Nonlinear Process Identification in the Presence of Multiple Correlated Hidden Scheduling Variables with Missing Data

Lei Chen
Engineering Research Center of Digitized Textile & Fashion Technology, Ministry of Education, Shanghai 201620, P.R. China

College of Information Sciences and Technology, Donghua University, Shanghai 201620, P.R. China

Key Laboratory of Advanced Process Control for Light Industry (Ministry of Education), Institute of Automation, Jiangnan University, Wuxi 214122, P.R. China

Shima Khatibisepehr and Biao Huang
Dept. of Chemical and Materials Engineering, University of Alberta, Edmonton, AB, Canada T6G 2G6

Fei Liu
Key Laboratory of Advanced Process Control for Light Industry (Ministry of Education), Institute of Automation, Jiangnan University, Wuxi 214122, P.R. China

Yongsheng Ding
Engineering Research Center of Digitized Textile & Fashion TechnologyMinistry of Education, Shanghai 201620 P.R. China

College of Information Sciences and Technology, Donghua University, Shanghai 201620, P.R. China

DOI 10.1002/aic.14866
Published online June 1, 2015 in Wiley Online Library (wileyonlinelibrary.com)

Identification of nonlinear processes in the presence of noise corrupted and correlated multiple scheduling variables with missing data is concerned. The dynamics of the hidden scheduling variables are represented by a state-space model with unknown parameters. To assure generality, it is assumed that the multiple correlated scheduling variables are corrupted with unknown disturbances and the identification dataset is incomplete with missing data. A multiple model approach is proposed to formulate the identification problem of nonlinear systems under the framework of the expectation-maximization algorithm. The parameters of the local process models and scheduling variable models as well as the hyperparameters of the weighting function are simultaneously estimated. The particle smoothing technique is adopted to handle the computation of expectation functions. The efficiency of the proposed method is demonstrated through several simulated examples. Through an experimental study on a pilot-scale multitank system, the practical advantages are further illustrated. © 2015 American Institute of Chemical Engineers *AIChE J*, 61: 3270–3287, 2015
Keywords: nonlinear system identification, multiple models, expectation maximization algorithm, particle smoother, missing observations, multiple scheduling variables

Introduction

Some real-life systems may be represented by linear models. In many cases, however, nonlinear models are required to approximate the dynamic behavior of the system under investigation. Due to the unique challenges of the complex stochastic nonlinear process systems, identification of nonlinear dynamic models in process industries has long attracted research interest.

Industrial processes are often operated along certain fixed operating trajectories, and several predetermined operating points are used to compose these operating trajectories.[1] Conventional single-model-based modeling techniques cannot sufficiently capture the underlying dynamics of such process systems. To overcome the shortcomings of the conventional techniques, researchers have developed various multiple modeling strategies.[2–5]

Gopaluni[6] represented nonlinear processes through combining several radial basis functions (RBFs), and the parameters of the RBFs are estimated using the expectation-maximization (EM) algorithm. Jin et al.[1] proposed an approach to construct a global model, and they synthesized multiple local models with transient data and estimated the parameters using the EM algorithm with consideration of a single scheduling variable. Huang et al.[7] studied the nonlinear system identification problem with two noise-free scheduling variables. In their study,

Correspondence concerning this article should be addressed to B. Huang at biao.huang@ualberta.ca.

different weighting functions such as linear, Gaussian, and polynomial were used and results were compared. Our earlier work[8,9] considered nonlinear identification problem with either single scheduling variable of known dynamics or multiple but exactly known scheduling variables under the EM framework.

However, in industrial settings, multiple scheduling variables often exist and may not be directly measured, which are correlated with each other and corrupted with unknown disturbances. Furthermore, the scheduling variables are hidden having unknown parameters of their own dynamics and may encounter missing observations. In this work, we consider the general cases in which the multiple and correlated scheduling variables are not measured directly, dynamics of the true scheduling variables are represented by an unknown stochastic state-space model, and the observed scheduling variables have missing data. Thus, the problem to be solved in this article consists of four components: (1) estimating the parameters of all the local ARX models and associate weighting function; (2) estimating the parameters of the state-space model of the hidden scheduling variables; (3) recovering the true scheduling variables for identifying the weighting function; and (4) handling missing data.

We use multiple auto regressive exogenous (ARX) models to approximate the local dynamics of nonlinear processes. The information contained in the noisy measurements that are related to the scheduling variables through the unknown dynamics are extracted to simultaneously identify different process operating conditions along with the corresponding local ARX models. A normalized weighting function is used to combine all the local ARX models and, consequently, obtain a global representation of the nonlinear process. Here, the weighting function is viewed as the probability associated with each of the local ARX models capturing the underlying dynamics.[1,5,10] The weighting function used in this work not only is a function of the hidden scheduling variables, but also contains unknown parameters.

EM algorithm is used to formulate and solve the aforementioned problems. In the proposed EM framework, the parameters of the local process models and scheduling variable models as well as the hyperparameters of the weighting function are simultaneously estimated. To estimate the parameters of all the local models, a closed form update rule is derived. Besides, the hyperparameters of the weighting function and the parameters of the state-space model of the scheduling variables are estimated by solving optimization problems. The efficacy of the proposed identification method is demonstrated on several simulation examples, including cases of linear and nonlinear hidden scheduling variables. Finally, the practicality of the method in real-life applications is illustrated through a pilot-scale experiment.

The remainder of this article is organized as follows: the second section discusses a multitank system as the motivation example. The third section presents the problem formulation. The fourth section starts from a brief review of the EM algorithm; this section is followed by formulating and solving the identification problem under the EM algorithm and the details of the derivations are also presented. The next section verifies the proposed method through several simulation examples considering both linear and nonlinear dynamics of the hidden scheduling variables with various percentage of missing observations. The experimental section presents the experiment conducted on a pilot-scale multitank system and renders verification results. Finally, concluding remarks are given for this article.

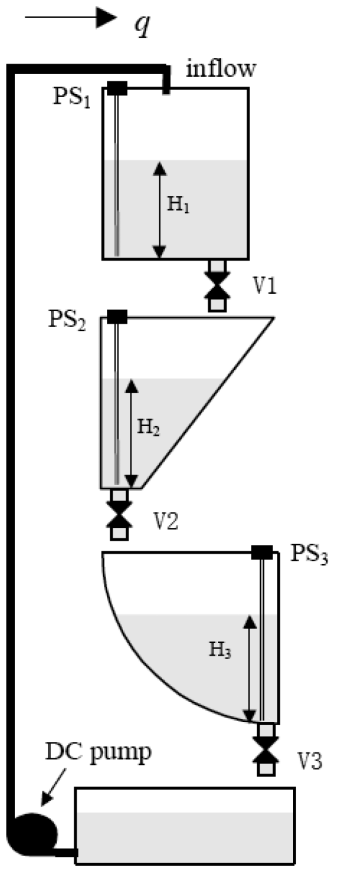


Figure 1. Schematic diagram of the multitank system.

Motivations

An illustrative example is presented in this section to motivate the problem to be addressed in this article. Consider a multiple tank system shown in Figure 1. There are three separate tanks, and each tank fitted with drain valves. In the base of the set-up, there is another tank as a reservoir for the system. The three tanks have different cross sections: constant, conical, and spherical. The nonlinearity of the system is a result of these variable cross sections. A variable speed pump is used to fill the upper tank. The water flows out of the three tanks by gravity, and the valves resist the water flow. The opening percentage of the valves can be controlled and used to vary the outflow characteristic. Each tank is equipped with a level sensor that operates on the basis of hydraulic pressure measurement.

A nonlinear model can be derived to describe the relations between the change in inflow q and the second tank level H_2, and two controlled valves V_1 and V_2 are the two correlated scheduling variables operating at miscellaneous steady states. Note that the true scheduling variables z_k are unknown, but can be inferred through other noisy measurements w_k, and the setpoint of the valve controllers, denoted by u'_k, is varied to reach the desired operating point. Let the inflow be the manipulated input u_k to the process, and let the second tank level be the process output y_k. The dynamics of the multiple tank system in Figure 1 or that of any other complex dynamical system can be described by the following model structure[8]

$$h(\dot{c}_k, c_k, y_k, u_k, z_k, w_k, u'_k, T, k, \epsilon_k) = 0 \tag{1}$$

where $y_k \in \mathscr{Y} \subseteq \mathbb{R}$ and $c_k \in \mathscr{C} \subseteq \mathbb{R}^r$ for $k \in \mathbb{N}$ are the measurement and state, respectively; $u_k \in \mathscr{U} \subseteq R^m$ is the nonlinear process input; $z_k = \{z_k^1, z_k^2, \ldots, z_k^s\} \in \mathscr{Z} \subseteq \mathbb{R}^s$ and $w_k = \{w_k^1, w_k^2, \ldots, w_k^s\} \in \mathscr{W} \subseteq \mathbb{R}^s$ are the hidden multivariate scheduling variables and observations related to the hidden scheduling variables through unknown dynamics; $\epsilon_k \in \mathbb{R}$ is the process noise; $u_k' \in \mathscr{U}' \subseteq \mathbb{R}^p$ is the source that drives the scheduling variables. $h(\cdot)$ is an unknown nonlinear mapping function. Define $T_{1:\lambda}^{\kappa} = \{T_1^{\kappa}, T_2^{\kappa}, \ldots, T_{\lambda}^{\kappa}\}$ as λ nominal operating points of the κth scheduling variable, where the superscript represents the κth scheduling variable, and the subscript represents λ different operating points. $T = \{T_{1:M_1}^1, T_{1:M_2}^2, \ldots, T_{1:M_s}^s\}$ is the collection of nominal operating points for all s different scheduling variables. Although z_k in Eq. 1 is hidden, it is related to an observation w_k, and the dynamics of z_k can be represented using the following state-space model

$$z_k = f_{\theta_f}(z_{k-1}, u_{k-1}', \gamma_{k-1}) \tag{2a}$$

$$w_k = g_{\theta_g}(z_k, u_k', v_k) \tag{2b}$$

where $f_{\theta_f}(\cdot)$ and $g_{\theta_g}(\cdot)$ are s-dimensional state and output mapping functions which can be either linear or nonlinear, and the structure of $f_{\theta_f}(\cdot)$ and $g_{\theta_g}(\cdot)$ are given but parameterized by a set of unknown parameters, $\theta_z = \{\theta_f, \theta_g\}$. Moreover, γ_k and v_k are mutually independent Gaussian noise terms represented by a probability distribution function $p_{\gamma}(\cdot)$ and $p_v(\cdot)$, respectively. Due to the stochastic nature of Eq. 2, random variables of interest can be represented as follows

$$Z_0 = z_0 \sim p_{\theta_f}(z_0) \tag{3}$$

$$Z_k|(Z_{k-1} = z_{k-1}) \sim p_{\theta_f}(\cdot|z_{k-1}) \tag{4}$$

$$W_k|(Z_k = z_k) \sim p_{\theta_g}(\cdot|z_k) \tag{5}$$

where u_k' is dispensed in Eqs. 4 and 5 for brevity as it is a known input.

For any generic sequence, χ_k, let $\chi_{i:j} = \{\chi_i, \chi_{i+1}, \ldots, \chi_j\}$ with $\chi_{i:j} = 0$ for $i > j$. The historical data for scheduling variables are denoted by $w_{1:N} = \{w_{\text{obs}}, w_{\text{mis}}\}$, where $w_{\text{mis}} = \{w_{m_1}, w_{m_2}, \ldots, w_{m_a}\}$ is the missing data, hereinafter denoted as $w_{m_1:m_a}$, and $w_{\text{obs}} = \{w_{o_1}, w_{o_2}, \ldots, w_{o_b}\}$ is the observed data, hereinafter denoted as $w_{o_1:o_b}$. It is assumed that we have the historical data for input and output variables, $\{u_{1:N}, y_{1:N}\}$, as well as the operating points of the scheduling variables, T, shown in Eq. 1. It is also assumed that the dataset $\{u'_{1:N}, w_{o_1:o_b}\}$ is given. Then, $C_{\text{obs}} = \{y_{1:N}, u_{1:N}, u'_{1:N}, w_{o_1:o_b}, T\}$ can represent the observed dataset.

Problem Formulation

Nonlinear dynamics in different operating conditions is represented by concatenating among the multiple local models. The local dynamics of Eq. 1 are approximated using ARX models. Ljung[11] has justified that any linear dynamics may be approximated using ARX models. The ARX model is given by

$$y_k = \theta_{I_k}^{\mathrm{T}} x_k + e_k \tag{6}$$

where $e_k \in \mathbb{R}$ is a zero mean Gaussian white noise, and the variance σ^2 is unknown; $x_k \in \mathscr{X} \subseteq \mathbb{R}^n$ is the regressor expressed as

$$x_k \triangleq [y_{k-1}, y_{k-2}, \ldots, y_{k-n_a}, u_{k-1}^{\mathrm{T}}, u_{k-2}^{\mathrm{T}}, \ldots, u_{k-n_b}^{\mathrm{T}}]^{\mathrm{T}} \tag{7}$$

where n_a and n_b are the orders of the process output and input polynomial, respectively, such that $n = n_a + mn_b$. Here, it is assumed that $u_k = 0$, $y_k = 0$ for $k \leq 0$. I_k represents the identity of the local model at sampling time k, and is a hidden variable, so that $\Theta_{I_k} = \{\theta_{I_k}, \sigma\}$ are the local model parameters to be estimated. Let $I_k = (m_1, m_2, \ldots, m_s)$ represent the local model identity at $w_k^1 = T_{m_1}^1, w_k^2 = T_{m_2}^2, \ldots, w_k^s = T_{m_s}^s$ operating points, where $m_1 = \{1, 2, \ldots, M_1\}, m_2 = \{1, 2, \ldots, M_2\}, \ldots, m_s = \{1, 2, \ldots, M_s\}$. For notational simplicity, we use $I_k = \pi$ to denote the local model in the remainder of this article.

Let $M_1, M_2, \ldots, M_s$ operating points be equipped with $\mathscr{M}$ local ARX models where $\mathscr{M} = M_1 \times M_2 \times \cdots \times M_s$, such that

$$Y_k|(X_k = x_k, I_k = \pi) \sim p_{\Theta_\pi}(\cdot|x_k), \text{ for all } 1 \leq \pi \leq \mathscr{M} \tag{8}$$

Within a relatively small region around each of the operating points, the local process dynamics can be approximated by an identified local model. Given all the past information, the probability of the observed output y_k can be determined from the law of total probability as

$$\begin{aligned} &p_\Theta(y_k|y_{1:k-1}, c_{1:k}, u_{1:k}, z_{0:k}, w_{1:k}, u'_{1:k}, T) \\ &= \sum_{m_1=1}^{M_1} \sum_{m_2=1}^{M_2} \cdots \sum_{m_s=1}^{M_s} p_\Theta(y_k, I_k = \pi| \\ &\quad y_{1:k-1}, c_{1:k}, u_{1:k}, z_{0:k}, w_{1:k}, u'_{1:k}, T) \end{aligned} \tag{9a}$$

$$\begin{aligned} &= \sum_{m_1=1}^{M_1} \sum_{m_2=1}^{M_2} \cdots \sum_{m_s=1}^{M_s} \alpha_{k,\pi} p_\Theta(y_k|y_{1:k-1}, \\ &\quad c_{1:k}, u_{1:k}, z_{0:k}, w_{1:k}, u'_{1:k}, T, I_k = \pi) \end{aligned} \tag{9b}$$

$$= \sum_{m_1=1}^{M_1} \sum_{m_2=1}^{M_2} \cdots \sum_{m_s=1}^{M_s} \alpha_{k,\pi} p_{\Theta_\pi}(y_k|x_k) \tag{9c}$$

where Θ is the unknown parameters to be estimated, and we have employed the chain rule of probability to Eq. 9a. Note π is a value reflecting various combinations of $m_1, \ldots, m_s$. Here, $\alpha_{k,\pi} = \Pr_\Theta(I_k = \pi|y_{1:k-1}, c_{1:k}, u_{1:k}, z_{0:k}, w_{1:k}, u'_{1:k}, T)$ is the probability that $I_k = \pi$th local ARX model takes effect at sampling time k.

To make a distinction of the probability between the discrete and continuous random variables, $\Pr(\cdot)$ is used to represent a discrete probability distribution, and $p(\cdot)$ is used to represent a probability density function (PDF).

Equation 8 states that given the identity of the local model I_k, y_k becomes independent of $c_{1:k}$, $z_{0:k}$, $w_{1:k}$, $u'_{1:k}$,, and T. That is, $p_\Theta(y_k|y_{1:k-1}, c_{1:k}, u_{1:k}, z_{0:k}, w_{1:k}, u'_{1:k}, T, I_k = \pi)$ in Eq. 9b can be simplified as $p_{\Theta_\pi}(y_k|x_k)$ in deriving Eq. 9c. Furthermore, we assume that the identity of local model I_k can be directly inferred from z_k and T. Thus, $\alpha_{k,\pi}$ can be simplified as $\Pr_\Theta(I_k = \pi|z_k, T)$.

In this article, a normalized weighting function of $\varpi_{k,\pi}$ is employed to model $\alpha_{k,\pi}$ as

$$\alpha_{k,\pi} = \Pr_\Theta(I_k = \pi|z_k, T) \tag{10a}$$

$$= \frac{\varpi_{k,\pi}}{\sum_{m_1=1}^{M_1} \sum_{m_2=1}^{M_2} \cdots \sum_{m_s=1}^{M_s} \varpi_{k,\pi}} \tag{10b}$$

where, $z_k^1, z_k^2, \ldots, z_k^s$ are correlated with each other such that

$$\varpi_{k,\pi} = \exp\left[-\frac{1}{2}\begin{pmatrix} z_k^1 - T_{m_1}^1 \\ z_k^2 - T_{m_2}^2 \\ \vdots \\ z_k^s - T_{m_s}^s \end{pmatrix}^{\mathrm{T}} \Sigma^{-1} \begin{pmatrix} z_k^1 - T_{m_1}^1 \\ z_k^2 - T_{m_2}^2 \\ \vdots \\ z_k^s - T_{m_s}^s \end{pmatrix}\right] \tag{11}$$

where Σ is a symmetric covariance matrix

$$\Sigma = \begin{bmatrix} o_{1,1} & o_{1,2} & \cdots & o_{1,s} \\ o_{2,1} & o_{2,2} & \cdots & o_{2,s} \\ \vdots & \vdots & \ddots & \vdots \\ o_{s,1} & o_{s,2} & \cdots & o_{s,s} \end{bmatrix} \tag{12}$$

with $o_{i,j} = \{o^{i,j}_{1,1,\ldots,1}, \ldots, o^{i,j}_{m_1,m_2,\ldots,m_s}, \ldots, o^{i,j}_{M_1,M_2,\ldots,M_s}\} \subseteq \mathcal{O} \in \mathbb{R}^{M_1} \times \mathbb{R}^{M_2} \times \cdots \times \mathbb{R}^{M_s}$, $o_\pi = [o^{i,j}_{m_1,m_2,\ldots,m_s}]_{s\times s}$ being the validity width of the $I_k = \pi$th local model, o_π is bounded between $o_{\min}$ (the lower bound) and $o_{\max}$ (the upper bound), and the parameters for the weighting function that need to be estimated is $O = [o_\pi]_{M_1\times M_2\times\cdots\times M_s}$.

Due to the stochastic nature of Eq. 10a, random variables of interest can be represented as

$$I_k|(Z_k = z_k, T) \sim \Pr_O(\cdot|z_k, T) \tag{13}$$

Since at each sampling time k, I_k and z_k are hidden variables, then $C_{\text{mis}} = \{I_{1:N}, z_{0:N}, w_{m_1:m_a}\}$ can be denoted as the missing dataset, $\{C_{\text{obs}}, C_{\text{mis}}\}$ is the complete dataset, and $\Theta = \{\Theta_{I_k}, O, \theta_z\} = \cup_{m_1=1}^{M_1}\cup_{m_2=1}^{M_2}\cdots\cup_{m_s=1}^{M_s}\{\theta_\pi, \sigma, o_\pi, \theta_z\}$ are the parameters to be estimated.

The estimated global model can thus be expressed as

$$\hat{y}_k = \sum_{m_1=1}^{M_1}\sum_{m_2=1}^{M_2}\cdots\sum_{m_s=1}^{M_s}\alpha_{k,\pi}\hat{y}_k^\pi \tag{14}$$

where $\hat{y}_k^\pi$ is the estimated output of the πth local ARX model (the conditional mean of each local output) at time k.

Estimating the parameter set Θ is the main problem to be addressed in this article, and C_{obs} and the function forms of Eq. 2a are given. To solve this problem, we resort to the EM algorithm.

Formulation of the Multiple Model Approach Based on the EM Algorithm

The EM algorithm was first proposed by Dempster for maximum-likelihood estimation, where the likelihood function involves hidden variables or missing observations.[12] After given some initial guess of the parameters, the EM algorithm iterates between two steps: the expectation step (E-step) and the maximization step (M-step). In the E-step, the Q-function is the expectation of the log likelihood function for the complete dataset $\{C_{\text{obs}}, C_{\text{mis}}\}$, and the expectation is computed with respect to C_{mis} given the observed data C_{obs} and the current parameter estimates Θ'. In the M-step, parameter estimates are updated through maximizing the Q-function. The E-step and the M-step are iteratively repeated until the parameter estimates converge.

The EM algorithm can be summarized as follows[13]

1. *Initialization*: Give some initial guess of the parameters for Θ'.
2. *E-step*: Use the current parameter Θ' to evaluate the approximated Q-function
$$Q(\Theta|\Theta') = \mathrm{E}_{C_{\text{mis}}|(C_{\text{obs}},\Theta')}\{\log p_\Theta(C_{\text{obs}}, C_{\text{mis}})\} \tag{15}$$
3. *M-step*: Obtain the new iterative parameter through maximizing the Q-function
$$\Theta = \arg\max_\Theta Q(\Theta|\Theta') \tag{16}$$
and then set $\Theta' = \Theta$.
4. *Iterate*: Evaluate the relative change of the estimated parameters
$$\delta = \left\|\frac{\Theta - \Theta'}{\Theta'}\right\| \tag{17}$$

If δ is larger than a predetermined tolerance, then repeat step 2 and step 3.

There are several different categories of stopping criteria for terminating the EM algorithm.[14] EM algorithm can converge to a stationary point under certain regularity conditions.[15]

Formulation of the multiple model approach under the EM algorithm

The input sequence of the scheduling variables $u'_{1:N}$ is considered to be known and, consequently, does not play a role in the following derivations. Hereinafter, $u'_{1:N}$ is omitted from the PDFs for notation simplicity.

From the chain rule of probability, the complete likelihood function $p_\Theta(C_{\text{obs}}, C_{\text{mis}})$ can be decomposed as

$$\begin{aligned} p_\Theta(C_{\text{obs}}, C_{\text{mis}}) &= p_\Theta(y_{1:N}, u_{1:N}, w_{m_1:m_a}, T, I_{1:N}, z_{0:N}, w_{o_1:o_b}) \\ &= p_\Theta(y_{1:N}, u_{1:N}, w_{1:N}, T, I_{1:N}, z_{0:N}) \qquad (18a) \\ &= p_\Theta(y_{1:N}|u_{1:N}, w_{1:N}, T, I_{1:N}, z_{0:N}) \\ &\quad \cdot \Pr_\Theta(I_{1:N}|u_{1:N}, w_{1:N}, T, z_{0:N}) \\ &\quad \cdot p_\Theta(w_{1:N}|u_{1:N}, T, z_{0:N}) \\ &\quad \cdot p_\Theta(z_{0:N}|u_{1:N}, T) \\ &\quad \cdot p_\Theta(u_{1:N}, T) \qquad (18b) \end{aligned}$$

Applying the chain rule of probability, following the first-order Markov property, and taking into account the assumptions of $\chi_{i:j} = 0$ for $i > j$ and $y_k = 0$ for $k \le 0$, Eq. 18b can be simplified as

$$\begin{aligned} p_\Theta(C_{\text{obs}}, C_{\text{mis}}) = &\prod_{k=1}^N p_{\Theta_{I_k}}(y_k|x_k) \cdot \prod_{k=1}^N \Pr_O(I_k|z_k, T) \\ &\cdot \prod_{k=1}^N p_{\theta_z}(w_k|z_k) \cdot \prod_{k=1}^N p_{\theta_z}(z_k|z_{k-1}) \cdot p_{\theta_z}(z_0) \cdot C_1 \end{aligned} \tag{19}$$

The detailed derivation of Eq. 19 is given in Appendix A.

Thus, the Q-function in the E-step shown in Eq. 15 can be determined as

$$\begin{aligned} Q(\Theta|\Theta') = &\int_{I_{1:N}, z_{0:N}, w_{m_1:m_a}} \sum_{k=1}^N \log p_{\Theta_{I_k}}(y_k|x_k) p_{\Theta'}(I_{1:N}, z_{0:N}, w_{m_1:m_a}|C_{\text{obs}}) dz_{0:N} dI_{1:N} dw_{m_1:m_a} \\ &+ \int_{I_{1:N}, z_{0:N}, w_{m_1:m_a}} \sum_{k=1}^N \log \Pr_O(I_k|z_k, T) p_{\Theta'}(I_{1:N}, z_{0:N}, w_{m_1:m_a}|C_{\text{obs}}) dz_{0:N} dI_{1:N} dw_{m_1:m_a} \end{aligned}$$

$$
\begin{aligned}
&+\int_{I_{1:N},z_{0:N},w_{m_1:m_a}}\sum_{k=1}^{N}\log p_{\theta_z}(w_k|z_k)p_{\Theta'}(I_{1:N},z_{0:N},w_{m_1:m_a}|C_{\text{obs}})dz_{0:N}dI_{1:N}dw_{m_1:m_a}\\
&+\int_{I_{1:N},z_{0:N},w_{m_1:m_a}}\sum_{k=1}^{N}\log p_{\theta_z}(z_k|z_{k-1})p_{\Theta'}(I_{1:N},z_{0:N},w_{m_1:m_a}|C_{\text{obs}})dz_{0:N}dI_{1:N}dw_{m_1:m_a}\\
&+\int_{I_{1:N},z_{0:N},w_{m_1:m_a}}\log p_{\theta_z}(z_0)p_{\Theta'}(I_{1:N},z_{0:N},w_{m_1:m_a}|C_{\text{obs}})dz_{0:N}dI_{1:N}dw_{m_1:m_a}\\
&+\int_{I_{1:N},z_{0:N},w_{m_1:m_a}}\log C_1\cdot p_{\Theta'}(I_{1:N},z_{0:N},w_{m_1:m_a}|C_{\text{obs}})dz_{0:N}dI_{1:N}dw_{m_1:m_a}
\end{aligned}
\tag{20}
$$

The detailed derivation of Eq. 20 is given in Appendix B. Since the integrands in the first two terms of Eq. 20 are only a function of I_k and z_k, the multidimensional integrals with respect to $p_{\Theta'}(I_{1:N},z_{0:N},w_{m_1:m_a}|C_{\text{obs}})$ can be simplified and rewritten as[8]

$$
\begin{aligned}
&\int_{I_k,z_k}\sum_{k=1}^{N}\log p_{\Theta_{I_k}}(y_k|x_k)\Pr_{\Theta'}(I_k|z_k,C_{\text{obs}})p_{\Theta'}(z_k|C_{\text{obs}})dz_kdI_k\\
&+\int_{I_k,z_k}\sum_{k=1}^{N}\log \Pr_O(I_k|z_k,T)\Pr_{\Theta'}(I_k|z_k,C_{\text{obs}})p_{\Theta'}(z_k|C_{\text{obs}})dz_kdI_k
\end{aligned}
\tag{21}
$$

where $p_{\Theta'}(I_k,z_k|C_{\text{obs}})$ is decomposed as $\Pr_{\Theta'}(I_k|z_k,C_{\text{obs}})p_{\Theta'}(z_k|C_{\text{obs}})$.

Since the local model identity I_k is a discrete random variable, Eq. 21 can be further expressed as

$$
\begin{aligned}
&\int_{z_k}\sum_{k=1}^{N}\sum_{m_1=1}^{M_1}\cdots\sum_{m_s=1}^{M_s}\log p_{\Theta_\pi}(y_k|x_k)\Pr_{\Theta'}(I_k=\pi|z_k,C_{\text{obs}})p_{\Theta'}(z_k|C_{\text{obs}})dz_k\\
&+\int_{z_k}\sum_{k=1}^{N}\sum_{m_1=1}^{M_1}\cdots\sum_{m_s=1}^{M_s}\log \Pr_{o_\pi}(I_k=\pi|z_k,T)\Pr_{\Theta'}(I_k=\pi|z_k,C_{\text{obs}})p_{\Theta'}(z_k|C_{\text{obs}})dz_k
\end{aligned}
\tag{22}
$$

As w_k has two subsets of observed data $w_{o_1:o_b}$ and missing data $w_{m_1:m_a}$, the third term in Eq. 20 can be further expressed as

$$
\begin{aligned}
&\int_{I_{1:N},z_{0:N},w_{m_1:m_a}}\sum_{k=1}^{N}\log p_{\theta_z}(w_k|z_k)p_{\Theta'}(I_{1:N},z_{0:N},w_{m_1:m_a}|C_{\text{obs}})dz_{0:N}dI_{1:N}dw_{m_1:m_a}\\
&=\int_{I_{1:N},z_{0:N},w_{m_1:m_a}}\sum_{k=o_1}^{o_b}\log p_{\theta_z}(w_k|z_k)p_{\Theta'}(I_{1:N},z_{0:N},w_{m_1:m_a}|C_{\text{obs}})dz_{0:N}dI_{1:N}dw_{m_1:m_a}\\
&\quad+\int_{I_{1:N},z_{0:N},w_{m_1:m_a}}\sum_{k=m_1}^{m_a}\log p_{\theta_z}(w_k|z_k)p_{\Theta'}(I_{1:N},z_{0:N},w_{m_1:m_a}|C_{\text{obs}})dz_{0:N}dI_{1:N}dw_{m_1:m_a}
\end{aligned}
\tag{23a}
$$

$$
\begin{aligned}
&=\int_{z_k}\sum_{k=o_1}^{o_b}\log p_{\theta_z}(w_k|z_k)p_{\Theta'}(z_k|C_{\text{obs}})dz_k\\
&\quad+\int_{z_k,w_k}\sum_{k=m_1}^{m_a}\log p_{\theta_z}(w_k|z_k)p_{\Theta'}(z_k,w_k|C_{\text{obs}})dz_kdw_k
\end{aligned}
\tag{23b}
$$

As the integrand in the fourth term of Eq. 20 is a function of z_{k-1} and z_k, the multidimensional integral with respect to $p_{\Theta'}(I_{1:N},z_{0:N},w_{m_1:m_a}|C_{\text{obs}})$ can be simplified and written as

$$
\int_{z_k,z_{k-1}}\sum_{k=1}^{N}\log p_{\theta_z}(z_k|z_{k-1})p_{\Theta'}(z_k,z_{k-1}|C_{\text{obs}})dz_kdz_{k-1}
\tag{24}
$$

Finally, the last term in Eq. 20 can be simplified and written as

$$
\int_{z_0}\log p_{\theta_z}(z_0)p_{\Theta'}(z_0|C_{\text{obs}})dz_0
\tag{25}
$$

Substituting Eqs. 22, 23b, 24, and 25 into Eq. 20, yields

$$
\begin{aligned}
Q(\Theta|\Theta') = & \int_{z_k} \sum_{k=1}^{N} \sum_{m_1=1}^{M_1} \cdots \sum_{m_s=1}^{M_s} \log p_{\Theta_\pi}(y_k|x_k) \Pr_{\Theta'}(I_k = \pi|z_k, C_{\text{obs}}) p_{\Theta'}(z_k|C_{\text{obs}}) dz_k \\
& + \int_{z_k} \sum_{k=1}^{N} \sum_{m_1=1}^{M_1} \cdots \sum_{m_s=1}^{M_s} \log \Pr_{o_\pi}(I_k = \pi|z_k, T) \Pr_{\Theta'}(I_k = \pi|z_k, C_{\text{obs}}) p_{\Theta'}(z_k|C_{\text{obs}}) dz_k \\
& + \int_{z_k} \sum_{k=o_1}^{o_b} \log p_{\theta_z}(w_k|z_k) p_{\Theta'}(z_k|C_{\text{obs}}) dz_k \\
& + \int_{z_k, w_k} \sum_{k=m_1}^{m_a} \log p_{\theta_z}(w_k|z_k) p_{\Theta'}(z_k, w_k|C_{\text{obs}}) dz_k dw_k \\
& + \int_{z_k, z_{k-1}} \sum_{k=1}^{N} \log p_{\theta_z}(z_k|z_{k-1}) p_{\Theta'}(z_k, z_{k-1}|C_{\text{obs}}) dz_k dz_{k-1} \\
& + \int_{z_0} \log p_{\theta_z}(z_0) p_{\Theta'}(z_0|C_{\text{obs}}) dz_0 + C_2
\end{aligned} \tag{26}
$$

where $C_2 = \int_{I_{1:N}, z_{0:N}, w_{m_1:m_a}} \log C_1 \cdot p_{\Theta'}(I_{1:N}, z_{0:N}, w_{m_1:m_a}|C_{\text{obs}})$ $dz_{0:N} dI_{1:N} dw_{m_1:m_a}$. The detailed derivation is given in Appendix C.

Now to compute the Q-function in Eq. 26, the probability mass and density functions that need to be evaluated are as following:

1. $p_{\Theta_\pi}(y_k|x_k)$;
2. $\Pr_{o_\pi}(I_k = \pi|z_k, T)$;
3. $p_{\theta_z}(w_k|z_k)$;
4. $p_{\theta_z}(z_k|z_{k-1})$;
5. $\Pr_{\Theta'}(I_k = \pi|z_k, C_{\text{obs}})$;
6. $p_{\Theta'}(z_k|C_{\text{obs}})$, when measurements of the observed scheduling variables are available at time instant k, that is, $w_k \in w_{\text{obs}}$ with $w_{\text{obs}} = w_{o_1:o_b}$;
7. $p_{\Theta'}(z_k, w_k|C_{\text{obs}})$, when measurements of the observed scheduling variables are missing at time instant k, that is, $w_k \in w_{\text{mis}}$ with $w_{\text{mis}} = w_{m_1:m_a}$;
8. $p_{\Theta'}(z_k, z_{k-1}|C_{\text{obs}})$.

As Gaussian white noise has been assumed for the ARX model in Eq. 6, the likelihood function $p_{\Theta_\pi}(y_k|x_k)$ is a Gaussian PDF and can be expressed as

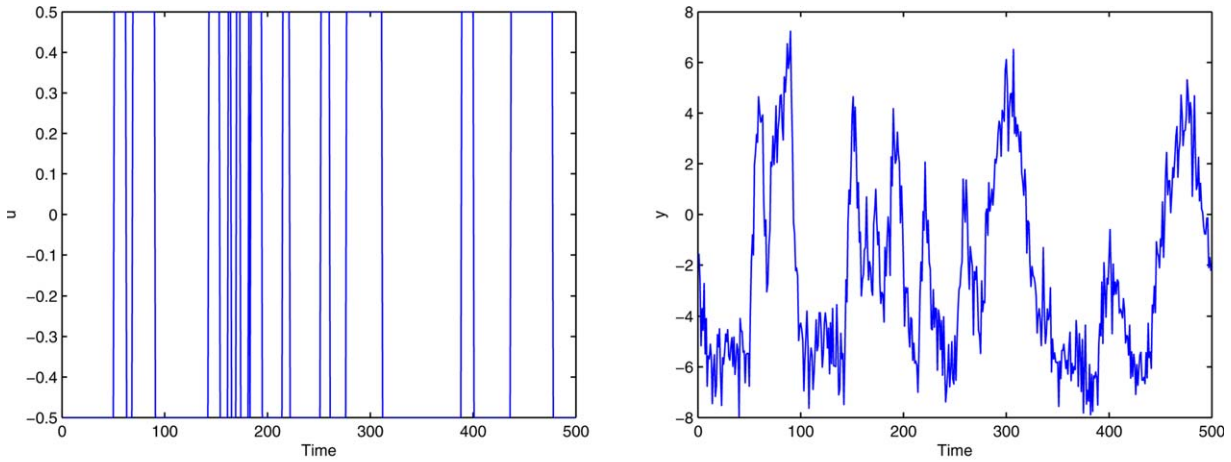
$$
p_{\Theta_\pi}(y_k|x_k) = \frac{1}{\sqrt{2\pi}\sigma} \exp \frac{-1}{2\sigma^2} (y_k - \theta_\pi^{\mathrm{T}} x_k)^{\mathrm{T}} (y_k - \theta_\pi^{\mathrm{T}} x_k) \tag{27}
$$

Similarly, for the state-space model of the scheduling variables, the state transition PDF of the hidden scheduling variables, $p_{\theta_z}(z_k|z_{k-1})$, and the PDF of the observed scheduling variables, $p_{\theta_z}(w_k|z_k)$, both are Gaussian.

Given z_k, C_{obs} and the current parameter estimate Θ', the posterior probability of the $I_k = \pi$th local model taking effect at sampling instant k can be derived using the Bayes' rule

$$
\Pr_{\Theta'}(I_k = \pi|z_k, C_{\text{obs}}) = \frac{p_{\Theta'_\pi}(y_k|x_k) \Pr_{o'_\pi}(I_k = \pi|z_k, T)}{\sum_{m_1=1}^{M_1} \cdots \sum_{m_s=1}^{M_s} p_{\Theta'_\pi}(y_k|x_k) \Pr_{o'_\pi}(I_k = \pi|z_k, T)} \tag{28}
$$

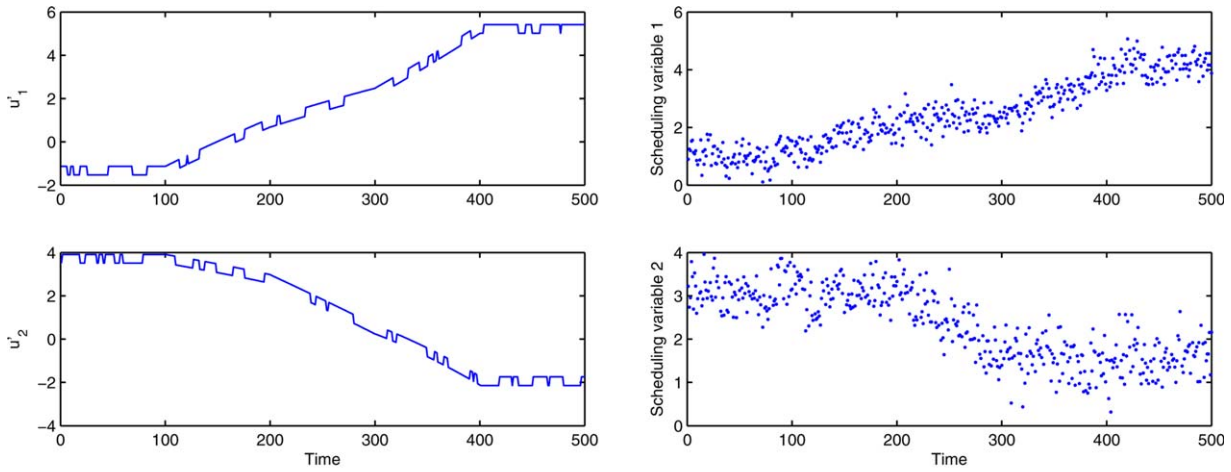
where Θ'_π and o'_π are the current parameter estimates. The detailed derivation of Eq. 28 is given in Appendix C. Given z_k, T, and the validity width of the $I_k = \pi$th local model, the conditional probability of the $I_k = \pi$th local model taking effect at sampling instant k, $\Pr_{o_\pi}(I_k = \pi|z_k, T)$, can be calculated using Eq. 10.



(a) The input data of the nonlinear process (b) The output data of the nonlinear process

Figure 2. The simulated data of the nonlinear process (numerical example).

[Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com.]



(a) The input data of the scheduling variables (b) The observed scheduling variables

Figure 3. The input and observed data of the scheduling variables with 10% missing measurements (numerical example).

[Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com.]

The density function $p_{\Theta'}(z_k|C_{\text{obs}})$ can be simplified as $p_{\theta'_z}(z_k|w_{o_1:o_b})$, where $p_{\theta'_z}(z_k|w_{o_1:o_b})$ is a smoothing density function, and θ'_z is the current parameter estimates of the state-space model of the scheduling variables. As the computation of the integrals in Eq. 26 is analytically intractable, to obtain numerical solutions, the use of Sequential Monte-Carlo (SMC) methods is necessary. The general idea behind the SMC methods is a set of random samples associated with weights to approximate the distributions of interest, and the random samples are propagated forward through sequential importance sampling and resampling steps. Following the procedure used in Gopaluni[16,17] to apply the SMC method, the smoothing density function $p_{\theta'_z}(z_k|w_{o_1:o_b})$ can be numerically calculated as

$$p_{\theta'_z}(z_k|w_{o_1:o_b}) \approx \frac{1}{L}\sum_{l=1}^{L}\delta(z_k - z_k^l) \tag{29}$$

where $\delta(\cdot)$ is a dirac-delta function, z_k^l represent the lth hidden scheduling variable particle, and L is the number of particles. The detailed derivation is shown in Appendix D.

The joint PDF of the hidden and observed scheduling variables $p_{\Theta'}(z_k, w_k|C_{\text{obs}})$ can be simplified as $p_{\theta'_z}(z_k, w_k|w_{o_1:o_b})$, which is a predictive density function when the observation w_k is missing. Use of the chain rule of probability yields

$$p_{\theta'_z}(z_k, w_k|w_{o_1:o_b}) = p_{\theta'_z}(w_k|z_k)p_{\theta'_z}(z_k|w_{o_1:o_b}) \tag{30}$$

The first term on the right-hand side, $p_{\theta'_z}(w_k|z_k)$, is the PDF of the observed scheduling variables given the states of the hidden scheduling variables. The second term, $p_{\theta'_z}(z_k|w_{o_1:o_b})$, is the smoothing PDF that can be calculated through Eq. 29. Since w_k is missing, we can obtain an estimate of w_k^l for a given z_k^l from the density function $p_{\theta'_z}(w_k|z_k)$. Therefore, through Eq. 30, we can obtain the following particle approximations after resampling according to the procedure used in Gopaluni[16,17]

$$p_{\theta'_z}(z_k, w_k|w_{o_1:o_b}) \approx \frac{1}{L}\sum_{l=1}^{L}\delta(z_k - z_k^l)\delta(w_k - w_k^l) \tag{31}$$

The joint PDF of the hidden scheduling variables $p_{\Theta'}(z_k, z_{k-1}|C_{\text{obs}})$ can be simplified as $p_{\theta'_z}(z_k, z_{k-1}|w_{o_1:o_b})$. The following particle approximations after resampling can be obtained[16]

$$p_{\theta'_z}(z_k, z_{k-1}|w_{o_1:o_b}) \approx \frac{1}{L}\sum_{l=1}^{L}\sum_{j=1}^{L}\delta(z_k - z_k^l)\delta(z_{k-1} - z_{k-1}^j) \tag{32}$$

Combining Eqs. 27–29, 31, and 32, we can determine the Q-function in Eq. 26 as

$$\begin{aligned}
Q(\Theta|\Theta') = {} & \frac{1}{L}\sum_{l=1}^{L}\sum_{k=1}^{N}\sum_{m_1=1}^{M_1}\cdots\sum_{m_s=1}^{M_s}\log p_{\Theta_\pi}(y_k|x_k)\Pr_{\Theta'}(I_k=\pi|z_k^l, C_{\text{obs}}) \\
& + \frac{1}{L}\sum_{l=1}^{L}\sum_{k=1}^{N}\sum_{m_1=1}^{M_1}\cdots\sum_{m_s=1}^{M_s}\log\Pr_{o_\pi}(I_k=\pi|z_k^l, T)\Pr_{\Theta'}(I_k=\pi|z_k^l, C_{\text{obs}}) \\
& + \frac{1}{L}\sum_{l=1}^{L}\sum_{k=o_1}^{o_b}\log p_{\theta_z}(w_k|z_k^l) \\
& + \frac{1}{L}\sum_{l=1}^{L}\sum_{k=m_1}^{m_a}\log p_{\theta_z}(w_k^l|z_k^l) \\
& + \frac{1}{L}\sum_{l=1}^{L}\sum_{j=1}^{L}\sum_{k=1}^{N}\log p_{\theta_z}(z_k^l|z_{k-1}^j) \\
& + \frac{1}{L}\sum_{l=1}^{L}\log p_{\theta_z}(z_0^l) + C_2
\end{aligned} \tag{33}$$

For the M-step shown in Eq. 16, derivatives of the Q-function are taken with respect to θ_π, o_π, σ, and θ_z. The estimated values of θ_π, o_π, and σ can be updated following the same approach as in Chen et al.[8] For simplicity, we skip the detailed derivations and only present the final results

$$\theta_\pi = \frac{\sum_{l=1}^{L}\sum_{k=1}^{N}\Pr_{\Theta'}(I_k=\pi|z_k^l, C_{\text{obs}})x_k^{\mathrm{T}}y_k}{\sum_{l=1}^{L}\sum_{k=1}^{N}\Pr_{\Theta'}(I_k=\pi|z_k^l, C_{\text{obs}})x_k^{\mathrm{T}}x_k} \tag{34}$$

$$(\sigma)^2 = \frac{\sum_{l=1}^{L}\sum_{k=1}^{N}\sum_{m_1=1}^{M_1}\cdots\sum_{m_s=1}^{M_s}\Pr_{\Theta'}(I_k=\pi|z_k^l, C_{\text{obs}})(y_k-\theta_\pi^{\mathrm{T}}x_k)^{\mathrm{T}}(y_k-\theta_\pi^{\mathrm{T}}x_k)}{\sum_{l=1}^{L}\sum_{k=1}^{N}\sum_{m_1=1}^{M_1}\cdots\sum_{m_s=1}^{M_s}\Pr_{\Theta'}(I_k=\pi|z_k^l, C_{\text{obs}})} \tag{35}$$

$$\max_{o_\pi} \sum_{l=1}^{L} \sum_{k=1}^{N} \sum_{m_1=1}^{M_1} \cdots \sum_{m_s=1}^{M_s} \log \Pr_{O_\pi}(I_k = \pi | z_k^l, T) \Pr_{\Theta'}(I_k = \pi | z_k^l, C_{\text{obs}})$$
$$\text{S.t.} \quad o_{\min} \le o_\pi \le o_{\max} \tag{36}$$

Since θ_z cannot take a close form solution, an optimization method is required. The mathematical formulation of the optimization problem in searching for the optimal value of θ_z can be expressed as

$$\max \frac{1}{L} \sum_{l=1}^{L} \sum_{k=o_1}^{o_b} \log p_{\theta_z}(w_k | z_k^l) + \frac{1}{L} \sum_{l=1}^{L} \sum_{k=m_1}^{m_a} \log p_{\theta_z}(w_k^l | z_k^l) + \frac{1}{L} \sum_{l=1}^{L} \sum_{j=1}^{L} \sum_{k=1}^{N} \log p_{\theta_z}(z_k^l | z_{k-1}^j) + \frac{1}{L} \sum_{l=1}^{L} \log p_{\theta_z}(z_0^l) \tag{37}$$

Summary of the proposed approach

The nonlinear system identification based on the multiple model approach using the EM algorithm can be summarized in the following steps:

1. *Initialization*: Set the initial value to Θ'.
2. *Particle smoother*: According to the current state-space model of the scheduling variables, apply the particle smoother to determine the distribution of the scheduling variables shown in Eq. 29.
3. *E-step*: According to Eq. 33, evaluate the approximate Q-function through particle smoothers using current parameter Θ'.
4. *M-step*: Maximize the Q-function with respect to Θ, and then set $\Theta' = \Theta$.
5. *Iterate*: Evaluate the relative change shown in Eq. 17, and according to the predetermined tolerance, decide whether to repeat steps 3 and 4 or terminate.

Simulations

In this section, a numerical example and a simulated high purity distillation column example are presented to verify the proposed method.

Numerical example with nonlinear dynamics of the scheduling variables

The simulated nonlinear process model is given by[9]

$$G(s,z) = \frac{K(z_1,z_2)}{\tau(z_1,z_2)s+1}$$
$$K(z_1,z_2) = 0.6 + z_1^2 + z_2^2$$
$$\tau(z_1,z_2) = 3 + 0.5z_1^3 + 0.5z_2 \tag{38}$$

In this simulated scenario, two scheduling variables are interacted. The operating points are $T^1 = [1,\ 2.25,\ 4]$ and $T^2 = [3,\ 1.5]$, respectively, and random excitation signals of a small magnitude are added to input, and the output is corrupted with white noise of variance of 1.

The scheduling variables are described as

$$z_k = A z_{k-1}^2 / (z_{k-1}+1) + B u'_{k-1} + \gamma_{k-1} \tag{39a}$$
$$w_k = C z_k + v_k \tag{39b}$$
$$\begin{bmatrix} \gamma_k \\ v_k \end{bmatrix} \sim N\left(0, \begin{bmatrix} Q & 0 \\ 0 & R \end{bmatrix}\right) \tag{39c}$$

where the true parameters are

$$A^* = \begin{bmatrix} 0.1 & 0.2 \\ 0.3 & 0.1 \end{bmatrix},\ B^* = \begin{bmatrix} 0.8 & 0.3 \\ 0.2 & 0.8 \end{bmatrix},\ C^* = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix},$$
$$Q^* = \begin{bmatrix} 0.01 & 0 \\ 0 & 0.01 \end{bmatrix},\ R^* = \begin{bmatrix} 0.1 & 0 \\ 0 & 0.1 \end{bmatrix} \tag{40}$$

The main objective is to identify the nonlinear process model (approximated by the combination of several local ARX models), along with the parameter set $\theta_z = [A, B, C]$ of the state-space model of the scheduling variables from an incomplete dataset, in which the observed scheduling variables have 10% missing data. The simulated data of the input and scheduling variables are shown in Figures 2 and 3. A set of validation data are generated to verify the identified models, and the operating points of the scheduling variables are $T^1 = [1.5,\ 2.5,\ 3.5]$ and $T^2 = [4,\ 1]$ for validation data.

The proposed method is used to identify the nonlinear system under investigation. To measure the quality of the identified model, the relative prediction error is used[7,8]

$$\text{Err} = \frac{\text{var}(y - \hat{y})}{\text{var}(y)} * 100\% \tag{41}$$

where var denotes the variance of the random variable, $\hat{y}$ is the estimated output, and y is the true output.

The parameter comparison results of the local model at operation points $T_1 = 1$ and $T_2 = 3$ are shown in Figure 4. The results for the training and validation datasets are shown in Figures 5 and 6, respectively. As the observed scheduling variables are noisy and have missing data, we only show the probabilities assigned to one local model. The relative errors for the training and validation datasets are 11.95 and 17.95%, and the mean square error (MSE) are 1.56 and 1.94, respectively. Therefore, it is concluded that under various operating conditions, the identified model can well capture the process dynamics.

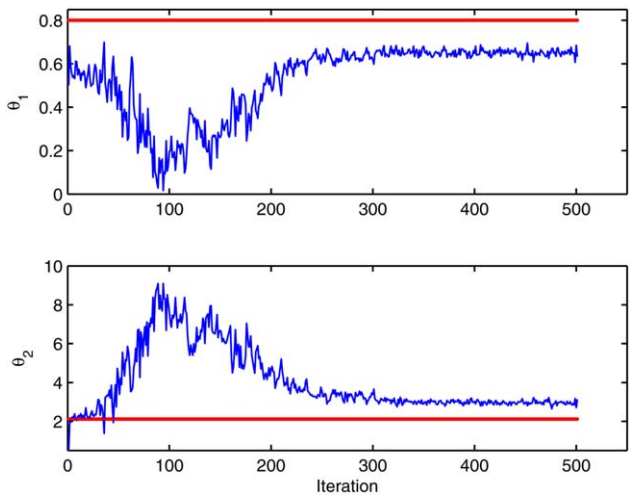
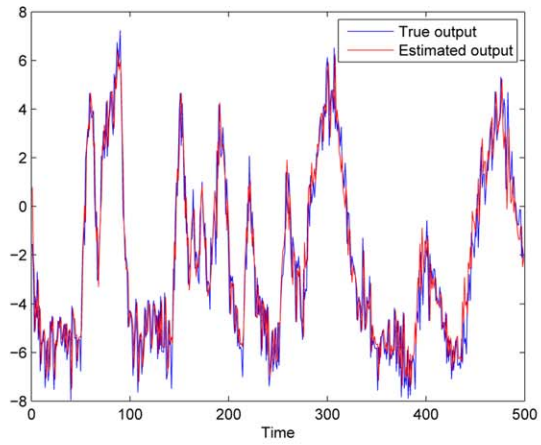
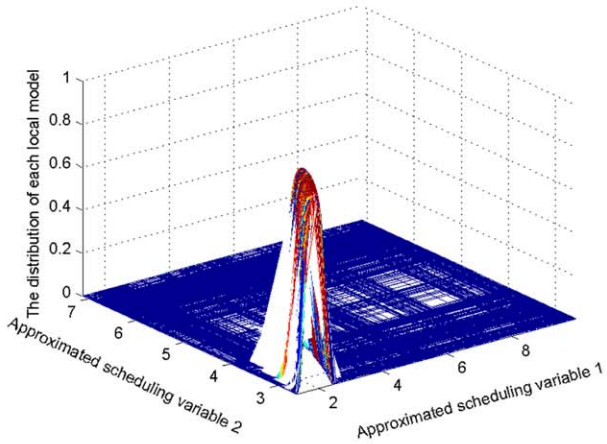


Figure 4. The results for the training dataset with 10% missing measurements (numerical example).

[Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com.]





(a) Prediction performance of the identified nonlinear model (b) The probabilities assigned to one local model

Figure 5. Parameter estimation of one local model.

[Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com.]

High purity distillation column with linear dynamics of the scheduling variables

In the process industry, distillation is one of the most common unit operations. Here, we present the LV-configuration of a distillation column yielding high purity products for a case study. The boil-up (V) and reflux (L) are the input variables manipulated to control the compositions of the bottom product (xB) and the top product (yD). Furthermore, the feed flow rate (F) is considered as the disturbance variable.[18] Figure 7 shows the schematic representation of the high purity distillation column.[7] The detailed dynamic equations of the distillation columns can be found in Skogestad.[18]

As the reflux ratio L/F is easily calculated and closely related to the purity of the top product, we select the feed flow rate F and the reflux ratio L/F as the scheduling variables, the composition of the top product yD as the output variable, and the reflux L and boil-up V as the two input variables.

The scheduling variables are described as

$$z_k = Az_{k-1} + Bu'_{k-1} + \gamma_{k-1} \tag{42a}$$

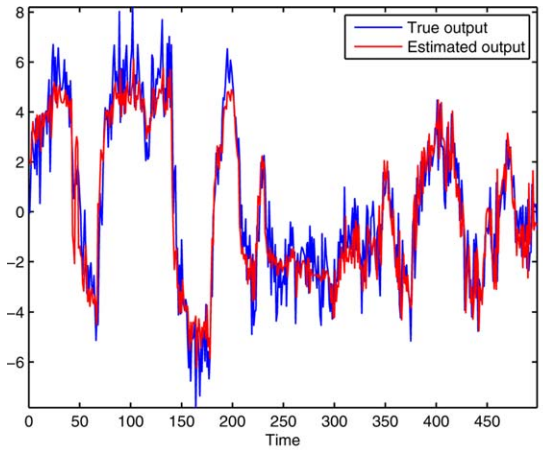
$$w_k = Cz_k + v_k \tag{42b}$$

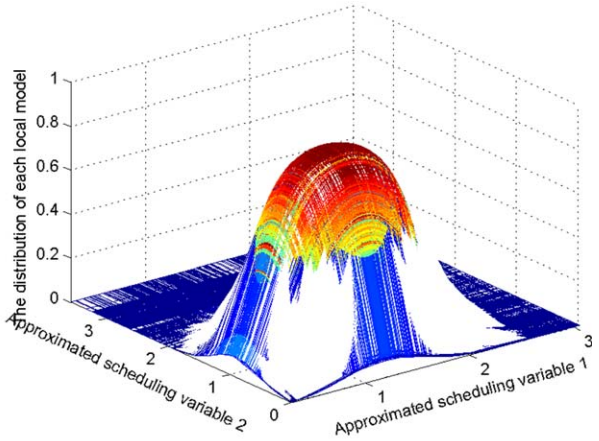
$$\begin{bmatrix} \gamma_k \\ v_k \end{bmatrix} \sim N\left(0, \begin{bmatrix} Q & 0 \\ 0 & R \end{bmatrix}\right) \tag{42c}$$

where $z = [L/F, F]^{\mathrm{T}}$, and the true parameters are

$$A^* = \begin{bmatrix} 0.02 & 0.01 \\ 0.02 & 0.03 \end{bmatrix}, \quad B^* = \begin{bmatrix} 0.08 & 0.02 \\ 0.01 & 0.07 \end{bmatrix}, \quad C^* = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix},$$
$$Q^* = \begin{bmatrix} 0.001 & 0 \\ 0 & 0.001 \end{bmatrix}, \quad R^* = \begin{bmatrix} 0.01 & 0 \\ 0 & 0.01 \end{bmatrix} \tag{43}$$

The proposed method is applied to estimate the nonlinear process model, along with parameters of the state-space model of the scheduling variables from incomplete datasets with different





(a) Prediction performance of the identified nonlinear model (b) The probabilities assigned to one local model

Figure 6. The results for the validation dataset with 10% missing measurements (numerical example).

[Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com.]

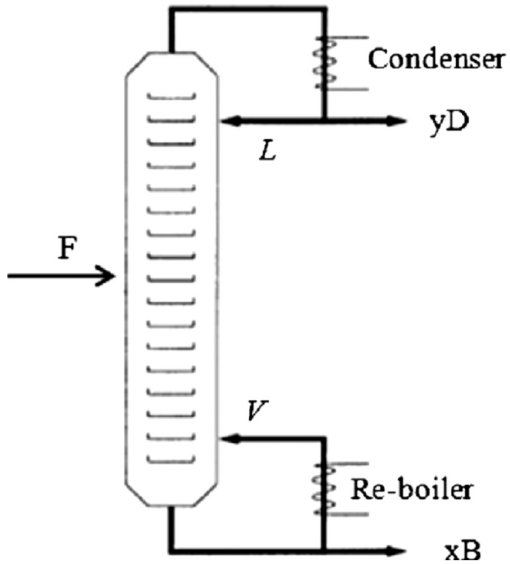


Figure 7. Schematic diagram of the high purity distillation column under LV configuration.

degrees of incompleteness. The results have been compared for cases in which 0, 10, 20, and 50% of the measurements of the scheduling variables are missing. The step responses and bode plots of the identified state-space models of the scheduling variables are compared in Figures 8 and 9. For the case with 10% missing measurements, the simulated data for the input and the scheduling variables are shown in Figures 10 and 11.

The relative errors for the training and validation datasets are 9.42 and 16.09%, and the MSE are 2.36e-3 and 2.78e-3, respectively. The graphic comparisons are shown in Figures 12 and 13.

To evaluate the true benefit of the proposed methodology, a comparison against the approach proposed by Huang et al.[7] is conducted. Because Huang et al. has not considered the missing data problem, in order to make a fair comparison, we consider to identify a high purity distillation column without missing data. In their study, the focus was on the nonlinear system identification problem with two noise-free and uncorrelated scheduling variables, while our focus is on the nonlinear system identification problem with multiple correlated and hidden scheduling variables.

In the simulation example, the feed flow rate F and the reflux ratio L/F are the scheduling variables, the composition of the top product yD is the output variable, and the reflux L and boil-up V are the two input variables, and the structure of the weighting function is shown in Eq. 10. The unknown parameters of the scheduling variables are estimated using prediction error method. The relative error and the MSE for the training dataset are 12.14% and 3.29e-3. Applying the proposed approach from our work, the results are 8.87% and 2.16e-3. These results illustrate that our method is superior to the conventional one, because it can provide more accurate prediction and has smaller MSE and the relative prediction error. When the process has missing data, our approach has even more advantages.

The results of these simulation examples show that the predictions obtained from the models identified from either complete or incomplete datasets can indeed track the true process output for both linear and nonlinear dynamics of the scheduling variables.

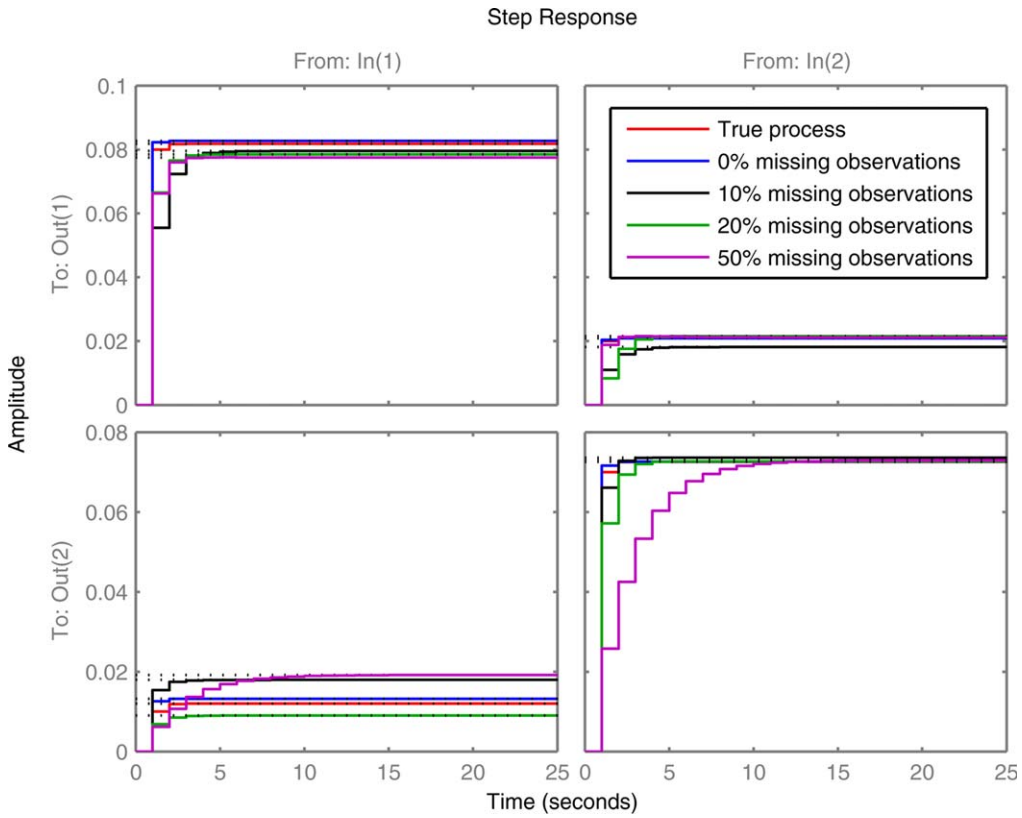


Figure 8. The step response of the state-space model of the scheduling variables (distillation column).

[Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com.]

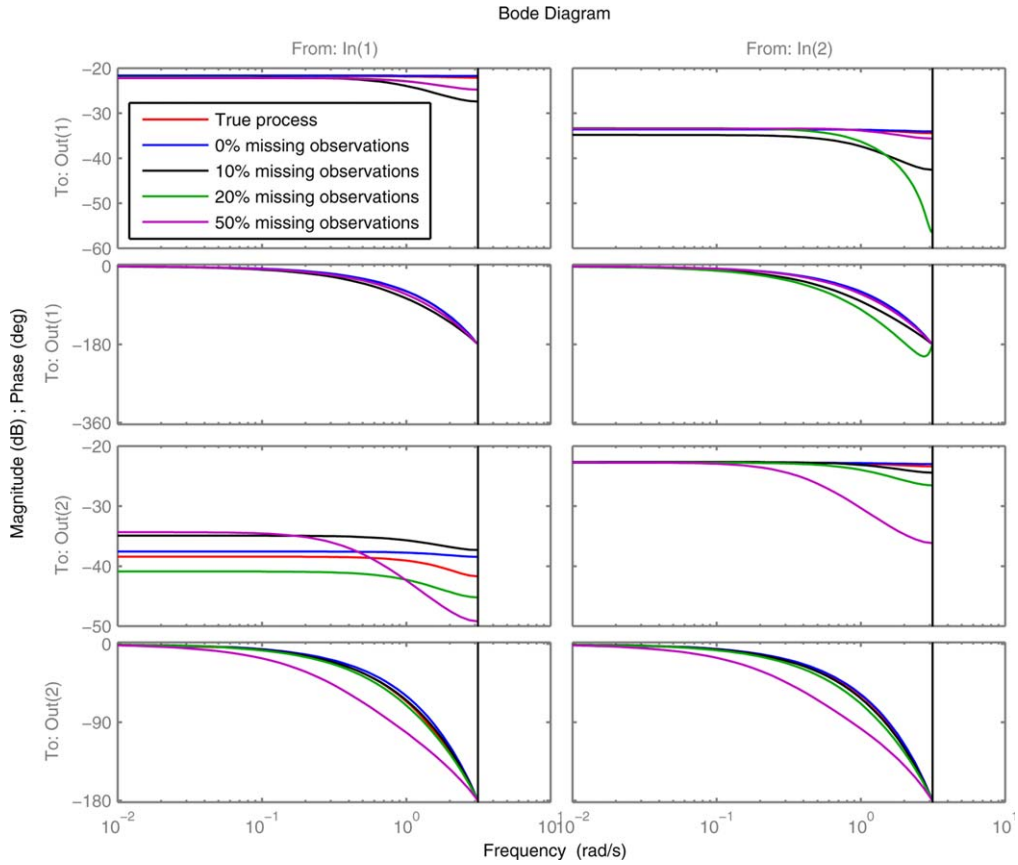
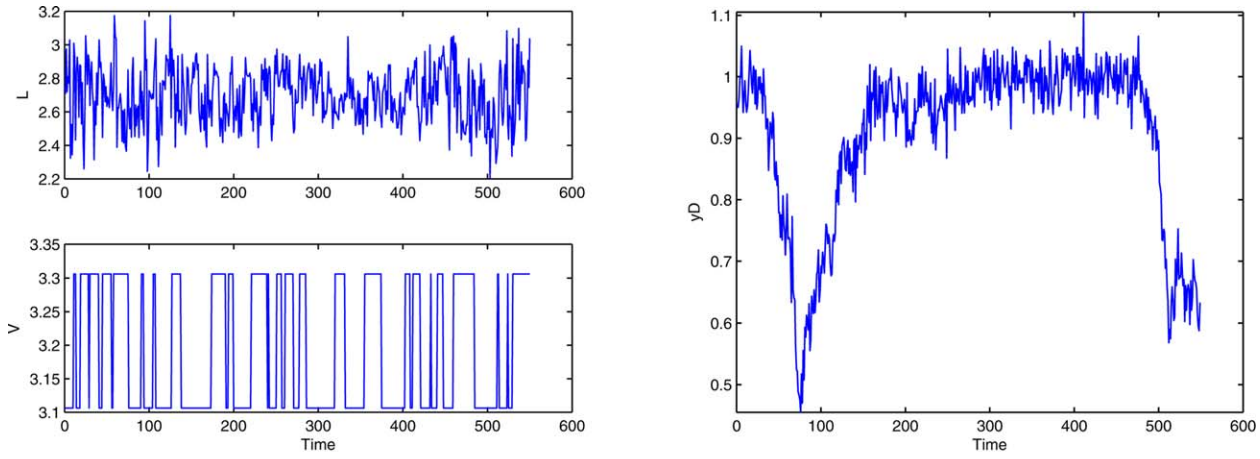


Figure 9. The bode plot of the identified state-space model of the scheduling variables (distillation column).

[Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com.]



(a) The input data of the nonlinear process (b) The output data of the nonlinear process

Figure 10. The simulated data of the nonlinear process with 10% missing measurements (distillation column).

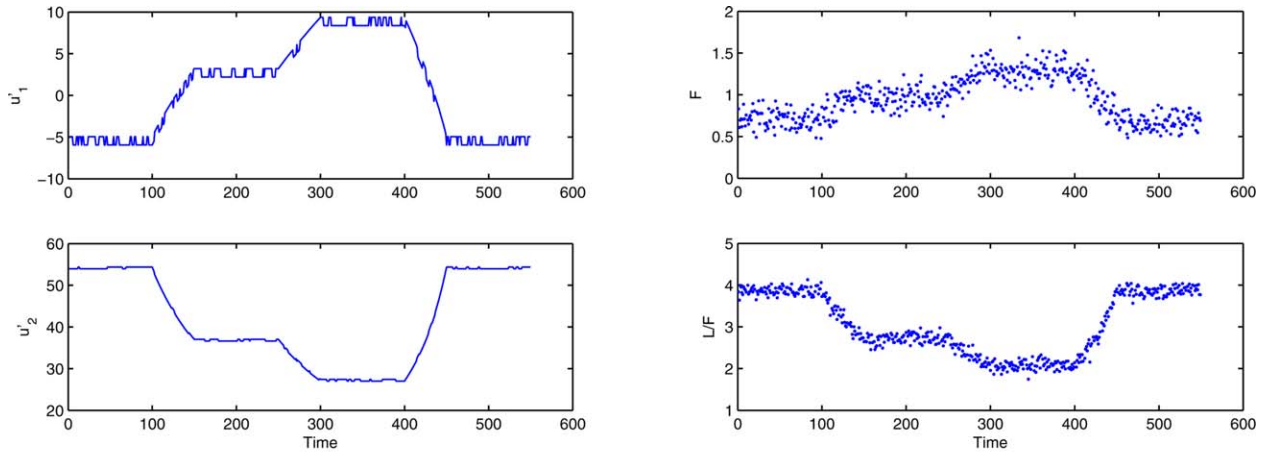
[Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com.]

Experimental Study of the Multitank System

A nonlinear system identification experiment performed on a pilot-scale multitank system is used to further illustrate the effectiveness of the proposed approach. The pilot-scale system is located in the Computer Process Control Laboratory in the Department of Chemical and Materials Engineering at the University of Alberta. A photograph of the experimental multitank system setup is shown in Figure 14 and the schematic diagram is displayed in Figure 1.

The nonlinear model of the multitank system is illustrated by Khatibisepehr and Huang[19] and Deng and Huang[20]

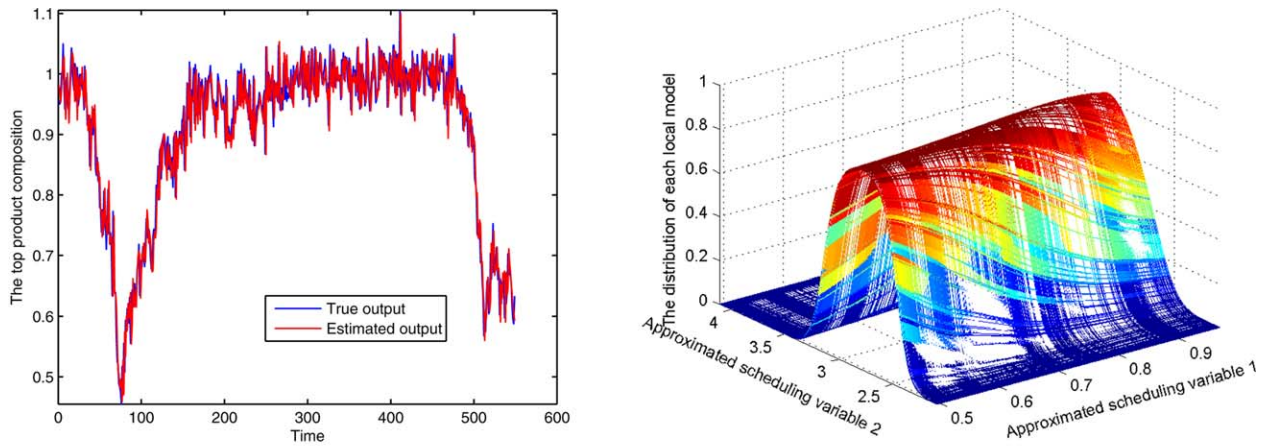
$$\frac{dH_1}{dt} = \frac{1}{\beta_1(H_1)}q - \frac{1}{\beta_1(H_1)}D_1 H_1^{\alpha_1} \tag{44a}$$



(a) The input data of the scheduling variables (b) The observed scheduling variables

Figure 11. The input and observed data of the scheduling variables with 10% missing measurements (distillation column).

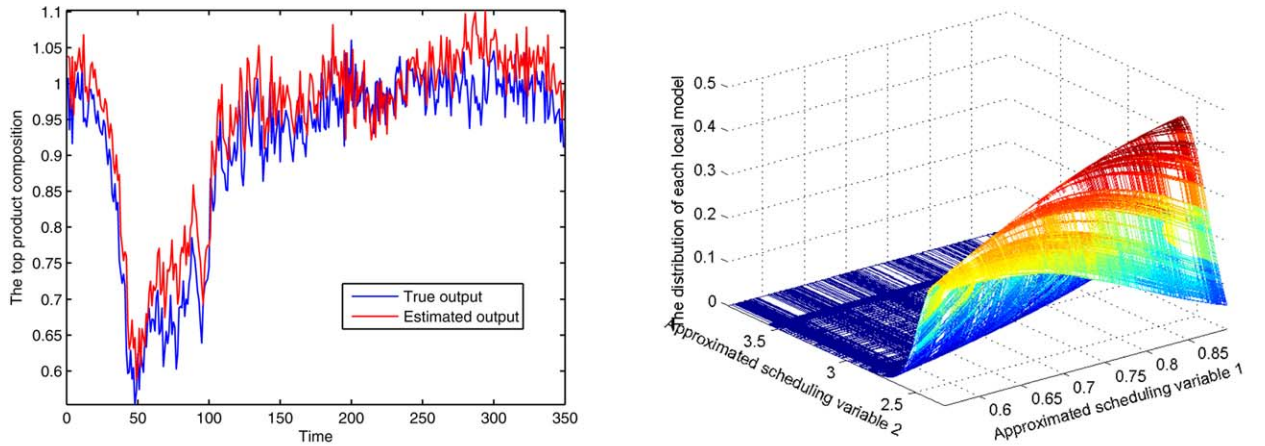
[Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com.]



(a) Prediction performance of the identified nonlinear model (b) The probabilities assigned to one local model

Figure 12. The results for the training dataset with 10% missing measurements (distillation column).

[Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com.]



(a) Prediction performance of the identified nonlinear model (b) The probabilities assigned to one local model

Figure 13. The results for the validation dataset with 10% missing measurements (distillation column).

[Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com.]

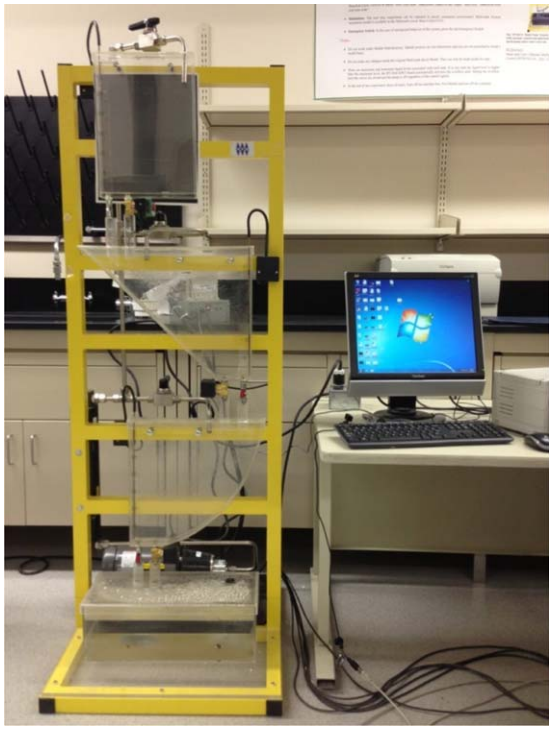


Figure 14. The experimental facility of the multitank system in the laboratory.

[Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com.]

$$\frac{dH_2}{dt} = \frac{1}{\beta_2(H_2)} D_1 H_1^{\alpha_1} - \frac{1}{\beta_2(H_2)} D_2 H_2^{\alpha_2} \tag{44b}$$

$$\frac{dH_3}{dt} = \frac{1}{\beta_3(H_3)} D_2 H_2^{\alpha_2} - \frac{1}{\beta_3(H_3)} D_3 H_3^{\alpha_3} \tag{44c}$$

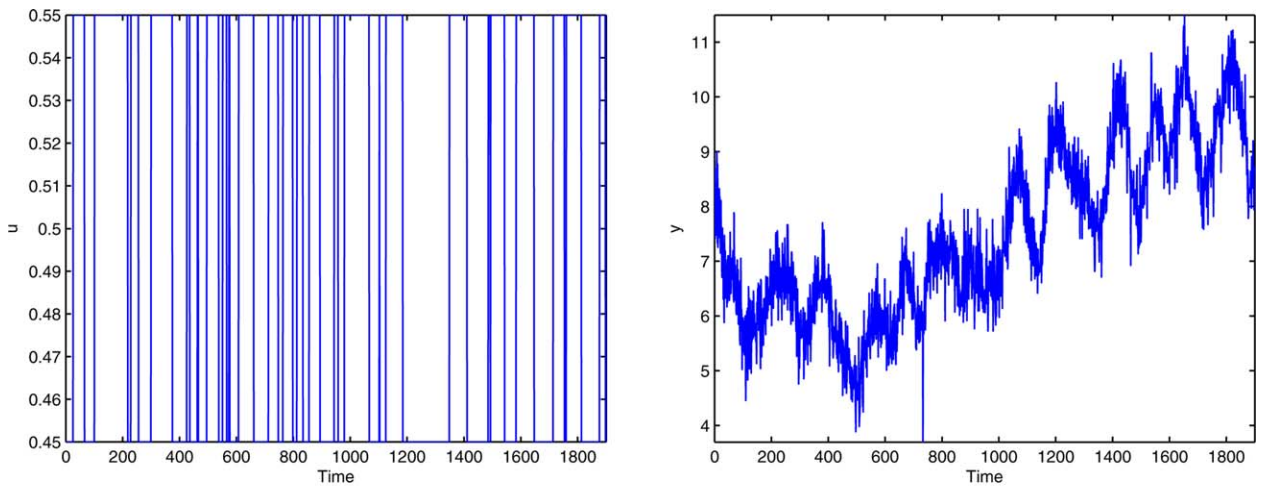
where H_i is the fluid levels in the ith tanks, D_i is resistance of the output orifice of ith tank, q is inflow to the upper tank, β_i is the cross sectional area of the ith tank, and α_i is the flow coefficient for the ith tank, where $i = 1,\ 2,\ 3$.

In this experiment, the inflow q is selected as the input, and the second tank level H_2 as the output. Two controlled valves V_1 and V_2 are the two correlated scheduling variables operating at steady states, and small excitation signals are added to the process. It is assumed that the two scheduling variables cannot be measured directly, but they can be inferred from an observation w_t governed through the state-space model given by Eq. 42, where $z = [D_1, D_2]^{\mathrm{T}}$, and the true parameters are

$$A^* = \begin{bmatrix} 0.02 & 0.04 \\ 0.03 & 0.02 \end{bmatrix},\ B^* = \begin{bmatrix} 0.08 & 0.03 \\ 0.02 & 0.08 \end{bmatrix},\ C^* = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix},$$
$$Q^* = 0,\ R^* = \begin{bmatrix} 0.005 & 0 \\ 0 & 0.005 \end{bmatrix} \tag{45}$$

In order not to over wear the equipment, we set $Q^* = 0$. White noise with variance of about 5% of the variations of the noise free output is added to the measured output H_2. The proposed method is applied to estimate the model parameters from various incomplete datasets in which 0, 10, 20, and 50% of the measurements of the scheduling variables are missing. For the case with 20% missing measurements, the observed input-output data for the multitank system is plotted in Figure 15 and the observed data of the scheduling variables are shown in Figure 16. The step responses and bode plots of the identified models of the scheduling variables are compared in Figures 17 and 18.

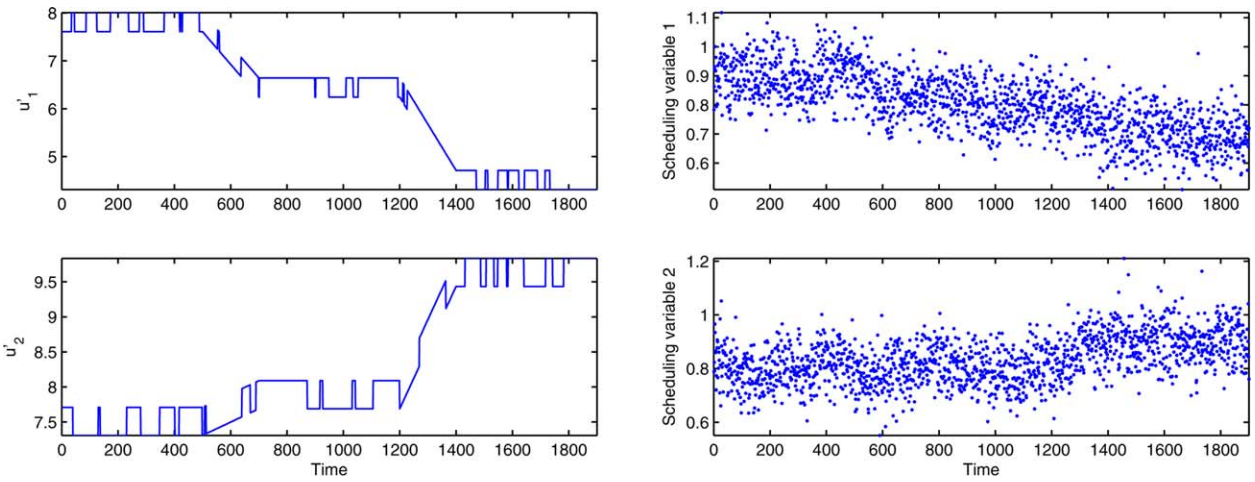
The relative errors for the training and validation datasets are 11.73 and 15.96%, and the MSE are 0.2217 and 0.2839, respectively. Furthermore, from Figures 19 and 20, it can be observed that the predictions obtained by the identified nonlinear process models are in a close agreement with the measured outputs. The results have once again confirmed the effectiveness of the proposed approach and demonstrated its practicality for practical industrial applications.



(a) The input data of the multitank system (b) The output data of the multitank system

Figure 15. The input-output data of the multitank system.

[Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com.]



(a) The input data of the scheduling variables (b) The observed scheduling variables

Figure 16. The input and observed data of the scheduling variables of the multitank system (20% missing measurements).

[Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com.]

Conclusions

A multiple model approach under the EM algorithm has been developed for identification of the nonlinear systems. Multiple correlated and noisy scheduling variables were modeled by a state-space model with unknown parameters. The nonlinear dynamics of the underlying system was represented by concatenating among multiple local ARX models identified in different process operating region. To obtain a complete representation of the nonlinear process, a normalized weighting function was employed to evaluate the probability of the local models taking effect and then combine them accordingly. The parameters of the local models, the weighting functions, and the model of the scheduling variables were estimated simultaneously. Particle smoother was applied to estimate the PDF of the hidden scheduling

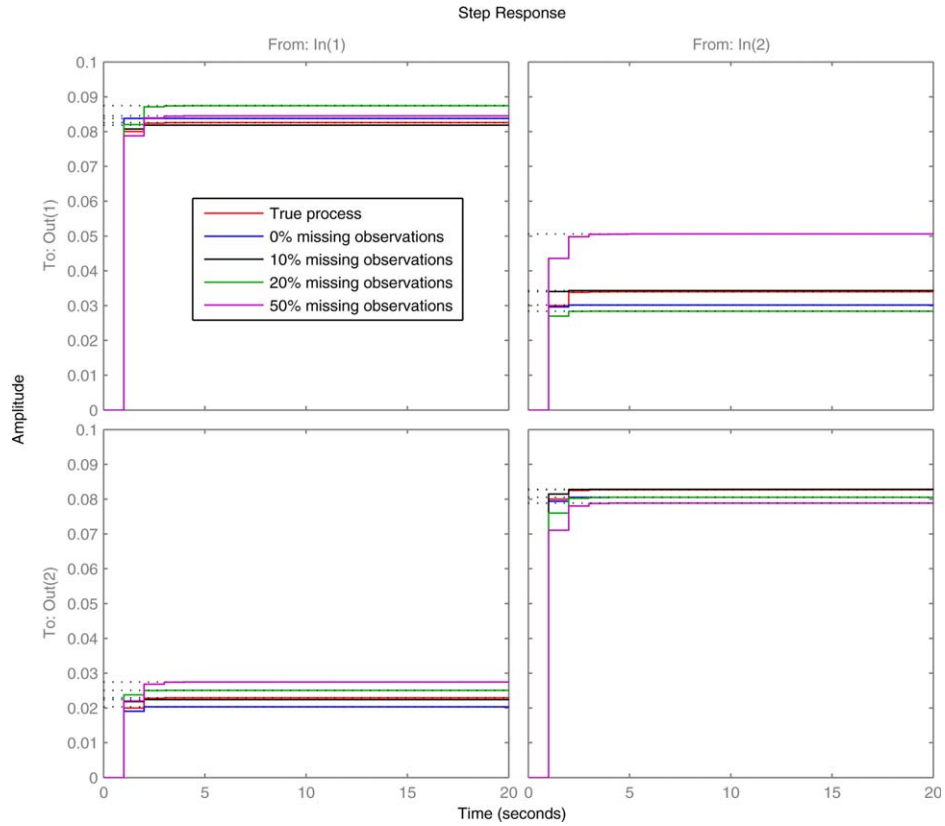


Figure 17. The step response of the state-space model of the scheduling variables.

[Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com.]

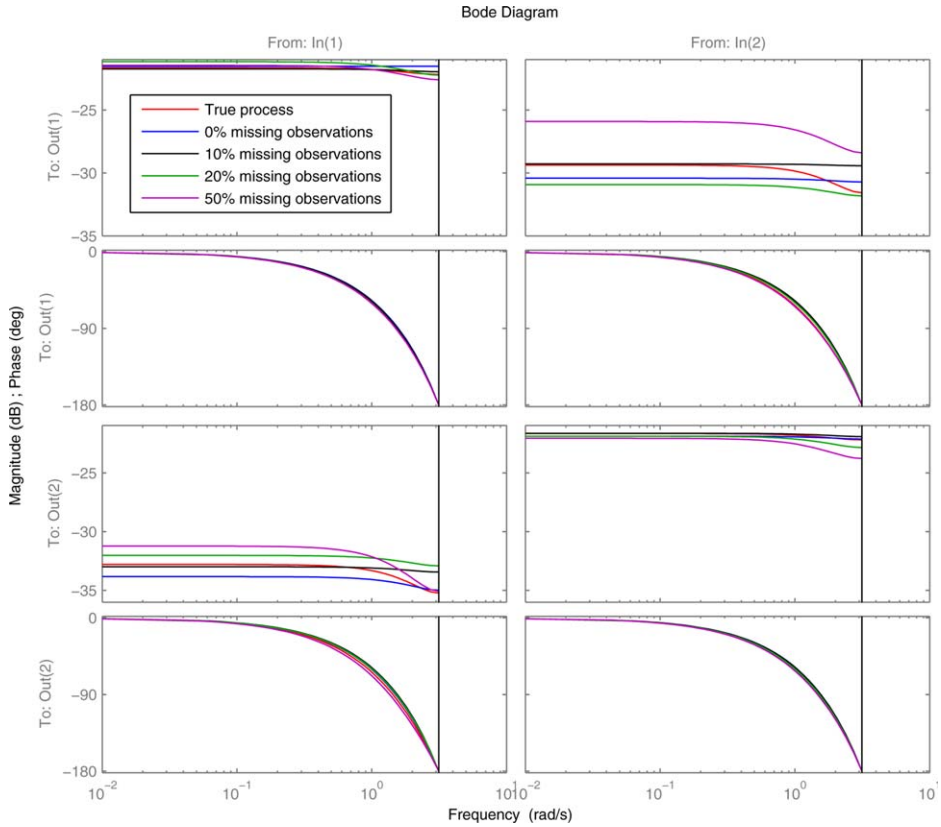


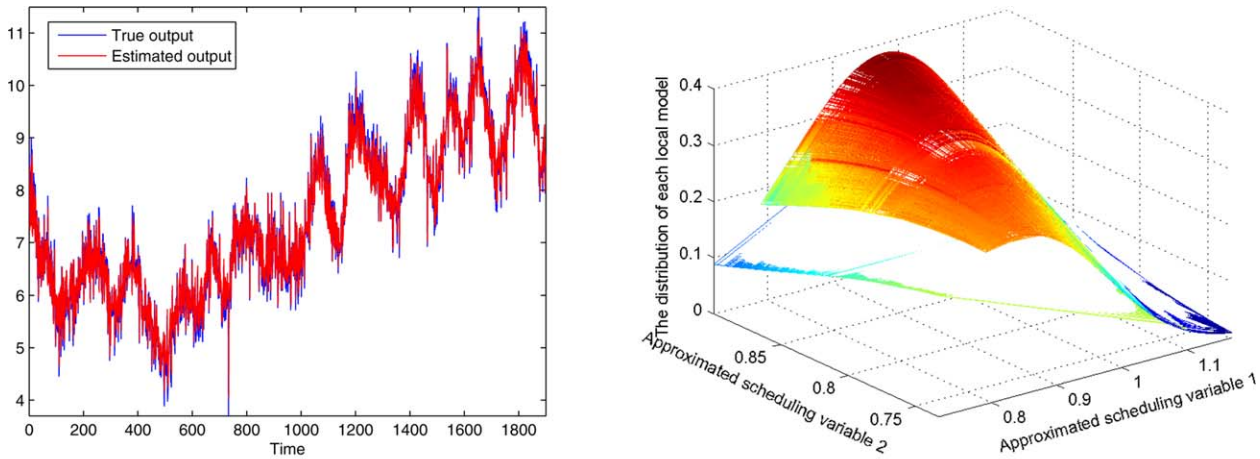
Figure 18. The bode plot of the estimated state-space model of the scheduling variables.

[Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com.]

variables with missing measurements. Simulation case studies for both linear and nonlinear dynamics of the scheduling variables were presented. The identification of a pilot-scale multitank system from experimental data was considered to further validate the effectiveness of the proposed approach for real-life applications. Satisfactory performance of the proposed approach was demonstrated on all the simulation and experimental studies.

Acknowledgments

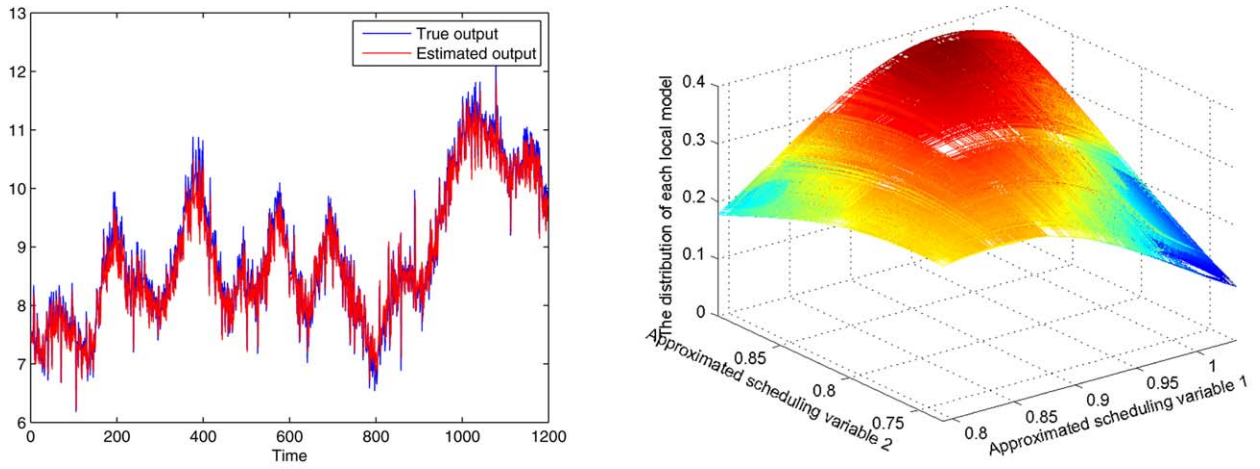
This work was supported by the Natural Sciences and Engineering Research Council (NSERC) of Canada, Alberta Innovates Technology Futures (AITF), the National Nature Science Foundation of China (Nos. 61273087, 61473077, 61473078), the Key Project of the National Nature Science Foundation of China (No. 61134009), Program for Changjiang Scholars from the Ministry of Education, Specialized



(a) Prediction performance of the identified nonlinear model (b) The probabilities assigned to one local model

Figure 19. The results for the training dataset of the multitank system (20% missing measurements).

[Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com.]



(a) Prediction performance of the identified nonlinear model (b) The probabilities assigned to one local model

Figure 20. The results for the validation dataset of the multitank system (20% missing measurements).

[Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com.]

Research Fund for Shanghai Leading Talents, Project of the Shanghai Committee of Science and Technology (No. 13JC1407500), and the Fundamental Research Funds for the Central Universities. The help from Da Zheng on the multitank system experiment is greatly appreciated.

Literature Cited

1. Jin X, Huang B, Shook DS. Multiple model LPV approach to nonlinear process identification with EM algorithm. *J Process Control*. 2011;21(1):182–193.
2. Li XR, Bar-Shalom Y. Model approach to noise identification. *IEEE Trans Aerosp Electron Syst*. 1994;30(3):671–684.
3. Banerjee A, Arkun Y, Ogunnaike B, Pearson R. Estimation of nonlinear systems using linear multiple. *AIChE J*. 1997;43(5):1204–1226.
4. Zhu Y, Xu Z. A method of LPV model identification for control. In: *Proceedings of the 17th IFAC World Congress*. Seoul, Korea, July 6–11, 2008:5018–5023.
5. Zhao Y, Huang B, Su H, Chu J. Prediction error method for identification of LPV models. *J Process Control*. 2012;22(1):180–193.
6. Gopaluni RB. Identification of nonlinear state-space models: the case of unknown model structure. *Adv Control Chem Process*. 2009;7(1):470–475.
7. Huang J, Ji G, Zhu Y, Van den Bosch P. Identification of multimodel LPV models with two scheduling variables. *J Process Control*. 2012;22(7):1198–1208.
8. Chen L, Tulsyan A, Huang B, Liu F. Multiple model approach to nonlinear system identification with uncertain scheduling variables using EM algorithm. *J Process Control*. 2013;23(10):1480–1496.
9. Chen L, Huang B, Liu F. Nonlinear system identification with multiple and correlated scheduling variables. In: *Proceedings of the 10th International Symposium on Dynamics and Control Process Systems*. Mumbai, India, December 18–20, 2013:319–324.
10. Khatibisepehr S, Huang B, Xu F, Espejo A. A Bayesian approach to design of adaptive multi-model inferential sensors with application in oil sand industry. *J Process Control*. 2013;22(10):1913–1929.
11. Ljung L. *System Identification: Theory for the User*. New Jersey: Prentice Hall, 1987.
12. Dempster AP, Laird NM, Rubin DB. Maximum likelihood from incomplete data via the EM algorithm. *J R Stat Soc Ser B*. 1977;39(1):1–38.
13. McLachlan GJ, Krishnan T. *The EM Algorithm and Extensions*, 2nd ed. Hoboken, NJ: John Willey & Sons, 2008.
14. Karlis D. A cautionary note about the EM algorithm for finite exponential mixtures. Technical Report 150, Department of Statistics, Athens University of Economics and Business, 2001.
15. Wu J. On the convergence properties of the EM algorithm. *Ann Stat*. 1983;11(1):95–103.
16. Gopaluni RB. Nonlinear system identification under missing observations: the case of unknown model structure. *J Process Control*. 2010;20(3):314–324.
17. Gopaluni RB. A particle filter approach to identification of nonlinear processes under missing observations. *Can J Chem Eng*. 2008;86(6):1081–1092.
18. Skogestad S. Dynamics and control of distillation columns - a tutorial introduction. *Chem Eng Res Des*. 1997;75:539–562.
19. Khatibisepehr S, Huang B. Dealing with irregular data in soft sensors: Bayesian method and comparative study. *Ind Eng Chem Res*. 2008;47(22):8713–8723.
20. Deng J, Huang B. Identification of nonlinear parameter varying systems with missing output data. *AIChE J*. 2012;58(11):3454–3467.
21. Klaas M, Briers M, de Freitas N, Doucet A, Maskell S, Lang D. Fast particle smoothing: if I had a million particles. In: *Proceedings of the 23rd International Conference on Machine Learning*. Pittsburgh, PA, 2006.
22. Doucet A, Godsill S, Andrieu C. On sequential Monte Carlo sampling methods for Bayesian filtering. *Stat Comput*. 2000;10(3):197–208.
23. Douc R, Cappé O, Moulines E. Comparison of resampling schemes for particle filtering. In: *Proceedings of the 4th International Symposium on Image and Signal Processing and Analysis*. Zagreb, Croatia, 2005.

Appendix A: Detailed Derivation of Eq. 19

Using Eq. 8, the first term in Eq. 18b can be simplified as

$$p_\Theta(y_{1:N}|u_{1:N},w_{1:N},T,I_{1:N},z_{0:N})$$

$$=\prod_{k=1}^{N}p_\Theta(y_k|y_{1:k-1},u_{1:N},w_{1:N},T,I_{1:N},z_{0:N}) \tag{A1a}$$

$$=\prod_{k=1}^{N}p_\Theta(y_k|x_k,I_k) \tag{A1b}$$

$$=\prod_{k=1}^{N}p_{\Theta_{I_k}}(y_k|x_k) \tag{A1c}$$

Using Eq. 13, the second term in Eq. 18b can be simplified as

$$\Pr_\Theta(I_{1:N}|u_{1:N},w_{1:N},T,z_{0:N})$$

$$= \prod_{k=1}^{N} \Pr_{\Theta}(I_k | I_{1:k-1}, u_{1:N}, w_{1:N}, T, z_{0:N}) \tag{A2a}$$

$$= \prod_{k=1}^{N} \Pr_{\Theta}(I_k | z_k, T) \tag{A2b}$$

$$= \prod_{k=1}^{N} \Pr_{O}(I_k | z_k, T) \tag{A2c}$$

Using Eq. 5, the third term in Eq. 18b can be simplified as

$$p_{\Theta}(w_{1:N} | u_{1:N}, T, z_{0:N})$$

$$= \prod_{k=1}^{N} p_{\Theta}(w_k | w_{1:k-1}, u_{1:N}, T, z_{0:N}) \tag{A3a}$$

$$= \prod_{k=1}^{N} p_{\Theta}(w_k | z_k) \tag{A3b}$$

$$= \prod_{k=1}^{N} p_{\theta_z}(w_k | z_k) \tag{A3c}$$

Using Eqs. 3 and 4, the fourth term in Eq. 18b can be simplified as

$$p_{\Theta}(z_{0:N} | u_{1:N}, T)$$

$$= \prod_{k=1}^{N} p_{\Theta}(z_k | z_{1:k-1}, u_{1:N}, T) \cdot p_{\Theta}(z_0 | u_{1:N}, T) \tag{A4a}$$

$$= \prod_{k=1}^{N} p_{\Theta}(z_k | z_{k-1}) \cdot p_{\Theta}(z_0) \tag{A4b}$$

$$= \prod_{k=1}^{N} p_{\theta_z}(z_k | z_{k-1}) \cdot p_{\theta_z}(z_0) \tag{A4c}$$

As the last term in Eq. 18b is independent of Θ, we define

$$C_1 = p_{\Theta}(u_{1:N}, T) \tag{A5}$$

By substituting Eqs. A1c, A2c, A3c, A4c, and A5 into $p_{\Theta}(C_{\text{obs}}, C_{\text{mis}})$ in Eq. 18a, we get Eq. 19.

Appendix B: Detailed Derivation of Eq. 20

As shown in Eq. 15, the E-step in the EM algorithm can be compute as

$$\begin{aligned} & Q(\Theta|\Theta') \\ & = \mathrm{E}_{p_{\Theta'}(I_{1:N}, z_{0:N}, w_{m_1:m_a} | C_{obs})} \Bigg\{ \log \prod_{k=1}^{N} p_{\Theta_{I_k}}(y_k|x_k) \\ & + \log \prod_{k=1}^{N} \Pr_O(I_k|z_k, T) + \log \prod_{k=1}^{N} p_{\theta_z}(w_k|z_k) \\ & + \log \prod_{k=1}^{N} p_{\theta_z}(z_k|z_{k-1}) + \log p_{\theta_z}(z_0) + \log C_1 \Bigg\} \end{aligned} \tag{B1a}$$

$$\begin{aligned} & = \int_{I_{1:N}, z_{0:N}, w_{m_1:m_a}} \Bigg\{ \sum_{k=1}^{N} \log p_{\Theta_{I_k}}(y_k|x_k) \\ & + \sum_{k=1}^{N} \log \Pr_O(I_k|z_k, T) + \sum_{k=1}^{N} \log p_{\theta_z}(w_k|z_k) \\ & + \sum_{k=1}^{N} \log p_{\theta_z}(z_k|z_{k-1}) + \log p_{\theta_z}(z_0) + \log C_1 \Bigg\} \end{aligned} \tag{B1b}$$

$$p_{\Theta'}(I_{1:N}, z_{0:N}, w_{m_1:m_a} | C_{\text{obs}}) dz_{0:N} dI_{1:N} dw_{m_1:m_a}$$

Decompose the above equation, we get Eq. 20.

Appendix C: Detailed Derivation of Eq. 28

Given z_k, C_{obs}, and the current parameter estimates Θ', the conditional probability of the $I_k = \pi$th local model taking effect at sampling time k can be stated as $\Pr_{\Theta'}(I_k = \pi | z_k, y_{1:N}, u_{1:N}, u'_{1:N}, w_{1:N}, T)$, and it can be simplified as $Pr_{\Theta'}(I_k = \pi | z_k, y_k, x_k, T)$. Using the Bayes' rule, it can be derived as

$$\Pr_{\Theta'}(I_k = \pi | z_k, y_k, x_k, T) = \frac{p_{\Theta'}(z_k, y_k, x_k, T | I_k = \pi) \Pr_{\Theta'}(I_k = \pi)}{\sum_{m_1=1}^{M_1} \cdots \sum_{m_s=1}^{M_s} p_{\Theta'}(z_k, y_k, x_k, T | I_k = \pi) \Pr_{\Theta'}(I_k = \pi)} \tag{C1}$$

Due to the chain rule of probability, Eq. C1 can be further expressed as

$$\frac{p_{\Theta'}(y_k | x_k, z_k, T, I_k = \pi) \Pr_{\Theta'}(I_k = \pi | x_k, z_k, T) p_{\Theta'}(x_k, z_k, T)}{\sum_{m_1=1}^{M_1} \cdots \sum_{m_s=1}^{M_s} p_{\Theta'}(y_k | x_k, z_k, T, I_k = \pi) \Pr_{\Theta'}(I_k = \pi | x_k, z_k, T) p_{\Theta'}(x_k, z_k, T)}$$

$$= \frac{p_{\Theta'}(y_k | x_k, I_k = \pi) \Pr_{\Theta'}(I_k = \pi | z_k, T)}{\sum_{m_1=1}^{M_1} \cdots \sum_{m_s=1}^{M_s} p_{\Theta'}(y_k | x_k, I_k = \pi) \Pr_{\Theta'}(I_k = \pi | z_k, T)} \tag{C2a}$$

$$= \frac{p_{\Theta'_\pi}(y_k | x_k) \Pr_{o'_\pi}(I_k = \pi | z_k, T)}{\sum_{m_1=1}^{M_1} \cdots \sum_{m_s=1}^{M_s} p_{\Theta'_\pi}(y_k | x_k) Pr_{o'_\pi}(I_k = \pi | z_k, T)} \tag{C2b}$$

Appendix D: Particle Smoother

If the observed hidden variable has a missing measurement at time k (i.e., $w_k \in w_{\text{mis}}$ with $w_{\text{mis}} = w_{m_1:m_a}$), the following filter equation is used recursively until an observation becomes available[17]

$$p_{\theta_z'}(z_k|w_{o_1:o_b}) = \int p_{\theta_z'}(z_k|z_{k-1}) \cdots p_{\theta_z'}(z_{o_b+1}|z_{o_b}) p_{\theta_z'}(z_{o_b}|w_{o_1:o_b}) dz_{o_b:k-1} \tag{D1}$$

where o_b is the last observation available up to time k. Now, assume that the following approximation of the filter at time o_b is available

$$p_{\theta_z'}(z_{o_b}|w_{o_1:o_b}) = \sum_{l=1}^{L} \omega_{o_b|o_b}^{l} \delta(z_{o_b} - z_{o_b}^{l}) \tag{D2}$$

where $\delta(\cdot)$ is a dirac-delta function, L is the number of particles, and z_k^l are chosen from an important sampling function $p_{\theta_z'}(z_k|w_{o_1:o_b})$.

Substituting Eq. D2 into Eq. D1, the filter density at time k can be approximated as

$$p_{\theta_z'}(z_{o_b}|w_{o_1:o_b}) = \sum_{l=1}^{L} \omega_{k|k}^{l} \delta(z_k - z_k^l) \tag{D3}$$

where $\omega_{k|k}^l$ is the filter weight given by

$$\omega_{k|k}^{l} = \frac{p_{\theta_z'}(w_k|z_k^l)}{\sum_{l=1}^{L} p_{\theta_z'}(w_k|z_k^l)} \tag{D4}$$

In Karlis,[21] several efficient particle smoothing methods for generalized state-space models have been proposed. In the forward-backward smoother algorithm used in this article, the smoothed density is factored as follows[22]

$$p_{\theta_z'}(z_k|w_{1:N}) = \int p_{\theta_z'}(z_k, z_{k+1}|w_{1:N}) dz_{k+1}$$

$$= \int p_{\theta_z'}(z_{k+1}|w_{1:N}) p_{\theta_z'}(z_k|z_{k+1}, w_{1:k}) dz_{k+1} \tag{D5a}$$

$$= p_{\theta_z'}(z_k|w_{1:k}) \int \frac{p_{\theta_z'}(z_{k+1}|w_{1:N}) p_{\theta_z'}(z_{k+1}|z_k)}{\int p_{\theta_z'}(z_{k+1}|z_k) p_{\theta_z'}(z_k|w_{1:k}) dz_k} dz_{k+1} \tag{D5b}$$

Unlike the forward recursion of the smoothing algorithm, the backward recursion does not depend on the observations. Therefore, missing observation weights of the filtering density are used. Then, the smoother density shown in Eq. 29 can be calculated using the particle approximation and the weight of the smoother density can be approximated recursively as[17]

$$\omega_{k|N}^{l} = \omega_{k|k}^{l} \left[\sum_{m=1}^{L} \omega_{k+1|N}^{m} \frac{p_{\theta_z'}(z_{k+1}^m|z_k^l)}{\sum_{n=1}^{L} \omega_{k|k}^{n} p_{\theta_z'}(z_{k+1}^m|z_k^n)} \right] \tag{D6}$$

To avoid the degeneracy phenomenon, the resampling step is necessary. The basic idea of resampling method is to eliminate trajectories which have small weights and only concentrate on trajectories with large weights. Several resampling approaches are proposed in Douc et al.[23] and the weight after resampling is $1/L$.

Again, the same procedure can be applied to Eqs. 31 and 32.

Manuscript received Jan. 13, 2015, and revision received Apr. 14, 2015.