Received: 4 January 2023 | Revised: 30 May 2023 | Accepted: 1 July 2023

DOI: 10.1002/acs.3657

RESEARCH ARTICLE

WILEY

Expectation-maximization algorithm for bilinear state-space models with time-varying delays under non-Gaussian noise

Xinyue Wang | Junxia Ma | Weili Xiong

Key Laboratory of Advanced Process Control for Light Industry (Ministry of Education), School of Internet of Things Engineering, Jiangnan University, Wuxi, People's Republic of China

Correspondence
Weili Xiong, Key Laboratory of Advanced Process Control for Light Industry (Ministry of Education), School of Internet of Things Engineering, Jiangnan University, Wuxi 214122, People's Republic of China.
Email: greenpre@163.com

Funding information
National Natural Science Foundation of China, Grant/Award Number: 61773182

Summary
In this paper, the parameter identification of bilinear state-space model (SSM) in the presence of random outliers and time-varying delays is investigated. Under the basis of the observable canonical form of the bilinear model, the system output can be written as a regressive form, and a bilinear state observer is applied to estimate the unknown states. To eliminate the influence of outliers and time-varying delays on parameter estimation, we employ the Student's t distribution to deal with the measurement noise and use a first-order Markov chain to model the delays. In the framework of expectation-maximization (EM) algorithm, the unknown parameters, delays, noise variance, states and transition probability matrix can be estimated iteratively. A numerical simulation and a continuous stirred tank reactor (CSTR) process demonstrate that the proposed algorithm has good immunity against outliers and time-varying delays and offers good estimation accuracy for the bilinear SSM.

KEYWORDS
bilinear state-space model, expectation-maximization algorithm, outliers, parameter identification, time-varying delays

1 | INTRODUCTION

Considerable efforts have been made on the parameter identification for nonlinear systems in recent years.[1-3] As a particular kind of nonlinear system, the bilinear SSM contains the coupled term of the state vector and the control vector, and has an obvious changeable structure, which can describe the nuclear reaction process, population evolution process, cell division, and others.[4-6] Furthermore, the bilinear system can theoretically approximate a class of input-affine dynamic system and is more accurate than traditional linear approximation.[7-9] Thus, the research on the identification algorithm of bilinear SSM has high academic value.

In recent years, many approaches for identifying the parameters of bilinear SSM have been put forward. For example, Liu et al. derived an iterative identification algorithm for bilinear systems corrupted by moving average noise based on the dynamical moving data window.[10] Li et al. proposed a least squares iterative algorithm for input and output representation of bilinear systems by eliminating the unknown state variables.[11] Gu et al. discussed a hierarchical multi-innovation stochastic gradient identification algorithm based on the Kalman filtering, and given the convergence analyzes to ensure the stability of the algorithm.[12] However, the parameter estimation methods proposed in the above literature do not take into account the case of the bilinear SSM may be corrupted by outliers and time delays in practical applications.

As we all know, the precision of parameter identification is closely related to the quality of data set.[13] If we use the contaminated measurements in the experiment, imprecise parameters will be introduced.[14] Time delays are common in many practical production processes due to uncertainties such as long distance transmission, network transmission delays and manual measurement errors.[15] Ma et al. proposed a parameter estimation method based EM approach for time-varying Hammerstein models, where delays are described by uniformly distributed.[16] Unfortunately, time delays are often not uniformly distributed as described above, but the any two consecutive delays are correlated. Considering the correlation and variability of time delay, we adopt a first-order Markov chain to describe the unknown time delay.

In addition, the collected data often contain outliers due to some random factors, such as sensor failure, operation failure, or other unknown interference.[17,18] The traditional outlier treatment method is to detect the outliers directly, and then applies the data after eliminating the outliers to identify the unknown parameters. However, this may lead to information loss, which is not conducive to parameter estimation.[19] As an alternative, Student's t distribution can resist the influence of outliers on parameter identification, because Student's t distribution has a more enormous tail than Gaussian distribution by changing the degree of freedom (DOF).[20,21] When DOF is smaller, the Student's t distribution curve is smoother and the tail of both sides is higher. On the contrary, when DOF is larger, the Student's t distribution is closer to the standard Gaussian distribution.[22,23] In recent years, many scholars have been applied Student's t distribution to decay the impact of outliers. For example, Chen et al. presented a robust recursive EM algorithm for ARX models with Student's t distribution noise.[24] Assuming that the noise follows the Student's t distribution, Zhao et al. developed an FIR filter to estimate state for time-invariant SSMs contaminated by outliers.[25]

On the basis of the work in Reference 24, this paper proposes a parameter identification algorithm for the bilinear SSM in the presence of random outliers and time-varying delays to further promote the application of system identification in practical process. Considering that there is a coupled term of state variables and input variables in bilinear SSM, a bilinear state observer is applied to estimate the unknown states in the information vector, and the estimated information vector is derived. The main contributions of this paper are as follows.

1. To eliminate the influence of outliers and time-varying delays on parameter estimation, Student's t distribution is introduced to model the measurement noise, and the time delays are assumed to follow a first-order Markov chain.
2. Taking the unknown time delays and variance scale of noise as hidden variables, the mathematical formula of the EM algorithm for the identification of bilinear SSM with time-varying delays and outliers are derived, and the estimation of parameters, states, noise variance, time delays, and transition probability matrix are simultaneously achieved.

The remainder of this paper is organized as follows. In Section 2, we offer the bilinear SSM and formulate the problem. Section 3 briefly reviews the EM algorithm. Section 4 derives an open-loop state observer, based on which the EM approach is developed to accurately estimate the parameters. Section 5 provides a numerical simulation and a CSTR process to verify the effectiveness of the proposed algorithm. Section 6 gives the conclusions.

2 | PRELIMINARIES AND PROBLEM FORMULATION

In this article, some symbols are introduced for convenience. "$\boldsymbol{P} := \boldsymbol{C}$" denotes $\boldsymbol{C}$ is defined as $\boldsymbol{P}$; $\mathrm{E}(\boldsymbol{x})$ stands for the expectation of $\boldsymbol{x}$; $\hat{\boldsymbol{x}}_t$ symbolizes the estimate of state $\boldsymbol{x}$ at time t; z^{-1} means a unit backward shift operation like $z^{-1}\boldsymbol{x}_t = \boldsymbol{x}_{t-1}$; $\boldsymbol{\theta}^s$ represents the estimate of $\boldsymbol{\theta}$ at iteration variable s; $\boldsymbol{B}^{\mathrm{T}}$ represents the transpose of the vector or matrix $\boldsymbol{B}$.

Consider the following single-input single-output (SISO) bilinear SSM in the presence of random outliers and time-varying delays:[26,27]

$$\boldsymbol{x}_{t+1} = \boldsymbol{A}\boldsymbol{x}_t + \boldsymbol{D}\boldsymbol{x}_t u_t + \boldsymbol{f}u_t, \tag{1}$$

$$y_t = \boldsymbol{c}\boldsymbol{x}_{t-\lambda_t} + e_t, \tag{2}$$

where $\{u_{t|t=1,2,\dots,N}\} \in \mathbb{R}$, $\{y_{t|t=1,2,\dots,N}\} \in \mathbb{R}$ are the input, output data of the bilinear SSM, N is the data length, $\boldsymbol{x}_t := [x_{1,t}, x_{2,t}, \dots, x_{n,t}]^{\mathrm{T}} \in \mathbb{R}^n$ is the state vector, n is the order of the considered model and be assumed known, e_t is the measurement noise, $\{\lambda_{t|t=1,2,\dots,N}\} \in \mathbb{R}$ is the random time delays. Suppose that the bilinear SSM is stable, observable and controllable. Taking into account the bilinear SSM with an observer canonical form, $\boldsymbol{A} \in \mathbb{R}^{n\times n}$, $\boldsymbol{D} \in \mathbb{R}^{n\times n}$, $\boldsymbol{f} \in \mathbb{R}^n$, and

$\boldsymbol{c} \in \mathbb{R}^{1\times n}$ are time-invariant parameter matrices/vectors:

$$\boldsymbol{A} := \begin{bmatrix} -a_1 & 1 & 0 & \cdots & 0 \\ -a_2 & 0 & 1 & & 0 \\ \vdots & \vdots & & \ddots & \vdots \\ -a_{n-1} & 0 & \cdots & 0 & 1 \\ -a_n & 0 & \cdots & 0 & 0 \end{bmatrix} \in \mathbb{R}^{n\times n}, \quad \boldsymbol{D} := \begin{bmatrix} d_{11} & d_{12} & \cdots & d_{1n} \\ d_{21} & d_{22} & \cdots & d_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ d_{n1} & d_{n2} & \cdots & d_{nn} \end{bmatrix} = \begin{bmatrix} \boldsymbol{d}_1 \\ \boldsymbol{d}_2 \\ \vdots \\ \boldsymbol{d}_n \end{bmatrix} \in \mathbb{R}^{n\times n},$$

$$\boldsymbol{f} := \begin{bmatrix} f_1 \\ f_2 \\ \vdots \\ f_n \end{bmatrix} \in \mathbb{R}^{n}, \quad \boldsymbol{c} := \begin{bmatrix} 1, 0, \ldots, 0, 0 \end{bmatrix} \in \mathbb{R}^{1\times \mathrm{n}}.$$

Substituting the parameter matrices/vectors $\boldsymbol{A}, \boldsymbol{D}, \boldsymbol{f}, \boldsymbol{c}$ into (1) and (2), we get the following equations:

$$x_{m,t+1} = -a_m x_{1,t} + x_{m+1,t} + \boldsymbol{d}_m \boldsymbol{x}_t u_t + f_m u_t, \quad m = 1, 2, \ldots, n-1, \tag{3}$$

$$x_{n,t+1} = -a_n x_{1,t} + \boldsymbol{d}_n \boldsymbol{x}_t + f_n u_t, \tag{4}$$

$$y_t = x_{1,t-\lambda_t} + e_t, \tag{5}$$

Multiply both sides of (3) by z^{-m}, we have

$$\begin{cases} x_{1,t} = -a_1 x_{1,t-1} + x_{2,t-1} + \boldsymbol{d}_1 \boldsymbol{x}_{t-1} u_{t-1} + f_1 u_{t-1}, \\ x_{2,t-1} = -a_2 x_{1,t-2} + x_{3,t-2} + \boldsymbol{d}_2 \boldsymbol{x}_{t-2} u_{t-2} + f_2 u_{t-2}, \\ \vdots \\ x_{n-1,t-n+2} = -a_{n-1} x_{1,t-n+1} + x_{n,t-n+1} + \boldsymbol{d}_{n-1} \boldsymbol{x}_{t-n+1} u_{t-n+1} + f_{n-1} u_{t-n+1}, \end{cases} \tag{6}$$

Also, multiplying both sides of (4) by z^{-n} gives

$$x_{n,t-n+1} = -a_n x_{1,t-n} + \boldsymbol{d}_n \boldsymbol{x}_{t-n} + f_n u_{t-n}, \tag{7}$$

Combing (6) and (7), we can get

$$x_{1,t} = -\sum_{m=1}^{n} a_m x_{1,t-m} + \sum_{m=1}^{n} \boldsymbol{d}_m \boldsymbol{x}_{t-m} u_{t-m} + \sum_{m=1}^{n} f_m u_{t-m}, \tag{8}$$

According to (8), we can obtain $x_{1,t-\lambda_t}$ as

$$x_{1,t-\lambda_t} = -\sum_{m=1}^{n} a_m x_{1,t-m-\lambda_t} + \sum_{m=1}^{n} \boldsymbol{d}_m \boldsymbol{x}_{t-m-\lambda_t} u_{t-m-\lambda_t} + \sum_{m=1}^{n} f_m u_{t-m-\lambda_t}. \tag{9}$$

Submitting (9) into (5), the output variables $y(t)$ can be rewritten as the following regressive form:

$$y_t = \boldsymbol{\phi}^{\mathrm{T}}_{t-\lambda_t} \boldsymbol{\theta} + e_t, \tag{10}$$

where the parameter vector $\boldsymbol{\theta}$ to be identified is given by

$$\boldsymbol{\theta} := [a_1, a_2, \ldots, a_n, \boldsymbol{d}_1, \boldsymbol{d}_2, \ldots, \boldsymbol{d}_n, f_1, f_2, \ldots, f_n]^{\mathrm{T}} \in \mathbb{R}^{n^2+2n}, \tag{11}$$

the information vector $\boldsymbol{\phi}_{t-\lambda_t}$ with unknown states and delays is defined as

$$\boldsymbol{\phi}_{t-\lambda_t} := [-x_{1,t-1-\lambda_t}, \ \dots \ , -x_{1,t-n-\lambda_t}, \boldsymbol{x}^{\mathrm{T}}_{t-1-\lambda_t} u_{t-1-\lambda_t}, \ \dots \ , \boldsymbol{x}^{\mathrm{T}}_{t-n-\lambda_t} u_{t-n-\lambda_t}, u_{t-1-\lambda_t}, \ \dots \ , u_{t-n-\lambda_t}]^{\mathrm{T}} \in \mathbb{R}^{n^2+2n}. \tag{12}$$

Considering that the time delays are often variable and the delay at the current time is related to the delay at the previous time, we use a first-order Markov chain instead of the traditional uniform distribution to model the delays. Markov process includes an initial probability and a fixed probability transition matrix.[28,29] The transition probability matrix dominates the jump of time delays. Define the upper limit of delays as d. The transition probability α_{ij} and the initial probability ρ_i are denoted as

$$\alpha_{ij} := p(\lambda_t = i|\lambda_{t-1} = j), \quad t = 2, 3, \ \dots \ , N, \quad 1 \le i, j \le d, \quad s.t. \sum_{i=1}^{d} \alpha_{ij} = 1,$$

$$\rho_i := p(\lambda_1 = i), \quad 1 \le i \le d, \quad s.t. \sum_{i=1}^{d} \rho_i = 1.$$

Moreover, it is usually assumed that the measurement noise obeys the Gaussian distribution. However, if the output data is polluted by outliers, the estimated parameters obtained by using the Gaussian distribution will deviate from the actual values. In other words, the Gaussian distribution is vulnerable to outliers. An advanced way to mitigate the negative effects of outliers is to use Student's t distribution. Compared with Gaussian distribution, the probability density function (PDF) of the Student's t distribution has a heavier tail, which allocates more accommodate for outliers.[30] In this paper, the measurement noise e_t is assumed to follow the Student's t distribution with zero mean, variance σ^2 and the degree of freedom γ, e_t can be described as[31]

$$e_t \sim St(0, \sigma^2, \gamma) = \frac{\Gamma(\frac{\gamma+1}{2})}{\Gamma(\frac{\gamma}{2})\sqrt{\pi\gamma\sigma^2}} \left[1 + \frac{\delta(e_t|0, \sigma^2)}{\gamma}\right]^{-\frac{1+\gamma}{2}},$$

where $\delta(a|b, c)$ represents the squared Mahalanobis distance which is calculated as $\delta(a|b, c) = (a-b)^2/c$, $\Gamma(\cdot)$ is the Gamma function specified as $\Gamma(t) = \int_0^\infty z^{t-1} e^{-z} dz$. Based on (10), the output y_t also follows the Student's t distribution, that is

$$y_t \sim St(\boldsymbol{\phi}^{\mathrm{T}}_{t-\lambda_t}\boldsymbol{\theta}, \sigma^2, \gamma) = \frac{\Gamma(\frac{\gamma+1}{2})}{\Gamma(\frac{\gamma}{2})\sqrt{\pi\gamma\sigma^2}} \left[1 + \frac{\delta(y_t|\boldsymbol{\phi}^{\mathrm{T}}_{t-\lambda_t}\boldsymbol{\theta}, \sigma^2)}{\gamma}\right]^{-\frac{1+\gamma}{2}}. \tag{13}$$

Essentially, the Student's t distribution can be decomposed into scaled Gaussian distributions where the variance scale υ_t is a Gamma distributed latent variable determined by the DOF,[32] that is,

$$\begin{aligned} p(y_t|\boldsymbol{\phi}_{t-\lambda_t}, \boldsymbol{\theta}, \sigma^2, \gamma) &= \int p(y_t|\boldsymbol{\phi}_{t-\lambda_t}, \boldsymbol{\theta}, \sigma^2, \gamma, \upsilon_t) p(\upsilon_t|\boldsymbol{\phi}_{t-\lambda_t}, \boldsymbol{\theta}, \sigma^2, \gamma) d\upsilon_t \\ &= \int p(y_t|\boldsymbol{\phi}_{t-\lambda_t}, \boldsymbol{\theta}, \sigma^2, \upsilon_t) p(\upsilon_t|\gamma) d\upsilon_t, \end{aligned} \tag{14}$$

where

$$(y_t|\boldsymbol{\phi}_{t-\lambda_t}, \boldsymbol{\theta}, \sigma^2, \upsilon_t) \sim N\left(\boldsymbol{\phi}^{\mathrm{T}}_{t-\lambda_t}\boldsymbol{\theta}, \frac{\sigma^2}{\upsilon_t}\right),$$

$$(\upsilon_t|\gamma) \sim \mathrm{Gamma}\left(\frac{\gamma}{2}, \frac{\gamma}{2}\right),$$

The PDF of the Gamma distribution has the form:

$$\mathrm{Gamma}(\upsilon|a, b) = \frac{b^a \upsilon^{a-1}}{\Gamma(a)} \exp\{-b\upsilon\}, \quad \upsilon > 0, a > 0, b > 0. \tag{15}$$

Therefore, the conditional PDF of y_t and υ_t is computed as

$$p(y_t|\boldsymbol{\phi}_{t-\lambda_t}, \boldsymbol{\theta}, \sigma^2, \upsilon_t) = \frac{\sqrt{\upsilon_t}}{\sqrt{2\pi\sigma^2}} \exp\left\{ -\frac{\upsilon_t \delta(y_t|\boldsymbol{\phi}^{\mathrm{T}}_{t-\lambda_t}\boldsymbol{\theta}, \sigma^2)}{2} \right\}, \tag{16}$$

$$p(\upsilon_t|\gamma) = \frac{\frac{\gamma}{2}^{\frac{\gamma}{2}} \upsilon_t^{\frac{\gamma}{2}-1}}{\Gamma\left(\frac{\gamma}{2}\right)} \exp\left\{ -\frac{\gamma}{2}\upsilon_t \right\}. \tag{17}$$

The purpose of this paper is to estimate the unknown the parameter vector $\boldsymbol{\theta}$, time delay λ_t, the transition probability α_{ij}, noise variance σ^2, the degree of freedom γ by using the observable input sequence $\{u_t\}$, output sequence $\{y_t\}$ polluted by outliers and time-varying delays. There is no doubt that unknown delays and outliers greatly increase the difficulty of parameter estimation. Next, we will elaborate an iterative identification algorithm based on the EM algorithm to estimate the parameters of the considered model in this paper.

Remark 1. It should be noted that the information vector $\boldsymbol{\phi}_{t-\lambda_t}$ contains not only the unknown state $\boldsymbol{x}_t$, but also random time delay λ_t. Thus, traditional identification algorithms such as least squares and gradient iteration can not be directly applied to identify the parameters of the considered model.

3 THE EM ALGORITHM REVISITED

In this section, we briefly introduce the EM algorithm. EM algorithm was first proposed by Dempster et al. in 1977 to solve the parameter estimation problem of data loss.[33] The basic principle of the EM algorithm is to iteratively execute the Expectation step (E-step) to estimate the hidden variables and to apply the Maximization step (M-step) to update the system parameters according to the available input and output data.[34] Its convergence was proved by Russell A. Boyles.[35] Suppose the complete data set C contains the observed data set C_{obs} and the hidden data set C_{mis}, Θ stands for the parameter set to be estimated, the specific algorithm steps are as follows.[36]

(1) Initialization: Let $s = 1$, set the initial value Θ^0 and the maximum number of the iteration variable $S_{\max}$.
(2) E-step: Θ^{s-1} represents the parameter estimate obtained from the previous iteration or the initial value, $\mathrm{E}_{f(\cdot)}\{g(\cdot)\}$ denotes the expectation of $g(\cdot)$ taken with respect to $f(\cdot)$, then the expectation of complete data log-likelihood function (Q-function) can be obtained by the following formula

$$Q(\Theta|\Theta^{s-1}) = \mathrm{E}_{C_{\mathrm{mis}}|C_{\mathrm{obs}},\Theta^{s-1}} \left\{ \log p\left(C_{\mathrm{mis}}, C_{\mathrm{obs}}|\Theta\right) \right\}. \tag{18}$$

(3) M-step: Maximizing the Q-function with respect Θ to update the parameter Θ, that is

$$\Theta^s = \underset{\Theta}{argmax}\; Q\left(\Theta|\Theta^{s-1}\right). \tag{19}$$

(4) If $s = S_{\max}$, stop; Otherwise, set $s = s + 1$ and go back to E-step.

4 EM ALGORITHM FOR BILINEAR SSM

In this section, an open-loop state observer is designed to estimate unknown state in the information vector, and then the bilinear state observer based EM algorithm using Student's t distribution (BSO-EM-T algorithm) is derived to estimate the unknown parameter of bilinear SSM contaminated by random outliers and time-varying delays.

4.1 E-step

The information vector $\boldsymbol{\phi}_{t-\lambda_t}$ contains unknown states, which increases the difficulty of identification. Therefore, we should first estimate the states before estimating the unknown parameters. Kalman filter and state observer are often adopted to estimate the unknown states of SSM. When applying the Kalman filter to estimate the state, however, one always faces a dilemma that the noise obeys white Gaussian distribution. That is to say, Kalman filter is not suitable for estimating the state of this paper. Thus, we adopt an open-loop state observer to estimate the unknown states in this paper. Based on (1), the open-loop state observer is designed as

$$\hat{\boldsymbol{x}}_{t+1} = \boldsymbol{A}\hat{\boldsymbol{x}}_t + \boldsymbol{D}\hat{\boldsymbol{x}}_t u_t + \boldsymbol{f} u_t. \tag{20}$$

Since the state observer cannot work when the system parameters $\boldsymbol{A}, \boldsymbol{D}, \boldsymbol{f}$ are unknown, we use the estimated parameter $\boldsymbol{A}^{s-1}, \boldsymbol{D}^{s-1}, \boldsymbol{f}^{s-1}$ to replace them, respectively. Then, we can compute the estimate of $\hat{\boldsymbol{x}}_{t+1}$ with

$$\hat{\boldsymbol{x}}_{t+1} = \boldsymbol{A}^{s-1}\hat{\boldsymbol{x}}_t + \boldsymbol{D}^{s-1}\hat{\boldsymbol{x}}_t u_t + \boldsymbol{f}^{s-1} u_t, \tag{21}$$

where

$$\boldsymbol{A}^{s-1} = \begin{bmatrix} -a_1^{s-1} & 1 & 0 & \cdots & 0 \\ -a_2^{s-1} & 0 & 1 & & 0 \\ \vdots & \vdots & & \ddots & \vdots \\ -a_{n-1}^{s-1} & 0 & \cdots & 0 & 1 \\ -a_n^{s-1} & 0 & \cdots & 0 & 0 \end{bmatrix}, \quad \boldsymbol{D}^{s-1} = \begin{bmatrix} \boldsymbol{d}_1^{s-1} \\ \boldsymbol{d}_2^{s-1} \\ \vdots \\ \boldsymbol{d}_{n-1}^{s-1} \\ \boldsymbol{d}_n^{s-1} \end{bmatrix}, \quad \boldsymbol{f}^{s-1} = \begin{bmatrix} f_1^{s-1} \\ f_2^{s-1} \\ \vdots \\ f_{n-1}^{s-1} \\ f_n^{s-1} \end{bmatrix},$$

Based on the idea of auxiliary model identification, we use the estimated state $\hat{\boldsymbol{x}}_t$ to replace the true state $\boldsymbol{x}_t$ in the information vector $\boldsymbol{\phi}_{t-\lambda_t}$, then, the estimated information vector $\hat{\boldsymbol{\phi}}_{t-\lambda_t}$ can be achieved by

$$\begin{aligned}\hat{\boldsymbol{\phi}}_{t-\lambda_t} := [&-\hat{x}_{1,t-1-\lambda_t}, \dots, -\hat{x}_{1,t-n-\lambda_t}, \hat{\boldsymbol{x}}^{\mathrm{T}}_{t-1-\lambda_t} u_{t-1-\lambda_t}, \dots, \hat{\boldsymbol{x}}^{\mathrm{T}}_{t-n-\lambda_t} u_{t-n-\lambda_t}, \\ &u_{t-1-\lambda_t}, \dots, u_{t-n-\lambda_t}]^{\mathrm{T}}.\end{aligned} \tag{22}$$

For the discussed bilinear SSM, the observed data set C_{obs} include input data u_t and output data y_t, the variance scale υ_t and the unknown delay λ_t are considered as the hidden data set C_{mis}. The overall parameter set Θ to be estimated includes the unknown parameters $\boldsymbol{\theta}$, noise variance σ^2, the degree of freedom γ, the initial distribution of delay ρ_i and the transition probability of delay α_{ij}. Concretely defined as follow:

$$\begin{aligned} C_{\mathrm{obs}} &= \{Y, U\} = \{y_1, y_2, \dots, y_N, u_1, u_2, \dots, u_N\}, \\ C_{\mathrm{mis}} &= \{\Lambda, \Upsilon\} = \{\lambda_1, \lambda_2, \dots, \lambda_N, \upsilon_1, \upsilon_2, \dots, \upsilon_N\}, \\ \Theta &= \{\boldsymbol{\theta}, \sigma^2, \gamma, \alpha_{ij}, \rho_i\}, 1 \le i, j \le d. \end{aligned}$$

Based on (18), the Q-function can be expressed as

$$\begin{aligned} Q\left(\Theta|\Theta^{s-1}\right) &= \mathrm{E}_{C_{\mathrm{mis}}|C_{\mathrm{obs}},\Theta^{s-1}}\left\{\log p\left(C_{\mathrm{mis}}, C_{\mathrm{obs}}|\Theta\right)\right\} \\ &= \mathrm{E}_{\Lambda,\Upsilon|Y,U,\Theta^{s-1}}\left\{\log p\left(Y, U, \Lambda, \Upsilon|\Theta\right)\right\}. \end{aligned} \tag{23}$$

On the basis of Bayesian theory, the term $\log p\,(Y, U, \Lambda, \Upsilon|\Theta)$ in (23) can be decomposed as

$$\begin{aligned} \log p\,(Y, U, \Lambda, \Upsilon|\Theta) &= \log p\,(Y|U, \Lambda, \Upsilon, \Theta)\, p\,(\Lambda|U, \Upsilon, \Theta)\, p\,(\Upsilon|U, \Theta)\, p\,(U|\Theta) \\ &= \sum_{t=1}^{N} \log p(y_t|\hat{\boldsymbol{\phi}}_{t-\lambda_t}, \lambda_t, \upsilon_t, \Theta) + \sum_{t=2}^{N} \log p(\lambda_t|\lambda_{t-1}, \Theta) + \log p(\lambda_1|\Theta) \\ &\quad + \sum_{t=1}^{N} \log p(\upsilon_t|\Theta) + \mathrm{K}, \end{aligned} \tag{24}$$

where $\mathrm{K} := \log p(U|\Theta)$ is constant since the input set is known and independent with the parameter Θ. Then, substituting (24) into (23), the Q-function can be further expressed as

$$\begin{aligned} Q\left(\Theta|\Theta^{s-1}\right) = &\sum_{t=1}^{N}\int\sum_{i=1}^{d} p(v_t|Y,U,\lambda_t=i,\Theta^{s-1})p(\lambda_t=i|Y,U,\Theta^{s-1})\log p(y_t|\hat{\boldsymbol{\phi}}_{t-\lambda_t},\lambda_t,v_t,\Theta)dv_t \\ &+\sum_{t=2}^{N}\sum_{i=1}^{d}\sum_{j=1}^{d} p(\lambda_t=i,\lambda_{t-1}=j|Y,U,\Theta^{s-1})\log p(\lambda_t=i|\lambda_{t-1}=j,\Theta) \\ &+\sum_{i=1}^{d} p(\lambda_1=i|Y,U,\Theta^{s-1})\log p(\lambda_1=i|\Theta) \\ &+\sum_{t=1}^{N}\int\sum_{i=1}^{d} p(v_t|Y,U,\lambda_t=i,\Theta^{s-1})p(\lambda_t=i|Y,U,\Theta^{s-1})\log p(v_t|\Theta)dv_t+\mathrm{K}. \end{aligned} \tag{25}$$

For convenience, we divide the Q-function in (25) into four parts as follows:

$$\begin{aligned} Q_1 = &\sum_{t=1}^{N}\int\sum_{i=1}^{d} p(v_t|Y,U,\lambda_t=i,\Theta^{s-1})p(\lambda_t=i|Y,U,\Theta^{s-1}) \\ &\times\log p(y_t|\hat{\boldsymbol{\phi}}_{t-\lambda_t},\lambda_t,v_t,\Theta)dv_t, \end{aligned} \tag{26}$$

$$Q_2 = \sum_{t=2}^{N}\sum_{i=1}^{d}\sum_{j=1}^{d} p(\lambda_t=i,\lambda_{t-1}=j|Y,U,\Theta^{s-1})\log p(\lambda_t=i|\lambda_{t-1}=j,\Theta), \tag{27}$$

$$Q_3 = \sum_{i=1}^{d} p(\lambda_1=i|Y,U,\Theta^{s-1})\log p(\lambda_1=i|\Theta), \tag{28}$$

$$Q_4 = \sum_{t=1}^{N}\int\sum_{i=1}^{d} p(v_t|Y,U,\lambda_t=i,\Theta^{s-1})p(\lambda_t=i|Y,U,\Theta^{s-1})\log p(v_t|\Theta)dv_t+\mathrm{K}. \tag{29}$$

In order to calculate Q_1, Q_2, Q_3, Q_4, the following posteriori PDFs about hidden variables should be computed in advance:

(1) $p(\lambda_1=i|Y,U,\Theta^{s-1})$, $p(\lambda_t=i,\lambda_{t-1}=j|Y,U,\Theta^{s-1})$, $p(\lambda_t=i|Y,U,\Theta^{s-1})$,
(2) $p(v_t|Y,U,\lambda_t=i,\Theta^{s-1})$.

First, according to Bayesian rules, the posterior probability of the initial time delay can be obtained by

$$\begin{aligned} p(\lambda_1=i|Y,U,\Theta^{s-1}) &= \frac{p(\lambda_1=i,y_1|u_1,\Theta^{s-1})}{\sum_{i=1}^{d} p(\lambda_1=i,y_1|u_1,\Theta^{s-1})} \\ &= \frac{p(y_1|\hat{\boldsymbol{\phi}}_{1-\lambda_1},\lambda_1=i,\Theta^{s-1})(\rho_i)^{s-1}}{\sum_{i=1}^{d} p(y_1|\hat{\boldsymbol{\phi}}_{1-\lambda_1},\lambda_1=i,\Theta^{s-1})(\rho_i)^{s-1}}. \end{aligned} \tag{30}$$

The joint probability distribution $p(\lambda_t=i,\lambda_{t-1}=j|Y,U,\Theta^{s-1})$ for $t\geq 2$ of time delays can be calculated as

$$\begin{aligned} p(\lambda_t=i,\lambda_{t-1}=j|Y,U,\Theta^{s-1}) &= \frac{p(\lambda_t=i,\lambda_{t-1}=j,y_t|y_{1:t-1},u_{1:t},\Theta^{s-1})}{\sum_{i=1}^{d}\sum_{j=1}^{d} p(\lambda_t=i,\lambda_{t-1}=j,y_t|y_{1:t-1},u_{1:t},\Theta^{s-1})} \\ &= \frac{\Pi_{t,i}(\alpha_{ij})^{s-1}\Psi_{t-1,j}}{\sum_{i=1}^{d}\sum_{j=1}^{d}\Pi_{t,i}(\alpha_{ij})^{s-1}\Psi_{t-1,j}}, \end{aligned} \tag{31}$$

where $\Pi_{t,i} := p(y_t|\hat{\boldsymbol{\phi}}_{t-\lambda_t},\lambda_t=i,\Theta^{s-1})$, $\Psi_{t-1,j} := p(\lambda_{t-1}=j|y_{1:t-1},u_{1:t-1},\Theta^{s-1})$. From the above joint probability distribution, we can get the distribution of delays for time $t\geq 2$ by

$$p(\lambda_t=i|Y,U,\Theta^{s-1}) = \sum_{j=1}^{d} p(\lambda_t=i,\lambda_{t-1}=j|Y,U,\Theta^{s-1}). \tag{32}$$

Then, according to the Bayesian formula, the posterior PDF of variance scale υ_t can be written as

$$p(\upsilon_t|Y,U,\lambda_t=i,\Theta^{s-1}) = \frac{p(y_t|\upsilon_t,\hat{\boldsymbol{\phi}}_{t-\lambda_t},\lambda_t=i,\Theta^{s-1})p(\upsilon_t|\Theta^{s-1})}{\int_0^\infty p(y_t|\upsilon_t,\hat{\boldsymbol{\phi}}_{t-\lambda_t},\lambda_t=i,\Theta^{s-1})p(\upsilon_t|\Theta^{s-1})d\upsilon_t}. \tag{33}$$

For notation simplicity, we define the molecule of (33) as

$$\begin{aligned} H &:= p(y_t|\upsilon_t,\hat{\boldsymbol{\phi}}_{t-\lambda_t},\lambda_t=i,\Theta^{s-1})p(\upsilon_t|\Theta^{s-1}) \\ &= \frac{\sqrt{\upsilon_t}}{\sqrt{2\pi(\sigma^2)^{s-1}}}\exp\left\{-\frac{\upsilon_t\delta(y_t|\hat{\boldsymbol{\phi}}^{\mathrm{T}}_{t-\lambda_t}\boldsymbol{\theta}^{s-1},(\sigma^2)^{s-1})}{2}\right\}\frac{\eta^\eta\upsilon_t^{\eta-1}}{\Gamma(\eta)}\exp\left\{-\eta\upsilon_t\right\}, \end{aligned} \tag{34}$$

where $\eta := \frac{\gamma^{s-1}}{2}$, the denominator of (33) can be obtained by

$$\begin{aligned} \int_0^\infty H d\upsilon_t &= \frac{\eta^\eta}{\Gamma(\eta)\sqrt{2\pi(\sigma^2)^{s-1}}}\int \upsilon_t^{\eta-\frac{1}{2}}\exp\left\{-\upsilon_t\frac{\gamma^{s-1}+\delta(y_t|\hat{\boldsymbol{\phi}}^{\mathrm{T}}_{t-\lambda_t}\boldsymbol{\theta}^{s-1},(\sigma^2)^{s-1})}{2}\right\}d\upsilon_t \\ &= \frac{\eta^\eta}{\Gamma(\eta)\sqrt{2\pi(\sigma^2)^{s-1}}}\Gamma(\frac{\gamma^{s-1}+1}{2})\left\{\frac{\gamma^{s-1}+\delta(y_t|\hat{\boldsymbol{\phi}}^{\mathrm{T}}_{t-\lambda_t}\boldsymbol{\theta}^{s-1},(\sigma^2)^{s-1})}{2}\right\}^{\frac{\gamma^{s-1}+1}{2}}. \end{aligned} \tag{35}$$

Substituting (34) and (35) into (33), we can get that the posterior distribution of υ_t follows a Gamma distribution, that is, $(\upsilon_t|Y,U,\lambda_t=i,\Theta^{s-1}) \sim \mathrm{Gamma}(\upsilon_t|\kappa,\beta)$, where $\kappa = \frac{\gamma^{s-1}+1}{2}$, $\beta = \frac{\gamma^{s-1}+\delta(y_t|\hat{\boldsymbol{\phi}}^{\mathrm{T}}_{t-\lambda_t}\boldsymbol{\theta}^{s-1},(\sigma^2)^{s-1})}{2}$. Therefore, the expectation of υ_t and $\log\upsilon_t$ can be written as

$$\bar{\upsilon}_{t,i} := \mathrm{E}(\upsilon_t) = \frac{\gamma^{s-1}+1}{\gamma^{s-1}+\delta(y_t|\hat{\boldsymbol{\phi}}^{\mathrm{T}}_{t-\lambda_t}\boldsymbol{\theta}^{s-1},(\sigma^2)^{s-1})}, \tag{36}$$

$$\tilde{\upsilon}_{t,i} := \mathrm{E}(\log\upsilon_t) = \Psi\left(\frac{\gamma^{s-1}+1}{2}\right) - \log\frac{\gamma^{s-1}+1}{2\bar{\upsilon}_{t,i}}, \tag{37}$$

where $\Psi(\gamma)$ is the derivative of the logarithm of the Gamma function, that is, $\Psi(\gamma) = \frac{1}{\Gamma(\gamma)}\frac{\partial\Gamma(\gamma)}{\partial\gamma}$.

According to (16) and (17), we get

$$\log p(y_t|\hat{\boldsymbol{\phi}}_{t-\lambda_t},\lambda_t,\upsilon_t,\Theta) = -\frac{1}{2}\log 2\pi\sigma^2 + \frac{1}{2}\log\upsilon_t - \frac{\upsilon_t}{2\sigma^2}\left(y_t - \hat{\boldsymbol{\phi}}^{\mathrm{T}}_{t-\lambda_t}\boldsymbol{\theta}\right)^2, \tag{38}$$

$$\log p(\upsilon_t|\Theta) = -\log\Gamma\left(\frac{\gamma}{2}\right) + \frac{\gamma}{2}\log\frac{\gamma}{2} + \left(\frac{\gamma}{2}-1\right)\log\upsilon_t - \frac{\gamma}{2}\upsilon_t. \tag{39}$$

Substituting (30)–(32) and (36)–(39) into the four separated terms Q_1, Q_2, Q_3 and Q_4 in (26)–(29), respectively, gives,

$$\begin{aligned} Q_1 = -\frac{1}{2}\sum_{t=1}^{N}\sum_{i=1}^{d}p(\lambda_t=i|Y,U,\Theta^{s-1})\bigg\{&\log 2\pi\sigma^2 - \Psi\left(\frac{\gamma^{s-1}+1}{2}\right) + \log\frac{\gamma^{s-1}+1}{2\bar{\upsilon}_{t,i}} \\ &+ \frac{\bar{\upsilon}_{t,i}}{\sigma^2}\left(y_t - \hat{\boldsymbol{\phi}}^{\mathrm{T}}_{t-\lambda_t}\boldsymbol{\theta}\right)^2\bigg\}, \end{aligned} \tag{40}$$

$$Q_2 = \sum_{t=2}^{N}\sum_{i=1}^{d}\sum_{j=1}^{d}p(\lambda_t=i,\lambda_{t-1}=j|Y,U,\Theta^{s-1})\log\alpha_{ij}, \tag{41}$$

$$Q_3 = \sum_{i=1}^{d} p(\lambda_1 = i|Y, U, \Theta^{s-1}) \log \rho_i, \tag{42}$$

$$Q_4 = \frac{1}{2}\sum_{t=1}^{N}\sum_{i=1}^{d} p(\lambda_t = i|Y, U, \Theta^{s-1}) \left\{-2\log\Gamma\left(\frac{\gamma}{2}\right) + \gamma\log\frac{\gamma}{2} - \gamma\overline{\upsilon}_{t,i} + (\gamma-2)\tilde{\upsilon}_{t,i}\right\}. \tag{43}$$

4.2 | M-step

The objective of the M-step is to obtain the estimates of parameter Θ based on E-step. First, we let the partial derivative of $Q\left(\Theta|\Theta^{s-1}\right)$ with respect to $\boldsymbol{\theta}$ and σ^2 be zero, respectively,

$$\frac{\partial Q(\Theta|\Theta^{s-1})}{\partial \boldsymbol{\theta}} = \frac{\partial Q_1}{\partial \boldsymbol{\theta}} = \mathbf{0},$$
$$\frac{\partial Q(\Theta|\Theta^{s-1})}{\partial \sigma^2} = \frac{\partial Q_1}{\partial \sigma^2} = 0,$$

then, the estimates of parameters $\boldsymbol{\theta}$, noise variance σ^2 can be obtained by

$$\boldsymbol{\theta}^s = \frac{\sum_{t=1}^{N}\sum_{i=1}^{d} p(\lambda_t = i|Y, U, \Theta^{s-1})\overline{\upsilon}_{t,i}\hat{\boldsymbol{\phi}}_{t-\lambda_t} y_t}{\sum_{t=1}^{N}\sum_{i=1}^{d} p(\lambda_t = i|Y, U, \Theta^{s-1})\overline{\upsilon}_{t,i}\hat{\boldsymbol{\phi}}_{t-\lambda_t}\hat{\boldsymbol{\phi}}_{t-\lambda_t}^{\mathrm{T}}}, \tag{44}$$

$$(\sigma^2)^s = \frac{1}{N}\sum_{t=1}^{N}\sum_{i=1}^{d} p(\lambda_t = i|Y, U, \Theta^{s-1})\overline{\upsilon}_{t,i}\left(y_t - \hat{\boldsymbol{\phi}}_{t-\lambda_t}^{\mathrm{T}}\boldsymbol{\theta}^s\right)^2. \tag{45}$$

For Markov jump time delays, under the constraints $\sum_{i=1}^{d}\alpha_{ij} = 1$ and $\sum_{i=1}^{d}\rho_i = 1$, Lagrange multiplier Λ_α and Λ_ρ are introduced. The corresponding Lagrange function are formulated as follows:

$$L(\alpha_{ij}) = Q_2 + \Lambda_\alpha\left(\sum_{i=1}^{d}\alpha_{ij} - 1\right),$$
$$L(\rho_i) = Q_3 + \Lambda_\rho\left(\sum_{i=1}^{d}\rho_i - 1\right).$$

Then the estimates of the transition probability α_{ij}, the estimated initial probability ρ_i at iteration s can be computed by

$$\alpha_{ij}^s = \frac{\sum_{t=2}^{N} p(\lambda_t = i, \lambda_{t-1} = j|Y, U, \Theta^{s-1})}{\sum_{t=2}^{N}\sum_{i=1}^{d} p(\lambda_t = i, \lambda_{t-1} = j|Y, U, \Theta^{s-1})}, \tag{46}$$

$$\rho_i^s = p(\lambda_1 = i|Y, U, \Theta^{s-1}). \tag{47}$$

For measurement noise subject to Student's t distribution, letting the partial derivative of $Q\left(\Theta|\Theta^{s-1}\right)$ with respect to γ be zero,

$$\frac{\partial Q\left(\Theta|\Theta^{s-1}\right)}{\partial \gamma} = \frac{\partial Q_4}{\partial \gamma} = 0, \tag{48}$$

the estimated DOF of noise γ^s at iteration variable s can be calculated by means of the Matlab toolbox. According to the posteriori PDF of delays at each sampling time obtained by E-step, The estimated value of the time delay can be given by

$$\lambda_t^s = \underset{\lambda_t}{argmax}\{p(\lambda_t = i|Y, U, \Theta^{s-1})\}, \quad i = 1, 2, \ldots, d. \tag{49}$$

Based on the above derivation, the BSO-EM-T algorithm for bilinear SSM in the presence of random time delays and outliers is summarized in Algorithm 1.

Algorithm 1. The BSO-EM-T algorithm

Input: $\{u_t, y_t, t = 1, \dots, N\}$ contaminated by random outliers and time delays
Output: $\boldsymbol{\theta}^s$, $(\sigma^2)^s$, α_{ij}^s, ρ_i^s, γ^s, $\hat{\boldsymbol{x}}_t$, λ_t^s
Initialization: Set $s = 1$. Set the initial value $\Theta^0 = \{\boldsymbol{\theta}^0, (\sigma^2)^0, \gamma^0, \alpha_{ij}^0, \rho_i^0\}$ and the initial state $\boldsymbol{x}^0$, the maximum number of the iteration variable $S_{\max}$.
for $s = 1 : S_{\max}$ **do**
E-step:
(1) Compute the unknown states $\hat{\boldsymbol{x}}_t$ by (21),
(2) Construct the information vector $\hat{\boldsymbol{\phi}}_{t-\lambda_t}$ by (22),
(3) Evaluate $p(\lambda_1 = i|Y, U, \Theta^{s-1}), p(\lambda_t = i, \lambda_{t-1} = j|Y, U, \Theta^{s-1}), p(\lambda_t = i|Y, U, \Theta^{s-1})$ and $\bar{v}_{t,i}, \tilde{v}_{t,i}$ by (30)-(32) and (38)-(39) with the parameters obtained by the previous iteration.
M-step:
(1) Update $\boldsymbol{\theta}^s$, $(\sigma^2)^s$, α_{ij}^s, ρ_i^s, γ^s by (44)–(48), respectively,
(2) Obtain λ_t^s by (49).
end for

Remark 2. To the best of the authors' knowledge, there is almost no study which takes into account the effects of time-varying delays and outliers on parameter identification for the bilinear SSM. Compared with previous algorithms, the BSO-EM-T algorithm considers not only the time-varying delays, but also the random outliers, which has wider applicability.

Remark 3. The BSO-EM-T algorithm is an iterative identification algorithm, which realizes the interactive estimation of system parameters, unknown states, noise variance, random delays and transition probability matrix. The hidden variables are computed in E-step with the system parameters got in the previous iteration, the parameters are updated in M-step based on the estimated hidden variables.

5 | SIMULATIONS

In this section, a numerical simulation and a CSTR example are adopted to verify that the proposed BSO-EM-T algorithm in this paper can accurately estimate the unknown parameters of the bilinear SSM in the presence of random outliers and time-varying delays.

5.1 | Numerical simulation

Consider a second-order bilinear SSM with random outliers and time-varying delays,

$$\boldsymbol{x}_{t+1} = \begin{bmatrix} -0.54 & 1.00 \\ -0.15 & 0.00 \end{bmatrix} \boldsymbol{x}_t + \begin{bmatrix} 0.04 & 0.28 \\ 0.45 & 0.20 \end{bmatrix} \boldsymbol{x}_t u_t + \begin{bmatrix} 1.20 \\ 0.40 \end{bmatrix} u_t,$$
$$y_t = \begin{bmatrix} 1, 0 \end{bmatrix} \boldsymbol{x}_{t-\lambda_t} + e_t.$$

The parameter vector $\boldsymbol{\theta}$ to be estimated in this numerical example is

$$\boldsymbol{\theta} = \begin{bmatrix} 0.54, & 0.15, & 0.04, & 0.28, & 0.45, & 0.20, & 1.20, & 0.40 \end{bmatrix}^{\mathrm{T}}.$$

The upper limit of delay is set to $d = 4$, the true transition probability matrix is preset as

$$\boldsymbol{P} = [\alpha_{ij}] = \begin{bmatrix} 0.950 & 0.010 & 0.030 & 0.020 \\ 0.010 & 0.920 & 0.020 & 0.010 \\ 0.020 & 0.050 & 0.930 & 0.030 \\ 0.020 & 0.020 & 0.020 & 0.940 \end{bmatrix}. \tag{50}$$

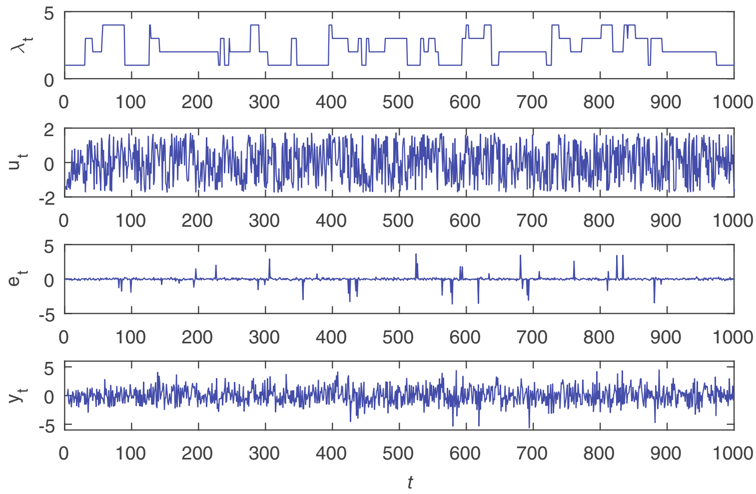

FIGURE 1 The simulated time delay, input, noise, and output data in the presence of 5% outliers.

TABLE 1 The parameter estimates and errors versus s with 5% outliers.

s	a_1	a_2	d_{11}	d_{12}	d_{21}	d_{22}	f_1	f_2	$\delta(\%)$
1	0.33038	0.35185	0.37515	0.44838	0.37416	0.37682	0.53889	0.51304	56.43048
2	0.25987	0.24942	0.22972	0.39226	0.30472	0.33125	0.56822	0.42058	50.69731
5	0.41765	0.18571	0.07563	0.44747	0.43461	0.40029	0.86358	0.36492	29.92761
8	0.49557	0.14126	0.06374	0.29470	0.50436	0.21135	1.11124	0.30466	10.11454
10	0.52387	0.15155	0.05594	0.26744	0.48114	0.15977	1.19460	0.35036	5.07299
20	0.54383	0.15440	0.02933	0.28304	0.44330	0.20405	1.20353	0.40754	1.13405
True values	0.54000	0.15000	0.04000	0.28000	0.45000	0.20000	1.20000	0.40000	

In the simulation, the input data $\{u_t\}$ is chosen as a persistent excitation signal with zero mean and unit variance. A Gaussian noise sequence with zero mean, variance $\sigma^2 = 0.1^2$ and random 5% outliers are added to the measurement noise e_t. Specifically, the outliers are generated by drawing samples from the uniform distribution $\mu[-4, 4]$ randomly. Under the control of the true transition probability matrix, the time delay switches among $\{1, 2, 3, 4\}$. 1000 sets of simulated data are employed for the identification. The simulated time delay, input, noise and output data contaminated 5% outliers are shown in Figure 1.

Based on the simulation data, the proposed BSO-EM-T algorithm is used to identify the unknown parameter of the second-order bilinear SSM. The maximal iteration variable $S_{\max}$ is set to 20. The parameter estimates and their errors $\delta := ||\theta^s - \theta||/||\theta||$ versus iterative variable s are presented in Table 1. The curves of the parameter estimates versus s are shown in Figure 2. The results of estimated state of $\hat{x}_{1,t}$ and $\hat{x}_{2,t}$ obtained by the state observer are display in Figure 3. About measurement noise, the estimated DOF is $\hat{\gamma} = 1.9897$ and the estimated variance is $\hat{\sigma}^2 = 0.0089$. The estimated transition probability matrix as follows

$$\hat{\boldsymbol{P}} = [\hat{\alpha}_{ij}] = \begin{bmatrix} 0.9421 & 0.0144 & 0.0320 & 0.0046 \\ 0.0137 & 0.9204 & 0.0340 & 0.0137 \\ 0.0283 & 0.0309 & 0.9056 & 0.0495 \\ 0.0159 & 0.0342 & 0.0285 & 0.9321 \end{bmatrix}. \tag{51}$$

Apparently, the estimated value of each element in (51) is very close to the true value in (50). The true delays and the estimated delays are compared in Figure 4 with the accuracy of time delay estimation(ADE) is 96.2%.

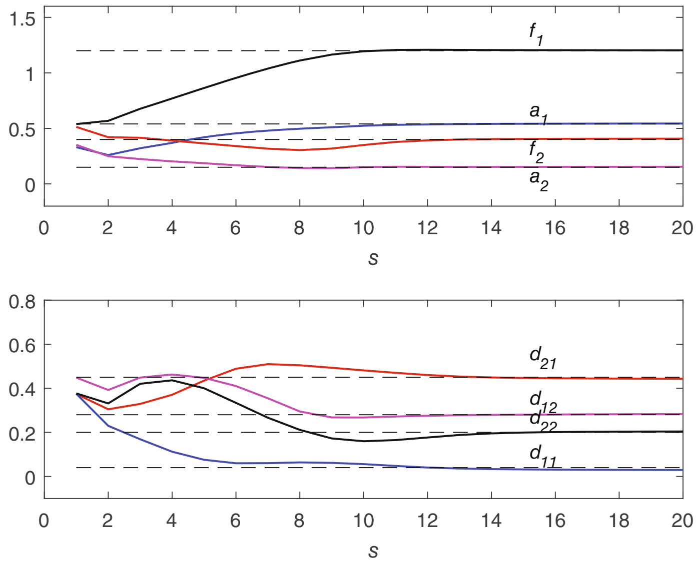

FIGURE 2 The parameter estimates with 5% outliers.

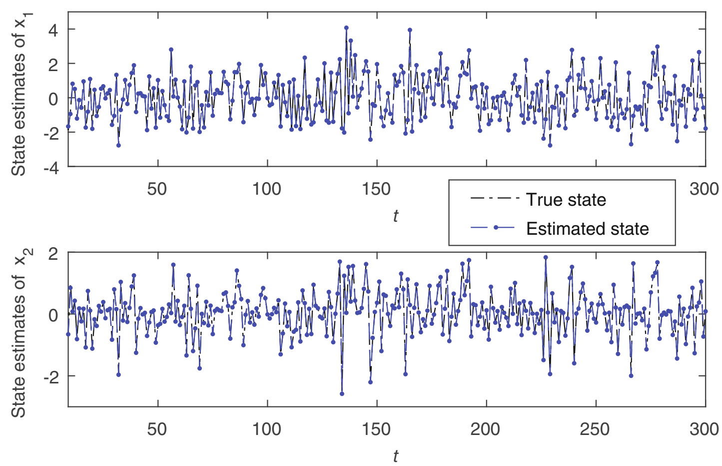

FIGURE 3 Comparison of estimated states and true states with 5% outliers.

Furthermore, in order to verify the capability of the proposed BSO-EM-T algorithm in handling outliers, 20 Monte Carlo(MC) runs are carried out when the output data contaminated by 5%, 10%, and 20% outliers, respectively. Table 2 presents the parameter estimates, δ and ADE under 20 MC runs with different outliers percentages. The parameter estimation error $\delta(\%)$ with the proportion of outliers(PO) and the iterative variable s is shown in Figure 5.

From Figures 1–5 and Tables 1,2, we can draw the following conclusions.

1. With the increasing of iteration variable s, the parameter estimates gradually approach the real value, and the errors become smaller. Those indicate that the proposed algorithm can well suppress the influence of outliers and time delays and exhibit good estimation performance–see Table 1 and Figure 2.
2. The estimated states $\hat{x}_{1,t}$ and $\hat{x}_{2,t}$ can well fit the real states, which indicate the effectiveness of the open-loop state observer–see Figure 3.
3. The estimated time delay can be well fitted to the real time delay, and the value of the estimated state transition matrix is very close to the real value. Those indicate that the BSO-EM-T algorithm can accurately estimate the unknown time delays–see Figure 4.
4. In the case of different outlier ratios, the estimated parameters are still very close to the actual values under 20 MC runs, which further proves that the BSO-EM-T algorithm can effectively estimate parameters of bilinear system under outliers and time-varying delays–see Table 2.

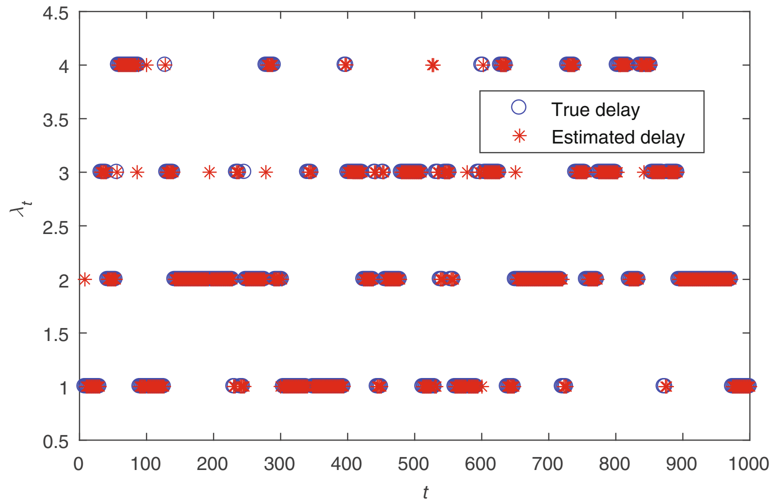

FIGURE 4 The comparison of true delays and estimated delays with 5% outliers.

TABLE 2 The parameter estimates, δ and ADE under 20 MC runs with different outliers percentages.

Outliers	5%	10%	20%	True values
a_1	0.5400 ± 0.0096	0.5369 ± 0.0110	0.5423 ± 0.0087	0.5400
a_2	0.1482 ± 0.0052	0.1465 ± 0.0057	0.1519 ± 0.0052	0.1500
d_{11}	0.0394 ± 0.0059	0.0408 ± 0.0059	0.0424 ± 0.0092	0.0400
d_{12}	0.2826 ± 0.0085	0.2817 ± 0.0086	0.2703 ± 0.0135	0.2800
d_{21}	0.4498 ± 0.0083	0.4496 ± 0.0090	0.4498 ± 0.0074	0.4500
d_{22}	0.1993 ± 0.0155	0.2017 ± 0.0186	0.1901 ± 0.0113	0.2000
f_1	1.2014 ± 0.0080	1.2029 ± 0.0087	1.2151 ± 0.0077	1.2000
f_2	0.4054 ± 0.0164	0.4033 ± 0.0179	0.4114 ± 0.0123	0.4000
$\delta(\%)$	1.1197 ± 0.5757	1.1641 ± 0.5998	1.7453 ± 0.7349	
ADE(%)	95.90 ± 1.00	93.81 ± 1.40	86.13 ± 1.60	

5. The lower the proportion of outliers, the higher the accuracy of the parameter estimation. This shows that the precision of parameter estimation is related to the quality of data–see Figure 5.

Finally, in order to the advantages of the proposed algorithm, we introduce the identification algorithm under the EM algorithm (EM-G) in Reference 36 and the bilinear state observer based gradient iterative algorithm(BSO-GI) in Reference 37 for comparison. Considering that BSO-GI algorithm does not involve time delays, the output data with known time delay is applied to identify this model. The maximum iteration variable, true transition probability matrix, and initial values of the BSO-GI and EM-G are set the same as the above BSO-EM-T algorithm. Figure 6 shows the estimated output and true output under three algorithms with 5% outliers. As the results show, compared to the BSO-GI algorithm and EM-G algorithm, The estimated output by BSO-EM-T algorithm can better match the real data, which is because Student's t distribution with smaller DOF can assign higher probability density to the tails.

The root-mean-square error (RMSE) is general applied to explain the relative error between the true outputs and the predicted outputs. The RMSE is defined as follow,

$$\mathrm{RMSE} = \sqrt{\sum_{t=1}^{N} (y_t - \hat{y}_t)^2 / N},$$

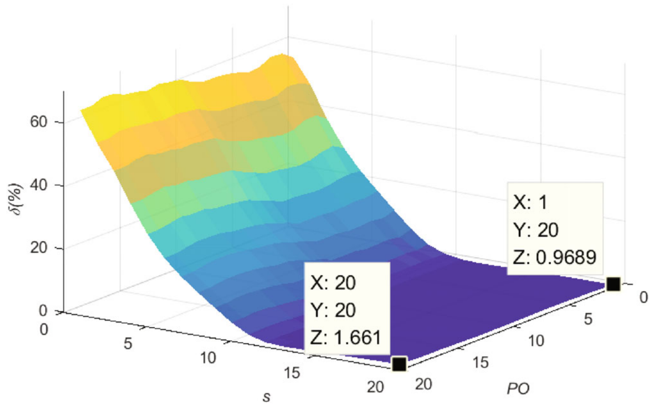

FIGURE 5 $\delta(\%)$ under different data quality.

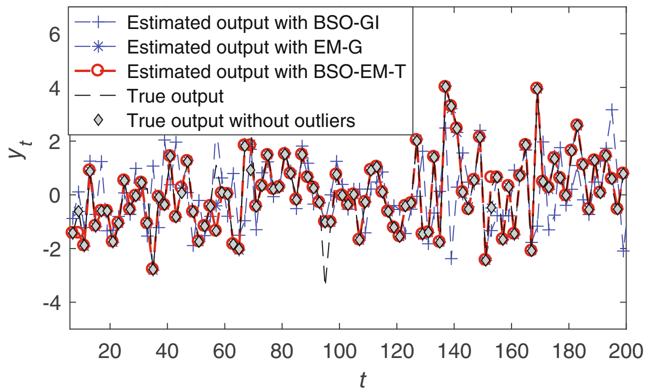

FIGURE 6 The estimated output under three algorithm with 5% outliers.

where N means the data length, y_t symbols the true output, and $\hat{y}_t$ means the predicted output. The result of parameter estimates, δ and RMSE of the three algorithms under 20 MC runs with 5% outliers is summarized in Table 3. As can be seen, when the output data is disturbed by outliers, the parameters estimates based on BSO-GI algorithm and EM-G algorithm deviate greatly from the true value, the proposed BSO-EM-T algorithm yields significant improvements in estimating the parameters and predicting output.

5.2 | Continuous stirred tank reactor

In this section, to re-verify the above observations, CSTR process is employed to verify the effectiveness of the BSO-EM-T algorithm. The dynamic differential models of the CSTR process are display as follows:[38,39]

$$\frac{dC_A(t)}{dt} = \frac{q(t)}{V}[C_{A0}(t) - C_A(t)] - k_0 C_A(t)\exp\left(\frac{-E}{RT(t)}\right),$$

$$\frac{dT(t)}{dt} = \frac{q(t)}{V}[T_0(t) - T(t)] - \frac{(-\Delta H)k_0 C_A(t)}{\rho C_p} \times \exp\left(\frac{-E}{RT(t)}\right) + \frac{\rho C_{pc}}{\rho C_p V} q_c(t) \times \left\{1 - \exp\left(\frac{-hA}{q_c(t)\rho C_p}\right)\right\}[T_{c0}(t) - T(t)],$$

where $q_c(t)$, $C_A(t)$ and $T(t)$ represent the coolant flow rate, the concentration of outlet reagent and the temperature of the reactor, respectively. Table 4 shows nominal parameter values of the CSTR process.

TABLE 3 The parameter estimates, δ and RMSE under 20 MC runs with 5% outliers for three algorithms.

Algorithms	BSO-EM-T	EM-G	BSO-GI	True values
a_1	0.5400 ± 0.0096	0.4905 ± 0.0083	0.1155 ± 0.0073	0.5400
a_2	0.1482 ± 0.0052	0.1423 ± 0.0042	0.1515 ± 0.0049	0.1500
d_{11}	0.0396 ± 0.0059	0.0502 ± 0.0039	0.1629 ± 0.0041	0.0400
d_{12}	0.2826 ± 0.0085	0.2643 ± 0.0074	0.2217 ± 0.0079	0.2800
d_{21}	0.4498 ± 0.0083	0.4601 ± 0.0062	0.4671 ± 0.0071	0.4500
d_{22}	0.1993 ± 0.0155	0.2006 ± 0.0121	-0.0022 ± 0.0152	0.2000
f_1	1.2014 ± 0.0080	1.1947 ± 0.0063	1.2002 ± 0.0079	1.2000
f_2	0.4064 ± 0.0164	0.3283 ± 0.0094	0.1104 ± 0.0107	0.4000
$\delta(\%)$	1.1197 ± 0.5757	6.0617 ± 0.0076	38.0490 ± 0.8711	
RMSE	0.3254 ± 0.0034	0.3567 ± 0.0032	0.6436 ± 0.0159	

TABLE 4 Nominal parameter values of the CSTR process.

Parameter	Value	Unit
Coolant flow rate, q_c	Input	L/min
Product concentration of component A, C_A	Output1	mol/L
Temperature of the reactor, T	Output2	K
Process flow rate, q	100	L/min
Feed concentration of component A, C_{A0}	1	mol/L
Feed temperature, T_0	350.0	K
Inlet coolant temperature, T_{c0}	350.0	K
Specific heats, C_p, C_{pc}	1	cal/(g k)
Heat transfer term, hA	7×10^5	cal/(min k)
Reactor volume, V	100	L
Liquid density, ρ, ρ_c	1×10^3	g/L
Heat of reactor, $-\Delta H$	-2×10^5	cal/mol
Activation energy term, E/R	1×10^4	k
Reaction rate constant, k_0	7.2×10^{10}	min^{-1}

In this simulation, we regard $q_c(t)$ as the input variable, $C_A(t)$ as the output variable. According to the process requirement,[40] the operating condition $q_c(t)$ =100 L/min is chosen to simulate the variation of the dynamic. A random binary signal with amplitude of 1 is added on $q_c(t)$ to generate the input data. Using the differential equation of CSTR to generate the output $C_A(t)$.

Set the unknown time delay switches among $\{1, 2, 3\}$. the transform of time delay is governed by the following Markov transition probability matrix:

$$\boldsymbol{P} = [\alpha_{ij}] = \begin{bmatrix} 0.950 & 0.100 & 0.020 \\ 0.010 & 0.870 & 0.080 \\ 0.040 & 0.030 & 0.900 \end{bmatrix}.$$

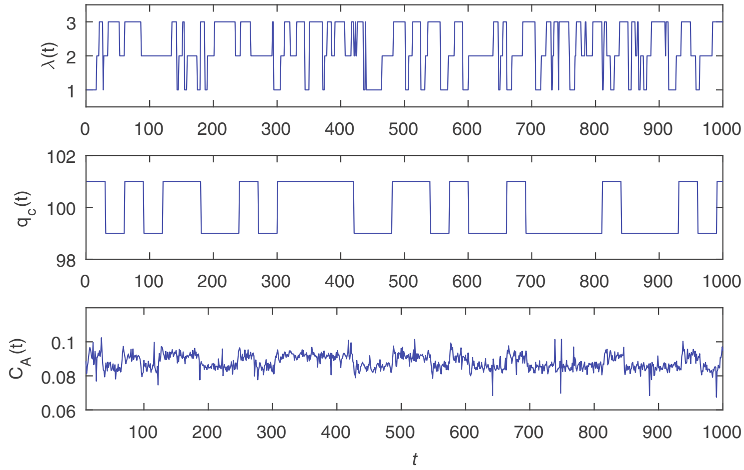

FIGURE 7 The delays, inputs, outputs of CSTR with outputs polluted by random 5% outliers.

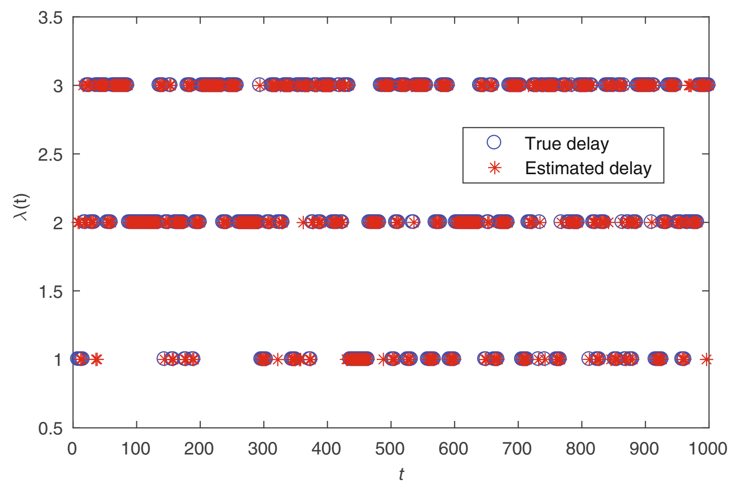

FIGURE 8 Comparison of true delays and that estimated of CSTR with 5% outliers.

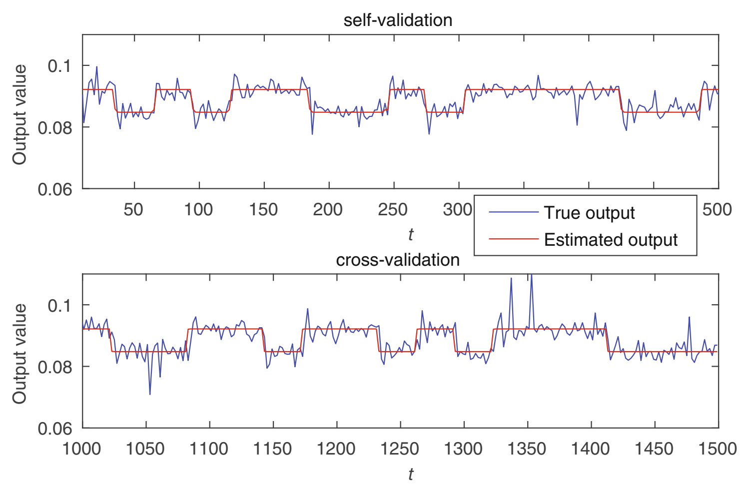

FIGURE 9 The results of self-validation and cross-validation with 5% outliers.

In order to verify the ability of the algorithm to deal with outliers, 5% outliers generated from the uniform distribution $\mu[-0.02, 0.02]$ and a Gaussian noise sequence with zero mean, variance $\sigma^2 = 0.001^2$ are added to the output data. Thousand sets of simulation signal with outputs polluted by random 5% outliers are presented in Figure 7.

According the collected data, the proposed algorithm is adopted to estimate the parameters of CSTR. Take the maximum iteration variable $S_{\max} = 30$, the estimated transition probability matrix is

$$\hat{\boldsymbol{P}} = \begin{bmatrix} 0.9322 & 0.0939 & 0.0331 \\ 0.0073 & 0.8811 & 0.0958 \\ 0.0524 & 0.0232 & 0.8867 \end{bmatrix}.$$

Obviously, the value of each element in the estimated Markov transition probability matrix is very close to the value in the true transition probability matrix. The true delay and the estimated delay are compared in Figure 8 with ADE=95.2%. It can be seen that the proposed BSO-EM-T algorithm can generate good delay estimation performance. The results of self-validation and cross-validation of CSTR with 5% outliers are given in Figure 9, which indicate that estimated output can capture the dynamics of the CSTR process well and validate the effectiveness of the algorithm.

6 | CONCLUSIONS

This paper proposes a BSO-EM-T algorithm for identifying the unknown parameters of the bilinear state-space model in the presence of time-varying delays and outliers. Based on the idea of the auxiliary model, the unknown state in the information vector is replaced by the estimated state obtained from the bilinear state observer. In order to handle the irregular outliers and time-varying delays, the Students' t distribution is applied to describe the measurement noise, a first-order Markov chain is adopted to model the delays. Under the framework of EM algorithm, the system parameters, unknown states, noise variance, undetermined delays and transition probability matrix are estimated iteratively. The effectiveness of the BSO-EM-T algorithm is validated by a numerical simulation and a CSTR process. All results show that the proposed BSO-EM-T algorithm can well overcome the effects of outliers and time-varying delays on parameter identification and generate good parameter estimation performance.

ACKNOWLEDGMENTS

This research was funded by the National Natural Science Foundation of China (no. 61773182).

ORCID

Xinyue Wang https://orcid.org/0000-0002-9958-2463
Junxia Ma https://orcid.org/0000-0002-0151-3188

REFERENCES

1. Liu Q, Tang X, Li J, et al. Identification of wiener-Hammerstein models based on variational Bayesian approach in the presence of process noise. *J Franklin Inst*. 2021;358(10):5623-5638.
2. Liu Q, Ding F. Gradient-based recursive parameter estimation for a periodically nonuniformly sampled-data Hammerstein–wiener system based on the key-term separation. *Int J Adapt Control Signal Process*. 2021;35:1970-1989.
3. Gan M, Chen GY, Chen L, et al. Term selection for a class of separable nonlinear models. *IEEE Trans Neural Netw Learn Syst*. 2020;31(2):445-451.
4. Patten BC, StraKraba M, Jorgensen SE. Ecosystems emerging: 1. Conservation. *Ecol Model*. 1997;96(1-3):221-284.
5. Mohler RR, Shen CN. *Optimal Control of Nuclear Reactors*. Academic; 1970.
6. Williamson D. Observation of bilinear systems with application to biological control. *Automatica*. 1977;13(3):243-254.
7. Zhang X, Ding F, Alsaadi FE, et al. Recursive parameter identification of the dynamical models for bilinear state space systems. *Nonlinear Dyn*. 2017;89(4):415-2429.

8. Fei QL, Ma JX, Xiong WL. Variational Bayesian identification for bilinear state space models with Markov-switching time delays. *Int J Robust Nonlinear Control*. 2020;30(17):7478-7495.
9. Liu WT, Xiong WL. Generalized continuous mixed p-norm based sliding window algorithm for a bilinear system with impulsive noise. *Int J Robust Nonlinear Control*. 2022;32(13):7663-7681. doi:10.1002/rnc.6236
10. Liu SY, Ding F, Hayat T. Moving data window gradient-based iterative algorithm of combined parameter and state estimation for bilinear systems. *Int J Robust Nonlinear Control*. 2020;30(6):2413-2429.
11. Li MH, Liu XM. Iterative identification methods for a class of bilinear systems by using the particle filtering technique. *Int J Adapt Control Signal Process*. 2021;35(10):2056-2074.
12. Gu Y, Dai W, Zhu QM, et al. Hierarchical multi-innovation stochastic gradient identification algorithm for estimating a bilinear state-space model with moving average noise. *J Comput Appl Math*. 2023;420:114794.
13. Liu XP, Yang XQ. Variational identification of linearly parameterized nonlinear state-space systems. *IEEE Trans Control Syst Technol*. 2022;31(4):1844-1854. doi:10.1109/TCST.2023.3249042
14. Liu XP, Yang XQ. Identification of nonlinear state-space systems with skewed measurement noises. *IEEE Trans Circuits Syst I-Regul Pap*. 2022;69(11):4654-4662.
15. Liu XP, Yang XQ. Robust variational inference for LPV dual-rate systems with randomly delayed outputs. *IEEE Trans Instrum Meas*. 2021;70:1-9. doi:10.1109/TIM.2021.3067242
16. Ma J, Xiong W, Chen J, et al. Expectation maximization estimation algorithm for Hammerstein models with non-Gaussian noise and random time delay from dual-rate sampled-data. *Digit Signal Process*. 2018;73:135-144.
17. Liu X, Yang X, Yin S. Nonlinear system identification with robust multiple model approach. *IEEE Trans Control Syst Technol*. 2019;28(6):2728-2735.
18. Liu X, Yang X. A variational Bayesian approach for robust identification of linear parameter varying systems using mixture Laplace distributions. *Neurocomput*. 2020;395(5):15-23.
19. Liu YY, Wang HW, Zhang W. Robust parameter estimation with outlier-contaminated correlated measurements and applications to aerodynamic coefficient identification. *Aerosp Sci Technol*. 2021;118:106995.
20. Liu X, Yang XQ, Zhu PB, et al. Robust multimodel identification of LPV systems with missing observations based on t-distribution. *IEEE Trans Syst Man Cybern*. 2019;51(8):5254-5260.
21. Yang XQ, Yin S, Kaynak O. Robust identification of LPV time-delay system with randomly missing measurements. *IEEE Trans Syst Man Cybern Syst*. 2018;48(12):2198-2208.
22. Yang XQ, Liu X, Yin S. Robust identification of nonlinear systems with missing observations: the case of state-space model structure. *IEEE Trans Ind Inform*. 2019;15(5):2763-2774.
23. Xu C, Zhao SY, Liu F. Sensor fault detection and diagnosis in the presence of outliers. *Neurocomput*. 2019;349:156-163.
24. Chen X, Zhao SY, Liu F. Robust identification of linear ARX models with recursive EM algorithm based on student's t-distribution. *J Franklin Inst*. 2021;358(1):1103-1121.
25. Zhao SY, Ma Y, Huang B. Robust FIR state estimation of dynamic processes corrupted by outliers. *IEEE Trans Ind Inform*. 2019;15(1):139-147.
26. Liu S, Zhang Y, Ding F, et al. Three-stage least squares-based iterative estimation algorithms for bilinear state-space systems based on the bilinear state estimator. *Int J Adapt Control Signal Process*. 2020;34:1501-1518.
27. Zhang X, Liu Q, Ding F, et al. Recursive identification of bilinear time-delay systems through the redundant rule. *J Franklin Inst*. 2020;357(1):726-747.
28. Chen X, Zhao SY, Liu F. Identification of jump Markov autoregressive exogenous systems with missing measurements. *J Franklin Inst*. 2019;357(6):3498-3523.
29. Chen X, Zhao SY, Liu F. Online identification of time-delay jump Markov autoregressive exogenous systems with recursive expectation-maximization algorithm. *Int J Adapt Control Signal Process*. 2020;34(1):407-426.
30. Yan AJ, Guo JC, Wang DH. Robust stochastic configuration networks for industrial data modelling with student's-t mixture distribution. *Inform Sci*. 2022;607:493-505.
31. Zhang TY, Zhao SY, Luan XL, Liu F. Bayesian inference for state-space models with student-t mixture distributions. *IEEE Trans Cybern*. 2022;53(7):4435-4445. doi:10.1109/TCYB.2022.3183104
32. Svensen M, Bishop CM. Robust Bayesian mixture modelling. *Neurocomput*. 2005;64:235-252.
33. Dempster AP. Maximum likelihood from incomplete data via the EM algorithm. *J R Stat Soc*. 1977;39(1):1-38.
34. Liu X, Yang X, Zhu P, et al. Robust identification of nonlinear time-delay system in state-space form. *J Franklin Inst*. 2019;356(16):9953-9971.
35. Boyles RA. On the convergence of the EM algorithm. *J R Stat Soc*. 1983;45(1):47-50.
36. Dong XW, Sun F, Xu FC, et al. Three-parameter estimation method of multiple hybrid Weibull distribution based on the EM optimization algorithm. *Mathematics*. 2022;10(22):4337. doi:10.3390/math10224337
37. Ding F, Xu L, Meng D, et al. Gradient estimation algorithms for the parameter identification of bilinear systems using the auxiliary model. *J Comput Appl Math*. 2019;369:112575.
38. Senthil R, Janarthanan K, Prakash J. Nonlinear state estimation using fuzzy Kalman filter. *Ind Eng Chem Res*. 2006;45(25): 8678-8688.

39. Guo F, Wu O, Hariprasad K, Ding Y, Huang B. An augmented model approach for identification of nonlinear errors-in-variables systems using the EM algorithm. *IEEE Trans Syst Man Cybern*. 2017;48(11):1968-1978.
40. Fang MQ, Kodamana H, Huang B. Real-time mode diagnosis for processes with multiple operating conditions using switching conditional random fields. *IEEE Trans Ind Electron*. 2020;67(6):5060-5070.

How to cite this article: Wang X, Ma J, Xiong W. Expectation-maximization algorithm for bilinear state-space models with time-varying delays under non-Gaussian noise. *Int J Adapt Control Signal Process*. 2023;37(10):2706-2724. doi: 10.1002/acs.3657