Operating Condition Diagnosis Based on HMM with Adaptive Transition Probabilities in Presence of Missing Observations

Nima Sammaknejad and Biao Huang
Dept. of Chemical and Materials Engineering, University of Alberta, Edmonton, Alberta T6G 2V4, Canada

Weili Xiong
Key Laboratory of Advanced Process Control for Light Industry (Ministry of Education), Jiangnan University, Wuxi, Jiangsu 214122, China

Alireza Fatehi
Faculty of Electrical Engineering, K.N. Toosi University of Technology, Tehran, Iran

Fangwei Xu and Aris Espejo
Syncrude Canada Ltd., Fort McMurray, Alberta, Canada T9H 3L1

DOI 10.1002/aic.14661
Published online November 5, 2014 in Wiley Online Library (wileyonlinelibrary.com)

A new approach for modeling and monitoring of the multivariate processes in presence of faulty and missing observations is introduced. It is assumed that operating modes of the process can transit to each other following a Markov chain model. Transition probabilities of the Markov chain are time varying as a function of the scheduling variable. Therefore, the transition probabilities will be able to vary adaptively according to different operating modes. In order to handle the problem of missing observations and unknown operating regimes, the expectation maximization algorithm is used to estimate the parameters. The proposed method is tested on two simulations and one industrial case studies. The industrial case study is the abnormal operating condition diagnosis in the primary separation vessel of oil-sand processes. In comparison to the conventional methods, the proposed method shows superior performance in detection of different operating conditions of the process. © 2014 American Institute of Chemical Engineers *AIChE J*, 61: 477–493, 2015
Keywords: fault diagnosis, hidden Markov models, adaptive transition probabilities, expectation maximization algorithm

Introduction

Regime switching systems have been of a great interest in the field of economics since 1990. It starts with Hamilton's study in 1988 where a nonlinear filter was developed to learn about the changes in regime and find the maximum likelihood estimation of the parameters.[1] Later, he developed an expectation maximization (EM) algorithm framework to find the maximum likelihood estimation of the process parameters with discrete shifts.[2] The shifts were modeled as the outcome of a Markov process. The goal of this study was to find the main changes of the asset prices from observable events.

Since then, applications of switching Markov regimes have been widely used in the field of economics. Bollen et al. studied applications of the regime-switching models to analyze the dynamic behavior of foreign exchange rates. They found that the prices do not obey a Markov regime switching behavior.[3] Ang et al. performed a similar study in interest rates of United States, Germany, and the United Kingdom and concluded that regime switching models have better forecasts in comparison to single-regime models.[4] Pelettier introduced a regime switching dynamic correlation model for the variance between different time series. It is shown in their empirical case study that the developed model has a better performance in comparison to the previous studies.[5] In the same year, Mount et al. showed that a stochastic regime switching system is able to model the behavior of wholesale electricity prices and the price spikes.[6]

The idea of considering time-varying transition probabilities in regime switching systems was initially proposed by Diebold et al.[7] They developed an EM algorithm framework for parameter estimation in cases where the transition probabilities are function of underlying economic fundamentals. However, there are several limitations in their study. First, it is assumed that the scheduling variable (economic fundamentals) can only accept some limited discrete values. In other words, the transition behavior of the scheduling variable between different operating modes, which is an important factor in industrial processes, is not considered. Second, they solve the problem considering only two possible hidden modes for the process. Finally, their proposed optimization procedure for the nonlinear terms is very dependent to the initial values. Filardo et al. made a clear picture of the advantages in considering time varying transition probabilities over fixed transition probabilities afterwards.[8] Later,

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

Otranto considered a specific multichain Markov switching model where the transition probabilities are dependent to the regime of other variables. This approach was successful in predicting the regime of the analyzed variables.[9]

Although hidden Markov models (HMMs) and regime switching systems have been widely studied in the field of economics, their applications in the field of system identification and fault detection are sparse. Discrete-time Markov jump linear systems and sudden failures have been previously reviewed in literature.[10] Jin et al. developed a solution strategy for identification of switched Markov autoregressive exogenous systems under the EM framework.[11] Ghasemi et al. proposed a parameter estimation method for a condition monitoring equipment with a certain failure rate structure. Hidden states (modes) were assumed to transit following a hidden Markov model and observations were assumed to be imprecise. They used a maximum likelihood estimation framework to obtain the parameters.[12] Wong et al. proposed hidden Markov models as a generalization to the mixture of Gaussian approach in order to model the faults due to sensor malfunctions. They showed that these types of faults can be detected more appropriately using a HMM-based model.[13] Jiang et al. proposed a new method for fault detection of gear transition system. They modeled the system behavior as a three state continuous time Markov process. Parameter estimation was based on the EM algorithm framework.[14] The detailed proof of their mathematical derivations is available in their article.[15]

There has also been a great effort in handling the problem of missing data in recent years. Deng et al. studied identification of nonlinear parameter varying systems with missing output data using particle filter. The model is appropriate for the processes which work in multiple operating conditions.[16] Different approaches to handle the problem of missing data using the EM algorithm is discussed in detail in literature.[17] Multivariate process monitoring methods have also been broadly reviewed in literature. Keshavarz et al. compared the application of Bayesian and EM methods in multivariate change point detection.[18]

In this article, we propose a new modeling and monitoring strategy for multivariate processes which follow a Markov regime switching behavior with time varying transition probabilities in the presence of missing and faulty observations. Since the scheduling variable is usually a good indication to the current operating mode of the process, transition probabilities are considered to be time-varying as a function of the scheduling variable. Therefore, transitions of the process between different operating modes are taken into account by defining the transition probabilities as distributions which are function of the underlying scheduling variable. In comparison to conventional HMMs, this structure shows a far better performance for the processes which have an asymmetric time-varying transition behavior between different operating modes, that is, when some of the modes are far from the majority and the scheduling variable provides more flexibility in the modeling and filtering steps. Furthermore, a certain structure is considered for the operating modes in the transition probability matrix as the operating modes can transit to each other only in a logical manner, for example, normal, abnormal, and then the faulty mode. This structure will reduce the computational cost and provide an appropriate framework for the industrial processes with continuous transitions from the normal to faulty modes.

Since industrial data are usually subject to missing observations and unknown operating regimes, the problem is solved under the EM algorithm framework. In the maximization step (M-step) of the EM algorithm, the nonlinear interior point local optimization algorithm is adopted to find the optimal value of some of the unknown parameters numerically. Since local numerical nonlinear optimization methods are usually sensitive to the initial guess, the initial values of the local optimization problem for the first iteration of the EM algorithm are obtained from an optimization based on the first few generations of the Genetic Algorithm (GA). Other initial values for the unknown parameters in the EM algorithm are obtained by assuming that the observations follow a mixture of multivariate Gaussian distributions.

After the parameter estimation step, Hamilton's filter is applied to infer the hidden operating mode of the process for the test dataset.[2] The accuracy of the algorithm is tested on both simulation and industrial case studies and compared to conventional HMMs. The industrial case study is the abnormal operating condition diagnosis in the primary separation vessel (PSV) which is an important early separation step in oil-sands processes. The method shows satisfactory predictions in recognition of different operating conditions of the process.

The remainder of this article is organized as follows: Problem Statement is the section where the model structure and unknown parameters are introduced. Parameter Estimation based on the Expectation Maximization Algorithm section reviews the steps of the EM algorithm for parameter estimation. Operating Mode Recognition section is the application of Hamilton's filter to infer the hidden operating mode of the process. Results and Discussion section includes the results of the simulation and industrial case studies and Conclusion section concludes the article.

Problem Statement

The dataset for the process variables are considered as follows

$$Y=\begin{pmatrix} y_{11} & y_{12} & \cdots & y_{1N} \\ y_{21} & y_{22} & \cdots & y_{2N} \\ \vdots & \vdots & \vdots & \vdots \\ y_{P1} & y_{P2} & \cdots & y_{PN} \end{pmatrix}=(Y_1, Y_2, \ldots, Y_N)$$

where P is the number of process variables and N is the number of sample times.

The observable data (C_{obs}) include $Y_O=\{Y_{t_1}, \ldots, Y_{t_\alpha}\}$, and $H=\{H_1, \ldots, H_N\}$, where H is the scheduling variable. The missing data (C_{miss}) include $Y_M=\{Y_{m_1}, \ldots, Y_{m_\beta}\}$, and the hidden operating modes of the process at different sample times, that is, $I=\{I_1, \ldots, I_N\}$. This hidden operating mode corresponds to the operating condition of the process, for example, normal, abnormal, and fault.

One could easily compare this terminology with the corresponding terms in a state-space model. The introduced operating mode (operating condition) in this article, corresponds to the states of a state-space model. Similar to a state-space model, Observations (Y), which correspond to the process outputs, are functions of the underlying states (Eq. 1). However, the dataset is only partially available (Y_O). The transition probability matrix (Eq. 5) corresponds to the state, or system matrix (A matrix) in the state-space representation.

It is obvious that the union of Y_O and Y_M is Y. The missing data (Y_M) might come from different sources such as computer disconnections, sensor failures, and data collection errors.[19] Three main types of missing data have been introduced in literature[20]: (1) Missing not at random (MNAR) (2) Missing at random (MAR) (3) Missing completely at random (MCAR). In the MNAR case, the probability of missingness depends on the missing data. In the MAR case, the probability of missingness does not depend on the missing data, but depends on the observed data. In the MCAR case, the probability of missingness is independent of both the missing and observed data. The missing data in the simulation case studies of this paper are assumed to be completely missed at random.

Observations at the *kth* time step are assumed to follow a multivariate normal distribution with mean vector μ_i and covariance matrix Σ_i given the operating mode i, that is

$$f(Y_k|I_k=i) \sim N_P(\mu_i, \Sigma_i) \tag{1}$$

where

$$N_P(\mu_i, \Sigma_i)=(2\pi)^{-P/2}|\Sigma_i|^{-1/2}exp\left(-\frac{1}{2}(Y_k-\mu_i)^T\Sigma_i^{-1}(Y_k-\mu_i)\right)$$

where $1 \le k \le N, 1 \le i \le M$, and M is the number of available hidden operating modes in the process.

π_i is the initial state (operating mode) distribution of the Markov chain ($\pi_i=P(I_1=i)$) and the time varying transition probabilities are defined as a function of the scheduling variable at the previous time step, that is

$$\alpha(k)_{ij}=P(I_k=j|I_{k-1}=i, H_{k-1}) \tag{2}$$

When $i=j$, $\alpha(k)_{ii}$ follows the distribution in Eq. 3. Otherwise, it follows the distribution in Eq. 4

$$\alpha(k)_{ii}=\frac{2\gamma_{ii}\exp\left(\frac{-(H_{k-1}-H_i)^2}{2\sigma_{H_i}^2}\right)}{1+\exp\left(\frac{-(H_{k-1}-H_i)^2}{2\sigma_{H_i}^2}\right)} \tag{3}$$

$$\alpha(k)_{ij,i\neq j}=\gamma_{ij}(1-\alpha(k)_{ii}) \tag{4}$$

where $1 \le k \le N, 1 \le i,j \le M, \sigma_{H_i}$ is an indicator for the validity of the scheduling variable in operating mode i, H_i is the mean value of the scheduling variable in operating mode i, and γ_{ij}'s provide more flexibility in estimation of the distribution for the transition probability $\alpha(k)_{ij}$.

Having such distributions in Eqs. 3 and 4, continuous transitions of the scheduling variable, and consequently the process, between different operating modes are taken into account. One should note that in the case when $H_{k-1}=H_i, \alpha(k)_{ij}$ turns to a constant value and the transition probability matrix behaves as a conventional Markov chain model. When H_{k-1} starts to deviate from H_i, the distribution in Eq. 3 decreases, while the distributions in Eq. 4 start to grow, that is, the probability to stay is the same mode decreases and, as a result, the probability of transition to other modes increases. This framework provides an appropriate structure for modeling and monitoring of the processes with time-varying asymmetric transitions between different operating modes, that is, when some of the operating modes are far from the majority and the scheduling variable can provide more flexibility in the modeling and filtering steps. Industrial examples of such cases are provided in the case studies of this paper. Another advantage of considering such distributions is the appearance of the linear constraints on $\gamma_{ij,i\neq j}$'s in Eqs. 6–8. Having such linear constraints, makes the optimization problem to find $\gamma_{ij,i\neq j}$ analytically tractable using Lagrange multipliers (Eqs. 27, 28).

Moreover, the $M \times M$ transition probability matrix is assumed to follow the structure in Eq. 5

$$\begin{bmatrix} \alpha(k)_{11} & \cdots & \alpha(k)_{1Q} & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots & \cdots & \vdots \\ \alpha(k)_{P1} & \cdots & \alpha(k)_{PQ} & 0 & \cdots & 0 \\ \alpha(k)_{(P+1)1} & \cdots & \cdots & \cdots & \cdots & \alpha(k)_{(P+1)M} \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ \alpha(k)_{Q1} & \cdots & \cdots & \cdots & \cdots & \alpha(k)_{QM} \\ 0 & \cdots & 0 & \alpha(k)_{(Q+1)(Q+1)} & \cdots & \alpha(k)_{(Q+1)(M)} \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & \cdots & 0 & \alpha(k)_{(M)(Q+1)} & \cdots & \alpha(k)_{(M)(M)} \end{bmatrix} \tag{5}$$

This structure imposes a logical transition between different operating modes of the process, that is, the process cannot enter the faulty modes right after leaving the normal modes, and the abnormal modes are some intermediate modes which have the capability to transit to both normal and faulty modes. Also, from the faulty modes the process can just transit to the abnormal modes. This transition behavior is graphically presented in Figure 1. This structure makes the model more appropriate for a wide class of process industry applications where the process continuously changes between different operating modes rather than sudden discrete jumps from normal to faulty. Furthermore, it reduces the number of required parameters and the computational time in both parameter estimation and online operating mode recognition (filtering) steps.

According to the structure introduced in Eq. 5, there are three types of constraints on the parameters of the different operating modes.

Constraints for the normal operating modes

$$\begin{gathered} if\, 1 \le i \le P, 1 \le j \le Q : 0 \le \gamma_{ij} \le 1, \sum_{j=1,j\neq i}^{j=Q} \gamma_{ij}=1 \\ if\, 1 \le i \le P, Q+1 \le j \le M : \gamma_{ij}=\alpha_{ij}=0 \end{gathered} \tag{6}$$

Constraints for the abnormal operating modes

$$if\, P+1 \le i \le Q, 1 \le j \le M : 0 \le \gamma_{ij} \le 1, \sum_{j=1,j\neq i}^{j=M} \gamma_{ij}=1 \tag{7}$$

Constraints for the faulty operating modes

$$\begin{gathered} if\, Q+1 \le i \le M, 1 \le j \le Q : \gamma_{ij}=\alpha_{ij}=0 \\ if\, Q+1 \le i \le M, Q+1 \le j \le M : 0 \le \gamma_{ij} \le 1, \sum_{j=Q+1,j\neq i}^{j=M} \gamma_{ij}=1 \end{gathered} \tag{8}$$

These constraints are used later in the derivations of the EM algorithm.

In summary, the unknown parameters to be estimated from the EM algorithm are the mean vectors and covariance matrices of the different modes (μ_i and Σ_i), parameters of

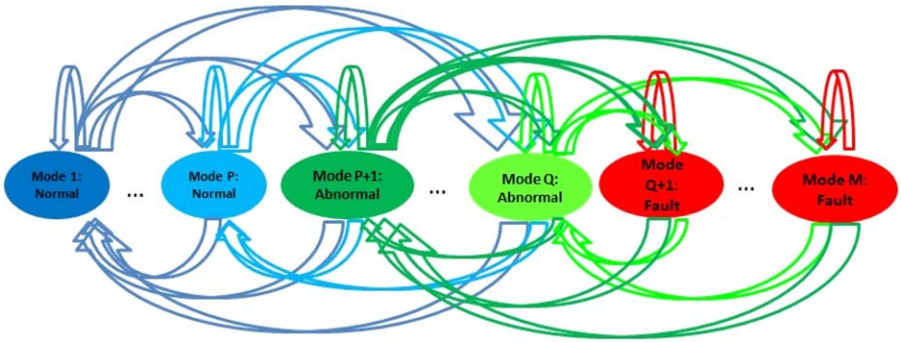

Figure 1. The diagram of operating mode transitions used for fault detection purpose of this article (normal modes: 1 to *P*, abnormal modes: *P* + 1 to *Q*, faulty modes: *Q* + 1 to *M*).

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

the transition probabilities (γ_{ii}'s and γ_{ij}'s) and the validity of the scheduling variable in different operating modes (σ_{H_i}). The mean value of the scheduling variable at different operating modes (H_i's) are assumed to be known from the historical data (find more information in Section Maximization Step).

Parameter Estimation Based on the Expectation Maximization Algorithm

EM algorithm finds the maximum likelihood estimation of the unknown parameters by iteratively switching between the expectation (E) and maximization (M) steps.[21]

In the E-step of the EM algorithm, the Q-function, which is the conditional expectation of the complete data, is calculated

$$Q(\theta|\theta^{\text{old}}) = E_{C_{\text{miss}}|(\theta^{\text{old}},C_{\text{obs}})}\{\log f(C_{\text{obs}}, C_{\text{miss}}|\theta)\} \tag{9}$$

where θ^{old} is the vector of parameters for the previous iteration, C_{miss} is the missing dataset and C_{obs} is the observed dataset.

In the M-step, the set of parameters that maximizes the Q-function are calculated

$$\theta^{\text{new}} = \underset{\theta}{\text{argmax}}\, Q(\theta|\theta^{\text{old}}) \tag{10}$$

This procedure is iteratively repeated until some stopping criterion is satisfied.

Expectation step

In the expectation step, the expected value of the complete-data log-likelihood function is calculated. Observed and missing data and the unknown parameters have been previously introduced in the problem statement section

$$\begin{aligned} Q(\theta|\theta^{\text{old}}) &= E_{C_{\text{miss}}|(\theta^{\text{old}},C_{\text{obs}})}\{\log f(C_{\text{obs}}, C_{\text{miss}}|\theta)\} \\ &= E_{I_{1:N},Y_M|(\theta^{\text{old}},C_{\text{obs}})}\{log f(Y_{1:N}, H_{1:N}, I_{1:N}|\theta)\} \end{aligned} \tag{11}$$

Using the chain rule, the probability density function in Eq. 11 can be decomposed as follows

$$\begin{aligned} &f(Y_{1:N}, H_{1:N}, I_{1:N}|\theta) \\ &= f(Y_{1:N}|H_{1:N}, I_{1:N}, \theta) P(I_{1:N}|H_{1:N}, \theta) P(H_{1:N}|\theta) \end{aligned} \tag{12}$$

Each term of Eq. 12 is explained in Eqs. 13 to 15, respectively

$$\begin{aligned} f(Y_{1:N}|H_{1:N}, I_{1:N}, \theta) &= \prod_{k=1}^{N} f(Y_k|Y_{k-1}, \ldots, Y_1, H_{1:N}, I_{1:N}, \theta) \\ &= \prod_{k=1}^{N} f(Y_k|I_k, \theta) \end{aligned} \tag{13}$$

In Eq. 13, we have used the fact that given the model identity I, the conditional distribution of Y is independent of the scheduling variable H. Furthermore, Y_k follows the multivariate normal distribution in Eq. 1 given the hidden operating mode at time k. Also

$$\begin{aligned} P(I_{1:N}|H_{1:N}, \theta) &= \prod_{k=1}^{N} P(I_k|I_{k-1}, \ldots, I_1, H_{1:N}, \theta) \\ &= P(I_1) \prod_{k=2}^{N} P(I_k|I_{k-1}, H_{k-1}, \theta) \end{aligned} \tag{14}$$

Equation 14 is derived based on the Markov property of the model.

H_i is independent of θ, and therefore, the last term in Eq. 12 can be considered as a constant in the Q function, that is

$$P(H_{1:N}|\theta) = \text{Const} \tag{15}$$

Following Eqs. 12 to 15, the Q-function in Eq. 11 can be written as

$$\begin{aligned} Q(\theta|\theta^{\text{old}}) = E_{I_{1:N},Y_M|(\theta^{\text{old}},C_{\text{obs}})} \Bigg\{ &\sum_{k=1}^{N} \log f(Y_k|I_k, \theta)) \\ &+ \sum_{k=2}^{N} \log P(I_k|I_{k-1}, H_{k-1}, \theta)) + \log P(I_1) + \log(\text{Const}) \Bigg\} \end{aligned} \tag{16}$$

In the first step, the expected value in Eq. 16 is calculated with respect to the hidden operating modes (I_k), that is

$$\begin{aligned} &Q(\theta|\theta^{\text{old}}) = \\ &E_{Y_M|(\theta^{\text{old}},C_{\text{obs}},I)} \sum_{I_1} \cdots \sum_{I_N} P(I_1, \ldots, I_N|\theta^{\text{old}}, C_{\text{obs}}) \\ &\Bigg\{ \sum_{k=1}^{N} \log f(Y_k|I_k, \theta) + \sum_{k=2}^{N} \log P(I_k|I_{k-1}, H_{k-1}, \theta) \\ &+ \log P(I_1) + \log(\text{Const}) \Bigg\} \end{aligned} \tag{17}$$

Equation 17 can be further simplified as

$$Q(\theta|\theta^{\text{old}})=$$
$$E_{Y_M|(\theta^{\text{old}},C_{\text{obs}},I)}\left\{\sum_{i=1}^{M}\sum_{k=1}^{N}P(I_k=i|\theta^{\text{old}},C_{\text{obs}})\log f(Y_k|I_k=i,\theta)\right.$$
$$+\sum_{i=1}^{M}\sum_{j=1}^{M}\sum_{k=1}^{N}P(I_k=j,I_{k-1}=i|\theta^{\text{old}},C_{\text{obs}})\log\alpha_{ij}(k) \quad (18)$$
$$\left.+\sum_{i=1}^{M}P(I_1=i|\theta^{\text{old}},C_{\text{obs}})\log\pi_i+\log\text{Const}\right\}$$

In the next step, expectation is calculated with respect to missing observations (Y_M), that is

$$Q(\theta|\theta^{\text{old}})=$$
$$\int_{Y_m}\sum_{i=1}^{M}\sum_{k=1}^{N}P(I_k=i|\theta^{\text{old}},C_{\text{obs}})\log f(Y_k|I_k=i,\theta)$$
$$\times P(Y_{m_1:m_\beta}|\theta^{\text{old}},C_{\text{obs}},I)dY_{m_1:m_\beta}$$
$$+\int_{Y_m}\sum_{i=1}^{M}\sum_{j=1}^{M}\sum_{k=1}^{N}P(I_k=j,I_{k-1}=i|\theta^{\text{old}},C_{\text{obs}})\log\alpha_{ij}(k) \quad (19)$$
$$\times P(Y_{m_1:m_\beta}|\theta^{\text{old}},C_{\text{obs}},I)dY_{m_1:m_\beta}$$
$$+\int_{Y_m}\sum_{i=1}^{M}P(I_1=i|\theta^{\text{old}},C_{\text{obs}})\log\pi_i$$
$$\times P(Y_{m_1:m_\beta}|\theta^{\text{old}},C_{\text{obs}},I)dY_{m_1:m_\beta}+\log\text{Const}$$

Since the integration is with respect to the missing observations, Eq. 19 can be simplified as

$$Q(\theta|\theta^{\text{old}})=$$
$$\sum_{i=1}^{M}\sum_{k=t_1}^{t_\alpha}\log f(Y_k|I_k=i,\theta)P(I_k=i|\theta^{\text{old}},C_{\text{obs}})$$
$$+\sum_{i=1}^{M}\sum_{k=m_1}^{m_\beta}P(I_k=i|\theta^{\text{old}},C_{\text{obs}})$$
$$\times\int P(Y_k|\theta^{\text{old}},C_{\text{obs}},I_k=i)\log f(Y_k|I_k=i,\theta)dY_k \quad (20)$$
$$+\sum_{i=1}^{M}\sum_{j=1}^{M}\sum_{k=1}^{N}P(I_k=j,I_{k-1}=i|\theta^{\text{old}},C_{\text{obs}})\log\alpha_{ij}(k)$$
$$+\sum_{i=1}^{M}P(I_1=i|\theta^{\text{old}},C_{\text{obs}})\log\pi_i$$
$$+\log\text{Const}$$

Assuming that Y_k follows the multivariate Gaussian distribution in Eq. 2 given the hidden operating mode at time k, and using the properties of the expected value of the quadratic form, the integral term in Eq. 20 can be derived as

$$\int P(Y_k|\theta^{\text{old}},C_{\text{obs}},I_k=i)\log f(Y_k|I_k=i,\mu_i,\Sigma_i)dY_k$$
$$=-\frac{1}{2}log((2\pi)^P|\Sigma_i|)-\frac{1}{2}(tr(\Sigma_i^{-1}\Sigma_i^{\text{old}}) \quad (21)$$
$$+(\mu_i^{\text{old}}-\mu_i)^T\Sigma_i^{-1}(\mu_i^{\text{old}}-\mu_i))$$

Details of the derivation are provided in Appendix. Finally, the Q-function is written as,

$$Q(\theta|\theta^{\text{old}})=\sum_{i=1}^{M}\sum_{t_1}^{t_\alpha}\log f(Y_k|I_k=i,\theta)P(I_k=i|\theta^{\text{old}},C_{\text{obs}})$$
$$+\sum_{i=1}^{M}\sum_{k=m_1}^{m_\beta}P(I_k=i|\theta^{\text{old}},C_{\text{obs}})$$
$$\times\left(-\frac{1}{2}\log((2\pi)^P|\Sigma_i|)-\frac{1}{2}(tr(\Sigma_i^{-1}\Sigma_i^{\text{old}})\right.$$
$$\left.+(\mu_i^{\text{old}}-\mu_i)^T\Sigma_i^{-1}(\mu_i^{\text{old}}-\mu_i))\right) \quad (22)$$
$$+\sum_{i=1}^{M}\sum_{j=1}^{M}\sum_{k=1}^{N}P(I_k=j,I_{k-1}=i|\theta^{\text{old}},C_{\text{obs}})\log\alpha_{ij}(k)$$
$$+\sum_{i=1}^{M}P(I_1=i|\theta^{\text{old}},C_{\text{obs}})\log\pi_i$$
$$+\log\text{Const}$$

Maximization step

In the maximization step, derivatives of the Q-function are taken with respect to the unknown parameters and then set to zero. For the parameters without an analytical solution, the optimal value of the parameters are found following a numerical optimization procedure.

In order to find the optimal mean vector of the different operating modes, derivatives of the first two terms in Eq. 22 are taken with respect to μ_i and then set to zero. Using the derivative properties of the vectors, the final result for the mean vector of each mode is obtained as in Eq. 23. More details are discussed in the Appendix.

$$\mu_i^{\text{new}}=\frac{\sum_{k=t_1}^{t_\alpha}Y_kP(I_k=i|\theta^{\text{old}},C_{\text{obs}})+\sum_{k=m_1}^{m_\beta}\mu_i^{\text{old}}P(I_k=i|\theta^{\text{old}},C_{\text{obs}})}{\sum_{k=1}^{N}P(I_k=i|\theta^{\text{old}},C_{\text{obs}})} \quad (23)$$

Derivatives with respect to the covariance matrix of each mode are taken and set to zero in a similar manner. The final result is presented in Eq. 24. Details of the derivations are provided in the Appendix.

$$(\Sigma_i)^{\text{new}}=\frac{\sum_{k=t_1}^{t_\alpha}(Y_k-\mu_i^{\text{new}})(Y_k-\mu_i^{\text{new}})^TP(I_k=i|\theta^{\text{old}},C_{\text{obs}})}{\sum_{k=1}^{N}P(I_k=i|\theta^{\text{old}},C_{\text{obs}})}$$
$$+\frac{\sum_{k=m_1}^{m_\beta}((\Sigma_i^{\text{old}})+(\mu_i^{\text{old}}-\mu_i^{\text{new}})(\mu_i^{\text{old}}-\mu_i^{\text{new}})^T)P(I_k=i|\theta^{\text{old}},C_{\text{obs}})}{\sum_{k=1}^{N}P(I_k=i|\theta^{\text{old}},C_{\text{obs}})} \quad (24)$$

The optimization problem to find π_i is constrained by $\sum_{i=1}^{M}\pi_i=1$ and as a results the Lagrange multiplier λ is introduced

$$\pi_i^{\text{new}}=\underset{\pi_i}{\text{argmax}}\left\{\sum_{i=1}^{M}P(I_1=i|\theta^{\text{old}},C_{\text{obs}})\log\pi_i+\lambda\left(\sum_{i=1}^{M}\pi_i-1\right)\right\} \quad (25)$$

Taking the derivative of Eq. 25 with respect to the Lagrange multiplier and π_i, and solving the set of linear equations, the expression in Eq. 26 is obtained

$$\pi_i^{\text{new}}=P(I_1=i|\theta^{\text{old}},C_{\text{obs}}) \quad (26)$$

Similarly, the optimization problem to find $\gamma_{ij,i\neq j}$ is constrained by $\sum_{j=1}^{M}\gamma_{ij,i\neq j}=1$. Therefore, Lagrange multiplier λ' is introduced

$$\gamma_{ij,i\neq j}^{\text{new}} = \underset{\gamma_{ij,i\neq j}}{\text{argmax}} \left\{ \sum_{i=1}^{M}\sum_{j\neq i=1}^{M}\sum_{k=2}^{N} P(I_k=j, I_{k-1}=i|\theta^{\text{old}}, C_{\text{obs}}) \times \log(\alpha_{ij}(k)) + \lambda' \left(\sum_{j\neq i=1}^{M} \gamma_{ij} - 1 \right) \right\} \tag{27}$$

where $\alpha_{ij}(k)$ is introduced as a function of γ_{ij} in Eqs. 3 and 4. Taking the derivative of Eq. 27 with respect to $\gamma_{ij,i\neq j}$ and the Lagrange multiplier and then solving the set of linear equations, the following result is obtained

$$(\gamma_{ij,i\neq j})^{\text{new}} = \frac{\sum_{k=2}^{N} P(I_k=j, I_{k-1}=i|\theta^{\text{old}}, C_{\text{obs}})}{\sum_{j\neq i=1}^{M}\sum_{k=2}^{N} P(I_k=j, I_{k-1}=i|\theta^{\text{old}}, C_{\text{obs}})} \tag{28}$$

Due to the existence of the exponential function in transition probabilities ($\alpha_{ii}(k)$), the unknown parameters σ_{H_i} and γ_{ii} cannot be obtained analytically when maximizing the cost function in Eq. 29

$$(\sigma_{H_i}, \gamma_{ii})_{1\leq i\leq M}^{\text{new}} = \underset{\sigma_{H_i},\gamma_{ii}}{\text{argmax}} \sum_{i=1}^{M}\sum_{j=1}^{M}\sum_{k=2}^{N} P(I_t=j, I_{t-1}=i|\theta^{\text{old}}, C_{\text{obs}}) \times \log(\alpha_{ij}(k)) \quad S.t. 0 \leq \gamma_{ii} \leq 1, \sigma_{H_{\min}} \leq \sigma_{H_i} \leq \sigma_{H_{\max}}, \gamma_{ij} \neq 0 \tag{29}$$

Local nonlinear optimization methods to find the optimal value of the unknowns (σ_{H_i} and γ_{ii}) are very sensitive to the initial values and it is possible that the optimization problem converges to a local optimum rather than a global. On the other hand, some of the moderate optimization techniques, like GA, do not require initial values, and if some certain criteria such as parallel searching, efficient interactions between different search trajectories and intelligent steps with appropriate step sizes are considered, the algorithm will normally reach a global solution, or a better local one. However, these methods are time consuming if used for global multivariate optimization of large databases in the presence of missing observations.[22,23]

In this article, we will use the following procedure in order to find the optimal value of the unknowns without analytical solution in the Maximization step. Similar procedures have been previously introduced in literature.[24,25] Using this procedure, an intelligent random sampling for initialization of some of the parameters in the EM algorithm is used. Unlike previous initialization techniques, which provide completely random initial values and select the one solution with the largest likelihood,[26] GA will provide the initial values based on the population's fitness and a target sampling rate. Consequently, the low performance initial values will be generated with very small probabilities,[23] and it will be more likely to have the appropriate initial values for the EM algorithm.

1. At the first iteration of the EM algorithm, start the nonlinear optimization problem in Eq. 29 with only a few generations of the GA.
2. Continue the optimization of the function in Eq. 29 with results of the GA as the initial values for the local interior point nonlinear constrained optimization algorithm (more details are available in the references.[27,28]
3. Having the optimal values from the previous step, continue the maximization step following Eqs. 23, 24, 26, and 28.
4. Save the calculated optimal values as θ^{old} for the next iteration which starts from step 2.

The initial values for the mean vectors and covariance matrices (μ_i and Σ_i) of the different modes in the EM algorithm can be obtained from an initial solution based on a mixture of multivariate Gaussian distributions assumption for the observations (Y_O). The initial values for $\gamma_{ij,i\neq j}$ can be selected to be equal to 0.5, assuming equal probability for all the transitions. In the cases where operating modes of the scheduling variable are unclear, the mean values of the scheduling variable at each operating mode ($H_i^{'}s$) and the initial values for the validity of the scheduling variable at each operating mode (σ_{H_i}) can be obtained assuming that the scheduling variable follows a mixture of Gaussian distributions.

The optimization problem will iterate between the E and M steps until the convergence criterion is satisfied. The convergence criterion in this article is set to be less than the absolute value of the likelihood change in two successive iterations.[17]

Furthermore, $P(I_k=j, I_{k-1}=i|\theta^{\text{old}}, C_{\text{obs}})$ and $P(I_k=i|\theta^{\text{old}}, C_{\text{obs}})$ are required to complete the maximization step in Eqs. 23 to 29. These terms are calculated as follows

$$\text{if } Y_k \text{ is observed,}$$

$$P(I_k=j, I_{k-1}=i|\theta^{\text{old}}, C_{\text{obs}}) = P(I_k=j, I_{k-1}=i|Y_{t_1}, \ldots, Y_{t_\alpha}, \theta^{\text{old}}, H_1, \ldots, H_N) = \frac{f(Y_{t_1}, \ldots, Y_{t_\alpha}|I_k=j, I_{k-1}=i, \theta^{\text{old}}, H_1, \ldots, H_N) \times P(I_k=j, I_{k-1}=i|\theta^{\text{old}}, H_1, \ldots, H_N)}{\sum_{i=1}^{M}\sum_{j=1}^{M} f(Y_{t_1}, \ldots, Y_{t_\alpha}|I_k=j, I_{k-1}=i, \theta^{\text{old}}, H_1, \ldots, H_N) \times P(I_k=j, I_{k-1}=i|\theta^{\text{old}}, H_1, \ldots, H_N)} = \frac{f(Y_k|I_k=j, \theta^{\text{old}}) P(I_k=j|I_{k-1}=i, H_{k-1}, \theta^{\text{old}}) \times P(I_{k-1}=i|\theta^{\text{old}}, H_1, \ldots, H_{k-2})}{\sum_{i=1}^{M}\sum_{j=1}^{M} f(Y_k|I_k=j, \theta^{\text{old}}) P(I_k=j|I_{k-1}=i, H_{k-1}, \theta^{\text{old}}) \times P(I_{k-1}=i|\theta^{\text{old}}, H_1, \ldots, H_{k-2})} \tag{30}$$

if Y_k is missing,

$$P(I_k=j, I_{k-1}=i|\theta^{\text{old}}, C_{\text{obs}})= P(I_k=j, I_{k-1}=i|\theta^{\text{old}}, H_1, \ldots, H_N)= P(I_k=j|I_{k-1}=i, \theta^{\text{old}}, H_{k-1})P(I_{k-1}=i|\theta^{\text{old}}, H_1, \ldots, H_{k-2}) \quad (31)$$

where in Eqs. 30 and 31, $P(I_k=j|I_{k-1}=i, \theta^{\text{old}}, H_{k-1})=\alpha_{ij}(k)^{\text{old}}$, $f(Y_k|I_k=j, \theta^{\text{old}})$ follows the multivariate normal distribution in Eq. 1 with mean vector and covariance matrix μ_i^{old} and Σ_i^{old} obtained from the previous iteration, and $P(I_{k-1}=i|\theta^{\text{old}}, H_1, \ldots, H_{k-2})$ is obtained through discrete-valued state propagation of Markov chain starting from the initial value of $P(I_1=i|\theta^{\text{old}}, C_{\text{obs}})=\pi_i^{\text{old}}$.

Finally, $P(I_k=i|\theta^{\text{old}}, C_{\text{obs}})$ can be obtained from summation of $P(I_k=i, I_{k-1}=j|\theta^{\text{old}}, C_{\text{obs}})$ over all the possible modes for I_{k-1}, that is

$$P(I_k=i|\theta^{\text{old}}, C_{\text{obs}})=\sum_{j=1}^{M} P(I_k=i, I_{k-1}=j|\theta^{\text{old}}, C_{\text{obs}}) \quad (32)$$

Operating Mode Recognition

Online operating mode (state) recognition is needed for fault detection. Through an online application, probability of the hidden process modes given the observations ($P(I_k|Y_k, \ldots, Y_1, H_k, \ldots, H_1)$) can be calculated using Hamilton's filtering strategy[2] as follows:

1. Calculate the joint probability of the modes I_k and I_{k-1} given the information up to time $k-1$

$$P(I_k, I_{k-1}|Y_{k-1}, \ldots, Y_1, H_{k-1}, \ldots, H_1)= P(I_k|I_{k-1}, H_{k-1})P(I_{k-1}|Y_{k-1}, \ldots, Y_1, H_{k-1}, \ldots, H_1) \quad (33)$$

where $P(I_k|I_{k-1}, H_{k-1})=\alpha_{ij}(k)$, and $P(I_{k-1}|Y_{k-1}, \ldots, Y_1, H_{k-1}, \ldots, H_1)$ is the previous output of the filter.

2. Update the probability of the modes I_k, I_{k-1} using the new observations at time k

$$P(I_k, I_{k-1}|Y_k, \ldots, Y_1, H_k, \ldots, H_1)= \frac{P(Y_k, I_k, I_{k-1}|Y_{k-1}, \ldots, Y_1, H_k, \ldots, H_1)}{P(Y_k|Y_{k-1}, \ldots, Y_1, H_k, \ldots, H_1)} \quad (34)$$

where

$$\begin{aligned} &P(Y_k, I_k, I_{k-1}|Y_{k-1}, \ldots, Y_1, H_k, \ldots, H_1)\\ &=P(Y_k|I_k, I_{k-1}, Y_{k-1}, \ldots, Y_1, H_k, \ldots, H_1)\\ &\times P(I_k, I_{k-1}|Y_{k-1}, \ldots, Y_1, H_k, \ldots, H_1)\\ &=f(Y_k|I_k)P(I_k, I_{k-1}|Y_{k-1}, \ldots, Y_1, H_{k-1}, \ldots, H_1) \end{aligned}$$

where $f(Y_k|I_k)$ follows the multivariate normal distribution in Eq. 1 and $P(I_k, I_{k-1}|Y_{k-1}, \ldots, Y_1, H_{k-1}, \ldots, H_1)$ is known from Eq. 33.

The denominator in Eq. 34 can be calculated as

$$P(Y_k|Y_{k-1}, \ldots, Y_1, H_k, \ldots, H_1) =\sum_{I_k=1}^{M}\sum_{I_{k-1}=1}^{M} P(Y_k, I_k, I_{k-1}|Y_{k-1}, \ldots, Y_1, H_k, \ldots, H_1)$$

3. The output of the filter will be

$$P(I_k|Y_k, \ldots, Y_1, H_k, \ldots, H_1)= \sum_{I_{k-1}=1}^{M} P(I_k, I_{k-1}|Y_k, \ldots, Y_1, H_k, \ldots, H_1) \quad (35)$$

Table 1. System Parameters to Generate the Simulation Data

$\pi_0=[0.25, 0.25, 0.25, 0.25]$
$\gamma_{11}=0.98, \gamma_{12}=0.5, \gamma_{21}=0.5, \gamma_{22}=0.95, \gamma_{23}=0.3$
$\gamma_{31}=0.2, \gamma_{32}=0.6, \gamma_{33}=0.89, \gamma_{42}=0.4, \gamma_{44}=0.92$
$\mu_1=[5, 3], \mu_2=[10, 8], \mu_3=[11, 9], \mu_4=[18, 16]$
$\Sigma_1=\begin{pmatrix} 1 & 0.5 \\ 0.5 & 2.5 \end{pmatrix}, \Sigma_2=\begin{pmatrix} 3 & 0.75 \\ 0.75 & 4.5 \end{pmatrix}$
$\Sigma_3=\begin{pmatrix} 4 & 0.4 \\ 0.4 & 5.5 \end{pmatrix}, \Sigma_4=\begin{pmatrix} 2 & 0.6 \\ 0.6 & 0.5 \end{pmatrix}$
$\sigma_{H_1}=5, \sigma_{H_2}=4, \sigma_{H_3}=4.5, \sigma_{H_4}=3$

Results and Discussion

As previously stated in the introduction and problem statement sections, the advantage of using the proposed structure of this article over conventional HMMs can be very well demonstrated when a process has an asymmetric time-varying transition behavior between different operating modes, that is, when some of the operating modes are far from the majority and the scheduling variable helps in the modeling and filtering steps by providing more flexibility. Examples of such situations are demonstrated in the case studies of this section.

A numerical case study

In this simulation case study, we consider a system that operates in four operating modes which are normal (Mode 1), abnormal 1 (Mode 2), abnormal 2 (Mode 3) and fault (Mode 4). Observations of each mode are assumed to follow different multivariate normal distributions. The abnormal modes (Modes 2 and 3) are assumed to be close to each other and far from the normal and faulty modes as in Table 1. Operating modes of the process follow the structure proposed in Figure 1. In order to test the validity of the algorithm for missing observations, some observations (10%) are randomly missed in the simulation at various sampling instants. The missing data are assumed to be completely MAR. The scheduling variable is assumed to linearly transit between different operating conditions. Parameters used for the simulation are presented in Table 1. Parameters γ_{13}, γ_{14}, γ_{24}, γ_{34}, γ_{41}, and γ_{43} can be further obtained from Eqs. 6, 7, and 8 respectively.

Following the expectation-maximization procedure introduced in Parameter Estimation Based on the Expectation Maximization Algorithm, the final estimated value of the parameters from a training dataset including 8000 data are obtained as presented in Table 2. As it is clear from this table, the estimated parameters are close to true parameters

Table 2. Estimated Parameters from the EM Algorithm

$\pi_0=[0.25, 0.25, 0.25, 0.25]$
$\gamma_{11}=0.9912, \gamma_{12}=0.6633, \gamma_{21}=0.6756, \gamma_{22}=0.9775, \gamma_{23}=0.2882$
$\gamma_{31}=0.3037, \gamma_{32}=0.5379, \gamma_{33}=0.8861, \gamma_{42}=0.3470, \gamma_{44}=0.9397$
$\mu_1=[5.1991, 3.2050], \mu_2=[9.8667, 7.8267]$
$\mu_3=[10.8162, 8.7122], \mu_4=[17.0019, 14.9031]$
$\Sigma_1=\begin{pmatrix} 1.2727 & 0.7264 \\ 0.7264 & 3.0525 \end{pmatrix}, \Sigma_2=\begin{pmatrix} 3.4465 & 1.0722 \\ 1.0722 & 4.7376 \end{pmatrix}$
$\Sigma_3=\begin{pmatrix} 4.7419 & 0.6518 \\ 0.6518 & 5.6666 \end{pmatrix}, \Sigma_4=\begin{pmatrix} 2.6854 & 0.6330 \\ 0.6330 & 0.6890 \end{pmatrix}$
$\sigma_{H_1}=5.2613, \sigma_{H_2}=4.1880, \sigma_{H_3}=4.3707, \sigma_{H_4}=3.4833$

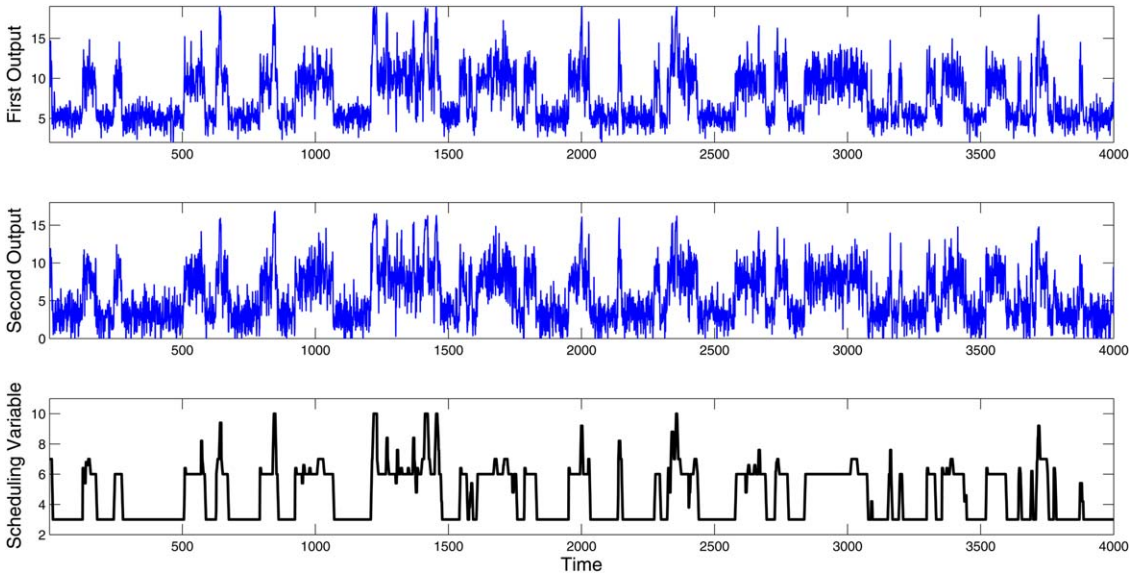

Figure 2. Validation dataset.

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

of the process. Due to the nature of the industrial process which is used in this article, with small sampling rates and large number of operating modes, using such large training datasets are more appropriate for the proposed method to provide robust process identification results. However, size of the training dataset might vary for other industrial applications according to their sampling-rate and number of operating modes.

In order to compare the performance of the proposed method in operating condition diagnosis with conventional HMMs,[12,13] first, another dataset including 8000 data are generated from the same model in Table 1 for training purposes. Since conventional hidden Markov models cannot deal with missing observations, the complete dataset is assumed to be observable and the performance is only compared in the adaptive property of the new technique rather than handling of the missing data. Next, a validation dataset is generated from the same model, which is presented in Figure 2. Results of the filtering procedure to find the probability of the hidden operating modes given observations based on the proposed method of this article (Operating Mode Recognition section) are presented in Figure 3. Based on these probabilities, true and the estimated operating modes of the process are presented in Figure 4.

As presented in Figure 4, except for some very fast changes in the dynamics of the process which cause some false alarms (time instants around 1700 and 2700 for example), the method is generally able to detect the true operating mode of the process.

Results of the probability calculation and operating mode recognition based on conventional multivariate HMMs are presented in Figures 5 and 6.

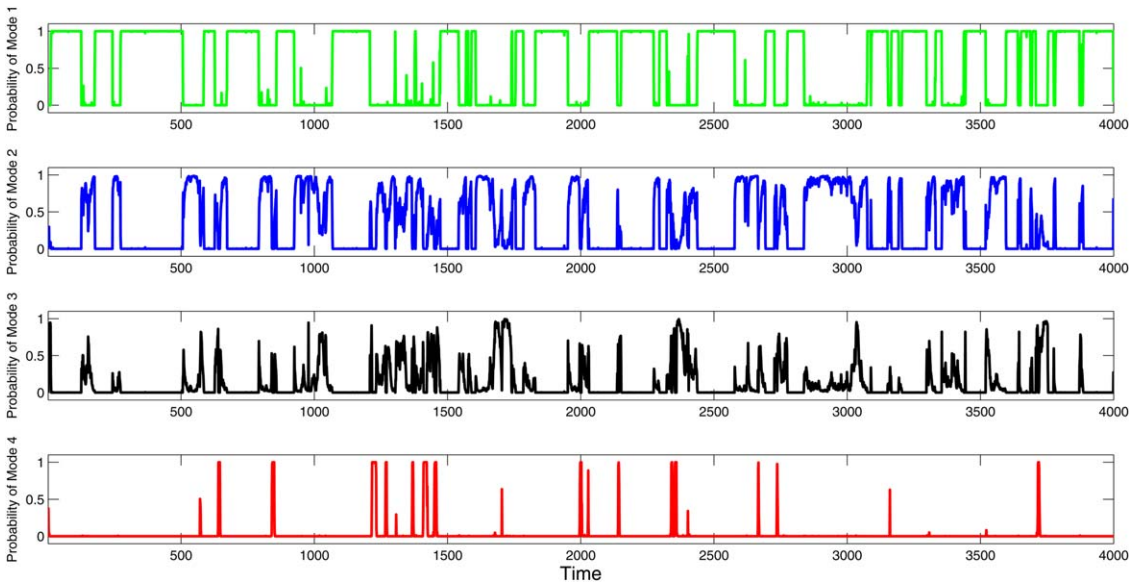

Figure 3. Probability of the hidden operating modes for the validation dataset using the proposed method of this article.

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

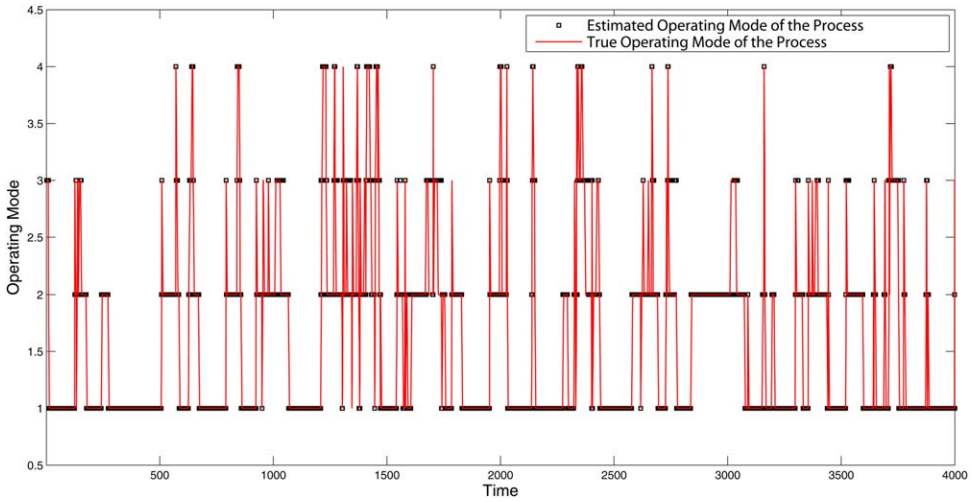

Figure 4. True and estimated hidden operating modes of the process for the validation dataset using the proposed method of this article.

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

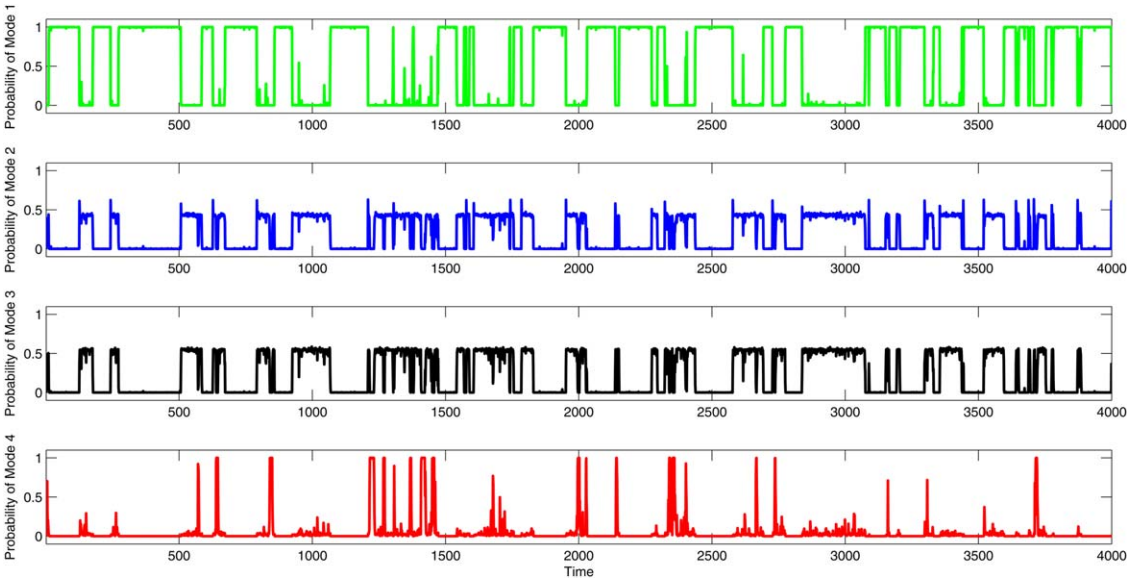

Figure 5. Probability of the hidden operating mode for the validation dataset based on a conventional multivariate hidden Markov model.

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

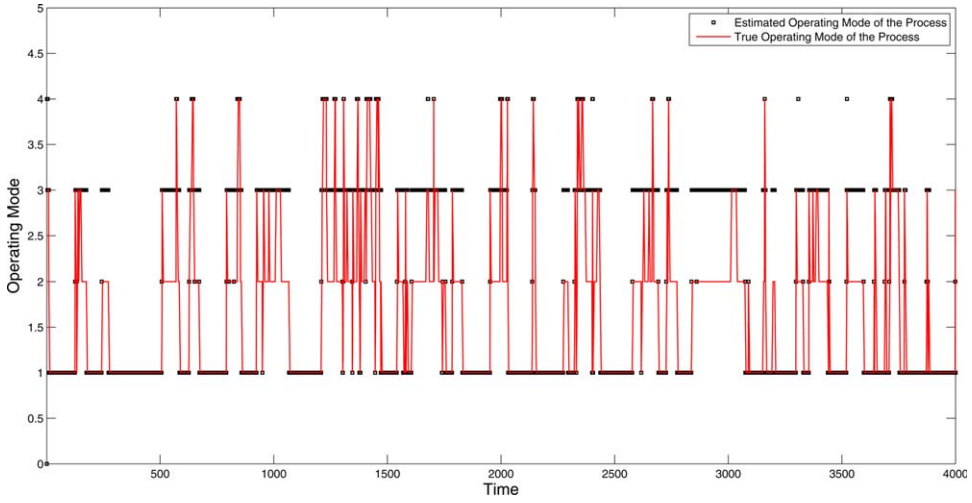

Figure 6. True and estimated hidden operating modes of the process for the validation dataset based on a conventional multivariate hidden Markov model.

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

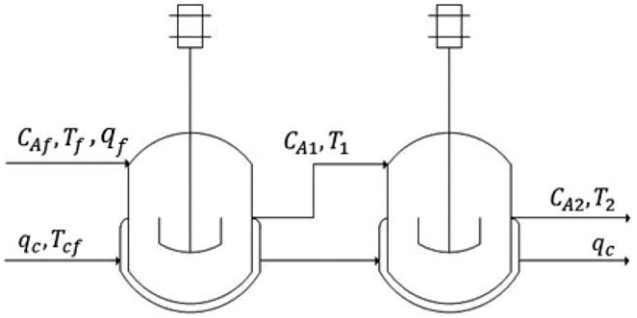

Figure 7. Two CSTRs in series.

As presented in Figure 5, applying conventional HMMs, the developed model is unable to distinguish between the two close abnormal operating modes (Modes 2 and 3), that is, probability of the observations given these modes is close to each other and close to 0.5. Consequently, in Figure 6, one could see that based on the probabilities in Figure 5, at several time instants, operating mode 2 is incorrectly categorized as 3.

This example illustrates one of the cases where applying time-varying transition probabilities based on the distributions introduced in Eqs. 3 and 4 can provide more accurate predictions for the true hidden operating mode of the process. As this example shows, existence of an accurate

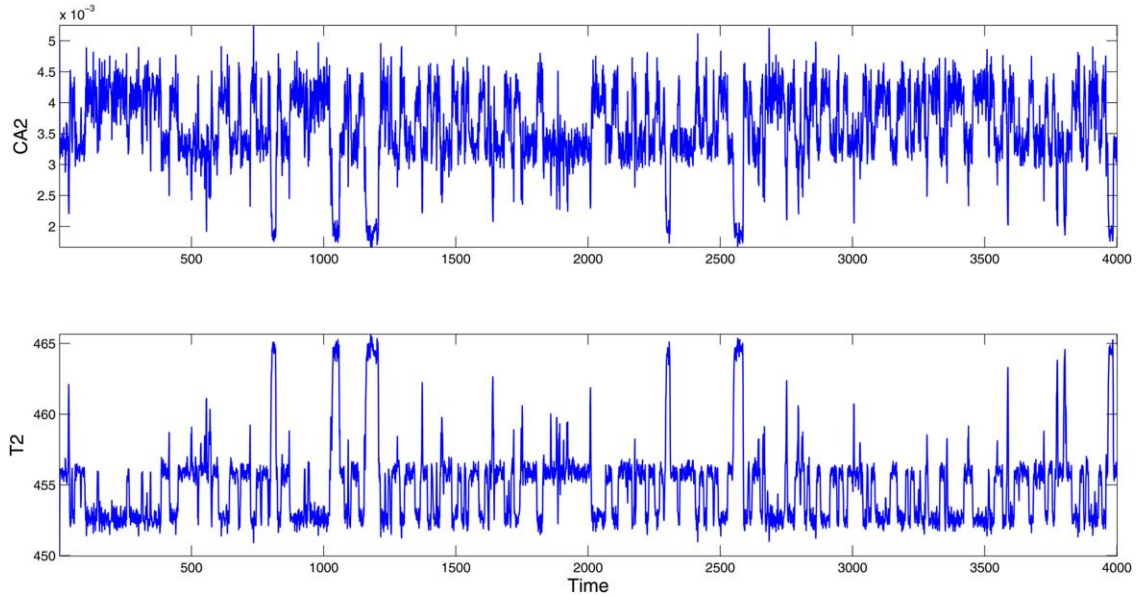

Figure 8. Different operating modes for the process variables (validation dataset).

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

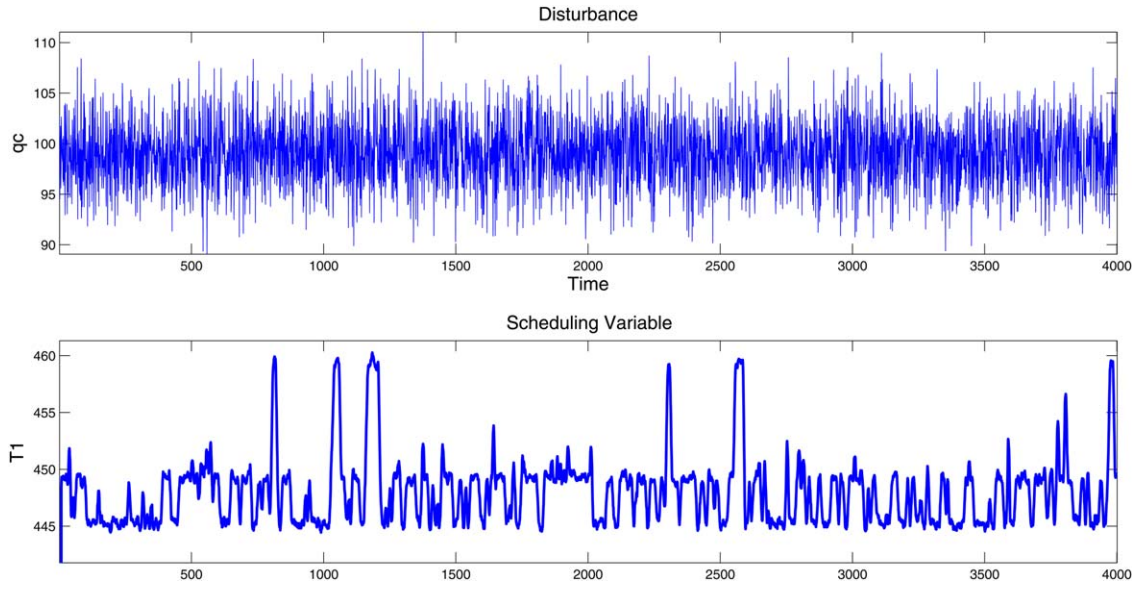

Figure 9. The scheduling variable and disturbance to the process (validation dataset).

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

Table 3. Estimated Parameters for the CSTRs in Series Using the EM Algorithm

$\pi_0 = [0.3333, 0.3333, 0.3333]$
$\gamma_{11} = 0.9472, \gamma_{21} = 0.6301, \gamma_{22} = 0.9142, \gamma_{33} = 0.9968,$
$\mu_1 = [0.0041, 452.7567], \mu_2 = [0.0033, 455.8058],$

$$\mu_3 = [0.0023, 461.5816]$$

$$\Sigma_1 = \begin{pmatrix} 9.2697\times10^{-8} & -2.0526\times10^{-4} \\ -2.0526\times10^{-4} & 0.5908 \end{pmatrix}$$
$$\Sigma_2 = \begin{pmatrix} 7.1064\times10^{-8} & -2.1382\times10^{-4} \\ -2.1382\times10^{-4} & 0.7967 \end{pmatrix}$$
$$\Sigma_3 = \begin{pmatrix} 3.0473\times10^{-7} & -1.8\times10^{-3} \\ -1.8\times10^{-3} & 11.3670 \end{pmatrix}$$

$\sigma_{H_1} = 0.9089, \sigma_{H_2} = 2.7259, \sigma_{H_3} = 9.8989$

scheduling variable can provide more flexibility to model and monitor the process transitions between different operating modes.

A simulation study

In this example, the proposed method is tested on the two continuous stirred tank reactors (CSTRs) in series introduced by Henson et al.[29] The irreversible exothermic first-order reaction $A \rightarrow B$ occurs in the two reactors in series. The feed enters the first reactor with flow rate q_f and temperature and concentration C_{Af} and T_f respectively. The product of the first reactor is then feed to the second reactor. A parallel flow (q_c) is used as the coolant. The process is illustrated in Figure 7.

The process works in open loop condition. Concentration of the product C_{A2} is the important output variable for control purposes. In the steady-state condition, C_{A2} is around $0.05\left(\frac{\text{mol}}{\text{L}}\right)$.

Temperature of the first reactor (T_1), which can provide a preindication to the operating condition of the process before receiving the final product, is selected as the scheduling variable. It is assumed that there is no measurement for the coolant flow rate $\left(q_c\left(\frac{\text{L}}{\text{min}}\right)\right)$ and coolant flow rate is selected as the disturbance with variance 10.

The main cause of the changes in operating condition of the process is the sudden changes in the feed flow rate (q_f). In this example, the feed flow rate is assumed to vary between 3 operating modes following the Markov switching model given in Eq. 5. Normal (Mode 1) and abnormal (mode 2) operations occur when the feed flow rate is around its steady-state value, that is, $q_{f(\text{Mode}_1)} = 105.4\left(\frac{\text{L}}{\text{min}}\right)$ and $q_{f(\text{Mode}_2)} = 112.6\left(\frac{\text{L}}{\text{min}}\right)$. The faulty mode occurs when feed flow rate suddenly increases $\left(q_{f(\text{Mode}_3)} = 134.3\left(\frac{\text{L}}{\text{min}}\right)\right)$ and the coolant flow rate is not enough to maintain a constant process temperature. In such situations, the output temperature (T_2) suddenly increases. This is followed by a very low product concentration (C_{A2}). Therefore, these two key variables (T_2 and C_{A2}) are selected as indicators of the operating condition of the process. An example of the normal, abnormal and faulty operating conditions of the process is illustrated in Figure 8. The output temperature is in Kelvin unit. Parameters of the transition probability matrix, which cause the switching behavior in the feed flow rate, are selected as $\gamma_{11} = 0.95, \gamma_{21} = 0.7, \gamma_{22} = 0.93, \gamma_{33} = 0.97, \sigma_{H1} = 15, \sigma_{H2} = 13,$ and $\sigma_{H3} = 10$. Parameters γ_{12}, γ_{23}, and γ_{32} can be further obtained from Eqs. 6, 7, and 8, respectively.

The scheduling variable and the disturbance to the process are presented in Figure 9. Applying a moving average filter, the scheduling variable is filtered to provide an overall indication for the transitions of the process between different operating modes. According to the previous discussion on the feed flow rate, in this example, the faulty mode is selected far from the normal and abnormal operations of the process. Similar to the previous example, this asymmetric mode transition provides a condition to more clearly demonstrate the advantage of the proposed method of this article over conventional techniques.

The historical dataset including 8000 data points, which are different but from the same model as Figure 8, is used for training purposes. Similar to previous example, 10% of the training data are assumed to be randomly missing.

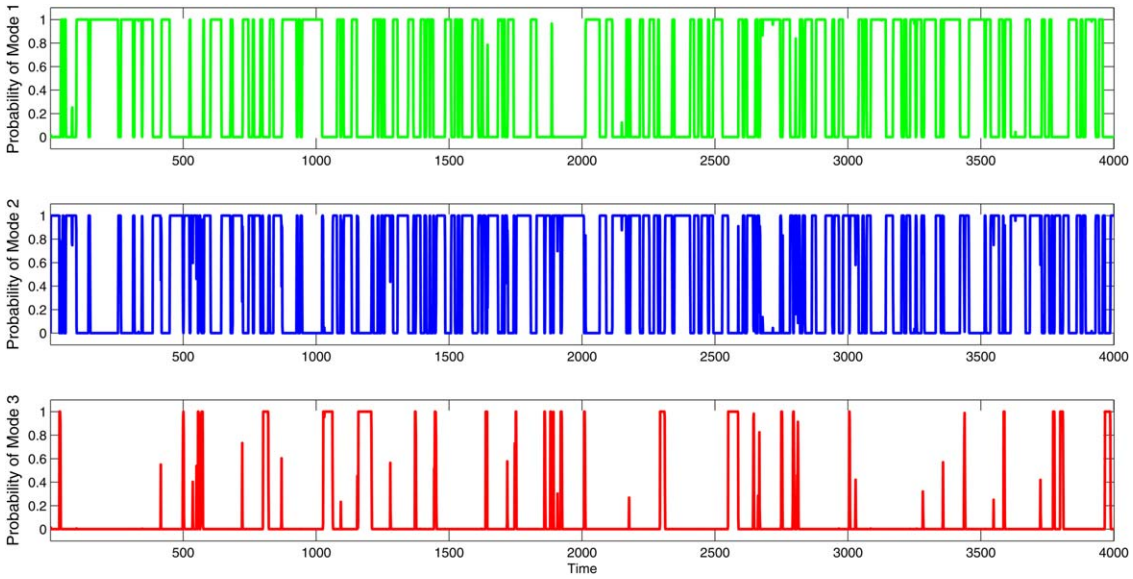

Figure 10. Probability of the hidden operating modes for the CSTRs in series based on the proposed method of this article and the validation dataset in Figure 8.

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

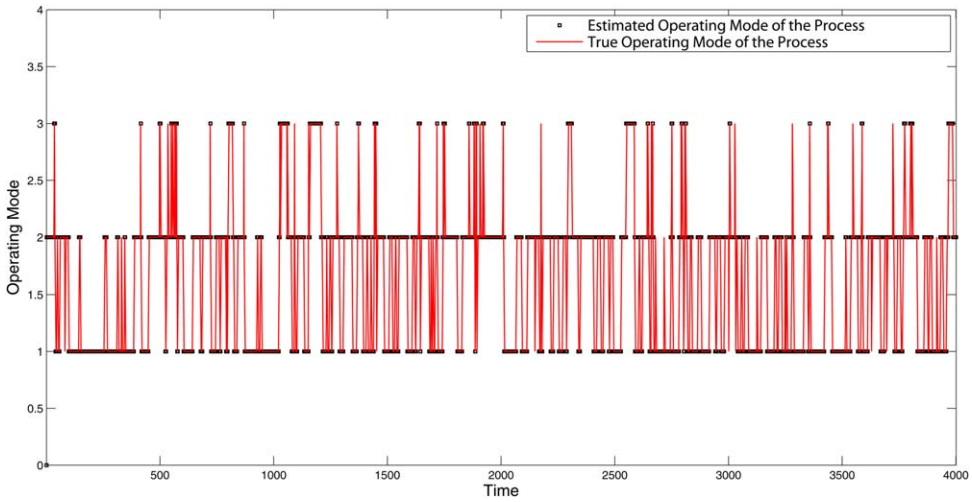

Figure 11. True and estimated operating modes of the process for the CSTRs in series based on the proposed method of this article and the validation dataset in Figure 8.

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

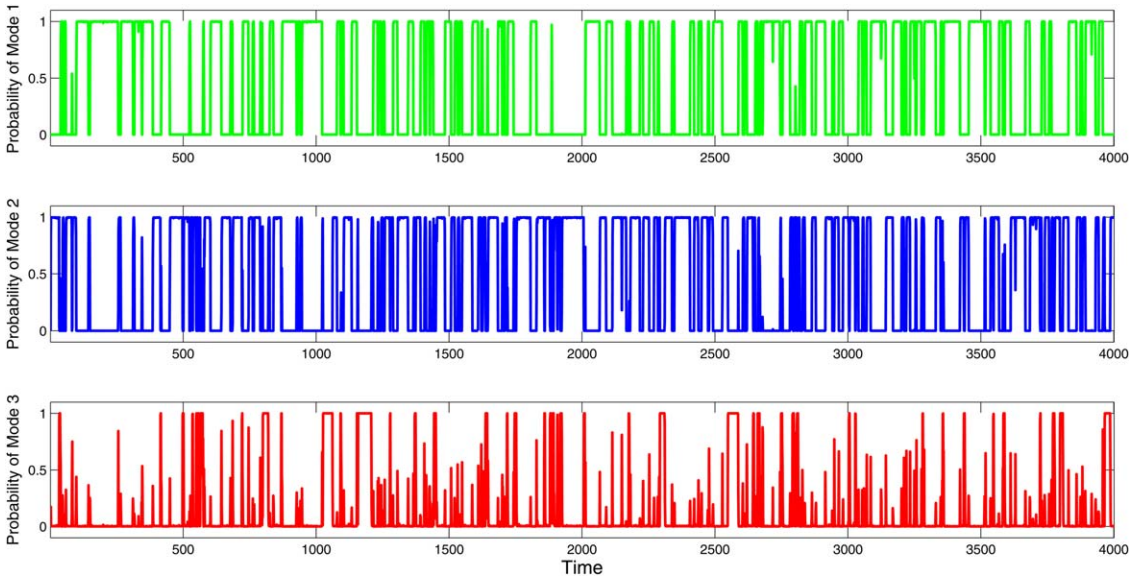

Figure 12. Probability of the hidden operating modes for the CSTRs in series using conventional HMMs and the validation dataset in Figure 8.

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

Results of the parameter estimation based on EM algorithm introduced in Parameter Estimation based on the Expectation Maximization Algorithm section are presented in Table 3.

In order to compare the performance of the proposed method in this article and conventional HMMs, both methods are tested on the data in Figures 8 and 9. The training dataset to train conventional HMMs is the same as the historical dataset that is used for parameter estimation in Table 3. However, here, the complete dataset is assumed to be observable since conventional HMMs cannot deal with the missing observations. Therefore, the comparison is made only in the time-varying property of the new model.

Using the proposed method of this article, probability of the different operating modes given new observations is presented in Figure 10. Based on these probabilities, true and estimated operating conditions of the process are demonstrated in Figure 11.

As presented in Figure 11, similar to the previous case study, other than some very fast changes in the process variables, the proposed method of this article is able to detect the different operating modes of the process.

Results of the filtering and operating condition diagnosis based on conventional HMMs are presented in Figures 12 and 13, respectively.

Comparing the results in Figures 11 and 13, one could easily observe that the conventional method provides several false alarms in detection of the faulty mode. This is another good example to show the merit of the proposed method of this article for the cases where the process operates among asymmetric operating modes and adaptive transition probabilities provide more flexibility for overall monitoring of the

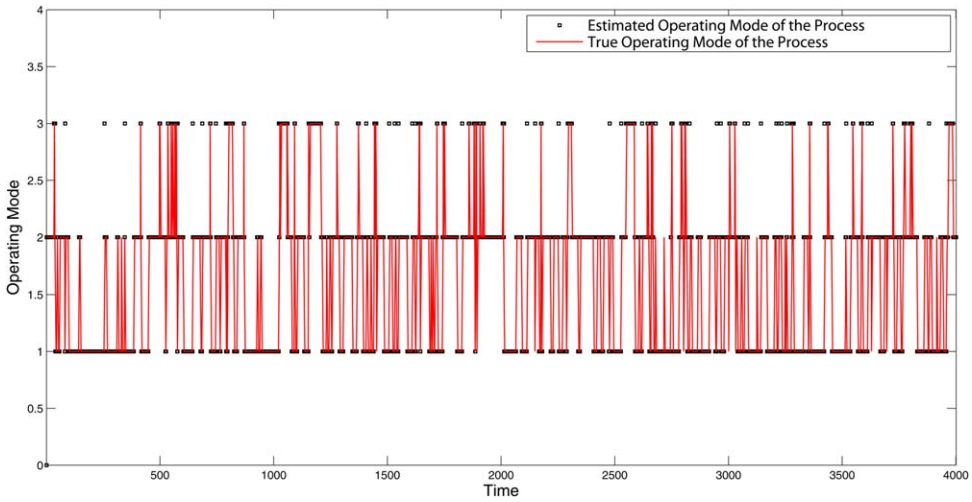

Figure 13. True and estimated operating modes of the process for the CSTRs in series using conventional HMMs and the validation dataset in Figure 8.

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

process. In this example, existence of the scheduling variable helps the model to adapt to the new modes. Therefore, the new framework eventually enables the model to better detect the faulty mode which behaves far from the normal and abnormal modes.

An industrial case study

In this section, the proposed process monitoring strategy is tested on the underflow of an industrial scale primary separation vessel (also known as the PSV unit) in the oil-sand industry. Due to the high solid concentration and sand deposition in the underflow, there is a concern of underflow plugging that can cause process upsets and interruptions. The schematic of the unit is presented in Figure 14.

The PSV unit is a cone shape vessel that separates the feed. The feed stream, which usually contains water, bitumen, coarse sand, and fine, enters the unit. The lighter part of the feed (bitumen) floats to the top section and creates the froth layer. Fines, water, and some bitumen aggregates are pumped through the middlings stream for recycling purposes. The underflow, which is also known as the tailings, contains coarse solids and some bitumen. When the flow rate of the tailings is below a critical value (critical minimum velocity) coarse particles start to deposit and cause upset operating conditions in the underflow. The complete plugging of the pipeline is referred to as "sanding" in oil-sand industries.

The critical minimum velocity is a function of different variables including carrier fluid's density, carrier fluid's viscosity, coarse particle volumetric fraction, coarse particle diameter, etc.[30] Carrier fluid is a portion of the fluid which contains fines. Therefore, the middlings stream is a good indicator for the carrier fluid's properties. On the other hand, the tailings stream, which usually contains coarse particles, is a good indicator for the depositing material's properties. Thus, the properties of these two layers play the key roles in sanding detection.

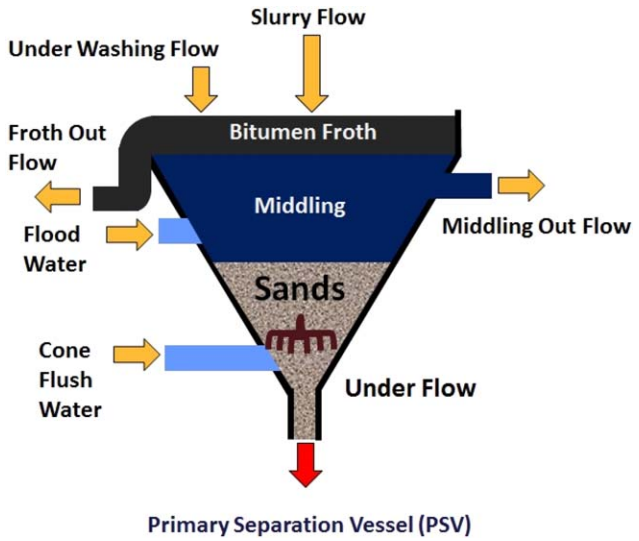

Figure 14. Schematic of the PSV unit.

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

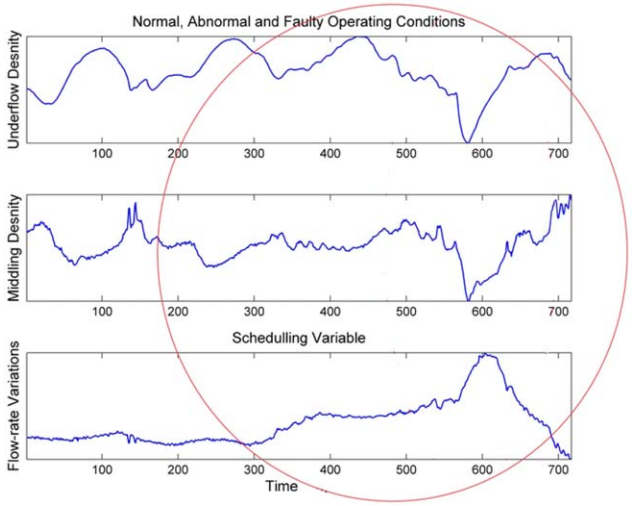

Figure 15. A case of upset operating condition in the primary separation vessel from historical data.

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

Table 4. Estimated Parameters for the Industrial Case Study Using the EM Algorithm

$\pi_0 = [0.3333, 0.3333, 0.3333]$
$\gamma_{11} = 0.9934, \gamma_{21} = 0.1987, \gamma_{22} = 0.9887, \gamma_{33} = 0.9929,$
$\mu_1 = [1.4483, 1.4965], \mu_2 = [1.2757, 1.5246], \mu_3 = [1.1490, 1.4227]$

$$\Sigma_1 = \begin{pmatrix} 0.0022 & 0.0003 \\ 0.0003 & 0.0013 \end{pmatrix}$$

$$\Sigma_2 = \begin{pmatrix} 0.0041 & -0.0018 \\ -0.0018 & 0.0032 \end{pmatrix}$$

$$\Sigma_3 = \begin{pmatrix} 0.0043 & 0.0014 \\ 0.0014 & 0.0085 \end{pmatrix}$$

$\sigma_{H_1} = 4.681 \times 10^3, \sigma_{H_2} = 1.257 \times 10^3, \sigma_{H_3} = 3.7051 \times 10^4$

Density of the middlings and underflow streams can be measured through online analyzers. Since all the other necessary variables which impact the critical minimum velocity are directly, or indirectly, functions of these two variables, the densities are frequently observed by the operators to infer the operating condition of the process and avoid sanding conditions. Therefore, these two variables are selected as the sanding indicators. The tailings flow rate can also provide a preindication to the operating condition of the process and is selected as the scheduling variable in this study.

When the process approaches the upset operating condition, underflow and middlings densities start to gradually increase. Since the tailings flow rate is in closed loop with underflow density, when the underflow density exceeds some high limit, the pump revolution per minute (RPM) starts to increase to remove the deposited sand and return the operation to its normal condition. This causes a sudden decrease in the middlings and underflow densities. If in the early stage of such circumstances, the operators can be notified about the operating mode of the process, they can add water through the cone flush water (Figure 14) to assist the suspension of solid deposits and avoid complete sanding of the pipeline. An example of an upset operating region which has occurred in the past is presented in Figure 15.

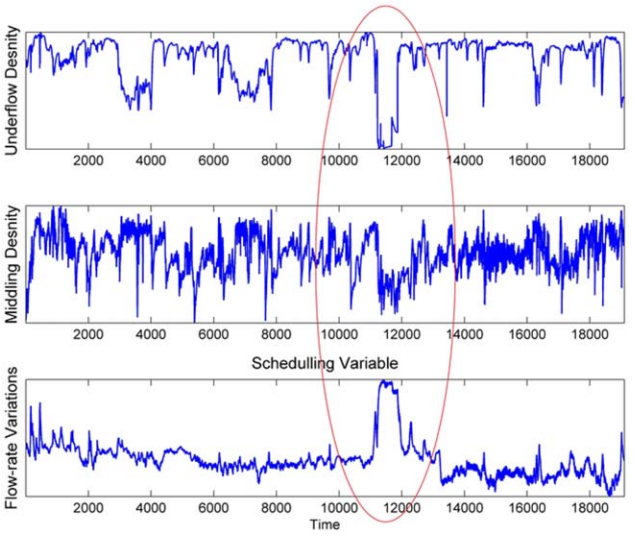

Figure 16. Validation dataset for the industrial case study.

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

From Figure 15, one could observe that the abnormal (intermediate) and beginning of the upset operating conditions, which are denoted by a red circle, are close to each other and far from the normal operation of the process. Existence of such asymmetric behavior provides a good example to show the merit of using time varying over fixed transition probabilities.

In this industrial case study, eight cases of upset operating conditions similar to Figure 15, which have occurred during a year in the historical data, are selected for the training purposes. A case of an upset operating region which has occurred in a month of a different year is selected as the test dataset. During the online measurement process, some observations appear as "Not a Number" or "NaN" value in the server. These observations are treated as the "missing observations" in this case study.

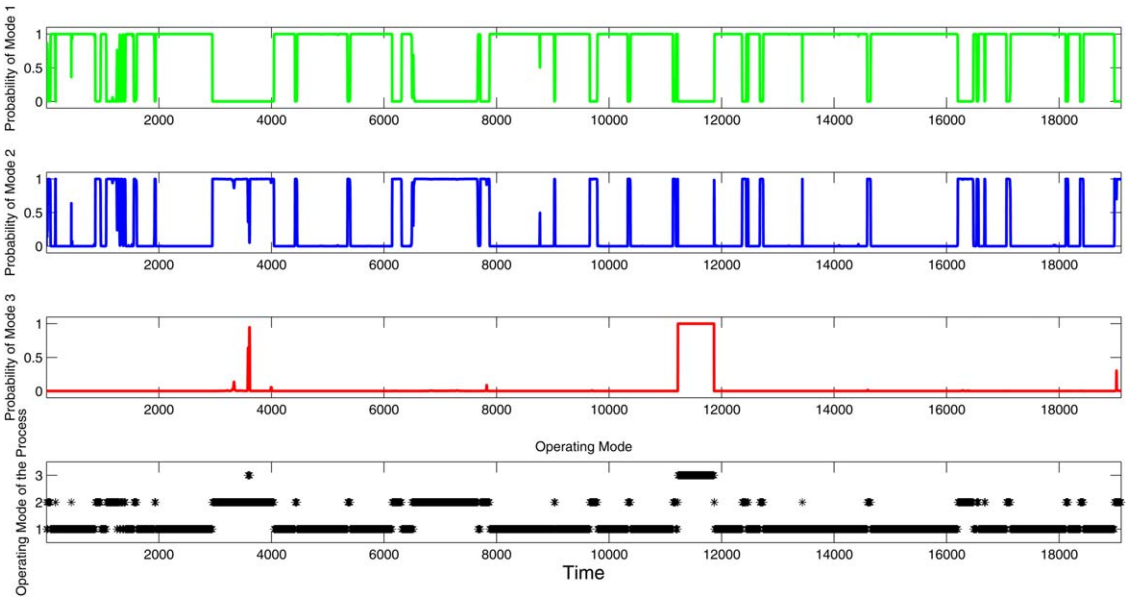

Figure 17. Operating modes of the process for the industrial case study based on the proposed method of this article.

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

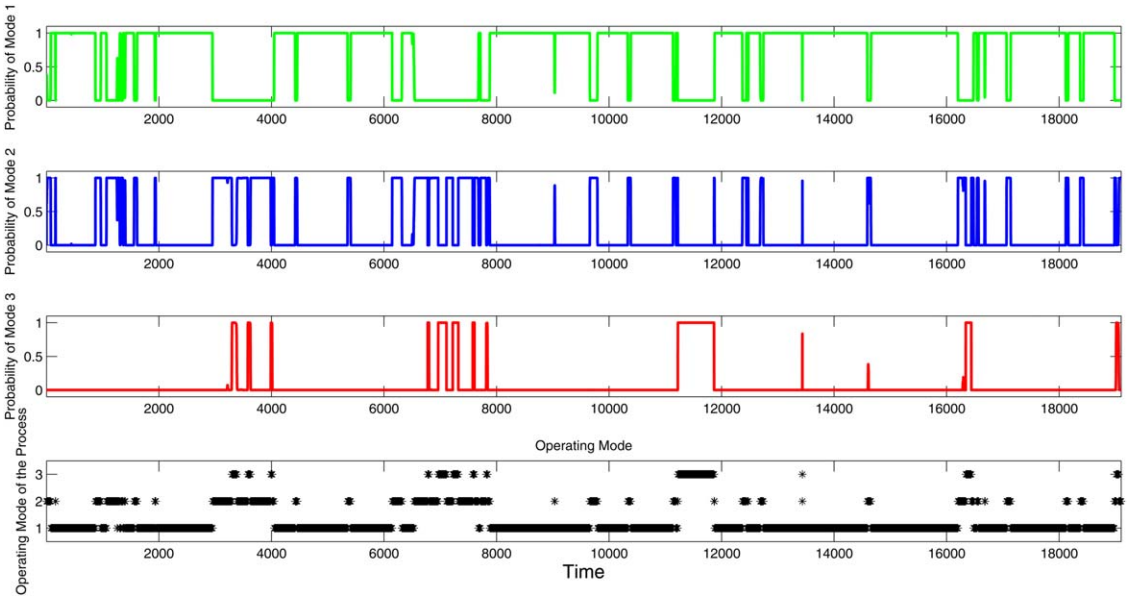

Figure 18. Operating modes of the process for the industrial case study based on conventional HMMs.

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

Following the proposed estimation method in this paper, the parameters are obtained as in Table 4.

The validation dataset is presented in Figure 16. The upset behavior occurs at the time period around 12,000 where the densities suddenly start to decrease due to the reaction of the pump to an abnormal event (red oval). Results of the operating mode recognition based on the Hamilton's filtering algorithm are presented in Figure 17.

In Figure 17, operating mode 1 is the normal, operating mode 2 is the abnormal and operating mode 3 is the upset operation in the process. As presented in Figures 16 and 17, a sudden decrease in the density of the middlings and underflow streams, which is accompanied by an increase in the underflow pump RPM, has ended to an alarm (operating mode 3) during the online diagnosis application.

In order to compare the results with the case of conventional HMMs, the same training and test datasets are utilized. However, only the observed part is used since conventional HMMs cannot deal with missing observations. Results of the operating mode diagnosis based on conventional HMMs are presented in Figure 18.

Again, it is observed from Figure 18 that a number of false alarms appear in the predictions and the abnormal and faulty modes are not clearly diagnosed from each other when using conventional HMMs. The better performance of the proposed method of this article (compare the results in Figures 17 and 18) can be understood from the behavior of the scheduling variable (Flow rate) in Figure 16. One could see that in the faulty mode, the scheduling variable suddenly increases and helps in operating condition diagnosis. However, in the normal and abnormal operations, the scheduling variable correctly affects the transition probabilities to remain in the normal and abnormal modes and avoid false alarms.

Conclusion

In this article, a novel approach for modeling and monitoring of the time-varying multivariate regime switching systems subject to fault and missing observations is introduced. Due to the existence of the missing observations and unknown operating modes, the EM algorithm is applied to find the unknown parameters. Also, an optimization procedure is introduced to reduce the sensitivity of the EM algorithm to its initial values.

The proposed method is tested on two simulations and one industrial case studies and shows a superior performance in detecting the different operating modes of the process in comparison to the conventional methods. In general, application of time varying transition probabilities, as introduced in this article, provides better classifications for different operating modes of the process. This improvement can be more clearly observed in the processes with asymmetric time varying transitions between different operating modes.

Literature Cited

1. Hamilton JD. Rational-expectations econometric analysis of changes in regime: an investigation of the term structure of interest rates. *J Econ Dyn Control*. 1988;12:385–423.
2. Hamilton JD. Analysis of time series subject to changes in regimes. *J Econometrics*. 1990;45:39–70.
3. Bollen NPB, Gray SF, Whaley RE. Regime switching in foreign exchange rates: evidence from currency option prices. *J Econometrics*. 2000;94:239–276.
4. Ang A, Bekert G. Regime switches in interest rates. *J Business Econ Stat*. 2002;20(2):163–182.
5. Pelletier D. Regime switching for dynamic correlations. *J Econometrics*. 2006;131:445–473.
6. Mount TD, Ning Y, Cai X. Predicting price spikes in electricity markets using a regime switching model with time varying parameters. *Energy Econ*. 2006;28:62–80.
7. Diebold FX, Lee J, Weinbach GC. Regime switching with time varying transition probabilities. In: Hargreaves C, editor, Nonstationary Time Series Analysis and Cointegration. Oxford: Oxford University Press, 1994;283–302.
8. Filardo AJ. Business-Cycle phases and their transitional dynamics. *J Business Econ Stat*. 1994;12(3):299–308.
9. Otrano E. The multi-chain Markov switching model. *J Forecasting*. 2005;24:523–537.
10. Costa OLV, Fragoso MD, Marques RP. Discrete-time Markov jump linear systems. New York: Springer, 2005.
11. Jin X, Huang B. Identification of switched Markov autoregressive exogenous systems with hidden switching state. *Automatica*. 2012; 48(2):436–441.
12. Ghasemi A, Yacout S, Ouali M. Parameter estimation methods for condition-based maintenance with indirect observations. *IEEE Trans Reliability*. 2010;59(2):426–439.

13. Wong WC, Lee JH. Fault detection and diagnosis using hidden Markov disturbance models. *Ind Chem Res*. 2010;49:7901–7908.
14. Jiang R, Yu J, Makis V. Optimal Bayesian estimation and control scheme for gear shaft fault detection. 2012;63:754–762.
15. Kim MJ, Makis V, Jiang R. Parameter estimation for partially observable systems subject to random failure. *Appl Stochastic Models Business Ind*. 2013;29:279–294.
16. Deng J, Huang B. Identification of nonlinear parameter varying systems with missing output data. *AIChE*. 2012;58(11):3454–3467.
17. McLachlan GJ, Krishnan T. The EM algorithm and Extensions. Hoboken, New Jersy: Wiley, 2008.
18. Keshavarz M, Huang B. Bayesian and Expectation methods for multivariate change point detection. *Comput Chem Eng*. 2014;60:339–353.
19. Imtiaz S, Shah S. Treatment of missing values in process data analysis. *Can J Chem Eng*. 2008;86:838–858.
20. Schafer JL, Grahm JW. Missing data: our view of the state of the art. *Psychol Methods*. 2008;7:2076–2089.
21. Dempster AP, Laird NM, Rubin DB. Maximum likelihood from incomplete data via the EM algorithm. *J R Stat Soc*. 1997;39(1):1–38.
22. Whitley DA. Genetic Algorithm Tutorial. Computer Science Department. Colorado State University, 1993.
23. Lucasius CB, Kateman G. Understanding and using genetic algorithms Part 1. *Concepts, properties and context*. *Chemom Intell Lab Syst*. 1993;19:1–33.
24. Martinez AM, Vitria J. Learning mixture models using a genetic version of the EM algorithm. *Pattern Recognit Lett*. 2000;21:759–769.
25. Pernkopf F, Bouchaffra D. Genetic-based EM algorithm for learning Gaussian mixture models. *IEEE Trans Pattern Anal Machine Intell*. 2005;27(8):1344–1348.
26. Karlis D, Xekalaki E. Choosing initial values for the EM algorithm for finite mixtures. *Comput Stat Data Anal*. 2003;41:557–590.
27. Byrd RH, Gilbert JC, Nocedal J. A trust region method based on interior point techniques for non-linear programming. *Math Programming*. 2000;89(1):149–185.
28. Byrd RH, Hribar ME, Nocedal J. An interior point algorithm for large-scale non-linear programming. *SIAM J Optimizat*. 1999;9(4):877–900.
29. Henson MA, Seborg DE. Input-output linearization of general nonlinear processes. *AIChE*. 1990;36(11):1753–1757.
30. Kokpinar MA, Gogus M. Critical flow velocity in slurry transporting horizontal pipelines. *J Hydraulic Eng*. 2011;127(9):763–771.

Appendix

Details of the derivations for Eq. 21

$$\int P(Y_k|\theta^{\text{old}}, C_{\text{obs}}, I_k=i)\log f(Y_k|I_k=i,\mu_i,\Sigma_i)dY_k$$

$$=\int P(Y_k|I_k=i,\mu_i^{\text{old}},\sigma_i^{\text{old}})\log((2\pi)^{-P/2}|\Sigma_i|^{-1/2})$$

$$\exp(-\frac{1}{2}(Y_k-\mu_i)^T\Sigma_i^{-1}(Y_k-\mu_i))dY_k$$

$$=-\frac{1}{2}\log((2\pi)^P|\Sigma_i|)-\frac{1}{2}\int P(Y_k|I_k=i,\mu_i^{\text{old}},\Sigma_i^{\text{old}})\times$$

$$(Y_k-\mu_i)^T\Sigma_i^{-1}(Y_k-\mu_i)dY_k$$

Using the properties of the expected value of the quadratic form, the integral can be calculated as

$$=-\frac{1}{2}\log((2\pi)^P|\Sigma_i|)-\frac{1}{2}(tr(\Sigma_i^{-1}\Sigma_i^{\text{old}}) + (\mu_i^{\text{old}}-\mu_i)^T\Sigma_i^{-1}(\mu_i^{\text{old}}-\mu_i)) \tag{A1}$$

Details of the derivations for Eq. 23

$$\frac{\partial\left(\sum_{i=1}^{M}\sum_{k=t_1}^{t_\alpha}P(I_k=i|\theta^{\text{old}},C_{\text{obs}})\times\left(-\frac{1}{2}\log((2\pi)^P|\Sigma_i|)-\frac{1}{2}(Y_k-\mu_i)^T\Sigma_i^{-1}(Y_k-\mu_i)\right)\right)}{\partial\mu_i}$$

$$+\frac{\partial\left(\sum_{i=1}^{M}\sum_{k=m_1}^{m_\beta}P(I_k=i|\theta^{\text{old}},C_{\text{obs}})\times\left(-\frac{1}{2}\log((2\pi)^P|\Sigma_i|)-\frac{1}{2}(tr(\Sigma_i^{-1}\Sigma_i^{\text{old}})+(\mu_i^{\text{old}}-\mu_i)^T\Sigma_i^{-1}(\mu_i^{\text{old}}-\mu_i))\right)\right)}{\partial\mu_i}=0$$

Using the derivative properties of the quadratic form we obtain

$$\Rightarrow\sum_{k=t_1}^{t_\alpha}P(I_k=i|\theta^{\text{old}},C_{\text{obs}})(Y_k-\mu_i)+\sum_{k=m_1}^{m_\beta}P(I_k=i|\theta^{\text{old}},C_{\text{obs}})(\mu_i^{\text{old}}-\mu_i)=0$$

$$\Rightarrow\mu_i^{\text{new}}=\frac{\sum_{k=t_1}^{t_\alpha}Y_kP(I_k=i|\theta^{\text{old}},C_{\text{obs}})+\sum_{k=m_1}^{m_\beta}\mu_i^{\text{old}}P(I_k=i|\theta^{\text{old}},C_{\text{obs}})}{\sum_{k=1}^{N}P(I_k=i|\theta^{\text{old}},C_{\text{obs}})} \tag{A2}$$

Details of the derivations for Eq. 24

$$\frac{\partial\left(\sum_{i=1}^{M}\sum_{k=t_1}^{t_\alpha}P(I_k=i|\theta^{\text{old}},C_{\text{obs}})\times\left(-\frac{1}{2}\log((2\pi)^P|\Sigma_i|)-\frac{1}{2}(Y_k-\mu_i)^T\Sigma_i^{-1}(Y_k-\mu_i)\right)\right)}{\partial\Sigma_i}$$

$$+\frac{\partial\left(\sum_{i=1}^{M}\sum_{k=m_1}^{m_\beta}P(I_k=i|\theta^{\text{old}},C_{\text{obs}})\times\left(-\frac{1}{2}\log((2\pi)^P|\Sigma_i|)-\frac{1}{2}(tr(\Sigma_i^{-1}\Sigma_i^{\text{old}})+(\mu_i^{\text{old}}-\mu_i)^T\Sigma_i^{-1}(\mu_i^{\text{old}}-\mu_i))\right)\right)}{\partial\Sigma_i}=0$$

Using the derivative properties of the trace, determinant and inverse we obtain

$$\Rightarrow \sum_{k=t_1}^{t_\alpha} P(I_k=i|\theta^{\text{old}}, C_{\text{obs}})(-\Sigma_i^{-1}+\Sigma_i^{-1}(Y_k-\mu_i^{\text{new}})(Y_k-\mu_i^{\text{new}})^T\Sigma_i^{-1})$$

$$+\sum_{k=m_1}^{m_\beta} P(I_k=i|\theta^{\text{old}}, C_{\text{obs}})(-\Sigma_i^{-1}+\Sigma_i^{-1}\Sigma_i^{\text{old}}\Sigma_i^{-1}+\Sigma_i^{-1}(\mu_i^{\text{old}}-\mu_i^{\text{new}})(\mu_i^{\text{old}}-\mu_i^{\text{new}})^T\Sigma_i^{-1})=0$$

$$\Rightarrow (\Sigma_i)^{\text{new}} = \frac{\sum_{k=t_1}^{t_\alpha}(Y_k-\mu_i^{\text{new}})(Y_k-\mu_i^{\text{new}})^T P(I_k=i|\theta^{\text{old}}, C_{\text{obs}})}{\sum_{k=1}^{N} P(I_k=i|\theta^{\text{old}}, C_{\text{obs}})} + \frac{\sum_{k=m_1}^{m_\beta}(\Sigma_i^{\text{old}}+(\mu_i^{\text{old}}-\mu_i^{\text{new}})(\mu_i^{\text{old}}-\mu_i^{\text{new}})^T)P(I_k=i|\theta^{\text{old}}, C_{\text{obs}})}{\sum_{k=1}^{N} P(I_k=i|\theta^{\text{old}}, C_{\text{obs}})} \quad \text{(A3)}$$

Manuscript received July 22, 2014, and revision received Sep. 23, 2014.