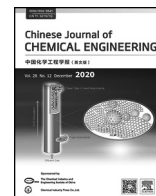




Contents lists available at ScienceDirect

Chinese Journal of Chemical Engineering

journal homepage: www.elsevier.com/locate/CJChE

Article

Multimodal process monitoring based on transition-constrained Gaussian mixture model

Shutian Chen, Qingchao Jiang^{*}, Xuefeng Yan

Key Laboratory of Advanced Control and Optimization for Chemical Processes of Ministry of Education, East China University of Science and Technology, Shanghai 200237, China

ARTICLE INFO

Article history:

Received 24 February 2020

Received in revised form 5 July 2020

Accepted 5 August 2020

Available online 8 September 2020

Keywords:

Multimodal process monitoring

Gaussian mixture model

State transition matrix

Process control

Process systems

Systems engineering

ABSTRACT

Reliable process monitoring is important for ensuring process safety and product quality. A production process is generally characterized by multiple operation modes, and monitoring these multimodal processes is challenging. Most multimodal monitoring methods rely on the assumption that the modes are independent of each other, which may not be appropriate for practical application. This study proposes a transition-constrained Gaussian mixture model method for efficient multimodal process monitoring. This technique can reduce falsely and frequently occurring mode transitions by considering the time series information in the mode identification of historical and online data. This process enables the identified modes to reflect the stability of actual working conditions, improve mode identification accuracy, and enhance monitoring reliability in cases of mode overlap. Case studies on a numerical simulation example and simulation of the penicillin fermentation process are provided to verify the effectiveness of the proposed approach in multimodal process monitoring with mode overlap.

© 2020 The Chemical Industry and Engineering Society of China, and Chemical Industry Press Co., Ltd.

All rights reserved.

1. Introduction

Since the end of the previous century, process monitoring has been widely explored by the academia and industries [1,2]. Modern industries must ensure the safety of production operations, stability of product quality, and sustainability of factory profits. Establishing mathematical models of processes is difficult due to the increasing complexity of modern industries, and accurately monitoring all types of faults is challenging. Data-driven modeling methods have been widely used with the development of sensing and computing technologies. Methods represented by multivariate statistical process monitoring (MSPM) are the most widely utilized among data-driven approaches [3–5].

Many studies have been conducted on MSPM, but most of their methods assume that process data originate from an operating mode. The data are assumed to follow a single-mode distribution. However, the operation mode in industrial processes often changes due to modifications in raw materials, product specifications, set values, manufacturing strategies, and other factors. When operating conditions change, the statistics of the data produced by the process operation may vary considerably. Therefore, process data present a multimodal distribution, which renders traditional MSPM technology unsuitable. Many improved methods have been proposed to enhance the monitoring of multimodal processes. These improved methods based on MSPM can generally be divided into the following categories: global model, robust

adaptive model, multi-model, and mixed-model monitoring methods. Global model monitoring methods [6–10] use the concept of global modeling to build a unified model of multimodal processes. Based on this idea, Hwang and Han proposed a super principal component analysis (PCA) model based on hierarchical clustering [6]. Robust adaptive model monitoring methods [11–16] enable new normal data to adapt to changes in modes. For example, Wold *et al.* [11] proposed a method to update PCA and partial least squares (PLS) models using the concept of exposure weighted movement. Ge *et al.* [14] used the just-in-time learning (JITL) strategy to solve the multimodal problem which solved the multimodal process monitoring problem by updating the local model online. By combining process knowledge, Lee *et al.* [16] used if-then rules to identify operational mode changes. Multi-model monitoring methods [17,18] establish a corresponding sub-model for each sub-mode and subsequently construct a model library that can represent all modes of complex multimodal industrial processes. A local batch process monitoring method developed by Kyoo [17] that uses MPCA model and multi-probability model is a typical multi-model monitoring method. Mixed-model monitoring methods establish Gaussian mixture model (GMM) [19–25], implicit Hidden Markov model (HMM) [26–31], and other mixed model for certain processes.

Most current studies assume that the modes in multimodal processes are independent of each other, and the data of each mode do not overlap [32–34]. However, separating the modes completely is difficult in practice. Therefore, this study focuses on improving the identification rate in the case of mode overlap and proposes a multimodal

^{*} Corresponding author.E-mail address: qchjiang@ecust.edu.cn (Q. Jiang).

process monitoring method. This approach is based on a transition-constrained Gaussian mixture model (TCGMM). By considering the time series of process data, TCGMM improves the mode identification results of historical and online data when overlapping exists.

First, the Figueiredo-Jain (F-J) algorithm is used to establish GMM [35]. According to the clustering results of historical data, the mode of each sample are marked, and then the mode identification results of historical data are updated according to the fact that the actual operating conditions of the process do not change frequently. Second, a state transition matrix is calculated in accordance with the modes of historical data through the Bayesian method. Lastly, the transition matrix is utilized to improve online mode identification.

The rest of this paper is organized as follows. Section 2 briefly introduces the basics of GMM and the Bayesian method. Section 3 describes TCGMM in detail. Section 4 presents two case studies, namely, a numerical simulation experiment and a penicillin fermentation process simulation. Section 5 concludes this work.

2. Preliminaries

2.1. GMM

Suppose that $\mathbf{x} \in \mathbf{R}^m$ is an m -dimension sample from a multimodal process. The following GMM is used to represent the probability density function of $\mathbf{x}$.

$$p(\mathbf{x}|\boldsymbol{\theta}) = \sum_{i=1}^K \omega_i g(\mathbf{x}|\boldsymbol{\theta}_i) \quad (1)$$

where K represents the number of Gaussian components in GMM, ω_i represents the weight of the i th Gaussian component, and $\boldsymbol{\theta} = \{\boldsymbol{\theta}_1, \dots, \boldsymbol{\theta}_K\} = \{\boldsymbol{\mu}_1, \boldsymbol{\Sigma}_1, \dots, \boldsymbol{\mu}_K, \boldsymbol{\Sigma}_K\}$ and $\boldsymbol{\theta}_i = \{\boldsymbol{\mu}_i, \boldsymbol{\Sigma}_i\}$ represent global and local Gaussian model parameter sets, respectively (where $\boldsymbol{\mu}_i$ is the mean and $\boldsymbol{\Sigma}_i$ is the covariance). The multivariate Gaussian distribution density function for the i th Gaussian component C_i is

$$g(\mathbf{x}|\boldsymbol{\theta}_i) = \frac{1}{(2\pi)^{\frac{m}{2}} |\boldsymbol{\Sigma}_i|^{\frac{1}{2}}} \exp \left[-\frac{1}{2} (\mathbf{x} - \boldsymbol{\mu}_i)^T \boldsymbol{\Sigma}_i^{-1} (\mathbf{x} - \boldsymbol{\mu}_i) \right]. \quad (2)$$

The mixture density function $p(\mathbf{x}|\boldsymbol{\theta})$ is the weighted sum of all Gaussian components, and its weight ω_i reflects the prior probability that any sampling point belongs to the i th Gaussian component. The integral of the density function of the mixture model should satisfy

$$\int_{\mathbf{R}^m} p(\mathbf{x}|\boldsymbol{\theta}) d\mathbf{x} = \sum_{i=1}^K \omega_i \left(\int_{\mathbf{R}^m} g(\mathbf{x}|\boldsymbol{\theta}_i) d\mathbf{x} \right) = 1 \quad (3)$$

Given that $\int_{\mathbf{R}^m} g(\mathbf{x}|\boldsymbol{\theta}_i) d\mathbf{x} = 1$ is applicable to each Gaussian component, the following equation can be obtained.

$$\sum_{i=1}^K \omega_i = 1 \quad (4)$$

Thus, the prior probability is $0 \leq \omega_i \leq 1$. The population mean of multimodal operation data can also be calculated as follows:

$$\boldsymbol{\mu}_x = \int_{\mathbf{R}^m} \mathbf{x} p(\mathbf{x}|\boldsymbol{\theta}) d\mathbf{x} = \sum_{i=1}^K \omega_i \left(\int_{\mathbf{R}^m} \mathbf{x} g(\mathbf{x}|\boldsymbol{\theta}_i) d\mathbf{x} \right) = \sum_{i=1}^K \omega_i \boldsymbol{\mu}_i. \quad (5)$$

According to Eqs. (1) and (2), unknown parameter $\boldsymbol{\theta} = \{\omega_1, \boldsymbol{\mu}_1, \boldsymbol{\Sigma}_1, \dots, \omega_K, \boldsymbol{\mu}_K, \boldsymbol{\Sigma}_K\}$ needs to be estimated to construct GMM. In this study, the parameters are estimated with the F-J algorithm, which was proposed by Figueiredo and Jain [35].

Through repeated calculation of the expectation step (E-step) and maximization step (M-step), the F-J algorithm estimates the

corresponding distribution parameters until the convergence criterion of logarithmic likelihood function is met to achieve iteration. The algorithm can start from any given number of Gaussian components. Then, it adjusts the number of Gaussian components automatically by deleting components with unimportant weights.

The training data set is $\mathbf{X}$, and the initial parameter value is $\boldsymbol{\theta}^{(0)} = \{\omega_1^{(0)}, \boldsymbol{\mu}_1^{(0)}, \boldsymbol{\Sigma}_1^{(0)}, \dots, \omega_K^{(0)}, \boldsymbol{\mu}_K^{(0)}, \boldsymbol{\Sigma}_K^{(0)}\}$. E-step and M-step are alternated.

E-step:

$$P^{(s)}(C_k|\mathbf{x}_j) = \omega_k^{(s)} g(\mathbf{x}_j|\boldsymbol{\mu}_k^{(s)}, \boldsymbol{\Sigma}_k^{(s)}) / \sum_{i=1}^K \omega_i^{(s)} g(\mathbf{x}_j|\boldsymbol{\mu}_i^{(s)}, \boldsymbol{\Sigma}_i^{(s)}) \quad (6)$$

where $P^{(s)}(C_k|\mathbf{x}_j)$ is the posterior probability of $\mathbf{x}_j$ belonging to C_k during the s th iteration.

M-step:

$$\boldsymbol{\mu}_k^{(s+1)} = \sum_{j=1}^n P^{(s)}(C_k|\mathbf{x}_j) \mathbf{x}_j / \sum_{j=1}^n P^{(s)}(C_k|\mathbf{x}_j), \quad (7)$$

$$\boldsymbol{\Sigma}_k^{(s+1)} = \sum_{j=1}^n P^{(s)}(C_k|\mathbf{x}_j) (\mathbf{x}_j - \boldsymbol{\mu}_k^{(s+1)}) (\mathbf{x}_j - \boldsymbol{\mu}_k^{(s+1)})^T / \sum_{j=1}^n P^{(s)}(C_k|\mathbf{x}_j), \quad (8)$$

$$\omega_k^{(s+1)} = \max \left\{ 0, \left(\sum_{j=1}^n P^{(s)}(C_k|\mathbf{x}_j) \right) - \frac{V}{2} \right\} / \sum_{i=1}^K \max \left\{ 0, \left(\sum_{j=1}^n P^{(s)}(C_k|\mathbf{x}_j) \right) - \frac{V}{2} \right\}, \quad (9)$$

where $\boldsymbol{\mu}_k^{(s+1)}$, $\boldsymbol{\Sigma}_k^{(s+1)}$, and $\omega_k^{(s+1)}$ are the mean value, covariance, and prior probability, respectively, of C_k at the $(s+1)$ th iteration. $V = \frac{1}{2} m^2 + \frac{3}{2} m$ is the number of scalar parameters of each specified Gaussian component.

2.2. Bayesian method

Bayesian method is an efficient tool in probability estimation which takes both the historical data and prior knowledge into consideration [36]. Let $\psi_{\boldsymbol{\theta}_i|\boldsymbol{\theta}_j}$ be defined as the probability that in historical data $\mathbf{D}$, the mode of operation of the data at the previous time is $\boldsymbol{\theta}_j$. Then, the probability of the data at this time being in mode $\boldsymbol{\theta}_i$ is

$$P(\mathbf{x}_{new} \in \boldsymbol{\theta}_i | \boldsymbol{\theta}_j, \mathbf{D}) = P(\boldsymbol{\theta}_i | \mathbf{D}_{\boldsymbol{\theta}_j}) = \psi_{\boldsymbol{\theta}_i|\boldsymbol{\theta}_j}. \quad (10)$$

$\boldsymbol{\Psi}_{\boldsymbol{\theta}_j}$ is the vector $\boldsymbol{\Psi}_{\boldsymbol{\theta}_j} = \{\psi_{\boldsymbol{\theta}_1|\boldsymbol{\theta}_j}, \psi_{\boldsymbol{\theta}_2|\boldsymbol{\theta}_j}, \dots, \psi_{\boldsymbol{\theta}_K|\boldsymbol{\theta}_j}\}$ of $\psi_{\boldsymbol{\theta}_j}$, where K represents the mode number. The following can be obtained in accordance with Bayes' rule.

$$f(\boldsymbol{\Psi}_{\boldsymbol{\theta}_j} | \mathbf{D}_{\boldsymbol{\theta}_j}) = P(\boldsymbol{\Psi}_{\boldsymbol{\theta}_j} | \mathbf{D}_{\boldsymbol{\theta}_j}) f(\boldsymbol{\Psi}_{\boldsymbol{\theta}_j}) / P(\mathbf{D}_{\boldsymbol{\theta}_j}) \quad (11)$$

A priori model is established by using the Dirichlet distribution [36] with parameter $\alpha(\boldsymbol{\theta}_i | \mathbf{D}_{\boldsymbol{\theta}_j})$, and the following likelihood parameter model is obtained.

$$f(\boldsymbol{\Psi}_{\boldsymbol{\theta}_j} | \boldsymbol{\theta}_j) = \Gamma \left[\sum_{i=1}^K \alpha(\boldsymbol{\theta}_i | \mathbf{D}_{\boldsymbol{\theta}_j}) \right] \cdot \prod_{i=1}^K \psi_{\boldsymbol{\theta}_i|\boldsymbol{\theta}_j}^{\alpha(\boldsymbol{\theta}_i | \mathbf{D}_{\boldsymbol{\theta}_j}) - 1} / \prod_{i=1}^K \Gamma \left[\sum_{i=1}^K \alpha(\boldsymbol{\theta}_i | \mathbf{D}_{\boldsymbol{\theta}_j}) \right] \quad (12)$$

Let $n(\boldsymbol{\theta}_i | \mathbf{D}_{\boldsymbol{\theta}_j})$ be numbers of data in historical data, with the previous data running in mode $\boldsymbol{\theta}_j$ and the current data running in mode $\boldsymbol{\theta}_i$. Then,

$$P(\mathbf{D}_{\boldsymbol{\theta}_j} | \boldsymbol{\Psi}_{\boldsymbol{\theta}_j}) = \prod_{i=1}^K \psi_{\boldsymbol{\theta}_i|\boldsymbol{\theta}_j}^{n(\boldsymbol{\theta}_i | \mathbf{D}_{\boldsymbol{\theta}_j})} \quad (13)$$

The parameter $n(\boldsymbol{\theta}_i | \mathbf{D}_{\boldsymbol{\theta}_j}) + \alpha(\boldsymbol{\theta}_i | \mathbf{D}_{\boldsymbol{\theta}_j})$ is obtained after updating the posterior distribution (11) with (12) and (13). Then, $\psi_{\boldsymbol{\theta}_i|\boldsymbol{\theta}_j}$ is calculated by replacing theoretical values with values considering prior knowledge.

$$\psi_{\theta_i|\theta_j} = (n(\theta_i|\mathbf{D}_{\theta_j}) + \alpha(\theta_i|\mathbf{D}_{\theta_j})) / \sum_{l=1}^K [n(\theta_l|\mathbf{D}_{\theta_j}) + \alpha(\theta_l|\mathbf{D}_{\theta_j})] \quad (14)$$

That is,

$$P(\mathbf{x}_{new} \in \theta_i | \theta_j, \mathbf{D}) = (n(\theta_i|\mathbf{D}_{\theta_j}) + \alpha(\theta_i|\mathbf{D}_{\theta_j})) / \sum_{l=1}^K [n(\theta_l|\mathbf{D}_{\theta_j}) + \alpha(\theta_l|\mathbf{D}_{\theta_j})]. \quad (15)$$

3. TCGMM-based process monitoring

The TCGMM method considers the time series of data, which is mainly embodied in the improvement of the mode identification of historical and online data.

When improving the mode identification of historical data, it is assumed that the operating conditions of the process do not change frequently, which means that the mode of historical data does not change frequently. Therefore, in the evaluation of the mode of historical data, mode transfer is assumed to occur only when the mode of successive sample points is transferred to the same mode; otherwise, the mode is unchanged. In consideration of the correlation between modes and time, the transfer frequency of identified modes can be reduced, and the mode of each sample point in the historical data with mode overlap characteristics can be identified well.

The mode identification of historical data mainly involves calculating the state transition matrix, which is then applied to the mode identification of online data. The calculated state transition matrix stores the probability of each sub-mode transition. This probability is introduced into the probability calculation of online data mode identification to identify the mode of the overlapped data and improve the mode identification rate.

The following are the monitoring steps.

Step 1 Mode identification of historical data

There are two steps in the mode identification of historical data. The first step is clustering by using F-J algorithm and marking mode. The second step is updating the mode identification result by considering the time series. A set of training data that needs to contain data under

all operational conditions is generated. The model parameter set $\Theta = \{\{\omega_1, \mu_1, \Sigma_1\}, \dots, \{\omega_K, \mu_K, \Sigma_K\}\}$ of historical data is estimated with the F-J algorithm, and GMM is obtained. The probability that each sample point of historical data belongs to each mode is calculated according to Eq. (16). For each sample $\mathbf{x}$, θ_i with the maximum probability $g(\mathbf{x}|\theta_i)$ is the current sample mode. The mode for each sample is temporarily marked.

$$g(\mathbf{x}|\theta_i) = \exp\left[-\frac{1}{2}(\mathbf{x}-\mu_i)^T \Sigma_i^{-1}(\mathbf{x}-\mu_i)\right] / (2\pi)^{\frac{m}{2}} |\Sigma_i|^{\frac{1}{2}} \quad (16)$$

Given the time series of data, mode transition is assumed to occur only when the modes of three consecutive sample points are transferred to the same mode. Otherwise, the mode is unchanged. In this way, the mode of each sample point temporarily marked is updated, and the modes of the frequently transferred sample points are changed to obtain the final historical data modes.

Step 2 Calculation of the state transition matrix

Here the state transition of a process is assumed to be governed by a Markov process. Then the state transition matrix can be obtained by calculating the probabilities of mode transition. The state transition matrix based on the Bayesian method is calculated by using the obtained historical data mode in combination with prior knowledge $\alpha(\theta_i|\mathbf{D}_{\theta_j})$ in accordance with

$$P(\mathbf{x} \in \theta_i | \theta_j, \mathbf{D}) = \psi_{\theta_i|\theta_j} = (n(\theta_i|\mathbf{D}_{\theta_j}) + \alpha(\theta_i|\mathbf{D}_{\theta_j})) / \sum_{l=1}^K [n(\theta_l|\mathbf{D}_{\theta_j}) + \alpha(\theta_l|\mathbf{D}_{\theta_j})],$$

where, $n(\theta_i|\mathbf{D}_{\theta_j})$ represents the number of sample points transferred from mode θ_j to θ_i in the historical data, and K represents the number of modes in the historical data.

Step 3 Mode identification of online data

Mode identification of online data considers the possibility of mode from two dimensions. The first dimension is the spatial dimension. By using the probability density function, the probability of test data $\mathbf{x}_{new}$ in each mode is calculated without considering the time series.

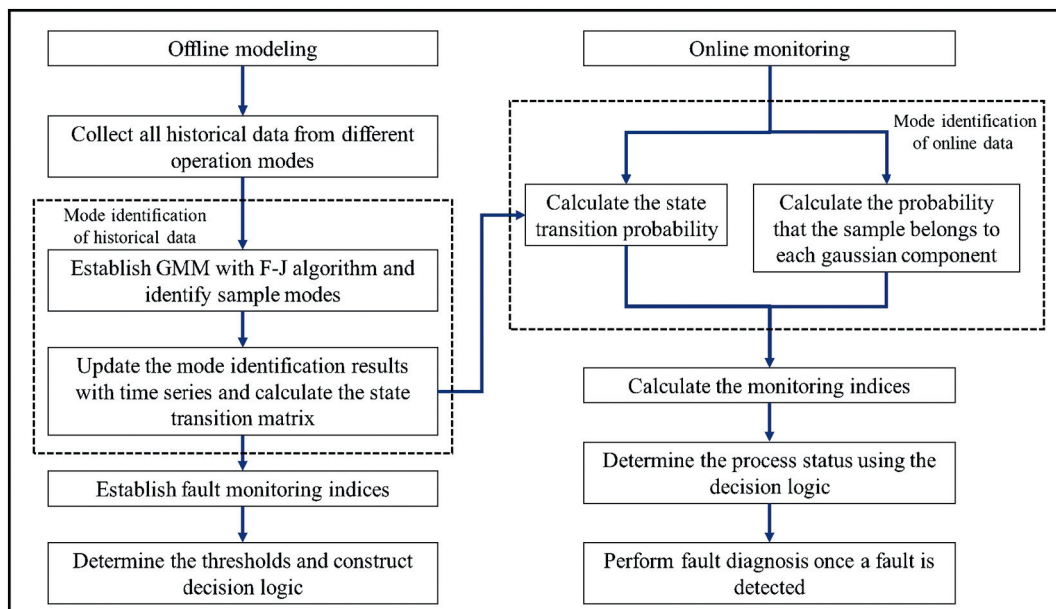


Fig. 1. Flowchart of the TCGMM monitoring.

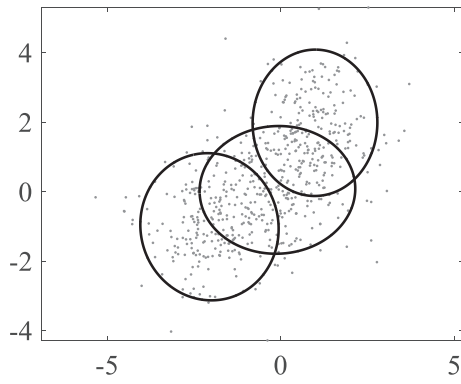


Fig. 2. Numerical example of multimodal simulation based on the F-J algorithm.

$$g(\mathbf{x}_{new}|\theta_i) = \exp\left[-\frac{1}{2}(\mathbf{x}_{new}-\boldsymbol{\mu}_i)^T \boldsymbol{\Sigma}_i^{-1}(\mathbf{x}_{new}-\boldsymbol{\mu}_i)\right] / \left((2\pi)^{\frac{m}{2}} \cdot |\boldsymbol{\Sigma}_i|^{\frac{1}{2}}\right) \quad (17)$$

The second dimension is the temporal dimension. According to the mode at the previous time and the state transition matrix obtained in the online modeling step, the probability that $\mathbf{x}_{new}$ belongs to each mode from the temporal dimension can be obtained, which is denoted as $\psi_{\theta_i|\theta_j}$.

Based on the above two dimensions, the transition-constrained probability is defined and calculated as $P_t(\mathbf{x}_{new}|\theta_i, t) = g(\mathbf{x}_{new}|\theta_i) \times \psi_{\theta_i|\theta_j}$. The mode θ_i with the largest transition-constrained probability $P_t(\mathbf{x}_{new}|\theta_i, t)$ is identified as the current operation mode. Here we explain the rationality of introducing this transition-constrained probability for identifying the operating mode. For the time instant $t_{x_{new}} = k$, we can obtain the $g(\mathbf{x}_{new}|\theta_i)$. The possibility of the current state θ_i from the temporal dimension is estimated by using state θ_n at the time instant $t = k - 2$, namely $P(\mathbf{x}_k \in \theta_i | \mathbf{x}_{k-2} \in \theta_n) = \psi(\hat{\mathbf{x}}_k \in \theta_i | \hat{\mathbf{x}}_{k-1} \in \theta_j) \times \psi(\hat{\mathbf{x}}_{k-1} \in \theta_j | \mathbf{x}_{k-2} \in \theta_n) = \psi_{\theta_i|\theta_j} \times \psi_{\theta_j|\theta_n}$. Since it is assumed that the state transition is governed by a Markov process, $g(\mathbf{x}_{new}|\theta_i)$ and $P(\mathbf{x}_k \in \theta_i | \mathbf{x}_{k-2} \in \theta_n)$ are independent of each other. It is reasonable to form a new probability as $P(\mathbf{x}_{new}|\theta_i, t) = g(\mathbf{x}_{new}|\theta_i) \times P(\mathbf{x}_k \in \theta_i | \mathbf{x}_{k-2} \in \theta_n)$. At the time instant k , since the state θ_j at the time $k - 1$ and the state θ_n at the time $k - 2$ are known, the $\psi_{\theta_j|\theta_n}$ can be obtained. Then $\theta_i = \arg \max_{\theta_i} [P(\mathbf{x}_{new}|\theta_i, t)] = \arg \max_{\theta_i} [g(\mathbf{x}_{new}|\theta_i) \times \psi_{\theta_i|\theta_j} \times \psi_{\theta_j|\theta_n}] = \arg \max_{\theta_i} [g(\mathbf{x}_{new}|\theta_i) \times \psi_{\theta_i|\theta_j}]$.

Step 4 Fault detection

Online mode identification is completed, and the process status is determined by calculating the monitoring index and using judgment logic. T^2 and BIP indexes are constructed for evaluation and comparison.

Construction of the T^2 index:

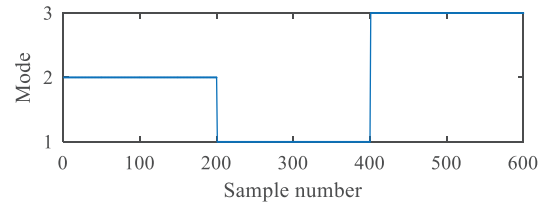
Let $\mathbf{x} \in \mathbf{R}^m$, where $\mathbf{X} \in \mathbf{R}^{n \times m}$ is a normalized matrix composed of n samples $\mathbf{x}$. Spectral decomposition of the covariance matrix $cov(\mathbf{x})$ is conducted.

$$cov(\mathbf{x}) \approx (\mathbf{X}^T \mathbf{X}) / (n-1) = \mathbf{P} \boldsymbol{\Lambda} \mathbf{P}^T, \quad (18)$$

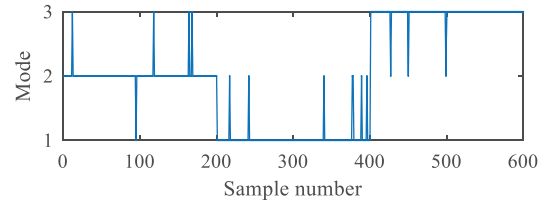
where $\boldsymbol{\Lambda} = \text{diag} \{\lambda_1, \lambda_2, \dots, \lambda_m\}$ represents the eigenvalue matrix. $\mathbf{P} = [\mathbf{P}\mathbf{P}]$ contains the subspaces of the principal component load matrix and residual subspaces. Assuming that $\boldsymbol{\Lambda}_l = \text{diag} \{\lambda_1, \lambda_2, \dots, \lambda_l\}$, statistical metrics can be used for process monitoring. Then, the T^2 index is calculated as follows:

$$T^2 = \mathbf{x}^T \hat{\mathbf{P}} \boldsymbol{\Lambda}_l^{-1} \hat{\mathbf{P}}^T \mathbf{x}. \quad (19)$$

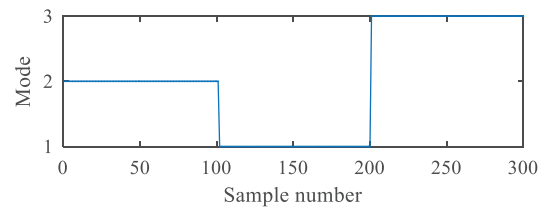
When the number of samples is sufficiently large, the corresponding T^2 index is approximately χ^2 distribution with l degrees of freedom. Therefore, the T^2 control limit with $100(1 - \alpha)\%$ is



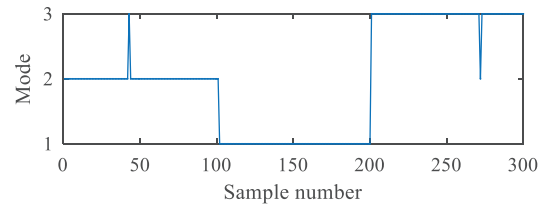
(a)



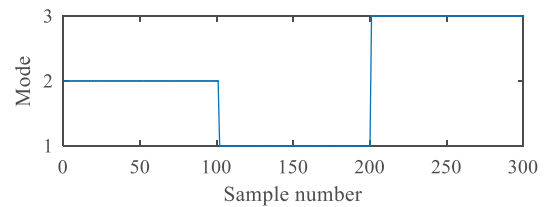
(b)



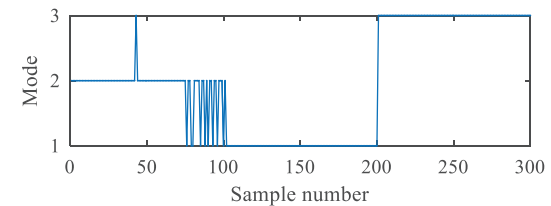
(c)



(d)



(e)



(f)

Fig. 3. Mode identification results for the numerical example. (a) Historical data mode identification results based on TCGMM. (b) Historical data mode identification results based on GMM. (c) Normal online data mode identification results based on TCGMM. (d) Normal online data mode identification results based on GMM. (e) Online fault data mode identification results based on TCGMM. (f) Online fault data mode identification result based on GMM.

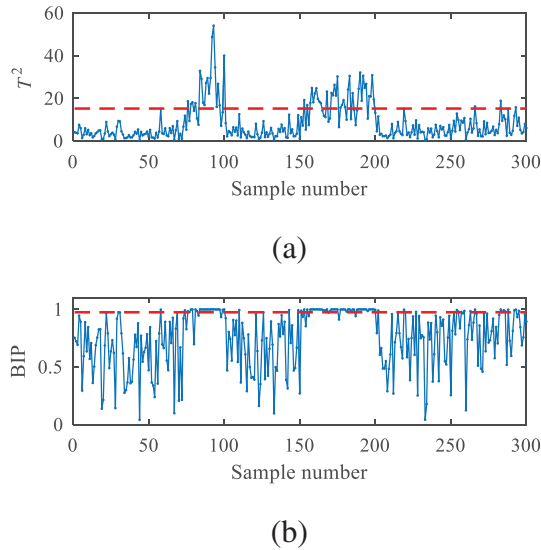


Fig. 4. Monitoring results of online fault data based on TCGMM. (a) T^2 control diagram. (b) BIP control diagram.

$$T_{\alpha}^2 = \chi_{l,\alpha}^2. \quad (20)$$

When $T^2 \leq T_{\alpha}^2$, the process is considered running normally; when $T^2 > T_{\alpha}^2$, the process is deemed abnormal.

Construction of the BIP index:

Literature [20] provided a monitoring index that is based on the Bayesian inference strategy, namely, the BIP index. For online sample data $\mathbf{x}_{new}$, the posterior probability is

$$P(C_k|\mathbf{x}_{new}) = P(\mathbf{x}_{new} \in C_k) = (\omega_k \cdot g(\mathbf{x}_{new}|\boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k)) / \sum_{i=1}^K \omega_i g(\mathbf{x}_{new}|\boldsymbol{\mu}_i, \boldsymbol{\Sigma}_i). \quad (21)$$

The formula of the BIP index of sample data $\mathbf{x}_{new}$ is derived as follows:

$$BIP = \sum_{k=1}^K P(C_k|\mathbf{x}_{new}) P_L^{(k)}(\mathbf{x}_{new}). \quad (22)$$

When $BIP \leq 1 - \alpha$, the process is considered running normally; when $BIP > 1 - \alpha$, the opposite is true.

The flowchart of the proposed TCGMM monitoring is shown in Fig. 1.

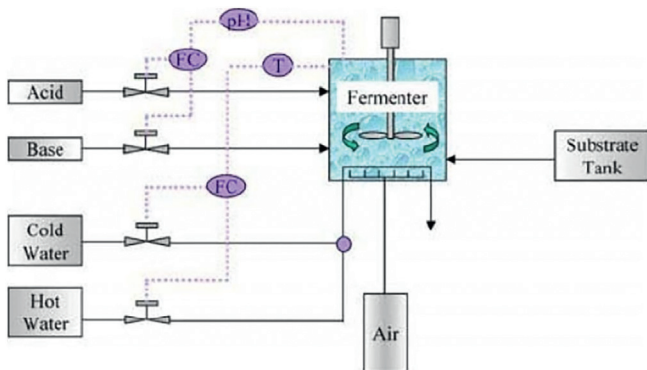


Fig. 5. Penicillin production flowchart [37,38].

Table 1
Measured variables of BPFP [37,38] (1 cal = 4.1868 J)

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

4. Case Study

4.1. Numerical example

In order to obtain the multimode data with mode overlap, a group of data subject to the multivariate Gaussian distribution is generated, which contains three Gaussian components, and the data under each component is generated according to $\mathbf{x}_i = N(\boldsymbol{\mu}_i, \boldsymbol{\Sigma}_i) + \boldsymbol{\varepsilon}_i$, where $\mathbf{x}_i = [m_{i1} \ m_{i2} \ m_{i3} \ m_{i4} \ m_{i5} \ m_{i6}]^T$; $N(\boldsymbol{\mu}_i, \boldsymbol{\Sigma}_i)$ is a Gaussian distribution whose covariance is $\boldsymbol{\Sigma}_i$ and mean is $\boldsymbol{\mu}_i$. The mean values of the three Gaussian components are [0, 0, 0, 0, 0], [1, 2, 1.5, 2.5, 3, 0.5], and [2, 1, 1.5, 0.5, 2.5, 3], respectively. The covariances are

$$\begin{bmatrix} 0.9 & 0.5 & 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 \\ 0.9 & 0.4 & 0 & 0 & 0 & 1 \\ 0.5 & 0.4 & 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix}.$$

Under these mean and covariance conditions, the training data are generated, and each Gaussian component contains 200 points. These sets of data are regarded as normal, unlabeled, and mixed modes and used as training samples. The F-J algorithm is utilized to estimate the GMM results of the training samples. As shown in Fig. 2, the training samples are automatically identified into three modes.

Table 2
Initial conditions and set points to simulate BPFP [37,38]

Initial conditions	Range
Biomass concentration/g·L ⁻¹	0.1
Carbon dioxide concentration/g·L ⁻¹	0.5–1
Dissolved oxygen concentration/g·L ⁻¹	1–1.2
Penicillin concentration/g·L ⁻¹	0
Substrate concentration/g·L ⁻¹	15–17
Culture volume/L	101–103
Bioreactor temperature/K	277–279
Generated heat/kcal	0
Hydrogen ion concentration:[H ⁺]/mol·L ⁻¹	10 ⁻⁵
Set points	Range
Aeration rate/L·h ⁻¹	8–9
Substrate feed rate/L·h ⁻¹	0.039–0.045
Bioreactor temperature/K	297–298
Substrate feed temperature/K	295–296
Agitator power/W	29–31
PH	4.95–5.05

The mode identification results of historical data based on the TCGMM method are shown in Fig. 3(a), and the mode identification results of historical data based on the GMM method are shown in Fig. 3(b). Comparison of the effect diagrams of the mode identification of the two methods shows that the mode of historical data identified by the TCGMM method is not frequently transferred. The accuracy of mode identification is enhanced, and the identification effect is better than that of the original method.

Using the same mean and variance as the historical data, one hundred normal, unlabeled, and mixed samples are generated for the test data in each mode as the normal online data. The mode identification results obtained by the TCGMM method for this set of test samples are shown in Fig. 3(c). The mode identification results obtained by the GMM method are shown in Fig. 3(d). Comparison of the two results indicates that mode identification of the test data has the same optimization effect as that of historical data.

The following test data are generated to evaluate the effectiveness of the proposed TCGMM method for monitoring different faults. In the normally generated test data, in the first mode from 51 to 100 sample points on the second property, we introduce a slope of change with a value of $0.1 \times (t - 50)$. In the second mode from 151 to 200 sample points on the second property, we introduce a step change with a value of 4.5. In the third mode from 251 to 300 sample points on the fifth property, we introduce a step change with a value of 4.5.

The mode identification results of the proposed TCGMM method are presented in Fig. 3(e), and the results of GMM are shown in Fig. 3(f). The identification results indicate that the TCGMM method can perform accurate mode identification of fault data, but the GMM method is ineffective because of its frequent mode transfer. Fig. 4(a) and (b) show the T^2 and BIP control diagrams, respectively, of the TCGMM monitoring method. The two indexes are not sensitive to the fault signals of the

first type of slope variation and have many undetected fault points. As for the fault signal of the second step change, both indexes can be detected well, and the detection rate of the BIP index is better than that of the T^2 index. The positions of the third type of fault signal with step change differ from those of the second type. The control diagram shows that the control effect of using the BIP index is much better than that of using the T^2 index. The undetected rate of the T^2 index is high, whereas that of the BIP index is low. The three fault cases indicate that the monitoring effect of the proposed method differs for various fault types, but the monitoring effect of the BIP index is generally better than that of the T^2 index.

One hundred Monte Carlo simulations are performed to further test the availability of the TCGMM monitoring method and the superiority of the BIP monitoring index over the T^2 monitoring index. The undetected rate based on the BIP and T^2 indexes of the two methods is calculated, and the findings show that the fault rate of the fault samples based on the BIP monitoring index is significantly lower than that based on the T^2 monitoring index. In other words, the BIP control diagram has a better monitoring effect. No obvious difference is observed in the detection rates of fault samples between TCGMM and GMM monitoring methods. However, the data mode identification effect of TCGMM is better than that of GMM; the former provides more accurate information on the real process status.

4.2. Batch-fed penicillin fermentation process

Penicillin is an antibiotic with high clinical value, and the multimodality of its production process is remarkable. As a secondary microbial metabolism process, penicillin fermentation uses specific production bacteria to grow and reproduce under certain conditions. When the production bacteria reproduce to a certain concentration, penicillin

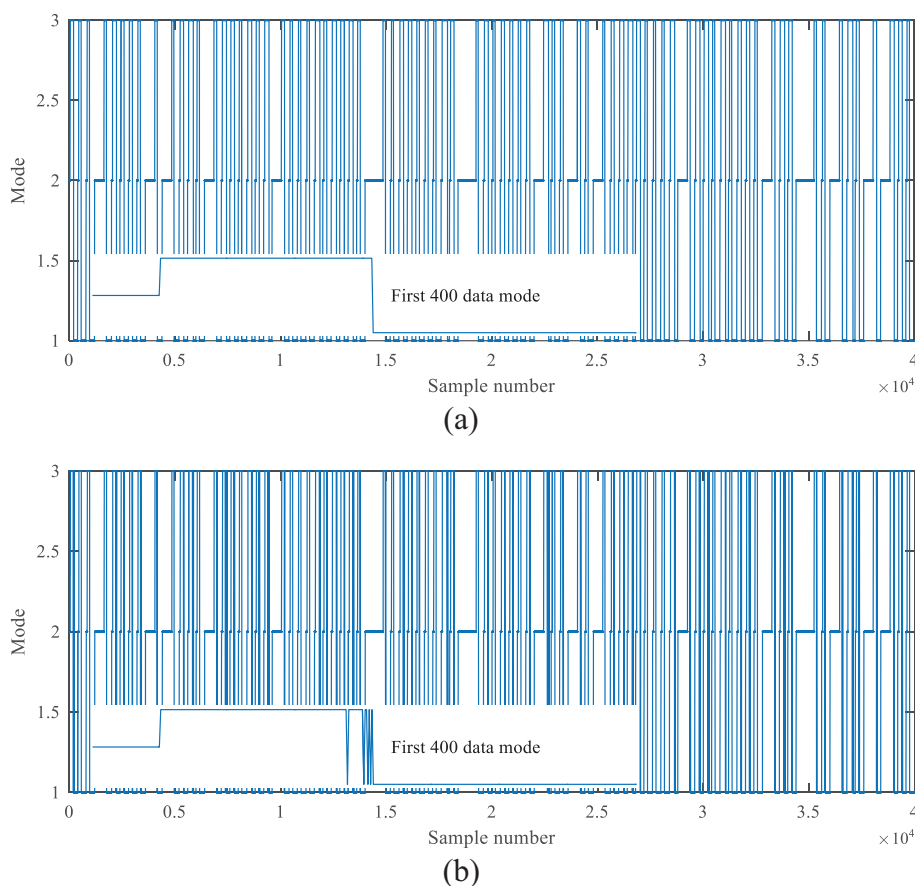


Fig. 6. Historical data mode identification results of BPFP. (a) TCGMM. (b) GMM.

begins to be produced as a metabolite. In the production of penicillin, the concentration of the producing bacteria must be maintained above a certain level, during which nutrients (e.g., nitrogen and sugar) must be continuously supplemented to ensure the final yield of penicillin. The production flowchart is shown in Fig. 5. The production cycle of penicillin is generally believed to comprise three parts: penicillium growth, penicillin synthesis, and penicillium autolysis. In addition to the most important variable, namely, penicillin concentration, the other operation and monitoring variables involved in the penicillin fermentation process include bacterial concentration, substrate concentration, fermentation tank temperature, feed substrate flow rate, and dissolved oxygen concentration.

In this work, the simulation package proposed in Literature [37] is used to generate experimental process data to test the proposed method. The 16 variables shown in Table 1 are considered for process monitoring to assess the availability of the TCGMM algorithm. Specific data are generated in accordance with Literature [39]. The response cycle of each batch is set to 400 h, and the initial simulation conditions and set values are shown in Table 2. The sampling frequency is once every hour, and each batch contains 400 samples. Under normal

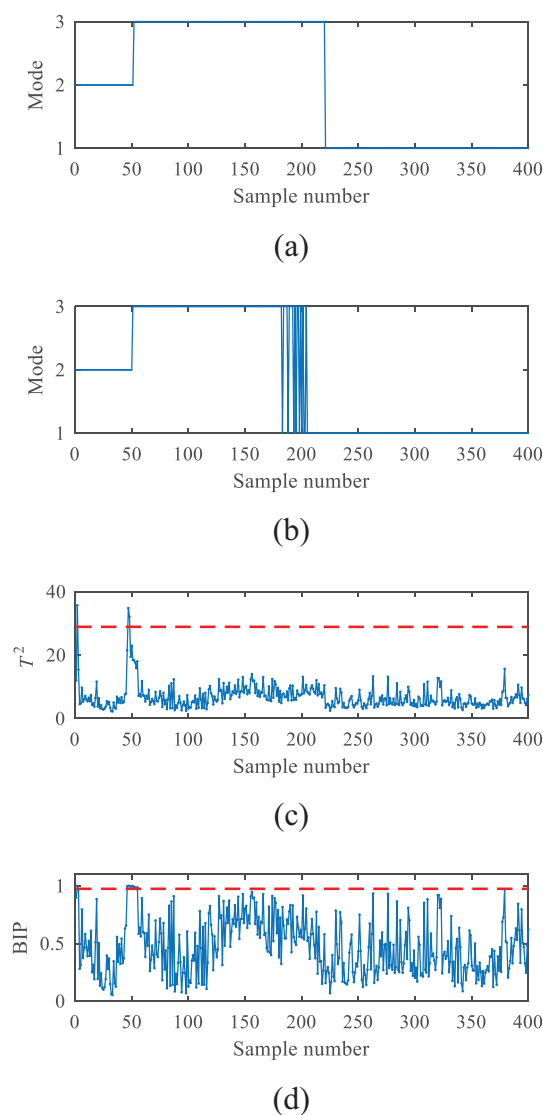


Fig. 7. Monitoring results for BPF normal batch. (a) Mode identification results based on TCGMM. (b) Mode identification results based on GMM. (c) T^2 control diagram based on TCGMM. (d) BIP control diagram based on TCGMM.

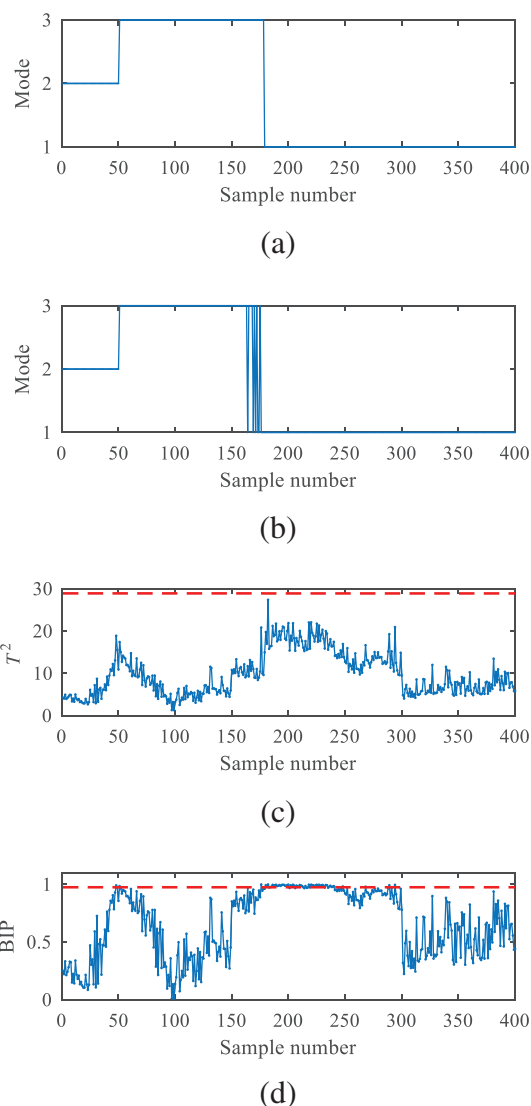


Fig. 8. Monitoring results for BPF Fault 1. (a) Mode identification results based on TCGMM. (b) Mode identification results based on GMM. (c) T^2 control diagram based on TCGMM. (d) BIP control diagram based on TCGMM.

operating conditions, 100 batches are simulated, and stored process simulation data are collected as training data. The TCGMM method is used to train the training data, and the mode identification effects of the training data are shown in Fig. 6. In view of the large number of training samples, this study selected the first batch of 400 sample points to identify the results, so as to show and compare the effects of the two methods more clearly. The identification results of other batches of training samples have the same characteristics. Figs. 6(b) shows the mode identification effects of historical data based on the GMM method for comparison. All training data are divided into three operation modes. These modes are described by a GMM composed of three Gaussian elements with various weights. Comparison of Fig. 6(a) and 6(b) reveals that the three modes identified by the TCGMM method are relatively stable, and no frequent mode transition occurs. However, the mode information identified by the GMM method is frequently transferred between the 200th and 220th samples. The fact that the operating conditions of a normal production process do not change so frequently further confirms the superiority of the TCGMM method in mode identification.

Four test situations are designed to evaluate the availability of the TCGMM monitoring method. One test situation involves a normal

batch, and the three others involve fault batches. The introduced faults are as follows. In the fault 1, a step change with amplitude 3 is introduced from the 150th to the 300th sampling points of the agitator power. In the fault 2, a ramp change of 0.05 ($t-150$), where t is the time instant, is introduced into the agitator power from the 150th to the 300th sampling points. In the fault 3, the ramp change of 0.05 ($t-150$) in the aeration rate occurs between the 150th and the 300th points.

Mode identification of the normal batch test data is conducted with the TCGMM and GMM methods mentioned in this research, and the results are shown in Fig. 7(a) and (b), respectively. Fig. 7(c) and (d) show the T^2 control diagram and the BIP control diagram, respectively, of the normal batch test data on the basis of the TCGMM multimodal monitoring method. The experimental results of each fault batch are presented in Figs. 8 to 10.

Fig. 7(a) shows a mode identification result based on the TCGMM method for normal batches. The divided modes do not have frequent mode transitions and thus maintain an excellent and stable effect. By contrast, the mode identification results based on the GMM method have frequent mode transitions, as shown in Fig. 7(b). This observation

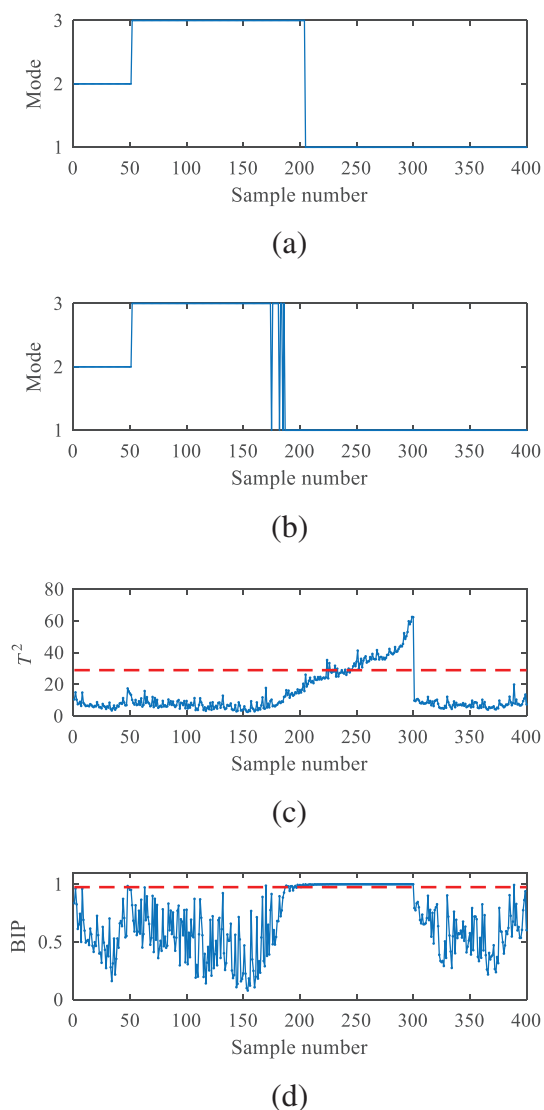


Fig. 9. Monitoring results for BPF Fault 2. (a) Mode identification results based on TCGMM. (b) Mode identification results based on GMM. (c) T^2 control diagram based on TCGMM. (d) BIP control diagram based on TCGMM.

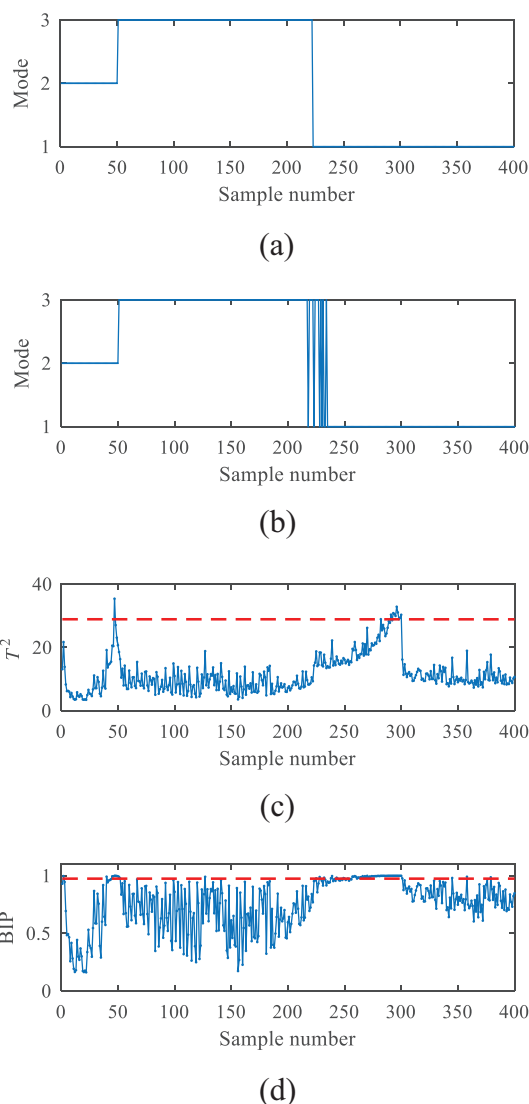


Fig. 10. Monitoring results for BPF Fault 3. (a) Mode identification results based on TCGMM. (b) Mode identification results based on GMM. (c) T^2 control diagram based on TCGMM. (d) BIP control diagram based on TCGMM.

indicates that the results identified by the new method after the introduction of the time series become increasingly stable and consistent with the timing characteristics of each stage in the actual process.

For fault batches 1, 2 and 3, as shown in Figs. 8(a)–(b), 9(a)–(b) and 10(a)–(b), the mode identification results inherit this advantage well. Moreover, the mode identification results based on the TCGMM method can achieve superior results by updating the prior knowledge introduced in the state transition matrix calculation. Overall, mode identification based on the TCGMM method is much better than that based on GMM. Except for differences in some data (such as the initial sample point of mode transfer), the mode identification result of the fault sample is nearly identical to that of the normal batch, which indicates that the mode identification of the fault batch is successful. In conclusion, this new method is successful in mode identification of multimodal processes.

As can be seen from the T^2 and BIP control diagrams, both indexes can monitor the normal batch samples well and maintain a low false alarm rate. The fault in fault batch 1 is not detected by the T^2 index monitoring method, while the monitoring effect of BIP index is relatively better. For fault batch 2, the BIP control diagram continues to maintain its good monitoring effect, but the time of detection of the fault is the 187th sampling

point, and the time of detection lags behind the time of introduction of the fault. This is because the initial amplitude of ramp fault is small and difficult to detect. The T^2 index monitoring effect in fault batch 2 is improved, but it is still not as good as that based on the BIP index. The T^2 monitoring effect in fault batch 3 is roughly similar to that in fault batch 1, and the monitoring effect of BIP control diagram is relatively better.

This analysis shows that the TCGMM monitoring method is successful in the mode identification of the penicillin fermentation process, and the identified mode is in line with the actual process. The fault can be identified well by BIP monitoring. However, the monitoring method using the T^2 index is not good due to the complex correlation between the variables of the penicillin fermentation process. The effectiveness of the TCGMM monitoring method using the BIP monitoring index is verified.

5. Conclusions

The TCGMM method is proposed for the problem of mode identification of multimodal processes with mode overlap and applied to multimodal process monitoring. This approach introduces time series information into the mode identification of historical and online data. The feasibility of the TCGMM monitoring method is analyzed using numerical and BPPF examples. The two groups of experiments reveal that the proposed method considerably improves the mode identification of multimodal processes and significantly increases the mode identification accuracy. For the penicillin fermentation process, mode identification can reflect and assist in evaluating the process' progress. Although the monitoring effect is not improved significantly as indicated by the experiments, the improvement in mode identification accuracy is expected to enhance the monitoring effect significantly for multimodal processes with more overlapping data points.

Acknowledgements

This work was supported in part by National Natural Science Foundation of China under Grants 61973119 and 61603138, in part by Shanghai Rising-Star Program under Grant 20QA1402600, in part by the Open Funding from Shandong Key Laboratory of Big-data Driven Safety Control Technology for Complex Systems under Grant SKDN202001, and in part by the Programme of Introducing Talents of Discipline to Universities (the 111 Project) under Grant B17017.

References

- [1] T. Kourti, J.F. MacGregor, Multivariate SPC methods for process and product monitoring, *J. Qual. Tech.* 28 (1996) 409–428.
- [2] Z. Ge, Process data analytics via probabilistic latent variable models: a tutorial review, *Ind. Eng. Chem. Res.* 57 (2018) 12646–12661.
- [3] S.J. Qin, Statistical process monitoring: basics and beyond, *J. Chemom.* 17 (2003) 480–502.
- [4] J. Feng, J. Wang, H. Zhang, 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.* 13 (2016) 122–133.
- [5] Q. Jiang, X. Yan, B. Huang, Review and perspectives of data-driven distributed monitoring for industrial plant-wide processes, *Ind. Eng. Chem. Res.* 58 (2019) 12899–12912.
- [6] D.H. Hwang, C. Han, Real-time monitoring for a process with multiple operating modes, *Control. Eng. Pract.* 7 (1999) 891–902.
- [7] S. Lane, E.B. Martin, R. Kooijmans, A.J. Morris, Performance monitoring of a multi-product semi-batch process, *J. Process Control* 11 (2001) 1–11.
- [8] H. Ma, H. Yi, H. Shi, A novel local neighborhood standardization strategy and its application in fault detection of multimode processes, *Chemom. Intell. Lab. Syst.* 118 (2012) 287–300.
- [9] C. Tong, A. Palazoglu, X. Yan, An adaptive multimode process monitoring strategy based on mode clustering and mode unfolding, *J. Process Control* 23 (2013) 1497–1507.
- [10] Y. Yang, Y. Ma, B. Song, H. Shi, An aligned mixture probabilistic principal component analysis for fault detection of multimode chemical processes, *Chin. J. Chem. Eng.* 23 (2015) 1357–1363.
- [11] S. Wold, Exponentially weighted moving principal components analysis and projections to latent structures, *Chemom. Intell. Lab. Syst.* 23 (1994) 149–161.
- [12] H. Dae, Y.H. Lee, G. Lee, C. Han, Robust recursive principal component analysis modeling for adaptive monitoring, *Ind. Eng. Chem. Res.* 45 (2006) 696–703.
- [13] S.W. Choi, E.B. Martin, A.J. Morris, I.B. Lee, Adaptive multivariate statistical process control for monitoring time-varying processes, *Ind. Eng. Chem. Res.* 45 (2006) 3108–3118.
- [14] Z. Ge, Z. Song, Online monitoring of nonlinear multiple mode processes based on adaptive local model approach, *Control. Eng. Pract.* 16 (2008) 1427–1437.
- [15] M. Kano, S. Hasebe, I. Hashimoto, H. Ohno, Evolution of multivariate statistical process control: application of independent component analysis and external analysis, *Comput. Chem. Eng.* 28 (2004) 1157–1166.
- [16] Y.H. Lee, H.D. Jin, C. Han, On-line process state classification for adaptive monitoring, *Ind. Eng. Chem. Res.* 45 (2006) 3095–3107.
- [17] K.C. Yoo, K. Villez, I.B. Lee, C. Rosén, P.A. Vanrolleghem, Multi-model statistical process monitoring and diagnosis of a sequencing batch reactor, *Biotechnol. Bioeng.* 96 (2007) 687–701.
- [18] J. Liu, Modeling a large-scale nonlinear system using adaptive Takagi–Sugeno fuzzy model on PCA subspace, *Ind. Eng. Chem. Res.* 46 (2007) 788–800.
- [19] S.W. Choi, J.H. Park, I.B. Lee, Process monitoring using a Gaussian mixture model via principal component analysis and discriminant analysis, *Comput. Chem. Eng.* 28 (2004) 1377–1387.
- [20] Y. Jie, S.J. Qin, Multimode process monitoring with Bayesian inference-based finite Gaussian mixture models, *AIChE J.* 54 (2008) 1811–1829.
- [21] X. Xie, H. Shi, Dynamic multimode process modeling and monitoring using adaptive Gaussian mixture models, *Ind. Eng. Chem. Res.* 51 (2012) 5497–5505.
- [22] J. Yu, A nonlinear kernel Gaussian mixture model based inferential monitoring approach for fault detection and diagnosis of chemical processes, *Chem. Eng. Sci.* 68 (2012) 506–519.
- [23] J. Yu, A new fault diagnosis method of multimode processes using Bayesian inference based Gaussian mixture contribution decomposition, *Eng. Appl. Artif. Intell.* 26 (2013) 456–466.
- [24] Q. Jiang, B. Huang, X. Yan, GMM and optimal principal components-based Bayesian method for multimode fault diagnosis, *Comput. Chem. Eng.* 84 (2016) 338–349.
- [25] S. Ren, Z. Song, M. Yang, J. Ren, A novel multimode process monitoring method integrating LCGMM with modified LFDA, *Chin. J. Chem. Eng.* 23 (2015) 1970–1980.
- [26] W. Sun, A. Palazoglu, J.A. Romagnoli, Detecting abnormal process trends by wavelet-domain hidden Markov models, *AIChE J.* 49 (2003) 140–150.
- [27] A. Bakhtazad, A. Palazoglu, J.A. Romagnoli, Detection and classification of abnormal process situations using multidimensional wavelet domain hidden Markov trees, *Comput. Chem. Eng.* 24 (2000) 769–775.
- [28] J. Chen, W.J. Chang, Applying wavelet-based hidden Markov tree to enhancing performance of process monitoring, *Chem. Eng. Sci.* 60 (2005) 5129–5143.
- [29] Z. Ge, Z. Song, Maximum-likelihood mixture factor analysis model and its application for process monitoring, *Chemom. Intell. Lab. Syst.* 102 (2010) 53–61.
- [30] Z. Ge, Z. Song, Mixture Bayesian regularization method of PPCA for multimode process monitoring, *AIChE J.* 56 (2010) 2838–2849.
- [31] M. Rashid, J. Yu, Hidden Markov model based adaptive independent component analysis approach for complex chemical process monitoring and fault detection, *Ind. Eng. Chem. Res.* 51 (2012) 5506–5514.
- [32] Q. Jiang, X. Yan, B. Huang, Neighborhood variational bayesian multivariate analysis for distributed process monitoring with missing data, *IEEE Trans. Control Syst. Technol.* 27 (6) (2019) 2330–2339.
- [33] Q. Jiang, X. Yan, B. Huang, Performance-driven distributed PCA process monitoring based on fault-relevant variable selection and Bayesian inference, *IEEE Trans. Ind. Electron.* 63 (2016) 377–386.
- [34] Q. Jiang, F. Gao, H. Yi, X. Yan, Multivariate statistical monitoring of key operation units of batch processes based on time-slice CCA, *IEEE Trans. Control Syst. Technol.* 27 (3) (2019) 1368–1375.
- [35] M.A.T. Figueiredo, A.K. Jain, Unsupervised learning of finite mixture models, *IEEE Trans. Pattern Anal. Mach. Intell.* 24 (2002) 381–396.
- [36] F. Qi, B. Huang, E.C. Tamayo, A Bayesian approach for control loop diagnosis with missing data, *AIChE J.* 56 (2010) 179–195.
- [37] G. Birol, C. Ündey, A. Çinar, A modular simulation package for fed-batch fermentation: penicillin production, *Comput. Chem. Eng.* 26 (2002) 1553–1565.
- [38] J.M. Lee, C.K. Yoo, I.B. Lee, Statistical process monitoring with independent component analysis, *J. Process Control* 14 (2004) 467–485.
- [39] X. Yan, Q. Jiang, Multimode process monitoring using variational Bayesian inference and canonical correlation analysis, *IEEE Trans. Autom. Sci. Eng.* 16 (2019) 1814–1824.