AIChE

Multiway Discrete Hidden Markov Model-Based Approach for Dynamic Batch Process Monitoring and Fault Classification

Jie Yu
Dept. of Chemical Engineering, McMaster University, Hamilton, ON, Canada L8S 4L7

DOI 10.1002/aic.12794
Published online November 2, 2011 in Wiley Online Library (wileyonlinelibrary.com).

A new multiway discrete hidden Markov model (MDHMM)-based approach is proposed in this article for fault detection and classification in complex batch or semibatch process with inherent dynamics and system uncertainty. The probabilistic inference along the state transitions in MDHMM can effectively extract the dynamic and stochastic patterns in the process operation. Furthermore, the used multiway analysis is able to transform the three-dimensional (3-D) data matrices into 2-D measurement-state data sets for hidden Markov model estimation and state path optimization. The proposed MDHMM approach is applied to fed-batch penicillin fermentation process and compared to the conventional multiway principal component analysis (MPCA) and multiway dynamic principal component analysis (MDPCA) methods in three faulty scenarios. The monitoring results demonstrate that the MDHMM approach is superior to both the MPCA and MDPCA methods in terms of fault detection and false alarm rates. In addition, the supervised MDHMM approach is able to classify different types of process faults with high fidelity. © 2011 American Institute of Chemical Engineers *AIChE J,* 58: 2714–2725, 2012
Keywords: batch process, fault detection, fault classification, discrete hidden Markov model, multiway analysis, system uncertainty, dynamic randomness, penicillin fermentation process

Introduction

Batch process is increasingly important as it is being widely applied to chemical, pharmaceutical, semiconductor and material industries for high-throughput, high-yield, and high-value-added products.[1–5] Compared to the continuous process in petrochemical industry, batch process is of higher complexity due to several reasons: (i) the steady-state condition that can be assumed in continuous process is usually invalid in batch process operation and thereby system dynamics have to be taken into account; (ii) there often exists strong nonlinearity in batch process so that the linearized treatment for continuous process does not apply any more; (iii) the batch-to-batch variations may be quite significant, and thus the process uncertainty cannot be ignored; (iv) the batch process sensitivity to small perturbations can be very high and then the product yield and quality would be substantially impacted. The batch process monitoring is very critical in industry because it ensures safe and stable operation of process, maintains high product quality and yield, minimizes production wastes by reducing rejected batches and increase the overall economic profits.[6–8] Early and accurate detection of process abnormalities can effectively prevent any serious production incidents, proactively correct process and measurement faults, and rapidly restore faulty operation back to normal status.[9–14]

In literature, plenty of study has been performed to apply multivariate statistical techniques to batch process monitoring. The commonly used methods include the extension of principal component analysis (PCA) and partial least squares (PLS) through multiway analysis for fault detection and diagnosis in batch or semi-batch process.[15–17] The multiway PCA and PLS approaches are essentially unsupervised techniques, which require only the normal operating data to build the process model and then identify the faulty operation from monitored batches. Though the data unfolding can deal with the batch dynamics along the time trajectory, the underlying assumption of both methods requires the unfolded data matrix to follow multivariate Gaussian distribution approximately so that the confidence limits of the detection statistics can hold. Furthermore, the conventional PCA and PLS approaches are restricted to the linear process and lack of capability in handling the nonlinearity that can be significant in certain kinds of batch process. Though the nonlinear type of PCA and PLS algorithms based upon neural networks or kernel functions have been developed,[18,19] they are still unable to effectively handle the non-Gaussianity that is often present in batch process. Meanwhile, dynamic principal component analysis (DPCA) and dynamic partial least squares (DPLS)-based monitoring approaches have been developed and applied to continuous and batch processes.[17,20] The time-lag shift method is used to explore the dynamic behavior in the process data so that the disturbance sources of dynamic systems can be detected and isolated. However, such dynamic extension based on time lags may lead to excessively large number of variables in condition monitoring. Therefore, a subspace identification-based monitoring approach is proposed for dynamic processes with a reduced set of state variables, and the error in variable identification is integrated with PCA to extract the significant variations in process variables.[21] Both the DPCA/DPLS and

Correspondence concerning this article should be addressed to J. Yu at jieyu@mcmaster.ca.

the subspace approaches rely on the linear dynamic model structure, though the latter has more concise model and provides clear monitoring results. More recently, various kinds of dynamic model structure-based monitoring methods have been extended and applied to chemical processes.[22–24] Other types of unsupervised monitoring approaches that are proposed recently include the multiway independent component analysis (MICA) and multiway Gaussian mixture model (MGMM).[25–28] The MICA approach uses the higher-order statistics to characterize the non-Gaussian features in batch process and identifies the statistically independent components for on-line fault detection. The MGMM method, conversely, assumes that the unfolded data consist of multiple Gaussian components with different means and covariances under certain prior probabilities. This approach is able to precisely capture the multimodality within multiphase batch process, and thus isolate the abnormal events from the normal operation with small errors. The other class of process monitoring techniques are supervised methods that require both the normal and faulty operating data to build the model and identify the separating plane between normal and faulty clusters. One popular supervised monitoring approach is Fisher discriminant analysis, in which the optimal separating plane is searched by maximizing the between-class scattering while minimizing the within-class scattering.[29–33] Another well-known supervised monitoring method is support vector machines that maps the data samples into high-dimensional feature space through nonlinear kernel function and then identify the separating hyperplane with the largest class separation margin.[34–37]

Although a range of supervised and unsupervised monitoring approaches have been developed for fault detection and diagnosis in batch process, they are mainly established within deterministic framework and lack of stochastic feature to tackle the randomness and uncertainty in batch processes. The DPCA method is able to handle the system dynamics but the underlying process uncertainty is not specifically taken into account. Even the MGMM method is a kind of stochastic model with the estimated probabilistic distributions, it is more focused on the multi-Gaussianity of batch process due to operating phase shift but does not take into account the dynamic randomness during state transition along the time trajectories. As an alternative approach, hidden Markov model (HMM)[38,39] can estimate the probability distributions of state transitions and the probabilities of measurement outputs in dynamic process given the invisible states. Recently, the broader community of process systems engineering has witnessed a growing interest in HMM technique for various application fields including process modeling, control, and monitoring.[40–42] In practice, the process measurements should be viewed as realizations or observations of underlying stochastic processes. The abrupt changes in operating conditions, dynamic interactions, uncertain drifts, and nonstationary disturbances all pose challenges for conventional monitoring approaches. In contrast, the strong stochastic features of HMM makes it an excellent candidate to monitor the batch process with inherent dynamics and system uncertainty. The unobservable latent variables in the process can be inferred from past measurements, and then the process output probabilities can be estimated. For process monitoring, the unknown status of normal or abnormal operation and even the particular fault type can be inferred by HMM within a probabilistic framework so that the system uncertainty and dynamic variability are taken into full consideration. In this study, a new multiway discrete hidden Markov model (MDHMM)-based batch process monitoring approach is developed to evaluate the probabilities of different measurement variables with abnormal events at each sampling instant and further determine if any kind of process fault occurs in the batch operation. The proposed method can not only detect the faulty operation in batch process, but also identify which specific type of fault the abnormal operation belongs to.

The remainder of the article is organized as follows. ''Review of Hidden Markov Model'' Section introduces the HMM, its statistical properties and learning algorithm. Then the novel MDHMM-based batch process monitoring method is developed in ''Multiway Discrete HMM-Based Batch Process Monitoring Approach'' Section. ''Application Example'' Section demonstrates the application of the proposed MDHMM monitoring method in the penicillin fermentation process and compares its results with those of the conventional multiway PCA method. The conclusions of this work are summarized in ''Conclusions'' Section.

Review of HMM

HMM is a kind of probabilistic model extended from Markov chains to generate the statistically inferential information on a series of state sequences. The regular Markov model assumes that the states are directly observable, and thus the model objective is to estimate the state transition probabilities. In the HMM, however, the states become invisible while the outputs are measurable and their values depend upon the states. Therefore, the probability distribution of each state instead of its transition probability is estimated in HMM.[38]

First, consider a discrete system with N distinct states $S_1,S_2,\ldots,S_N$. The actual state of the system at time instant t is denoted by x_t ($t = 1,2,3,\ldots$). If the discrete state sequence is a Markov chain, then the probability of the current state only relies upon the state at the previous time instant and independent of the earlier states, that is

$$P\{x_t = S_j | x_{t-1} = S_i,\ x_{t-2} = S_k, \ldots\} = P\{x_t = S_j | x_{t-1} = S_i\} \tag{1}$$

Thus, the state transition probability matrix $A = [a_{ij}]$ ($1 \leq i,j \leq N$) can be expressed as

$$a_{ij} = P\{x_t = S_j | x_{t-1} = S_i\} \tag{2}$$

with $\sum_{j=1}^{N} a_{ij} = 1$.

In the above Markov model, each state x_t corresponds to an observable measurement or event and the objective is to estimate the state transition probabilities. However, its assumption of observable states is often invalid in practice, and then the idea of HMM has been proposed to handle the unobservable (hidden) states that are reflected through a series of measurements. The major components of HMM include

i. the total number of states, N, in the model;

ii. the number of observable measurement variables $\{y_1,y_2,\ldots,y_M\}$ at each state in the model;

iii. the state transition probability distribution $a_{ij} = P\{x_t = S_j | x_{t-1} = S_i\}$;

iv. the measurement probability distribution function $B = \{b_j(k)\}$ in state j as

$$b_j(k) = p\{y_{k,t}|x_t = S_j\}\ (1 \le j \le N, 1 \le k \le M) \quad (3)$$

where $y_{k,t}$ represents the value of the k-th measurement variable at sampling time t.

v. The probability distribution function π_i of the initial state x_1 as

$$\pi_i = p\{x_1 = S_i\}\ (1 \le i \le N) \quad (4)$$

As listed above, the three main components of a HMM are the state transition probability matrix A, the measurement probability distribution matrix B and the initial state probability distribution π. Then, the functional form of HMM can be expressed as

$$H = f(A, B, \pi) \quad (5)$$

In a HMM, the following three basic problems need to be solved:

(A) Evaluation: with the following measurement sequence

$$\begin{aligned} Y &= \{Y_1, Y_2, \ldots, Y_T\} \\ &= \{\{y_{1,1}, y_{2,1}, \ldots, y_{M,1}\}, \\ &\quad \{y_{1,2}, y_{2,2}, \ldots, y_{M,2}\}, \ldots, \{y_{1,T}, y_{2,T}, \ldots, y_{M,T}\}\} \end{aligned}$$

how can one estimate the probability of measurement sequence $P\{\{Y_1,Y_2,\ldots,Y_T\}|H\}$ given the HMM H?

(B) Recognition: with the above measurement sequence and model H, how can one determine a corresponding optimal state sequence $\{x_1,x_2,\ldots,x_T\}$ that explains the observable measurements in a best way?

(C) Training: with the measurement sequence, how can one tune the model H to maximize the measurement sequence probability $P\{\{Y_1,Y_2,\ldots,Y_T\}|H\}$?

The first problem is actually equivalent to assessing the fitness between a given model and the sequence of measurements. Then, the second problem is defined to discover the "hidden" part of the model, which corresponds to the unobservable state sequence. As for the third problem, its objective is trying to optimize the model parameters so that the measurement sequence can be best described.

Various kinds of solutions have been proposed to address the above three problems.[38] For the first problem, a direct solution is to enumerate every possible state sequence $X = \{x_1,x_2,\ldots,x_T\}$, and then calculate the conditional probability of a measurement sequence Y. However, the computation load of such brute-force search method is too expensive to be feasible as the magnitude of calculations is in the order of $2TN^T$. As a more efficient approach to the first problem, the forward–backward procedure is available to calculate the measurement sequence probability with much less computation effort.[43] First define a forward variable $\alpha_t(i)$ as

$$\alpha_t(i) = P\{\{Y_1, Y_2, \ldots, Y_t\}, x_t = S_i|H\} \quad (6)$$

which can be solved by the inductive steps as follows

1. Initialization

$$\alpha_1(i) = \pi_i b_i(Y_1) \quad (7)$$

with $1 \le i \le N$.

2. Induction

$$\alpha_{t+1}(j) = \left[\sum_{j=1}^{N} \alpha_t(i)a_{ij}\right] b_j(Y_{t+1}) \quad (8)$$

where $1 \le t \le T - 1$ and $1 \le j \le N$.

3. Termination

$$P(Y|H) = \sum_{i=1}^{N} \alpha_T(i) \quad (9)$$

Similarly, a backward variable can be defined as

$$\beta_t(i) = P(Y_{t+1}, Y_{t+2}, \ldots, Y_T|x_t = S_i, H) \quad (10)$$

which corresponds to the measurement sequence probability since sampling time $t + 1$ until the end. The above backward variable can be solved in a similar way as the forward variable through the following steps:

1. Initialization

$$\beta_T(i) = 1 \quad (11)$$

2. Induction

$$\beta_t(i) = \sum_{j=1}^{N} a_{ij} b_j(Y_{t+1}) \beta_{t+1}(j) \quad (12)$$

where $t = T - 1, T - 2,\ldots,1$, and $1 \le i \le N$.

For the second problem, there are several different criteria on "optimality" so that the solution is subject to change. The most widely used criterion is to find the single best state sequence so as to maximize the probability $P(X|Y,H)$ with $X = \{x_1,x_2,\ldots,x_T\}$. A dynamic programming-based method, namely Viterbi algorithm, has been proposed to find the best state sequence with the above optimality.[44] First, define the following probabilistic function

$$\delta_t(i) = \max_{x_1,x_2,\ldots,x_{t-1}} P\{x_1 x_2 \ldots x_t = S_i, Y_1 Y_2 \ldots Y_t|H\} \quad (13)$$

which denotes the highest probability of a single state sequence at time t ending in state S_i. The probabilistic inference can be established as

$$\delta_{t+1}(j) = [\max_i \delta_t(i)a_{ij}]b_j(Y_{t+1}) \quad (14)$$

The best state sequence can be found through the path backtracking to keep track of the maximized argument for each t and j. In Viterbi training, the solution convergence is guaranteed as the algorithm does not converge to the parameter set with maximized likelihood under given observation sequence but rather the parameter set with locally maximized contribution to the likelihood from the most probable observation sequence.

The last problem to adjust the model parameters for maximizing the measurement sequence probability is more difficult than the other two. There is no analytical and optimal way to solve for the model. In practice, a model $H = f(A,B,\pi)$ is chosen to locally maximize $P(Y|H)$ using an iterative procedure like expectation–maximization (EM)

algorithm or gradient technique. A probability variable, $\phi_t(i,j)$, is defined as

$$\phi_t(i,j) = P(x_t = S_i, x_{t+1} = S_j | Y, H) \tag{15}$$

which represents the probability of being in state S_i at time t and in state S_j at time $t+1$ and can be rewritten as

$$\phi_t(i,j) = \frac{\alpha_t(i) a_{ij} b_j(Y_{t+1}) \beta_{t+1}(j)}{\sum_{i=1}^{N} \sum_{j=1}^{N} \alpha_t(i) a_{ij} b_j(Y_{t+1}) \beta_{t+1}(j)} \tag{16}$$

Then the reestimation formulas are given by

$$\bar{\pi}_i = \sum_{j=1}^{N} \phi_1(i,j) \tag{17}$$

$$\bar{a}_{ij} = \frac{\sum_{t=1}^{T-1} \phi_t(i,j)}{\sum_{t=1}^{T-1} \sum_{j=1}^{N} \phi_t(i,j)} \tag{18}$$

and

$$\bar{b}_j(k) = \frac{\sum_{Y_t = y_k} \sum_{j=1}^{N} \phi_t(i,j)}{\sum_{t=1}^{T} \sum_{j=1}^{N} \phi_t(i,j)} \tag{19}$$

The maximum likelihood estimate of the HMM can be obtained by iterative estimation, though the algorithm leads to local maxima.

Multiway Discrete HMM-Based Batch Process Monitoring Approach

Different from continuous process, batch process has inherent system dynamics along the time trajectory so that the steady-state assumption does not apply. Furthermore, batch process such as the complex bioprocess is often of strong stochastic property during the dynamic transitions due to various reasons such as inherent uncertainty, phase shifts, operation perturbations and unmeasured disturbances. In this section, a MDHMM is used to monitor the dynamic batch process in a stochastic fashion. The Markov property of HMM applies if the sampling frequency of batch process is slower than the underlying state dynamics. Thus, the probability distribution of future states rely upon only the present instead of past states. In practice, the batch sampling interval is usually longer than the hidden state duration so that the Markov property can approximately hold.

First, define the state variable x as the operating status of batch process and there exist total N different states $\{S_1,S_2,\ldots,S_N\}$, which correspond to the normal operation and all $N-1$ kinds of potential faults. Therefore, the N discrete states in a batch process can be defined as follows

$$\begin{aligned} &S_1 : \text{normal batch operation} \\ &S_2 : \text{the first type of operation fault} \\ &S_3 : \text{the second type of operation fault} \\ &\vdots \\ &S_N : \text{The } (\mathrm{N}-1)\text{th type of operation fault} \end{aligned} \tag{20}$$

At any time instant t, the value of state variable x_t can be one of the N hidden states as $x_t \in \{S_1,S_2,\ldots,S_N\}$. Further assume a batch process has M measurement variables, T sampling instants and L batches. Then, all the available measurements $\{y_1,y_2,\ldots,y_M\}$ in the process are included as the observable measurement variables to build HMM. The collected training data form a three-way observation matrix as $Y = \{y_{i,j}^{(l)}\}$, where $1 \le i \le M$, $1 \le j \le T$, and $1 \le l \le L$. Rearrange the three-way data matrix in the format of sequences as

$$\{\underbrace{\{y_{1,1}^{(l)}, y_{2,1}^{(l)}, \ldots, y_{M,1}^{(l)}\}}_{t=1}, \underbrace{\{y_{1,2}^{(l)}, y_{2,2}^{(l)}, \ldots, y_{M,2}^{(l)}\}}_{t=2}, \ldots, \underbrace{\{y_{1,T}^{(l)}, y_{2,T}^{(l)}, \ldots, y_{M,T}^{(l)}\}}_{t=T}\} \tag{21}$$

for the l-th collected batch. The above data sequence can be viewed as T sequential blocks and each of them contains the measurement data at the same sampling instant. Then, an unfolded two-dimensional data (2-D) matrix can be expressed as follows

$$\tilde{Y} = \begin{bmatrix} Y^{(l=1)} \\ Y^{(l=2)} \\ \vdots \\ Y^{(l=L)} \end{bmatrix} = \left[\begin{array}{ccc|ccc|c|ccc} y_{1,1}^{(1)} & \cdots & y_{M,1}^{(1)} & y_{1,2}^{(1)} & \cdots & y_{M,2}^{(1)} & \cdots & y_{1,T}^{(1)} & \cdots & y_{M,T}^{(1)} \\ y_{1,1}^{(2)} & \cdots & y_{M,1}^{(2)} & y_{1,2}^{(2)} & \cdots & y_{M,2}^{(2)} & \cdots & y_{1,T}^{(2)} & \cdots & y_{M,T}^{(2)} \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots \\ y_{1,1}^{(L)} & \cdots & y_{M,1}^{(L)} & y_{1,2}^{(L)} & \cdots & y_{M,2}^{(L)} & \cdots & y_{1,T}^{(L)} & \cdots & y_{M,T}^{(L)} \end{array}\right] \tag{22}$$

The corresponding state matrix of the training set is given by

$$X = \begin{bmatrix} x_1^{(l=1)} & x_2^{(l=1)} & \cdots & x_T^{(l=1)} \\ x_1^{(l=2)} & x_2^{(l=2)} & \cdots & x_T^{(l=2)} \\ \vdots & \vdots & \ddots & \vdots \\ x_1^{(l=L)} & x_2^{(l=L)} & \cdots & x_T^{(l=L)} \end{bmatrix} \tag{23}$$

The unfolded training data set $\{X, \tilde{Y}\}$ should include both the normal operation and the $N-1$ faulty events. In

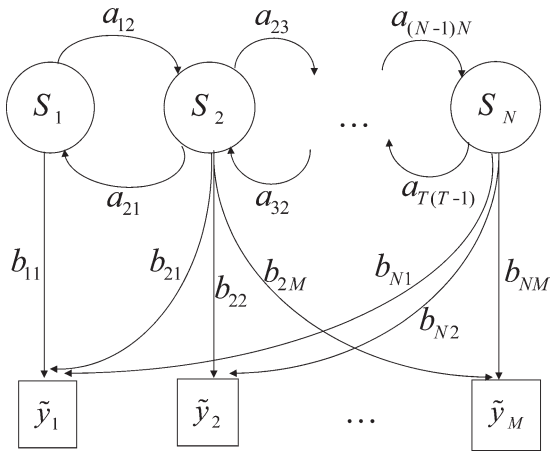

Figure 1. Illustration of DHMM estimated from batch process data.

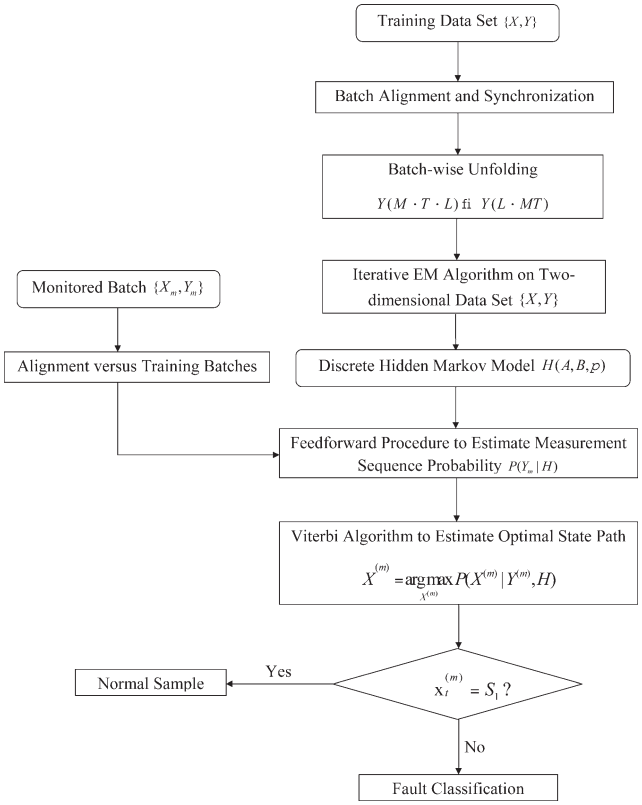

Figure 2. Schematic diagram of the MDHMM-based batch process monitoring and fault classification approach.

practice, the training set can be collected from the process data historian with all kinds of faulty scenarios.

To deal with batches with unequal lengths or unaligned time series, the dynamic time warping (DTW) technique is used in data preprocessing step for batch synchronization and data reconciliation.[45] During batch operation, the process status is assumed to follow Markov property, that is, the normal or faulty operation status at current time instant only depends upon the status at the previous sampling time but has nothing to do with the older statuses. With the state sequences X and observation sequences $\tilde{\boldsymbol{Y}}$ in training set, the HMM H can be first estimated using the iterative EM algorithm, as illustrated in Figure 1. Then, for each monitored batch, its measurement data sequence is arranged as follows

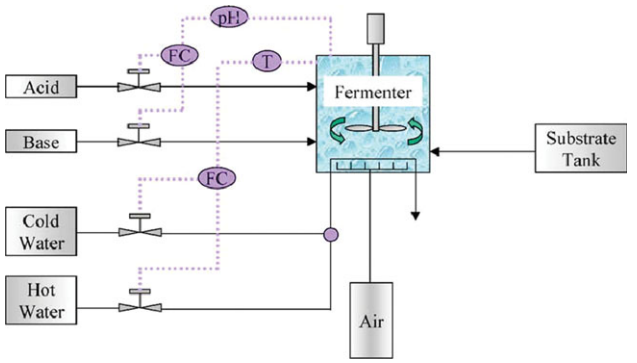

Figure 3. Flow diagram of fed-batch penicillin fermentation process.

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

Table 1. Monitored Variables in the Fed-Batch Penicillin Fermentation Process

Variable No.	Monitored Variables
1	Aeration rate (L/h)
2	Agitator power (W)
3	Substrate feed rate (L/h)
4	Substrate feed temperature (K)
5	Substrate concentration (g/L)
6	pH
7	Dissolved oxygen concentration (g/L)
8	Carbon dioxide concentration (g/L)
9	Biomass concentration (g/L)
10	Penicillin concentration (g/L)
11	Fermentor temperature (K)
12	Cooling water flow rate (L/h)
13	Generated heat (kcal)
14	Acid flow rate (L/h)
15	Base flow rate (L/h)
16	Culture volume (L)

$$Y^{(m)} = \{\{y_{1,1}^{(m)}, y_{2,1}^{(m)}, \ldots, y_{M,1}^{(m)}\}, \{y_{1,2}^{(m)}, y_{2,2}^{(m)}, \ldots, y_{M,2}^{(m)}\}, \ldots, \{y_{1,T}^{(m)}, y_{2,T}^{(m)}, \ldots, y_{M,T}^{(m)}\}\} \quad (24)$$

where the monitored batch has been synchronized vs. the training batches through DTW technique. With the dynamic programming based Viterbi algorithm, the following optimal state transition path of the highest probability to account for the monitored observations can be estimated as

$$\hat{X}^{(m)} = \arg\max_{X^{(m)}} P(X^{(m)}|Y^{(m)},H) = \{\hat{x}_1^{(m)}, \hat{x}_2^{(m)}, \ldots, \hat{x}_T^{(m)}\} \quad (25)$$

where the optimal state of the monitored batch at every sampling instant t can be determined as $\hat{x}_t^{(m)} = S_n$ $(1 \le n \le N)$. Thus, the abnormal events along the time trajectory of the monitored batch can be detected and the fault types are further identified according to the estimated optimal state sequence.

The procedure of the MDHMM-based batch process monitoring approach is summarized below and the schematic diagram of this method is also plotted in Figure 2.

i. Collect training data set $\{X,Y\}$ with a 3-D $M \times T \times L$ observation matrix Y and the corresponding 2-D $L \times T$ state matrix X that indicate the operation status of different training batches along the time trajectory.

Table 2. Ranges of Initial Conditions and Set Points of Operation Parameters in the Fed-Batch Penicillin Fermentation Process

Initial condition	
Substrate concentration	13–17 (g/L)
Dissolved oxygen concentration	1.05–1.25 (g/L)
Biomass concentration	0.05–0.15 (g/L)
Penicillin concentration	0 (g/L)
Culture volume	99–104 (L)
Carbon dioxide concentration	0.5–1.0 (g/L)
pH	4.5–5.5
Fermentor temperature	293–303 (K)
Generated heat	0 (kcal)
Set point	
Aeration rate	8–9 (L/h)
Agitator power	28.5–31.5 (W)
Substrate feed flow rate	0.035–0.048 (L/h)
Substrate feed temperature	294–298 (K)
Fermentor temperature	296–300 (K)
pH	4.8–5.2

Table 3. List of Three Types of Faults in the Fed-Batch Penicillin Fermentation Process

Fault No.	Fault Description
1	Drift fault: linearly increased aeration rate at the slope of 0.05 (L/h)/h
2	Step fault: step-decreased glucose feed flow by 0.02 L/h
3	Drift fault: linearly decreased agitator power at the slope of 0.12 W/h

Table 4. Summary of Three Abnormal Operation Scenarios in the Fed-Batch Penicillin Fermentation Process

Scenario No.	Description of abnormal Events
1	Drift fault on aeration rate from the 100th to 200th hour Step fault on glucose feed flow from the 280th to 350th hour
2	Drift fault on agitator power from the 110th to 160th hour Drift fault on aeration rate from the 250th to 340th hour
3	Drift fault on agitator power from the 80th to 140th hour Step fault on glucose feed flow from the 200th to 240th hour Drift fault on aeration rate from the 290th to 380th hour

ii. Conduct the data alignment and synchronization across the training batches with unequal time length through DTW technique.

iii. Perform batch-wise unfolding on the 3-D observation matrix Y and convert it to a 2-D $L \times MT$ matrix $\tilde{Y}$ as in Eq. 22.

iv. With data set $\{X, \tilde{Y}\}$, perform iterative EM algorithm to estimate the maximum likelihood solution of the discrete hidden Markov model (DHMM) $H = f(A,B,\pi)$.

v. For each of the monitored batches, align and synchronize the time series vs. the training batches.

vi. Arrange the aligned observation data in the format of $1 \times MT$ sequence array $Y^{(m)}$ as in Eq. 24.

vii. Run forward–backward procedure to compute the measurement sequence probability $P(Y^{(m)}|H)$ given the estimated HMM H.

viii. Apply Viterbi algorithm to estimate the optimal state path $\hat{X}^{(m)}$ with the highest conditional probability $P(X^{(m)}|Y^{(m)},H)$ given the estimated HMM H and the measurement data $Y^{(m)}$.

ix. At each sampling instant t of the monitored batch, determine if the operation condition is normal or faulty based on whether the estimated state $\hat{x}_t^{(m)} = S_1$ or not.

x. For all the detected faulty points in the monitored batch, the fault type can be further identified according to which of $\{S_2,S_3,\ldots,S_N\}$ the estimated state $\hat{x}_t^{(m)}$ belongs to.

It is noted that the process data in HMM are measurements of time series with inherent dynamics and the data set can be in wide range with significant uncertainty. In the batch process, the unmeasured disturbances, noises, and

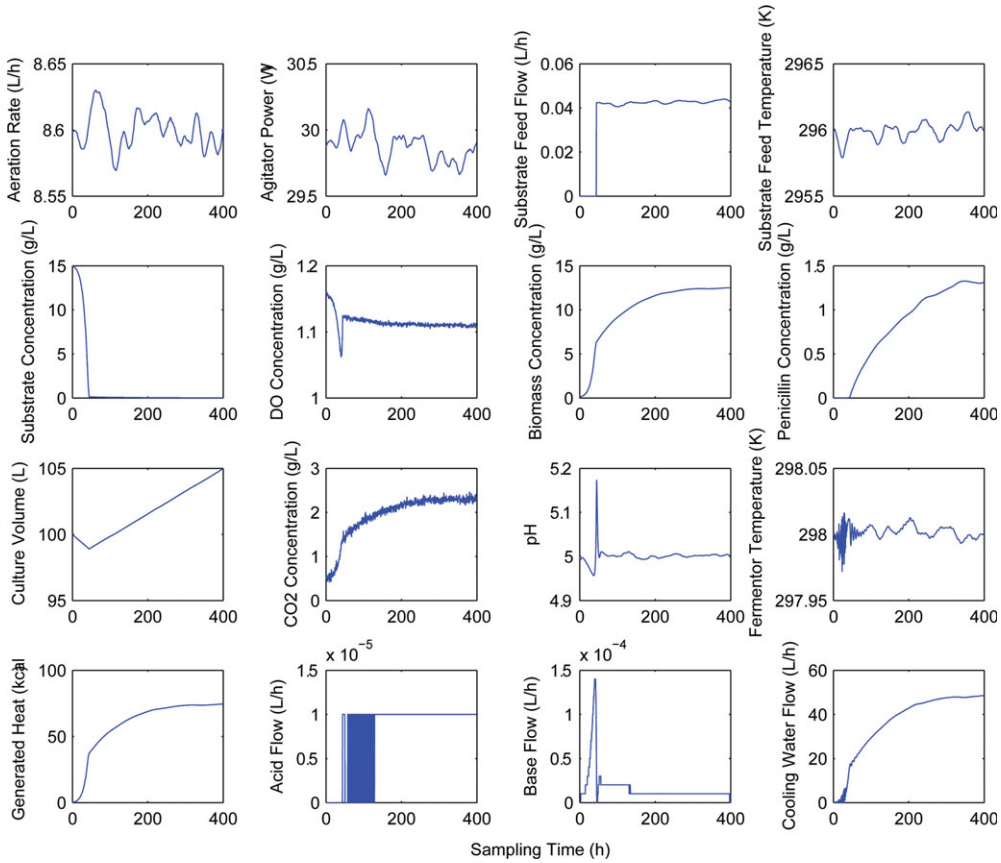

Figure 4. Normal batch data on 16 monitored variables of the fed-batch penicillin fermentation process.

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

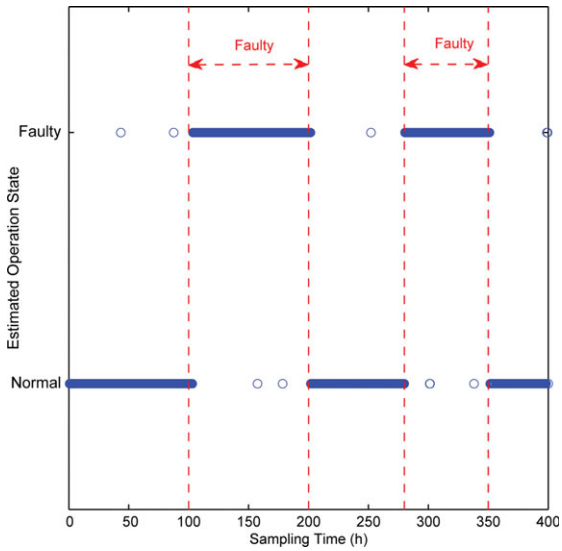

Figure 5. The first test case: Fault detection results of the proposed MDHMM approach.

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

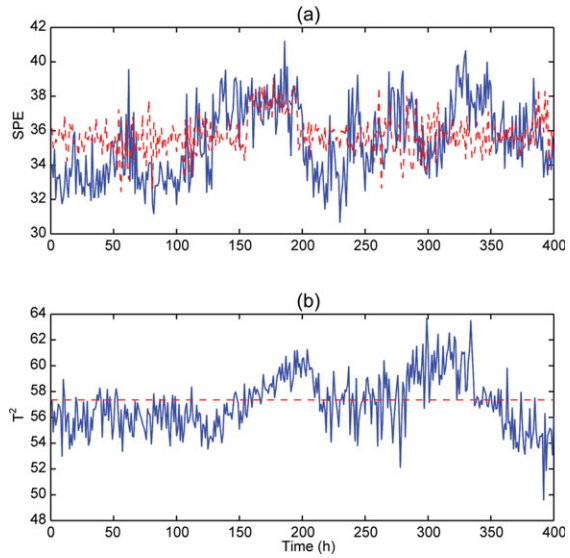

Figure 7. The first test case: Fault detection results of MDPCA based (a) SPE and (b) T^2 control charts.

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

system uncertainty can be handled by the iterative learning. In addition, the bad-quality data may be identified and distinguished from different types of process faults by the DHMM-based monitoring approach.

Application Example

Fed-batch penicillin fermentation process

The proposed MDHMM batch process monitoring approach is applied to the fed-batch penicillin fermentation benchmark process, which has been developed by process modeling, monitoring, and control group at Illinois Institute of Technology. As the formation of the target product, the antibiotic, is usually not associated with cell growth, the microorganisms is cultivated in a batch mode and then followed by a fed-batch operation to promote the synthesis of the antibiotic.[46,47] In the penicillin production process, the bioreactor is switched to the fed-batch mode of operation after 40 h of batch growth phase when the cells have entered their stationary phase. The developed first-principle models of penicillin production process include various aspects such as the cell growth, substrate consumption, antibiotic production, and carbon dioxide generation. The fed-batch penicillin cultivation process is characterized by inherent dynamics, model uncertainty, and natural disturbances. The flow

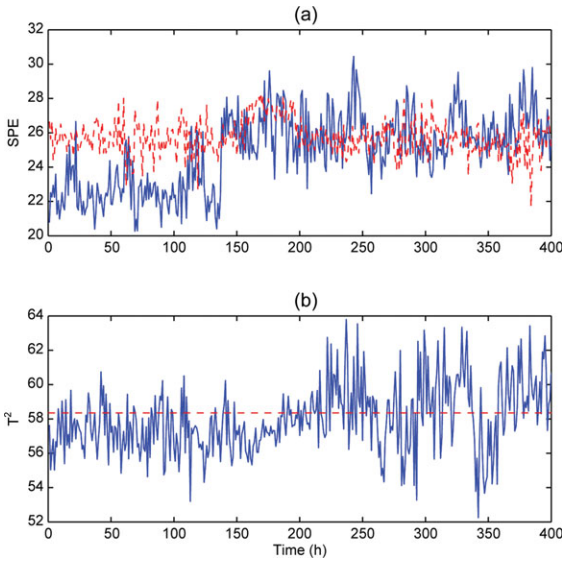

Figure 6. The first test case: fault detection results of MPCA-based (a) SPE and (b) T^2 control charts.

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

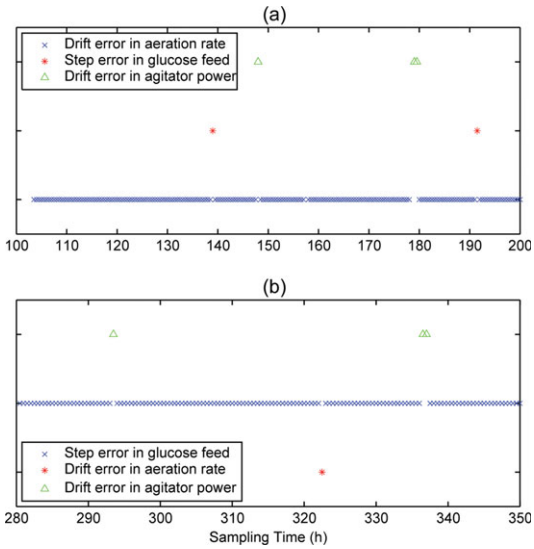

Figure 8. The first test case: fault classification results of the proposed MDHMM approach in the periods of (a) the 100th to 200th hour and (b) the 280–350 h.

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

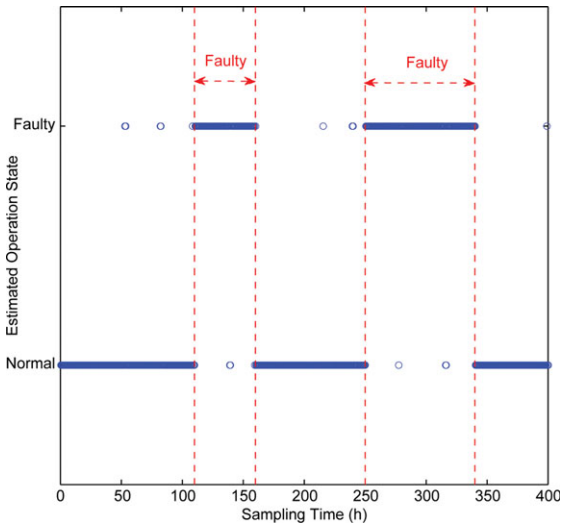

Figure 9. The second test case: Fault detection results of the proposed MDHMM approach.

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

diagram of the process is given in Figure 3 and the web-based simulator is available at http://www.chbe.iit.edu/cinar/. The fermentor is equipped with two cascade loops to control the reactor temperature and pH by manipulating the hot water/cold water flow ratio and acid/base flow ratio, respectively. Conversely, the substrate addition and aeration are operated under open-loop condition.

For the monitoring purpose, the sampling data of total 16 measurement variables are collected to form the observation matrix Y. The monitored variables are listed in Table 1 while the ranges of initial conditions and set points used to simulate the penicillin fermentation process are given in Table 2. The sampling interval of the process simulation is set to 0.5 h, and the duration of each batch is 400 h. As shown in Table 2, the batch-to-batch variations and model uncertainty

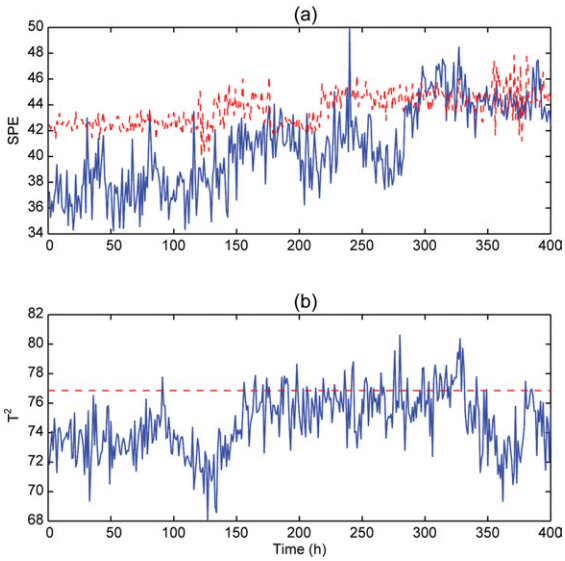

Figure 10. The second test case: fault detection results of MPCA-based (a) SPE and (b) T^2 control charts.

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

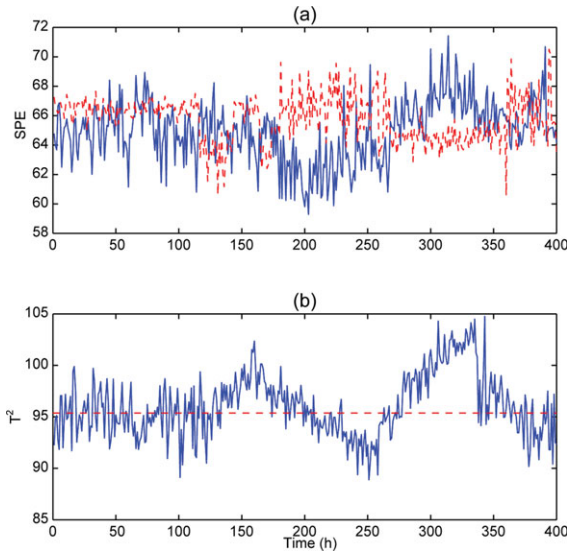

Figure 11. The second test case: fault detection results of MDPCA-based (a) SPE and (b) T^2 control charts.

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

are added to mimic the process perturbations and random features. To examine the fault detection and classification capability of the proposed approach, three types of process faults including step and drift errors are designed and listed in Table 3. The normal operation is represented by the state S_0, whereas the three kinds of faults are denoted as S_1, S_2, and S_3. With the three predefined process faults, three abnormal operation scenarios in Table 4 are further tested to assess the monitoring performance of the proposed approach vs. the conventional multiway principal component analysis (MPCA) and multiway dynamic principal component analysis (MDPCA) method.

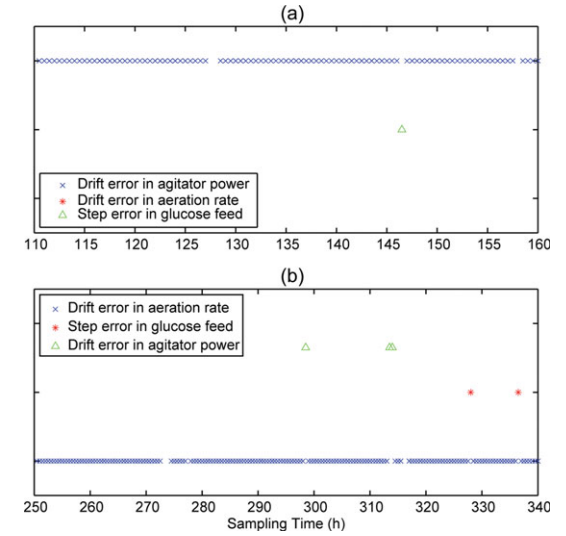

Figure 12. The second test case: fault classification results of the proposed MDHMM approach in the periods of (a) the 110th–160th hour and (b) the 250–340 h.

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

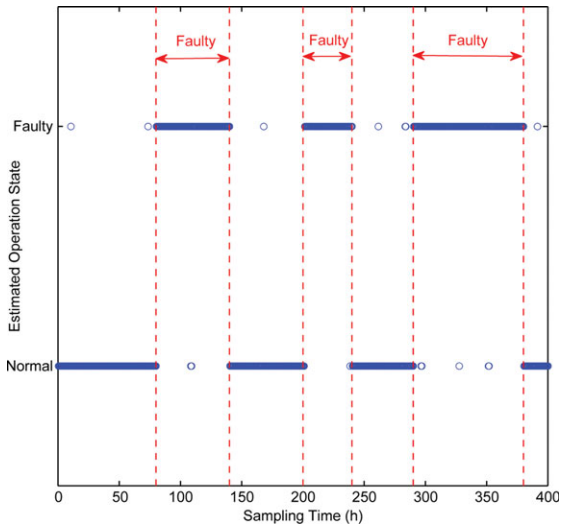

Figure 13. The third test case: Fault detection results of the proposed MDHMM approach.

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

Batch process monitoring and fault classification results

To first build the DHMM, total 120 training batches are collected and they include 60 normal batches and 20 abnormal batches from each type of process faults. The 3-D measurement matrix Y along with the corresponding operation status matrix X are used to estimate the DHMM $H(A,B,\pi)$. As a comparison, the measurement data matrix alone can be used to build the multiway PCA and DPCA models. The normal operation data on the 16-monitored variables are plotted in Figure 4. In each test scenario, 10 faulty batches are generated and the mean values of their measurement data are fed into MPCA, MDPCA, and DHMM models, respectively.

In the first test case, the drift error in aeration rate occurs from the 100th hour and lasts 100 hour. Then, the process

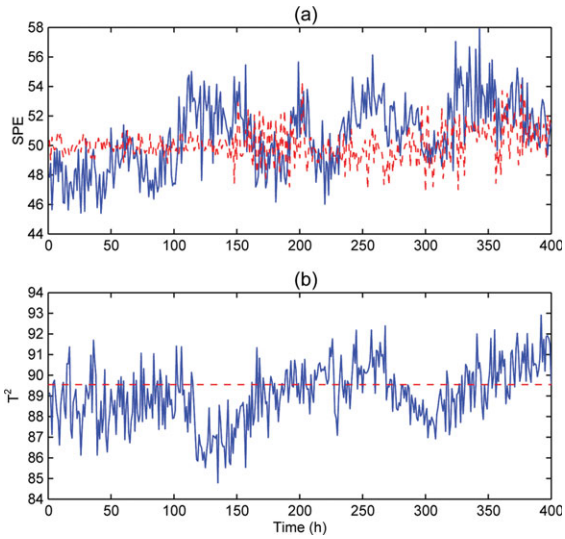

Figure 14. The third test case: fault detection results of MPCA-based (a) SPE and (b) T^2 control charts.

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

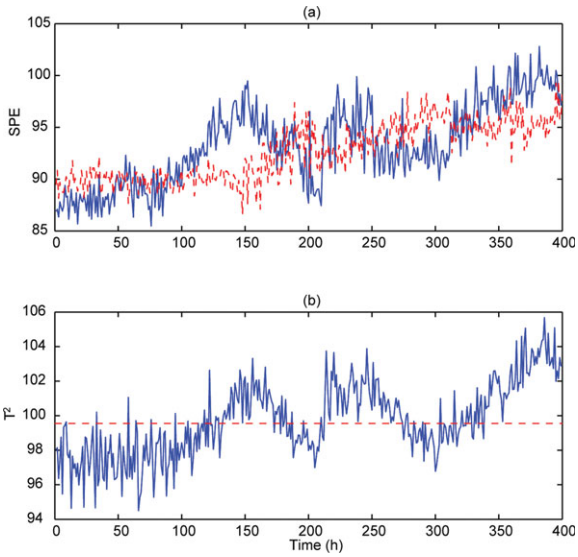

Figure 15. The third test case: fault detection results of MDPCA-based (a) SPE and (b) T^2 control charts.

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

returns to normal operation until the 280th hour when the step error in glucose feed rate starts. Such feed fault disappears from the process after 70-h duration. The fault detection results of the multiway DHMM method are shown in Figure 5. As a comparison, the MPCA and MDPCA-based SPE and T^2 control charts are depicted in Figures 6 and 7. It can be seen that most of the faulty samples in the monitored-batch are not detected in either SPE or T^2 control charts of MPCA approach. The MDPCA method has higher detection rate and fewer false alarms than those of MPCA approach. Nevertheless, both type-I and type-II errors by MDPCA method are still fairly high since 34.35% of the faulty points are unable to be detected while 42.51% of the normal points triggers false alarms incorrectly. In contrast, the MDHMM approach is able to capture the majority of abnormal points with the corresponding fault detection rate of 96.76% and only 11 among 340 faulty samples are missed. Conversely, the false alarm rate of the MDHMM approach is as low as 2.17%, which represents only 10 fault alarms being incorrectly triggered on normal samples. The poor performance of the conventional MPCA and MDPCA methods is mainly due to the strong system uncertainty and perturbations in the process. Though the MDPCA approach has the capability of handling dynamics, the process uncertainty is not specifically taken into account so that the faulty events and normal perturbations cannot be well differentiated by MDPCA approach. However, the MDHMM method is capable of discovering the dynamic patterns and its probabilistic inference on state transitions can effectively handle the stochastic feature of the process. The fault classification results of two operation periods with abnormal events are further shown in Figures 8a, b. During the faulty period from the 100th through 200th hour, only two samples are misclassified as step error in glucose feed rate and three points are wrongly categorized into the fault type of drift error in agitator power. In the abnormal period of the 280th–350th hour, one sample is incorrectly identified as drift error in aeration rate and three additional samples as drift fault in agitator power. The overall

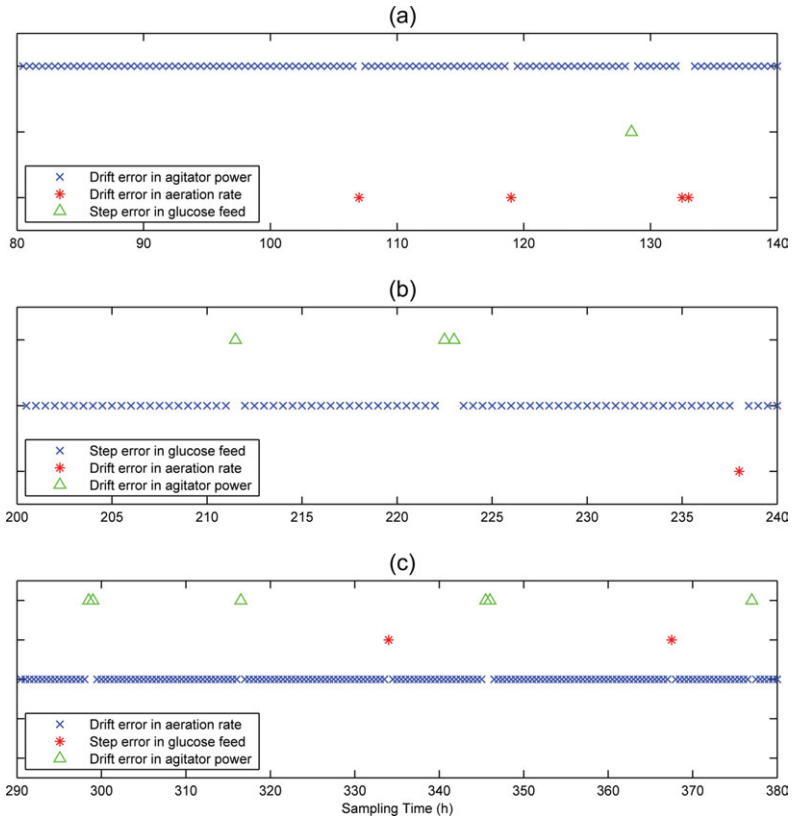

Figure 16. The third test case: fault classification results of the proposed MDHMM approach in the periods of (a) the 80th–140th hour; (b) the 200–240 h, and (c) the 290–380 h.

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

fault classification accuracy in both periods is as high as 97.35%, while the unsupervised MPCA and MDPCA methods are unable to classify multiple fault types.

In the second test scenario, two types of drift faults are applied to the process during different periods. The drift error in agitator power is first added to the fed-batch operation in the 110th hour and remains for 50 h. The other drift error in aeration rate is then activated from the 250th to the 340th hour. The rest of operation periods are under normal conditions. As shown in Figure 10a, the SPE control chart of MPCA method misses most of the faulty samples in the first abnormal period. Even in the second faulty period, the fault alarm is not triggered until almost 50 h after the drift error in aeration rate begins. Likewise, the T^2 control chart of MPCA in Figure 10b reveals that most of the abnormal samples in the first faulty period and over half of the abnormal ones in the second period are not detected. The SPE and T^2 charts of MDPCA method in Figure 11 show better detection rates and false alarm rates than those of MPCA method because the batch process dynamics are captured by MDPCA. The MHDMM approach, however, still gives the best fault detections results in Figure 9. During the two faulty periods, total six abnormal samples are undetected and lead to a high fault detection rate of 97.86%. Meanwhile, only nine normal samples are misidentified as faulty ones and cause a fairly low false alarm rate of 1.73%. The fault classification results of both abnormal periods are further plotted in Figures 12a, b. It is obvious that three samples are misclassified into the type of drift error in aeration rate and one as step error in glucose feed rate during the period of the 110th–160th hour with the occurrence of drift error in agitator power. In the second faulty period with drift error in aeration rate, two samples are mistakenly marked as step error in glucose feed rate and three as drift error in agitator power. The fault classification accuracies in both periods are 96 and 97.22%, respectively. This case further verifies the strong capacity of the MDHMM approach in fault detection and classification of dynamic batch process with significant uncertainty and randomness.

The third test case includes the step error and both drift faults, which occur in different operation periods. The drift fault in agitator power first occurs from the 80th to 140th hour followed by 60-h normal operation. Then, the step error in glucose feed flow takes place and lasts 40 h before the operation is back to normal. During the last faulty period between the 290th and the 380th hour, the drift error in aeration rate is applied. The MPCA, MDPCA, and MDHMM-based fault detection results are shown in Figures 13–15, respectively. In the SPE control chart of MPCA method, the initial detection of both drift faults in aeration rate and agitator power are delayed by over 20 h. Though the step error in glucose feed rate is captured rapidly by SPE index, the fault detection rate is still low with about 20-h faulty period completely missed. The SPE chart of MDPCA method encounter similar delay in triggering the fault alarms. The T^2 control charts of both MPCA and MDPCA methods perform even worse with large number of faulty samples undetected and normal points falsely alarmed. Nevertheless, the identified faulty states by MDHMM method match well the different fault occurrences. The overall fault detection rate across three abnormal periods is 94.74% while the false alarm rate

is as low as 2.62%. As this test scenario consists of all three types of process faults, the type-I and type-II errors of MDHMM-based fault detection are slightly higher than those of the first two cases. The fault classification results of MDHMM approach are further given in Figure 16. In the first faulty period with drift error in agitator power, only four samples are misclassified as drift error in aeration rate and one as step error in glucose feed, which result in the fault classification accuracy of 95.83%. Such high fault classification rates are consistently observed in both the second and third faulty periods, which are 95 and 95.56%, respectively. The MDHMM method is further demonstrated to be capable of handling complex faulty scenarios in dynamic batch or fed-batch operation with underlying uncertainty and stochastic behavior.

Conclusions

In this article, the DHMM is integrated with multiway analysis to monitor batch process with inherent uncertainty and randomness. The probability inference on the state and measurement variables in HMM is used to capture the stochastic nature during batch operation while the state transitions are adopted to characterize the process dynamics along the time trajectories with uncertainty. The incorporated multiway analysis can convert the 3-D data matrix from batch process into 2-D data set. The DHMM model parameters are then estimated and the state path can be further optimized with the highest conditional probability. The inferred state sequences are used to detect and classify different types of process faults.

The proposed MDHMM method is applied to three simulated faulty scenarios in a fed-batch penicillin fermentation process. The comparison of fault detection results among MPCA, MDPCA, and MDHMM methods indicates that the MDHMM approach has the highest fault detection rates with the fewest false alarms while the conventional MPCA and MDPCA methods both have worse performance due to the dynamic randomness and uncertainty in process. Furthermore, the MDHMM method is able to classify different types of process faults with high accuracy. The present work highlights that the supervised MDHMM approach is promising in dynamic batch process monitoring and fault identification.

Literature Cited

1. Wise BM, Gallagher NB, Butler SW, White DD, Barna GG. A comparison of principal component analysis, multiway principal component analysis, trilinear decomposition and parallel factor analysis for fault detection in a semiconductor etch process. *J Chemom.* 1999;13:379–396.
2. Ündey C, Çinar A. Statistical monitoring of multistage, multiphase batch processes. *IEEE Control Syst Mag.* 2002;10:40–52.
3. Lennox B, Kipling K, Glassey J, Montague G, Willis M, Hiden H. Automated production support for the bioprocess industry. *Biotechnol Prog.* 2002;18:269–275.
4. Clementschitsch F, Bayer K. Improvement of bioprocess monitoring: development of novel concepts. *Microb Cell Fact.* 2006;5:1–11.
5. Cherry G, Qin S. Multiblock principal component analysis based on a combined index for semiconductor fault detection and diagnosis. *IEEE Trans Semicond Manuf.* 2006;19:159–172.
6. Nomikos P, MacGregor JF. Monitoring of batch processes using multi-way principal component analysis. *AIChE J.* 1994;40:1361–1375.
7. Westerhuis JA, Kourti T, MacGregor J. Comparing alternative approaches for multivariate statistical analysis of batch process data. *J Chemom.* 1999;13:397–413.
8. Ündey C, Ertunç S, Çinar A. Online batch/fed-batch process performance monitoring, quality prediction, and variable-contribution analysis for diagnosis. *Ind Eng Chem Res.* 2003;42:4645–4658.
9. Wise B, Gallagher N. The process chemometrics approach to process monitoring and fault detection. *J Process Control.* 1996;6:329–348.
10. Chiang L, Russell E, Braatz R. *Fault Detection and Diagnosis in Industrial Systems. Advanced Textbooks in Control and Signal Processing. London*, Great Britain: Springer-Verlag, 2001.
11. Qin SJ. Statistical process monitoring: basics and beyond. *J Chemom.* 2003;17:480–502.
12. Venkatasubramanian V. Prognostic and diagnostic monitoring of complex systems for product lifecycle management: challenges and opportunities. *Comput Chem Eng.* 2005;29:1253–1263.
13. Yu J, Qin SJ. Statistical MIMO controller performance monitoring. I. Data driven covariance benchmark. *J Process Control.* 2008;18:277–296.
14. Yu J, Qin SJ. Statistical MIMO controller performance monitoring. II. Performance diagnosis. *J Process Control.* 2008;18:297–319.
15. Kourti T, Nomikos P, MacGregor JF. Analysis, monitoring, and fault diagnosis of batch processes using multi-block and multi-way PLS. *J Process Control.* 1995;5:277–284.
16. Lennox B, Montague G, Hiden H, Kornfeld G, Goulding P. Process monitoring of an industrial fed-batch fermentation. *Biotechnol Bioeng.* 2001;74:125–135.
17. Chen J, Liu K. On-line batch process monitoring using dynamic PCA and dynamic PLS models. *Chem Eng Sci.* 2002;57:63–75.
18. Dong D, McAvoy T. Nonlinear principal component analysis based on principal curves and neural networks. *Comput Chem Eng.* 1996;20:65–78.
19. Lee JM, Yoo CK, Lee IB. Statistical process monitoring with independent component analysis. *J Process Control.* 2004;14:467–485.
20. Ku W, Storer R, Georgakis C. Disturbance detection and isolation by dynamic principal component analysis. *Chemom Intell Lab Syst.* 1995;30:179–196.
21. Treasure R, Kruger U, Cooper J. Dynamic multivariate statistical process control using subspace identification. *J Process Control.* 2004;14:279–292.
22. Yao Y, Gao F. Batch process monitoring in score space of two-dimensional dynamic principal component analysis. *Ind Eng Chem Res.* 2007;46:8033–8043.
23. Yao Y, Gao F. Subspace identification for two-dimensional dynamic batch process statistical monitoring. *Chem Eng Sci.* 2008;63:3411–3418.
24. Zhang Y, Li Z, Zhou H. Statistical analysis and adaptive technique for dynamical process monitoring. *Chem Eng Res Des.* 2010;88:1381–1392.
25. Yoo CK, Lee DS, Vanrolleghem PA. Application of multiway ICA for on-line process monitoring of a sequencing batch reactor. *Water Res.* 2004;38:1715–1732.
26. Yoo CK, Lee JM, Vanrolleghem PA, Lee IB. On-line monitoring of batch processes using multiway independent component analysis. *Chemom Intell Lab Syst.* 2004;71:151–163.
27. Yu J, Qin SJ. Multimode process monitoring with Bayesian inference-based finite Gaussian mixture models. *AIChE J.* 2008;54:1811–1829.
28. Yu J, Qin SJ. Multiway Gaussian mixture model based multiphase batch process monitoring. *Ind Eng Chem Res.* 2009;48:8585–8594.
29. Chiang L, Russell E, Braatz R. Fault diagnosis in chemical processes using fisher discriminant analysis, discriminant partial least squares, and principal component analysis. *Chemom Intell Lab Syst.* 2000;50:243–252.
30. He QP, Qin SJ, Wang J. A new fault diagnosis method using fault directions in Fisher discriminant analysis. *AIChE J.* 2005;51:555–571.
31. Zhang X, Yan W, Zhao X, Shao H. Nonlinear biological batch process monitoring and fault identification based on kernel fisher discriminant analysis. *Process Biochem.* 2007;42:1200–1210.
32. Yu J. Localized fisher discriminant analysis based complex chemical process monitoring. *AIChE J.* 2011;57:1817–1828.
33. Yu J. nonlinear bioprocess monitoring using multiway kernel localized fisher discriminant analysis. *Ind Eng Chem Res.* 2011;50:3390–3402.
34. Chiang L, Kotanchek M, Kordon A. Fault diagnosis based on Fisher discriminant analysis and support vector machines. *Comput Chem Eng.* 2004;28:1389–1401.

35. Kulkarni A, Jayaraman V, Kulkarni B. Knowledge incorporated support vector machines to detect faults in Tennessee Eastman Process. *Comput Chem Eng*. 2005;29:2128–2133.
36. Yélamos I, Graells M, Puigjaner L, Escudero G. Simultaneous fault diagnosis in chemical plants using a multilabel approach. *AIChE J*. 2007;53:2871–2884.
37. Yélamos I, Escudero G, Graells M, Puigjaner L. Performance assessment of a novel fault diagnosis system based on support vector machines. *Comput Chem Eng*. 2009;33:244–255.
38. Rabiner L. Atutorial on hidden Markov models and selected applications in speech recognition. In: *Proceedings of IEEE*, 1989:257–286.
39. Rabiner L, Juang B. An introduction to hidden Markov models. *IEEE ASSP Mag*. 1986;3:4–16.
40. Chen J, Chang W. Applyingwavelet-based hidden Markov tree to enhancing performance of process monitoring. *Chem Eng Sci*. 2005;60:5129–5143.
41. Wong W, Lee J. Realistic disturbance modeling using hidden Markov models: applications in model-based process control. *J Process Control*. 2009;19:1438–1450.
42. Chen J, Hsu TJ, Chen CC, Cheng YC. Online predictive monitoring using dynamic imaging of furnaces with the combinational method of multiway principal component analysis and hidden Markov model. *Ind Eng Chem Res*. 2011;50:2946–2958.
43. Baum L, Egon J. An inequality with applications to statistical estimation for probabilistic function of a Markov process and to a model for ecology. *Bull Am Meteorol Soc*. 1967;73:360–363.
44. Viterbi A. Error bounds for convolutional codes and an asymptotically optimal decoding algorithm. *IEEE Trans Informat Theory*. 1967; pp. 260–269.
45. Kassidas A, MacGregor J, Taylor P. Synchronization of batch trajectories using dynamic time warping. *AIChE J*. 1998;44:864–875.
46. Birol I, Ündey C, Birol G, Tatara E, Çinar A. A web-based simulator for penicillin fermentation. *Int J Eng Simul*. 2001;2:24–30.
47. Birol G, Ündey C, Çinar A. A modular simulation package for fed-batch fermentation: penicillin production. *Comput Chem Eng*. 2002;26:1553–1565.

Manuscript received Jan. 31, 2011, and revision received Sept. 30, 2011.