An improved mixture of probabilistic PCA for nonlinear data-driven process monitoring

Article

Accepted Version

An improved mixture of probabilistic PCA for nonlinear data-driven process monitoring

Jingxin Zhang, *Student Member, IEEE*, Hao Chen, *Member, IEEE*, Songhang Chen, and Xia Hong, *Senior Member, IEEE*

***Abstract*—An improved mixture of probabilistic principal component analysis (PPCA) has been introduced for nonlinear data-driven process monitoring in this paper. To realize this purpose, the technique of a mixture of probabilistic principal component analysers is utilized to establish the model of the underlying nonlinear process with local PPCA models, where a novel composite monitoring statistic is proposed based on the integration of two monitoring statistics in modified PPCA-based fault detection approach. Besides, the weighted mean of the monitoring statistics aforementioned is utilised as a metrics to detect potential abnormalities. The virtues of the proposed algorithm have been discussed in comparison with several unsupervised algorithms. Finally, Tennessee Eastman process and an autosuspension model are employed to demonstrate the effectiveness of the proposed scheme further.**

***Index Terms*—Data driven, process monitoring, nonlinear systems, mixture of probabilistic principal component analysis**

I. Introduction

For the sake of system reliability and operational safety, large-scale industrial systems and applications increasingly demand improved process monitoring technologies, which have been extensively researched in recent decades [1–4]. However, often quantitative models are difficult to establish due to lack of prior knowledge. Alternatively sensing measurements which can replicate the desired process behavior are generally available and are utilized to design data-driven models [5, 6]. Thus, data-driven process monitoring techniques are becoming more prevalent and have been recognized as powerful tools for fault diagnosis purpose by comparison with knowledge-based approaches and approaches based on analytical model [7–9].

Data-driven process monitoring methods have been extensively researched [10, 11]. S. X. Ding has presented a comprehensive introduction on data-driven design of fault diagnosis and summarized several multivariate analysis techniques to fault diagnosis [12], i.e., principal component analysis (PCA), dynamic PCA, partial least squares (PLS), canonical variate analysis (CCA), etc. Besides, benchmark applications such as three-tank system, continuous stirred tank heater, Tennessee Eastman process, are utilized to demonstrate the effectiveness of these approaches. Yin et al also summarized a variety of data-driven techniques for multivariate statistical process monitoring (MSPM), and analysed their computational complexities as well as data assumption [13], including independent component analysis (ICA), fisher discriminant analysis (FDA), subspace aided approach, etc. However, these data-driven techniques are mainly applied to linear systems. Aimed at nonlinear applications, various variants of these basic approaches are developed, e.g., standard techniques under Gaussian model or probabilistic model [14–17].

Amongst many data-driven approaches, PCA serves as a classical technique for feature extraction due to its simplicity and effectiveness [18, 19]. Therefore, PCA has been widely employed for MSPM in recent years [20–22]. Several extensions of traditional PCA have been proposed to settle different issues, e.g., parameter variation [23], practical batch process [24], large-scale process [25], detecting slowly developing drifts [26].

However, traditional PCA scheme still has various limitations, e.g. the basic assumption of multivariate Gaussian distributed data. Besides, the efficiency and detectability of PCA-based technique would be greatly discounted for nonlinear applications. Numerous sophisticated variants of PCA have been intensively studied to tackle these problems. For instance, the locally weighted projection regression (LWPR) is a nonlinear regression method [27], where PCA-based process monitoring model can be computed under the locally weighted framework [28]. In [29, 30], just-in-time learning (JITL) or neural network (NN) serves as process model to account for the nonlinear as well as dynamic behavior of the process, followed by PCA which analyzes the residuals from the difference between predicted outputs and process outputs. In addition, PCA under Gaussian mixture model can be free from the assumption of Gaussian distribution [31]. In work of [32–34], kernel PCA (KPCA) is mainly aimed at tackling nonlinearity problem. Furthermore, other forms of nonlinear PCA are also discussed in [35, 36]. Nevertheless, the effectiveness of KPCA is largely dependent on the option of kernel functions and the corresponding critical parameters. In addition, traditional PCA algorithm performs badly when data values are incomplete. Aimed at this problem, probabilistic PCA (PPCA) was proposed, where expectation maximization (EM) algorithm could estimate the principal subspace iteratively [37]. Note that the applications of PPCA should satisfy the basic assumption that process data follow multivariate Gaussian distribution.

In order to deal with nonlinearities that are inherent in many underlying systems, it is of practical interest to integrate multiple PCA models to get complicated projection

Jingxin Zhang, Hao Chen and Songhang Chen are with the Department of Quanzhou Institute of Equipment Manufacturing, Haixi Institutes, Chinese Academy of Sciences, Jinjiang 362200, China. (e-mail: jx.zhang@fjirsm.ac.cn; chenhao@fjirsm.ac.cn; songhang.chen@fjirsm.ac.cn).

Xia Hong is with the Department of Computer Science, School of Mathematical, Physical and Computational Sciences, University of Reading, RG6 6AY, U.K.(email: x.hong@reading.ac.uk).

schemes [38]. Taking the virtues of PPCA aforementioned into consideration, a complicated model is readily implemented as a combination of such PPCA models via the technique of a mixture of probabilistic principal component analysers, namely, MPPCA approach [39, 40]. MPPCA enables dealing with any probability density function and can figure out the global linearity of PCA. Generally speaking, MPPCA inherits the benefits of PPCA and can be applied to nonlinear systems.

Due to virtues aforementioned, MPPCA has been utilized for process monitoring [41, 42]. In [41], MPPCA is utilized for sensor fault diagnosis purpose, but detailed theory of fault detection logic is unavailable. A mixture Bayesian regularization method of PPCA was proposed for multimode process monitoring [42]. However, it can not be applied to nonlinear and non-Gaussian systems. Therefore, it is valuable and necessary to present a specific description about improved MPPCA approach for nonlinear process monitoring.

In this paper, we proposed an improved nonlinear data-driven process monitoring algorithm based on MPPCA, referred to as I-MPPCA. A new monitoring statistic is introduced based on the integration of two monitoring statistics in modified PPCA-based fault detection approach. Besides, the weighted mean of the monitoring statistics aforementioned is developed to detect potential abnormalities. The major advantages of I-MPPCA are summarized as follows:

1) It can cope with the global linearity of PCA, which is appropriate and effective to monitor nonlinear process;

2) The weight of a new data point belonging to a certain local PPCA model can be interpreted by the probability of each local model of being chosen;

3) It can process with any probability density function and owns lower computational complexity as well as stronger parameter robustness than kernel approaches;

4) It can deliver an optimal monitoring performance even when some data values are missing;

5) Compared with traditional MPPCA technique, only one or two global monitoring statistics are developed for process monitoring, which is more convenient for practical industrial applications.

The rest of this paper is organized below. Section II reviews concepts and mathematical formulations of the probabilistic PCA and MPPCA as preliminaries of our proposed approach. Section III details the proposed I-MPPCA approach, in which we propose model selection for I-MPPCA, novel monitoring statistics as well as corresponding thresholds. Besides, the monitoring performance is discussed between the proposed approach and the existing unsupervised approaches. In Section IV, Tennessee Eastman (TE) process is employed to illustrate the rationality and virtues of the proposed approach in contrast with traditional MPPCA scheme. Then, an autosuspension model is adopted to demonstrate the superiorities of the proposed approach in comparison with other unsupervised schemes in Section V. Concluding remarks are given in Section VI.

II. Preliminaries

A. Latent variable models and PCA

Consider using the following model

$$\boldsymbol{t} = \boldsymbol{y}(\boldsymbol{x}; \boldsymbol{w}) + \boldsymbol{\xi} \tag{1}$$

to describe the observation process vector $\boldsymbol{t} \in \Re^d$, where $\boldsymbol{x} \in \Re^q$ is the vector of latent variables, $\boldsymbol{w}$ is the associated model parameter vector. $\boldsymbol{\varepsilon}$ is an independent noise vector. $\boldsymbol{y}(\boldsymbol{x}; \boldsymbol{w})$ is the unknown function of the system. For example, it can be interpreted by a linear model used in statistical factor analysis, given by

$$\boldsymbol{t} = \boldsymbol{W}\boldsymbol{x} + \boldsymbol{\mu} + \boldsymbol{\xi} \tag{2}$$

By means of defining a prior distribution over $\boldsymbol{x}$ and of $\boldsymbol{\xi}$, a related distribution is induced in the data space according to (2). Then, maximum-likelihood approach is utilized to determine the model parameters given a set observational data.

We assume $\boldsymbol{x} \sim \mathrm{N}(\boldsymbol{0}, \boldsymbol{I})$, $\boldsymbol{\xi} \sim \mathrm{N}(\boldsymbol{0}, \boldsymbol{\Psi})$. $\boldsymbol{0}$ and $\boldsymbol{I}$ denote the vector of all zeros and identity matrix with appropriate dimensions respectively. $\boldsymbol{\Psi} \in \Re^{d\times d}$ is assumed to be a diagonal matrix, $\boldsymbol{\mu} \in \Re^d$ is the output mean vector, $\boldsymbol{W} \in \Re^{d\times q}$ is the matrix of loading factors. Based on (2), it can be shown that the observation vector obeys Gaussian distribution $\boldsymbol{t} \sim \mathrm{N}(\boldsymbol{\mu}, \boldsymbol{C})$ with $\boldsymbol{C} = \boldsymbol{\Psi} + \boldsymbol{W}\boldsymbol{W}^{\mathrm{T}} \in \Re^{d\times d}$.

Consider that a data set of N output data samples $\{\boldsymbol{t}_n\}_{n=1}^N$ is available. Over the data set, (2) can be represented in matrix form as

$$\boldsymbol{T} = \boldsymbol{W}\boldsymbol{X} + \boldsymbol{u}\boldsymbol{1}^{\mathrm{T}} + \boldsymbol{\Xi} \tag{3}$$

where $\boldsymbol{T} = [\boldsymbol{t}_1, ..., \boldsymbol{t}_N] \in \Re^{d\times N}$, $\boldsymbol{X} = [\boldsymbol{x}_1, ...\boldsymbol{x}_N] \in \Re^{q\times N}$, $\boldsymbol{\Xi} = [\boldsymbol{\xi}_1, ...\boldsymbol{\xi}_N] \in \Re^{d\times N}$. $\boldsymbol{1}$ denotes the vector of all ones with appropriate dimension. Let $\boldsymbol{u} = \frac{1}{N}\sum_{n=1}^N \boldsymbol{t}_n$, and the sample covariance matrix be denoted by $\boldsymbol{S} = \frac{1}{N}(\boldsymbol{T} - \boldsymbol{u}\boldsymbol{1}^{\mathrm{T}})(\boldsymbol{T} - \boldsymbol{u}\boldsymbol{1}^{\mathrm{T}})^{\mathrm{T}}$. There exist certain links between factor analysis and PCA, which have been demonstrated in [43]. PCA problem can be settled by factor analysis.

Denote the eigenvalue decomposition of $\boldsymbol{S} = \tilde{\boldsymbol{W}}\boldsymbol{\Lambda}\tilde{\boldsymbol{W}}^{\mathrm{T}}$, where $\boldsymbol{\Lambda} = \mathrm{diag}\{\lambda_1, ..., \lambda_q, \lambda_{q+1}, ..., \lambda_d\}$, $\lambda_1 > \lambda_2 > ... > \lambda_q > 0$ are nonzero eigenvalues $\boldsymbol{S}$, the $d - q$ smallest eigenvalues are minor and negligible. Specifically we find q-dimensional vectors

$$\boldsymbol{x}_n = \boldsymbol{W}^{\mathrm{T}}(\boldsymbol{t}_n - \boldsymbol{\mu}) \tag{4}$$

to represent $\boldsymbol{t}_n$, $n = 1, ...N$, where $\boldsymbol{W}$ is the first q columns of $\tilde{\boldsymbol{W}}$ and q is determined by accumulating contribution rate. It can be shown that projection onto most dominant eigenvectors leads to finding lower dimensional latent variables whilst retaining maximal variance of the original variables $\boldsymbol{t}_n$.

B. The probabilistic PCA

Providing that noise $\boldsymbol{\varepsilon} \sim \mathrm{N}(\boldsymbol{0}, \sigma^2\boldsymbol{I})$, a probability distribution over $\boldsymbol{t}$-space is revealed for a specific $\boldsymbol{x}$ by the following formula

$$p(\boldsymbol{t}|\boldsymbol{x}) = (2\pi\sigma^2)^{-d/2}\exp\left\{-\frac{1}{2\sigma^2}\|\boldsymbol{t} - \boldsymbol{W}\boldsymbol{x} - \boldsymbol{\mu}\|^2\right\} \tag{5}$$

A Gaussian prior probability over $\boldsymbol{x}$ can be defined by

$$p(\boldsymbol{x}) = (2\pi)^{-q/2} \exp\left\{-\frac{1}{2}\boldsymbol{x}^{\mathrm{T}}\boldsymbol{x}\right\} \tag{6}$$

Then, the marginal distribution of $\boldsymbol{t}$ can be acquired in the form of

$$\begin{aligned} p(\boldsymbol{t}) &= \int p(\boldsymbol{t}|\boldsymbol{x})p(\boldsymbol{x})d\boldsymbol{x} \\ &= (2\pi)^{-d/2}|\boldsymbol{C}|^{-1/2} \exp\left\{-\tfrac{1}{2}(\boldsymbol{t}-\boldsymbol{\mu})^{\mathrm{T}}\boldsymbol{C}^{-1}(\boldsymbol{t}-\boldsymbol{\mu})\right\} \end{aligned} \tag{7}$$

where $|\cdot|$ denotes matrix determinant. The model covariance is given by

$$\boldsymbol{C} = \sigma^2\boldsymbol{I} + \boldsymbol{W}\boldsymbol{W}^{\mathrm{T}} \tag{8}$$

In accordance with Bayesian theory, given the observation vector $\boldsymbol{t}$, the corresponding posterior distribution of $\boldsymbol{x}$ may be calculated:

$$\begin{aligned} p(\boldsymbol{x}|\boldsymbol{t}) = \exp\Bigg\{-\frac{1}{2}\left\{\boldsymbol{x}-\boldsymbol{M}^{-1}\boldsymbol{W}^{\mathrm{T}}(\boldsymbol{t}-\boldsymbol{\mu})\right\}^{\mathrm{T}}(\sigma^{-2}\boldsymbol{M}) \\ \left\{\boldsymbol{x}-\boldsymbol{M}^{-1}\boldsymbol{W}^{\mathrm{T}}(\boldsymbol{t}-\boldsymbol{\mu})\right\}\Bigg\} \times (2\pi)^{-q/2}\left|\sigma^{-2}\boldsymbol{M}\right|^{1/2} \end{aligned} \tag{9}$$

where the posterior covariance satisfies

$$\sigma^2\boldsymbol{M}^{-1} = \sigma^2(\sigma^2\boldsymbol{I} + \boldsymbol{W}^{\mathrm{T}}\boldsymbol{W})^{-1} \tag{10}$$

with $\boldsymbol{M} \in \Re^{q\times q}$, $\boldsymbol{C} \in \Re^{d\times d}$.

The log-likelihood of the observation vector in this model is

$$\begin{aligned} L &= \sum_{n=1}^{N} \ln\left\{p\left(\boldsymbol{t}_n\right)\right\} \\ &= -\tfrac{N}{2}\left\{d\ln(2\pi) + \ln|\boldsymbol{C}| + tr\left(\boldsymbol{C}^{-1}\boldsymbol{S}\right)\right\} \end{aligned} \tag{11}$$

It can be shown [39] that the log-likelihood (11) is maximized when the columns of $\boldsymbol{W}$ span the principal subspace of the data. Analytical solutions can also be obtained via the eigen-decomposition of $\boldsymbol{S}$, together with the estimation of noise variance σ^2 (based on the smallest eigenvalues of $\boldsymbol{S}$). Alternatively iterative method of the EM algorithm can be used to generate the following complete-data log-likelihood as

$$\begin{aligned} L_c &= \sum_{n=1}^{N} \ln\left\{p(\boldsymbol{t}_n, \boldsymbol{x}_n)\right\} \\ &= \sum_{n=1}^{N} \ln\Bigg\{(2\pi\sigma^2)^{-d/2}\exp\left\{-\frac{1}{2\sigma^2}\|\boldsymbol{t}_n - \boldsymbol{W}\boldsymbol{x}_n - \boldsymbol{\mu}\|^2\right\} \\ &\quad (2\pi)^{-q/2}\exp\left\{-\frac{1}{2}\boldsymbol{x}_n^{\mathrm{T}}\boldsymbol{x}_n\right\}\Bigg\} \end{aligned} \tag{12}$$

For convenience, we point out that the following qualities are used in EM algorithm [39]

$$\langle \boldsymbol{x}_n \rangle = \boldsymbol{M}^{-1}\boldsymbol{W}^{\mathrm{T}}\left(\boldsymbol{t}_n - \boldsymbol{\mu}\right) \tag{13}$$

$$\left\langle \boldsymbol{x}_n\boldsymbol{x}_n^{\mathrm{T}}\right\rangle = \sigma^2\boldsymbol{M}^{-1} + \langle \boldsymbol{x}_n\rangle\langle \boldsymbol{x}_n\rangle^{\mathrm{T}} \tag{14}$$

denoting the expected posterior mean and covariance vectors based on the latent model (2).

C. The mixture of probabilistic PCA

In order to be able to model more complex data, the mixture of probabilistic PCA (MPPCA) has been introduced [39], which takes advantage of an integration of local PCA models via defining a mixture of probabilistic densities on the predicted output from K local PCA models [39]. Our proposed I-MPPCA approach for process monitoring is based on MPPCA since it provides a more powerful base to handle nonlinearities and missing data.

Instead of representing the system (1) by a single model of (2), we will consider, in accordance to probability rules, a mixture of K local PCA models, as

$$\begin{aligned} p(\boldsymbol{t}) &= \sum_{i=1}^{K} p(i)p(\boldsymbol{t}|i) \\ &= \sum_{i=1}^{K} \pi_i p(\boldsymbol{t}|i)) \end{aligned} \tag{15}$$

where the constraints on the mixing coefficients are $\pi_i \geq 0$ and $\sum \pi_i = 1$. $p(i)$ is interpreted as probability of choosing the ith local model. Each of $p(\boldsymbol{t}|i)$ is the local PCA model given by

$$\boldsymbol{t} = \boldsymbol{W}_i\boldsymbol{x} + \boldsymbol{\mu}_i + \boldsymbol{\xi}_i, \quad i = 1, ..., K \tag{16}$$

which is similar to (2), and has individual projection matrice $\boldsymbol{W}_i$, mean vector $\boldsymbol{\mu}_i$, as well as $\boldsymbol{\xi}_i \sim \mathrm{N}(\boldsymbol{0}, \sigma_i^2\boldsymbol{I})$.

By comparing with equations (13) and (14), the expected posterior mean and covariance vectors based on each of K local PCA can be evaluated as

$$\left\langle \boldsymbol{x}_n^{(i)}\right\rangle = \boldsymbol{M}_i^{-1}\boldsymbol{W}_i^{\mathrm{T}}\left(\boldsymbol{t}_n - \boldsymbol{\mu}_i\right) \in \Re^q \tag{17}$$

$$\left\langle \boldsymbol{x}_n^{(i)}\boldsymbol{x}_n^{(i)\mathrm{T}}\right\rangle = \sigma_i^2\boldsymbol{M}_i^{-1} + \left\langle \boldsymbol{x}_n^{(i)}\right\rangle\left\langle \boldsymbol{x}_n^{(i)}\right\rangle^{\mathrm{T}} \in \Re^{q\times q} \tag{18}$$

and are given here for convenience. Note that $\boldsymbol{x}_n^{(i)}$ is nth sample for each ith model.

The solution procedure of MPPCA is presented in Appendix A, where two-stage EM scheme is adopted to improve convergence speed and reduce computational cost in Appendix B.

III. The proposed I-MPPCA approach for process monitoring

The major procedure of MPPCA-based fault detection scheme can be summarized thereinafter. First, the model framework including the basic information of local PPCA models is determined when the log-likelihood reaches the maximum. Then, with regard to each local PPCA model, two monitoring statistics, i.e., Hotelling's T-squared (T^2) and squared prediction error (SPE), are calculated by utilizing the probability density of the score vector. To this end, a procedure for on-line data-driven fault diagnosis is developed for nonlinear process.

According to the procedure aforementioned, it is evidently observed that standard MPPCA has several local models and each local PPCA model has two monitoring charts. Thus, there are excessive monitoring charts to be observed for MPPCA,

TABLE I
OVERALL PROCEDURE OF I-MPPCA APPROACH FOR PROCESS MONITORING

Phase 1: Off-line learning
1. Collect N normal sensing measurements from a nonlinear process.
2. For I-MPPCA with K local models
3. Estimate the model parameters.
   3a. Initialize the model parameters $\boldsymbol{W}_i$ and $\boldsymbol{\mu}_i$ $(i = 1 : K)$ via traditional PCA.
   3b. Calculate the number of principal components through accumulation contribution rate.
   3c. Discover the optimal model parameters $\boldsymbol{\theta} = \left\{\boldsymbol{W}_i, \boldsymbol{\mu}_i, \sigma_i^2, \pi_i\right\}_{i=1:K}$ by implementation of the two-stage EM schedule.
   3d. Calculate $H(K)$ and go to step 2 until $K = K_{\max}$.
4. Discover the optimal number of local models K° among $K = \{1, 2, \ldots, K_{\max}\}$ via the criterion given in (19).
5. Calculate the global monitoring statistics from equations (24)-(26) under normal operation condition.
6. Calculate the thresholds of the monitoring statistics from (29) based on the calculated statistics aforementioned.

Phase 2: On-line monitoring
1. Get new sample and preprocess input data according to $\boldsymbol{\mu}_i$ $(i = 1 : K^\circ)$ aforementioned.
2. Calculate local monitoring statistics in each local model according to equations (21)-(23).
3. Calculate the global monitoring statistics from equations (24)-(26).
4. Detect faults according to the fault detection logic according to (30) or (31).

which is difficult for staff to acquire the accurate information timely and quickly.

This paper is mainly dedicated to this issue, where just one or two global monitoring statistics are designed to monitor the nonlinear process. To realize this aim, T^2 and SPE monitoring statistics are integrated to achieve reliable monitoring performance for each local PPCA model. Then, the weighted mean of the combined monitoring statistics aforementioned in PPCA models is developed to detect underlying abnormalities. Besides, kernel density estimation (KDE) scheme is applied to calculate thresholds in order to reduce computational overload and enhance generality. Detailed introduction about the proposed fault diagnosis approach is described below.

A. Optimal number of the local models

It can be evidently discovered that key model parameters are largely affected by the number of local models K. Besides, the computational cost increases with K. Therefore, it is essential to seek for a reliable criteria to determine an optimal value of K.

In this paper, the optimal number of local models, K°, can be determined by the following criteria [44]:

$$K^\circ = \arg\min_i H\left(i\right) \tag{19}$$

$$H\left(i\right) \equiv -\frac{1}{N}\sum_{n=1}^{N}\sum_{i=1}^{K} p\left(i|\boldsymbol{t}_n, \boldsymbol{\theta}\right)\ln\left(p\left(\boldsymbol{t}_n|i\right)\right) - \sum_{i=1}^{K}\pi_i \ln \pi_i \tag{20}$$

where $\boldsymbol{\theta}$ is the set of all model parameter vectors, containing $\boldsymbol{\theta} = \left\{\boldsymbol{W}_i, \boldsymbol{\mu}_i, \sigma_i^2, \pi_i\right\}_{i=1,\ldots,K}$.

Given a value of K, $H(i)$ is calculated after the implementation of two-stage EM schedule as the model parameter $\boldsymbol{\theta}$ is contained in $H(i)$. Thus, the model structure and parameters are embedded in each other. The value of K can be determined alternatively through discovering the smallest integer satisfying $|H\left(i\right) - H\left(i+1\right)| > \delta$, where δ is a predefined threshold.

Note that the maximum number of local models $K_{\max}$ varies relying on data pattern. However, $K_{\max} = 5 \sim 10$ is typical.

B. Monitoring scheme of I-MPPCA

I-MPPCA is proposed to solve nonlinear fault diagnosis problem, which is exactly based on a mixture of PPCA-based fault detection models.

For each ith local model, T_i^2 and SPE_i are monitoring statistics for principal component subspace and residual component subspace respectively. As two subspaces are mutually orthogonal, T^2 statistic can not detect faults that occur in the residual component subspace and vice versa [12]. The above two monitoring statistics utilize the identical measurement unit, i.e. Mahalanobis norm, and they can be integrated into one chart. Therefore, $T_{c,i}$ is proposed based on the integration of T_i^2 and SPE_i for each local improved PPCA model. Three monitoring statistics are computed as follows

$$T_i^2 = \left\|\boldsymbol{M}_i\boldsymbol{W}_i^{\mathrm{T}}\boldsymbol{t}_n\right\|^2 \tag{21}$$

$$SPE_i = \left\|\sigma_i^{-1}(\boldsymbol{I} - \boldsymbol{W}_i\boldsymbol{M}_i\boldsymbol{W}_i^{\mathrm{T}})\boldsymbol{t}_n\right\|^2 \tag{22}$$

$$T_{c,i}^2 = \boldsymbol{t}_n^{\mathrm{T}}(\sigma_i^2\boldsymbol{I} + \boldsymbol{W}_i\boldsymbol{W}_i^{\mathrm{T}})^{-1}\boldsymbol{t}_n \tag{23}$$

In the proposed approach, R_{ni} is regarded as weight, which measures the degree of nth sample belonging to ith local model. Thus, the associated global monitoring statistics can be formulated as

$$T^2 = \sum_{i=1}^{K} R_{ni}T_i^2 \Big/ \sum_{i=1}^{K} R_{ni} \tag{24}$$

$$SPE = \sum_{i=1}^{K} R_{ni}SPE_i \Big/ \sum_{i=1}^{K} R_{ni} \tag{25}$$

$$T_c^2 = \sum_{i=1}^{K} R_{ni}T_{c,i}^2 \Big/ \sum_{i=1}^{K} R_{ni} \tag{26}$$

TABLE II
A BRIEF COMPARATIVE STUDY OF FOUR APPROACHES

Approach	Computational cost	Parameter	Robustness	Outliers	Big data	fault information
K-means	Low: $O(N)$	no	sensitive	sensitive	appropriate	only detecting whether faults occur
FPCM	Low: $O(N)$	no	sensitive	insensitive	appropriate	only detecting whether faults occur
KPCA	High: $O(N^2)$	No. of PCs	sensitive	insensitive	inappropriate	fault detection in two subspaces
I-MPPCA	Medium: $O(KN)$	No. of PCs, K	insensitive	insensitive	appropriate	fault detection in two subspaces

In consideration of the constraint that $\sum_i R_{ni} = 1$, the global statistics aforementioned can be much simplified without the necessity of calculating denominator.

C. Summary of the proposed approach

For the fault diagnosis purpose, the thresholds under normal condition are recognized as reference to detect underlying faults. With regard to traditional PCA-based approach, thresholds are calculated under the assumption of Gaussian distributed data. However, process variables can hardly satisfy this requirement in practical applications. In order to reduce computational cost and enhance generality, KDE technique is employed to calculate the associated thresholds [45], which can be applied to both Gaussian distributed data and non-Gaussian distributed data. The basic theory is described as follows.

$$\hat{p}(z) = \frac{1}{Nh}\sum_{n=1}^{N}\psi(\frac{z-z_n}{h}) \tag{27}$$

where $z_n (n = 1, \ldots, N)$ are the values of monitoring statistics, h is the bandwidth of kernel function $\psi(\cdot)$. The selection of h is significant because the consequences of $p(z)$ estimation would be rough if h is small, whereas the density curve would be smooth. In this paper, the optimal bandwidth h_{opt} is determined by minimizing the approximation of the mean integrated square error, as depicted in (28), where s is the standard deviation [45].

$$h_{opt} = 1.06sN^{-1/5} \tag{28}$$

Given a confidence level α, the associated threshold J_{th} of the monitoring statistic J can be calculated by

$$\int_{-\infty}^{J_{th}} p(J)dJ = \alpha \tag{29}$$

J can be replaced by T^2, SPE and T_c^2 to acquire the corresponding threshold, namely, J_{th,T^2}, $J_{th,SPE}$, J_{th,T_c^2}. Thus, the fault detection logic follows

$$Fault\ alarm = \begin{cases} 0, & T^2 \le J_{th,T^2}\ \text{and}\ \mathrm{SPE} \le \mathrm{J_{th,SPE}} \\ 1, & \text{others} \end{cases} \tag{30}$$

or

$$Fault\ alarm = \begin{cases} 0, & T_c^2 \le J_{th,T_c^2} \\ 1, & \text{others} \end{cases} \tag{31}$$

Eventually, the overall procedure of I-MPPCA approach for nonlinear data-driven process monitoring can be summarized in Table I.

Generally, the novel monitoring statistic T_c^2 by (23) as well as (26) and the corresponding threshold are adopted to improve fault detectability, which is considerably simple and effective. As regard to the proposed approach, missing alarm rates (MARs) and false alarm rates (FARs) are mainly considered therein to evaluate the performance. It is expected that two indexes are better to approach zero.

$$\mathrm{MAR} = \frac{\text{number of samples } (J \le J_{th} | f \ne 0)}{\text{total samples } (f \ne 0)} \times 100\% \tag{32}$$

$$\mathrm{FAR} = \frac{\text{number of samples } (J > J_{th} | f = 0)}{\text{total samples } (f = 0)} \times 100\% \tag{33}$$

D. Comparison with other approaches

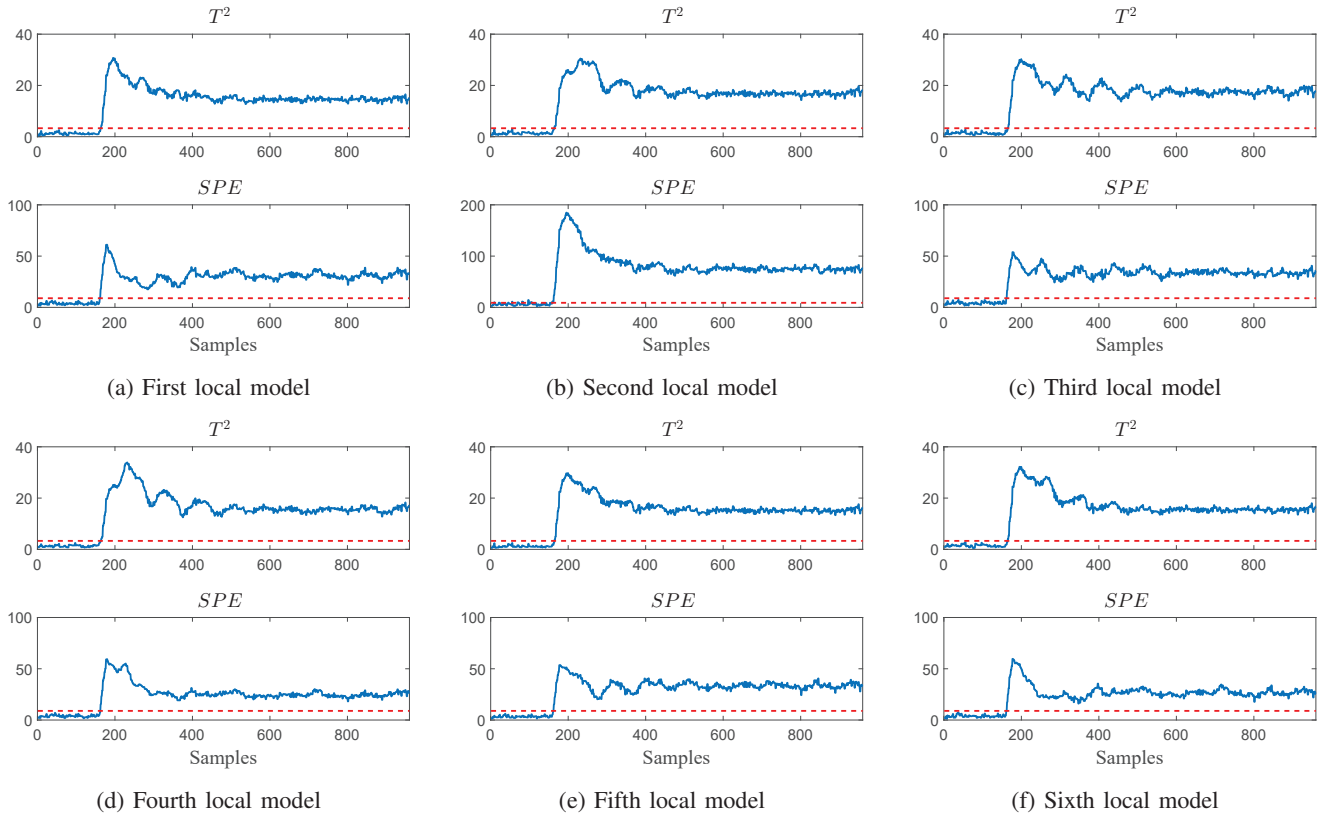
Due to the absence of data labels in most cases, several unsupervised techniques are discussed to illustrate the superior performance of I-MPPCA in this subsection. Since clustering and feature extraction techniques are typical unsupervised schemes, K-means [46, 47], fuzzy-possibilistic c-means (FPCM) [48] and KPCA are expected to compare with the proposed approach. The core of process monitoring is a binary classification issue. K-means and FPCM are popular classification schemes and thus can be used for fault diagnosis purpose. The number of clusters is set to be 2.

Several performance indicators are discussed among these approaches, including computational cost, key parameters, robustness, etc. Notice that the computational complexity of on-line monitoring phase is critical and is valuable because this phase is implemented to monitor real-time operations. Besides, practical industry has high requirement of real-time performance. Therefore, computational cost mainly refers to on-line monitoring procedure in this paper.

K-means is a partition scheme through seeking for certain clustering centers iteratively. It is considerably simple and easy to implement for massive data. K-means is sensitive to outliers since cluster centers are seriously influenced by outliers. Nevertheless, this approach has a high sensitivity of initial cluster centers. Once initial centers are chosen improperly, it is unlikely to acquire effective classification results. The complexity for general clustering problems is $O(kN)$, and k is the number of clusters, with $k = 2$ for fault diagnosis. Thus, the complexity for process monitoring is $O(N)$.

FPCM has more preferable reliability than K-means. The major spirit of FPCM is to acquire the membership vector via minimizing the objective function. The classification consequences are also sensitive to initial cluster centers. FPCM can solve the noise sensitivity of defect [48]. However, it is unable to detect which subspace the fault occurs.

The technical core of KPCA is to map low-dimensional data into high-dimensional linear space and PCA is performed in



(a) First local model (b) Second local model (c) Third local model

(d) Fourth local model (e) Fifth local model (f) Sixth local model

Fig. 1. Monitoring charts of each PPCA model using traditional MPPCA

the high-dimensional feature space. However, KPCA is sensitive to parameter tuning, especially the kernel bandwidth. The computational complexity is $O(N^2)$ owing to the calculation of Gaussian kernels, which makes it inappropriate for real-time process monitoring. In addition, as PCA is insensitive to outliers, this property is inherent in KPCA and I-MPPCA.

I-MPPCA approach partitions the data into several models via the technique of a mixture of probabilistic principal component analysers. It can deal with missing data while the other three approaches are lack of this property [49, 50]. Two key parameters, the number of local models K and the number of principal components (PCs), are simply determined with less computational cost. The computational complexity of the proposed approach is $O(KN)$. In practical applications, it is obvious that $K \ll N$, which indicates that I-MPPCA is less complicated than KPCA and suitable to process large data. Besides, it is insensitive to parameter tuning since K has limited impact on monitoring consequences when K varies in a certain range. Moreover, I-MMPCA can estimate which subspace faults occur by selecting appropriate monitoring statistic T^2 or SPE.

Major characteristics are concluded in Table II, which implies that I-MPPCA is superior to the others to some extent.

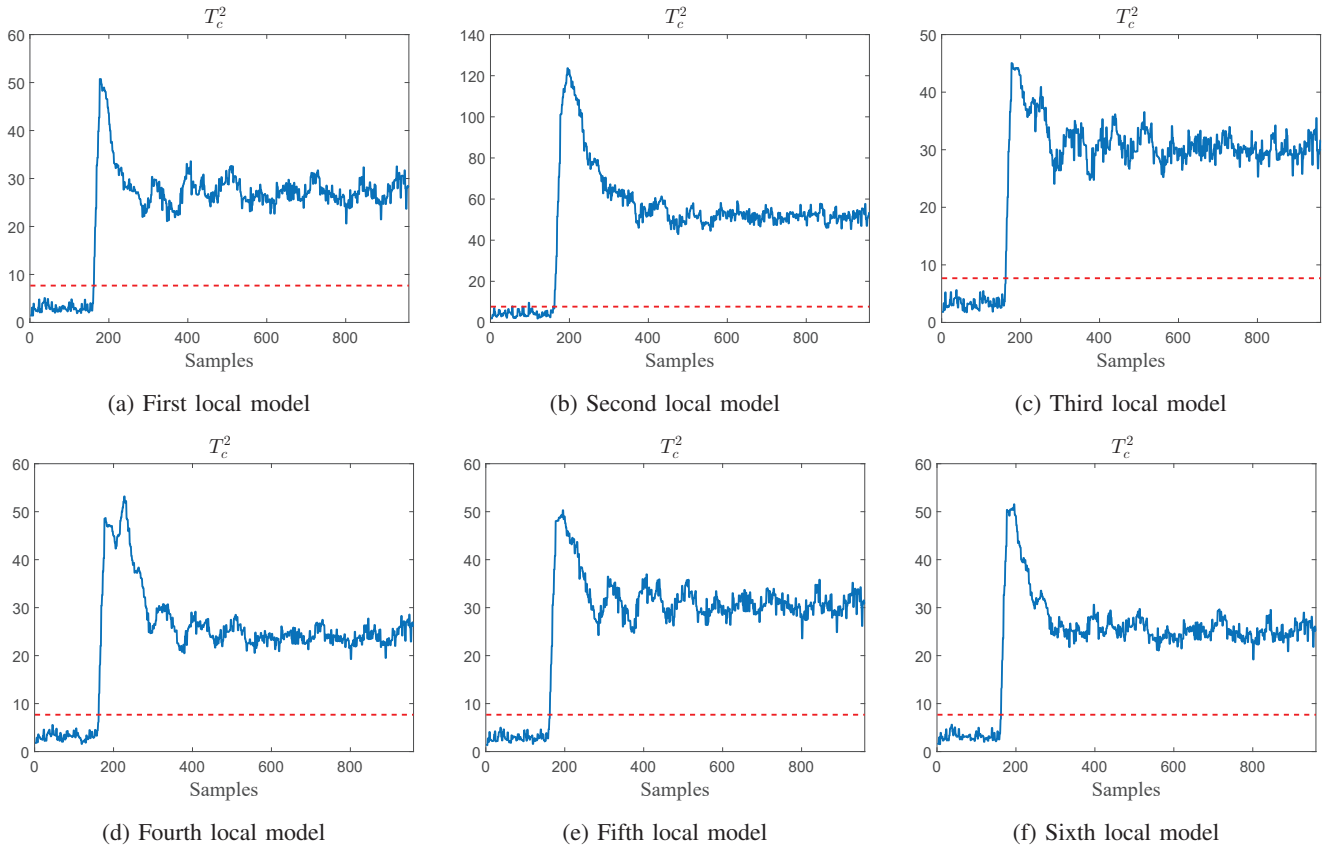
IV. Case Study on TE Process

The proposed I-MPPCA is exemplified through the TE process in this section. Local improved PPCA models are embedded in traditional MPPCA framework and global monitoring statistic is calculated. This case study is utilized to evaluate the rationality and superiority of I-MMPCA approach in comparison with traditional MPPCA approach. The results are obtained through simulations on MATLAB.

TE process model is a realistic chemical plant simulator that serves as a preferred benchmark for monitoring study [51, 52]. Since prior knowledge about the mathematical model of TE process is unavailable, the monitoring approach can be designed only based on sensing measurements. 20 process faults were initially defined and are adopted in this study, namely, IDV(1)-IDV(20). More detailed introduction was described in [53].

In this simulation, 22 control variables and 11 manipulated variables are chosen as the samples. 960 normal samples are utilized to acquire off-line learning model. 960 testing samples, including the first 160 normal samples and 800 subsequent faulty samples, are adopted to evaluate the performance. The confidence level is set to be 0.99. The number of PCs is selected as 6 based on accumulation contribution rate. According to the criterion described by (19) and (20), the number of local PPCA models is 6.

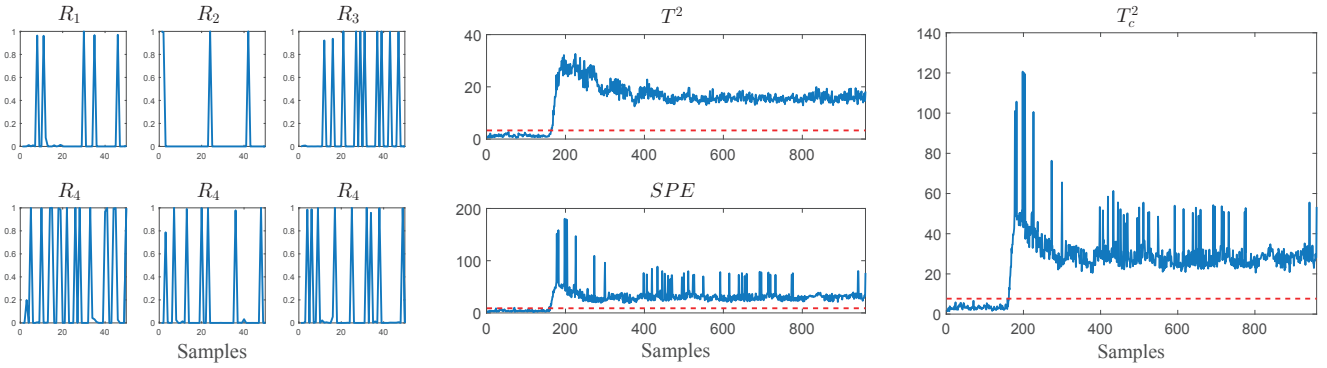
Then, fault IDV(1) is adopted to illustrate the rationality of I-MPPCA approach. Detailed monitoring consequences of traditional MPPCA approach are shown in Fig.1, which are calculated by (21) and (22). With regard to each local improved PPCA model, another monitoring statistic based on the integration of T^2 and SPE is described by (23), as illustrated in Fig. 2. According to Fig. 1 and Fig. 2, fault can be detected timely and accurately by 12 or 6 monitoring



(a) First local model (b) Second local model (c) Third local model

(d) Fourth local model (e) Fifth local model (f) Sixth local model

Fig. 2. Monitoring charts of each local improved PPCA model under traditional MPPCA framework



(a) Weights of local PPCA models (b) Two global monitoring statistics (c) Global combined monitoring statistic

Fig. 3. Monitoring charts using I-MPPCA approach

TABLE III
MARs (%) BASED ON TE DATA GIVEN IN [53]

Fault	IDV(1)	IDV(2)	IDV(3)	IDV(4)	IDV(5)	IDV(6)	IDV(7)	IDV(8)	IDV(9)	IDV(10)
MPPCA (T^2 or SPE)	0.75	1.63	83.88	6.88	7.75	0	0	1.25	84.37	39.12
I-MPPCA (T_c^2)	0	0.63	83.63	0.5	6.13	0	0	0.88	83.75	36.63
Fault	IDV(11)	IDV(12)	IDV(13)	IDV(14)	IDV(15)	IDV(16)	IDV(17)	IDV(18)	IDV(19)	IDV(20)
MPPCA (T^2 or SPE)	53.50	3.50	3.62	21.88	84.25	16.88	36.88	9.00	84.00	18.50
I-MPPCA (T_c^2)	56.87	1.75	3.50	12.13	63.50	16.13	36.00	8.38	87.75	16.75

graphs, respectively.

The monitoring consequences of I-MPPCA are represented in Fig. 3. Posterior probability R is regarded as weight, which measures membership degree a data point belonging to a certain local model of being chosen. In order to observe clearly, weights of the first 50 samples are revealed in Fig. 3a. The weights are almost 1 or 0, which indicates that the samples belong to the certain local PPCA model completely or not, and proves the rationality of the proposed global monitoring statistics. And the monitoring charts of statistics (24)-(25) are demonstrated in Fig. 3b and the statistic (26) is illustrated in Fig. 3c. The global monitoring statistics may change sharply at several points because the weights of the second local model are nearly 1 and the corresponding values of SPE as well as T_c^2 are relatively larger than those of other local models. According to the comparison among Fig. 1, Fig. 2, Fig. 3b and Fig. 3c, the accuracy rates of 4 sorts of calculation approaches are almost close. However, the number of monitoring charts would be reduced from 12 to 1 gradually, which is exactly the major advantage of I-MPPCA approach.

In addition, faults IDV(1)-IDV(20) are employed to demonstrate the superior performance of the global combined statistic in respect of false alarm rates and missing alarm rates. The detailed MARs are listed in Table III. It can be evidently acquired that the MARs of T_c^2 are almost smaller than those of T^2 or SPE. Besides, the FAR of T^2 or SPE is 7.5% while the FAR of T_c^2 is 2.5%. Moreover, the computational complexity is almost similar. Therefore, I-MPPCA approach provides significant improvement compared with traditional MPPCA technique in terms of accuracy and convenience.

V. Case study on Autosuspension benchmrk

In this section, autosuspension benchmark is employed to demonstrate the superiority through comparison with K-means, FPCM and KPCA. The basic description of autosuspension model is present briefly in Subsection V-A. Then, complete data and missing data cases are considered to illustrate the optimal performance of I-MPPCA even when some data values are missing, respectively.

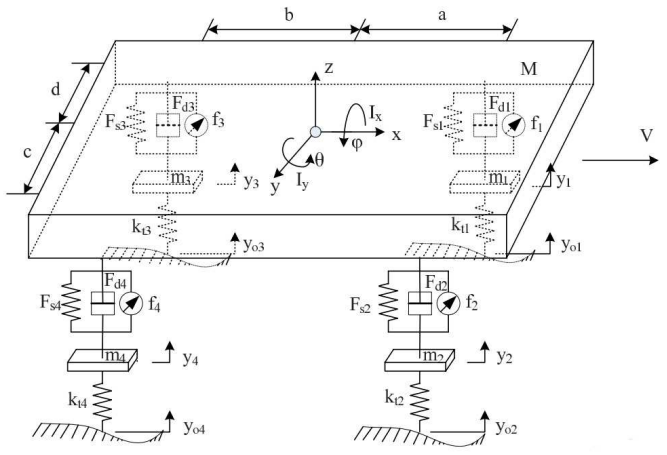


Fig. 4. Flow diagram of autosuspension model

A. Brief description of Autosuspension system

The basic flow diagram of this model is presented in Fig. 4 and a comprehensive introduction was described in [54]. Besides, the parameters are listed in Table IV and specific values can be found in [54].

To our best knowledge, the front wheels (suspension 1 and suspension 2) have the identical configuration and the rear wheels (suspension 3 and suspension 4) share a different one. In other words, researchers just need to study suspension 1 and suspension 3. In practice, most common faults originate from the aging of suspension components, for instance, the parameter reductions from spring and damper. It is difficult to establish specific mathematical model owing to lack of sufficient process knowledge. Therefore, it is essential to implement data-driven techniques for autosuspension monitoring.

Several sorts of sensors are available in industrial applications, e.g., laser sensor, accelerometers, grometer and linear variable displacement transducer. For process monitoring task, only accelerometers are useful and sensing measurements from four accelerometers are adopted in this study. That is, the dimension of data is 4.

B. Simulation with complete data

In this section, suspension coefficient reduction is taken as an example in this paper. These techniques can also be applied to damper coefficient reduction case. 750 normal samples are generated to train I-MPPCA model or KPCA model. The threshold of FPCM should be calculated by normal samples. K-means need not to train model in advance. Then, 750 testing samples are generated as follows:

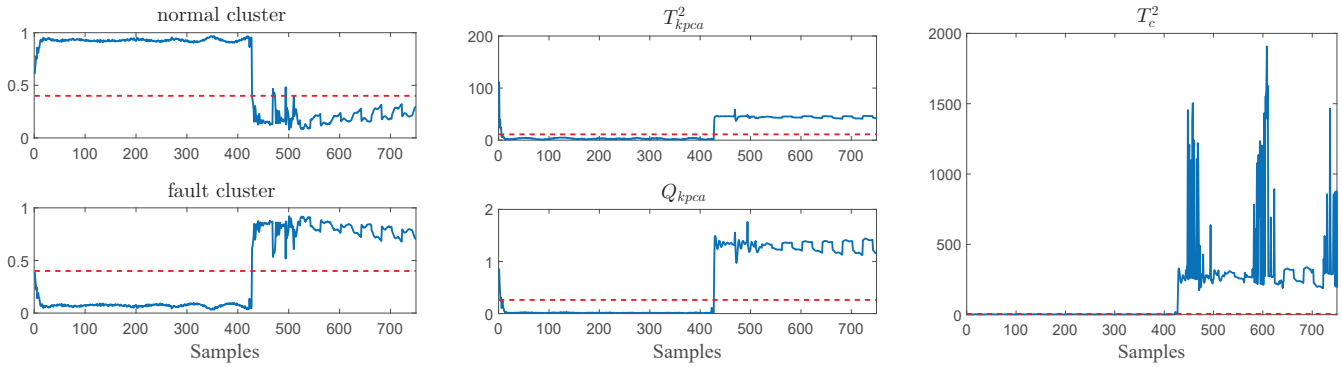
1) Fault 1, the spring coefficient of suspension 1 is reduced by 30% from the 429th sample;

2) Fault 2, the spring coefficient of suspension 3 is reduced by 30% from the 529th sample.

Since K-means just provides two classification labels, only the accuracy rates (MARs and FAR) are given in this paper.

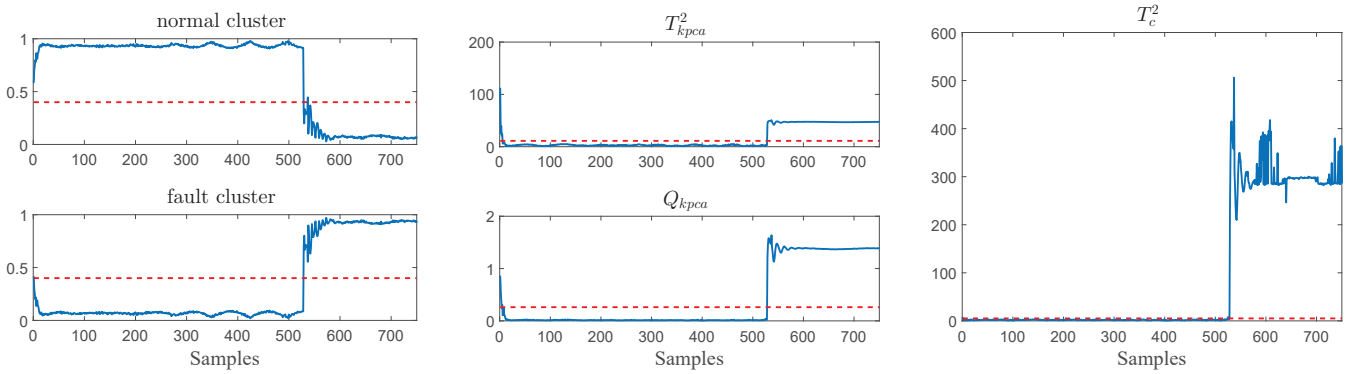
TABLE IV
The parameters of the autosuspension model

Notation	Description
M	Vehicle body mass
m_i	Unsprung mass
θ	Pitch motion
φ	Roll motion
I_x	Roll motion rotary inertias
I_y	Pitch motion rotary inertias
y_i	Unsprung mass displacement
Δy_i	Suspension deflection
$\Delta \dot{y}_i$	Deflection velocity
y_{0i}	Road input
k_i	Linear stiffness parameter
k_{ni}	Nonlinear stiffness parameter
b_{ei}	Extension movement damping parameter
b_{ci}	Compression movement damping parameter
k_{ti}	Stiffness of the tire
F_{si}	Force produced by the spring
F_{di}	Force produced by the damper
f_i	Force produced by the related actuator
a, b, c, d	Distances



(a) FPCM (b) KPCA (c) I-MPPCA

Fig. 5. Monitoring charts of Fault 1



(a) FPCM (b) KPCA (c) I-MPPCA

Fig. 6. Monitoring charts of Fault 2

Therefore, the monitoring charts of the other three approaches are illustrated in Fig. 5 and Fig. 6. It can be evidently observed that FPCM, KPCA and I-MPPCA can detect faults timely.

With regard to FPCM, there would be some misleading places where two membership values are both larger than the predefined threshold. Besides, FPCM can not detect which subspace faults happen. KPCA has two monitoring statistics for residual component subspace and principal component subspace. And the consequences reveal that faults appear in both subspaces. For I-MPPCA approach, it enables providing this information when global T^2 and SPE statistics are selected.

Besides, the MARs and FARs are listed in Table V and Table VI, respectively. For this simulation case, KPCA and I-MPPCA can discover faults completely accurately while K-means and FPCM have several missing alarm points. Moreover, the FAR of I-MPPCA approaches to zero, lower than those of K-means and KPCA. Generally speaking, I-MPPCA delivers optimal monitoring performance among four techniques through the trade-off of FARs and MARs.

In conclusion, I-MPPCA has relatively higher detection accuracy rates based on the analysis above.

C. Simulation with incomplete data

In order to compare conveniently, autosuspension data generated in Subsection V-B with different artificial missing schemes are employed to demonstrate that I-MPPCA can deliver optimal performance when some data values are missing.

Note that missing data values are generated randomly to simulate the practical systems. Besides, with regard to the proposed approach, the tolerant maximal missing rate for modeling data is 30% in this paper and the training accuracy rates basically remain the same with the increasing missing rates of modeling data before the tolerant maximal missing rate. Therefore, 750 training samples with 15% missing data are taken as an example and utilized to establish model in this study. Then, testing samples with missing data values are generated as follows:

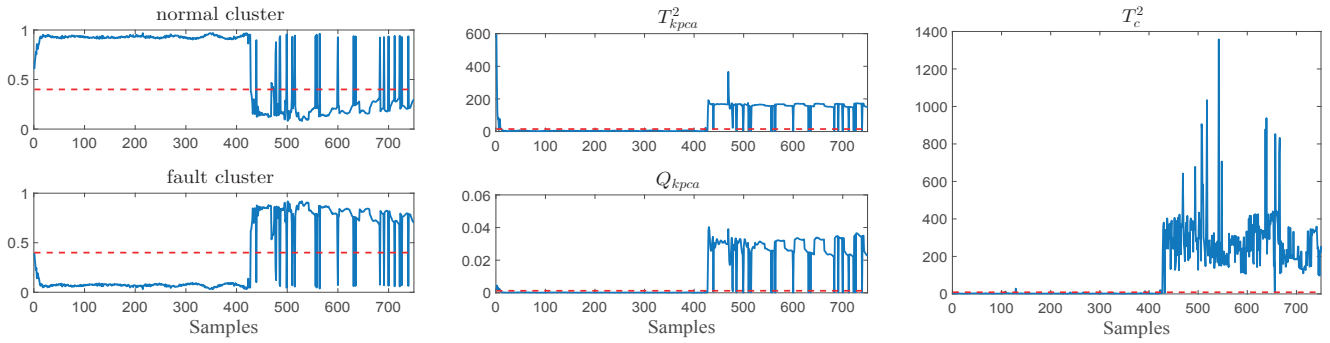
1) Fault 3, the spring coefficient of suspension 1 is reduced by 30% from the 429th sample with 5% missing data;

2) Fault 4, the spring coefficient of suspension 3 is reduced by 30% from the 529th sample with 5% missing data;

3) Fault 5, the spring coefficient of suspension 1 is reduced by 30% from the 429th sample with 10% missing data;

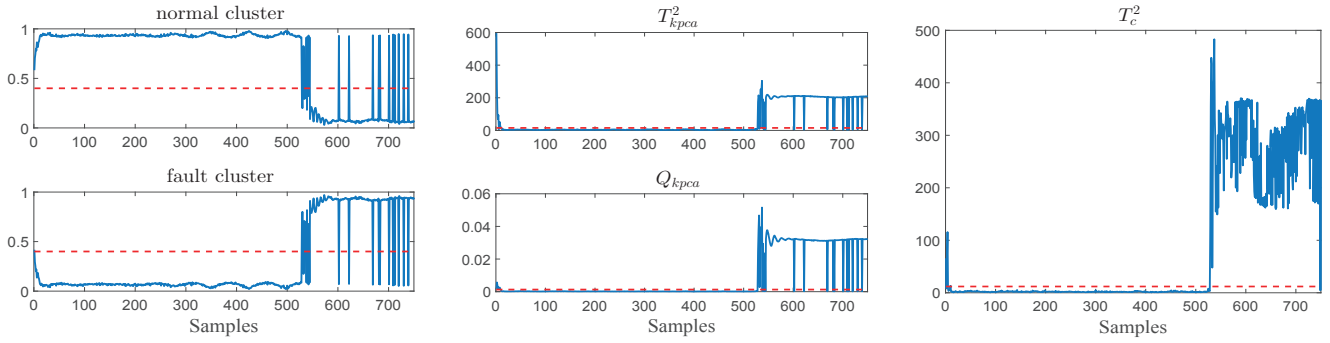
4) Fault 6, the spring coefficient of suspension 3 is reduced by 30% from the 529th sample with 10% missing data.

MARs and FARs are concluded in Table V and Table VI. As to I-MPPCA approach, it can be observed clearly that missing data values have the least influence on monitoring performance and the FAR as well as MARs are basically the lowest. Especially, according to the detection accuracy rates



(a) FPCM (b) KPCA (c) I-MMPCA

Fig. 7. Monitoring charts of Fault 3



(a) FPCM (b) KPCA (c) I-MMPCA

Fig. 8. Monitoring charts of Fault 4

of 6 faults, the accuracy rates will be reduced rapidly with the increasing missing rate.

Detailed monitoring charts of four faults are shown in Figs.7-10. According to the comparison of Fault 1, Fault 3 and Fault 5, Fault 2, Fault 4 and Fault 6, it can be seen that missing rate has important effect on K-means, FPCM and KPCA. Besides, the missing alarm points are exactly the positions of missing samples. And MARs are seriously influenced by which variable occurs missing data. However, the FAR and MARs of I-MPPCA approach are slightly affected by these missing data values.

In conclusion, with regard to missing data, I-MPPCA delivers optimal performance by comparison with three approaches. Furthermore, the proposed approach owns extra virtues, for instance, low computational complexity, insensitivity to parameter tuning, being able to monitor two subspaces, etc. Therefore, I-MPPCA is superior to the others and prior for nonlinear process monitoring both in practical applications and in academic study.

TABLE V
MARs (%) BASED ON AUTOSUSPENSION DATA

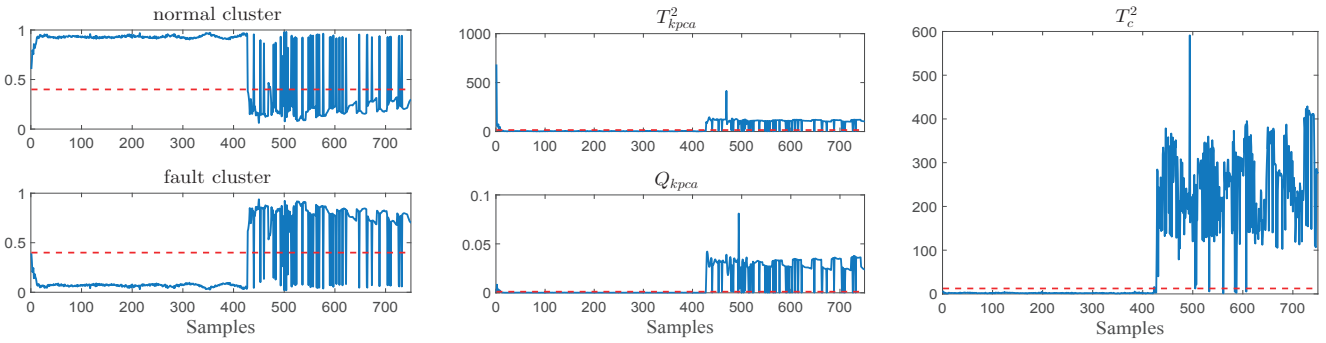
Approach	K-means	FPCM	KPCA	I-MPPCA
Fault 1	2.82	1.41	0	**0**
Fault 2	2.85	1.90	0	**0**
Fault 3	7.14	6.52	6.21	**0.62**
Fault 4	8.11	7.66	6.76	**0.45**
Fault 5	13.66	13.04	12.11	**1.86**
Fault 6	15.77	13.96	13.06	**3.15**

TABLE VI
FARs (%) BASED ON AUTOSUSPENSION DATA

Approach	K-means	FPCM	KPCA	I-MPPCA
Complete data	2.82	0.18	1.41	**0**
Missing data	5.64	1.04	2.57	**0.47**

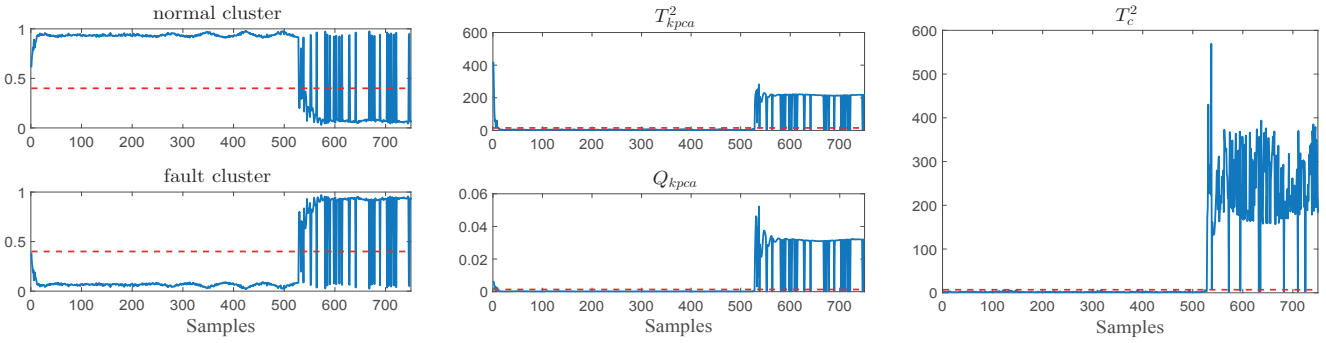
VI. Conclusion

This paper has proposed a novel computing method of monitoring statistics under the framework of traditional MPPCA for nonlinear data-driven process monitoring. Appropriate partitioning of sensing measurements and the parameters of local PPCA models are automatically acquired via the technique of a mixture of probabilistic principal component analysers. Besides, a two-stage EM schedule is employed to improve the convergence speed and reduce computational cost. A novel composite monitoring statistic has been introduced and calculated in each PPCA model aforementioned. It is shown that the posterior probability can be regarded as a weight to data point belonging to a certain PPCA model of being



(a) FPCM (b) KPCA (c) I-MMPCA

Fig. 9. Monitoring charts of Fault 5



(a) FPCM (b) KPCA (c) I-MMPCA

Fig. 10. Monitoring charts of Fault 6

chosen. Therefore, in order to provide an optimal fault detection performance and observational convenience, the global monitoring statistics are acquired based on the weighted mean of all local monitoring statistics. Moreover, several typical unsupervised schemes including feature extraction algorithms and clustering approaches have been discussed to highlight the virtues of the I-MPPCA including low computational cost, parameter robustness, the capability of dealing with incomplete data, etc. Finally, simulation studies in comparison with several known approaches have been carried out based on TE data and an autosuspension model which have demonstrated the superior performance of the proposed approach.

Appendix A
Solution procedure of MPPCA

According to (15), the log-likelihood of observation data for a mixture model can be depicted as:

$$\begin{aligned} L &= \sum_{n=1}^{N} \ln \{p(\boldsymbol{t}_n)\} \\ &= \sum_{n=1}^{N} \ln \left\{ \sum_{i=1}^{K} \pi_i p(\boldsymbol{t}_n|i) \right\} \end{aligned} \tag{34}$$

The maximum-likelihood is utilized to determine key parameters of the model, where the proper segmentation of the data occurs automatically when the log-likelihood reaches its maximum. An iterative EM algorithm is developed to optimize the model parameters π_i, $\boldsymbol{\mu}_i$, $\boldsymbol{W}_i$ and σ_i^2, which was first introduced in [55].

Suppose that $R_{ni} = p(i|\boldsymbol{t}_n)$ is the posterior probability of the ith local model for generating data $\boldsymbol{t}_n$, it can be estimated using Bayesian rule

$$R_{ni} = \frac{p(\boldsymbol{t}_n|i)\,\pi_i}{p(\boldsymbol{t}_n)} \tag{35}$$

With regard to this posterior distribution, the expectation of L_C can be acquired in the form of

$$\begin{aligned} \langle L_c \rangle = \sum_{n=1}^{N}\sum_{i=1}^{K} R_{ni} \Bigg\{ & \ln \pi_i - \frac{d}{2}\ln \sigma_i^2 - \frac{1}{2} tr \left\{ \left\langle \boldsymbol{x}_n^{(i)} \boldsymbol{x}_n^{(i)\mathrm{T}} \right\rangle \right\} \\ & - \frac{1}{2\sigma_i^2} \|\boldsymbol{t}_n - \boldsymbol{\mu}_i\|^2 + \frac{1}{\sigma_i^2} \left\langle \boldsymbol{x}_n^{(i)} \right\rangle^{\mathrm{T}} \boldsymbol{W}_i^{\mathrm{T}} (\boldsymbol{t}_n - \boldsymbol{\mu}_i) \\ & - \frac{1}{2\sigma_i^2} tr \left\{ \boldsymbol{W}_i^{\mathrm{T}} \boldsymbol{W}_i \left\langle \boldsymbol{x}_n^{(i)} \boldsymbol{x}_n^{(i)\mathrm{T}} \right\rangle \right\} \Bigg\} \end{aligned} \tag{36}$$

In order to obtain the optimal values of key model parameters aforementioned, a Lagrange multiplier λ is utilized to achieve the maximum value of (36). Thus, the solution of

maximum likelihood can be transformed into the following optimization problem

$$\begin{cases} \max & \langle L_C \rangle + \lambda \left(\sum_{i=1}^{K} \pi_i - 1 \right) \\ s.t. & \sum \pi_i = 1 \end{cases} \tag{37}$$

To our best knowledge, traditional EM algorithm is considerably complicated due to iterative convergence process. In this paper, a two-stage EM schedule is adopted, where generalized EM (GEM) is utilized in M-step to improve convergence speed and reduce computational complexity [56]. The two-stage EM algorithm for MPPCA is described in detail in Appendix B.

Appendix B
A two-stage EM for MPPCA

The log-likelihood function we expect to maximize is described as the likelihood (34).

The relevant expected complete-data log-likelihood is interpreted as

$$\widehat{L}_C = \sum_{n=1}^{N} \sum_{i=1}^{K} R_{ni} \ln \left\{ \pi_i p(\boldsymbol{t}_n, \boldsymbol{x}_n^{(i)}) \right\} \tag{38}$$

where R_{ni} is calculated by (35). The first stage of the two-stage EM schedule (E-step) is maximizing (38) to acquire $\widetilde{\boldsymbol{\mu}}_i$ and $\widetilde{\pi}_i$.

The second stage (M-step) takes advantage of generalised EM (or GEM) to update $\boldsymbol{W}_i$ and σ_i^2. The typical feature of GEM is to increase the value of $\widehat{L}_C$ and not to maximize it during the iteration process. Regarding $\widehat{L}_C$ as the likelihood of interest, one cycle of EM is performed about $\boldsymbol{W}_i$ and σ_i^2.

M-step procedure can be simplified further when (36) is expanded for $\left\langle \boldsymbol{x}_n^{(i)} \right\rangle$ and $\left\langle \boldsymbol{x}_n^{(i)} \boldsymbol{x}_n^{(i)\mathrm{T}} \right\rangle$, only terms in $\widetilde{\boldsymbol{\mu}}_i$ appear. Thus, the expected complete-data log-likelihood now can be obtained by inspection of (36) as follows

$$\begin{aligned} \langle L_c \rangle = \sum_{n=1}^{N} \sum_{i=1}^{K} R_{ni} \Bigg\{ & \ln \widetilde{\pi}_i - \frac{d}{2} \ln \sigma_i^2 - \frac{1}{2} tr \left\{ \left\langle \boldsymbol{x}_n^{(i)} \boldsymbol{x}_n^{(i)\mathrm{T}} \right\rangle \right\} \\ & - \frac{1}{2\sigma_i^2} \| \boldsymbol{t}_n - \widetilde{\boldsymbol{\mu}}_i \|^2 + \frac{1}{\sigma_i^2} \left\langle \boldsymbol{x}_n^{(i)} \right\rangle^{\mathrm{T}} \boldsymbol{W}_i^{\mathrm{T}} (\boldsymbol{t}_n - \widetilde{\boldsymbol{\mu}}_i) \\ & - \frac{1}{2\sigma_i^2} tr \left\{ \boldsymbol{W}_i^{\mathrm{T}} \boldsymbol{W}_i \left\langle \boldsymbol{x}_n^{(i)} \boldsymbol{x}_n^{(i)\mathrm{T}} \right\rangle \right\} \Bigg\} \end{aligned} \tag{39}$$

Much simplified M-step formulas can be acquired when (39) reaches the maximum with respects to $\boldsymbol{W}_i$ and σ_i^2 (keeping $\widetilde{\boldsymbol{\mu}}_i$ fixed)

$$\widetilde{\boldsymbol{W}}_i = \boldsymbol{S}_i \boldsymbol{W}_i \left({\sigma_i}^2 \boldsymbol{I} + \boldsymbol{M}_i^{-1} \boldsymbol{W}_i^{\mathrm{T}} \boldsymbol{S}_i \boldsymbol{W}_i \right)^{-1} \tag{40}$$

$$\widetilde{\sigma_i}^2 = \frac{1}{d} tr \left(\boldsymbol{S}_i - \boldsymbol{S}_i \boldsymbol{W}_i \boldsymbol{M}_i^{-1} \widetilde{\boldsymbol{W}}_i^{\mathrm{T}} \right) \tag{41}$$

where

$$\boldsymbol{S}_i = \frac{1}{\widetilde{\pi}_i N} \sum_{n=1}^{N} R_{ni} (\boldsymbol{t}_n - \widetilde{\boldsymbol{\mu}}_i)(\boldsymbol{t}_n - \widetilde{\boldsymbol{\mu}}_i)^{\mathrm{T}} \tag{42}$$

$$\widetilde{\boldsymbol{\mu}}_i = \frac{\sum_{n=1}^{N} R_{ni} \boldsymbol{t}_n}{\sum_{n=1}^{N} R_{ni}} \tag{43}$$

$$\widetilde{\pi}_i = \frac{1}{N} \sum_{n=1}^{N} R_{ni} \tag{44}$$

Obviously, the symbol $\widetilde{}$ indicates new variables that may be updated in the M-step. Iteration of (42)-(44) as well as (35) followed by (40) and (41) in turns is ensured to reach a local maximum of the likelihood (34).

Acknowledgment

This work was supported by the Chunmiao Project of Haixi Institute of Chinese Academy of Sciences (Program No.CMZX-2016-005), National Natural Science Foundation of China (No.61603369) and National Natural Science Foundation of China (No.61703388).

References

[1] R. H. Kwong and D. L. Yonge-Mallo, "Fault diagnosis in discrete-event systems with incomplete models: learnability and diagnosability," *IEEE Trans. Cybern.*, vol. 45, no. 7, pp. 1236–1249, Jul. 2015.

[2] E. Alizadeh, N. Meskin, and K. Khorasani, "A negative selection immune system inspired methodology for fault diagnosis of wind turbines," *IEEE Trans. Cybern.*, vol. 47, no. 11, pp. 3799–3813, Nov. 2017.

[3] Y.-L. Wang, C.-C. Lim, and P. Shi, "Adaptively adjusted event-triggering mechanism on fault detection for networked control systems," *IEEE Trans. Cybern.*, vol. 47, no. 8, pp. 2299–2311, Aug. 2017.

[4] J. Han, H. Zhang, Y. Wang, and X. Sun, "Robust fault detection for switched fuzzy systems with unknown input," *IEEE Trans. Cybern.*, vol. DOI: 10.1109/TCYB.2017.2755864, 2017.

[5] C. Mu, Z. Ni, C. Sun, and H. He, "Data-driven tracking control with adaptive dynamic programming for a class of continuous-time nonlinear systems," *IEEE Trans. Cybern.*, vol. 47, no. 6, pp. 1460–1470, Jun. 2017.

[6] T. Li, B. Ni, M. Xu, M. Wang, Q. Gao, and S. Yan, "Data-driven affective filtering for images and videos," *IEEE Trans. Cybern.*, vol. 45, no. 10, pp. 2336–2349, Oct. 2015.

[7] Z. Hou, R. Chi, and H. Gao, "An overview of dynamic-linearization-based data-driven control and applications," *IEEE Trans. Ind. Electron.*, vol. 64, no. 5, pp. 4076–4090, May 2017.

[8] Y.-L. Wang, P. Shi, C.-C. Lim, and Y. Liu, "Event-triggered fault detection filter design for a continuous-time networked control system," *IEEE Trans. Cybern.*, vol. 46, no. 12, pp. 3414–3426, Dec. 2016.

[9] L. Li, S. X. Ding, J. Qiu, and Y. Yang, "Real-time fault detection approach for nonlinear systems and its asynchronous T-S fuzzy observer-based implementation," *IEEE Trans. Cybern.*, vol. 47, pp. 283–294, Feb. 2017.

[10] S. Yin, H. Gao, J. Qiu, and O. Kaynak, "Fault detection for nonlinear process with deterministic disturbances: a just-in-time learning based data driven method," *IEEE Trans. Cybern.*, vol. 47, no. 11, pp. 3649–3657, Nov. 2017.

[11] S. Yin, S. X. Ding, X. Xie, and H. Luo, "A review on basic data-driven approaches for industrial process monitoring," *IEEE Trans. Ind. Electron.*, vol. 61, no. 11, pp. 6418–6428, Nov. 2014.

[12] S. X. Ding, *Data-driven design of fault diagnosis and fault-tolerant control systems*, ser. Advances in Industrial Control. Springer London, 2014.

[13] S. Yin, S. X. Ding, A. Haghani, H. Hao, and P. Zhang, "A comparison study of basic data-driven fault diagnosis and process monitoring methods on the benchmark Tennessee Eastman process," *J. Process Control*, vol. 22, no. 9, pp. 1567–1581, Oct. 2012.

[14] L. P. Queiroz, F. C. M. Rodrigues, J. P. P. Gomes, F. T. Brito, I. C. Chaves, M. R. P. Paula, M. R. Salvador, and J. C. Machado, "A fault detection method for hard disk drives based on mixture of Gaussians and non-parametric statistics," *IEEE Trans. Ind. Informat.*, vol. 13, no. 2, pp. 542–550, Apr. 2017.

[15] M. Kristan and A. Leonardis, "Online discriminative kernel density estimator with Gaussian kernels," *IEEE Trans. Cybern.*, vol. 44, no. 3, pp. 355–365, Mar. 2014.

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

[17] X. Lu, C. Liu, and M. Huang, "Online probabilistic extreme learning machine for distribution modeling of complex batch forging processes," *IEEE Trans. Ind. Informat.*, vol. 11, no. 6, pp. 1277–1286, Dec. 2015.

[18] H. Wang, M. Chen, X. Shi, and N. Li, "Principal component analysis for normal-distribution-valued symbolic data," *IEEE Trans. Cybern.*, vol. 46, no. 2, pp. 356–365, Feb. 2016.

[19] Y. Yin, D. Xu, X. Wang, and M. Bai, "Online state-based structured SVM combined with incremental PCA for robust visual tracking," *IEEE Trans. Cybern.*, vol. 45, no. 9, pp. 1988–2000, Sep. 2015.

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

[21] D. You, X. Gao, and S. Katayama, "WPD-PCA-based laser welding process monitoring and defects diagnosis by using FNN and SVM," *IEEE Trans. Ind. Electron.*, vol. 62, no. 1, pp. 628–636, Jan. 2015.

[22] K. Honda and H. Ichihashi, "Regularized linear fuzzy clustering and probabilistic PCA mixture models," *IEEE Trans. Fuzzy Syst.*, vol. 13, no. 4, pp. 508–516, Aug. 2005.

[23] W. Li, H. H. Yue, S. Valle-Cervantes, and S. J. Qin, "Recursive PCA for adaptive process monitoring," *J. Process Control*, vol. 10, no. 5, pp. 471–486, Oct. 2000.

[24] T. J. Rato, J. Blue, J. Pinaton, and M. S. Reis, "Translation-invariant multiscale energy-based PCA for monitoring batch processes in semiconductor manufacturing," *IEEE Trans. Autom. Sci. Eng.*, vol. 14, no. 2, pp. 894–904, Apr. 2017.

[25] J. Zhu, Z. Ge, and Z. Song, "Distributed parallel PCA for modeling and monitoring of large-scale plant-wide processes with big data," *IEEE Trans. Ind. Informat.*, vol. 13, no. 4, pp. 1877–1885, Aug. 2017.

[26] X. B. He and Y. P. Yang, "Variable MWPCA for adaptive process monitoring," *Ind. Eng. Chem. Res.*, vol. 47, no. 2, pp. 419–427, Jan. 2008.

[27] S. Yin, X. Xie, J. Lam, K. C. Cheung, and H. Gao, "An improved incremental learning approach for KPI prognosis of dynamic fuel cell system," *IEEE Trans. Cybern.*, vol. 46, no. 12, pp. 3135–3144, Dec. 2016.

[28] S. Yin, C. Yang, J. Zhang, and Y. Jiang, "A data-driven learning approach for nonlinear process monitoring based on available sensing measurements," *IEEE Trans. Ind. Electron.*, vol. 64, no. 1, pp. 643–653, Jan. 2017.

[29] C. Cheng and M.-S. Chiu, "Nonlinear process monitoring using JITL-PCA," *Chemom. Intell. Lab. Syst.*, vol. 76, no. 1, pp. 1–13, Mar. 2005.

[30] J. Chen and C.-M. Liao, "Dynamic process fault monitoring based on neural network and PCA," *J. Process Control*, vol. 12, no. 2, pp. 277–289, Feb. 2002.

[31] L. Cai, X. Tian, and S. Chen, "Monitoring nonlinear and non-Gaussian processes using Gaussian mixture model-based weighted kernel independent component analysis," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 28, no. 1, pp. 122–135, Jan. 2017.

[32] J. Ni, C. Zhang, and S. X. Yang, "An adaptive approach based on KPCA and SVM for real-time fault diagnosis of HVCBs," *IEEE Trans. Power Del.*, vol. 26, no. 3, pp. 1960–1971, Jul. 2011.

[33] C. F. Alcala and S. J. Qin, "Reconstruction-based contribution for process monitoring with kernel principal component analysis," *Ind. Eng. Chem. Res.*, vol. 49, no. 17, pp. 7849–7857, Sep. 2010.

[34] Z. Ge, C. Yang, and Z. Song, "Improved kernel PCA-based monitoring approach for nonlinear processes," *Chem. Eng. Sci.*, vol. 64, no. 9, pp. 2245–2255, May 2009.

[35] X. Liu, K. Li, M. McAfee, and G. W. Irwin, "Improved nonlinear PCA for process monitoring using support vector data description," *J. Process Control*, vol. 21, no. 9, pp. 1306–1317, Oct. 2011.

[36] D. Iacoviello, A. Petracca, M. Spezialetti, and G. Placidi, "A classification algorithm for electroencephalography signals by self-induced emotional stimuli," *IEEE Trans. Cybern.*, vol. 46, no. 12, pp. 3171–3180, Dec. 2016.

[37] M. E. Tipping and C. M. Bishop, "Probabilistic principal component analysis," *J. Roy. Statist. Soc. Ser. B*, vol. 61, no. 3, pp. 611–622, 1999.

[38] H. Lu, K. N. Plataniotis, and A. N. Venetsanopoulos, "MPCA: Multilinear principal component analysis of tensor objects," *IEEE Trans. Neural Netw.*, vol. 19, no. 1, pp. 18–39, Jan. 2008.

[39] M. E. Tipping and C. M. Bishop, "Mixtures of probabilistic principal component analysers," *Neural Comput.*, vol. 11, no. 2, pp. 443–482, 1999.
[40] J. Zhao, "Efficient model selection for mixtures of probabilistic PCA via hierarchical BIC," *IEEE Trans. Cybern.*, vol. 44, no. 10, pp. 1871–1883, Oct. 2014.
[41] R. Sharifi and R. Langari, "Nonlinear sensor fault diagnosis using mixture of probabilistic PCA models," *Mech. Syst. Signal Process.*, vol. 85, pp. 638–650, Feb. 2017.
[42] Z. Ge and Z. Song, "Mixture Bayesian regularization method of PPCA for multimode process monitoring," *AIChE J.*, vol. 56, no. 11, pp. 2838–2849, Nov. 2010.
[43] B. Thompson, *Exploratory and confirmatory factor analysis: Understanding concepts and applications*. American Psychological Association, 2004.
[44] L. Xu, "Bayesian Ying–Yang machine, clustering and number of clusters," *Pattern Recognit. Lett.*, vol. 18, no. 11–13, pp. 1167–1178, Nov. 1997.
[45] P.-E. P. Odiowei and Y. Cao, "Nonlinear dynamic process monitoring using canonical variate analysis and kernel density estimations," *IEEE Trans. Ind. Informat.*, vol. 6, no. 1, pp. 36–45, Feb. 2010.
[46] C. Yiakopoulos, K. C. Gryllias, and I. A. Antoniadis, "Rolling element bearing fault detection in industrial environments based on a K-means clustering approach," *Expert Syst. Appl.*, vol. 38, no. 3, pp. 2888–2911, Mar. 2011.
[47] C. Boutsidis and M. Magdon-Ismail, "Deterministic feature selection for K-means clustering," *IEEE Trans. Inf. Theory*, vol. 59, no. 9, pp. 6099–6110, Sep. 2013.
[48] R. Khanchana and M. Punithavalli, "Web usage mining for predicting users' browsing behaviors by using FPCM clustering," *Int. J. Eng. Tech.*, vol. 3, no. 5, pp. 491–496, Oct. 2011.
[49] K. Wagstaff, "Clustering with missing values: No imputation required," *Classification, Clustering, and Data Mining Applications*, pp. 649–658, 2004.
[50] H. Xiao, D. Huang, Y. Pan, Y. Liu, and K. Song, "Fault diagnosis and prognosis of wastewater processes with incomplete data by the auto-associative neural networks and ARMA model," *Chemom. Intell. Lab. Syst.*, vol. 161, pp. 96–107, Feb. 2017.
[51] G. Wang and S. Yin, "Quality-related fault detection approach based on orthogonal signal correction and modified PLS," *IEEE Trans. Ind. Informat.*, vol. 11, no. 2, pp. 398–405, Apr. 2015.
[52] X. Xie, W. Sun, and K. C. Cheung, "An advanced PLS approach for key performance indicator-related prediction and diagnosis in case of outliers," *IEEE Trans. Ind. Electron.*, vol. 63, no. 4, pp. 2587–2594, Apr. 2016.
[53] J. J. Downs and E. F. Vogel, "A plant-wide industrial process control problem," *Comput. Chem. Eng.*, vol. 17, no. 3, pp. 245–255, Mar. 1993.
[54] G. Wang and S. Yin, "Data-driven fault diagnosis for an automobile suspension system by using a clustering based method," *J. Franklin Inst.*, vol. 351, no. 6, pp. 3231–3244, Jun. 2014.
[55] A. P. Dempster, N. M. Laird, and D. B. Rubin, "Maximum-likelihood from incomplete data via the EM algorithm," *J. Roy. Statist. Soc. Ser. B*, vol. 39, pp. 1–38, 1977.
[56] O. Onur, T. Wimalajeewa, and B. Dulek, "Asynchronous linear modulation classification with multiple sensors via generalized EM algorithm," *IEEE Trans. Wireless Commun.*, vol. 14, no. 11, pp. 6389–6400, Nov. 2015.

Jingxin Zhang received her B.E. degree in electrical engineering and automation from Harbin Engineering University, Harbin, China, the M.E. degree in control science and engineering from Harbin Institute of Technology, Harbin, China, in 2014 and 2016, respectively. She is currently an Assistant Engineer with Quanzhou Institute of Equipment Manufacturing, Haixi Institutes, Chinese Academy of Sciences, Jinjiang, China.

Her research interests include data-driven fault detection and diagnosis, fault prognosis, performance monitoring, dimensionality reduction and their applications in the industrial process.

Hao Chen received the B.Eng. degree in automatic control from the National University of Defense Technology, Changsha, China, in 2006, the M.Sc. degree in control systems with distinction from the University of Sheffield, Sheffield, U.K., in 2009, and the Ph.D. degree in cybernetics from the School of Systems Engineering, University of Reading, Reading, U.K., in 2014, sponsored by the Engineering and Physical Sciences Research Council and Defence Science and Technology Laboratory from the British Government.

From 2014 to 2015, he was a Post-Doctoral Research Fellow with the Department of Chemical and Materials Engineering, University of Alberta, Edmonton, AB, Canada, and Syncrude Canada Ltd., Fort McMurray, AB, Canada. He is currently an Associate Professor with Haixi Institutes, Chinese Academy of Sciences, Jinjiang, China. His current research interest include online learning, soft sensors, system identification, neural networks, machine learning, signal processing, big data analysis and their applications in the industrial process.

Dr. Chen was a recipient of the Chinese Government Award for Outstanding Self-financed Students Abroad in 2012.

Songhang Chen received the Ph.D. degree in control theory and control engineering from the Institute of Automation, Chinese Academy of Sciences, Beijing, China, in 2014. From 2014 to 2016, he worked as an Assistant Researcher with the State Key Laboratory for Management and Control of Complex Systems, China. He is currently an Associate Professor with Quanzhou Institute of Equipment Manufacturing, Haixi Institutes, Chinese Academy of Sciences, Jinjiang, China. His research interests include multiobjective optimization, high-performance computing, big data, and Internet of things.

Xia Hong received the B.Sc. and M.Sc. degrees from the National University of Defense Technology, China, in 1984 and 1987, respectively, and the Ph.D. degree from The University of Sheffield, U.K., in 1998, all in automatic control. She was a Research Assistant with the Beijing Institute of Systems Engineering, Beijing, China, from 1987 to 1993. She was a Research Fellow with the Department of Electronics and Computer Science, University of Southampton, from 1997 to 2001.

She is currently a Professor with the Department of Computer Science, School of Mathematical, Physical and Computational Sciences, University of Reading. She is actively involved in research into nonlinear systems identification, data modelling, estimation and intelligent control, neural networks, pattern recognition, learning theory, and their applications. She has authored over 170 research papers, and co-authored a research book. Dr. Hong received the Donald Julius Groen Prize from IMechE in 1999.