Multimode Process Monitoring Using Variational Bayesian Inference and Canonical Correlation Analysis

Qingchao Jiang, *Member, IEEE*, and Xuefeng Yan

***Abstract*— Industrial processes generally have various operation modes, and fault detection for such processes is important. This paper proposes a method that integrates a variational Bayesian Gaussian mixture model with canonical correlation analysis (VBGMM-CCA) for efficient multimode process monitoring. The proposed VBGMM-CCA method maximizes the advantage of VBGMM in automatic mode identification and the superiority of CCA in local fault detection. First, VBGMM is applied to unlabeled historical process data to determine the number of operation modes and cluster the data in each mode. Second, local CCA models that explore input and output relationships are established. Fault detection residuals are generated in each local CCA model, and monitoring statistics are derived. Finally, a Bayesian inference probability index that integrates monitoring results from all local models is developed to increase the monitoring robustness. The effectiveness of the proposed monitoring scheme is verified through experimental studies on a numerical example and the multiphase batch-fed penicillin fermentation process.**

***Note to Practitioners*—Process monitoring is important in guaranteeing process safety and improving product quality. Large amounts of unlabeled process data with multiple operation modes generally exist in industrial processes. Labeling these data is difficult or costly. Hence, this paper presents a VBGMM-CCA method for monitoring multimode processes. The key advantage of the proposed method is that it automatically identifies the number of operation modes in historical data and clusters the data. Then, local CCA monitors are established to model the process input and output relationships. During online monitoring, the running-on operation mode is identified through a density function, and the process status is evaluated by the corresponding CCA monitor. A probabilistic monitoring index is also developed to increase the robustness of the monitoring. In comparison with the results of conventional methods, the monitoring results of the proposed approach are more reliable and informative because the process status and the type of the detected fault are presented.**

***Index Terms*— Canonical correlation analysis (CCA), fault detection, multimode process monitoring, variational Bayesian Gaussian mixture model (GMM).**

I. Introduction

PROCESS monitoring is crucial in maintaining long-term operations and is currently eliciting increasing attention [1]–[4]. A modern industrial process may be highly complex, and the corresponding mathematical process model may be unavailable. Meanwhile, abundant process data that contain meaningful process information are available [5]–[7]. Data-driven methods have become well-known, and the multivariate statistical process monitoring (MSPM) method is the most widely used among these methods [8]–[11]. Principal component analysis (PCA), partial least squares (PLS), and canonical correlation analysis (CCA) are basic MSPM methods. PCA models are used for extracting the major variance information of process data and are generally used for removing collinearity [8], [12]–[14]. PLS is commonly utilized in quality-related or key performance indicator-oriented process monitoring [15], [16]. A MATLAB toolbox that summarizes various PLS extensions is developed in [11]. CCA is an efficient multivariate analysis technique for characterizing the correlation between two sets of random variables and has been eliciting increasing attention recently [17]–[20].

CCA is generally used for process monitoring via two approaches. The first one is the use of CCA as a feature extraction and dimension reduction technique; the features are then examined for process status identification [21]–[24]. The second approach is the use of CCA to characterize two sets of process variables and generate fault detection residuals for monitoring [17], [19]. In [17], CCA was used to model the process input and output (the manipulated and the measured variables, respectively) and generate fault detection residuals. In [19], CCA was utilized to characterize the correlation between different operation units for distributed monitoring. The optimality of the approach was validated. The missing data problem in distributed monitoring was addressed using a neighborhood modeling strategy [25]. Similarly, CCA was used for non-Gaussian and batch process monitoring in [26] and [20], respectively. A mixture CCA was

Manuscript received December 19, 2018; revised January 18, 2019; accepted January 31, 2019. Date of publication February 26, 2019; date of current version October 4, 2019. This paper was recommended for publication by Associate Editor H. Wang and Editor F.-T. Cheng upon evaluation of the reviewers' comments. This work was supported in part by the National Natural Science Foundation of China under Grant 61603138 and Grant 21878081, in part by Shanghai Pujiang Program under Grant 17PJD009, in part by Fundamental Research Funds for the Central Universities under Grant 222201717006 and Grant 222201714027, and in part by the Programme of Introducing Talents of Discipline to Universities (the 111 Project) under Grant B17017. *(Corresponding author: Xuefeng Yan.)*

The authors are with the Key Laboratory of Advanced Control and Optimization for Chemical Processes, Ministry of Education, East China University of Science and Technology, Shanghai 200237, China (e-mail: qchjiang@ecust.edu.cn; xfyan@ecust.edu.cn).

Color versions of one or more of the figures in this article are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TASE.2019.2897477

developed in [27] for quality-relevant monitoring. Recently, a deep correlated representation learning-based monitoring scheme that incorporates deep belief neural network and CCA is developed for nonlinear process monitoring [28]. Although successful CCA applications have been reported, current CCA-based monitoring is limited to a single operation mode.

At present, a process generally operates in different modes because of various market demands, different reaction periods, and change in the production environment [29]–[31]. Monitoring such a multimode process is imperative, and several monitoring strategies have been developed [32], [33]. The main challenge in monitoring a multimode process is the manner in which the number of modes in historical data and the operation mode of an online sample are determined. Many clustering methods, such as K-means [34] and Gaussian mixture model (GMM) [35], have been utilized to solve this problem. However, conventional clustering methods generally require knowledge of the number of clusters, which is difficult to determine in practice.

The finite GMM has been proposed and applied to process monitoring to determine the number of modes automatically [36], [37]. In [36], the well-known Figueiredo–Jain (F–J) algorithm, which selects the number of components using the minimum message length criterion, was proposed. On the basis of the F–J GMM, a probabilistic multimode monitoring scheme was developed in [37]. Recently, an efficient unsupervised GMM learning algorithm, which uses variational methods to determine the number of components, has become a well-known research topic [38]. Variational Bayesian inference considers mode parameters and latent variables as unobserved variables [38]. The variational Bayesian GMM (VBGMM) regards the number of clusters as a parameter that can be adjusted during the expectation–maximization (EM) iteration process [38], [39]. VBGMM also determines the number of clusters automatically and has the potential to solve the mode identification problem.

In this paper, a method that integrates VBGMM and CCA (herein, VBGMM-CCA) is developed for efficient fault detection in multimode processes. The proposed method maximizes the advantage of VBGMM in automatic mode identification and the superiority of CCA in characterizing the correlation of local models. The main contributions of this paper are as follows.

First, this paper proposes to use multiple local linear models to describe an entire multimode process, which extends the CCA-based monitoring to handle complex process characteristics.

Second, variational Bayesian inference, which automatically identifies the number of local linear models that are necessary for modeling, is used to handle the historical process data. The influence of outliers is automatically eliminated.

Third, a Bayesian inference probability (BIP) index that integrates results from different models is developed for the CCA-based monitoring to increase the monitoring robustness.

This paper is structured as follows. Section II introduces the fundamentals of VBGMM and CCA. The multimode process monitoring problem is also formulated. Section III presents the proposed VBGMM-CCA method comprehensively. Application examples that involve a numerical multimode process and the batch-fed penicillin fermentation process (BPFP) are provided in Section IV. Section V presents the conclusion.

II. Preliminaries and Motivation

The notations throughout this paper are provided in this section. K represents the number of Gaussian components. N denotes the number of observations. a and b denote the number of variables. $\mathbf{u} \in \Re^a$ and $\mathbf{y} \in \Re^b$ denote the process input and output, respectively, and $\mathbf{x} = [\,\mathbf{u}^T\ \mathbf{y}^T\,]^T \in \Re^m$ ($m = a + b$) denotes a measurement with process input and output. π_k represents the mixing coefficient, $\boldsymbol{T}_k$ is the precision matrix, $\boldsymbol{\theta}_k = \{\boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k\}$ represents a parameter set with mean $\boldsymbol{\mu}_k$ and covariance variance $\boldsymbol{\Sigma}_k$, and $\boldsymbol{\Theta} = \{\pi_k, \boldsymbol{\theta}_k\}$. $N(\mathbf{x}_n|\boldsymbol{\theta}_k)$ is the normal distribution density function. $\boldsymbol{X}$ indicates the observed variables. $\boldsymbol{Z}$ denotes the hidden variables. $\boldsymbol{J}$ and $\boldsymbol{L}$ represent canonical vectors. $\mathbf{r}$ is the residual vector. rank($\mathbf{A}$) is the rank of matrix $\boldsymbol{A}$. $\boldsymbol{I}_a$ indicates the identity matrix with a columns. $E(\bullet)$ represents the expectation. prob(A) is the probability of event A. $\boldsymbol{\varepsilon}$ is the process noise vector. α is the level of significance. υ is the degree of freedom, and $\mathbf{V}$ is the scale matrix in the VBGMM algorithm. The other notations are standard.

A. VBGMM

A GMM comprises a finite number of normal distributions in a probabilistic manner. Given a set of samples $\boldsymbol{X}$ with $\{\mathbf{x}_n\}_{n=1,\ldots,N}$, a GMM with K Gaussian components is expressed as [38]

$$p(\mathbf{x}_n|\boldsymbol{\Theta}) = \sum_{k=1}^{K} \pi_k N(\mathbf{x}_n|\boldsymbol{\theta}_k) \tag{1}$$

where $0 \le \pi_k \le 1$ and $\sum_{k=1}^{K} \pi_k = 1$. The joint density of $\boldsymbol{X}$ can be derived as

$$p(\boldsymbol{X}|\boldsymbol{\Theta}) = \prod_{n=1}^{N} p(\mathbf{x}_n|\boldsymbol{\Theta}) = \prod_{n=1}^{N} \left(\sum_{k=1}^{K} \pi_k N(\mathbf{x}_n|\boldsymbol{\theta}_k) \right). \tag{2}$$

The EM algorithm is used to obtain the maximum likelihood estimation of the parameters, that is, $\boldsymbol{\Theta}_{\mathrm{ML}} = \arg\max(\ln p(\boldsymbol{X}|\boldsymbol{\Theta}))$.

In the E-step, the log-likelihood function is calculated as

$$\ln p(\boldsymbol{X}|\boldsymbol{\Theta}) = \sum_{n=1}^{N} \ln \left(\sum_{k=1}^{K} \pi_k N(\mathbf{x}_n|\boldsymbol{\theta}_k) \right). \tag{3}$$

Given that $\boldsymbol{\Theta}_{\mathrm{ML}}$ has no analytical solution, the maximum *a posteriori* probability estimate $\Theta_{\mathrm{MAP}} = \arg\max(\ln p(\boldsymbol{X}|\boldsymbol{\Theta}) + \ln p(\boldsymbol{\Theta}))$ given some prior $p(\boldsymbol{\Theta})$ is used. With z_{nk} denoting the probability of sample $\mathbf{x}_n$ belonging to the kth component, the conditional expectation of z_{nk} is calculated as

$$z_{nk} = \frac{\pi_k N(\mathbf{x}_n|\boldsymbol{\theta}_k)}{\sum_{j=1}^{K} \pi_j N(\mathbf{x}_n|\boldsymbol{\theta}_j)}. \tag{4}$$

In the M-step, setting the derivatives of $\ln p(\boldsymbol{X}|\boldsymbol{\Theta})$ with respect to $\boldsymbol{\mu}_k$, $\boldsymbol{\Sigma}_k$, and π_k yields

$$\boldsymbol{\mu}_k = \frac{\sum_{n=1}^{N} z_{nk}\mathbf{x}_N}{\sum_{n=1}^{N} z_{nk}} \tag{5}$$

$$\boldsymbol{\Sigma}_k = \frac{1}{\sum_{n=1}^{N} z_{nk}} \sum_{n=1}^{N} z_{nk}(\mathbf{x}_n - \boldsymbol{\mu}_k)(\mathbf{x}_n - \boldsymbol{\mu}_k)^T \tag{6}$$

$$\pi_k = \frac{1}{N} \sum_{n=1}^{N} z_{nk}. \tag{7}$$

The E- and M-steps alternate until a local maximum is obtained. When EM is used, selecting the number of components K is critical. An extremely large K may cause overfitting and an extremely small K may not be sufficiently flexible to describe the true underlying models.

The variational Bayesian method determines the number of components automatically by controlling the complexity in mixture models and reducing the mixing coefficient to nearly zero. Given the governing parameter set $\boldsymbol{\Theta}$, the log-likelihood of $p(\boldsymbol{X}|\boldsymbol{\Theta}) = \sum_Z p(\boldsymbol{X}, \boldsymbol{Z}|\boldsymbol{\Theta})$ is written as

$$\ln p(\boldsymbol{X}|\boldsymbol{\Theta}) = L(q, \Theta) + KL(q||P) \tag{8}$$

where

$$L(q, \boldsymbol{\Theta}) = \int q(\boldsymbol{Z}) \ln \frac{p(\boldsymbol{X}, \boldsymbol{Z})}{q(\boldsymbol{Z})} d\boldsymbol{Z} \tag{9}$$

$$\mathrm{KL}(q||p) = -\int q(\boldsymbol{Z}) \ln \frac{p(\boldsymbol{Z}|\boldsymbol{X})}{q(\boldsymbol{Z})} d\boldsymbol{Z} \tag{10}$$

in which $q(\boldsymbol{Z})$ is a density function that approximates the true joint conditional density $p(\boldsymbol{Z}|\boldsymbol{X})$. $\mathrm{KL}(q||P)$ is the Kullback–Leibler (KL) divergence between $p(\boldsymbol{Z}|\boldsymbol{X})$ and $q(\boldsymbol{Z})$. Given that the KL divergence is positive, maximizing the lower bound of (8) with regard to $q(\boldsymbol{Z})$ is equal to minimizing the KL divergence. By partitioning the elements of $\boldsymbol{Z}$ into disjoint groups, $q(\boldsymbol{Z})$ can be expressed as

$$q(Z) = \prod_{i=1}^{K} q_i(z_i). \tag{11}$$

$L(q, \boldsymbol{\Theta})$ can be expressed as

$$\begin{aligned} L(q, \boldsymbol{\Theta}) &= \int \prod_i q_i \left(\ln p(\boldsymbol{X}, \boldsymbol{Z}) - \sum_i \ln q_i \right) d\boldsymbol{Z} \\ &= -KL(q_j||\widetilde{p}) - \sum_{i \neq j} \int q_i \ln q_i dz_i. \end{aligned} \tag{12}$$

The bound in (12) is maximized when $KL(q_j||\widetilde{p})$ becomes zero. The optimal distribution $q_j^*(z_j)$ is expressed as

$$\ln q_j^*(z_j) = \langle \ln p(\boldsymbol{X}, \boldsymbol{Z}) \rangle_{i \neq j} + \text{const.} \tag{13}$$

Overall, the E-step evaluates $q(\boldsymbol{Z})$ to maximize $L(q, \boldsymbol{\Theta})$, and the M-step derives $\boldsymbol{\Theta}_{\mathrm{VB}} = \arg\max L(q, \boldsymbol{\Theta})$. The two steps are performed iteratively and converge to a local maximum. During the iteration process, several mixing coefficients converge to zero, which indicates that the corresponding components should be removed. The number of models is automatically determined. The detailed derivations of VBGMM are provided in the Appendix [38]–[40].

B. CCA

Given two sets of zero-mean random variables $\boldsymbol{U}$ and $\boldsymbol{Y}$, where

$$\boldsymbol{U} = \begin{bmatrix} \boldsymbol{u}_1 \\ \boldsymbol{u}_2 \\ \vdots \\ \boldsymbol{u}_N \end{bmatrix} = \begin{bmatrix} u_{11} & u_{12} & \dots & u_{1a} \\ u_{21} & u_{22} & \dots & u_{2a} \\ \vdots & \vdots & \ddots & \vdots \\ u_{N1} & u_{N2} & \dots & u_{Na} \end{bmatrix} \quad (N \text{ samples, } a \text{ variables})$$

and

$$\boldsymbol{Y} = \begin{bmatrix} \boldsymbol{y}_1 \\ \boldsymbol{y}_2 \\ \vdots \\ \boldsymbol{y}_N \end{bmatrix} = \begin{bmatrix} y_{11} & y_{12} & \dots & y_{1b} \\ y_{21} & y_{22} & \dots & y_{2b} \\ \vdots & \vdots & \ddots & \vdots \\ y_{N1} & y_{N2} & \dots & y_{Nb} \end{bmatrix} \quad (b \text{ variables})$$

CCA attempts to determine canonical vectors $\boldsymbol{J}$ and $\boldsymbol{L}$, which realize the high correlation of $\boldsymbol{J}^T\boldsymbol{U}^T$ and $\boldsymbol{L}^T\boldsymbol{Y}^T$, respectively, shown as follows [1], [41]:

$$(\boldsymbol{J}, \boldsymbol{L}) = \arg\max_{(\boldsymbol{J},\boldsymbol{L})} \frac{\boldsymbol{J}^T \boldsymbol{\Sigma}_{\mathbf{UY}} \boldsymbol{L}}{(\boldsymbol{J}^T \boldsymbol{\Sigma}_{\boldsymbol{U}} \boldsymbol{J})^{1/2} (\boldsymbol{L}^T \boldsymbol{\Sigma}_{\boldsymbol{U}} \boldsymbol{L})^{1/2}} \tag{14}$$

where $\boldsymbol{\Sigma}_{(\cdot)}$ is the covariance matrix. The solution to (14) is obtained by singular-value decomposition on matrix $\boldsymbol{K}$, shown as follows [1], [41]:

$$\boldsymbol{K} = \boldsymbol{\Sigma}_{\boldsymbol{U}}^{-1/2} \boldsymbol{\Sigma}_{\boldsymbol{UY}} \boldsymbol{\Sigma}_{\boldsymbol{Y}}^{-1/2} = \boldsymbol{R} \boldsymbol{\Sigma} \boldsymbol{V}^T \tag{15}$$

where

$$\boldsymbol{\Sigma} = \begin{bmatrix} \mathrm{diag}(\sigma_1, \dots, \sigma_l) & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix} \in \Re^{a \times b}$$

and $l = \mathrm{rank}(\boldsymbol{\Sigma})$. Then, we derive [1], [41]

$$\begin{aligned} \boldsymbol{J} &= \boldsymbol{\Sigma}_{\boldsymbol{U}}^{-1/2} \boldsymbol{R} \\ \boldsymbol{L} &= \boldsymbol{\Sigma}_{\boldsymbol{Y}}^{-1/2} \boldsymbol{V}. \end{aligned} \tag{16}$$

For samples $\mathbf{u} \in \Re^a$ and $\mathbf{y} \in \Re^b$, the residual vector is generated as [19]

$$\mathbf{r} = \boldsymbol{J}^T \mathbf{u} - \boldsymbol{\Sigma} \boldsymbol{L}^T \mathbf{y}. \tag{17}$$

Then, the T^2 statistic for the residual is established as [19]

$$T^2 = \mathbf{r}^T \boldsymbol{\Sigma}_r^{-1} \mathbf{r} \tag{18}$$

where $\boldsymbol{\Sigma}_r$ is the covariance matrix of the residual $\boldsymbol{r}$. Under the Gaussian assumption, the T^2 statistic is optimal for detecting a fault that affects only $\mathbf{u}$ [19]. Similarly, the T^2 test for fault detection that affects only $\mathbf{y}$ can be established.

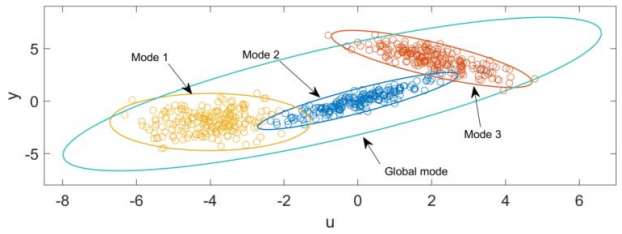


Fig. 1. Illustration of multimode process data.

C. Problem Formulation

The CCA monitoring model effectively handles linearly correlated process data from a single operation mode. However, a modern process can be characterized by multiple operation modes. Establishing a single CCA model for all operation modes can be probabilistic because the variable relationships in a local mode may not be well explored. Thus, local CCA models must be established to explore local linear correlations among variables. For example, a process is operated in three different modes during a period, and the process data are distributed, as shown in Fig. 1. Modeling the process with a global model ignores the local process behavior, particularly in exploring the correlation between input $\mathbf{u}$ and output $\mathbf{y}$. The local operation modes must be identified, and local monitoring models must be constructed.

The following challenges exist in exploring local variable relationships and establishing local monitoring models. First, the number of operation modes from the historical data set should be determined, and the operation mode where each sample belongs to should be identified. Historical process data are generally unlabeled, and an unsupervised clustering method should be utilized. Second, local linear relationships should be well characterized, and efficient fault detection residuals should be generated in each identified operation mode. Third, the running-on operation mode for an online sample should be identified, and an appropriate monitoring model should be used. Finally, a probability index that integrates the results from all modes should be developed to increase the monitoring robustness and avoid the potential risk of misclassification [37]. Here, a VBGMM-CCA monitoring scheme is proposed for efficient multimode process monitoring.

III. VBGMM-CCA Process Monitoring

A. VBGMM-CCA Monitoring Procedures

The basic idea of the VBGMM-CCA monitoring method is to identify the operation mode through VBGMM and model the local process input and output relationship through CCA. We present the monitoring scheme in a step-by-step manner.

Step 1 (Historical Operation Mode Identification): VBGMM is applied to the historical process data. A slightly large number of operation modes is set, and the number of operation modes is automatically determined by VBGMM. The operation model where each sample belongs to is also identified. Notably, several of the identified modes contain only a few samples, which may be caused by gross errors or outliers. These modes are not the actual existing modes in the process history. Moreover, these modes with the corresponding samples should be removed from the identified operation modes to reduce online computation complexity.

Step 2 (Local CCA Model Establishment): Once the samples at each operation mode are collected, the local CCA monitoring model is established for each mode. For samples at one operation mode, the following procedures are implemented to establish the CCA monitoring model.

First, the data are mean–variance normalized.

Second, the data are arranged as process input $\boldsymbol{U}$ and process output $\boldsymbol{Y}$. CCA is conducted between $\boldsymbol{U}$ and $\boldsymbol{Y}$ to obtain local canonical correlation vectors $\boldsymbol{J}$ and $\boldsymbol{L}$.

Third, the following residuals are generated for fault detection:

$$\mathbf{r}_u = \boldsymbol{J}^T\mathbf{u} - \boldsymbol{\Sigma}\boldsymbol{L}^T\mathbf{y} \tag{19}$$

$$\mathbf{r}_y = \boldsymbol{L}^T\mathbf{y} - \boldsymbol{\Sigma}^T\boldsymbol{J}^T\mathbf{u}. \tag{20}$$

We have

$$E(\mathbf{r}_u) = \boldsymbol{J}^T E(\mathbf{u}) - \boldsymbol{\Sigma}\boldsymbol{L}^T E(\mathbf{y}) = \mathbf{0} \tag{21}$$

$$E(\mathbf{r}_y) = \boldsymbol{L}^T E(\mathbf{y}) - \boldsymbol{\Sigma}^T\boldsymbol{J}^T E(\mathbf{u}) = \mathbf{0}. \tag{22}$$

Fourth, monitoring statistics are constructed. T^2 tests are performed on the following residuals:

$$T_u^2 = \mathbf{r}_u^T\boldsymbol{\Sigma}_{ru}^{-1}\mathbf{r}_u \tag{23}$$

$$T_y^2 = \mathbf{r}_y^T\boldsymbol{\Sigma}_{ry}^{-1}\mathbf{r}_y \tag{24}$$

where $\boldsymbol{\Sigma}_{ru} = \boldsymbol{I}_a - \sum\sum^T$ and $\boldsymbol{\Sigma}_{ry} = \boldsymbol{I}_b - \sum^T\sum$ are the covariance of $\mathbf{r}_u$ and $\mathbf{r}_y$, respectively. T_u^2 and T_y^2 examine a fault in the process input and output in consideration of the correlation information from the process output and input, respectively.

Fifth, the thresholds are determined. On the basis of the local Gaussian assumption, thresholds $T_{u,cl}^2$ and $T_{y,cl}^2$ can be determined as follows:

$$T_{u,cl}^2 = \chi_\alpha^2(m_u) \tag{25}$$

$$T_{y,cl}^2 = \chi_\alpha^2(m_y) \tag{26}$$

where $m_u = \text{rank}(\boldsymbol{\Sigma}_{ru})$ and $m_y = \text{rank}(\boldsymbol{\Sigma}_{ry})$.

These steps are repeated until CCA models are established for all operating conditions.

Step 3 (Online Operation Model Identification): When a new sample arrives, the density of the sample at operation mode i is calculated as

$$p(\mathbf{x}_{\text{new}}|\boldsymbol{\mu}_i, \boldsymbol{\Sigma}_i) = \frac{1}{(2\pi)^{m/2}|\boldsymbol{\Sigma}_i|^{1/2}} \exp\left\{-\frac{1}{2}(\mathbf{x}_{\text{new}} - \boldsymbol{\mu}_i)^T\boldsymbol{\Sigma}_i^{-1}(\mathbf{x}_{\text{new}} - \boldsymbol{\mu}_i)\right\}. \tag{27}$$

The mode with the largest density is determined as the current running-on mode.

Step 4 (Process Status Identification): After identifying the operation mode, the sample is evaluated by the corresponding local CCA model. The monitoring statistics are calculated,

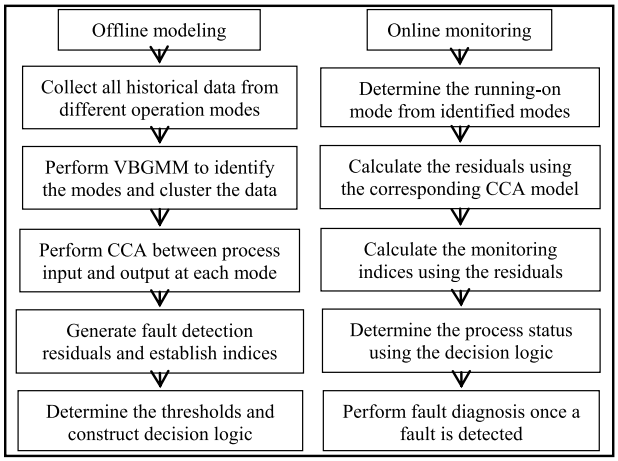


Fig. 2. Procedures of VBGMM-CCA monitoring scheme.

and the process status is determined according to the following decision logic:

$$\begin{cases} T_u^2 > T_{u,cl}^2 \text{ or } T_y^2 > T_{y,cl}^2 \Rightarrow \text{faulty} \\ T_u^2 \le T_{u,cl}^2 \text{ and } T_y^2 \le T_{y,cl}^2 \Rightarrow \text{fault-free.} \end{cases} \tag{28}$$

This is the deterministic mode identification approach. An integrated probability index should be established to avoid the potential risk of classifying a sample to a false mode. Yu and Qin [37] proposed a BIP monitoring index and demonstrated its advantages. However, the BIP does not focus on detecting a local fault that affects $\mathbf{u}$ or $\mathbf{y}$. Here, we use the Bayesian inference strategy in [37] and further extend it to CCA monitoring. For an online sample $\mathbf{x}_{\text{new}} = [\,\mathbf{u}_{\text{new}}^{\text{T}}\ \mathbf{y}_{\text{new}}^{\text{T}}\,]^T$, the posterior of $\mathbf{x}_{\text{new}}$ that belongs to the kth mode M_k is calculated as

$$P(\mathbf{x}_{\text{new}} \in M_k) = P(M_k|\mathbf{x}_{\text{new}}) = \frac{P(M_k)p(\mathbf{x}_{\text{new}}|\boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k)}{\sum_{i=1}^{K} P(M_i)p(\mathbf{x}_{\text{new}}|\boldsymbol{\mu}_i, \boldsymbol{\Sigma}_i)}. \tag{29}$$

Then, the final BIP index that concerns T_u^2 (BIP_u) can be obtained as follows:

$$\text{BIP}_u = \sum_{k=1}^{K} P(N_u|M_k)P(M_k|\mathbf{x}_{\text{new}}) \tag{30}$$

where $P(N_u|M_k)$ is defined as

$$P(N_u|M_k) = \text{prob}\big(T_{u,cl}^2|M_k < T_u^2|M_k\big) \tag{31}$$

which can be calculated by integrating the χ^2 distribution with corresponding degree of freedom. Similarly, the BIP index regarding the T_y^2 (BIP_y) can be obtained. Following the results in [37], the threshold of the statistics is determined as the confidence level 1-α. Then, the decision logic is established as follows:

$$\begin{cases} \text{BIP}_u > 1-\alpha \text{ or } \text{BIP}_y > 1-\alpha \Rightarrow \text{faulty} \\ \text{BIP}_u \le 1-\alpha \text{ and } \text{BIP}_y \le 1-\alpha \Rightarrow \text{fault-free.} \end{cases} \tag{32}$$

The procedures of the proposed monitoring scheme are summarized in Fig. 2.

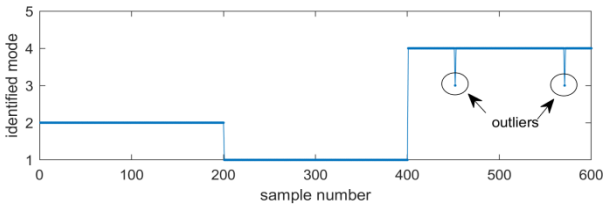


Fig. 3. Historical data mode identification result.

B. Characteristic Analysis

Characteristic 1 (Feasibility Analysis): The proposed VBGMM-CCA automatically identifies the number of local linear models for describing the entire process. The influence of outliers can be eliminated by removing the samples with the lowest probability. In each local model, optimal fault detection residual is generated, which increases the monitoring effectiveness in local fault detection.

Characteristic 2 (Computational Burden Analysis): The proposed multimode monitoring scheme is computational efficiency. The VBGMM algorithm considers the number of necessary models as a parameter that is optimized; this approach is more efficient than conventional GMM methods.

Characteristic 3 (Advantage in Local Fault Detection): The advantage of the VBGMM-CCA lies in the capability of CCA in detecting the local fault that affects only the process input or output. Additional information on the type of a detected fault is provided instead of using an entire statistic [19].

Characteristic 4 (Monitoring Robustness): The proposed VBGMM-CCA integrates the results from all local models into the BIP index. This feature takes the advantages of probabilistic monitoring in avoiding the effects of incorrect mode identification, thereby increasing monitoring robustness [37].

Notably, the VBGMM-CCA establishes monitoring models using historical process data. In practice, the process may be nonstationary due to time-varying operation conditions, where the process changes with time and does not remain at some steady model. Once the varied operation condition is removed in process history, the VBGMM-CCA may regard the condition change as a fault. In this situation, a full-condition analysis can be helpful to understand the process behavior [42].

IV. Experimental Studies

A. Numerical Example

The simulated numerical example presented in [19], which consists of process input $\mathbf{u}$ and process output $\mathbf{y}$, is expressed as follows:

$$\mathbf{x} = N(\boldsymbol{\mu}, \boldsymbol{\Sigma}_x) + \boldsymbol{\varepsilon} \tag{33}$$

where $\mathbf{x} = [\,u_1\ u_2\ y_1\ y_2\ y_3\ y_4\,]^T$ denotes the measured variables, and $N(\boldsymbol{\mu}, \boldsymbol{\Sigma}_x)$ represents the multivariate Gaussian distributed variables with mean $\boldsymbol{\mu}$ and covariance $\boldsymbol{\Sigma}_x$. Three different modes are simulated according to the different values of $\boldsymbol{\mu}$. The first, second, and third modes are $\boldsymbol{\mu}_1 = [\,0\ 0\ 0\ 0\ 0\ 0\,]^T$, $\boldsymbol{\mu}_2 = [\,2\ 4\ 3\ 5\ 6\ 1\,]^T$, and $\boldsymbol{\mu}_3 = [\,-4\ -2\ -3\ -1\ -5\ -6\,]^T$, respectively.

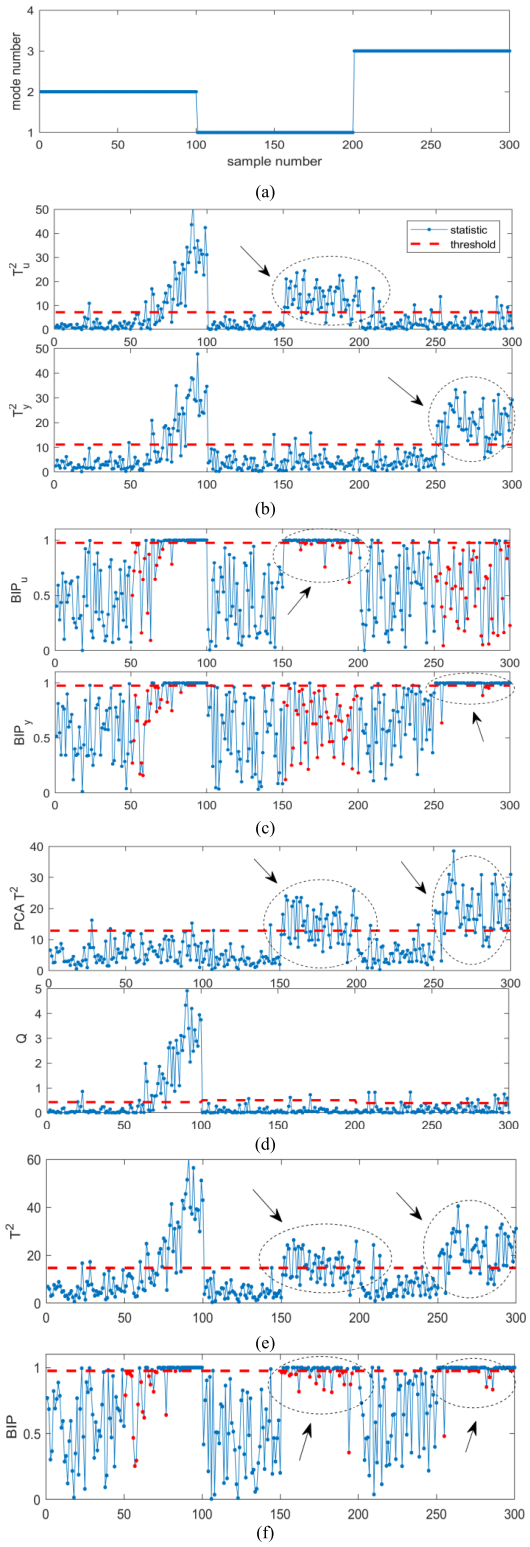


Fig. 4. Monitoring results for the numerical example. (a) Online monitoring mode identification result. (b) VBGMM-CCA. (c) VBGMM-CCA with BIP. (d) VBGMM-PCA. (e) VBGMM-T^2. (f) F–J GMM with BIP index.

Under normal operating conditions, 200 samples of each mode are generated as training data. The data are unlabeled and mixed. The mode identification results obtained based on VBGMM are presented in Fig. 3. The data are automatically identified as four modes. However, in the third mode, only

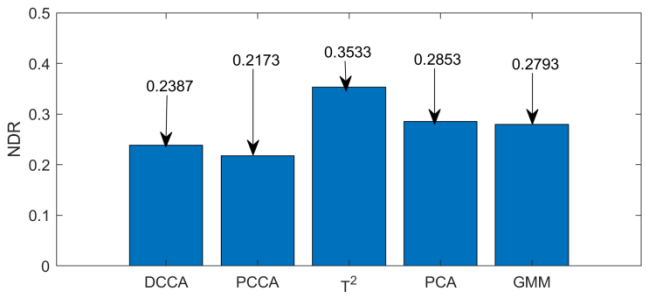


Fig. 5. Monte Carlo test results of different methods.

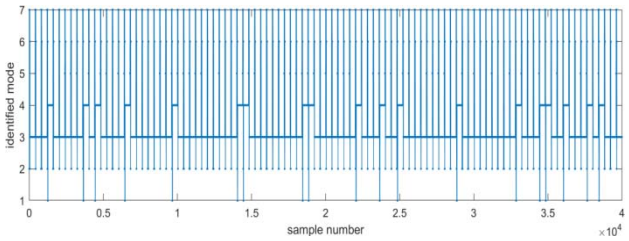


Fig. 6. Mode identification results of BPFP historical data.

a few points are identified because of gross errors; however, the mode is not real. The modes with only few samples (less than 2% of all samples) are omitted. A CCA monitoring model is established in each mode. The following testing data set is generated to test the monitoring performance.

The process is operated under the first mode from the 1st to the 100th point and a ramp change of $0.05 \times (t - 50)$, where t denotes the time instant, which is introduced to u_1 from the 51st to the 100th point. The process is operated under the second mode from the 101st to the 200th point, and a step change of 4.5 is introduced to u_2 from the 151st to the 200th point. The process is operated under the third mode from the 201st to the 300th point, and a step change of 4.5 is introduced to y_3 from the 251st to the 300th point. The mode identification results are presented in Fig. 4(a), which shows that the operation mode is successfully identified. The monitoring results using VBGMM-CCA with deterministic mode identification (T_u^2 and T_y^2), VBGMM-CCA with BIP index (BIP_u and BIP_y), VBGMM-PCA (T^2 and Q), VBGMM T^2, and the F–J GMM (BIP) in [37] are presented in Fig. 4(b)–(f), respectively. The first fault affects the process input; however, the affected part is related to the output. All methods detect the fault. The second fault affects the process input. T_u^2 and BIP_u detect the fault. This result indicates that the fault is a local one and affects only the process input. The third fault affects only the process output. T_y^2 and BIP_y also detect the fault. In comparison with PCA and T^2 test, the nondetection rate (NDR) of the CCA method is reduced. The type of the detected fault, that is, the fault that affects the input or output, can also be identified according to the status of the statistics or indices in CCA monitoring.

A Monte Carlo simulation is conducted (1000 times), and the average NDRs of VBGMM–CCA (with deterministic mode identification, DCCA), VBGMM-CCA with BIP (PCCA), VBGMM-T^2, VBGMM-PCA, and the GMM in [37]

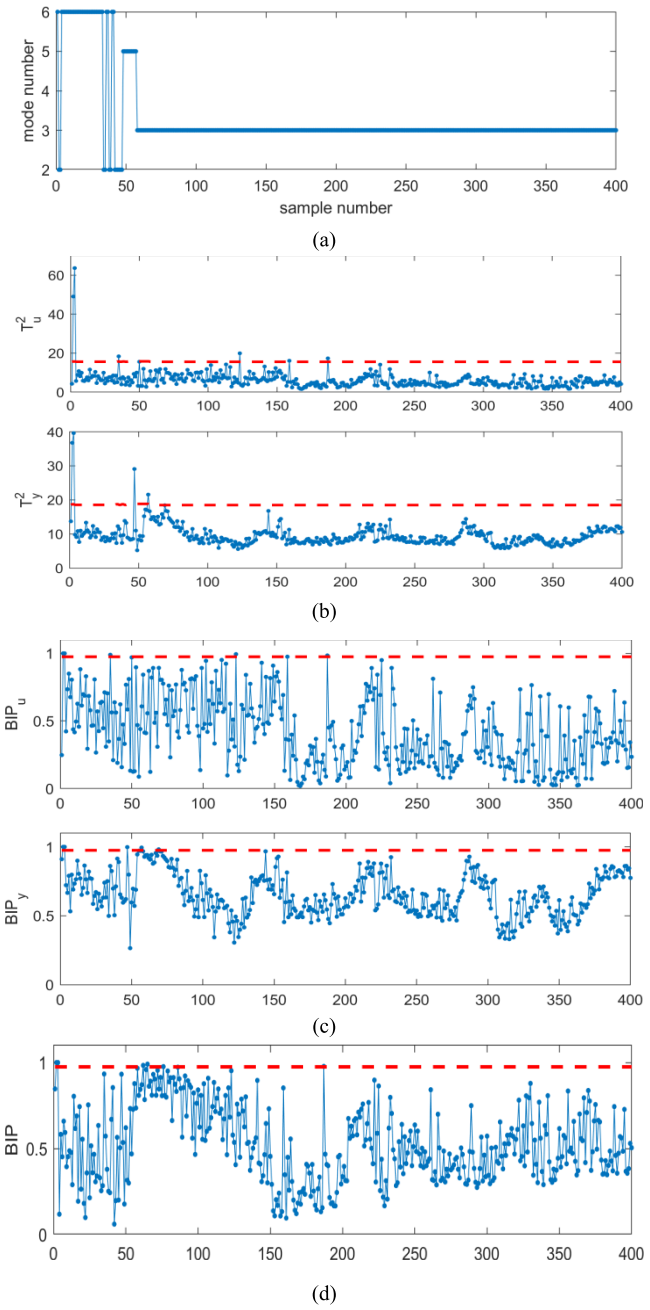


Fig. 7. Monitoring results for BPFP normal batch. (a) Online mode identification results. (b) VBGMM-CCA. (c) VBGMM-CCA with BIP. (d) Conventional GMM.

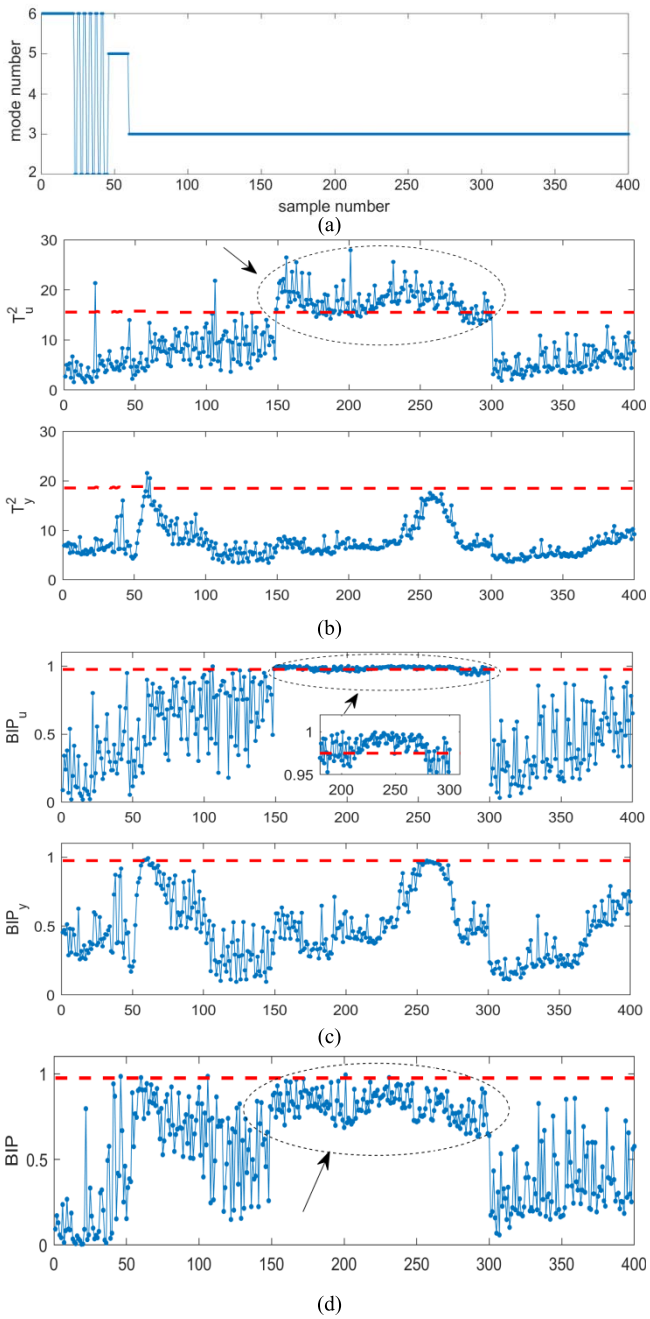


Fig. 8. Monitoring results for BPFP Fault 1. (a) Online mode identification results. (b) VBGMM-CCA. (c) VBGMM-CCA with BIP. (d) Conventional GMM.

are presented in Fig. 5. The proposed VBGMM with BIP indices exhibits the best monitoring performance with the lowest NDR. We compare the F–J GMM and VBGMM in terms of modeling complexity. The former takes approximately 1.69 s, whereas the latter takes approximately 0.17 s in one time operation. The VBGMM is more computationally efficient than the F–J GMM in this paper.

B. Batch-Fed Penicillin Fermentation Process

BPFP is a benchmark problem that is widely used to test the performance of a monitoring scheme [43]–[45]. The process is characterized by evident nonlinearity and multiple phases. The process of a batch run can generally be divided into three phases, namely, cell growth, penicillin synthesis, and cell autolysis. Throughout the entire fermentation process, many factors, such as temperature, pH, substrate concentration, and dissolved oxygen concentration, affect the effectiveness of penicillin fermentation. BPFP monitoring is imperative but still remains challenging.

A simplified flowchart is presented in Fig. 11. The simulation package is used to generate process data [43]. Sixteen variables are considered and listed in Table I. The entire duration of each batch is 400 h. The initial conditions and set points of the simulation are presented in Table II. The sampling interval is 1 h, and 400 samples are collected from

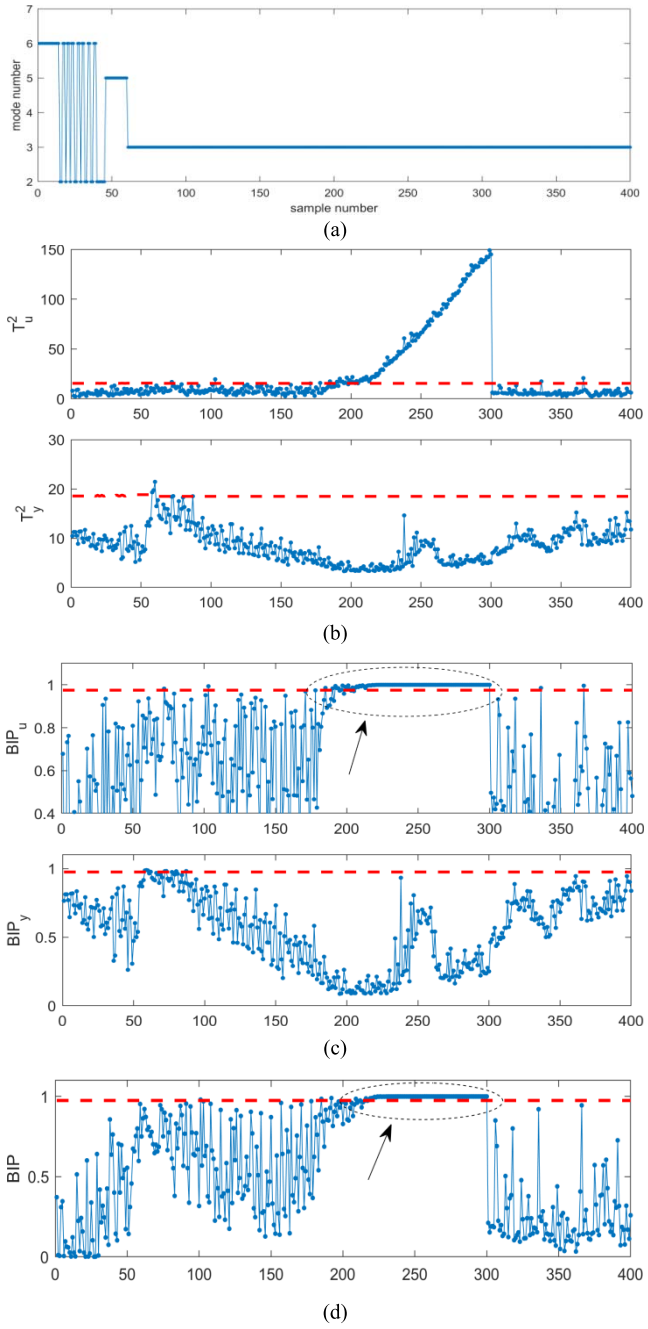


Fig. 9. Monitoring results for BPFP Fault 2. (a) Online mode identification results. (b) VBGMM-CCA. (c) VBGMM-CCA with BIP. (d) Conventional GMM.

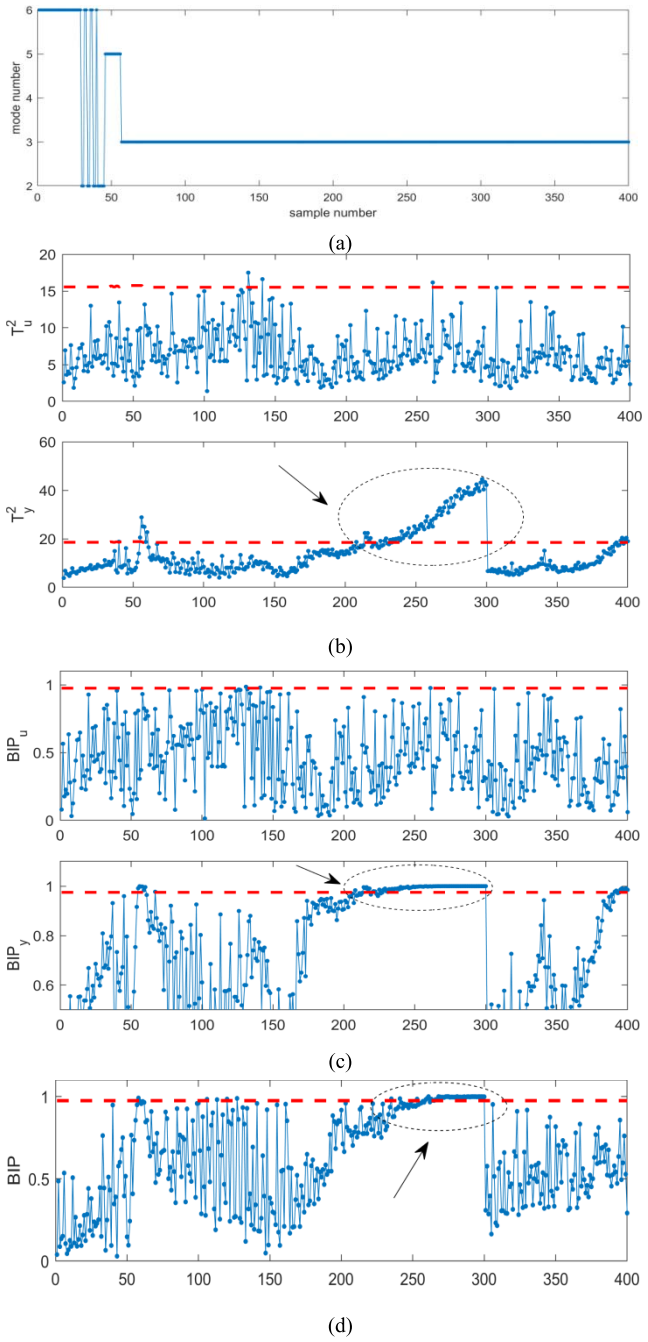


Fig. 10. Monitoring results for BPFP Fault 3. (a) Online mode identification results. (b) VBGMM-CCA, (c) VBGMM-CCA with BIP. (d) Conventional GMM.

a batch. A total of 100 batches are simulated under normal operating conditions, and the process data are collected as training data.

The VBGMM-based historical mode identification result is presented in Fig. 6. All process data are clustered into seven modes. Thus, seven Gaussian components are necessary to describe the historical process status. Although only three typical phases exist in a batch run, more than one mode may be necessary to describe a phase because each phase may exhibit considerable nonlinearity.

One batch under normal operating condition and three batches with different faults are generated to test the monitoring performance. The batches with faults are as follows. In Fault 1, a step change is introduced to agitator power from the 150th to the 300th point. In Fault 2, a ramp change is introduced to agitator power from the 150th to the 300th point. In Fault 3, a ramp change is introduced to aeration rate from the 150th to the 300th point.

The mode identification results of the normal operating batch are presented in Fig. 7(a), which shows that the entire process has six different operation periods. The monitoring results using VBGMM–CCA, VBGMM-CCA with BIP, and the GMM in [37] are presented in Fig. 7(b)–(d), respectively. The monitoring performance under normal operating condition

is clearly not degraded by the VBGMM monitoring strategies. The mode identification results of the Fault 1 batch are presented in Fig. 8(a), which also shows that the entire process has six different operation periods. The monitoring results for the Fault 1 batch obtained using VBGMM–CCA, VBGMM–CCA with BIP, and the GMM in [37] are presented in Fig. 8(b)–(d), respectively, from which CCA-based methods are observed to provide better monitoring results than conventional GMM for the fault. Notably, T^2 performs poorly because of the complex correlation among variables. Thus, its results are not listed here. The mode identification results for the Fault 2 batch are presented in Fig. 9(a). The monitoring results for the Fault 2 batch obtained using VBGMM-CCA, VBGMM-CCA with BIP, and the GMM in [37] are presented in Fig. 9(b)–(d), respectively. Fig. 9(b)–(d)shows that CCA-based methods detect the fault at the earliest time. The mode identification results for the Fault 3 batch are presented in Fig. 10(a). The monitoring results for the Fault 3 batch obtained using VBGMM-CCA, VBGMM-CCA with BIP, and the GMM in [37] are presented in Fig. 10(b)–(d), respectively. Fig. 10(b)–(d) also shows that CCA-based methods detect the fault at the earliest time. Thus, the efficiency of the proposed monitoring scheme is validated.

V. Conclusion

In this paper, a VBGMM–CCA-based fault detection method is proposed for multimode process monitoring. The basic idea is to maximize the advantage of VBGMM in automatic mode identification and the superiority of CCA in modeling the local operation mode. First, the number of operation modes is automatically determined by VBGMM from the historical process data. Second, the historical data are clustered, and a local CCA fault detection model is established. Third, fault detection residuals are established, and monitoring statistics are constructed. Finally, a BIP index is developed to increase the monitoring robustness. The effectiveness of the proposed monitoring scheme is validated by case studies on a numerical example and multiphase BPFP.

The proposed VBGMM-CCA is superior in three aspects. First, the number of operation modes, which is important for practical applications, can be automatically identified. Second, the CCA-based local monitoring identifies the process status and the property of a detected fault. Third, the proposed monitoring scheme performs process monitoring in a probabilistic manner, which increases the monitoring robustness.

Appendix

A. VBGMM Details

In VBGMM, Gaussian and Wishart priors are assumed for mixing weight $\boldsymbol{\pi}$ and precision matrix $\boldsymbol{T} = \{\boldsymbol{T}_k\}$. Then

$$p(\boldsymbol{\mu}) = \prod_{k=1}^{K} N(\boldsymbol{\mu}_k|\beta \boldsymbol{I}) \tag{A1}$$

$$p(\boldsymbol{T}) = \prod_{k=1}^{K} W(\boldsymbol{T}_k|\upsilon, \mathbf{V}) \tag{A2}$$

where υ is degrees of freedom and $\mathbf{V}$ represents the scale matrix. The expected value of $\boldsymbol{T}_k$ is $\langle \boldsymbol{T}_k \rangle = \upsilon \mathbf{V}^{-1}$. The marginal likelihood $p(\boldsymbol{X}, \boldsymbol{\pi})$ can be obtained by integrating the latent variable $\widehat{Z} = \{Z, \boldsymbol{\mu}, \boldsymbol{T}\}$ as follows:

$$p(\boldsymbol{X}, \boldsymbol{\pi}) = \int p(\boldsymbol{X}, \boldsymbol{\pi}, \widehat{\mathbf{Z}}) d\widehat{\mathbf{Z}}. \tag{A3}$$

VB approximation maximizes the lower bound of the logarithmic marginal likelihood as follows:

$$L(q, \boldsymbol{\pi}) = \int q(\widehat{\mathbf{Z}}) \ln \frac{p(\boldsymbol{X}, \boldsymbol{\pi}, \widehat{\mathbf{Z}})}{q(\widehat{\mathbf{Z}})} d\widehat{\mathbf{Z}} \leq p(\boldsymbol{X}, \boldsymbol{\pi}) \tag{A4}$$

where $q(\widehat{\mathbf{Z}})$ is a distribution approximating the posterior $p(\widehat{\mathbf{Z}}|\boldsymbol{X})$. We assume that $q(\widehat{\mathbf{Z}})$ has the following product form:

$$q(\widehat{\mathbf{Z}}) = q_z(\mathbf{Z}) q_\mu(\boldsymbol{\mu}) q_T(\boldsymbol{T}). \tag{A5}$$

After several calculation procedures, the following results are obtained [38]:

$$q_z(\mathbf{Z}) = \prod_{n=1}^{N} \prod_{k=1}^{K} r_{nk}^{z_{nk}} \tag{A6}$$

$$q_\mu(\boldsymbol{\mu}) = \prod_{k=1}^{K} N(\boldsymbol{\mu}_k|\boldsymbol{m}_k, \boldsymbol{S}_k) \tag{A7}$$

$$q_T(\boldsymbol{T}) = \prod_{k=1}^{K} W(\boldsymbol{T}_k|\eta_k, \boldsymbol{U}_k) \tag{A8}$$

$$r_{nk} = \frac{\widehat{r}_{nk}}{\sum_{j=1}^{K} \widehat{r}_{nj}} \tag{A9}$$

$$\widehat{r}_{nk} = \pi_k \exp\left(\frac{1}{2} \ln |\boldsymbol{T}_k| - \frac{1}{2} tr\left\{\langle \boldsymbol{T}_k \rangle \left(\mathbf{x}_n \mathbf{x}_n^T - \mathbf{x}_n \langle \mathbf{u}_k \rangle^T - \langle \mathbf{u}_k \rangle \mathbf{x}_n^T + \langle \mathbf{u}_k \mathbf{u}_k^T \rangle\right)\right\}\right) \tag{A10}$$

$$\boldsymbol{m}_k = \boldsymbol{S}_k^{-1} \langle \boldsymbol{T}_k \rangle \sum_{n=1}^{N} \langle z_{nk} \rangle \mathbf{x}_n \tag{A11}$$

$$\boldsymbol{S}_k = \beta \boldsymbol{I} + \langle \boldsymbol{T}_k \rangle \sum_{n=1}^{N} \langle z_{nk} \rangle \tag{A12}$$

$$\eta_k = \upsilon + \sum_{n=1}^{N} \langle z_{nk} \rangle \tag{A13}$$

$$\boldsymbol{U}_k = \mathbf{V} + \sum_{n=1}^{N} \langle z_{nk} \rangle \times \left(\mathbf{x}_n \mathbf{x}_n^T - \mathbf{x}_n \langle \mathbf{u}_k \rangle^T - \langle \mathbf{u}_k \rangle \mathbf{x}_n^T + \langle \mathbf{u}_k \mathbf{u}_k^T \rangle\right). \tag{A14}$$

In the M-step, $\boldsymbol{\pi}$ is updated as

$$\pi_k = \frac{\sum_{n=1}^{N} r_{nk}}{\sum_{j=1}^{K} \sum_{n=1}^{N} r_{nj}}. \tag{A15}$$

E-step and M-step are performed iteratively until the algorithm converges.

B. BPFP Details

See Fig. 11 and Tables I and II.

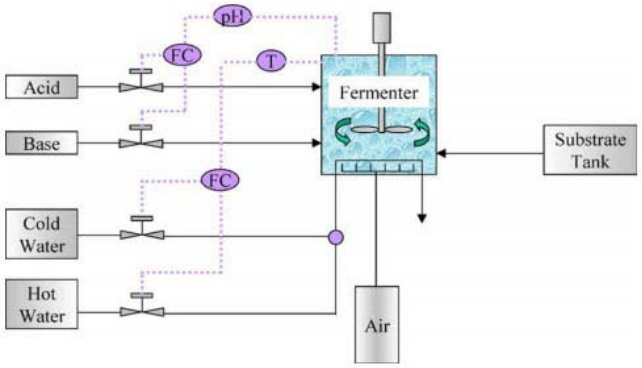


Fig. 11. Flowsheet of the penicillin cultivation process [43], [46].

TABLE I

MEASURED VARIABLES OF THE FED-BATCH PENICILLIN FERMENTATION PROCESS [43], [46]

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

TABLE II

INITIAL CONDITIONS AND SET POINTS TO SIMULATE THE PENICILLIN CULTIVATION PROCESS [43], 46]

Initial conditions	Range
Substrate concentration (g/L)	15-17
Dissolved oxygen concentration (mmol/L)	1-1.2
Biomass concentration (g/L)	0.1
Penicillin concentration (g/L)	0
Culture volume (L)	101-103
Carbon dioxide concentration (mmol/L)	0.5-1
Hydrogen ion concentration: $[H^+]$ (mol/L)	10^{-5}
Bioreactor temperature (K)	297-299
Generated heat (kcal)	0
Set points	
Aeration rate (g/L)	8-9
Agitator power (W)	29-31
Substrate feed flow rate (L/h)	0.039-0.045
Substrate feed temperature (K)	295-296
Bioreactor temperature (K)	297-298
PH	4.95-5.05

REFERENCES

[1] S. X. Ding, *Data-Driven Design of Fault Diagnosis and Fault-Tolerant Control Systems*. London, U.K.: Springer, 2014.

[2] Z. Gao, S. X. Ding, and C. Cecati, "Real-time fault diagnosis and fault-tolerant control," *IEEE Trans. Ind. Electron.*, vol. 62, no. 6, pp. 3752–3756, Jun. 2015.

[3] C. Zhao, "A quality-relevant sequential phase partition approach for regression modeling and quality prediction analysis in manufacturing processes," *IEEE Trans. Autom. Sci. Eng.*, vol. 11, no. 4, pp. 983–991, Oct. 2014.

[4] Q. Jiang and X. Yan, "Parallel PCA–KPCA for nonlinear process monitoring," *Control Eng. Pract.*, vol. 80, pp. 17–25, Nov. 2018.

[5] Z. Ge, "Review on data-driven modeling and monitoring for plant-wide industrial processes," *Chemometrics Intell. Lab. Syst.*, vol. 171, pp. 16–25, Dec. 2017.

[6] S. J. Qin, "Process data analytics in the era of big data," *AIChE J.*, vol. 60, no. 9, pp. 3092–3100, Sep. 2014.

[7] Q. Jiang and B. Huang, "Distributed monitoring for large-scale processes based on multivariate statistical analysis and Bayesian method," *J. Process Control*, vol. 46, pp. 75–83, Oct. 2016.

[8] S. J. Qin, "Statistical process monitoring: Basics and beyond," *J. Chemometrics*, vol. 17, nos. 8–9, pp. 480–502, Aug./Sep. 2003.

[9] J. Zhu, Z. Ge, and Z. Song, "Non-Gaussian industrial process monitoring with probabilistic independent component analysis," *IEEE Trans. Autom. Sci. Eng.*, vol. 14, no. 2, pp. 1309–1319, Apr. 2017.

[10] J. Feng, J. Wang, H. Zhang, and Z. Han, "Fault diagnosis method of joint fisher discriminant analysis based on the local and global manifold learning and its Kernel version," *IEEE Trans. Autom. Sci. Eng.*, vol. 13, no. 1, pp. 122–133, Jan. 2016.

[11] Y. Jiang and S. Yin, "Recent advances in key-performance-indicator oriented prognosis and diagnosis with a MATLAB toolbox: DB-KIT," *IEEE Trans. Ind. Informat.*, to be published. doi: 10.1109/TII.2018.2875067.

[12] P. Nomikos and J. F. Macgregor, "Multivariate SPC charts for monitoring batch processes," *Technometrics*, vol. 37, no. 1, pp. 41–59, Feb. 1995.

[13] Y. Jiang and S. Yin, "Recursive total principle component regression based fault detection and its application to vehicular cyber-physical systems," *IEEE Trans. Ind. Informat.*, vol. 14, no. 4, pp. 1415–1423, Apr. 2018.

[14] Q. Jiang, X. Yan, and B. Huang, "Performance-driven distributed PCA process monitoring based on fault-relevant variable selection and Bayesian inference," *IEEE Trans. Ind. Electron.*, vol. 63, no. 1, pp. 377–386, Jan. 2016.

[15] G. Li, S. J. Qin, and D. Zhou, "Geometric properties of partial least squares for process monitoring," *Automatica*, vol. 46, no. 1, pp. 204–210, Jan. 2010.

[16] Y. Jiang and S. Yin, "Recent results on key performance indicator oriented fault detection using the DB-KIT toolbox," presented at the-43rd Annu. Conf. IEEE Ind. Electron. Soc. (IECON), Beijing, China, Oct./Nov. 2017, pp. 7103–7108.

[17] Z. Chen, S. X. Ding, K. Zhang, Z. Li, and Z. Hu, "Canonical correlation analysis-based fault detection methods with application to alumina evaporation process," *Control Eng. Pract.*, vol. 46, pp. 51–58, Jan. 2016.

[18] Z. Chen, K. Zhang, S. X. Ding, Y. A. W. Shardt, and Z. Hu, "Improved canonical correlation analysis-based fault detection methods for industrial processes," *J. Process Control*, vol. 41, pp. 26–34, May 2016.

[19] Q. Jiang, S. X. Ding, Y. Wang, and X. Yan, "Data-driven distributed local fault detection for large-scale processes based on the GA-regularized canonical correlation analysis," *IEEE Trans. Ind. Electron.*, vol. 64, no. 10, pp. 8148–8157, Oct. 2017.

[20] Q. Jiang, F. Gao, H. Yi, and X. Yan, "Multivariate statistical monitoring of key operation units of batch processes based on time-slice CCA," *IEEE Trans. Control Syst. Technol.*, to be published. doi: 10.1109/TCST.2018.2803071.

[21] L. H. Chiang, E. L. Russell, and R. D. Braatz, *Fault Detection and Diagnosis in Industrial Systems*. London, U.K.: Springer Verlag, 2001.

[22] Q. Liu, Q. Zhu, S. J. Qin, and T. Chai, "Dynamic concurrent kernel CCA for strip-thickness relevant fault diagnosis of continuous annealing processes," *J. Process Control*, vol. 67, pp. 12–22, Jul. 2017.

[23] B. Jiang, D. Huang, X. Zhu, F. Yang, and R. D. Braatz, "Canonical variate analysis-based contributions for fault identification," *J. Process Control*, vol. 26, pp. 17–25, Feb. 2015.

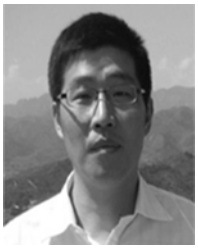
[24] Q. Zhu, Q. Liu, and S. J. Qin, "Concurrent quality and process monitoring with canonical correlation analysis," *J. Process Control*, vol. 60, pp. 95–103, Dec. 2017.

[25] Q. Jiang, X. Yan, and B. Huang, "Neighborhood variational Bayesian multivariate analysis for distributed process monitoring with missing data," *IEEE Trans. Control Syst. Technol.*, to be published. doi: 10.1109/TCST.2018.2870570.

[26] Z. Chen, S. X. Ding, T. Peng, C. Yang, and W. Gui, "Fault detection for non-Gaussian processes using generalized canonical correlation analysis and randomized algorithms," *IEEE Trans. Ind. Electron.*, vol. 65, no. 2, pp. 1559–1567, Feb. 2018.
[27] Y. Liu, B. Liu, X. Zhao, and M. Xie, "A mixture of variational canonical correlation analysis for nonlinear and quality-relevant process monitoring," *IEEE Trans. Ind. Electron.*, vol. 65, no. 8, pp. 6478–6486, Aug. 2018.
[28] Q. Jiang and X. Yan, "Learning deep correlated representations for nonlinear process monitoring," *IEEE Trans. Ind. Informat.*, to be published. doi: 10.1109/TII.2018.2886048.
[29] X. Peng, Y. Tang, W. Du, and F. Qian, "Multimode process monitoring and fault detection: A sparse modeling and dictionary learning method," *IEEE Trans. Ind. Electron.*, vol. 64, no. 6, pp. 4866–4875, Jun. 2017.
[30] W. Du, Y. Fan, and Y. Zhang, "Multimode process monitoring based on data-driven method," *J. Franklin Inst.*, vol. 354, no. 6, pp. 2613–2627, Apr. 2016.
[31] L. Zhou, J. Zheng, Z. Ge, Z. Song, and S. Shan, "Multimode process monitoring based on switching autoregressive dynamic latent variable model," *IEEE Trans. Ind. Electron.*, vol. 65, no. 10, pp. 8184–8194, Oct. 2018.
[32] C. Zhao, Y. Yao, F. Gao, and F. Wang, "Statistical analysis and online monitoring for multimode processes with between-mode transitions," *Chem. Eng. Sci.*, vol. 65, no. 22, pp. 5961–5975, Nov. 2010.
[33] C. Zhao, W. Wang, Y. Qin, and F. Gao, "Comprehensive subspace decomposition with analysis of between-mode relative changes for multimode process monitoring," *Ind. Eng. Chem. Res.*, vol. 54, no. 12, pp. 3154–3166, Mar. 2015.
[34] K. Peng, K. Zhang, B. You, J. Dong, and Z. Wang, "A quality-based nonlinear fault diagnosis framework focusing on industrial multimode batch processes," *IEEE Trans. Ind. Electron.*, vol. 63, no. 4, pp. 2615–2624, Apr. 2016.
[35] W. C. Sang, H. P. Jin, and I.-B. Lee, "Process monitoring using a Gaussian mixture model via principal component analysis and discriminant analysis," *Comput. Chem. Eng.*, vol. 28, no. 8, pp. 1377–1387, Jul. 2004.
[36] M. A. T. Figueiredo and A. K. Jain, "Unsupervised learning of finite mixture models," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 24, no. 3, pp. 381–396, Mar. 2002.
[37] J. Yu and S. J. Qin, "Multimode process monitoring with Bayesian inference-based finite Gaussian mixture models," *AIChE J.*, vol. 54, no. 7, pp. 1811–1829, Jul. 2010.
[38] C. Bishop, *Pattern Recognition and Machine Learning* (Information Science and Statistics). New York, NY, USA: Springer-Verlag, 2006.
[39] N. Nasios and A. G. Bors, "Variational learning for Gaussian mixture models," *IEEE Trans. Syst., Man, Cybern. B, Cybern.*, vol. 36, no. 4, pp. 849–862, Aug. 2006.
[40] M.-S. Chen, H.-F. Wang, C.-P. Hwang, T.-Y. Ho, and C.-H. Hung, "A variational Bayesian approach for unsupervised clustering," in *Frontier Computing*. Singapore: Springer, 2016, pp. 651–660.
[41] R. A. Johnson and D. W. Wichern, *Applied Multivariate Statistical Analysis*, vol. 5. Upper Saddle River, NJ, USA: Prentice-Hall, 2002.
[42] C. Zhao and B. Huang, "A full-condition monitoring method for nonstationary dynamic chemical processes with cointegration and slow feature analysis," *AIChE J.*, vol. 64, no. 5, pp. 1662–1681, May 2018.
[43] G. Birol, C. Ündey, and A.Çinar, "A modular simulation package for fed-batch fermentation: Penicillin production," *Comput. Chem. Eng.*, vol. 26, no. 11, pp. 1553–1565, Nov. 2002.
[44] X. Zhang, M. Kano, and Y. Li, "Locally weighted kernel partial least squares regression based on sparse nonlinear features for virtual sensing of nonlinear time-varying processes," *Comput. Chem. Eng.*, vol. 104, pp. 164–171, Sep. 2017.
[45] Y. Wang, Q. Jiang, B. Li, and L. Cui, "Joint-individual monitoring of parallel-running batch processes based on MCCA," *IEEE Access*, vol. 6, pp. 13005–13014, 2017. doi: 10.1109/ACCESS.2017.2784097.
[46] J.-M. Lee, C. K. Yoo, and I.-B. Lee, "Statistical process monitoring with independent component analysis," *J. Process Control*, vol. 14, pp. 467–485, 2004.

Qingchao Jiang (M'17) received the B.E. and Ph.D. degrees from the Department of Automation, East China University of Science and Technology, Shanghai, China, in 2010 and 2015, respectively.

In 2015, he joined the Department of Chemical and Materials Engineering, University of Alberta, Edmonton, AB, Canada, as a Post-Doctoral Fellow. From 2015 to 2016, he was a Humboldt Research Fellow with the Institute for Automatic Control and Complex Systems, University of Duisburg–Essen, Duisburg, Germany. From 2016 to 2017, he was a Visiting Research Fellow with the Department of Chemical and Biomolecular Engineering, The Hong Kong University of Science and Technology, Hong Kong. He is currently an Associate Professor with the East China University of Science and Technology. His research interests include data mining, soft sensing, multivariate statistical process monitoring, and deep learning-based process modeling and monitoring.



Xuefeng Yan received the B.S. degree in biochemical engineering and the Ph.D. degree in control theory engineering from Zhejiang University, Hangzhou, China, in 1995 and 2002, respectively.

He is currently a Professor with the East China University of Science and Technology, Shanghai, China. His current research interests include complex chemical process modeling, optimizing and controlling, process monitoring, fault diagnosis, and intelligent information processing.