

Identification of Linear Parameter Varying Systems with Missing Output Data Using Generalized Expectation-Maximization Algorithm ^{*}

X. Yang ^{*} W. Xiong ^{**} B. Huang ^{***} H. Gao ^{****}

^{*} *Research Institute of Intelligent Control and Systems, Harbin Institute of Technology, Harbin, Heilongjiang 150080, China (e-mail: xianqiangyang@gmail.com).*

^{**} *Key Laboratory of Advanced Process Control for Light Industry (Ministry of Education), Jiangnan University, Wuxi, Jiangsu 214122, China (e-mail: greenpre@163.com).*

^{***} *Department of Chemical and Materials Engineering, University of Alberta, Edmonton, Alberta T6G 2G6, Canada (e-mail: biao.huang@ualberta.ca).*

^{****} *Research Institute of Mechatronics and Automation, Bohai University, Jinzhou 121013, China, and Research Institute of Intelligent Control and Systems, Harbin Institute of Technology, Harbin, Heilongjiang 150080, China (e-mail: huijungao@gmail.com).*

Abstract: This paper is concerned with the identification problems of linear parameter varying (LPV) systems with randomly missing output data. Since one local linearized model cannot capture the global dynamics of the nonlinear industrial process, the multiple-model LPV model in which the global model is constructed by smoothly weighted combination of multiple local models is considered here. The problem of missing output variables data is commonly encountered in practice. In order to handle the multiple-model identification problems of LPV systems with incomplete data, the local model is taken to have a finite impulse response (FIR) model structure and the generalized expectation-maximization (EM) algorithm is adopted to estimate the unknown parameters of the global LPV model. To avoid the problems of ill-conditioned matrices and high sensitivity of parameters to noise, the prior information on the coefficients of each local FIR model is employed to construct the prior probability of unknown parameters. Then the maximum a posteriori (MAP) estimates of the global model parameters are derived via the generalized EM algorithm. The numerical example is presented to demonstrate the effectiveness of the proposed method.

Keywords: Linear parameter varying systems; Generalized EM algorithm; Missing output data; Maximum a posteriori estimates.

1. INTRODUCTION

In process industry, many advanced control strategies have been developed to meet various process requirements, such as improved process safety and reliability, consistent production, economic optimization, and so on. Typically, the application of these control strategies relies heavily on the understanding of the process behaviors and the existence of an accurate process model (Yin et al., 2012, 2014). Therefore, process modeling is a prerequisite for controller design and system analysis.

Modern processes are often operated in a wide working range, and exhibit strong nonlinearity and/or parameter

varying property (Zhao et al., 2012). Since the process dynamic changes inevitably when the process shifts from one working point to the other, one local linearized model cannot capture the global dynamics of the process. Meanwhile, the closed-loop system with the controller designed based on the local linearized model cannot meet the expected performance requirements within a large working range. To deal with these problems, great efforts have been made to find a flexible model structure to facilitate the modeling and controller design for modern processes. Among the results obtained in the literature, the linear parameter varying (LPV) model has attracted great attentions of many researchers (Bamieh and Giarre, 2002; Laurain et al., 2010; Jin et al., 2011). The LPV model is characterized by its linear model structure and is capable of approximating complicated nonlinear/parameter varying processes.

^{*} This work was supported by the Natural Sciences and Engineering Research Council (NSERC) of Canada, the National Natural Science Foundation of China (No: 21206053), and China Scholarship Council (CSC).

In the past decade, many LPV model identification methods have been developed. These methods can be roughly divided into the global approach (Bamieh and Giarre, 2002; Laurain et al., 2010; Tóth et al., 2012) and the local approach (Wassink et al., 2005; Xu et al., 2009; Jin et al., 2011). The global approach typically requires the control input to be manipulated throughout the entire operating range, so that the control inputs and the scheduling parameters can be both excited. This may not be allowed by many processes for the consideration of safety, consistent production, and so on. In the local approach, the control input is only required to be excited around each prespecified working point or for the fixed scheduling variables which is easy to implement in practice. Wassink et al. (2005) proposed to build an LPV model with polynomial dependence on parameters for the wafer stage with position-dependent dynamics. The position of the stage in the horizontal plane was selected as the scheduling variable. The local linear model was identified for each preselected position, and then the parameters of the coefficient polynomials of the final parameter interpolation LPV model were derived by solving two least squares optimization problems. Jin et al. (2011) put forward to deal with the LPV modeling problem using the expectation-maximization (EM) algorithm originally developed by Dempster et al. (1977). Based on the assumption of smooth transitions, the normalized exponential function was chosen as the validity function. The local model parameters and the validity widths were estimated simultaneously (Jin et al., 2011).

The problem of missing output variables data is commonly encountered in process industry due to sensor failure, irregularly sampling, packet losses or disorder in network communication system, malfunction of the data recording system, data deletion caused by outlier detection and so on. The conventional identification methods may suffer from performance deterioration when applied to the incomplete dataset directly. In order to handle missing data in the modeling process, many techniques, such as casewise deletion, mean substitution, last observation carried forward method, regression imputation, EM-based Bayesian algorithm, etc., have been developed (Khatibisepehr and Huang, 2008). Among all the techniques, the EM algorithm has already been widely applied owing to its attractive statistical properties (Jin et al., 2012). For example, the identification problems of the nonlinear process with missing observation were considered by Gopaluni (2008). The EM algorithm was employed to handle the missing data and hidden state by integrating them into the missing dataset. In the expectation step (E-step), the particle filters based smoothers were utilized to calculate the approximations of the expectation functions. The standard optimization routines were then performed in the maximization step (M-step) to derive the formulas of parameter estimates. Deng and Huang (2012) extended this work to parameter varying nonlinear system modeling with missing data. The missing output data and parameter varying problems in nonlinear state space model parameter estimation were solved simultaneously by using the EM algorithm (Deng and Huang, 2012). The particle filter, rather than the smoother, was used to calculate the approximations of the expectation functions in E-step. Since particle filter or smoother is used, both proposed

algorithms are computationally intensive which makes it difficult for on-line implementation.

The work introduced in this paper aims at developing an algorithm for identifying the LPV input-output models with missing output data. The generalized EM algorithm is adopted to handle the identification problems. Since the modeling of the LPV system is often obtained in input-output form, the input-output LPV model is considered here (Laurain et al., 2010). The local approach is adopted and the process data are collected by exciting the process following a predesigned operating trajectory. The normalized exponential functions are employed to combine the local models to approximate the transition dynamic periods of the process. In order to handle the missing output data in the generalized EM algorithm, the local models are taken to have a finite impulse response (FIR) model structure. Since the order of the local FIR model is typically very high, especially for slow process, in order to well capture the dynamics of the process, this may result in the problem of ill-conditioned matrices in the identification process. Though the FIR model is very flexible, the FIR model parameters are very sensitive to the noise. To avoid these problems, the prior information that the coefficients of the each local FIR model vary smoothly is employed to construct the prior probability of the unknown process parameters. Then the identification problems is formulated in a Bayesian framework and the generalized EM algorithm is utilized to derive the maximum a posteriori (MAP) estimates of the global model parameters.

The reminder of this paper is organized as follows: A brief revisit of the generalized EM algorithm is given in Section 2. The mathematical formulation for LPV model identification with incomplete dataset is presented in Section 3. In Section 4, the simulation examples are given to verify the efficiency of the proposed algorithm. The conclusions are given in Section 5.

2. GENERALIZED EM ALGORITHM REVISIT

The generalized EM algorithm is a well-known iterative optimization algorithm to calculate the maximum likelihood (ML) estimate and the MAP estimate in incomplete-data problems. Denote C_{obs} as the observation data set and C_{mis} as the missing data set. Then the complete data C can be defined as $C \triangleq \{C_{obs}, C_{mis}\}$. If the prior probability $p(\Theta)$ for unknown parameter vector Θ is known, the MAP estimate of Θ can be calculated in the Bayesian framework as

$$\begin{aligned} \hat{\Theta}_{MAP} &= \underset{\Theta}{\operatorname{argmax}} \log p(C_{obs}, \Theta) \\ &= \underset{\Theta}{\operatorname{argmax}} [\log p(C_{obs}|\Theta) + \log p(\Theta)] \quad (1) \end{aligned}$$

However, it is typically not tractable to calculate the MAP estimate directly in incomplete-data problem. In this case, the generalized EM algorithm can be easily modified to derive the MAP estimate. The procedures for the generalized EM algorithm to calculate the MAP estimate can be described as (McLachlan and Krishnan, 2007):

1. E-step: Given the observation data set C_{obs} and the current MAP estimate $\Theta^{(s)}$ of Θ , the conditional expect-

tation of the log complete data posterior density function $J(\Theta|\Theta^{(s)})$ can be determined by

$$\begin{aligned} J(\Theta|\Theta^{(s)}) &= E_{C_{mis}|C_{obs},\Theta^{(s)}}\{\log p(C|\Theta)\} + \log p(\Theta) \\ &= Q(\Theta|\Theta^{(s)}) + \log p(\Theta) \end{aligned} \quad (2)$$

where $Q(\Theta|\Theta^{(s)}) = E_{C_{mis}|C_{obs},\Theta^{(s)}}\{\log p(C|\Theta)\}$.

2. M-step: Choose $\Theta^{(s+1)}$ to increase $J(\Theta|\Theta^{(s)})$ over its value at $\Theta = \Theta^{(s)}$. That is

$$J(\Theta^{(s+1)}|\Theta^{(s)}) \geq J(\Theta^{(s)}|\Theta^{(s)}) \quad (3)$$

holds.

The E-step and M-step iterate until certain stop criterion is met.

3. MATHEMATICAL FORMULATION FOR LPV MODELING WITH MISSING OUTPUT

Many industrial processes are often operated along certain operating trajectory to meet different production objectives (Deng and Huang, 2012). The operating trajectory is composed of several typical working points of the process with smooth transition period between different working points (Xu et al., 2009). Obviously, one local linearized model fails to capture the global dynamic behaviors of the process under multiple working conditions. The generalized EM algorithm is employed in this section to identify a multiple-mode LPV model for the process. To handle the missing output in the generalized EM algorithm, each local model is taken to have an FIR model structure which can be described by:

$$Y_{tm} = \phi_t^T \theta_m + e(t) \quad (4)$$

where Y_{tm} is the output of the m th local model at time instant t , $\phi_t^T = [U_{t-1} \cdots U_{t-n} \ 1]$ is the model regressor, θ_m is the parameter vector of the m th local model, and $e(t)$ is zero mean Gaussian noise with variance σ^2 . Then the prediction \hat{Y}_t of the process output Y_t can be expressed as:

$$\hat{Y}_t = \sum_{m=1}^M \mu_{tm} \hat{Y}_{tm} \quad (5)$$

where $\hat{Y}_{tm} = \phi_t^T \theta_m$ is the prediction from the m th local model at time instant t and μ_{tm} is the corresponding weight, and M is the number of local models. Based on the assumption of smooth transition, the normalized exponential functions are chosen as the weighting functions for model interpolation. The exponential function has the following form:

$$\omega_{tm} = \exp\left(-\frac{(H_t - H_m)^2}{2(o_m)^2}\right) \quad (6)$$

and then the normalized exponential function can be expressed as:

$$\mu_{tm} = \frac{\omega_{tm}}{\sum_{m=1}^M \omega_{tm}} \quad (7)$$

where H_t is the measurable scheduling variable, H_m is the m th pre-specified working point, and o_m represents the validity width of the m th local model. Therefore, the LPV model identification problem is transformed into estimating unknown parameters, $\Theta \triangleq \{\theta_m, o_m, \sigma\}_{m=1, 2, \dots, M}$, by using the generalized EM algorithm.

As mentioned in Section 1, the high order FIR model may result in ill-conditioned matrices problem and the parameters of the FIR model are sensitive to the noise. The prior information of the parameters can be utilized to handle these problems. Since the first n parameters of the local FIR model (4) are the impulse response coefficients of the local process and the impulse response of the local process tends to vary smoothly, this prior information can be employed to construct the prior probability of the local model parameters by considering second-order derivatives of the parameters. Here, the second-order derivative is calculated by second-order finite difference approximation. The $(n+1)$ th parameter of the local FIR model (4) is the bias term of the local model and no prior information is available for this term. Here, we assume that the second-order derivative of impulse response coefficient of the m th local process is a zero mean Gaussian white noise with variance σ_{dm}^2 (Thomassin et al., 2009) and bias term of the m th local model is a zero mean Gaussian white noise with variance σ_{bm}^2 . Define matrix D with dimension $(n+1) \times (n+1)$ as:

$$D = \begin{pmatrix} 2 & -1 & 0 & 0 & \cdots & 0 & 0 & 0 & 0 \\ -1 & 2 & -1 & 0 & \cdots & 0 & 0 & 0 & 0 \\ 0 & -1 & 2 & -1 & \cdots & 0 & 0 & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & 0 & \cdots & -1 & 2 & -1 & 0 \\ 0 & 0 & 0 & 0 & \cdots & 0 & -1 & 2 & 0 \\ 0 & 0 & 0 & 0 & \cdots & 0 & 0 & 0 & 1 \end{pmatrix}$$

Then we have $\theta_{Dm} = D\theta_m$ that follows a zero mean Gaussian distribution with covariance matrix Σ_m as

$$\Sigma_m = \begin{pmatrix} \sigma_{dm}^2 I_{n \times n} & 0 \\ 0 & \sigma_{bm}^2 \end{pmatrix}$$

where $I_{n \times n}$ is an identity matrix with dimension $n \times n$. Therefore, the vector θ_m follows a zero mean Gaussian distribution with covariance matrix $\Sigma_m(D^T D)^{-1}$.

The value of σ_{dm}^2 can be determined by performing step test of the local process or by following the various initial guesses strategy. Since no prior information of the bias term is available, the parameter σ_{bm}^2 can be set to a large value.

Hereafter, the generalized EM algorithm is applied to solve the LPV identification problems described above. Assume we have got the data sample $\{Y_t, U_t, H_t\}_{t=1, \dots, N}$. Denoting Y as $\{Y_t\}_{t=1, \dots, N}$, U as $\{U_t\}_{t=1, \dots, N}$ and H as $\{H_t\}_{t=1, \dots, N}$. Since part of the output data are randomly missing, dataset Y can be divided into missing output $Y_{mis} = \{Y_t\}_{t=m_1, \dots, m_\alpha}$ and observed output $Y_{obs} = \{Y_t\}_{t=t_1, \dots, t_\beta}$. Here, we introduce a hidden variable $I = \{I_t\}_{t=1, \dots, N}$ in which I_t denotes the model identity of Y_t . Then the observed data set and the missing data set can be defined as $C_{obs} = \{Y_{obs}, U, H\}$ and $C_{mis} = \{Y_{mis}, I\}$, respectively.

Using the Bayesian theory, the log likelihood function of the complete data set, $\log p(Y, U, H, I|\Theta)$, can be decomposed into:

$$\begin{aligned}
 & \log p(Y, U, H, I|\Theta) \\
 &= \log p(Y|U, H, I, \Theta) + \log p(I|U, H, \Theta) + \log p(H, U|\Theta) \\
 &= \log \prod_{t=1}^N p(Y_t|Y_{t-1}, \dots, Y_1, U_t, \dots, U_1, I_t, \dots, I_1, \Theta) \\
 &+ \log \prod_{t=1}^N p(I_t|I_{t-1}, \dots, I_1, U_t, \dots, U_1, H_t, \dots, H_1, \Theta) \\
 &+ \log p(H, U|\Theta) \tag{8}
 \end{aligned}$$

According to (4) and (5), the Y_t depends only on the previous input $\{U_{t-1}, \dots, U_1\}$, the model identity I_t , and the parameters Θ . To simplify the expression, we define $Z_{t-1} = \{U_{t-1}, \dots, U_1\}$. The model identity of each data point depends only on the measured scheduling variable at the same time instant. Since H and U are measurable process data and they are independent of the parameters Θ , $p(H, U|\Theta)$ is a constant which can be denoted as C_1 . Therefore, the log likelihood function of the complete data (9) can be further written as:

$$\begin{aligned}
 & \log p(Y, U, H, I|\Theta) \\
 &= \log \prod_{t=1}^N p(Y_t|Z_{t-1}, I_t, \Theta) + \log \prod_{t=1}^N p(I_t|H_t, \Theta) \\
 &+ \log C_1 \\
 &= \sum_{t=1}^N \log p(Y_t|Z_{t-1}, I_t, \Theta) + \sum_{t=1}^N \log p(I_t|H_t, \Theta) \\
 &+ \log C_1 \tag{9}
 \end{aligned}$$

where $C_1 = p(H, U|\Theta)$.

The log prior density $\log p(\Theta)$ can be decomposed into:

$$\log p(\Theta) = \sum_{m=1}^M \log p(\theta_m) + \sum_{m=1}^M \log p(o_m) + \log p(\sigma) \tag{10}$$

Since no prior information of the σ is available, we can simply set $p(\sigma)$ as a constant (e.g. uniformly distributed). We assume the validity width o_m is uniformly distributed in the range $[o_{min}, o_{max}]$. Therefore, the last two terms in (11) will not play a role in the following derivation and then the summation of these two terms can be denoted as C_2 . Then the log prior density (11) can be reduced to:

$$\log p(\Theta) = \sum_{m=1}^M \log p(\theta_m) + C_2 \tag{11}$$

Therefore, the conditional expectation of the log complete data posterior density $J(\Theta|\Theta^{(s)})$ in (2) can be written as:

$$\begin{aligned}
 J(\Theta|\Theta^{(s)}) &= E_{C_{mis}|C_{obs}, \Theta^{(s)}} \{\log p(Y, U, H, I|\Theta)\} + \log p(\Theta) \\
 &= E_{Y_{mis}, I|C_{obs}, \Theta^{(s)}} \left\{ \sum_{t=1}^N \log p(Y_t|Z_{t-1}, I_t, \Theta) \right. \\
 &+ \sum_{t=1}^N \log p(I_t|H_t, \Theta) + \log C_1 \left. \right\} \\
 &+ \sum_{m=1}^M \log p(\theta_m) + C_2 \tag{12}
 \end{aligned}$$

The expectation is taken over $J(\Theta|\Theta^{(s)})$ with respect to Y_{mis} and I , then

$$\begin{aligned}
 J(\Theta|\Theta^{(s)}) &= E_{Y_{mis}, I|C_{obs}, \Theta^{(s)}} \left\{ \sum_{t=1}^N \log p(Y_t|Z_{t-1}, I_t, \Theta) \right. \\
 &+ \sum_{t=1}^N \log p(I_t|H_t, \Theta) + \log C_1 \left. \right\} \\
 &+ \sum_{m=1}^M \log p(\theta_m) + C_2 \\
 &= \sum_{t=m_1}^{m_\alpha} \sum_{m=1}^M p(I_t = m|C_{obs}, \Theta^{(s)}) \\
 &\int p(Y_t|C_{obs}, \theta_m^{(s)}) \log p(Y_t|Z_{t-1}, I_t = m, \Theta) dY_t \\
 &+ \sum_{t=t_1}^{t_\beta} \sum_{m=1}^M p(I_t = m|C_{obs}, \Theta^{(s)}) \log p(Y_t|Z_{t-1}, I_t = m, \Theta) \\
 &+ \sum_{t=1}^N \sum_{m=1}^M p(I_t = m|C_{obs}, \Theta^{(s)}) \log p(I_t = m|H_t, \Theta) \\
 &+ \sum_{m=1}^M \log p(\theta_m) + \log C_1 + C_2 \tag{13}
 \end{aligned}$$

In order to calculate $J(\Theta|\Theta^{(s)})$, the unknown terms in (13) should be calculated first.

$$\begin{aligned}
 & \int p(Y_t|C_{obs}, \theta_m^{(s)}) \log p(Y_t|Z_{t-1}, I_t = m, \Theta) dY_t \\
 &= -\frac{1}{2} \log(2\pi\sigma^2) - \frac{1}{2\sigma^2} \int p(Y_t|C_{obs}, \theta_m^{(s)}) (Y_t - \phi_t^T \theta_m)^2 dY_t \\
 &= -\frac{1}{2} \log(2\pi\sigma^2) - \frac{1}{2\sigma^2} ((\sigma^{(s)})^2 + (\phi_t^T \theta_m^{(s)})^2) \\
 &+ \frac{1}{\sigma^2} (\phi_t^T \theta_m) (\phi_t^T \theta_m^{(s)}) - \frac{1}{2\sigma^2} (\phi_t^T \theta_m)^2 \tag{14}
 \end{aligned}$$

$$\begin{aligned}
 p(I_t = m|C_{obs}, \Theta^{(s)}) &= p(I_t = m|H_t, \Theta^{(s)}) \\
 &= \frac{\exp\left(-\frac{(H_t - H_m)^2}{2(o_m^{(s)})^2}\right)}{\sum_{m=1}^M \exp\left(-\frac{(H_t - H_m)^2}{2(o_m^{(s)})^2}\right)} \\
 &= \mu_{tm}^{(s)} \tag{15}
 \end{aligned}$$

Therefore, the $J(\Theta|\Theta^{(s)})$ can be further written as:

$$\begin{aligned}
 J(\Theta|\Theta^{(s)}) &= \sum_{t=m_1}^{m_\alpha} \sum_{m=1}^M \mu_{tm}^{(s)} \left\{ -\frac{1}{2} \log(2\pi\sigma^2) - \frac{1}{2\sigma^2} ((\sigma^{(s)})^2 \right. \\
 &+ (\phi_t^T \theta_m - \phi_t^T \theta_m^{(s)})^2) \left. \right\} \\
 &+ \sum_{t=t_1}^{t_\beta} \sum_{m=1}^M \mu_{tm}^{(s)} \log p(Y_t|Z_{t-1}, I_t = m, \Theta) \\
 &+ \sum_{t=1}^N \sum_{m=1}^M \mu_{tm}^{(s)} \log p(I_t = m|H_t, \Theta) \\
 &+ \sum_{m=1}^M \log p(\theta_m) + \log C_1 + C_2 \tag{16}
 \end{aligned}$$

In order to calculate the parameter estimates, the gradients should be taken over $J(\Theta|\Theta^{(s)})$ with respect to the unknown parameters θ_m and σ^2 .

Taking the gradient of $J(\Theta|\Theta^{(s)})$ with respect to θ_m with fixed $(\sigma^{(s)})^2$ and setting it to zeros, the parameter estimate of θ_m can be derived as:

$$\hat{\theta}_m = \frac{\sum_{t=m_1}^{m_\alpha} \mu_{tm}^{(s)} \phi_t^T \theta_m^{(s)} + \sum_{t=t_1}^{t_\beta} \mu_{tm}^{(s)} \phi_t^T y_t}{\sum_{t=1}^N \mu_{tm}^{(s)} \phi_t^T \phi_t^T + (\sigma^{(s)})^2 \Sigma_m^{-1} D^T D} \quad (17)$$

To derive the estimate of noise variance σ^2 , the gradient of $J(\Theta|\Theta^{(s)})$ is taken over σ^2 with fixed $\hat{\theta}_m$ and then set to zero.

$$\begin{aligned} \hat{\sigma}^2 &= \frac{\sum_{t=m_1}^{m_\alpha} \sum_{m=1}^M \mu_{tm}^{(s)} ((\phi_t^T \theta_m^{(s)} - \phi_t^T \hat{\theta}_m)^2 + (\sigma^{(s)})^2)}{\sum_{t=1}^N \sum_{m=1}^M \mu_{tm}^{(s)}} \\ &+ \frac{\sum_{t=t_1}^{t_\beta} \sum_{m=1}^M \mu_{tm}^{(s)} (y_t - \phi_t^T \hat{\theta}_m)^2}{\sum_{t=1}^N \sum_{m=1}^M \mu_{tm}^{(s)}} \\ &= \frac{\sum_{t=m_1}^{m_\alpha} \sum_{m=1}^M \mu_{tm}^{(s)} ((\phi_t^T \theta_m^{(s)} - \phi_t^T \hat{\theta}_m)^2 + (\sigma^{(s)})^2)}{N} \\ &+ \frac{\sum_{t=t_1}^{t_\beta} \sum_{m=1}^M \mu_{tm}^{(s)} (y_t - \phi_t^T \hat{\theta}_m)^2}{N} \end{aligned} \quad (18)$$

In order to derive the estimates of the $\{o_m\}_{m=1, \dots, M}$, a nonlinear optimization problem is formulated as follows:

$$\begin{aligned} \max_{o_m, m=1, \dots, M} \sum_{t=1}^N \sum_{m=1}^M \mu_{tm}^{(s)} \log p(I_t = m | H_t, \Theta) \\ \text{s.t. } o_{min} \leq o_m, m = 1, \dots, M \leq o_{max} \end{aligned} \quad (19)$$

The constrained nonlinear optimization function ‘fmincon’ provided by Matlab software can be employed to solve this problem (Jin et al., 2011).

The parameters $\{o_m\}_{m=1, \dots, M}$, $\{\theta_m\}_{m=1, \dots, M}$, and σ^2 should be updated in each iteration until the convergence condition of the generalized EM algorithm is met.

4. SIMULATION EXAMPLES

Consider an LPV process described by the following LPV model (Deng and Huang, 2012):

$$G(s, H) = \frac{K(H)}{\tau(H)s + 1} \quad (20)$$

where

$$K(H) = 0.6 + H^2, \quad \tau(H) = 3 + 0.5H^3, \quad H \in [1, 4] \quad (21)$$

Three working points, $H_1 = 1$, $H_2 = 2.25$, $H_3 = 4$, are selected and the process is tested at these three working points. For the transition periods between neighboring working points, the scheduling variable H is linearly increased with a fixed small interval. The sampling period is 1s and 1500 samples are recorded to construct the training set. The measurement white noise $e(t)$ with zero mean and variance 0.1 is added to the output data. The input data, output data and the scheduling variable data are shown in Fig. 1. In the simulation, 25% output data are randomly missing. The order of the local FIR model is set to 45 and all the unknown parameters of the local FIR model are initialized to 0.2 which is selected arbitrarily. For comparison, the simulation is firstly performed to calculate the ML estimates of local model parameters without using prior information. Since the estimated FIR coefficients are very noisy, they are not shown here. Then, the proposed algorithm is used to estimate a multiple-model LPV model

for this process. σ_{d1}^2 , σ_{d2}^2 , and σ_{d3}^2 are set to 0.0000813, 0.0001, and 0.00006, respectively. $\{\sigma_{bm}^2\}_{m=1,2,3}$ are set to 0.0002. The normalized weight of each local model is shown in Fig. 2. The estimated FIR coefficients are presented in Fig. 3. Self-validation and cross-validation are performed to verify the accuracy of the global LPV model and the results are shown in Fig. 4 and Fig. 5, respectively. In the cross-validation, the process is tested at three other working points which are $H_1 = 1.5$, $H_2 = 2.7$, $H_3 = 3.4$. It can be seen from these figures that the estimated LPV model can capture the dynamics of the process accurately.

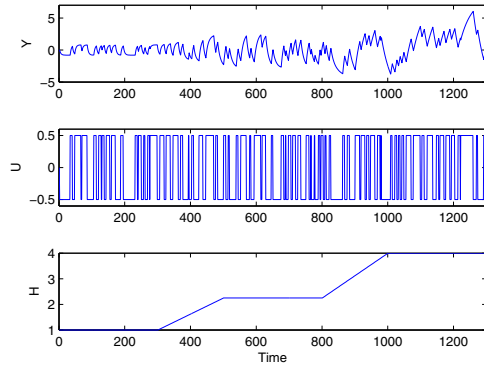


Fig. 1. The training set data.

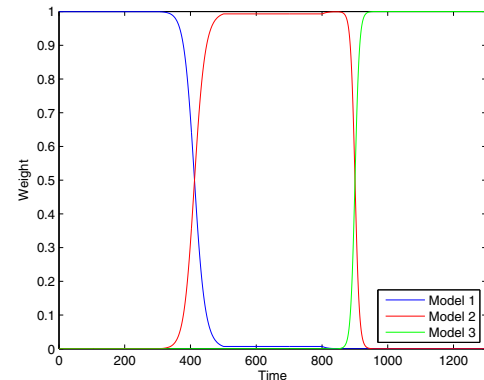


Fig. 2. The normalized weight of each local model.

5. CONCLUSION

This paper considered the modeling problems of an LPV system with missing data. Since many industrial processes are designed to conduct multiple production tasks, the multiple-model LPV model is considered here. To handle randomly missing data in output, the local model is assumed to have an FIR model structure. To avoid the potential problems of ill-conditioned matrices and sensitivity of parameters to process noise induced by using FIR model, the prior information of the FIR coefficients is employed to construct the prior density function of the parameters. The generalized EM algorithm is then modified to derive the MAP estimate of the global LPV model parameters. The effectiveness of the proposed method is demonstrated

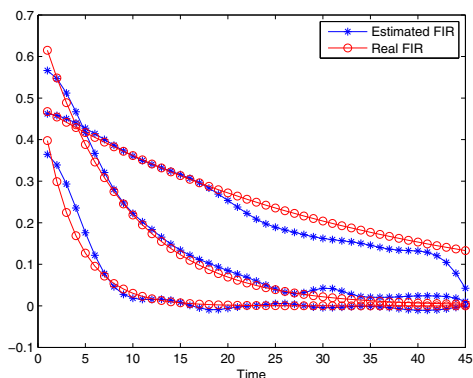


Fig. 3. The comparison of real FIR coefficients and MAP estimates of the FIR coefficients.

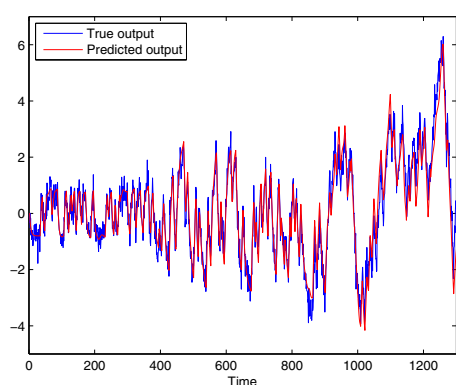


Fig. 4. The self-validation of the identified LPV model.

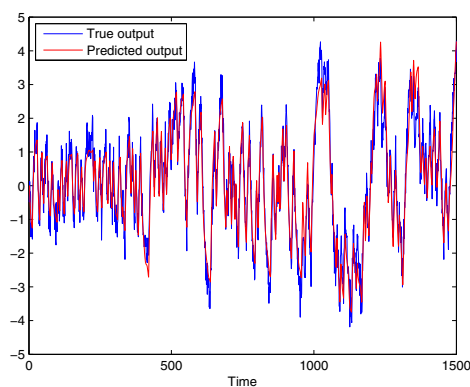


Fig. 5. The cross-validation of the identified LPV model.

through the numerical example. If the FIR model is used as an intermediate model, further work can be done to identify a general model, such as the Output Error (OE) model, based on the estimated FIR model.

REFERENCES

Bamieh, B. and Giarre, L. (2002). Identification of linear parameter varying models. *International Journal of Robust and Nonlinear Control*, 12(9), 841–853.

Dempster, A., Laird, N., and Rubin, D. (1977). Maximum likelihood from incomplete data via the em algorithm. *Journal of the Royal statistical Society*, 39(1), 1–38.

Deng, J. and Huang, B. (2012). Identification of nonlinear parameter varying systems with missing output data. *AIChE Journal*, 58, 3454–3467.

Gopaluni, R. (2008). A particle filter approach to identification of nonlinear process under missing observations. *The Canadian Journal of Chemical Engineering*, 86, 1081–1092.

Jin, X., Huang, B., and Shook, D. (2011). Multiple model lpv approach to nonlinear process identification with em algorithm. *Journal of Process Control*, 21, 182–193.

Jin, X., Wang, S., Huang, B., and Forbes, F. (2012). Multiple model based lpv soft sensor development with irregular/missing process output measurement. *Control Engineering Practice*, 20, 165–172.

Khatibisepehr, S. and Huang, B. (2008). Dealing with irregular data in soft sensors: bayesian method and comparative study. *Industrial & Engineering Chemistry Research*, 47, 8713–8723.

Laurain, V., Gilson, M., Tóth, R., and Garnier, H. (2010). Refined instrumental variable methods for identification of lpv box–jenkins models. *Automatica*, 46(6), 959–967.

McLachlan, G. and Krishnan, T. (2007). *The EM Algorithm and Extensions*. John Wiley & Sons.

Thomassin, M., Bastogne, T., and Richard, A. (2009). Identification of a managed river reach by a bayesian approach. *IEEE Transactions on Control Systems Technology*, 17, 353–365.

Tóth, R., Laurain, V., Gilson, M., and Garnier, H. (2012). Instrumental variable scheme for closed-loop lpv model identification. *Automatica*, 48(9), 2314–2320.

Wassink, M., van de Wal, M., Scherer, C., and Bosgra, O. (2005). Lpv control for a wafer stage: beyond the theoretical solution. *Control Engineering Practice*, 13, 231–245.

Xu, Z., Zhao, J., Qian, J., and Zhu, Y. (2009). Nonlinear mpc using identified lpv model. *Industrial & Engineering Chemistry Research*, 48, 3043–3051.

Yin, S., Ding, S., Haghani, A., Hao, H., and Zhang, P. (2012). A comparison study of basic data-driven fault diagnosis and process monitoring methods on the benchmark tennessee eastman process. *Journal of Process Control*, 22(9), 1567–1581.

Yin, S., Luo, H., and Ding, S. (2014). Real-time implementation of fault-tolerant control systems with performance optimization. *IEEE Transactions on Industrial Electronics*, 61(5), 2402–2411.

Zhao, Y., Huang, B., Su, H., and Chu, J. (2012). Prediction error method for identification of lpv models. *Control Engineering Practice*, 22, 180–193.