Identification of Nonlinear Parameter Varying Systems with Missing Output Data

Jing Deng and Biao Huang
Dept. of Chemical and Materials Engineering, University of Alberta, Edmonton, Alberta, Canada T6G 2G6

DOI 10.1002/aic.13735
Published online March 12, 2012 in Wiley Online Library (wileyonlinelibrary.com).

An identification of nonlinear parameter varying systems using particle filter under the framework of the expectation-maximizaiton (EM) algorithm is described. In chemical industries, processes are often designed to perform tasks under various operating conditions. To circumvent the modeling difficulties rendered by multiple operating conditions and the transitions between different working points, the EM algorithm, which iteratively increases the likelihood function, is applied. Meanwhile the missing output data problem which is common in real industry is also considered in this work. Particle filters are adopted to deal with the computation of expectation functions. The efficiency of the proposed method is illustrated through simulated examples and a pilot-scale experiment. © 2012 American Institute of Chemical Engineers *AIChE J,* 58: 3454–3467, 2012
Keywords: parameter identification, EM algorithm, missing data, particle filter, multiple model

Introduction

Over the past few decades, the research of parameter estimation for nonlinear processes has witnessed rapid progress as it plays a key role in the development of mathematical models to describe process behavior. Aspects of parameter estimation and system identification have been discussed extensively in literature.[1,2] Linear modeling techniques have been quite mature over the past few decades. Nonlinear process modeling such as nonlinear autoregressive exogenous, artificial neural network and Wiener or Hammerstein models have also been widely applied. However, studies on parameter estimation for nonlinear parameter varying (non-LPV) system have been sparse. The inborn nonlinearity of the chemical processes and the production complexity brought by various working conditions have both increased the estimation difficulties. In general, chemical processes may behave differently when performing different production tasks. This includes the feed raw material property changes, a varying grade in polymer plants, or reaction load changes, etc.

To overcome the limitations of conventional single-model based modeling techniques, researchers have developed various multiple modeling strategies. Shamma et al.[3] first introduced a LPV modeling method, which is featured by its linear structure and varying model parameters. Due to its capability in approximating nonlinear process, the LPV modeling method has drawn growing attention from researchers. An LPV modeling method was put forwarded by Xu et al.[4] in their study of nonlinear model predictive control (MPC). The process is tested around each operating trajectory and the global LPV model is identified by interpolating each linear local model using all available data. Jin and Huang[5] proposed an LPV modeling method under the framework of the expectation and maximization (EM) algorithm, which identified the LPV models using all data points collected from the experiments. Their work only considered the linear input–output ARX model as the local models to approximate a global nonlinear model.

Nonlinear state space model is a general class of models to represent nonlinear dynamic systems. Maximum likelihood estimation of nonlinear parameter invariant state space models has been studied by Schon et al.[6] However, the process often operates over various conditions which render different model parameters. The work conducted in this article aims at the identification of parameter varying nonlinear state space models.

On the other hand, missing data or irregularly sampled data is commonly observed in industrial practice. Parameter estimation of nonlinear dynamic models in the presence of missing observation has not been well studied. Missing data could be caused by a sudden mechanical breakdown, hardware sensor failure or data acquisition system malfunction, etc. Another increasing common source for this missing data problem is the integration of communication networks in process control systems and the subsequent potential for data losses and packet dropouts.

Some common approaches in dealing with missing data have been presented and summarized in Khatibisepehr's work.[7] One intuitive way known as case-wise deletion is to simply exclude the records that contain missing values. Its major drawback is that some informative data may also be thrown out in the meantime. In many chemical industries, for example, process variables such as flow rate, temperature, stream density are frequently sampled while the key quality variable like composition which often is of the most interest can only be obtained after hours of laboratory analysis. Arbitrarily removing the records where lab data is not

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

available leads to a loss of useful information contained in fast sampled variables. Another popular treatment of incomplete data set is called imputation, including mean substitution, regression imputation, multiple imputation, etc. These methods, by their names, replace all missing value with the mean of that variable or the prediction using information gained from other data, which appears to be attractive in the sense that it preserves the complete data size. Nevertheless, as pointed out by Khatibisepehr,[7] the variances of the data may be considerably changed with imputation.

The work by Gopaluni[8] is an important step towards identification of parameter invariant nonlinear models with missing observations, where the EM algorithm is adopted for dealing with missing data and hidden state, and the particle filter based smoother is applied for computation of the expectation functions. This article extends the work of Gopaluni[8] by considering parameter varying nonlinear systems. In addition, missing output and parameter varying problems in nonlinear state space model parameter estimation are solved simultaneously under the framework EM algorithm. Particle filters are used for computation of expectation functions. The use of particle filter rather than the smoother significantly reduces the computation load.

The remainder of this article is organized as follows: The ''Problem Statement'' section states the identification problem of parameter varying nonlinear state space models with missing output data. The ''Expectation-Maximization Algorithm'' section begins with a revisit of the EM algorithm and the derivation of the expression for Q function for non-LPV state space models with missing output data is given. The ''Computation Through Particle Filtering'' section provides a brief description of particle filters and the detail of evaluating the Q function using particle filters is presented. Numerical simulations as well as an experimental example are illustrated in the ''Simulations and Pilot-Scale Experiment'' section which aim at demonstrating the effectiveness of the proposed method in non-LPV system identification with missing output data. The ''Discussion and Conclusion'' section draws the conclusion based on the results obtained in this article.

Problem Statement

Many industrial processes are often operated in certain ''orderly'' ways to meet different production objectives. Such orderly ways are also referred as operating trajectory which consists of several predesigned operating points. In this article, we use ''H'' to denote the operating variable according to which the process is operated.

Consider the nonlinear state space model given by

$$x_t = f(x_{t-1}, u_{t-1}, \Theta) + \omega_t \tag{1}$$

$$y_t = h(x_t, \Theta) + v_t \tag{2}$$

where the system parameters are Θ which are functions of the scheduling variable H such that $\Theta = g(H)$. Assume that J operating points are predefined such that at each H_i, $i = 1,2,\ldots,J$, the process has different parameters in its (nonlinear) model, and each local set of parameters θ_i, $i = 1,2,\ldots,J$ are to be estimated. x_t, u_t, y_t, ω_t, and v_t are state, measured input, measured output, process noise and measurement noise, respectively; ω_t and v_t are independent and identically distributed Gaussian noises with covariance matrices Q and R, respectively. The input sequence $\{u_1,\ldots,u_T\}$ and the trajectory of scheduling variable $\{H_1,\ldots,H_T\}$ are known.

Let X denote the sequence of hidden states $\{x_1,\ldots x_T\}$. The outputs are available at time $\{t_1,\ldots,t_\alpha\}$ while missing at time $\{m_1,\ldots,m_\beta\}$. $Y_o = \{y_{t_1}, \ldots, y_{t_\alpha}\}$ and $Y_m = \{y_{m_1}, \ldots, y_{m_\beta}\}$ stand for the corresponding observed outputs set and missing outputs set. It is assumed that the data is missing completely at random (MCAR).[7] In other words, the probability that data missing mechanism does not depend on any part of the observed data or missing data. The nonlinear model structures in 1 and 2 are known a priori.

Due to the varying operating condition, a single nonlinear model is not sufficient to represent the process dynamics. Therefore, a global nonlinear model which is a weighted interpolation of each local nonlinear model is adopted as follows

$$x_t = \sum_{j=1}^{J} \alpha_{tj} x_{tj} \tag{3}$$

$$y_t = \sum_{j=1}^{J} \alpha_{tj} y_{tj} \tag{4}$$

An exponential weighting function is used here to denote the weight for each local model[5]

$$\omega_{tj} = \exp\left(\frac{-(H_t - H_j)^2}{2(\sigma_j)^2}\right) \tag{5}$$

and the normalized weight α can be derived as

$$\alpha_{tj} = \frac{\omega_{tj}}{\sum_1^J \omega_{tj}} \tag{6}$$

where σ_j represents the validity width for each local model which is bounded by $\sigma_{\min}$ (the lower bound for $\sigma_j, j = 1{:}\, J$) and $\sigma_{\max}$ (the upper bound for $\sigma_j, j = 1{:}\, J$). Hence, the parameters of each local state space model, θ_j, $j = 1{:}\, J$, as well as the model validity $\sigma_j, j = 1{:}\, J$, are of interest. In the following, we will show how to formulate the parameter estimation problem under the scheme of the EM algorithm.

Expectation-Maximization Algorithm

EM algorithm revisit

Expectation-maximization (EM) algorithm[9] is a well-known maximum likelihood based method, which iterates between two steps, the expectation step and maximization step. The basic principle behind the EM algorithm is that instead of performing a direct optimization of the likelihood of the observed data, which is typically not tractable, one augments the observed data set C_{obs} with missing data set C_{mis} to perform a series of iterative optimizations. In the EM procedure both the complete data log-likelihood, $\log[(C_{\text{obs}}, C_{\text{mis}}|\Theta)]$ and the conditional predictive distribution, $p(C_{\text{mis}}|C_{\text{obs}},\Theta)$, are calculated. Consisting of two steps, namely the expectation step (E-step) and the maximization step (M-step), the EM algorithm proceeds as follows:

Let Θ^k be the current best approximation to the mode of the observed posterior or the best estimated parameters using all available data. With the parameters currently available and data that are observed, the distribution function of the missing observations may be determined. Based on the

distribution function the expectation of the complete data with the expectation taking over the missing observation can be derived, which is known as the Q function. The E-step is to compute the Q function which is defined by

$$Q(\Theta|\Theta^k) = E_{C_{\text{mis}}|C_{\text{obs}},\Theta^k}\{\log[p(C_{\text{obs}}, C_{\text{mis}}|\Theta)]\} = \int_{C_{\text{mis}}} \log[p(C_{\text{obs}}, C_{\text{mis}})|\Theta]p(C_{\text{mis}}|C_{\text{obs}}, \Theta^k)dC_{\text{mis}} \tag{7}$$

and the M-step is to maximize the Q function with respect to Θ to obtain

$$\Theta^{k+1} = \arg\max_{\Theta} Q(\Theta|\Theta^k) \tag{8}$$

The E-step and M-step iterate until convergence.

Formulation of the multiple model parameter estimation based on the EM algorithm

Consider the state-space model described in Eqs. 1 and 2. A hidden variable I_t is introduced to represent the identity of the sub model which takes effect at time t. The observed data set C_{obs} are $Y_o = \{y_{t_1}, \ldots, y_{t_\alpha}\}$, $\{u_1,\ldots,u_T\}$, and $\{H_1,\ldots,H_T\}$, while the hidden states $X = \{x_1,\ldots,x_T\}$, the hidden model identity $I = \{I_1,\ldots,I_T\}$ and the missing outputs $Y_m = \{y_{m_1}, \ldots, y_{m_\beta}\}$ can be viewed as the latent or missing data C_{mis}. Since input sequence $\{u_1,\ldots,u_T\}$ are considered known, it will not play a role in the following derivation and will be omitted for simplicity. Let $p(C_{\text{obs}},C_{\text{mis}}|\Theta)$ denote the complete likelihood function including both the hidden states and observations. The Q function is defined as the expectation of the log-likelihood function $\log[p(x_{1:T},I_{1:T},H_{1:T},y_{1:T}|\Theta)]$ with respect to all latent variables or data which is given by

$$\begin{aligned} Q(\Theta|\Theta^k) &= E_{C_{\text{mis}}|C_{\text{obs}},\Theta^k}\{\log[p(C_{\text{obs}}, C_{\text{mis}}|\Theta)]\} \\ &= E_{x_{1:T},I_{1:T},Y_m|C_{\text{obs}},\Theta^k}\{\log[p(Y_o, Y_m, H_{1:T}, x_{1:T}, I_{1:T}|\Theta)]\} \\ &= E_{x_{1:T},I_{1:T},Y_m|C_{\text{obs}},\Theta^k}\{\log[p(y_{1:T}, H_{1:T}, x_{1:T}, I_{1:T}|\Theta)] \end{aligned} \tag{9}$$

In Eq. 9, the term $p(y_{1:T},H_{1:T},x_{1:T},I_{1:T}|\Theta)$ which is the joint density function of the full data set can be decomposed using the Bayesian property as

$$\begin{aligned} &p(y_{1:T}, H_{1:T}, x_{1:T}, I_{1:T}|\Theta) \\ &= p(y_{1:T}|H_{1:T}, x_{1:T}, I_{1:T}, \Theta)p(H_{1:T}, x_{1:T}, I_{1:T}|\Theta) \\ &= p(y_{1:T}|H_{1:T}, x_{1:T}, I_{1:T}, \Theta)p(x_{1:T}|H_{1:T}, I_{1:T}, \Theta)p(H_{1:T}, I_{1:T}|\Theta) \\ &= p(y_{1:T}|H_{1:T}, x_{1:T}, I_{1:T}, \Theta)p(x_{1:T}|H_{1:T}, I_{1:T}, \Theta). \\ &\quad p(I_{1:T}|H_{1:T}, \Theta) \cdot p(H_{1:T}|\Theta) \end{aligned} \tag{10}$$

The first term can be further written as

$$\begin{aligned} &p(y_{1:T}|H_{1:T}, x_{1:T}, I_{1:T}, \Theta) \\ &= p(y_T|y_{1:T-1}, H_{1:T}, x_{1:T}, I_{1:T}, \Theta)p(y_{1:T-1}|H_{1:T}, x_{1:T}, I_{1:T}, \Theta) \\ &= p(y_T|y_{1:T-1}, H_{1:T}, x_{1:T}, I_{1:T}, \Theta)p(y_{T-1}|y_{1:T-2}, H_{1:T}, x_{1:T}, I_{1:T}, \Theta)\ldots \\ &p(y_2|y_1, H_{1:T}, x_{1:T}, I_{1:T}, \Theta)p(y_1|H_{1:T}, x_{1:T}, I_{1:T}, \Theta) \\ &= p(y_T|x_T, I_T, \Theta)p(y_{T-1}|x_{T-1}, I_{T-1}, \Theta)\dot{s}p(y_1|x_1, I_1, \Theta) \\ &= \prod_{t=1}^{T} p(y_t|x_t, \Theta_{I_t}) \end{aligned} \tag{11}$$

where in the derivation of Eq. 11, we have used Markov property and the relation that given the model identity I, the conditional distribution of y is independent of the scheduling variable H. Similarly the second term can be simplified to

$$\begin{aligned} &p(x_{1:T}|H_{1:T}, I_{1:T}, \Theta) \\ &= p(x_T|x_{1:T-1}, H_{1:T}, I_{1:T}, \Theta)p(x_{1:T-1}|H_{1:T}, I_{1:T}, \Theta) \\ &= p(x_T|x_{1:T-1}, H_{1:T}, I_{1:T}, \Theta)p(x_{T-1}|x_{1:T-2}, H_{1:T}, I_{1:T}, \Theta)\ldots \\ &\quad p(x_2|x_1, H_{1:T}, x_{1:T}, I_{1:T}, \Theta)p(x_1|H_{1:T}, x_{1:T}, I_{1:T}, \Theta) \\ &= p(x_T|x_{T-1}, I_T, \Theta)p(x_{T-1}|x_{T-2}, I_{T-1}, \Theta)\ldots \\ &\quad p(x_2|x_1, I_1, \Theta) \cdot p(x_1|I_1, \Theta) \\ &= p(x_1|\Theta_{I_1}) \prod_{t=2}^{T} p(x_t|x_{t-1}, \Theta_{I_t}) \end{aligned} \tag{12}$$

where in the derivation of Eq. 12, the Markov property about the state has also been applied. The third term can be derived below

$$\begin{aligned} &p(I_{1:T}|H_{1:T}, \Theta) \\ &= p(I_T|I_{1:T-1}, H_{1:T}, \Theta)p(I_{1:T-1}|H_{1:T}, \Theta) \\ &= p(I_T|I_{1:T-1}, H_{1:T}\Theta)p(I_{T-1}|I_{1:T-2}, H_{1:T}, \Theta)\cdots p(I_1|H_{1:T}, \Theta) \\ &= p(I_T|H_T, \Theta)p(I_{T-1}|H_{T-1}, \Theta)\ \ldots\ p(I_1|H_1, \Theta) \\ &= \prod_{t=1}^{T} p(I_t|H_t, \Theta_{I_t}) \end{aligned} \tag{13}$$

Derivation of the third term has used the fact that the distribution of the model identity is completely determined by the scheduling variable H_t. Substituting Eqs. 11, 12, and 13 in 10, the joint density of the likelihood of the full data set can be rewritten as

$$\begin{aligned} p(y_{1:T}, H_{1:T}, x_{1:T}, I_{1:T}|\Theta) = &\prod_{t=1}^{T} (p(y_t|x_t, \Theta_{I_t})p(I_t|H_t, \Theta_{I_t}))\cdot \\ &p(x_1|\Theta_{I_1}) \prod_{t=2}^{T} p(x_t|x_{t-1}, \Theta_{I_t}) \cdot C \end{aligned} \tag{14}$$

where $C = p(H_{1:T}|\Theta)$ is considered as a constant since the trajectory of the scheduling variable $\{H_1\ldots H_T\}$ is known and does not depend on Θ. Furthermore, substituting Eq. 14 in 9, the Q function can be rearranged as

$$\begin{aligned} Q(\Theta|\Theta^k) &= E_{x_{1:T},I_{1:T},Y_m|C_{\text{obs}},\Theta^k}\left\{\log\left[\prod_{t=1}^{T}(p(y_t|x_t, \Theta_{I_t})\right.\right. \\ &\left.\left. p(I_t|H_t, \Theta_{I_t})) \cdot p(x_1|\Theta_{I_1}) \prod_{t=2}^{T} p(x_t|x_{t-1}, \Theta_{I_t}) \cdot C\right]\right\} \\ &= E_{x_{1:T},I_{1:T},Y_m|C_{\text{obs}},\Theta^k}\left\{\sum_{t=1}^{T}[\log(p(y_t|x_t, \Theta_{I_t}) + \log p(I_t|H_t, \Theta_{I_t})]\right. \\ &\left. + \log p(x_1|\Theta_{I_1}) + \sum_{t=2}^{T} \log p(x_t|x_{t-1}, \Theta_{I_t}) + \log C\right\} \end{aligned} \tag{15}$$

Since the possible set of the operating points are known a priori, the expectation can be taken over the discrete variable I_t first as

$$Q(\Theta|\Theta^k) = E_{x_{1:T},Y_m|C_{\text{obs}},\Theta^k}\left\{\sum_{t=1}^{T}\sum_{j=1}^{J}\log p(y_t|x_t,\Theta_j)\cdot p(I_t=j|C_{\text{obs}},\Theta^k) + \sum_{t=1}^{T}\sum_{j=1}^{J}\log p(I_t=j|\sigma_j,H_t)\cdot p(I_t=j|C_{\text{obs}},\Theta^k)\right.$$
$$\left. + \sum_{j=1}^{J}\log p(x_1|\Theta_1)\cdot p(I_t=j|C_{\text{obs}},\Theta^k) + \sum_{t=2}^{T}\sum_{j=1}^{J}\log p(x_t|x_{t-1},\Theta_j)\cdot p(I_t=j|C_{\text{obs}},\Theta^k) + \sum_{j=1}^{J}\log C\cdot p(I_t=j|C_{\text{obs}},\Theta^k)\right\} \quad (16)$$

Then the expectation is further taken over continuous variables states $X(x_{1:T})$ and missing observations Y_m

$$\begin{aligned}
&= \int_{X,Y_m}\sum_{t=1}^{T}\sum_{j=1}^{J}\log p(y_t|x_t,\Theta_j)\cdot p(I_t=j|C_{\text{obs}},\Theta^k)\cdot p(x_{1:T},y_{m_1:m_\beta}|C_{\text{obs}},\Theta^k)dx_{1:T}dy_{m_1:m_\beta}\\
&+ \int_{X,Y_m}\sum_{t=1}^{T}\sum_{j=1}^{J}\log p(I_t=j|\sigma_j,H_t)\cdot p(I_t=j|C_{\text{obs}},\Theta^k)\cdot p(x_{1:T},y_{m_1:m_\beta}|C_{\text{obs}},\Theta^k)dx_{1:T}dy_{m_1:m_\beta}\\
&+ \int_{X,Y_m}\sum_{j=1}^{J}\log p(x_1|\Theta_1)\cdot p(I_t=j|C_{\text{obs}},\Theta^k)\cdot p(x_{1:T},y_{m_1:m_\beta}|C_{\text{obs}},\Theta^k)dx_{1:T}dy_{m_1:m_\beta}\\
&+ \int_{X,Y_m}\sum_{t=2}^{T}\sum_{j=1}^{J}\log p(x_t|x_{t-1},\Theta_j)\cdot p(I_t=j|C_{\text{obs}},\Theta^k)\cdot p(x_{1:T},y_{m_1:m_\beta}|C_{\text{obs}},\Theta^k)dx_{1:T}dy_{m_1:m_\beta}\\
&+ \int_{X,Y_m}\sum_{j=1}^{J}\log C\cdot p(I_t=j|C_{\text{obs}},\Theta^k)\}\cdot p(x_{1:T},y_{m_1:m_\beta}|C_{\text{obs}},\Theta^k)dx_{1:T}dy_{m_1:m_\beta}
\end{aligned} \quad (17)$$

Given the previous division of the output into observed and missing subset Y_o and Y_m, the derivation can be continued as

$$\begin{aligned}
&= \sum_{t=t_1}^{t=t_z}\sum_{j=1}^{J}\int_X \log p(y_t|x_t,\Theta_j)\cdot p(I_t=j|C_{\text{obs}},\Theta^k)\cdot\left[\int_{Y_m}p(x_{1:T},y_{m_1:m_\beta}|C_{\text{obs}},\Theta^k)dy_{m_1:m_\beta}\right]dx_{1:T}\\
&+ \sum_{t=m_1}^{t=m_\beta}\sum_{j=1}^{J}\int_{X,Y_m}\log p(y_t|x_t,\Theta_j)\cdot p(I_t=j|C_{\text{obs}},\Theta^k)\cdot p(x_{1:T},y_{m_1:m_\beta}|C_{\text{obs}},\Theta^k)dx_{1:T}dy_{m_1:m_\beta}\\
&+ \sum_{t=1}^{t=T}\sum_{j=1}^{J}\int_X \log p(I_t=j|\sigma_j,H_t)\cdot p(I_t=j|C_{\text{obs}},\Theta^k)\cdot\left[\int_{Y_m}p(x_{1:T},y_{m_1:m_\beta}|C_{\text{obs}},\Theta^k)dy_{m_1:m_\beta}\right]\cdot dx_{1:T}\\
&+ \sum_{j=1}^{J}\int_X \log p(x_1|\Theta_1)\cdot p(I_t=j|C_{\text{obs}},\Theta^k)\cdot\left[\int_{Y_m}p(x_{1:T},y_{m_1:m_\beta}|C_{\text{obs}},\Theta^k)dy_{m_1:m_\beta}\right]dx_{1:T}\\
&+ \sum_{t=2}^{t=T}\sum_{j=1}^{J}\int_X \log p(x_t|x_{t-1},\Theta_j)\cdot p(I_t=j|C_{\text{obs}},\Theta^k)\cdot\left[\int_{Y_m}p(x_{1:T},y_{m_1:m_\beta}|C_{\text{obs}},\Theta^k)dy_{m_1:m_\beta}\right]\cdot dx_{1:T}\\
&+ \sum_{j=1}^{J}\int_X \log C\cdot p(I_t=j|C_{\text{obs}},\Theta^k)\cdot\left[\int_{Y_m}p(x_{1:T},y_{m_1:m_\beta}|C_{\text{obs}},\Theta^k)dy_{m_1:m_\beta}\right]dx_{1:T}\\
&= \sum_{t=t_1}^{t=t_z}\sum_{j=1}^{J}\int_X \log p(y_t|x_t,\Theta_j)\cdot p(I_t=j|C_{\text{obs}},\Theta^k)\cdot p(x_{1:T}|C_{\text{obs}},\Theta^k)dx_{1:T}\\
&+ \sum_{t=m_1}^{t=m_\beta}\sum_{j=1}^{J}\int_{X,Y_m}\log p(y_t|x_t,\Theta_j)\cdot p(I_t=j|C_{\text{obs}},\Theta^k)\cdot p(x_{1:T},y_{m_1:m_\beta}|C_{\text{obs}},\Theta^k)\cdot dx_{1:T}dy_{m_1:m_\beta}\\
&+ \sum_{t=1}^{t=T}\sum_{j=1}^{J}\log p(I_t=j|\sigma_j,H_t)\cdot p(I_t=j|C_{\text{obs}},\Theta^k) + \sum_{j=1}^{J}\int_X\log p(x_1|\Theta_1)\cdot p(I_t=j|C_{\text{obs}},\Theta^k)\cdot p(x_1|C_{\text{obs}},\Theta^k)dx_1\\
&+ \sum_{t=2}^{t=T}\sum_{j=1}^{J}\int_X\log p(x_t|x_{t-1},\Theta_j)\cdot p(I_t=j|C_{\text{obs}},\Theta^k)\cdot p(x_{1:T}|C_{\text{obs}},\Theta^k)dx_{1:T} + \sum_{j=1}^{J}\log C\cdot p(I_t=j|C_{\text{obs}},\Theta^k)
\end{aligned} \quad (18)$$

The probability of the jth local model taking effect at the tth sampling time $p(I_t=j|C_{\text{obs}},\Theta^k)$ can be calculated as

$$p(I_t=j|C_{\text{obs}},\Theta^k) = \exp\left(\frac{-(H_t-H_j)^2}{2(\sigma_j)^2}\right) \quad (19)$$

where H_t denotes the measurement of the scheduling variable at time t, H_j is the jth operating point and σ_j represents the validity width of the j_{th} local model.

To evaluate the Q function in Eq. 18, the values of density functions $p(x_{1:T}|C_{\text{obs}},\Theta^k)$ and $p(x_{1:T},y_{m1:m\beta}|C_{\text{obs}},\Theta^k)$ are needed. Since direct calculations are intractable, those

density functions are to be numerically calculated using particle filter in the next section.

Computation Through Particle Filtering

Particle filters revisit

The basic idea of particle filters is to represent the desired posterior density function by a series of particles with associated weights, i.e., $\{x_t^i, w_t^i\}_{i=1}^N$. Then the density function of the states given the current estimation of parameters Θ^k can be discretely approximated as[10]

$$p(x_t|y_{t_1:t_\beta}, \Theta^k) \approx \sum_{i=1}^{N} \omega_t^i \delta(x_t - x_t^i) \tag{20}$$

where $\delta(\cdot)$ is the Dirac delta function, $t_\beta \le t$; N is the number of particles; ω_t^i is the normalized weight associated with the ith particle such that $\sum_{i=1}^N \omega_t^i = 1$. Suppose that at time $t-1$, a set of particles $\{x_{1:t-1}^i\}_{i=1}^N$ are available and we want to obtain N particles which represent the hidden state for time t. Since it is usually difficult to directly draw samples from the true posterior density $p(x_{1:t}|y_{t1:t\beta},\Theta^k)$, the principle of importance sampling[11] is adopted. The idea is to use a so called importance density $q(\cdot)$ from which one can easily draw samples $x_{1:t}^i, i = 1,\ldots,N$. Then the posterior is obtained by resampling important sampling. It has been shown that, as long as the support region of the posterior density belongs to that of the importance density, the particle approximation is unbiased.[12] The importance sampling is commonly chosen as the probability of state transition, i.e.

$$q(x_t|y_{t_1:t_\beta}, \Theta^k) = p(x_t|x_{t-1}, \Theta^k) \tag{21}$$

With this choice, the weight for each particle can be derived as[9]

$$\omega_t^i \propto \omega_{t-1}^i p(y_t|x_t^i, \Theta^k) \tag{22}$$

For time instants $t = m_1,\ldots,m_\beta$, when the outputs are not available, draw particles from the importance density $p(x_t|x_{t-1}^i,\Theta^k)$ and keep the weights unchanged, i.e.

$$\omega_t^i = \omega_{t-1}^i \tag{23}$$

To avoid the degeneracy problem,[10] the importance sampling step is usually followed by a resampling procedure. The idea is to discard the particles with small weights and concentrate on those with large weights. After resampling, each particle's weight will be reset to $\omega_t^i = \frac{1}{N}$.

Particle filters approximation and cautious resampling

The problem brought by brute force resampling is that it reduces the diversity among particles. One solution is to resample the particles only when it is necessary instead of performing it at each step. To be specific, N_{eff} is introduced to represent the effective particle number[13]

$$N_{\text{eff}} = \frac{1}{\sum_{i=1}^{N} (\omega_t^i)^2} \tag{24}$$

where ω_t^i is the normalized weight obtained through 22. It implies that, as the variance of the weights grows very large, the effective sample size decreases to a small number which indicates a severe degeneracy problem. In practice, one uses resampling to eliminate useless particles only when a severe degeneracy problem occurs, say, N_{eff} falls below the threshold N_{thred}.

Given the current estimation of parameters, the particle filter algorithm is summarized as follows:

Step 1. Initialization. *Draw initial N particles* $\{x_0^i\}_{i=1}^N$ *from the prior density* $p(x_0|\Theta^k)$ *and set each particle's weight to* $\frac{1}{N}$ *Set* $t = 1$.

Step **2. Importance Sampling**. *Generate predicted particles* $\{x_t^i\}_{i=1}^N$ *from the importance density* $p(x_t|x_{t-1},\Theta^k)$.

Step 3. Assigning Weights. *Assign the weight to each particle using Eq. 22 when* y_t *is available. Otherwise, calculate the weights according to Eq. 23.*

Step 4. Resampling. *Compute the number of effective particles using Eq. 24. If* N_{eff} *is less than the threshold* N_{thred}, *then perform resampling and replace the predicted particles in Step 2 with resampled particles. Reset the weights of resampled particles uniformly as* $\omega_t^i = \frac{1}{N}$. *Otherwise, go to Step 5.*

Step 5. *Set* $t = t + 1$ *and repeat Step 2 to Step 4 for* $t \le T$.

Estimation of $p(x_{1:T}|C_{\text{obs}},\Theta^k)$ and $p(x_{1:T},y_{m1:m\beta}|C_{\text{obs}},\Theta^k)$ is a problem of smoothing all states with all available observations. Its computation with the iterative EM algorithm is intensive. With further marginalization of the states and the missing observations following the approach of Gopaluni,[8] the Q function obtained in 18 can be rewritten as

$$\begin{aligned}
Q(\Theta|\Theta^k) = &\sum_{t=t_1}^{t=t_\alpha} \sum_{j=1}^{J} \int_X \log p(y_t|x_t,\Theta_j) \cdot p(I_t = j|C_{\text{obs}},\Theta^k) \cdot p(x_t|C_{\text{obs}},\Theta^k) dx_t \\
&+ \sum_{t=m_1}^{t=m_\beta} \sum_{j=1}^{J} \int_{X,Y_m} \log p(y_t|x_t,\Theta_j) \cdot p(I_t = j|C_{\text{obs}},\Theta^k) \cdot p(x_t,y_t|C_{\text{obs}},\Theta^k) \cdot dx_t dy_t \\
&+ \sum_{t=1}^{t=T} \sum_{j=1}^{J} \log p(I_t = j|\sigma_j,H_t) \cdot p(I_t = j|C_{\text{obs}},\Theta^k) + \sum_{j=1}^{J} \int_X \log p(x_1|\Theta_1) \cdot p(I_t = j|C_{\text{obs}},\Theta^k) \cdot p(x_1|C_{\text{obs}},\Theta^k) dx_1 \\
&+ \sum_{t=2}^{t=T} \sum_{j=1}^{J} \int_X \log p(x_t|x_{t-1},\Theta_j) \cdot p(I_t = j|C_{\text{obs}},\Theta^k) \cdot p(x_{t-1:t}|C_{\text{obs}},\Theta^k) dx_{t-1:t} + \sum_{j=1}^{J} \log C \cdot p(I_t = j|C_{\text{obs}},\Theta^k)
\end{aligned} \tag{25}$$

Calculation of $p(x_t|C_{\text{obs}},\Theta^k)$, $p(x_t,y_t|C_{\text{obs}},\Theta^k)$ and $p(x_{t-1},x_t|C_{\text{obs}},\Theta^k)$ is a smoothing problem, of which the computation cost is very high. A practical solution is to apply recursive state filtering such that $p(x_t|C_{\text{obs}},\Theta^k)$ is recursively approximated by $p(x_t|y_{t1:t\beta},\Theta^k)$ for $t = 1{:}\ T$, $p(x_t,y_t|C_{\text{obs}},\Theta^k)$ is recursively approximated by $p(x_t,y_t|y_{t1:t\beta},\Theta^k)$ for $t = 1{:}\ T$, and $p(x_t,x_{t+1}|C_{\text{obs}},\Theta^k)$ is recursively approximated by $p(x_t,x_{t+1}|y_{t1:t\beta},\Theta^k)$ for $t = 1{:}\ T-1$, where $t_\beta \le t$. This solution can significantly reduce the computation complexity and thus make the solution possible in real-time applications.

In Eq. 25, the density function $p(x_t|C_{\text{obs}},\Theta^k)$ is approximated using particle filters as

$$p(x_t|C_{\text{obs}},\Theta^k) \approx p(x_t|y_{t_1:t_\beta},\Theta^k) = \sum_{i=1}^{N} \omega_t^i \delta(x_t - x_t^i) \tag{26}$$

When the observation is missing, the joint density of x_t and y_t is required, which can be derived as

$$\begin{aligned} p(x_t,y_t|C_{\text{obs}},\Theta^k) &\approx p(x_t,y_t|y_{t_1:t_\beta},\Theta^k) \\ &= p(y_t|x_t,\Theta^k)p(x_t|y_{t_1:t_\beta},\Theta^k) \end{aligned} \tag{27}$$

Since y_t is missing, one can replace it with the predicted y_t^i which is the prediction using x_t^i such that

$$y_t^i = h(x_t^i,\Theta^k) \tag{28}$$

Therefore

$$p(x_t,y_t|C_{\text{obs}},\Theta^k) \approx \sum_{i=1}^{N} \omega_{t|x}^i \delta(x_t - x_t^i)\delta(y_t - y_t^i) \tag{29}$$

where

$$\omega_{t|x}^i = \frac{p(y_t^i|x_t^i,\Theta^k)p(x_t^i|y_{t_1:t_\beta},\Theta^k)}{\sum_{i=1}^{N} p(y_t^i|x_t^i,\Theta^k)p(x_t^i|y_{t_1:t_\beta},\Theta^k)} \tag{30}$$

Using Eq. 28

$$\begin{aligned} p(y_t^i|x_t^i,\Theta^k) &= \frac{1}{\sqrt{2\pi R}}\exp\left(-\frac{(y_t^i - h(x_t^i,\Theta^k))^2}{2R}\right) \\ &= \frac{1}{\sqrt{2\pi R}}\exp\left(-\frac{(h(x_t^i,\Theta^k) - h(x_t^i,\Theta^k))^2}{2R}\right) = 1 \end{aligned} \tag{31}$$

Hence

$$\omega_{t|x}^i = \frac{p(x_t^i|y_{t_1:t_\beta},\Theta^k)}{\sum_{i=1}^{N} p(x_t^i|y_{t_1:t_\beta},\Theta^k)} = \omega_t^i \tag{32}$$

As for the joint density function of x_t and x_{t+1}, it can be approximated as

$$\begin{aligned} p(x_t,x_{t+1}|C_{\text{obs}},\Theta^k) &\approx p(x_t,x_{t+1}|y_{t_1:t_\beta},\Theta^k) \\ &= p(x_{t+1}|x_t,\Theta^k)p(x_t|y_{t_1:t_\beta},\Theta^k) \\ &= \sum_{i=1}^{N} \omega_{t|t+1}^i \delta(x_t - x_t^i)\delta(x_{t+1} - x_{t+1}^i) \end{aligned} \tag{33}$$

where

$$\omega_{t|t+1}^i = \frac{p(x_{t+1}^i|x_t^i,\Theta^k)p(x_t^i|y_{t_1:t_\beta},\Theta^k)}{\sum_{i=1}^{N} p(x_{t+1}^i|x_t^i,\Theta^k)p(x_t^i|y_{t_1:t_\beta},\Theta^k)} \tag{34}$$

Substituting these approximated density functions, the Q function in Eq. 25 can be finally obtained.

$$\begin{aligned} Q(\Theta|\Theta^k) \approx & \sum_{t=t_1}^{t=t_\alpha}\sum_{j=1}^{J}\sum_{i=1}^{N} \omega_t^i \log p(y_t|x_t^i,\Theta_j)\cdot p(I_t = j|C_{\text{obs}},\Theta^k) \\ & + \sum_{t=m_1}^{t=m_\beta}\sum_{j=1}^{J}\sum_{i=1}^{N} \omega_t^i \log p(y_t^i|x_t^i,\Theta_j)\cdot p(I_t = j|C_{\text{obs}},\Theta^k) \\ & + \sum_{t=1}^{t=T}\sum_{j=1}^{J} \log p(I_t = j|\sigma_j,H_t)\cdot p(I_t = j|C_{\text{obs}},\Theta^k) \\ & + \sum_{j=1}^{J}\sum_{i=1}^{N} \omega_1^i \log p(x_1^i|\Theta_1)\cdot p(I_t = j|C_{\text{obs}},\Theta^k) \\ & + \sum_{t=2}^{t=T}\sum_{j=1}^{J}\sum_{i=1}^{N} \omega_{t-1|t}^i \log p(x_t^i|x_{t-1}^i,\Theta_j)\cdot p(I_t = j|C_{\text{obs}},\Theta^k) \\ & + \sum_{j=1}^{J} \log C\cdot p(I_t = j|C_{\text{obs}},\Theta^k) \end{aligned} \tag{35}$$

With the approximated Q function, the EM algorithm can hence be implemented. In the expectation step, the Q function is evaluated according to Eq. 35 with the current estimated parameters Θ_j^k, $j = 1{:}\ J$. In the next maximization step, the new parameters Θ_j^{k+1}, $j = 1{:}\ J$, are obtained by maximizing the Q function.

To maximize the Q function over parameters Θ, derivative operation is performed with respect to each parameter. Therefore, optimal estimation of system parameters at each iteration can be calculated by equating the derivatives to zero, i.e., $\frac{\partial Q}{\theta_{ji}} = 0$, where θ_{ji} is the ith system parameter for the jth local model.

The EM algorithm is summarized as follows:

Step 1. Initialization. *Start with the initial parameters* Θ_j^0, $j = 1{:}\ J$, *and set* $t = 0$.

Step 2. Expectation. *At time t, calculate the approximate Q function using Eq. 35, given the current estimation of the system parameters* Θ^k.

Step 3. Maximization. *Maximize the approximated Q function and get the new parameters* Θ_j^{k+1}, $j = 1{:}\ J$. *Set* $k = k + 1$.

Step 4. *Repeat Step 2 and Step 3 until the converge condition is satisfied, i.e., the change of the estimated parameters between two iterations is less than the tolerance*.

The validity for each local model σ_j, $j = 1{:}\ J$ also needs to be updated during each iteration. Due to the usage of the exponential function as it is shown in Eq. 5, an analytical expression is difficult to obtain when maximizing the Q function.[5] The mathematical formulation of the optimization problem in the search for optimal o_i, $i = 1,2\ldots J$ values can be expressed as

$$\begin{aligned} &\max_{o_j, j=1,2\ \ldots\ J} \sum_{t=1}^{t=T}\sum_{i=1}^{N}\sum_{j=1}^{J} \omega_t^i \log p(I_t = j|\sigma_j,H_t)\cdot P(I_t = j|C_{\text{obs}},\Theta^k) \\ &S.t. \quad o_{min} \le \sigma_j, j = 1,2\ \ldots\ J \le \sigma_{max} \end{aligned} \tag{36}$$

where $\log p(I_t = j \mid \sigma_j, H_t)$ can be calculated from 6. $p(I_t = j \mid C_{\text{obs}}, \Theta^k)$ represents the probability of the data point belonging to ith submodel at time t.

In this article, a constrained nonlinear optimization function named "fmincon" provided by "MATLAB" is adopted in a search for the optimal value for σ_j at each iteration of the EM algorithm.

Finally, having all the estimated model parameters Θ_j, $j = 1{:}\ J$ and validity σ_j, $j = 1{:}\ J$ for each local model, the global model can be obtained by substituting the estimated parameters into 4.

Simulations and Pilot-Scale Experiment

In this section, the proposed approach is evaluated through both numerical simulations as well as experimental verification. Its efficiency in handling missing outputs with less computational cost will be demonstrated. All the simulations were run on a 3.00 GHz CPU with 4 GB RAM PC using MATLAB 2009a.

A numerical simulation example

A first-order process with varying system parameters is utilized here to demonstrate the efficiency of the proposed parameter-varying model estimation method. This process was originally used in Zhu and Xu (2008)[14] as an illustrative example. It is described by the following equation

$$G(s,H) = \frac{K(H)}{\tau(H)s + 1} \tag{37}$$

where both the process gain $K(H)$ and the process time constant $\tau(H)$ are nonlinear functions of the scheduling variable H. The specific nonlinear relation is expressed as follows

$$K(H) = 0.6 + H^2, \quad H \in [1,4] \tag{38}$$

$$\tau(H) = 3 + 0.5H^3, \quad H \in [1,4] \tag{39}$$

By assuming that the observation y_t is a cosine function of the state, this process can be converted to the following nonlinear state space model

$$\begin{aligned} \dot{x}_t &= a(H)x_t + b(H)u_t + \omega_t \\ y_t &= cos(x_t) + v_t \end{aligned} \tag{40}$$

where

$$\begin{aligned} a(H) &= -\frac{1}{\tau(H)} \\ b(H) &= \frac{K(H)}{\tau(H)} \end{aligned} \tag{41}$$

Apparently, over the whole operating range of the process, the gain as well as the time constant changes dramatically and one single process model would hardly capture the dynamics of the process in its complete operating range. In other words, one local model cannot give a good approximation of the process behavior throughout the whole operating trajectory. Therefore, multiple models or a parameter-varying global model is required here to describe the behavior of the process under different operating conditions.

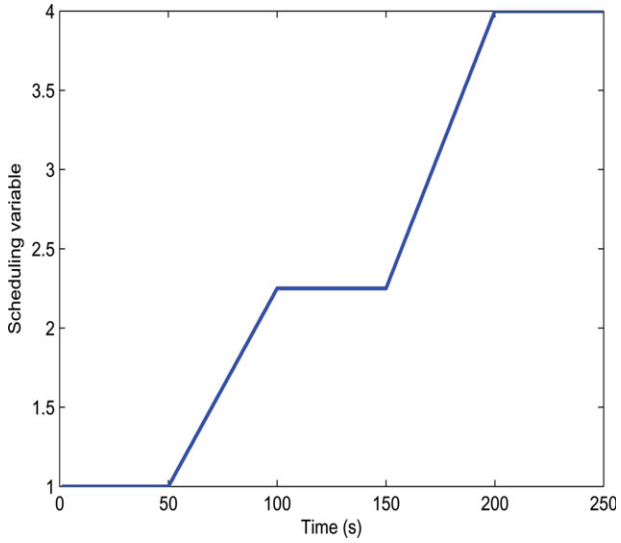

Figure 1. Trajectory of the scheduling variable H.

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

It is predetermined that the process is to be tested at three predesigned local operating points

$$H_1 = 1, \quad H_2 = 2.25, \quad H_3 = 4 \tag{42}$$

When transition from one operating point to the other, scheduling variable H is gradually increased by a fixed interval. Figure 1 shows the trajectory of the scheduling variable.

The process input u switches randomly among multiple levels throughout the whole experiment.

$T = 250$ measurements are collected from the simulation. Note that the relations between parameters and the scheduling variable expressed in Eqs. 38 and 39 are assumed unknown in the following identification process. The proposed multiple model parameter estimation method is then applied. To test the algorithm's capability in handling the missing data, different portions of the output data are randomly removed from the model training data set to simulate missing data problem. $N = 150$ particles are used for the particle filter computation.

In the expectation step of the EM algorithm, the Q function is calculated according to Eq. 35, where

$$\log[p(x_t^i | x_{t-1}^i, \Theta_j)] = \log\left[\frac{1}{\sqrt{2\pi Q_x}}\exp\left[-\frac{1}{2}\frac{(x_t^i - a_j x_{t-1}^i - b_j u_{t-1})^2}{Q_x}\right]\right] \tag{43}$$

For $t = t_1{:}\ t_\alpha$

$$\log[p(y_t | x_t^i, \Theta_j)] = \log\left[\frac{1}{\sqrt{2\pi Q_y}}\exp\left[-\frac{1}{2}\frac{(y_t - \cos x_t^i)^2}{Q_y}\right]\right] \tag{44}$$

For $t = m_1{:}\ m_\beta$

$$\log[p(y_t | x_t^i, \Theta_j)] = \log\left[\frac{1}{\sqrt{2\pi Q_y}}\exp\left[-\frac{1}{2}\frac{(y_t^i - \cos x_t^i)^2}{Q_y}\right]\right] \tag{45}$$

where $y_t^i = cos x_t^i$.

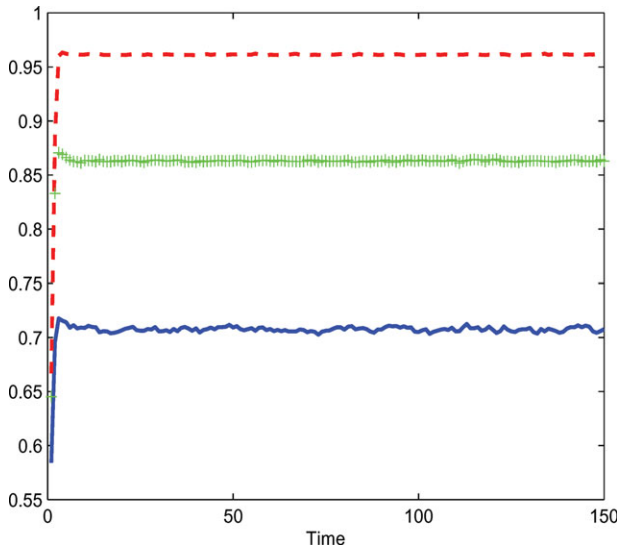

Figure 2. Trajectories of estimated parameter *a* for each local model when 25% observations are missing.

Blue solid line represents the trajectory of *a* for the first local model; green plus line represents the trajectory of *a* for the second local model; red dash line is the the trajectory of *a* for the third local model. [Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com.]

As for $p(I_t = j|C_{\text{obs}},\Theta^k)$, it can be calculated according to Eq. 6 as

$$p(I_t = j|C_{\text{obs}}, \Theta^k) = \exp\left(\frac{-(H_t - H_j)^2}{2(\sigma_j)^2}\right) \tag{46}$$

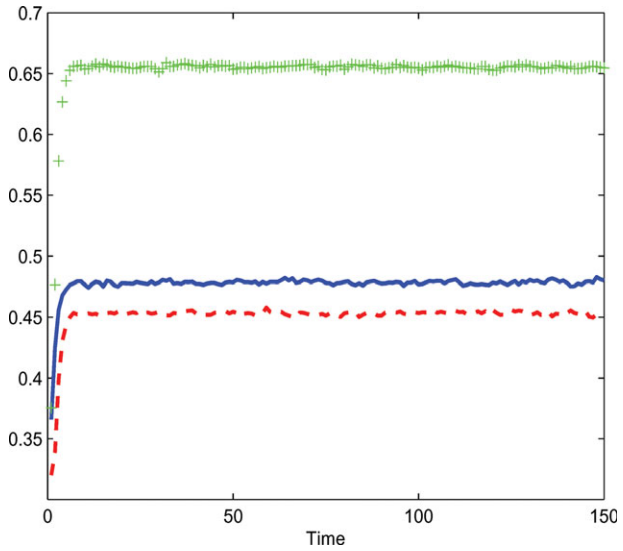

Figure 3. Trajectories of estimated parameter *b* for each local model when 25% observations are missing.

Blue solid line represents the trajectory of *b* for the first local model; green plus star line represents the trajectory of *b* for the second local model; red dash line is the the trajectory of *b* for the third local model. [Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com.]

Table 1. Estimated Parameter a After 150 Iterations

	True Value		
	$a_1 = 0.7143$	$a_2 = 0.8850$	$a_3 = 0.9715$
Proportion of missing output	a_1	a_2	a_3
Full data set	0.7258	0.8952	0.9628
25%	0.7074	0.8628	0.9615
50%	0.7021	0.8654	0.9751

In the maximization step of the EM algorithm, by taking derivative over the Q function and equating it to zero, each individual component of the parameters is hence calculated as

$$a_j^{\text{new}} = \frac{\sum_{t=2}^{T}\sum_{i=1}^{N}\omega_t^i \cdot \exp(\frac{-(H_t-H_j)^2}{2(\sigma_j)^2}) \cdot (x_t^i x_{t-1}^i - b^{\text{old}} x_{t-1}^i u_{t-1})}{\sum_{t=2}^{T}\sum_{i=1}^{N}\omega_t^i \cdot \exp(\frac{-(H_t-H_j)^2}{2(\sigma_j)^2}) \cdot (x_{t-1}^i)^2} \tag{47}$$

$$b_j^{\text{new}} = \frac{\sum_{t=2}^{T}\sum_{i=1}^{N}\omega_t^i \cdot \exp(\frac{-(H_t-H_j)^2}{2(\sigma_j)^2}) \cdot (x_t^i u_{t-1} - a^{\text{old}} x_{t-1}^i u_{t-1})}{\sum_{t=2}^{T}\sum_{i=1}^{N}\omega_t^i \cdot \exp(\frac{-(H_t-H_j)^2}{2(\sigma_j)^2}) \cdot u_{t-1}^2} \tag{48}$$

The trajectories of the estimated parameters for each local model when 25% output data are missing are shown in Figures 2 and 3. The estimated parameter values after 150 iterations are given in Tables 1 and 2.

The comparison result of the identified global model with the true output is displayed in Figure 4. Here to better test the validity of the identified model, model validation is conducted under other two different operating points

$$H_4 = 1.75, \quad H_5 = 3 \tag{49}$$

and the comparison of the global model prediction with the true process output is shown in figure 5.

Figure 6 provides a weighting map of each local model under different scheduling values. Based on this calculated weighting map as well as Eq. 4, model predictions can be calculated for all the H values.

Comparison result displayed in Figures 4 and 5 shows that the identified global model not only can well capture the process dynamics under the training operating conditions, but also perform well in capturing the process dynamics at other operating points that are different from the operating points within the training data. This confirms the effectiveness of the identified global model in approximating the real process dynamics throughout the operating range.

Table 2. Estimated Parameter b after 150 Iterations

	True Value		
Proportion of missing output	$b_1 = 0.4571$ b_1	$b_2 = 0.6512$ b_2	$b_3 = 0.4743$ b_3
Full data set	0.4650	0.6601	0.4798
25%	0.4779	0.6548	0.4534
50%	0.4716	0.6570	0.4681

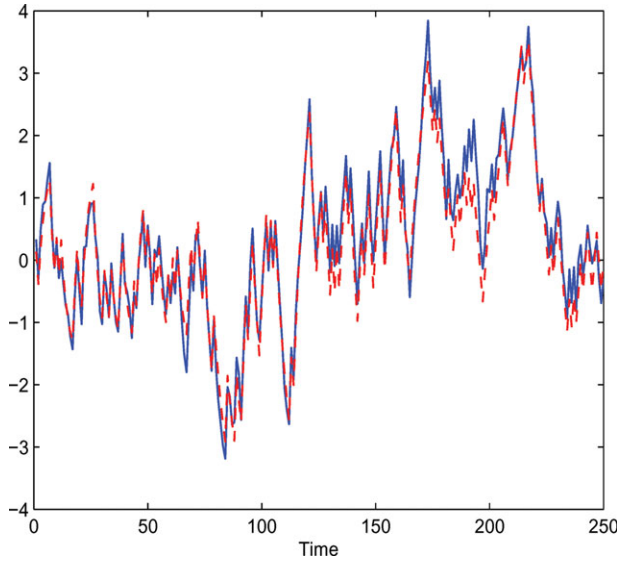

Figure 4. Validation of the identified global model against the model training data set.

Blue line represent the real process output and the red dash line is the simulated output from the identified global model. [Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com.]

Continuous stirred tank reactor

This model has been utilized as the illustration examples in Gopaluni (2008)[8] and Jin and Huang (2010).[5] The system is described by the following set of differential equations

$$\frac{dC_A(t)}{dt} = \frac{q(t)}{V}(C_{A0}(t) - C_A(t)) - k_0 C_A(t)\exp\left(\frac{-E}{RT(t)}\right) \quad (50)$$

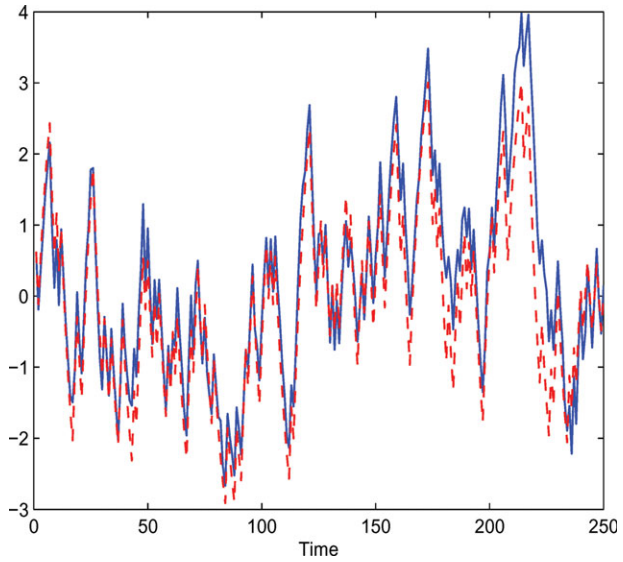

Figure 5. Cross validation of the identified global model.

Blue line represent the real process output and the red dash line is the simulated output from the identified global model. [Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com.]

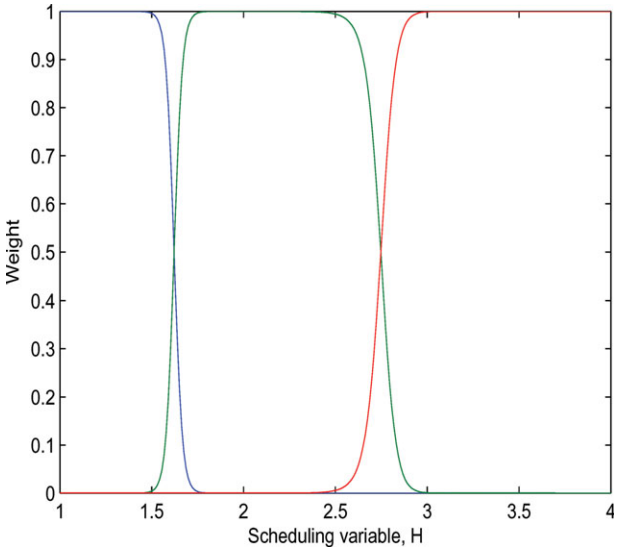

Figure 6. Weight of each local model at different operating points.

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

$$\frac{dT(t)}{dt} = \frac{q(t)}{V}(T_0(t) - T(t)) - \frac{(\Delta H)k_0 C_A(t)}{\rho C_p}\exp\left(\frac{-E}{RT(t)}\right) + \frac{\rho_c C_{pc}}{\rho C_p V}q_c(t)\left\{1 - \exp\left(\frac{-hA}{q_c(t)\rho C_p}\right)\right\}(T_{c_0}(t) - T(t)) \quad (51)$$

where C_A is the outlet reagent concentration (g/l); T is the reactor temperature (g/l); the inlet flow rate q is the system input. The explanations of the system variables and their corresponding steady state values are given in Table 3. We consider the temperature T as an interested output variable. The coolant flow rate q_c has a direct impact on the process dynamic; and it is the scheduling variable H from which the operation condition of the process is determined. The trajectory of the scheduling variable is given in Figure 7.

The system parameters to be estimated are θ_1 and θ_2 which are both functions of the scheduling variable q_c such that

$$\theta_1 = \frac{\rho_c C_{pc}}{\rho C_p}H(t) \quad (52)$$

Table 3. CSTR Model Parameters and Their Steady State Values

Parameters	Steady State Value
Production concentration of component A, C_A	$output_1$
Temperature of the reactor, T	$output_2$
Feed Concentration of component A, C_{A0}	1 mol/L
Feed temperature, T_0	350.0 K
Specific heats, C_p, C_pc	1 cal/(g K)
Liquid density, ρ, ρ_c	1×10^3 g/L
Heat of reaction, ΔH	-2×10^5 cal/mol
Activation energy term, E/R	1×10^4 K
Reaction rate constant, k_0	$7.2 \times 10^{10}\text{min}^{-1}$
Heat transfer term, hA	7×10^5 cal/(min K)
The reactor volume, V	100 L
Inlet coolant temperature, T_{c0}	350.0 K
Process flow rate, q	Input
Coolant flow rate, q_c	Scheduling variable

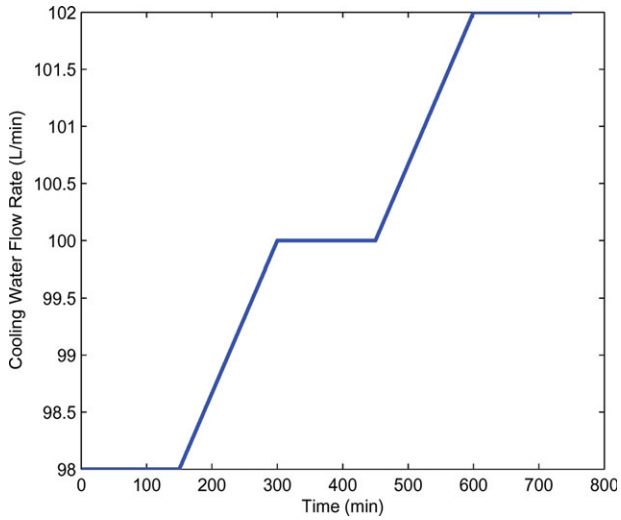

Figure 7. Trajectory of the scheduling variable H.

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

$$\theta_2 = \frac{hA}{H(t)\rho C_p} \tag{53}$$

Finally the differential equations can be written in the following form with θ_1 and θ_2 as varying unknown parameters

$$\frac{dC_A(t)}{dt} = \frac{q(t)}{V}(C_{A0}(t) - C_A(t)) - k_0 C_A(t)\exp\left(\frac{-E}{RT(t)}\right) \tag{54}$$

$$\frac{dT(t)}{dt} = \frac{q(t)}{V}(T_0(t) - T(t)) - \frac{(\Delta H)k_0 C_A(t)}{\rho C_p}\exp\left(\frac{-E}{RT(t)}\right) + \theta_1\{1 - \exp(-\theta_2)\}(T_{c_0}(t) - T(t)) \tag{55}$$

The process is operated at three different operating points $H_1 = 98$ L/min; $H_2 = 100$ L/min; $H_3 = 102$ L/min; during the transition period the dilution factor is increased by a fixed step size and no additional excitation signal is added to the dilution factor. System parameters θ_1 and θ_2 vary with the scheduling variable, resulting in a time varying nonlinear process which cannot be adequately described by a single nonlinear model. The experiment is performed under these three operation conditions and the training data collected with 25% output data randomly erased.

The algorithm is applied to the training data and an approximate global model is obtained afterwards. Once again the explicit relations expressed in Eqs. 52 and 53 are assumed unknown in the following identification process. The result of the model validation for the training data is shown in Figure 9.

Figure 10 provides a weighting map of each local model under different scheduling values. Based on this calculated

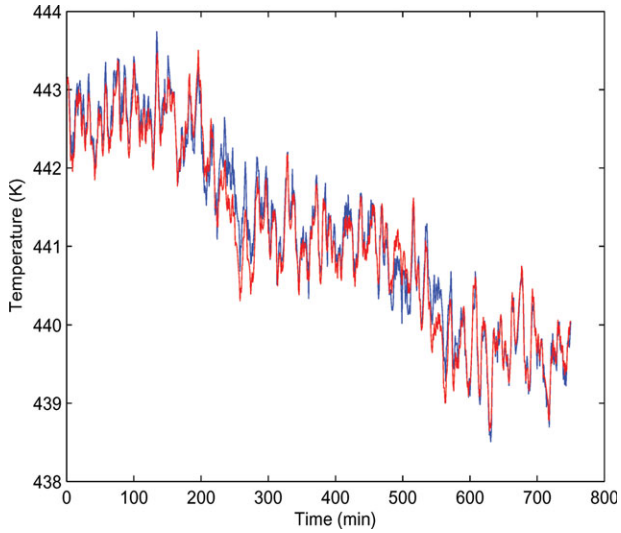

Figure 8. Validation of the identified global model against the model training data set.

Blue line represent the real process output and the red line is the simulated output from the identified global model. [Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com.]

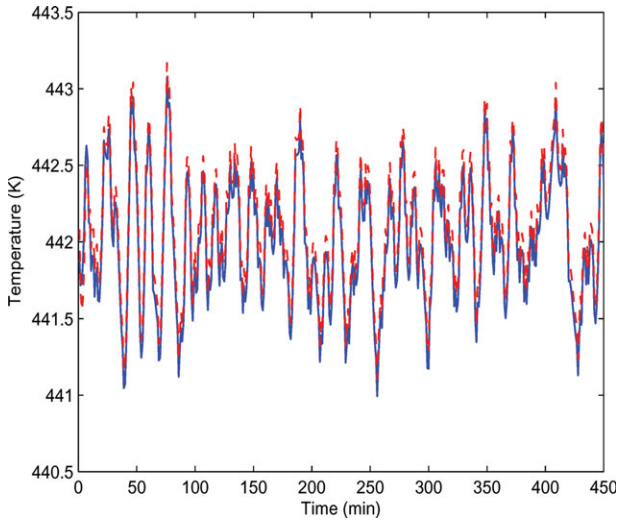

Figure 9. Cross validation of the identified global model.

Blue line represent the real process output and the red line is the simulated output from the identified global model. [Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com.]

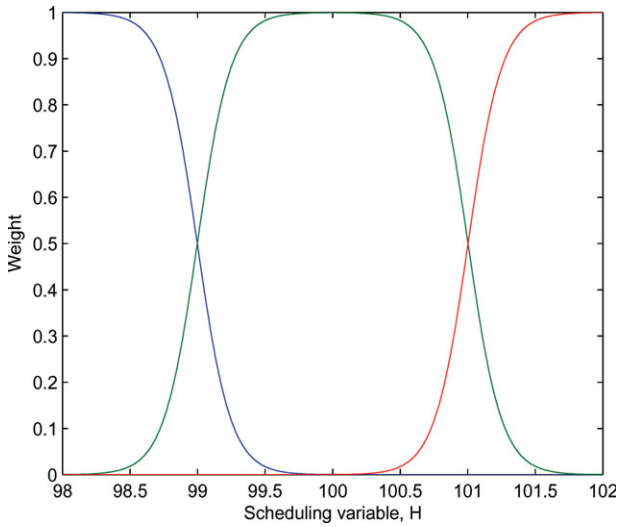

Figure 10. Weight of each local model at different operating points.

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

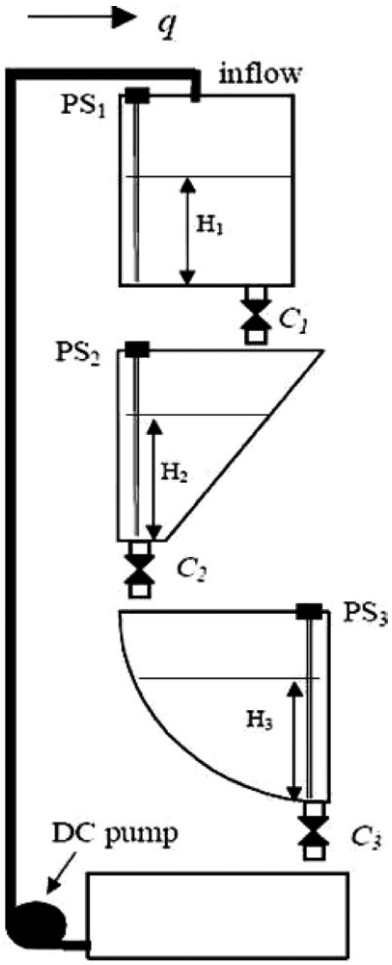

Figure 11. Three tank system schematic.[6]

weighting map as well as Eq. 4, model predictions can be calculated under all the H values within the operating range.

Comparison results displayed in Figures 8 and 9 show that the identified global model can well capture the process dynamics at both training operating points as well as other operating points that are not included in the training operating points.

Experimental evaluation: a multitank system

In this section, an experimental evaluation on a three-tank system is performed to further verify the effectiveness of the proposed algorithm. Figure 11 illustrates the simplified process setup of the three-tank system which consists of three tanks placed on top of each other. During the experiment, water is pumped from the bottom supply tank into the top tank and then flows to each tank by gravity. There are three valves, one for each tank, by adjusting which the outflow rate from each tank can be controlled. The nonlinear model describing the process dynamics is given by

$$\begin{aligned}\frac{dH_1}{dt} &= \frac{1}{\beta_1}q - \frac{1}{C_1 H_1^{\alpha_1}} C_1 H_1^{\alpha_1} \\ \frac{dH_2}{dt} &= \frac{1}{\beta_2} C_1 H_1^{\alpha_1} - \frac{1}{\beta_2} C_2 H_2^{\alpha_2} \\ \frac{dH_3}{dt} &= \frac{1}{\beta_3} C_2 H_2^{\alpha_2} - \frac{1}{\beta_3} C_3 H_3^{\alpha_3}\end{aligned} \tag{56}$$

Table 4. Designed Operating Point for the Experiment

Operating points, $H_{1m}, m = 1,2,3,4$
5 cm
10 cm
14 cm

where
- q is the inlet flow rate into the upper tank;
- H_i is the water level of the ith tank, $i = 1,2,3$;
- C_i is the resistance of the output orifice of the ith tank, $i = 1,2,3$;
- β_i is the cross sectional area of the ith tank, $i = 1,2,3$;
- α_i is the flow coefficient of the ith tank, $i = 1,2,3$.

The process dynamics of the second tank is of interest in this study. The state space model with the water level in the second tank H_2 as the state is given by

$$\begin{aligned}\frac{dH_2}{dt} &= \frac{1}{\beta_2} C_1 H_1^{\alpha_1} - \frac{1}{\beta_2} C_2 H_2^{\alpha_2} + w_t \\ y_t &= x_t + v_t\end{aligned} \tag{57}$$

From Eq. 57, it can be seen that water level of the first tank has a direct impact on the second tank water level. Therefore, H_1 is chosen as the scheduling variable of the system. Three different operating points are selected as shown in Table 4. The water level of the top tank H_1 is maintained at the desired value through a PID controller by manipulating the inlet flow rate into the top tank.

When transition from one operating point to the other, the water level in the top tank increases at a fixed step over the transition period until reaches the value of the next operating point. The trajectory of the scheduling variable H_1 is shown in Figure 12.

The value of α_2 is chosen as 0.5 according to the multi-tank system experiment manual. This is an appropriate assumption as the inlet flow rate is fairly small so that the water dynamic in each tank can be considered as laminar

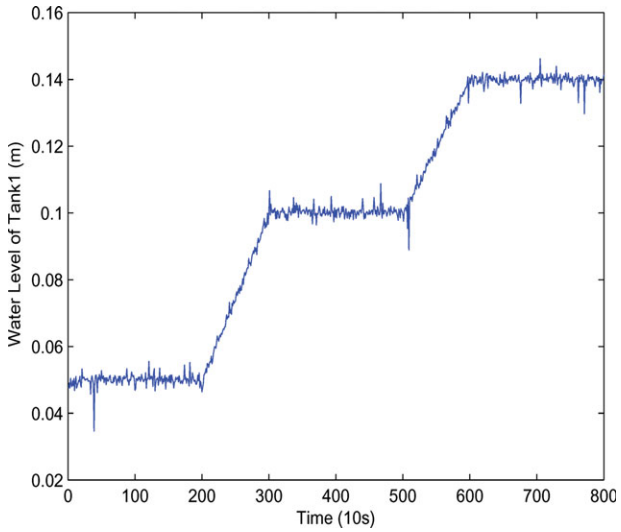

Figure 12. Three tank system scheduling variable for the self validation.

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

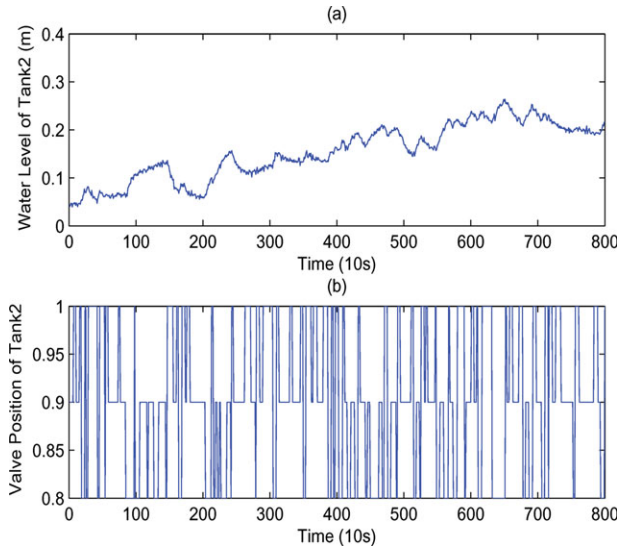

Figure 13. Three tank system input–output data (a) water level of the second tank, process output (b) valve position V2, process input.

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

flow. The output orifice resistance C_2 is directly affected by the valve position of the second tank V_2; hence C_2 can be approximated as linear function of V_2 such that

$$C_2 \approx \alpha \cdot V_2 \tag{58}$$

The sectional area of the second tank is given below

$$\beta_2 = c \cdot w + \frac{H_2}{H_{2\max}} \cdot b \cdot w \tag{59}$$

where c, w, b, $H_{2\max}$ are the geometrical parameters of the second tank with each value given as

$$c = 10\,\text{cm}, \quad w = 3.5\,\text{cm}, \quad b = 34.5\,\text{cm}, \quad H_{2\max} = 35\,\text{cm} \tag{60}$$

By substituting Eqs. 59 and 58 in 57, the state equation becomes

$$\frac{dH_2}{dt} = \frac{H_{2\max} \cdot C_1 \cdot H_1^{\alpha_1}}{w(cH_{2\max} + bH_2)} - \frac{H_{2\max} \cdot \alpha \cdot V_2 H_2^{0.5}}{w(cH_{2\max} + bH_2)} + w_t \tag{61}$$

Therefore, the state equation of H_2 can be rearranged as

$$\frac{dH_2}{dt} = \frac{\theta_1}{cH_{2\max} + bH_2} - \frac{\theta_2 \cdot V_2 \cdot H_2^{0.5}}{cH_{2\max} + bH_2} \tag{62}$$

where θ_1 is function of H_1 such that

$$\theta_1 = \frac{H_{2\max} \cdot C_1 \cdot H_1^{\alpha_1}}{w} \tag{63}$$

$$\theta_2 = \frac{H_{2\max} \cdot \alpha}{w} \tag{64}$$

A multiple level random signal is designed for the valve position V_2 as the system input. The process input and output data are given in Figure 13.

The relations between parameters and the scheduling variable expressed in Eqs. 63 and 64 are assumed unknown in identification. The collected experimental data is then passed through the proposed algorithm with 20% output data randomly removed. The self validation result is given in Figure 14.

It can be seen from Figure 14 that the estimated global model performs well in the self validation. Another experiment is conducted for the cross validation to test the global model's capability in predicting the dynamics of the process at different operating points. In the cross validation experiment, the scheduling variable H_1 is maintained at 8 cm during the whole cross validation experiment. The input signal and the output data for cross validation is shown in Figure 15.

Figure 16 gives the result of the cross validation which shows a good match between the real process output and the prediction of the identified global model, confirming the effectiveness of the proposed algorithm.

Discussion and Conclusion

This article described a Bayesian approach for identifying parameter-varying nonlinear state space model with missing output data within the framework of the EM algorithm. Particle filter is used for the calculation in the Expectation step. The capability of the proposed algorithm in handling missing observations in the presence of varying parameters is demonstrated through numerical examples as well as a pilot-scale experiment.

What have been discussed in this article so far mainly focus on the parameter estimation using a complete data set which is usually considered as an off-line analysis. The particle filtering approximation used in this article reduced the computation burden compared with the particle smoothers. The EM algorithm is an iterative method which takes certain time for the estimation to reach convergence. For example, for the data set with the length of 500 data point, the EM algorithm would take a couple of minutes to reach

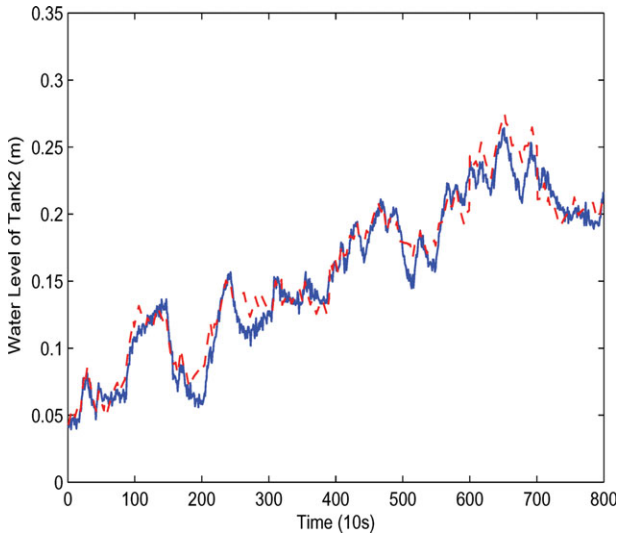

Figure 14. Self-validation result.

Blue solid line is the collected process data, red dash line is the simulated output of the identified global model. [Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com.]

convergence. When it comes to on-line implementation, it is possible to apply the proposed method for model parameter estimation with proper sampling time and data length.

The model identification is an important step towards process monitoring and process control. Obtaining a fairly accurate model not only help better understand the process behavior, but can also be used for controller design purpose. For example, in the MPC which is a multivariable control algorithm, an internal dynamic model of the process is a prerequisite. For nonlinear systems with multiple operating points, single model may not be sufficient to describe the process and thus not suitable for controller design. The algorithm proposed in this article takes all the operating conditions into account and provides estimation of the global model parameters. Therefore, it is possible to design a MPC controller based on this identified global model and the control signal can be accordingly calculated.

It is doable to extend the proposed algorithm to deal with the parameter identification problem for hybrid systems. In this article, the parameter estimation problem is formulated under the framework of the EM algorithm. In addition to the hidden state $x_{1:T}$ and missing output Y_m, the hidden model identity I is introduced to denote which local model takes effect. As for hybrid system, take the linear piecewise autoregressive exogenous (PWARX) process as an example which is formulated as below

$$y_k = \begin{cases} \theta_1^T \begin{bmatrix} x_k \\ 1 \end{bmatrix} + e_k, & x_k \in \chi_1 \\ \vdots & , k = 1, 2 \ldots N \\ \theta_M^T \begin{bmatrix} x_k \\ 1 \end{bmatrix} + e_k, & x_k \in \chi_M \end{cases} \tag{65}$$

where N, M represent number of data points collected and number of submodels, respectively, y_k is the output, x_k is the regressor which consists of past input and output, e_k is the Gaussian distributed noise with zero mean and variance σ^2, θ_i is the parameter vector of the ith submodel. $Z_k = x_k, y_k, k = 1,2,\ldots,N$ is defined as the observed data set generated from a PWARX system. I is introduced as a "missing variable" to denote the submodel identity of each data point. With the defined observed data set and missing data set, the identification problem can be formulated under the framework of the EM algorithm. More discussion with detailed formulation of hybrid PWARX system can be found in Jin and Huang (2010).[15]

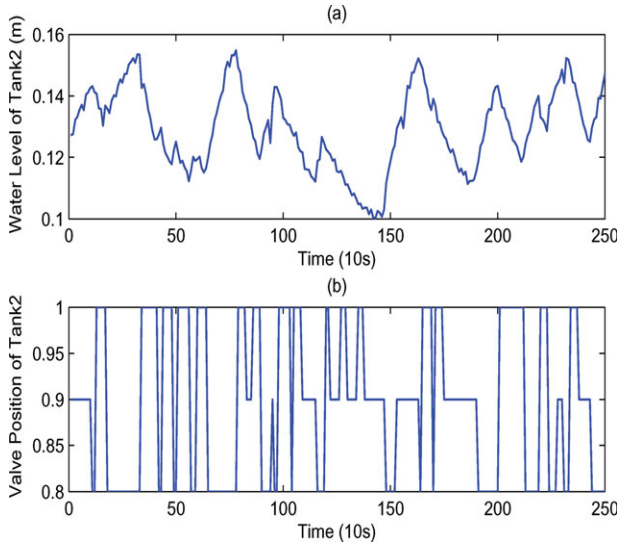

Figure 15. Three tank system input–output data (a) water level of the second tank, process output (b) valve position V2, process input.

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

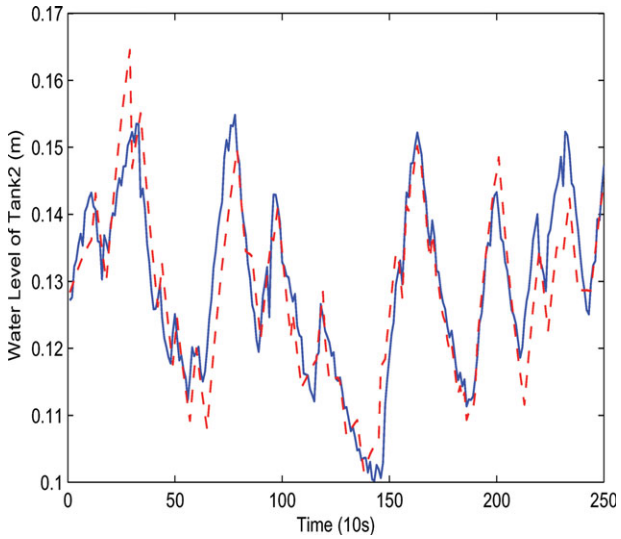

Figure 16. Cross-validation result.

Blue solid line is the collected process data, red dash line is the simulated output of the identified global model. [Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com.]

Acknowledgments

This work is supported in part by Natural Sciences and Engineering Research Council of Canada.

Literature Cited

1. Pearson PK. Control systems, identification. *Encyclopedia Phys Sci Technol*. 2004;687–707.
2. Feeny BF. Nonlinear system identification. *Encyclopedia Vib*. 2004;924–928.
3. Shamma J, Athans M. Guaranteed properties ff gain scheduled control for lienar parameter varying plants. *Automatica*. 1991;27:559–564.
4. Xu Z, Zhao J, Qian J. Nonlinear MPC using an identified LPV Model. *Ind Eng Chem Res*. 2009;49:3043–3051.
5. Jin X, Huang B. Multiple model LPV approach to nonlinear process identification with EM algorithm. *J Process Control*. 2010;21:182–193.
6. Schon TB, Wills A, Ninness B. System Identification of Nonlinear State-Space Models. *Automatica*. 2011;47:39–49.
7. Khatibisepehr S, Huang B. Dealing with irregular data in soft Sensors: Bayesian method and comparative study. *Ind Eng Chem Res*. 2008;22:8713–8723.
8. Gopaluni RB. A particle filter approach to identification of nonlinear process under missing observations. *Can J Chem Eng*. 2008;86:1081–1092.
9. Demspter AP, Larid NM, Rubin DB, Maximum likelihood from incomplete data via the EM algorithm. *J R Stat Soc, Ser B*. 1977;39:1–38.
10. Arulampalam MS, Maskell S, Gordon N, Clapp T, A tutorial on particle filters for onine nonlinear/non-Guassian Bayesian tracking. *IEEE Trans Signal Process*. 2002;50:174–188.
11. Bergman N. Recursive Bayesian estimation: navigation and tracking applications. Ph.D thesis, Linköping University, Linköping, Sweden, 1999.

12. Doucet A. On sequential Monte Carlo methods for Bayesian filtering, Department of Engineering, University of Cambridge, UK, Technical Report, 1998.
13. Liu JS, Chen R, Sequential Monte Carlo methods for dynamical systems. *J Am Stat Assoc*. 1998;93:1032–1044.
14. Zhu Y, Xu Z. A method of LPV model identification for control. In: *Proceeding of the 17th IFAC World Congress, Seoul, Korea*, 2008.
15. Jin X, Huang B. Robust identification of piecewise/switching autoregressive exogenous process. *Am Inst Chem Eng J*. 2010;56:1829–1844.

Manuscript received Apr. 10, 2011 and revision received Dec. 22, 2011.