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 analyzers 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 utilized as a metrics to detect potential abnormalities. The virtues of the proposed algorithm are 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, mixture of probabilistic principal component analysis (MPPCA), nonlinear systems, process monitoring.

I. INTRODUCTION

F OR 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]. Ding [12] has presented a comprehensive introduction on data-driven design of fault diagnosis and summarized several multivariate analysis techniques to fault diagnosis, i.e., principal component analysis (PCA), dynamic PCA, partial least squares, canonical variate analysis, etc. Besides, benchmark applications such as three-tank system, continuous stirred tank heater, and Tennessee Eastman (TE) process are utilized to demonstrate the effectiveness of these approaches. Yin et al. [13] also summarized a variety of data-driven techniques for multivariate statistical process monitoring (MSPM), and analyzed their computational complexities as well as data assumption, including independent component analysis, Fisher discriminant analysis, 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].

Among 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], and 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 is a nonlinear regression method [27], where PCA-based process monitoring model can be computed under the locally weighted framework [28]. In [29] and [30], just-in-time learning or neural network 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 models 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] and [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 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 analyzers, 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 PCA model can be interpreted by the posterior 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 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, 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:

$$ t = y(x; w) + \xi $$

(1)

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

$$ t = Wx + \mu + \xi. $$

(2)

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

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

$$ T = WX + u1^T + \Xi $$

(3)

where \( T = [t_1, \ldots, t_N] \in \mathbb{R}^{d \times N} \), \( X = [x_1, \ldots, x_N] \in \mathbb{R}^{q \times N} \), and \( \Xi = [\xi_1, \ldots, \xi_N] \in \mathbb{R}^{d \times N} \). \( 1 \) denotes the vector of all ones with appropriate dimension. Let \( u = (1/N) \sum_{n=1}^{N} t_n \), and the sample covariance matrix be denoted by \( S = (1/N)(T - u1^T)(T - u1^T)^T \). 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 \( S = \tilde{W} \Lambda \tilde{W}^T \), where \( \Lambda = \text{diag}(\lambda_1, \ldots, \lambda_q, \lambda_{q+1}, \ldots, \lambda_d) \), \( \lambda_1 \geq \lambda_2 \geq \cdots > \lambda_q > 0 \) are nonzero eigenvalues \( S \), the \( d - q \) smallest eigenvalues are minor and negligible. Specifically we find \( q \)-dimensional vectors

$$ x_n = W^T(t_n - \mu) $$

(4)

to represent \( t_n, n = 1, \ldots, N \), where \( W \) is the first \( q \) columns of \( \tilde{W} \) and \( q \) is determined by cumulative percent variance (CPV) approach. It can be shown that projection onto most dominant eigenvectors leads to finding lower dimensional latent variables while retaining maximal variance of the original variables \( t_n \).
B. Probabilistic PCA

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

\[
p(t|x) = \left(2\pi\sigma^2\right)^{-d/2} \exp\left\{- \frac{1}{2\sigma^2} \|t - Wx - \mu\|^2 \right\}.
\]  

A Gaussian prior probability over \( x \) can be defined by

\[
p(x) = (2\pi)^{-q/2} \exp\left\{- \frac{1}{2} x^T x \right\}.
\]  

Then, the marginal distribution of \( t \) can be acquired in the form of

\[
p(t) = \int p(t|x)p(x)dx = (2\pi)^{-d/2} |C|^{-1/2} \exp\left\{- \frac{1}{2} (t - \mu)^T C^{-1} (t - \mu) \right\}
\]  

where \(|.|\) denotes matrix determinant. The model covariance is given by

\[
C = \sigma^2 I + WW^T.
\]  

In accordance with Bayesian theory, given the observation vector \( t \), the corresponding posterior distribution of \( x \) may be calculated

\[
p(x|t) = \exp\left\{- \frac{1}{2} \left\{ x - M^{-1} W^T (t - \mu) \right\}^T \left( \sigma^{-2} M \right) \left\{ x - M^{-1} W^T (t - \mu) \right\} \right\} \times (2\pi)^{-q/2} |\sigma^{-2} M|^{1/2}
\]  

where the posterior covariance satisfies

\[
\sigma^2 M^{-1} = \sigma^2 \left( \sigma^2 I + WW^T \right)^{-1}
\]  

with \( M \in \mathbb{R}^{q \times d}, C \in \mathbb{R}^{d \times d} \).

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

\[
L = \sum_{n=1}^{N} \ln|p(t_n)| = -N/2 \left( d\ln(2\pi) + \ln|C| + \text{tr}(C^{-1}S) \right).
\]  

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

\[
(x_n) = M^{-1} W^T (t_n - \mu) \quad (13)
\]

\[
(x_n x_n^T) = \sigma^2 M^{-1} + (x_n)(x_n)^T \quad (14)
\]

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

C. Mixture of Probabilistic PCA

In order to be able to model more complex data, the 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

\[
p(t) = \sum_{i=1}^{K} \pi_i p(t|i) = \sum_{i=1}^{K} \pi_i p(t|i)
\]  

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

\[
t = W_i x + \mu_i + \xi_i, \quad i = 1, \ldots, K
\]  

which is similar to (2), and has individual projection matrix \( W_i \), mean vector \( \mu_i \), as well as \( \xi_i \sim N(0, \sigma_i^2 I) \).

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

\[
(x_n^{(i)}) = M_i^{-1} W_i^T (t_n - \mu_i) \quad (17)
\]

\[
(x_n^{(i)} x_n^{(i)T}) = \sigma_i^2 M_i^{-1} + (x_n^{(i)})(x_n^{(i)})^T \quad (18)
\]

and are given here for convenience. Note that \( x_n^{(i)} \) is \( n \)th sample for each \( i \)th 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. 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 \( T^2 \)-squared 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, 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, $T_i^2$ 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, $K^o$, can be determined by the following criteria [44]:

$$K^o = \arg \min_i H(i)$$

$$H(i) = \frac{1}{N} \sum_{n=1}^{N} \sum_{i=1}^{K} p(i|t_i, \theta) \ln(p(t_i|i)) - \sum_{i=1}^{K} \pi_i \ln \pi_i$$

where $\theta$ is the set of all model parameter vectors, containing $\theta = \{W_i, \mu_i, \sigma_i^2, \pi_i\}_{i=1,...,K}$.

Given a value of $K$, $H(i)$ is calculated after the implementation of two-stage EM schedule as the model parameter $\theta$ 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 $|H(i) - H(i+1)| > \delta$, where $\delta$ is a predefined threshold.

Note that the maximum number of local models $K_{\text{max}}$ varies relying on data pattern. However, $K_{\text{max}} = 5 \sim 10$ 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 $i$th local model, $T_i^2$ and SPE$_i$ are monitoring statistics for principal component (PC) subspace and residual component subspace, respectively. As two subspaces are mutually orthogonal, $T_i^2$ 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:

$$T_i^2 = \|M_iW_i^T t_i\|^2$$

$$\text{SPE}_i = \left\| \sigma_i^{-1}(I - W_iM_iW_i^T) t_i \right\|^2$$

$$T_{c,i}^2 = \frac{1}{\sigma_i^2} \left( I + \frac{\sigma_i^2}{\|W_i\|^2} \right) t_i$$

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

$$T^2 = \sum_{i=1}^{K} R_{ni} T_i^2$$

$$\text{SPE} = \sum_{i=1}^{K} R_{ni} \text{SPE}_i$$

$$T_{c}^2 = \sum_{i=1}^{K} R_{ni} T_{c,i}^2$$

In consideration of the constraint that $\sum_{i=1}^{K} R_{ni} = 1$, 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:

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

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 $h_{\text{opt}}$ is determined by minimizing the approximation of the mean integrated square error, as depicted in (28), where $s$ is the standard deviation [45]

$$h_{\text{opt}} = 1.06 s N^{-1/5}$$

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

$$\int_{-\infty}^{J_{\text{th}}} p(J) dJ = \alpha$$
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
   a. Initialize the model parameters $W_i$ and $\mu_i$ ($i = 1: K$) via traditional PCA.
   b. Calculate the number of principal components through CPV approach.
   c. Discover the optimal model parameters $\theta = \{W_i, \mu_i, \sigma_i^2, \pi_i\}_{i=1:K}$ by implementation of the two-stage EM schedule.
   d. Discover the optimal number of local models $K^*$ among $K = \{1, 2, \ldots, K_{\text{max}}\}$ via the criterion given in (19).
4. Discover the optimal number of local models $K^*$ among $K = \{1, 2, \ldots, K_{\text{max}}\}$ 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 $\mu_t$ ($i = 1: K^*$) 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).

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

$$\text{Fault alarm} = \begin{cases} 0, & T^2 \leq J_{\text{th}, T^2} \text{ and } \text{SPE} \leq J_{\text{th}, \text{SPE}} \\ 1, & \text{others} \end{cases} \quad (30)$$

or

$$\text{Fault alarm} = \begin{cases} 0, & T^2_c \leq J_{\text{th}, T^2_c} \\ 1, & \text{others} \end{cases} \quad (31)$$

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 $T^2_c$ 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

$$\text{MAR} = \frac{\text{number of samples (} J \leq J_{\text{th}} \mid f \neq 0 \text{)}}{\text{total samples (} f \neq 0 \text{)}} \times 100\% \quad (32)$$

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

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 section. 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 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 PC analyzers. 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 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-MMPA can estimate which subspace faults occur by selecting appropriate monitoring statistic $T^2$ or SPE.
TABLE II
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 |
| PPCM     | 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 |

Fig. 1. Monitoring charts of each PPCA model using traditional MPPCA. (a) First local model. (b) Second local model. (c) Third local model. (d) Fourth local model. (e) Fifth local model. (f) Sixth local model.

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-MPPCA 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 paper, 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 CPV approach. 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 Figs. 1 and 2, fault can be detected timely and accurately by 12 or 6 monitoring graphs, respectively.

The monitoring consequences of I-MPPCA are represented in Fig. 3. Posterior probability $R$ is regarded as weight, which
Fig. 2. Monitoring charts of each local improved PPCA model under traditional MPPCA framework. (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. 3. Monitoring charts using I-MPPCA approach. (a) Weights of local PPCA models. (b) Two global monitoring statistics. (c) Global combined monitoring statistic.

| 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      | 1.25   | 84.37  | 39.12  |
| 1-MPPCA (T^2)  | 0      | 0.63   | 83.63  | 0.5    | 6.13   | 0      | 0.88   | 83.75  | 36.63  |
| 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   |
| 1-MPPCA (T^2)  | 56.87  | 1.75   | 3.50   | 12.13  | 63.50  | 16.13  | 36.00  | 8.38   | 87.75  | 16.75   |

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. 3(a). 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) and (25) are demonstrated in Fig. 3(b) and the statistic (26) is illustrated in Fig. 3(c). 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 $T^2_c$ are relatively larger than those of other local models. According to the comparison among Figs. 1, 2, and 3(b) and (c), 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 FARs and MARs. The detailed MARs are listed in Table III. It can be evidently acquired that the MARs of $T^2_c$ 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 $T^2_c$ 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 BENCHMARK

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 Section V-A. Then, complete data and incomplete data cases are considered in Table IV. Besides, the parameters are listed in Table IV and specific values can be found in [54].

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 (suspensions 1 and 2) have the identical configuration and the rear wheels (suspensions 3 and 4) share a different one. In other words, researchers just need to study suspensions 1 and 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 paper. 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. Therefore, the monitoring charts of the other three approaches are illustrated in Figs. 5 and 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 PC 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 Tables V and 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 tradeoff 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 Section 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 incomplete 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% incomplete data are taken as an example and utilized to establish model in this paper. Then, testing samples with incomplete 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% incomplete data.
2) Fault 4, the spring coefficient of suspension 3 is reduced by 30% from the 529th sample with 5% incomplete data.
3) Fault 5, the spring coefficient of suspension 1 is reduced by 30% from the 429th sample with 10% incomplete data.
4) Fault 6, the spring coefficient of suspension 3 is reduced by 30% from the 529th sample with 10% incomplete data.

MARs and FARs are concluded in Tables V and VI. As to I-MPPCA approach, it can be observed clearly that incomplete 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 of 6 faults, the accuracy rates will be reduced rapidly with the increasing missing rate.
TABLE V  
MARs (%) BASED ON AUTOSUSPENSION DATA

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

TABLE VI  
FARS (%) BASED ON AUTOSUSPENSION DATA

| Simulation  | K-means | FPCM | KPCA | I-MPPCA |
|-------------|---------|------|------|---------|
| Complete    | 2.82    | 0.18 | 1.41 | 0  |
| Incomplete  | 5.64    | 1.04 | 2.57 | 0.47 |

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 incomplete data. However, the FAR and MARs of I-MPPCA approach are slightly affected by these incomplete data values.

In conclusion, with regard to incomplete 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.

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 PC analyzers. 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 weight to data point belonging to a certain PPCA model of being chosen. Therefore, in order to provide optimal fault detection performance and observational convenience, the global monitoring statistics are acquired based on the weighted means 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

$$L = \sum_{n=1}^{N} \ln[p(t_n)] = \sum_{n=1}^{N} \ln \left( \sum_{i=1}^{K} \pi_i p(t_n | i) \right).$$  (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 $\pi_i$, $\mu_i$, $W_i$, and $\sigma_i^2$, which was first introduced in [55].

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

$$R_{ni} = \frac{p(t_n|i) \pi_i}{p(t_n)}.$$  (35)

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

$$\langle L_C \rangle = \sum_{n=1}^{N} \sum_{i=1}^{K} R_{ni} \left\{ \ln \pi_i - \frac{d}{2} \ln \sigma_i^2 - \frac{1}{2} \text{tr} \left( \mathbf{x}_n (\mathbf{x}_n^T) \right) \right\}$$

$$- \frac{1}{2\sigma_i^2} \| t_n - \mu_i \|^2 + \frac{1}{2\sigma_i^2} \text{tr} \left( W_i^T (t_n - \mu_i) W_i^T \right)$$

$$- \frac{1}{2\sigma_i^2} \text{tr} \left[ W_i^T W_i (\mathbf{x}_n (\mathbf{x}_n^T) \right].$$  (36)

In order to obtain the optimal values of key model parameters aforementioned, a Lagrange multiplier $\lambda$ is utilized to achieve the maximum value of (36). Thus, the solution of maximum likelihood can be transformed into the following optimization problem:

$$\max \langle L_C \rangle + \lambda \left( \sum_{i=1}^{K} \pi_i - 1 \right)$$

$$\text{s.t.} \; \sum_{i=1}^{K} \pi_i = 1.$$  (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

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

$$\hat{L}_C = \sum_{i=1}^{N} \sum_{n=1}^{K} R_{ni} \ln \left\{ \pi_i p(t_n, x_n^{(i)}) \right\}$$

(38)

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

The second stage (M-step) takes advantage of GEM to update $W_i$ and $\sigma^2_i$. The typical feature of GEM is to increase $\mu$-step. Iteration of (42)–(44) along with (35) expands for $(x_n^{(i)})$ and $(x_n^{(i)}x_n^{(i)T})$, only terms in $\hat{\mu}_i$ appear. Thus, the expected complete-data log-likelihood now can be obtained by inspection of (36) as follows:

$$\langle L_c \rangle = \sum_{i=1}^{N} \sum_{n=1}^{K} R_{ni} \ln \left\{ \bar{\pi}_i - \frac{d}{2} \ln \sigma^2_i - \frac{1}{2} \text{tr} \left\{ (x_n^{(i)}x_n^{(i)T}) \right\} \right\}
- \frac{1}{2\sigma^2_i} \left\| t_n - \bar{\mu}_i \right\|^2
+ \frac{1}{\sigma^2_i} \left\{ (x_n^{(i)})^T W_i^T (t_n - \bar{\mu}_i) \right\}
- \frac{1}{2} \text{tr} \left\{ W_i^T W_i \left( x_n^{(i)}x_n^{(i)T} \right) \right\} \right\}.$$  

(39)

Much simplified M-step formulas can be acquired when (39) reaches the maximum with respect to $W_i$ and $\sigma^2_i$ (keeping $\bar{\mu}_i$ fixed)

$$\bar{W}_i = S_i W_i \left( \sigma^2_i I + M^{-1}W_i^T S_i W_i \right)^{-1}$$

(40)

$$\bar{\sigma}^2_i = \frac{1}{d} \text{tr} \left\{ S_i - S_i W_i M^{-1} \bar{W}_i \right\}$$

(41)

where

$$S_i = \frac{1}{\bar{\pi}_i} N \sum_{n=1}^{N} R_{ni} (t_n - \bar{\mu}_i)_i (t_n - \bar{\mu}_i)_i^T$$

(42)

$$\bar{\mu}_i = \frac{1}{\bar{\pi}_i} \sum_{n=1}^{N} R_{ni} t_n$$

(43)

$$\bar{\pi}_i = \frac{1}{N} \sum_{n=1}^{N} R_{ni}$$

(44)

Obviously, the symbol $\sim$ 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).

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., to be published, doi: 10.1109/TCYB.2017.2755864.

[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 et al., “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, no. 2, 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 (Advances in Industrial Control). London, U.K.: Springer, 2014.

[13] S. Yin, S. X. Ding, A. Haghighi, 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 et al., “A fault detection method for hard disk drives based on mixture of Gaussians and nonparametric 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. Zha, 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. 1389–1399, 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. Ichishia, “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. Pinatow, 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. Zha, Z. Ge, and Z. Song, “Distributed parallel PCA for modeling and monitoring of large-scale plant-wide processes with big data,” IEEE Trans. Cybern., 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.
