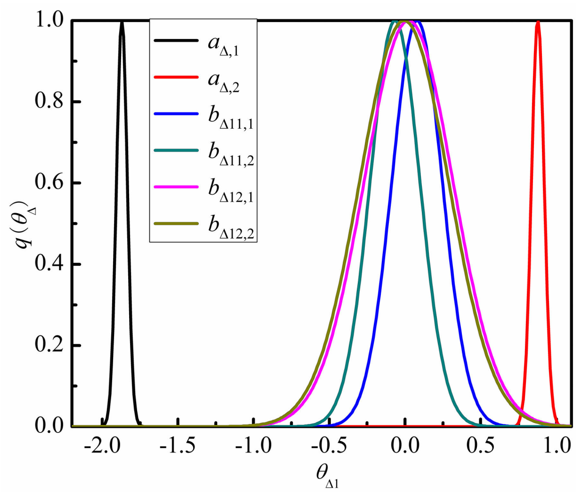

FIGURE 9. **Probability distribution of parameters** $\theta_{\Delta 1}$.

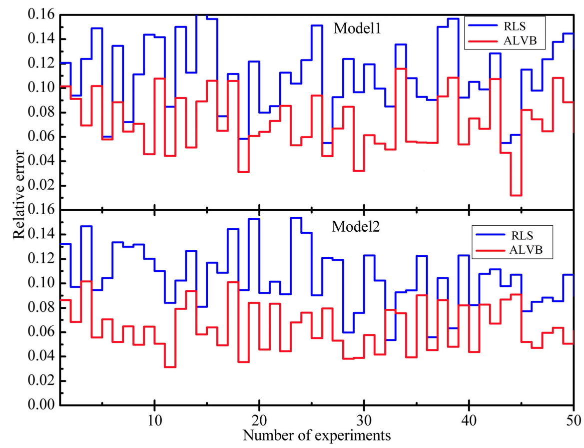

FIGURE 11. **Relative errors of repetitive experiments for Model-1 and Model-2.**

that the over-sampling closed-loop structure ALVB method not only does not need extra excitation but also achieve high parameter estimates accuracy compared with that of RLS. Fig. 4 and Fig. 5 show that the probability density reaches the maximum at the true value $\boldsymbol{\theta}_{\Delta 2}$.

VI. CONCLUSION

This article proposes a multivariable nonlinear over-sampling closed-loop structure model when the multivariable nonlinear traditional closed-loop structure model cannot be identifiable. A Variational Bayesian (VB) method based on the multivariable over-sampling closed-loop structure Hammerstein model is proposed, which improved the traditional VB method. Also, in this article, we propose an Amplitude-Limited Variational Bayesian (ALVB) method based on the multivariable nonlinear over-sampling closed-loop structure model which is applicable for colored noise. The simulations show that the VB method based on multivariable nonlinear over-sampling closed-loop structure model satisfy the identifiability, but also has a higher identification accuracy than RLS. Under the situation that the traditional VB method might not converge caused by the colord output noise, the ALVB method satisfy the convergence and has a higher accuracy advantage over the traditional VB method. Therefore, the VB and ALVB methods based on the multivariable nonlinear over-sampling closed-loop structure model are suitable for the large plant process identification.

REFERENCES

[1] M. V. Kothare, V. Balakrishnan, and M. Morari, "Robust constrained model predictive control using linear matrix inequalities," *Automatica*, vol. 32, no. 10, pp. 1361–1379, Oct. 1996.

[2] K. Li, J.-X. Peng, and G. W. Irwin, "A fast nonlinear model identification method," *IEEE Trans. Autom. Control*, vol. 50, no. 8, pp. 1211–1216, Aug. 2005.

[3] Q. B. Jin, R. G. Yang, and Z. Wang, "Description of unified nonlinear characteristics Hammerstein model and its identification method research," *J. Syst. Simul.*, vol. 26, no. 12, pp. 2887–2991, Dec. 2014.

[4] F. Wang, K. Y. Xing, and X. P. Xu, "Study on method for identification of Hammerstein model," *J. Syst. Simul.*, vol. 23, no. 6, pp. 1090–1092, Jun. 2011.

[5] Q. C. Wang and J. Z. Zhang, "Nonlinear predictive control for continuous stirred-tank reactor using Hammerstein model," *J. Nanjing Univ. Sci. Technol.*, vol. 34, no. 5, pp. 618–623, Oct. 2010.

[6] Z. Y. Zou, Y. Q. Guo, and Z. Z. Wang, "Nonlinear Hammerstein model predictive control strategy and its application in pH neutralization process," *J. Chem. Ind. Eng.*, vol. 63, no. 12, pp. 3965–3970, Dec. 2012.

[7] J. F. Zhao, X. Z. Ma, and S. Z. Zhao, "Improved Hammerstein system and its application in supercharged boilers," *J. Harbin Eng. Univ.*, vol. 34, no. 3, pp. 312–317, Mar. 2013.

[8] F. Ding, P. X. Liu, and J. Ding, "Iterative solutions of the generalized Sylvester matrix equations by using the hierarchical identification principle," *Appl. Math. Comput.*, vol. 197, no. 1, pp. 41–50, Mar. 2008.

[9] L. Xie, J. Ding, and F. Ding, "Gradient based iterative solutions for general linear matrix equations," *Comput. Math. Appl.*, vol. 58, no. 7, pp. 1441–1448, Oct. 2009.

[10] N. Haist, F. Chang, and R. Luus, "Nonlinear identification in the presence of correlated noise using a Hammerstein model," *IEEE Trans. Autom. Control*, vol. 18, no. 5, pp. 552–555, Oct. 1973.

[11] M. Verhaegen and D. Westwick, "Identifying MIMO Hammerstein systems in the context of subspace model identification methods," *Int. J. Control*, vol. 63, no. 2, pp. 331–349, Feb. 2007.

[12] I. Goethals, K. Pelckmans, J. A. K. Suykens, and B. De Moor, "Subspace identification of Hammerstein systems using least squares support vector machines," *IEEE Trans. Autom. Control*, vol. 50, no. 10, pp. 1509–1519, Oct. 2005.

[13] L. Vanbeylen, R. Pintelon, and J. Schoukens, *Blind maximum likelihood identification of Hammerstein systems*. London, U.K.: Pergamon, 2008.

[14] J. Wang, A. Sano, T. Chen, and B. Huang, "A blind approach to identification of Hammerstein systems," *Lect. Notes Control Inf. Sci.*, vol. 404, no. 2, pp. 293–312, Sep. 2010.

[15] D. H. Wu, "Identification method for nonlinear dynamic systems using Wiener neural network," *Control Theory Appl.*, vol. 26, no. 11, pp. 1192–1196, Nov. 2009.

[16] X. P. Xu, F. C. Qian, and D. Liu, "Method of system identification based on PSO algorithm," *J. Syst. Simul.*, vol. 20, no. 13, pp. 3525–3528. Jul. 2008.

[17] L. Ljung, *System Identification: Theory for the Users*. Beijing, China: Tsinghua Univ. Press, 2002.

[18] L. Ljung, I. Gustavsson, and T. Soderstrom, "Identification of linear, multivariable systems operating under linear feedback control," *IEEE Trans. Autom. Control*, vol. 19, no. 6, pp. 836–840, Dec. 1974.

[19] T. Soderstrom, L. Ljung, and I. Gustavsson, "Identifiability conditions for linear multivariable systems operating under feedback," *IEEE Trans. Autom. Control*, vol. 21, no. 6, pp. 837–840, Dec. 1976.

[20] M. Fang and Y. Zhu, "Reducing identification error using over-sampling technique," *IFAC-PapersOnLine*, vol. 48, no. 28, pp. 1142–1147, 2015.

[21] L. Sun, H. Ohmori, and A. Sano, "Output intersampling approach to direct closed-loop identification," *IEEE Trans. Autom. Control*, vol. 46, no. 12, pp. 1936–1941, Dec. 2001.

[22] J. Wang, T. Chen, and B. Huang, "Closed-loop identification via output fast sampling," *J. Process Control*, vol. 14, no. 5, pp. 555–570, Aug. 2004.

[23] L. Sun and A. Sano, "Cyclic spectral based approach to closed-loop identification," *IFAC Proc. Volumes*, vol. 38, no. 1, pp. 1161–1166, Jul. 2005.

[24] M. Fang, Y. Zhu, and H. Hjalmarsson, "On anti-aliasing filtering and over-sampling scheme in system identification," *Comput. Chem. Eng.*, vol. 106, pp. 572–581, Nov. 2017.

[25] L. Sun, W. Liu, and A. Sano, "Over-sampling approach to closed-loop identification," in *Proc. 36th IEEE Conf. Decis. Control*, vol. 2, pp. 1253–1258, Dec. 1997.

[26] L. Sun and A. Sano, "Output over-sampling approach to direct closed-loop identification and its performance," *IFAC Proc. Volumes*, vol. 42, no. 10, pp. 687–692, Jul. 2009.

[27] M. Fang and Y. Zhu, "Asymptotic variance expression in output over-sampling based closed-loop identification," *IFAC-PapersOnLine*, vol. 48, no. 28, pp. 110–115, 2015.

[28] Y. Lu, B. Huang, and S. Khatibisepehr, "A variational Bayesian approach to robust identification of switched ARX models," *IEEE Trans. Cybern.*, vol. 46, no. 12, pp. 3195–3208, Dec. 2016.

[29] Y. Zhao, A. Fatehi, and B. Huang, "Robust estimation of ARX models with time varying time delays using variational Bayesian approach," *IEEE Trans. Cybern.*, vol. 48, no. 2, pp. 532–542, Feb. 2018.

[30] J. Chen and Y. Liu, "Variational Bayesian-based iterative algorithm for ARX models with random missing outputs," *Circuits, Syst., Signal Process.*, vol. 37, no. 4, pp. 1594–1608, Apr. 2018.

BAOCHANG XU was born in Heilongjiang, China, in 1974. He received the Ph.D. degree in precision instrument and machinery from Beihang University, in 2005. From 2013 to 2014, he was a Visiting Scholar with the Department of Chemical Engineering, Texas A&M University. He is currently an Associate Professor of the China University of Petroleum, Beijing. His current research interests include the modeling and advanced control of complex systems and automatic control technology of drilling process.

LIKUN YUAN was born in Gansu, China, in 1992. She received the B.Eng. degree in automation from the China University of Petroleum, Beijing, in 2015, where she is currently pursuing the Ph.D. degree in control science and engineering. Her research interests include system identification, data mining, and deep learning.

YAXIN WANG was born in Xinjiang, China, in 1995. She received the B.Eng. degree in automation from the China University of Petroleum, Beijing, in 2017, where she is currently pursuing the Ph.D. degree in control science and engineering. Her research interests include system identification, data mining, and deep learning.