Frontiers of Information Technology & Electronic Engineering
www.zju.edu.cn/jzus; engineering.cae.cn; www.springerlink.com
ISSN 2095-9184 (print); ISSN 2095-9230 (online)
E-mail: jzus@zju.edu.cn

A novel multimode process monitoring method integrating LDRSKM with Bayesian inference*

Shi-jin REN†[1,2], Yin LIANG[2], Xiang-jun ZHAO[2], Mao-yun YANG[2]

([1]*National Laboratory of Industrial Control Technology, Zhejiang University, Hangzhou 310027, China*)

([2]*School of Computer Science & Technology, Jiangsu Normal University, Xuzhou 221116, China*)

†E-mail: sjren_phd@163.com

Received July 20, 2014; Revision accepted May 3, 2015; Crosschecked July 8, 2015

Abstract: A local discriminant regularized soft *k*-means (LDRSKM) method with Bayesian inference is proposed for multimode process monitoring. LDRSKM extends the regularized soft *k*-means algorithm by exploiting the local and non-local geometric information of the data and generalized linear discriminant analysis to provide a better and more meaningful data partition. LDRSKM can perform clustering and subspace selection simultaneously, enhancing the separability of data residing in different clusters. With the data partition obtained, kernel support vector data description (KSVDD) is used to establish the monitoring statistics and control limits. Two Bayesian inference based global fault detection indicators are then developed using the local monitoring results associated with principal and residual subspaces. Based on clustering analysis, Bayesian inference and manifold learning methods, the within and cross-mode correlations, and local geometric information can be exploited to enhance monitoring performances for nonlinear and non-Gaussian processes. The effectiveness and efficiency of the proposed method are evaluated using the Tennessee Eastman benchmark process.

Key words: Multimode process monitoring, Local discriminant regularized soft *k*-means clustering, Kernel support vector data description, Bayesian inference, Tennessee Eastman process

doi:10.1631/FITEE.1400263 **Document code:** A **CLC number:** TP277

1 Introduction

Data-driven chemical process monitoring methods have an important role in ensuring the safety of process operations and the quality of the product. Such methods have attracted the attention of researchers from many fields as they do not require extensive knowledge about the physical and chemical phenomena underlying the system under study. Industrial process data sets are commonly high dimensional and nonlinear. Their intrinsic dimensionality is often smaller than the dimensionality of the ambient space (Shen *et al.*, 2012; Deng and Tian, 2013). Therefore, removing inessential variables and extracting as much relevant information as possible from original variables are critical to process monitoring. Multivariate statistical process control (MSPC) has been intensively studied for fault detection and diagnosis. Some techniques, such as principal component analysis (PCA), partial least squares (PLS), and linear discriminant analysis (LDA), have been broadly applied in practice (Venkatasubramanian *et al.*, 2003; Kano *et al.*, 2007; Perez, 2011; Dong *et al.*, 2012). In real-world applications, chemical processes are conventionally characterized by their large scale, complexity, dynamics, and nonlinearity, posing a real challenge to process monitoring. To address these issues, various extensions to these MSPC methods, such as kernel PCA (KPCA), multi-scale PCA, batch dynamic KPCA (BDKPCA), and kernel entropy PCA, have been proposed for nonlinear process monitoring

* Project supported by the National Natural Science Foundation of China (No. 61272297)

ORCID: Shi-jin REN, http://orcid.org/0000-0002-8321-1879

(Zhang, 2009; Yu, 2012; Deng and Tian, 2013; Yang *et al.*, 2015). However, considering that process data do not follow a Gaussian distribution, independent component analysis (ICA) is employed to yield non-Gaussian features from process data, and a support vector data description (SVDD) or kernel density estimation approach is conducted to establish monitoring statistics and confidence limits. Conventional ICA cannot extract effective independent components (ICs) when more than one IC follows a Gaussian distribution. To address the issue, kernel time structure ICA and kernel ICA-PCA have been proposed for nonlinear dynamic process monitoring (Cai *et al.*, 2014). Most existing monitoring methods use all the measurement variables as a whole for modern industrial process monitoring. However, incorporating additional variables irrelevant to a specific fault degrades monitoring performance (Ghosh *et al.*, 2014). Identifying the optimal selected variables irrelevant to predefined and novel faults and reducing the missed detection rate and detection delay require further research.

Generally, industrial process systems exhibit multimode behaviors due to the multiple operating conditions incurred by the fluctuations of process raw materials, aging of the main components of the process, seasonal effects, and changing of the setting points. The monitoring performance of traditional MSPC methods may be degraded in these situations. To build an effective monitoring model for a multimode process, one feasible strategy is to develop a global nonlinear model such as kernel PCA, neural networks, or a Gaussian process latent vector model to capture the multimode process nonlinearities (Tan *et al.*, 2007; Zhang, 2009; Serradilla *et al.*, 2011; Deng and Tian, 2013; Deng *et al.*, 2013). Although these nonlinear modeling techniques are capable of multimode process monitoring, some critical model parameter values are required to be pre-specified through a trial-and-error method, which requires a large amount of computation and degrades the practicability of these techniques. Another feasible nonlinear modeling method called the linear approximation approach aims to partition multimode process data into multiple linear spaces (Ge *et al.*, 2010; Xu *et al.*, 2011; Xie and Shi, 2012; Feital *et al.*, 2013; Song *et al.*, 2014). As a result, a nonlinear model is approximated by combining multiple local linear models. In contrast to the global nonlinear modeling method, the linear approximation approach has more intuitive interpretations and works well in many applications (Ge *et al.*, 2010; 2013).

Clustering and mixture model approaches are typical of monitoring methods used for multimode processes (Teppola *et al.*, 1999; Xu *et al.*, 2011; Feital *et al.*, 2013; Ge *et al.*, 2013; Song *et al.*, 2014). The aim of data clustering is to group observations into multiple clusters that represent the underlying data patterns (Zang *et al.*, 2014). Due to a natural connection with multimode modeling, a clustering approach is employed to partition multimode process data into multiple linear spaces which are assumed to follow Gaussian distributions, and linear MSPC methods are then applied in each of these linear subspaces. Mixture model approaches, e.g., a mixture of probabilistic PCA (MPPCA), PCA, or a Gaussian mixture model (GMM), are capable of approximating any probability density function with multiple components. These methods and their variants (mixtures of factor analysis models, kernel GMM, infinite GMM) have been extensively explored to handle industrial processes with multiple operation regions (Ge and Song, 2010; Yu, 2012; Feital *et al.*, 2013; Song *et al.*, 2014). However, prior process knowledge is required to segment the entire process data into several groups corresponding to operating modes. Zhang *et al.* (2013) proposed a novel subspace extraction method with moderate computation loads for nonlinear multimode process monitoring. The subspaces consist of common and specific subspaces, and KPCA for process monitoring is conducted separately in each subspace. With a Bayesian inference based strategy, a global probabilistic index is constructed to enhance the reliability and comprehensiveness of fault detection by combining the different local monitoring results (Ge and Song, 2010). In general, most mixture model methods and clustering methods for multimode process monitoring work well under the assumption that each operation mode follows a Gaussian distribution. However, the local discriminant information from neighbors is omitted. Moreover, the process data do not exactly follow a Gaussian distribution in most applications, degrading the performance of these methods to some extent. Since the correlations between two adjacent modes and the potential characteristics of each mode are different, the separation of

common information, specific mode information, and transition information can enhance nonlinear multimode monitoring performance and understanding of multiple mode behaviors (Zhang and Li, 2014).

From the geometric point of view, manifolds are generalizations of curves and surfaces to an arbitrary number of dimensions. For convenience, manifolds are deemed the spaces that locally look like some Euclidean space, and the calculus can be conducted on these spaces. Recently, some manifold learning algorithms, such as locality preserving projection (LPP), locally linear embedding (LLE), and local tangent space alignment (LTSA), have been applied to reveal the underlying manifold structure of the process data for complex process monitoring (Zhang *et al.*, 2010; Xie and Shi, 2012; Deng and Tian, 2013; Miao *et al.*, 2015). Among them, LPP is a well-known local structure analysis method that is popular due to its simple extension to new data. Deng and Tian (2013) developed a sparse nonlinear LPP approach by using sample selection and kernel techniques. Song *et al.* (2014) defined a novel distance to eliminate the discrepancies of scales of within-mode and mode-to-mode variables, whilst the dynamic behaviors and neighborhood information of a single mode and correlations of mode-to mode are exploited. However, in practice, individual application of the manifold learning method to complex process monitoring may fail to detect process faults precisely and reliably. Integrating manifold learning approaches with MSPC methods is a feasible way to deal with this problem. Xie and Shi (2012) combined fuzzy c-means (FCM) with LPP for multimode process monitoring. Zhang *et al.* (2010) developed an SVDD and LTSA-based fault monitoring method for non-Gaussian processes, in which the monitoring statistics are constructed by SVDD in the low-dimensional space induced from LTSA. To exploit global and local geometric structure information hidden in process data, Deng *et al.* (2013) combined kernel PCA with LPP to extract linear or nonlinear features for complex process monitoring. Considering that maximum variance unfolding (MVU) can reveal global structure of the data via non-local variance information and that neighborhood preserving embedding can refine the local variance information, Miao *et al.* (2015) defined a dual-objective optimization problem that minimizes local scatter and maximizes non-local scatter simultaneously. Zhang *et al.* (2011) proposed a novel local-global structure analysis approach for process monitoring in the framework of PCA and LPP.

Noting that process variables often exhibit randomness due to the raw material, measurement noise, and changes of product specifications, data samples can be generated from multiple subspaces, and the probabilities of data samples forming the subspaces may be unknown. For instance, when one operating condition moves to another, the transition process may last a few hours, during which the correlations among process variables are changing. The characteristics of the process variables are also associated with multiple subspaces during the moving process. Therefore, it is necessary to simultaneously partition the data into multiple subspaces and find the low-dimensional underlying subspace to fit with instances associated with one cluster. Traditional clustering-based process monitoring approaches conduct process data clustering and dimensionality reduction separately (Xie and Shi, 2012; Tong *et al.*, 2013). However, the clustering performance of process data greatly relies on the performance of the dimensionality reduction approach, and vice versa.

In this study, we propose a local discriminant regularized soft k-means (LDRSKM) approach for multimode process monitoring. The merits of the LDRSKM algorithm can be summarized as follows: (1) A regularized soft k-means with locality preservation (LPRSKM) is proposed to perform clustering and subspace selection simultaneously, uncovering the structure of real-world data as faithfully as possible. Each cluster is assumed to be associated with a single operation mode. (2) A generalization linear discriminant analysis (GELDA) algorithm is introduced to select the most discriminant subspace based on the membership degrees by exploiting the local and non-local geometry of the data set. GELDA can not only separate clusters with maximum margin but also reveal the multimodalities within clusters, significantly improving the clustering accuracy. The implementation procedure of LDRSKM consists of two stages. First, LPRSKM is carried out to find the approximate cluster assignment in the low-dimensional optimal subspaces achieved by GELDA. Then GELDA is performed on the cluster assignments to obtain the optimal projection matrix, and original data points are projected onto the low-

dimensional discriminant subspaces. The two stages are carried out iteratively until convergence. After the optimal subspaces are obtained, the posterior probabilities of a new mapped data point within each subspace are calculated. Two global Bayesian inference probability indicators (BIPIs) for fault detection are then constructed in principal and residual subspaces by integrating posterior probabilities with priors over all the subspaces. Thus, all local monitoring results from individual clusters are combined to establish the global monitoring results, effectively enhancing the reliability and accuracy of the monitoring results of the proposed approach.

The main contributions of this study are as follows: (1) We propose a general discriminative clustering framework for process monitoring. The novelty lies in that it performs clustering and subspace selection simultaneously, which can reveal the multimodality within the multiple clusters and enhance the clustering accuracy. (2) A locality preserving technique in the probability space is introduced to regularized soft *k*-means clustering (ResKmeans) to ensure the smoothness of the membership function, and also to gain some new insights into the description of complex process data. (3) The proposed approach can be easily extended to a supervised or semi-supervised learning for complex process monitoring. (4) Global monitoring statistics are constructed by integrating the monitoring results of individual clusters.

2 A regularized soft *k*-means with locality preservation algorithm

In this section, we introduce our LDRSKM algorithm. The approach exploits the intrinsic geometry of the probability distribution and locality discriminative information, and the cluster assignments are found via membership degrees. First, the ResKmeans algorithm is briefly described.

2.1 A regularized soft *k*-means algorithm

Given a data set $X=\{\boldsymbol{x}_i\}_{i=1}^N$ with $\boldsymbol{x}_i\in\mathbb{R}^D$, for simplicity, the instances are assumed to have zero mean. Denote the data matrix by $\boldsymbol{X}=[\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_N]$, the ith column of which is represented by $\boldsymbol{x}_i$ (i=1, 2, …, N). The aim of hard k-means clustering is to obtain a cluster indicator matrix $\boldsymbol{H}\in\mathbb{R}^{N\times C}$ with elements H_{ic}=0 or 1. H_{ic}=1 if $\boldsymbol{x}_i$ is within cluster c, and H_{ic}=0 otherwise, where $c\in\{1, 2, \ldots, C\}$ is a cluster label and C is the number of the clusters. Contrary to a hard k-means clustering algorithm which assumes that each instance is within only a specific cluster, a soft k-means clustering algorithm assigns a sample to one cluster based on its membership degree, whose optimization problem is expressed as follows:

$$\begin{aligned}&\min J(\boldsymbol{U})=\sum_{c=1}^{C}\sum_{i=1}^{N}u_{ic}\,\|\,\boldsymbol{x}_i-\boldsymbol{m}_c^r\,\|^2\\&\text{s.t. }\sum_{c=1}^{C}u_{ic}=1\text{ and }u_{ic}\geq 0,\end{aligned}\tag{1}$$

where u_{ic} is the membership degree of $\boldsymbol{x}_i$ within cluster c with centroid $\boldsymbol{m}_c^r$, and $\boldsymbol{U}\in\mathbb{R}^{N\times C}$ with its elements $u_{ic}\in[0, 1]$ is the membership degree matrix. Since the membership degree u_{ic} can be deemed the probability of instance i within cluster c, the entropy of membership degree, as a regularization term, is incorporated into the objective function of the above optimization problem to avoid overfitting (Miyamoto and Mukaidono, 1997). As a consequence, a modified optimization problem is expressed as (Yin *et al.*, 2013)

$$\begin{aligned}&\min J'(\boldsymbol{U})=\sum_{c=1}^{C}\sum_{i=1}^{N}u_{ic}\,\|\,\boldsymbol{x}_i-\boldsymbol{m}_c^r\,\|^2+\eta\sum_{c=1}^{C}\sum_{i=1}^{N}u_{ic}\log u_{ic}\\&\text{s.t. }\sum_{c=1}^{C}u_{ic}=1\text{ and }u_{ic}\geq 0.\end{aligned}\tag{2}$$

The second term in the objective function is the entropy, which is used to avoid overfitting. Noting that the above optimization problem is non-convex, it is usually solved in an iterative way. During the optimization process, the optimization objective function (2) can be rewritten as

$$\begin{aligned}J'(\boldsymbol{U}\,|\,\boldsymbol{U}^{(t-1)})=&\sum_{c=1}^{C}\sum_{i=1}^{N}u_{ic}\,\|\,\boldsymbol{x}_i-\boldsymbol{m}_c^r\,|_{\boldsymbol{U}^{(t-1)}}\|^2\\&+\eta\sum_{c=1}^{C}\sum_{i=1}^{N}u_{ic}\log u_{ic}\\&\text{s.t. }\sum_{c=1}^{C}u_{ic}=1\text{ and }u_{ic}\geq 0,\end{aligned}\tag{3}$$

where $\boldsymbol{U}^{(t-1)}$ is the membership degree matrix at iteration $t-1$. Details about ResKmeans can be found in Yin *et al.* (2013).

2.2 LPRSKM

In view of the manifold assumption and probabilistic distribution, data points $\boldsymbol{x}\in X$ are generated from a certain probability distribution P which varies smoothly along the geodesics in the intrinsic geometry of marginal probability in most real-world applications (He *et al.*, 2011). Thus, if the two samples $\boldsymbol{x}_1$, $\boldsymbol{x}_2$ are close in the intrinsic geometry, their corresponding membership degree function values $f_c(\boldsymbol{x}_1)$ and $f_c(\boldsymbol{x}_2)$ should still be close under the assumption of manifold learning. Considering that the clustering performance can be effectively enhanced by exploiting the geometric information hidden in the data, in this study LPRSKM is developed by extending an established framework for learning soft k-means.

Suppose that there are C clusters in X and let $f_c(\boldsymbol{x})=u_{\boldsymbol{x},c}$ be the membership degree function. $\|f_c\|_M$ is used to measure the smoothness of the membership degree function f_c. A sufficiently smooth membership degree function can be achieved through the minimization of $\|f_c\|_M$. Similar to Liu *et al.* (2010), we establish an adjacent graph to approximate $\|f_c\|_M$. Due to the close connection between membership degree and conditional distribution, $f_c(\boldsymbol{x})$ can be deemed to be $p(c|\boldsymbol{x})$. Manifold regularization $\|f_c\|_M$ is approximated in the probabilistic space by the following formula:

$$\begin{aligned}\| f_c \|_M = \frac{1}{2}\sum_{i,j=1}^{N}\Big[&\mathrm{KL}\Big(p(C\,|\,\boldsymbol{x}_i)\,\|\,p(C\,|\,\boldsymbol{x}_j)\Big)\\ &+\mathrm{KL}\Big(p(C\,|\,\boldsymbol{x}_j)\,\|\,p(C\,|\,\boldsymbol{x}_i)\Big)\Big]\cdot S_{ij},\end{aligned} \tag{4}$$

where $\boldsymbol{S}=[S_{ij}]_{i,j=1}^{N,N}$ is computed by a heat kernel function defined on k_{NN} nearest neighbors, i.e.,

$$S_{ij}=\begin{cases}\exp(-\|\boldsymbol{x}_i-\boldsymbol{x}_j\|^2/t^0),\\ \qquad \boldsymbol{x}_i\in \mathrm{NN}(\boldsymbol{x}_j)\text{ or }\boldsymbol{x}_j\in \mathrm{NN}(\boldsymbol{x}_i),\\ 0,\qquad \text{otherwise},\end{cases} \tag{5}$$

where $\mathrm{NN}(\boldsymbol{x}_j)$ represents the neighborhood set of data points $\boldsymbol{x}_j$, and t^0 is a parameter of the heat kernel function which has an intrinsic connection to the Laplacian Beltrami operator on differential functions on a manifold (He *et al.*, 2011). An alternative option to heat kernel is defined by $S_{ij}=\exp[-\|\boldsymbol{x}_i-\boldsymbol{x}_j\|^2/(\sigma_i\sigma_j)]$, where $\sigma_i=\|\boldsymbol{x}_i-\boldsymbol{x}_i^{(k_{NN})}\|$ represents the local scaling of data samples around $\boldsymbol{x}_i$ with $\boldsymbol{x}_i^{(k_{NN})}$, and $\boldsymbol{x}_i^{(k_{NN})}$ (i=1, 2, …, k_{NN}) are the k_{NN} nearest neighbors of $\boldsymbol{x}_i$. This strategy takes the density of the data samples in different regions into account, and the heat kernel parameter is not required to be pre-specified.

The divergence in Eq. (4) is defined by

$$\begin{aligned}&\mathrm{KL}\Big(p(C\,|\,\boldsymbol{x}_i)\,\|\,p(C\,|\,\boldsymbol{x}_j)\Big)\\ &=\sum_{k=1}^{C}p(k\,|\,\boldsymbol{x}_i)\log\Big(p(k\,|\,\boldsymbol{x}_i)/\,p(k\,|\,\boldsymbol{x}_j)\Big).\end{aligned}$$

Although $\mathrm{KL}(p(C|\boldsymbol{x}_i)\|p(C|\boldsymbol{x}_j))$ is asymmetric,

$$\begin{aligned}\sum_{i,j=1}^{N}\Big[&\mathrm{KL}\Big(p(C\,|\,\boldsymbol{x}_i)\,\|\,p(C\,|\,\boldsymbol{x}_j)\Big)\\ &+\mathrm{KL}\Big(p(C\,|\,\boldsymbol{x}_j)\,\|\,p(C\,|\,\boldsymbol{x}_i)\Big)\Big]\cdot S_{ij}\end{aligned}$$

is symmetric due to the symmetry of $\boldsymbol{S}$. Considering that u_{ic} is equivalent to $p(c|\boldsymbol{x}_i)$ in the sense of classification, $\mathrm{KL}(p(C|\boldsymbol{x}_i)\|p(C|\boldsymbol{x}_j))$ is calculated via $\mathrm{KL}(u_{ic}\|u_{jc})$. Inspired by manifold learning, the optimization problem of LPRSKM proposed in this study is defined as follows:

$$\begin{aligned}\arg\min_{u_{ic}}&\sum_{c=1}^{C}\sum_{i=1}^{N}u_{ic}\,\|\,\boldsymbol{x}_i-\boldsymbol{m}_c^r\,\|^2+\eta\sum_{c=1}^{C}\sum_{i=1}^{N}u_{ic}\log u_{ic}\\ &+\gamma\sum_{i,j}^{N}\sum_{c=1}^{C}\left(u_{ic}\log\frac{u_{ic}}{u_{jc}}+u_{jc}\log\frac{u_{jc}}{u_{ic}}\right)\cdot S_{ij}\end{aligned} \tag{6}$$

$$\text{s.t. }\sum_{c=1}^{C}u_{ic}=1\ \text{ and }\ u_{ic}\ge 0,\quad i=1,2,\cdots,N, \tag{7}$$

where $\gamma>0$ is a regularization coefficient. The second term is used to avoid overfitting in the clustering process. The third term is introduced to describe the underlying geometry of the data. Because the computation of $\log u_{jc}$ is often infeasible when u_{jc} is nearly zero, u_{ic} and u_{jc} in $\log(u_{ic}/u_{jc})$ are approximated by posterior probabilities $p(c|\boldsymbol{x}_i)$ and $p(c|\boldsymbol{x}_j)$, respectively. Following the Bayesian principle, the posterior probability of data sample $\boldsymbol{x}$ is

$$p(c\,|\,\boldsymbol{x})=\frac{p(\boldsymbol{x}\,|\,c)\,p(c)}{p(\boldsymbol{x})}. \tag{8}$$

Let $u'_{ic}=p(c|\boldsymbol{x}_i)$ and $u'_{jc}=p(c|\boldsymbol{x}_j)$. The approximation of $\log(u_{ic}/u_{jc})$ is given by

$$\log\frac{u'_{ic}}{u'_{jc}}=\log\frac{p(\boldsymbol{x}_i\,|\,c)}{p(\boldsymbol{x}_j\,|\,c)}+\log\frac{p(\boldsymbol{x}_j)}{p(\boldsymbol{x}_i)}. \tag{9}$$

Considering constraint $\sum_c u_{ic}=1$, we have

$$\begin{aligned}&\sum_c\left(u_{ic}\log\frac{u_{ic}}{u_{jc}}+u_{jc}\log\frac{u_{jc}}{u_{ic}}\right)\\&=\sum_c(u_{ic}-u_{jc})\log\frac{p(\boldsymbol{x}_i\,|\,c)}{p(\boldsymbol{x}_j\,|\,c)}.\end{aligned} \tag{10}$$

Since the conditional probability $p(c|\boldsymbol{x}_i)$ is unknown, it is infeasible to compute Eq. (10). For simplicity, $p(\boldsymbol{x}|c)$ (c=1, 2, …, C) is assumed to be a Gaussian function in this paper. Therefore, Eq. (10) is calculated by

$$\begin{aligned}\log\frac{p(\boldsymbol{x}_i\,|\,c)}{p(\boldsymbol{x}_j\,|\,c)}=&\frac{1}{2}(\boldsymbol{x}_i-\boldsymbol{m}_c^r)^{\mathrm{T}}\boldsymbol{\Sigma}_c^{-1}(\boldsymbol{x}_i-\boldsymbol{m}_c^r)\\&-\frac{1}{2}(\boldsymbol{x}_j-\boldsymbol{m}_c^r)^{\mathrm{T}}\boldsymbol{\Sigma}_c^{-1}(\boldsymbol{x}_j-\boldsymbol{m}_c^r),\end{aligned} \tag{11}$$

where centroid $\boldsymbol{m}_c^r$ and covariance matrix $\boldsymbol{\Sigma}_c$ of cluster c are calculated by

$$\boldsymbol{m}_c^r=\sum_{i=1}^N u_{ic}\boldsymbol{x}_i\Big/\sum_{i=1}^N u_{ic}, \tag{12}$$

$$\boldsymbol{\Sigma}_c=\sum_{i=1}^N u_{ic}(\boldsymbol{x}_i-\boldsymbol{m}_c^r)(\boldsymbol{x}_i-\boldsymbol{m}_c^r)^{\mathrm{T}}\Big/\sum_{i=1}^N u_{ic}. \tag{13}$$

Although the manifold regularization approximated under the assumption that the data follow a Gaussian distribution can degrade the performance of the proposed clustering algorithm, the neighborhood information in the probabilistic space and original data space are both taken into account, and the geometric structure of the non-Gaussian data can be effectively revealed. Details can be found in Liu *et al.* (2010).

It is hard to obtain the optimal u_{ic} satisfying constraint (7) through minimization of the above optimization problem as given by expressions (6) and (7). Just like ResKmeans, the optimization problem as given by expressions (6) and (7) is solved in an iterative way. Using Eq. (10), the Lagrangian function associated with the optimization problem in Eqs. (6) and (7) is formulated by

$$\begin{aligned}L(u_{ic},\boldsymbol{m}_c^r,\lambda)=&\sum_{c=1}^C\sum_{i=1}^N u_{ic}\,\|\,\boldsymbol{x}_i-\boldsymbol{m}_c^r\,\|^2+\eta\sum_{c=1}^C\sum_{i=1}^N u_{ic}\log u_{ic}\\&-\sum_{i=1}^N\lambda_i\left(\sum_{c=1}^C u_{ic}-1\right)+\frac{\gamma}{2}\sum_{i,j}^N\sum_{c=1}^C l_{ij}^c(u_{ic}-u_{jc})S_{ij},\end{aligned} \tag{14}$$

where $l_{ij}^c=\log[p(\boldsymbol{x}_i\,|\,c)/p(\boldsymbol{x}_j\,|\,c)]$ and λ_i are Lagrange multipliers.

Taking the derivatives of $L(u_{ic},\lambda_i)$ with respect to u_{ic}, $\boldsymbol{m}_c^r$, and λ_i, and setting them to zeros, we have

$$\begin{aligned}&\frac{\delta L(u_{ic},\boldsymbol{m}_c^r,\lambda_i)}{\delta u_{ic}}=0\Rightarrow\\&\lambda_i=\|\,\boldsymbol{x}_i-\boldsymbol{m}_c^r\,\|^2+\eta\log u_{ic}+\eta+\gamma\sum_j^N l_{ij}^c S_{ij},\\&\frac{\delta L(u_{ic},\boldsymbol{m}_c^r,\lambda_i)}{\delta\lambda_i}=\sum_{c=1}^C u_{ic}-1=0,\\&\frac{\delta L(u_{ic},\boldsymbol{m}_c^r,\lambda_i)}{\delta\boldsymbol{m}_c^r}=0\Rightarrow\boldsymbol{m}_c^r=\sum_{i=1}^N u_{ic}\boldsymbol{x}_i\Big/\sum_{i=1}^N u_{ic}.\end{aligned}$$

After some processing for the above equations, the membership degree u_{ic} is updated by

$$u_{ic}=\frac{\exp\left(-\gamma\sum_j^N l_{ij}^c S_{ij}\Big/\eta\right)\exp(-\|\,\boldsymbol{x}_i-\boldsymbol{m}_c^r\,\|^2/\eta)}{\sum_{c=1}^C\exp\left(-\gamma\sum_j^N l_{ij}^c S_{ij}\Big/\eta\right)\exp(-\|\,\boldsymbol{x}_i-\boldsymbol{m}_c^r\,\|^2/\eta)}. \tag{15}$$

With the entropy and manifold regularization, the cluster assignments achieved by LPRSKM not only have a probabilistic interpretation, but also effectively reflect the inherent fuzzy nature and geometric structure of the data and enhance the discrepancy of data within different classes or clusters.

The proof of the convergence of LPRSKM is similar to that described by Miyamoto and Mukaidono (1997) and Jing *et al.* (2007), and the procedure is thus omitted.

3 Generalized linear discriminant analysis

Based on the membership degree matrix $\boldsymbol{U}$ and centroids $\boldsymbol{m}_c^r$ obtained through LPRSKM, the

optimal low-dimensional subspaces can be obtained via GELDA.

Just like traditional LDA, GELDA obtains low-dimensional subspaces using the membership degrees from soft clustering in a supervised situation. The between-cluster matrix $\boldsymbol{S}_\mathrm{b}$, within-cluster matrix $\boldsymbol{S}_\mathrm{w}$, and total scatter matrix $\boldsymbol{S}_\mathrm{t}$ are defined by

$$\boldsymbol{S}_\mathrm{b}=\frac{1}{N}\sum_{c=1}^{C}n_c(\boldsymbol{m}_c-\boldsymbol{m})(\boldsymbol{m}_c-\boldsymbol{m})^\mathrm{T}, \tag{16}$$

$$\boldsymbol{S}_\mathrm{w}=\sum_{c=1}^{C}\sum_{i=1}^{N}u_{ic}(\boldsymbol{x}_i-\boldsymbol{m}_c)(\boldsymbol{x}_i-\boldsymbol{m}_c)^\mathrm{T}, \tag{17}$$

$$\boldsymbol{S}_\mathrm{t}=\sum_{i=1}^{N}(\boldsymbol{x}_i-\boldsymbol{m})(\boldsymbol{x}_i-\boldsymbol{m})^\mathrm{T}, \tag{18}$$

where $n_c=\sum_{i=1}^{N}u_{ic}$ is the soft number of instances involved in cluster c and $\sum_{c=1}^{C}n_c=\sum_{c=1}^{C}\sum_{i=1}^{N}u_{ic}=N$, $\boldsymbol{m}_c=\sum_{i=1}^{N}u_{ic}\boldsymbol{x}_i\Big/\sum_{i=1}^{N}u_{ic}$ is the centroid of cluster c, and $\boldsymbol{m}=\frac{1}{N}\sum_{c=1}^{C}\sum_{i=1}^{N}u_{ic}\boldsymbol{x}_i=\frac{1}{N}\sum_{i=1}^{N}\boldsymbol{x}_i$ is the global centroid of the data.

When the training data are centered, $\boldsymbol{m}$ is equal to zero. Since GELDA is the generalization of LDA, they share some identical properties. Their global scatter matrices are identical and the scatter matrices of GELDA also satisfy the relationships rank($\boldsymbol{S}_\mathrm{t}$)=rank($\boldsymbol{S}_\mathrm{b}$)+rank($\boldsymbol{S}_\mathrm{w}$) and $\boldsymbol{S}_\mathrm{t}=\boldsymbol{S}_\mathrm{b}+\boldsymbol{S}_\mathrm{w}$. Now, we attempt to find the discriminative projection matrix $\boldsymbol{P}=[\boldsymbol{p}_1,\boldsymbol{p}_2,\ldots,\boldsymbol{p}_D]\in\mathbb{R}^{d\times D}$ ($D<d$) via GELDA. One approach seeks to find an optimal projection matrix by maximizing the following objective function:

$$\boldsymbol{P}^*=\arg\max_{\boldsymbol{P}}\ \frac{\mathrm{tr}(\boldsymbol{P}^\mathrm{T}\boldsymbol{S}_\mathrm{b}\boldsymbol{P})}{\mathrm{tr}\ (\boldsymbol{P}^\mathrm{T}\boldsymbol{S}_\mathrm{w}\boldsymbol{P})}, \tag{19}$$

where tr($\cdot$) represents the trace operator of a matrix.

Another approach is based on the idea that not all projection vectors obtained by Eq. (19) are helpful for classification (Howland *et al.*, 2006). Therefore, it is necessary to obtain projection vectors forming an efficient discriminative subspace. The projection matrix is obtained by

$$\boldsymbol{\Phi}=\{\varphi\,|\,\varphi^\mathrm{T}\boldsymbol{S}_\mathrm{b}\varphi-\varphi^\mathrm{T}\boldsymbol{S}_\mathrm{w}\varphi>0\}. \tag{20}$$

Projection matrix $\boldsymbol{\Phi}$ embodies the information within the null subspace of $\boldsymbol{S}_\mathrm{w}$ and thus contains the most efficient discriminative information. Based on the above address, we choose the most discriminative vectors sequentially according to Eq. (20).

To perform clustering and subspace selection, with fixed $\boldsymbol{U}$, $\boldsymbol{P}$ is computed by applying GELDA; with fixed $\boldsymbol{P}$, $\boldsymbol{U}$ is then computed by LPRSKM in the P-transformed space. In summary, the optimization is iteratively performed by fixing one of two components ($\boldsymbol{P}$ and $\boldsymbol{U}$) and optimizing the other component. The implementation process of LDRSKM is described as follows:

Input: Dataset X, cluster number C, model parameters γ and η, number of iterations t=1, stopping threshold δ.

Output: Membership degree matrix $\boldsymbol{U}$ and projection matrix $\boldsymbol{P}$.

Step 1: Carry out PCA on X to obtain the initial projection matrix $\boldsymbol{P}^{(t)}$.

Step 2: Set the reduced dimensionality to $d=C-1$, and initialize the cluster centroids $\boldsymbol{m}^{(t)}$ via the conventional FCM algorithm.

Repeat

Step 3: Perform LPRSKM by calculating matrix $\boldsymbol{U}^{(t)}$ in terms of Eq. (15) in the $\boldsymbol{P}^{(t)}$-transformed subspace.

Step 4: Construct adjacent graph matrix $\boldsymbol{S}$ and smoothing regularization term R_c on $\boldsymbol{U}^{(t)}$ and X. The membership degrees are then calculated by

$$u_{ic}^{(t+1)}=\frac{\exp\left(-\gamma\sum_{j}^{N}l_{ij}^{c}S_{ij}\Big/\eta\right)\exp(-\|\boldsymbol{x}_i-\boldsymbol{m}_c^{(t)}\|^2/\eta)}{\sum_{c=1}^{C}\exp\left(-\gamma\sum_{j}^{N}l_{ij}^{c}S_{ij}\Big/\eta\right)\exp(-\|\boldsymbol{x}_i-\boldsymbol{m}_c^{(t)}\|^2/\eta)},$$
$$c=1,2,\ldots,C,\ \ i=1,2,\ldots,N.$$

Step 5: Perform GELDA on the original data set with the obtained $\boldsymbol{U}^{(t+1)}$. Calculate $\boldsymbol{S}_\mathrm{t}$, $\boldsymbol{S}_\mathrm{b}$, and $\boldsymbol{S}_\mathrm{w}$ with $\boldsymbol{U}^{(t+1)}$ according to Eqs. (16)–(18) and compute projection matrix $\boldsymbol{P}^{(t+1)}$ according to Eq. (20).

Step 6: Evaluate the following objective function:

$$L(\boldsymbol{U}^{(t+1)})=\sum_{c=1}^{C}\sum_{i=1}^{N}u_{ic}\|\boldsymbol{x}_i-\boldsymbol{m}_c^r\|^2+\eta\sum_{c=1}^{C}\sum_{i=1}^{N}u_{ic}\log u_{ic}$$
$$+\gamma\sum_{i,j}^{N}\sum_{c=1}^{C}\left(u_{ic}\log\frac{u_{ic}}{u_{jc}}+u_{jc}\log\frac{u_{jc}}{u_{ic}}\right)S_{ij}.$$

Step 7: $i \leftarrow i+1$.
until $L(\boldsymbol{U}^{(t-1)})-L(\boldsymbol{U}^{(t)})\le\delta$.
return $\boldsymbol{P}^{(t)}$, $\boldsymbol{U}^{(t)}$.

The LDRSKM consists of two stages, GELDA and LPRSKM, and has three components (GELDA+ locality preserving terms+regularized soft k-means clustering). Therefore, the time complexity of LDRSKM is $O(DNt)$ for regularized soft k-means clustering and $O(D^2Nt)$ for GELDA.

As in Xie and Shi (2012), the value of the heat kernel parameter is chosen as $t^0=2^m\sigma_0$, where σ_0 represents the standard variance of the training data and $m\in\{-8, -7, \ldots, 7, 8\}$ is determined through cross-validation. The number of nearest neighbors denoted by N_n is determined according to Perez (2011) and Xie and Shi (2012).

4 Construction of monitoring statistics

For process monitoring methods, effective monitoring statistics play a very important role and should be carefully developed. Considering that data points forming a specific cluster often obey non-Gaussian distribution and SVDD is an effective approach to model non-Gaussian data (Xie *et al.*, 2009), SVDDs are extensively applied to establish monitoring statistics and the corresponding control limits. To improve the robustness and reliability of monitoring performance, global monitoring statistics are established by combining monitoring results of all the principal and residual subspaces in a probabilistic manner by the Bayesian inference method. Next, kernel SVDD (KSVDD) and Bayesian inference based monitoring statistics are briefly introduced for fault detection.

4.1 Kernel support vector data description

Considering a training data set $\{\boldsymbol{x}_i\}_{i=1}^N$ and using a nonlinear transformation ϕ to map $\boldsymbol{x}_i$ to a high-dimensional feature space, the minimum volume hypersphere characterized with its center $\boldsymbol{a}$ and radius R is constructed to capture most data points by solving the following optimization problem (Lee and Lee, 2007):

$$\min_{R,\boldsymbol{a},\xi} R^2 + C'\sum_{i=1}^N \xi_i$$
$$\text{s.t.}\ \|\phi(\boldsymbol{x}_i)-\boldsymbol{a}\|^2 \le R^2+\xi_i \text{ and } \xi_i\ge 0, i=1, 2, \cdots, N, \tag{21}$$

where C' is a trade-off between the volume of the hypersphere and the errors, and ξ_i are slack variables which allow for the soft boundaries. The solution to the primal optimization problem (21) can be achieved by solving the following dual optimization problem:

$$\arg\max_{\alpha_j} W=\sum_{j=1}^N \alpha_j k(\boldsymbol{x}_j,\boldsymbol{x}_j)-\sum_{i,j=1}^N \alpha_j\alpha_i k(\boldsymbol{x}_i,\boldsymbol{x}_j)$$
$$\text{s.t.}\ 0\le\alpha_j\le C', j=1, 2, \cdots, N \text{ and } \sum_{j=1}^N \alpha_j=1, \tag{22}$$

where $k(\boldsymbol{x}_i, \boldsymbol{x}_j)=\phi(\boldsymbol{x}_i)\cdot\phi(\boldsymbol{x}_j)$ is the Mercer kernel function and often selected as the Gaussian kernel function $k(\boldsymbol{x}_i, \boldsymbol{x}_j)=\exp(-\|\boldsymbol{x}_i-\boldsymbol{x}_j\|^2/(2\sigma^2))$. When $k(\boldsymbol{x}_i,\boldsymbol{x}_j)=\boldsymbol{x}_i^{\mathrm{T}}\boldsymbol{x}_j$, KSVDD degrades to a linear SVDD.

Through solving the dual optimization problem, centroid $\boldsymbol{a}$ and radius R of the hypersphere are calculated by

$$\boldsymbol{a}=\sum_{i=1}^{N_\mathrm{sv}} \alpha_i^*\phi(\boldsymbol{x}_i^*), \tag{23}$$

$$R^2=1-2\sum_{j=1}^{N_\mathrm{sv}}\alpha_j^* k(\boldsymbol{x}_j^*,\boldsymbol{x}_i^*)+\sum_{i,j=1}^{N_\mathrm{sv}}\alpha_j^*\alpha_i^* k(\boldsymbol{x}_j^*,\boldsymbol{x}_i^*), \tag{24}$$

where the data points $\boldsymbol{x}_i^*$ associated with $\alpha_i^*>0$ are named support vectors and N_sv is the number of support vectors. The distance between a test data point $\boldsymbol{z}$ and the centroid is calculated by

$$\begin{aligned} D^2(\boldsymbol{z}) &= \|\phi(\boldsymbol{z})-\boldsymbol{a}\|^2 \\ &= k(\boldsymbol{z},\boldsymbol{z})-2\sum_{j=1}^{N_\mathrm{sv}}\alpha_j^* k(\boldsymbol{x}_j^*,\boldsymbol{z})+\sum_{i,j=1}^{N_\mathrm{sv}}\alpha_j^*\alpha_i^* k(\boldsymbol{x}_j^*,\boldsymbol{x}_i^*) \\ &\le R^2. \end{aligned} \tag{25}$$

Finally, the confidence limits of these monitoring statistics can be determined. The local monitoring statistics involved in all the subspaces based on KSVDD models have been constructed.

4.2 Bayesian inference based monitoring statistics

Given N training data points $\{\boldsymbol{x}_i\}_{i=1}^{N}$ with C clusters, the optimal membership degree matrix $\boldsymbol{U}$ is obtained via LDRSKM. The data points within cluster c (c=1, 2, …, C) are represented by

$$X^c=\left\{\boldsymbol{x}\,|\,c=\max_{k=1,2,\cdots,C}\{\mu_{ik}(\boldsymbol{x}_i)\},i=1,2,\cdots,N\right\}.$$

The projection matrix $\boldsymbol{P}^c\in\mathbb{R}^{D\times d_c}$ is obtained by performing PCA on X^c. Given a new data point $\boldsymbol{x}_{\text{new}}\in\mathbb{R}^{D\times 1}$, $\boldsymbol{x}_{\text{new}}$ is firstly mapped into discriminant subspace c via projection matrix $\boldsymbol{P}^c$. The corresponding data points within the principal subspace c and residual subspace c are denoted as $\boldsymbol{x}_{pc,\text{new}}^c=(\boldsymbol{P}^c)^{\mathrm{T}}\boldsymbol{x}_{\text{new}}$ and $\boldsymbol{x}_{rs,\text{new}}^c=\boldsymbol{x}_{\text{new}}-(\boldsymbol{P}^c)\boldsymbol{x}_{pc,\text{new}}^c$, respectively. Then the corresponding monitoring statistics denoted by $D_{pc,\text{new}}^c$ and $D_{rs,\text{new}}^c$ are defined as the distances between the sample and the centers of KSVDD models.

Owing to the randomness of industrial process data, process data may be generated from multiple operating regions. Therefore, the probabilities of $\boldsymbol{x}_{\text{new}}$ under abnormal and normal conditions are estimated using the monitoring results from principal subspaces or residual subspaces. First, the posterior probabilities of $\boldsymbol{x}_{pc,\text{new}}^c$ and $\boldsymbol{x}_{rs,\text{new}}^c$ under abnormal condition in principal subspace c are calculated by

$$\begin{aligned}&P_{D_{pc}^c}(\text{AB}\,|\,\boldsymbol{x}_{pc,\text{new}}^c)\\&=P_{D_{pc}^c}(\boldsymbol{x}_{pc,\text{new}}^c\,|\,\text{AB})P_{D_{pc}^c}(\text{AB})\Big/P_{D_{pc}^c}(\boldsymbol{x}_{pc,\text{new}}^c),\end{aligned}\tag{26}$$

$$\begin{aligned}P_{D_{pc}^c}(\boldsymbol{x}_{pc,\text{new}}^c)&=P_{D_{pc}^c}(\boldsymbol{x}_{pc,\text{new}}^c\,|\,\text{AB})P_{D_{pc}^c}(\text{AB})\\&\quad+P_{D_{pc}^c}(\boldsymbol{x}_{pc,\text{new}}^c\,|\,\text{NM})P_{D_{pc}^c}(\text{NM}),\end{aligned}\tag{27}$$

where 'AB' and 'NM' represent abnormal and normal operating conditions, respectively, $P_{D_{pc}^c}(\text{AB})$ and $P_{D_{pc}^c}(\text{NM})$ are priors under abnormal and normal operating conditions in principal subspace c, respectively. The formulations of $P_{D_{rs}^c}(\text{AB}\,|\,\boldsymbol{x}_{pc,\text{new}}^c)$ and $P_{D_{rs}^c}(\boldsymbol{x}_{pc,\text{new}}^c)$ in residual subspace c are similar to Eqs. (26) and (27), respectively. The conditional probabilities of $\boldsymbol{x}_{pc,\text{new}}^c$ under abnormal and normal operating conditions are defined as in Ge *et al.* (2010):

$$P_{D_{pc}^c}(\boldsymbol{x}_{pc,\text{new}}^c\,|\,\text{AB})=\exp(-v_1D_{pc,\text{lim}}^c/D_{pc,\text{new}}^c),$$

$$P_{D_{pc}^c}(\boldsymbol{x}_{pc,\text{new}}^c\,|\,\text{NM})=\exp(-v_2\,D_{pc,\text{new}}^c\big/D_{pc,\text{lim}}^c),$$

where v_1, v_2>0, $D_{pc,\text{lim}}^c=r_{pc}^c$, $D_{pc,\text{new}}^c=D_{pc}^c(\boldsymbol{x}_{pc,\text{new}}^c)=\|\boldsymbol{\phi}(\boldsymbol{x}_{pc,\text{new}}^c)-\boldsymbol{a}_{pc}^c\|^2$, and $\boldsymbol{a}_{pc}^c$ and r_{pc}^c are the center and radius of the KSVDD model associated with principal subspace c, respectively. Similarly, $P_{D_{rs}^c}(\boldsymbol{x}_{rs,\text{new}}^c\,|\,\text{AB})$ and $P_{D_{rs}^c}(\boldsymbol{x}_{rs,\text{new}}^c\,|\,\text{NM})$ can be given. As with the computation procedure of $P_{D_{pc}^c}(\text{AB}|\,\boldsymbol{x}_{pc,\text{new}}^c)$ and $P_{D_{rs}^c}(\text{AB}\,|\,\boldsymbol{x}_{rs,\text{new}}^c)$, the posterior probabilities $P_{D_{pc}^c}(\text{NM}\,|\,\boldsymbol{x}_{pc,\text{new}}^c)$ and $P_{D_{rs}^c}(\text{NM}\,|\,\boldsymbol{x}_{rs,\text{new}}^c)$ can also be calculated. Given a significance level α, the priors under abnormal and normal operating conditions are simply chosen as α and $1-\alpha$, respectively.

To exploit the local monitoring results from principal and residual subspaces, the global Bayesian inference based fault detection indicators (BFDIs) are given by

$$\text{BFDI}_{D_{pc}}(\boldsymbol{x}_{\text{new}})=\sum_{c=1}^{C}P_{D_{pc}}(c\,|\,\boldsymbol{x}_{\text{new}})P_{D_{pc}^c}(\text{AB}\,|\,\boldsymbol{x}_{pc,\text{new}}^c),$$

$$\text{BFDI}_{D_{rs}}(\boldsymbol{x}_{\text{new}})=\sum_{c=1}^{C}P_{D_{rs}}(c\,|\,\boldsymbol{x}_{\text{new}})P_{D_{rs}^c}(\text{AB}\,|\,\boldsymbol{x}_{rs,\text{new}}^c),$$

where $P(c)$=N_c/N, and

$$P_{D_{pc}}(c\,|\,\boldsymbol{x}_{\text{new}})=P_{D_{pc}^c}(\boldsymbol{x}_{\text{new}})P(c)\Big/\left(\sum_{c=1}^{C}P_{D_{pc}^c}(\boldsymbol{x}_{\text{new}})P(c)\right),$$

$$P_{D_{rs}}(c\,|\,\boldsymbol{x}_{\text{new}})=P_{D_{rs}^c}(\boldsymbol{x}_{\text{new}})P(c)\Big/\left(\sum_{c=1}^{C}P_{D_{rs}^c}(\boldsymbol{x}_{\text{new}})P(c)\right).$$

The confidence control limits of global fault detection indicators, $\text{BFDI}_{D_{rs},\text{lim}}$ and $\text{BFDI}_{D_{pc},\text{lim}}$, are given by

$$\text{BFDI}_{D_{pc},\text{lim}}=\max_{\boldsymbol{x}_{\text{train}}}\{\text{BFDI}_{D_{pc}}(\boldsymbol{x}_{\text{train}})\},$$

$$\text{BFDI}_{D_{rs},\text{lim}}=\max_{\boldsymbol{x}_{\text{train}}}\{\text{BFDI}_{D_{rs}}(\boldsymbol{x}_{\text{train}})\}.$$

At a given significance level α and a new data point, the operating condition is treated as normality

if both of Bayesian inference fault detection indicators are no larger than their control limits. Otherwise, it is treated as abnormality. Details about Bayesian inference based monitoring statistics can be found in Ge *et al.* (2010).

A flow chart summarizing the proposed monitoring method is illustrated in Fig. 1.

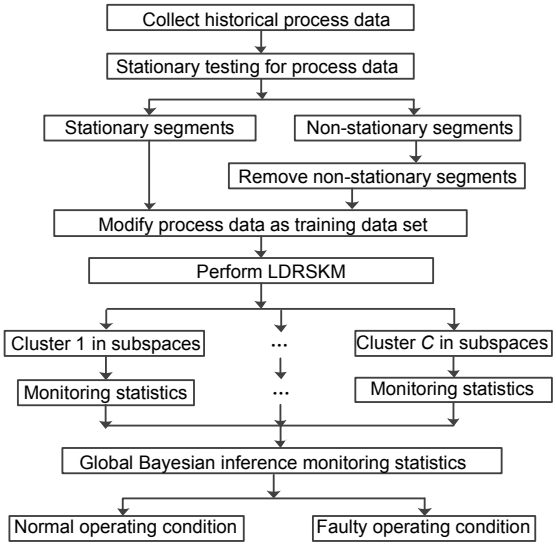

Fig. 1 Schematic diagram of the LDRSKM based process monitoring approach

5 Tennessee Eastman benchmark simulations

The Tennessee Eastman (TE) simulation plant has been widely used to evaluate the efficiency of process monitoring methods (Zhu *et al.*, 2010; Xu *et al.*, 2011; Ge and Song, 2012). A schematic diagram of TE is shown in Fig. 2 (Chiang *et al.*, 2000). The TE process consists of five major units: a reactor, a partial condenser, a recycle compressor, a vapor/liquid separator, and a stripper. The system also contains 41 measured and 12 manipulated variables. The control structure is according to the second structure proposed by Lyman and Georgakis (Molina *et al.*, 2011).

To obtain a better result from simulation, 16 variables listed in Table 1 were selected for monitoring purposes, as in Kano *et al.* (2002). All faults were introduced at the beginning of simulations. For the PCA-based method, the number of components is selected according to 95% of the total variance contained in the principal components.

There are six pre-specified operation modes in the platform to meet the demands of different production grades. Three operation modes shown in Table 2 were chosen in this study. The process data were collected by running the simulation for 60 h under normal conditions. The first 20 h was run under mode 1, the next 20 h under mode 2, and the last 20 h under mode 3. Each mode transition between two adjacent modes lasted 5 h. The sampling time interval was 0.05 h. Thus, 400 data points were acquired from each mode and 1200 data points constituted the historical data.

To achieve excellent results for comparison, one normal data set and a total of 20 programmable faults were simulated using the downloaded Simulink models. Twenty programmed faulty scenarios and a normal scenario were introduced to each process mode. Faults 1–7 were step changes of process variables, faults 8–12 were random changes of variables, fault 13 was a slow shift of reaction kinetics, faults 14 and 15 were related to valve sticking, and faults 16–20 were types of unknown faults. The details of

Table 1 The selected monitoring variables

No.	Measured variable
1	A feed
2	D feed
3	E feed
4	A and C feed
5	Recycle flow
6	Reactor feed rate
7	Reactor temperature
8	Purge rate
9	Product separator temperature
10	Product separator pressure
11	Product separator underflow
12	Stripper pressure
13	Stripper temperature
14	Stripper steam flow
15	Reactor cooling water outlet temperature
16	Separator cooling water outlet temperature

Table 2 Three operation modes used in simulation

Mode	Mass ratio (G/H)	Production rate (kg/h)
1	50/50	7038 (G), 7038 (H)
2	10/90	1111 (G), 10000 (H)
3	90/10	10000 (G), 1111 (H)

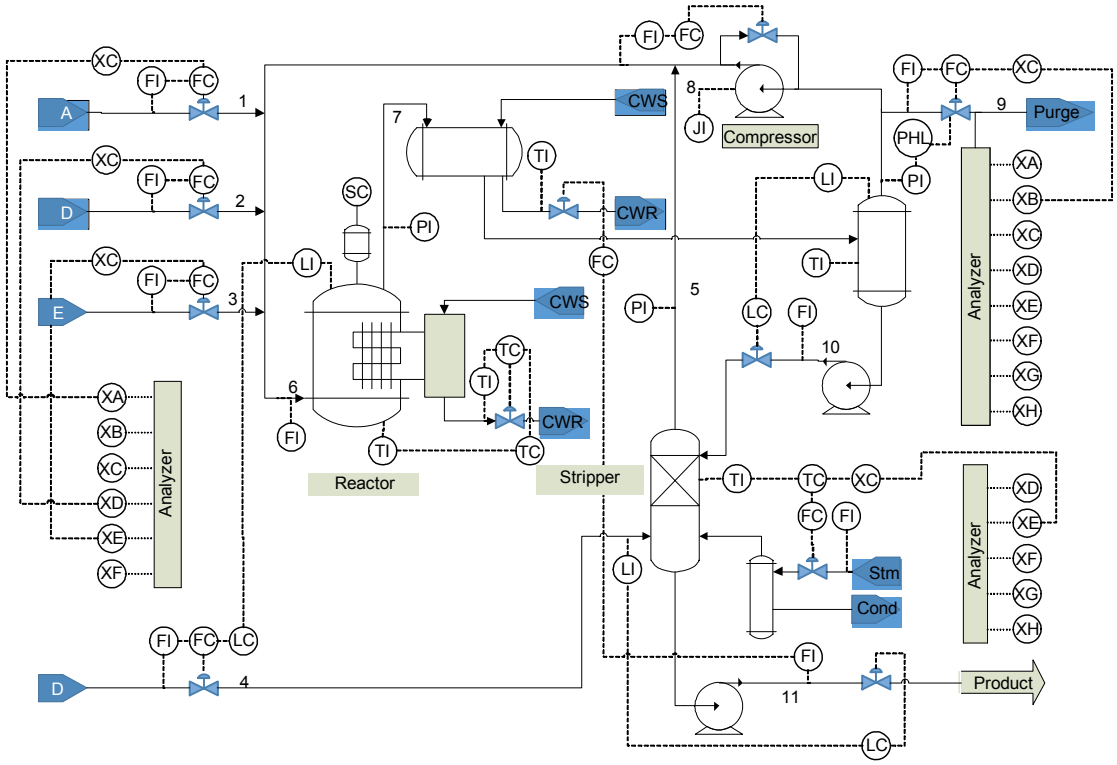

Fig. 2 Schematic diagram of the Tennessee Eastman process (descriptions of monitoring variables are given in Table 1)
Reprinted from Chiang *et al.* (2000), Copyright 2000, with permission from Elsevier

the 20 faults can be found in Zhu *et al.* (2012). The test data set also comprised 1200 samples, in which the first 200 samples were normal. Process faults were introduced from sample 201 to the end.

Since the efficiency and effectiveness of LPP-FCM for multimode process monitoring had been validated by comparing three monitoring approaches, adjoint principal component analysis (AdPCA) (Ng and Srinivasan, 2009), dynamic principal component analysis (DPCA), and Gaussian mixture model (GMM) (Xie and Shi, 2012), we validated our approach by comparing only to LPP-FCM. In the simulations, the confidence limit of the proposed approach was set at 99%. Before process monitoring, data points from the operation transition process and steady normal operations were first identified by the approach proposed by Zhu *et al.* (2012). Two hundred data points were collected from each of the steady normal operations. Then LDRSKM was performed on the 600 training points. In this simulation, the number of clusters C was determined by searching in interval [4, 8], the regularization parameters η and γ were determined by searching in intervals $[10^{-5}, 10^{-2}]$ and [0.1, 10], respectively, the reduced dimensionality of the data by LDRSKM was simply set as $C-1$, and the number of nearest neighbors was chosen in interval [7, 10]. Furthermore, KSVDDs were used to establish radii D_{pc} and D_{rs} in the principal and residual subspaces, respectively, associated with each of clusters. Although KSVDD can precisely describe non-Gaussian data sets, its performance depends heavily upon kernel parameters. For example, KSVDDs with different kernel parameter values conducted on the Banana data set are as shown in Fig. 3. The data show that a KSVDD with a small kernel parameter value is capable of describing complex nonlinear process data compactly. In this simulation, a compact searching interval was first specified for a kernel parameter by analyzing the characters of the operation mode associated with the cluster, and then the optimal kernel parameter could be obtained via cross-validation. The monitoring statistic parameters v_1 and v_2 were simply set as 1.

According to the cumulative percentage variance (CPV) rule, the number of principal components for PCA in the preprocessing step is 12, which can

explain 93% of the entire process data information. For simplicity, the Bayesian fault detection indicators calculated from principal and residual subspaces generated by LPP-FCM were denoted as T2 and SPE, respectively, and the Bayesian fault detection indicators by LDRSKM as Dpc and Drs, respectively. In this simulation, 100 test samples and 100 training samples were collected from each of the three stationary normal operation modes. The process monitoring results of LDRSKM and LPP-FCM for the training and test samples are given in Figs. 4 and 5, respectively. None of the control charts of the test and training data exceeded the control limits, and both the LDRSKM and LPP-FCM methods were capable of monitoring multiple normal operation modes. Moreover, the fault indicators of the LDRSKM approach could exhibit the distinctive characteristics of operation modes. As a critical prerequisite, a precise multimode process monitoring model enables the effective detection of abnormal operating conditions and transition processes in the monitoring stage.

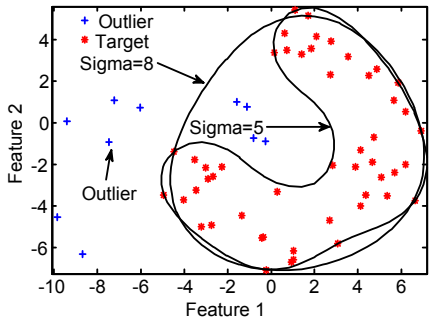

Fig. 3 KSVDD with different kernel parameter (Sigma) values on the Banana set (target class 1)

A multimode process often consists of multiple transition processes due to the changing of specifications of the products. In this simulation, two transition scenarios were considered. The first transition process switching from mode 1 to mode 2 began at sample 380 and ended around sample 430. The duration of the second transition process between mode 2 and mode 3 ranged from samples 780 to 830. The test samples for transition process monitoring were collected from samples 330 to 450 and from samples 750 to 930, respectively.

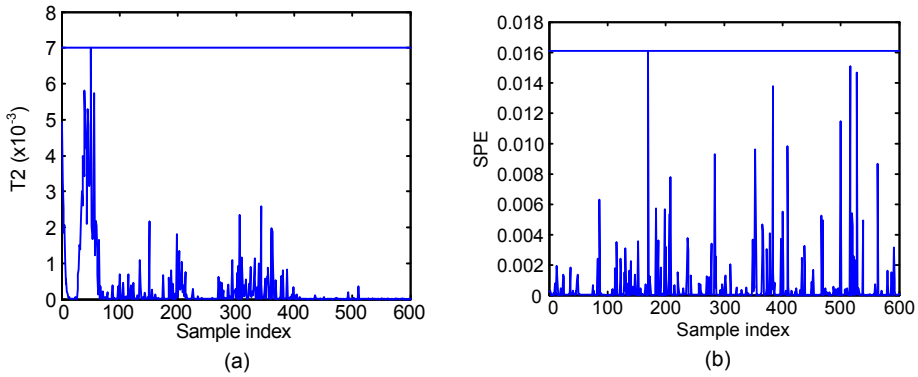

Fig. 4 LPP-FCM monitoring results for the three normal operation modes: (a) T2; (b) SPE

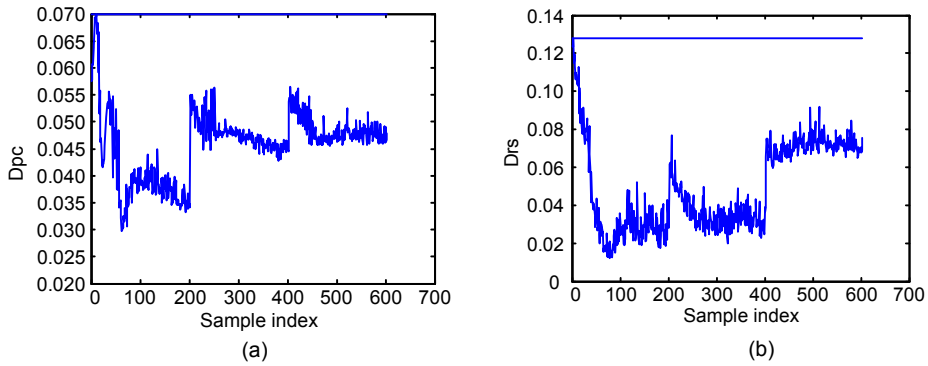

Fig. 5 LDRSKM monitoring results for the three normal operation modes: (a) Dpc; (b) Drs

The transition processes monitoring results of LPP-FCM and LDRSKM are shown in Figs. 6 and 7, respectively. Both of the control charts of LDRSKM were capable of detecting the first transition process with a short delay of around six samples; the second transition process could be detected precisely by the control chart in the residual subspace. Furthermore, the time of transition occurrence was very close to simulation settings and the transition durations were 48 and 52 samples, respectively. The time of transition occurrence also approximated to the identification results of LPP-FCM. However, both control charts of the LPP-FCM approach detected the abnormalities incurred by mode transition with a high miss detection rate, and could not show the duration of the transition process. The comparison of detection results for the transition process reveals that the proposed method has a higher sensitivity to mode transition in multimode processes than LPP-FCM and can trigger abnormal alarms in a timely manner.

To examine the effectiveness of the proposed approach for multimode industrial process monitoring, four test scenarios with different kinds of pre-specified faults were designed (Table 3). As pointed out by Downs and Vogel (1993), faults 3 and 9 are difficult to detect by the traditional PCA because they have little effect on the process and the fault amplitudes are quite small (Lee *et al.*, 2011). The monitoring results of our proposed method are shown in Figs. 8–12. From these five figures, we can make the following conclusions. Overall, the proposed algorithm gave better monitoring results, except for

Table 3 Four test cases with different kinds of process faults

Case	Description
1	Fault 2: 100 faulty samples under mode 1
	Fault 3: 290 faulty samples under each mode
2	Fault 4: 100 faulty samples under mode 1
	Fault 5: 290 faulty samples under each mode
3	Fault 7: 100 faulty samples under mode 1
	Fault 8: 290 faulty samples under each mode
4	Fault 10: 100 faulty samples under mode 1
	Fault 9: 290 faulty samples under each mode

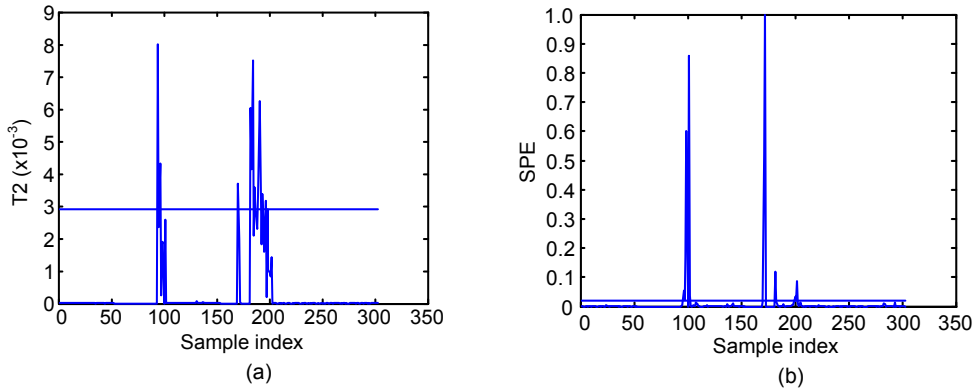

Fig. 6 LPP-FCM monitoring results for transition process data in principal and residual subspaces: (a) T2; (b) SPE

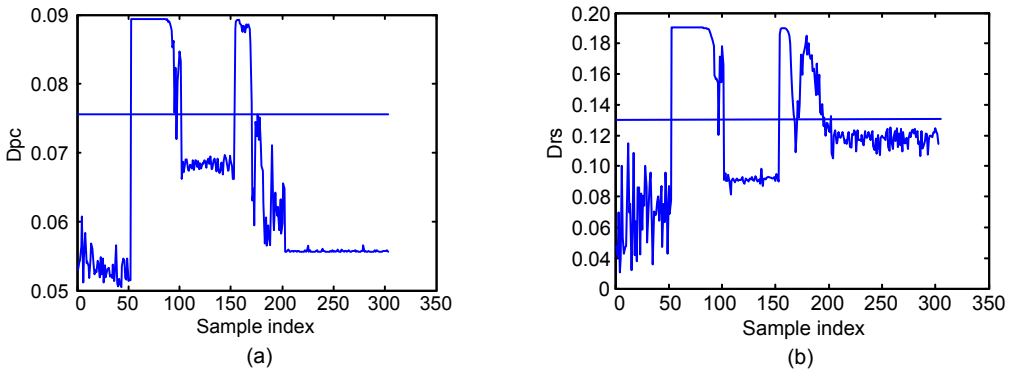

Fig. 7 LDRSKM monitoring results for transition process data in principal and residual subspaces: (a) Dpc; (b) Drs

fault 4, and had fewer detection delays than the LPP-FCM approach. PCA and its variants, such as AdPCA and DPCA, are prone to produce more detection delays than Bayesian-based monitoring approaches such as GMM, LPP-FCM, and LDRSKM. Compared to Bayesian-based multimode process monitoring approaches, our algorithm can perform clustering analysis with locality preservation and subspace selection based on discriminant analysis simultaneously, and the reduced data in the principal subspaces contain most of the useful information from the original data set. As a result, the faults can be detected reliably using the local monitoring results in the principal subspaces. Moreover, the fault detection

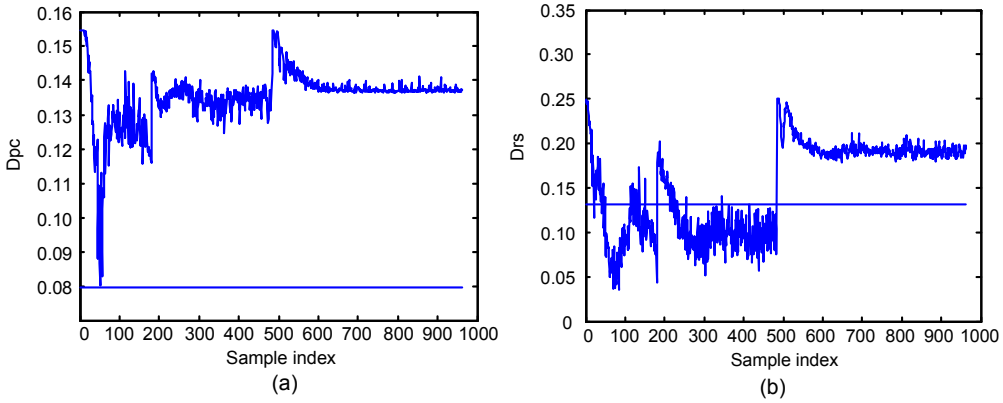

Fig. 8 LDRSKM monitoring results for fault 2 and fault 3: (a) Dpc; (b) Drs

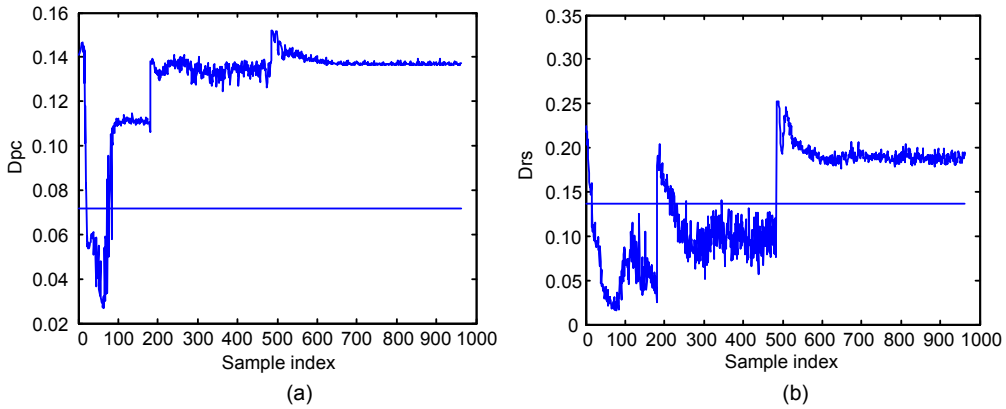

Fig. 9 LDRSKM monitoring results for fault 4 and fault 5: (a) Dpc; (b) Drs

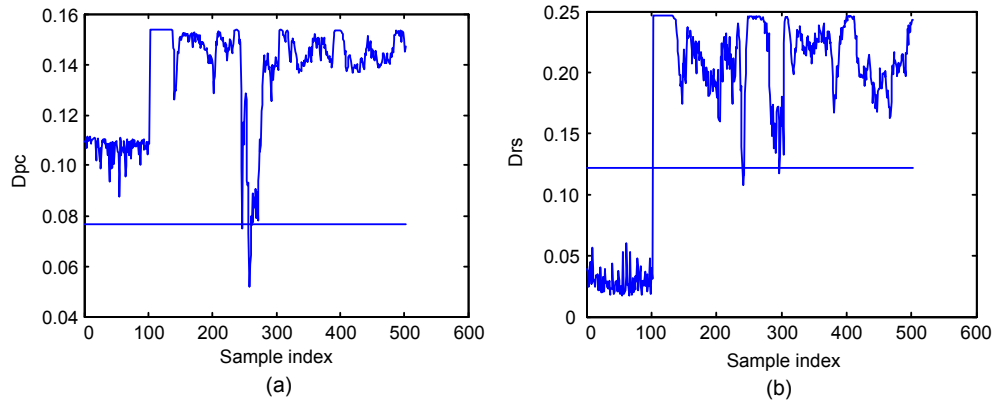

Fig. 10 LDRSKM monitoring results for fault 7 and fault 8: (a) Dpc; (b) Drs

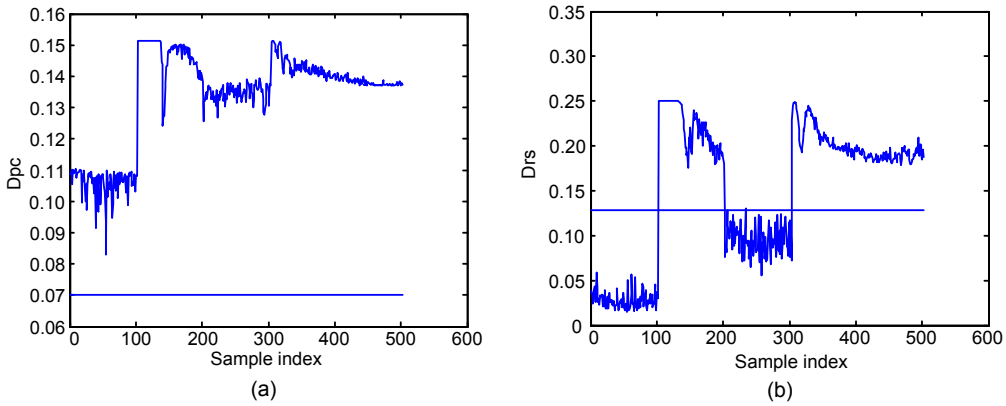

Fig. 11 LDRSKM monitoring results for fault 10 and fault 9: (a) Dpc; (b) Drs

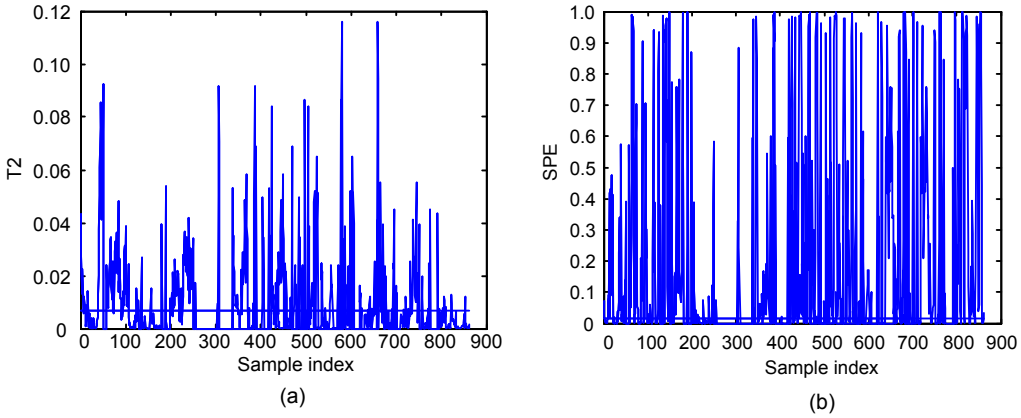

Fig. 12 LPP-FCM monitoring results for fault 8: (a) T2; (b) SPE

results (Figs. 8–13) show the efficiency of the proposed method and can reveal the characteristics of operating modes in the residual subspace, which can provide a feasible way to identify a working mode and help to diagnose faults.

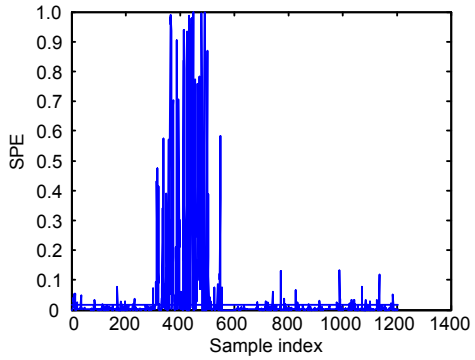

Fig. 13 LPP-FCM monitoring results for fault 9

6 Conclusions

In this paper, LDRSKM integrated with Bayesian inference is successfully developed for process monitoring. Aiming at multimode processes, LDRSKM is proposed to partition the overlapped multimodal operating data into optimal discriminant low-dimensional subspaces by iteratively performing soft clustering and discriminant dimensionality reduction. The local geometric structure of process data is preserved and multimodalities can be separated. Kernel support vector data description models are employed to construct local monitoring statistics due to the non-Gaussian nature of the data from each subspace. With the posterior probabilities of each monitored data point within the principal and residual subspaces under abnormal conditions, the global integrated monitoring statistic indicators are established to monitor multimode processes. Simulations

demonstrate that the proposed approach not only works for normal multimode processes but also detects transition processes with high precision and less delay. Compared to state-of-the-art multimode process monitoring approaches, our method can detect faults effectively and reliably and show satisfactory monitoring performance.

Although the proposed process monitoring approach provides a promising tool to monitor multimode processes without requiring prior process knowledge, the integrated monitoring statistics are sensitive to data autocorrelation due to system dynamics and thus require each normal operating mode be approximately at the steady state. Furthermore, our research is focused on fault detection of continuous processes and should be extended to fault diagnosis and the identification of root cause variables.

Acknowledgements

The authors sincerely thank Prof. Zhi-huan SONG for his advice and supervision.

References

Cai, L.F., Tian, X.M., Zhang, N., 2014. A kernel time structure independent component analysis method for nonlinear process monitoring. *Chin. J. Chem. Eng.*, **22**(11-12):1243-1253. [doi:10.1016/j.cjche.2014.09.021]

Chiang, L.H., Russell, E.L., Braatz, R.D., 2000. Fault diagnosis in chemical processes using Fisher discriminant analysis, discriminant partial least squares, and principal component analysis. *Chemometr. Intell. Lab. Syst.*, **50**(2):243-252. [doi:10.1016/S0169-7439(99)00061-1]

Deng, X.G., Tian, X.M., 2013. Sparse kernel locality preserving projection and its application in nonlinear process fault detection. *Chin. J. Chem. Eng.*, **21**(2):163-170. [doi:10.1016/S1004-9541(13)60454-1]

Deng, X.G., Tian, X.M., Chen, S., 2013. Modified kernel principal component analysis based on local structure analysis and its application to nonlinear process fault diagnosis. *Chemometr. Intell. Lab. Syst.*, **127**:195-209. [doi:10.1016/j.chemolab.2013.07.001]

Dong, W.W., Yao, Y., Gao, F.R., 2012. Phase analysis and identification method for multiphase batch processes with partitioning multi-way principal component analysis (MPCA) model. *Chin. J. Chem. Eng.*, **20**(6):1121-1127. [doi:10.1016/S1004-9541(12)60596-5]

Downs, J.J., Vogel, E.F., 1993. A plant-wide industrial process control problem. *Comput. Chem. Eng.*, **17**(3):245-255. [doi:10.1016/0098-1354(93)80018-I]

Feital, T., Kruger, U., Dutra, J., *et al.*, 2013. Modeling and performance monitoring of multivariate multimodal processes. *AIChE J.*, **59**(5):1557-1569. [doi:10.1002/aic.13953]

Ge, Z.Q., Song, Z.H., 2010. Maximum-likelihood mixture factor analysis model and its application for process monitoring. *Chemometr. Intell. Lab. Syst.*, **102**(1):53-61. [doi:10.1016/j.chemolab.2010.04.002]

Ge, Z.Q., Song, Z.H., 2012. A distribution-free method for process monitoring. *Exp. Syst. Appl.*, **38**(8):9812-9829. [doi:10.1016/j.eswa.2011.02.048]

Ge, Z.Q., Zhang, M.G., Song, Z.H., 2010. Nonlinear process monitoring based on linear subspace and Bayesian inference. *J. Process Contr.*, **20**(5):676-688. [doi:10.1016/j.jprocont.2010.03.003]

Ge, Z.Q., Song, Z.H., Gao, F.R., 2013. Review of recent research on data-based process monitoring. *Ind. Eng. Chem. Res.*, **52**(10):3543-3562. [doi:10.1021/ie302069q]

Ghosh, K., Ramteke, M., Srinivasan, R., 2014. Optimal variable selection for effective statistical process monitoring. *Comput. Chem. Eng.*, **60**:260-276. [doi:10.1016/j.compchemeng.2013.09.014]

He, X.F., Cai, D., Shao, Y.L., *et al.*, 2011. Laplacian regularized Gaussian mixture model for data clustering. *IEEE Trans. Knowl. Data Eng.*, **23**(9):1406-1418. [doi:10.1109/TKDE.2010.259]

Howland, P., Wang, J., Park, H., 2006. Solving the small sample size problem in face recognition using generalized discriminant analysis. *Patt. Recog.*, **39**(2):277-287. [doi:10.1016/j.patcog.2005.06.013]

Jing, L.P., Ng, M.K., Huang, J.Z., 2007. An entropy weighting *k*-means algorithm for subspace clustering of high-dimensional sparse data. *IEEE Trans. Knowl. Data Eng.*, **19**(8):1026-1041. [doi:10.1109/TKDE.2007.1048]

Kano, M., Nagao, K., Hasebe, S., *et al.*, 2002. Comparison of multivariate statistical process monitoring methods with applications to the Eastman challenge problem. *Comput. Chem. Eng.*, **26**(2):161-174. [doi:10.1016/S0098-1354(01)00738-4]

Kano, M., Fujioka, T., Tonomura, O., *et al.*, 2007. Data-based and model-based blockage diagnosis for stacked microchemical processes. *Chem. Eng. Sci.*, **62**(4):1073-1080. [doi:10.1016/j.ces.2006.11.011]

Lee, D., Lee, J., 2007. Domain described support vector classifier for multi-classification problems. *Patt. Recog.*, **40**(1):41-51. [doi:10.1016/j.patcog.2006.06.008]

Lee, J., Kang, B., Kang, S., 2011. Integrating independent component analysis and local outlier factor for plant-wide process monitoring. *J. Process Contr.*, **21**(7):1011-1021. [doi:10.1016/j.jprocont.2011.06.004]

Liu, J.L., Cai, D., He, X.F., 2010. Gaussian mixture model with local consistency. Proc. 24th AAAI Conf. on Artificial Intelligence, p.512-517.

Miao, A.M., Ge, Z.Q., Song, Z.H., *et al.*, 2015. Nonlocal structure constrained neighborhood preserving embedding model and its application for fault detection. *Chemometr. Intell. Lab. Syst.*, **142**:184-196. [doi:10.1016/j.chemolab.2015.01.010]

Miyamoto, S., Mukaidono, M., 1997. Fuzzy C-means as a regularization and maximum entropy approach. Proc.

IFSA, p.86-92.

Molina, G.D., Zumoffen, D.A.R., Basualdo, M.S., 2011. Plant-wide control strategy applied to the Tennessee Eastman process at two operating points. *Comput. Chem. Eng.*, **35**(10):2081-2097. [doi:10.1016/j.compchemeng.2010.11.006]

Ng, Y.S., Srinivasan, R., 2009. An adjoined multi-model approach for monitoring batch and transient operations. *Comput. Chem. Eng.*, **33**(4):887-902. [doi:10.1016/j.compchemeng.2008.11.014]

Perez, C.F.A., 2011. Fault Diagnosis with Reconstruction-Based Contributions for Statistical Process Monitoring. PhD Thesis, University of Southern California, USA.

Serradilla, J., Shi, J.Q., Morris, A.J., 2011. Fault detection based on Gaussian process latent variable models. *Chemometr. Intell. Lab. Syst.*, **109**(1):9-21. [doi:10.1016/j.chemolab.2011.07.003]

Shen, J.F., Bu, J.J., Ju, B., *et al.*, 2012. Refining Gaussian mixture model based on enhanced manifold learning. *Neurocomputing*, **87**:19-25. [doi:10.1016/j.neucom.2012.01.029]

Song, B., Ma, Y.X., Shi, H.B., 2014. Multimode process monitoring using improved dynamic neighborhood preserving embedding. *Chemometr. Intell. Lab. Syst.*, **135**:17-30. [doi:10.1016/j.chemolab.2014.03.013]

Tan, S.C., Lim, C.P., Rao, M.V.C., 2007. A hybrid neural network model for rule generation and its application to process fault detection and diagnosis. *Eng. Appl. Artif. Intell.*, **20**(2):203-213. [doi:10.1016/j.engappai.2006.06.007]

Teppola, P., Mujunen, S.P., Minkkinen, P., 1999. Adaptive fuzzy C-means clustering in process monitoring. *Chemometr. Intell. Lab. Syst.*, **45**(1-2):23-38. [doi:10.1016/S0169-7439(98)00087-2]

Tong, C.D., Palazoglu, A., Yan, X.F., 2013. An adaptive multimode process monitoring strategy based on mode clustering and mode unfolding. *J. Process Contr.*, **23**(10):1497-1507. [doi:10.1016/j.jprocont.2013.09.017]

Venkatasubramanian, V., Rengaswamy, R., Yin, K., *et al.*, 2003. A review of process fault detection and diagnosis: Part I: quantitative model-based methods. *Comput. Chem. Eng.*, **27**(3):293-311. [doi:10.1016/S0098-1354(02)00160-6]

Xie, L., Liu, X.Q., Zhang, J.M., *et al.*, 2009. Non-Gaussian process monitoring based on NGPP-SVDD. *Acta Autom. Sin.*, **35**(1):107-112 (in Chinese).

Xie, X., Shi, H.B., 2012. Multimode process monitoring based on fuzzy C-means in locality preserving projection subspace. *Chin. J. Chem. Eng.*, **20**(6):1174-1179. [doi:10.1016/S1004-9541(12)60604-1]

Xu, X., Xie, L., Wang, S., 2011. Multi-mode process monitoring method based on PCA mixture model. *CIESC J.*, **62**(3):743-752 (in Chinese).

Yang, Y.H., Li, X., Liu, X.Z., *et al.*, 2015. Wavelet kernel entropy component analysis with application to industrial process monitoring. *Neurocomputing*, **147**:395-402. [doi:10.1016/j.neucom.2014.06.045]

Yin, X.S., Chen, S.C., Hu, E.L., 2013. Regularized soft *K*-means for discriminant analysis. *Neurocomputing*, **103**:29-42. [doi:10.1016/j.neucom.2012.08.021]

Yu, J., 2012. A nonlinear kernel Gaussian mixture model based inferential monitoring approach for fault detection and diagnosis of chemical processes. *Chem. Eng. Sci.*, **68**(1):506-519. [doi:10.1016/j.ces.2011.10.011]

Zang, X., Vista Iv, F.P., Chong, K.T., 2014. Fast global kernel fuzzy c-means clustering algorithm for consonant/vowel segmentation of speech signal. *J. Zhejiang Univ.-Sci. C (Comput. & Electron.)*, **15**(7):551-563. [doi:10.1631/jzus.C1300320]

Zhang, M., Ge, Z.Q., Song, Z.H., *et al.*, 2011. Global-local structure analysis model and its application for fault detection and identification. *Ind. Eng. Chem. Res.*, **50**(11):6837-6848. [doi:10.1021/ie102564d]

Zhang, S.J., Wang, Z.L., Qian, F., 2010. FS-SVDD based on LTSA and its application to chemical process monitoring. *CIESC J.*, **61**(8):1894-1900 (in Chinese).

Zhang, Y.W., 2009. Enhanced statistical analysis of nonlinear processes using KPCA, KICA and SVM. *Chem. Eng. Sci.*, **64**(5):801-811. [doi:10.1016/j.ces.2008.10.012]

Zhang, Y.W., Li, S., 2014. Modeling and monitoring of nonlinear multi-mode processes. *Contr. Eng. Pract.*, **22**:194-204. [doi:10.1016/j.conengprac.2013.04.007]

Zhang, Y.W., An, J.Y., Li, Z.M., *et al.*, 2013. Modeling and monitoring for handling nonlinear dynamic processes. *Inform. Sci.*, **235**:97-105. [doi:10.1016/j.ins.2012.04.023]

Zhu, Z.B., Wang, P.L., Song, Z.H., 2010. PCA-SVDD based fault detection and self-learning identification. *J. Zhejiang Univ. (Eng. Sci.)*, **44**(4):652-658 (in Chinese).

Zhu, Z.B., Song, Z.H., Palazoglu, A., 2012. Process pattern construction and multi-mode monitoring. *J. Process Contr.*, **22**(1):247-262. [doi:10.1016/j.jprocont.2011.08.002]