IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received August 15, 2018, accepted September 6, 2018, date of publication September 17, 2018, date of current version October 12, 2018.

Digital Object Identifier 10.1109/ACCESS.2018.2870140

Principal Polynomial Analysis for Fault Detection and Diagnosis of Industrial Processes

XINMIN ZHANG [1], **MANABU KANO** [1], **(Member, IEEE), AND YUAN LI** [2]
[1]Department of Systems Science, Kyoto University, Kyoto 606-8501, Japan
[2]Department of Information Engineering, Shenyang University of Chemical Technology, Shenyang 110142, China
Corresponding author: Xinmin Zhang (zhang.xinmin.58s@st.kyoto-u.ac.jp)

ABSTRACT Real-time process monitoring is crucial to improve the productivity, process safety, and product quality. In this paper, a novel fault detection and diagnosis technique based on a principal polynomial analysis (PPA) is proposed. PPA is a nonlinear modeling technique, which describes the data using a set of flexible principal polynomial components. Compared with the PCA-based methods, PPA is more effective in capturing the intrinsic nonlinear geometry structure of the process data. Moreover, compared with other nonlinear methods, such as kernel-based and neural-networks-based methods, PPA has the appealing features of straightforward out-of-sample extension, volume-preservation, and invertibility. In addition, two new types of fault detection and diagnosis statistics are derived. The effectiveness of the proposed PPA-based monitoring method was verified through its applications to a nonlinear numerical example and an industrial benchmark process. The application results have demonstrated that the proposed method has superior fault detection and diagnosis performance than the conventional PCA-based and kernel PCA-based methods.

INDEX TERMS Fault detection and diagnosis, nonlinear processes, process monitoring, principal polynomial analysis.

I. INTRODUCTION

In modern industry, it is essential to monitor the production process in real time so as to increase overall equipment effectiveness and improve the process safety and product quality. Data-driven process monitoring and control techniques have been widely applied to various industrial processes including chemicals [1]–[4], pharmaceuticals [5], [6], ironmaking [7], [8], and semiconductor manufacturing [9]. Principal component analysis (PCA) is one of the most widely used multivariate statistical techniques [10]. Since its good features of simplicity, invertibility, energy compaction, and intuitive interpretation, PCA-based process monitoring approaches have been successfully used in many industrial processes [11]–[13]. However, PCA-based approaches assume that the process is linear, which restricts their applications to nonlinear industrial processes.

To monitor nonlinear processes, several nonlinear modeling methods such as nonlinear PCA (NLPCA) [14], [15] and kernel PCA (KPCA) [16], [17] have been used. NLPCA is a nonlinear generalization of PCA by using an auto-associative neural network to map the data into feature space. Since the nonlinear features are not explicit in the formulation, NLPCA is difficult to compute the contributions of the original process variables to the fault when performing fault diagnosis [17]. Moreover, the selection of network structure and model parameters directly affect the regularization ability of the network. KPCA is another popular nonlinear extension of PCA by using nonlinear kernel function. KPCA projects the original data onto a kernel feature space, where a linear PCA model is performed. The basic theory behind KPCA is that the nonlinear relationship among variables in the original space is most likely to be linear after kernel mapping. However, the possibility that the intrinsic nonlinear geometry structure of data may reside on a manifold is not explicitly considered by KPCA [18]. Moreover, KPCA cannot guarantee that the learned data representations are accurate in the minimum information loss term in the original input space [19]. Furthermore, KPCA is also not straightforward to invert the Hilbert feature representation to the original input space. Recently, some nonlinear manifold learning methods have been proposed, such as local linear embedding (LLE) [20], Isomap [21], diffusion maps (DM) [22], and Laplacian eigenmaps [23]. The basic idea of these methods is to find the lower-dimensional embedding representation of data lying on a nonlinear manifold. However, these techniques do not have a straightforward out-of-sample extension, which is important for online fault detection and diagnosis. In addition, these methods also need to specify the number of reduced dimensionality prior to modeling.

In this study, a novel fault detection and diagnosis technique based on principal polynomial analysis (PPA) is developed. PPA is a nonlinear modeling technique which describes data efficiently by using a set of flexible principal polynomial components. Two new types of fault detection and diagnosis statistics are derived in the PPA setting. Compared to the PCA-based methods, PPA is more effective in capturing the intrinsic nonlinear geometry structure of process data. Moreover, unlike other nonlinear approaches such as NLPCA, KPCA, Laplacian Eigenmaps, and LLE, PPA has the appealing features of straightforward out-of-sample extension, volume-preservation, and invertibility. In addition, PPA implements a simple and efficient sequential learning algorithm to extract the nonlinear features. Thus, it does not require to specify the number of reduced dimensions prior to modeling as compared to the nonlinear manifold learning methods. The effectiveness and advantages of the proposed PPA-based monitoring method were verified through its applications to a nonlinear numerical example and a benchmark of Tennessee Eastman process. The application results have shown that the proposed method has better performance than the standard PCA-based and KPCA-based methods in feature extraction, fault detection, and fault diagnosis.

The remainder of this paper is structured as follows. Section 2 provides a brief introduction of principal component analysis. In Section 3, a novel fault detection and diagnosis technique based on principal polynomial analysis is given. In Section 4, the superiority of the proposed PPA-based monitoring technique is illustrated through its application to a nonlinear numerical example and a benchmark of Tennessee Eastman process, and its application results are compared with the standard PCA-based and KPCA-based monitoring methods. Section 5 gives the conclusions of this paper.

II. PRINCIPAL COMPONENT ANALYSIS

PCA is an effective tool for dimensionality reduction or feature extraction. PCA aims at finding an orthogonal linear projection that projects the original data onto a low-dimensional space, known as principal component (PC) subspace, such that the variance of the projected data is maximized and the reconstruction error is minimal [10], [24]. Consider a data matrix $\mathbf{X} \in \Re^{N\times d}$ with N observations and d variables. PCA decomposes $\mathbf{X}$ into score matrix $\mathbf{T} \in \Re^{N\times \ell}$, loading matrix $\mathbf{P} \in \Re^{d\times \ell}$, and residual matrix $\mathbf{E} \in \Re^{N\times d}$, as follows:

$$\mathbf{X} = \sum_{i=1}^{\ell} t_i p_i^{\mathrm{T}} + \mathbf{E} = \mathbf{T}\mathbf{P}^{\mathrm{T}} + \mathbf{E} = \hat{\mathbf{X}} + \mathbf{E} \tag{1}$$

where ℓ is the number of retained PCs in the PCA model. The score vectors $t_i \in \Re^N$ are orthogonal and the loading vectors $p_i \in \Re^d$ are orthonormal. $\hat{\mathbf{X}} \in \Re^{N\times d}$ is an estimation of the original data (also known as the reconstructed observation data). Computationally, PCA can be calculated by using singular value decomposition of the data matrix or eigenvalue decomposition of the covariance matrix. If only the first few PCs are required, the sequential NIPALS algorithm can be used [24].

To perform process monitoring, Hotelling's T^2 and Q (or *SPE*) statistics are commonly utilized to monitor the PC subspace and the residual subspace, respectively. Let $x_{new} \in \Re^d$ be the new observation vector. The T^2 statistic is given by

$$T^2 = x_{new}^{\mathrm{T}} \mathbf{P} \Lambda^{-1} \mathbf{P}^{\mathrm{T}} x_{new} \tag{2}$$

where $\Lambda \in \Re^{\ell\times\ell}$ is the diagonal matrix, and its diagonal elements are the variances of the retained PC scores in the PCA model. The Q statistic is defined as follows:

$$Q = \|x_{new} - \hat{x}_{new}\|^2 = \|(\mathbf{I} - \mathbf{P}\mathbf{P}^{\mathrm{T}})x_{new}\|^2 \tag{3}$$

where $\hat{x}_{new}$ denotes the reconstructed measurement vector and $\mathbf{I} \in \Re^{d\times d}$ denotes an identity matrix. The thresholds for the Hotelling's T^2 and Q statistics can be easily determined by using the F-distribution [11], [25] and χ^2-distribution [25], [26] functions, respectively.

III. PRINCIPAL POLYNOMIAL ANALYSIS BASED FAULT DETECTION AND DIAGNOSIS TECHNIQUE

In this section, principal polynomial analysis, which is an effective nonlinear feature extraction technique, is first introduced. Then, fault detection and diagnosis technique based on principal polynomial analysis is proposed.

A. PRINCIPAL POLYNOMIAL ANALYSIS

In general, PCA works well when the process is linear. However, for the process with nonlinear nature, PCA performs poorly or inefficiently due to its assumption that the process is linear. This is primarily because the optimal linear decomposition of PCA in (1) can be acquired if and only if the conditional mean in each PC is constant along the considered dimension [19]. This assumption is also known as conditional mean independence restriction, as shown in Fig. 1. In Fig. 1(a), PCA presents a good solution because of the data meet the required symmetry; the conditional mean (red circles) in PC2 (zero for centered data) is independent of PC1. Inversely, PCA presents a poor solution in Fig. 1(b) where the process variables show nonlinear nature because the conditional mean in PC2 is not constant along PC1. So when PCA maps the data onto PC1 along its orthogonal direction, a large reconstruction error will be produced. To handle this problem, principal polynomial analysis (PPA) was proposed [19]. PPA learns a low-dimensional representation from process data using a set of principal polynomial components (PPCs). By replacing straight PCs in PCA with curved PPCs, the nonlinear characteristic of process variables is captured by the PPA model.

Mathematically, PPA implements a sequential algorithm to calculate the principal polynomial components. In each step, a leading vector that best projects the data is calculated. More specifically, given a random vector $x \in \Re^d$, on the p-th step

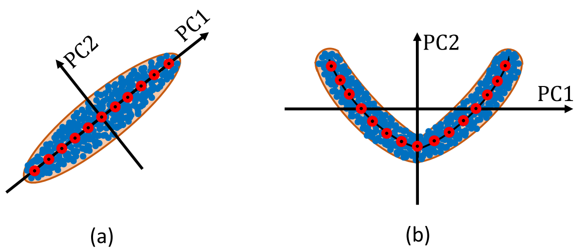

FIGURE 1. **The conditional mean independence restriction.**

of the sequential algorithm, PPA takes the following form:

$$\alpha_p = e_p^{\mathrm{T}} x_{p-1} \quad (4)$$
$$x_p = \mathbf{E}_p^{\mathrm{T}} x_{p-1} - \hat{m}_p \quad (5)$$

where α_p denotes the projection of x_{p-1} onto the leading vector e_p at step p; x_{p-1} is the residual derived from the previous step. When $p = 1$, $x_0 = x$ is just the original input data. $\hat{m}_p$ represents the estimated conditional mean and x_p is the residual that will be utilized for the next step. The leading vector $e_p \in \Re^{(d-p+1)}$ can be found through maximizing the variance of the projected data:

$$e_p = \arg \max_{\|e\|=1} \left\{ \mathbb{E}\left[(e^{\mathrm{T}} x_{p-1})^2 \right] \right\}. \quad (6)$$

$\mathbf{E}_p^{\mathrm{T}} \in \Re^{(d-p)\times(d-p+1)}$ is the matrix whose rows consist of $d-p$ orthonormal vectors, and the subspace spanned by $\mathbf{E}_p^{\mathrm{T}}$ is orthogonal to e_p. Mathematically, $\mathbf{E}_p$ and e_p meet

$$\mathbf{E}_p^{\mathrm{T}} e_p = \emptyset \quad (7)$$
$$\mathbf{E}_p^{\mathrm{T}} \mathbf{E}_p = \mathbf{I}_{(d-p)\times(d-p)} \quad (8)$$

where $\mathbf{I}$ is an identity matrix. Note that the information loss at the p-th step is x_p. Based on the minimum information loss criterion, PPA can also be written in the mean square error (MSE) term as

$$e_p = \arg \min_{\|e\|=1} \mathbb{E}\left[\|\mathbf{E}_p^{\mathrm{T}} x_{p-1} - \hat{m}_p\|^2 \right]. \quad (9)$$

Since the orthonormality of the projection vector e_p and $\mathbf{E}_p$, minimizing the MSE term of (9) equals to maximizing the variance term of (6).

Theoretically, the conditional mean can be estimated by utilizing any regression approach $\hat{m}_p = g(\alpha_p)$. PPA employs a polynomial function because of its enough flexibility in giving solutions through the use of the appropriate degree r_p. The estimation equation can be defined as

$$\hat{m}_p = \mathbf{W}_p v_p \quad (10)$$

where $\mathbf{W}_p \in \Re^{(d-p)\times(r_p+1)}$ represents the polynomial coefficients and $v_p = [1, \alpha_p, \alpha_p^2, \cdots, \alpha_p^{r_p}]^{\mathrm{T}}$ represents the Vandermonde vector of $\mathbf{e}_p^{\mathrm{T}} x_{p-1}$. $\mathbf{W}_p$ can be calculated by solving a least squares problem. Consider N input samples, which are formed in the matrix $X_0 \in \Re^{d\times N}$. Then, equations (4) and (5) can be written as the following matrix form

$$\boldsymbol{\alpha}_p = e_p^{\mathrm{T}} \mathbf{X}_{p-1} \quad (11)$$

$$\mathbf{X}_p = \mathbf{E}_p^{\mathrm{T}} \mathbf{X}_{p-1} - \hat{\mathbf{M}}_p \quad (12)$$

where $\hat{\mathbf{M}}_p = \mathbf{W}_p \mathbf{V}_p$ is the estimated conditional means formed in column-wise and $\mathbf{V}_p = [v_{p,1}, v_{p,2}, \cdots, v_{p,N}] \in \Re^{(r_p+1)\times N}$ denotes the Vandermonde matrix. The least squares solution for $\mathbf{W}_p$ is given by

$$\mathbf{W}_p = (\mathbf{E}_p^{\mathrm{T}} \mathbf{X}_{p-1}) \mathbf{V}_p^{\dagger} \quad (13)$$

where $\dagger$ denotes the pseudoinverse operation.

To sum up, the cost function for leading vector e_p in the PPA model can be expressed as

$$\begin{aligned} e_p = \ & \arg \min_{\|e\|=1} \mathbb{E}\left[\|\mathbf{E}_p^{\mathrm{T}} x_{p-1} - \mathbf{W}_p v_p\|^2 \right] \\ & \text{s.t. } \mathbf{E}_p^{\mathrm{T}} \mathbf{E}_p = \mathbf{I} \\ & \quad\ \ \mathbf{E}_p^{\mathrm{T}} e_p = \emptyset \\ & \quad\ \ \mathbf{W}_p = (\mathbf{E}_p^{\mathrm{T}} \mathbf{X}_{p-1}) \mathbf{V}_p^{\dagger} \end{aligned} \quad (14)$$

There are two ways to solve (14). One is the PCA-based solution method, which uses the first eigenvector of the sample covariance as leading vector e_p and uses the remaining eigenvectors as $\mathbf{E}_p^{\mathrm{T}}$. In such case, PPA could provide a smaller or at least the same truncation error as compared to PCA if the parameter r_p is tuned appropriately. This is because when $r_p = 1, \forall p$, PPA reduces to PCA. The other is the gradient descent optimation method, which solves a non-convex problem. Although the gradient descent optimation method may provide a better solution than the PCA-based solution method, its computational burden is greatly increased. In this work, the PCA-based solution method was chosen due to its simplicity.

PPA has the following advantages: (1) Compared to PCA, PPA is more powerful in identifying and capturing the nonlinear features of process data; (2) Compared to other nonlinear methods such as kernel-based and neural networks-based methods, PPA has the appealing features of straightforward out-of-sample extension, volume-preservation, and invertibility. (3) PPA does not require to specify the number of reduced dimensions prior to modeling as compared to the nonlinear manifold learning methods such as Laplacian Eigenmaps and DM. As a powerful feature extraction approach, PPA has been successfully used for remote sensing data processing [27], but it has not been investigated for process monitoring and fault diagnosis of industrial processes.

B. PRINCIPAL POLYNOMIAL ANALYSIS FOR FAULT DETECTION AND DIAGNOSIS

In this section, a novel principal polynomial analysis based fault detection and diagnosis method that inherits all the merits of the PPA feature extraction algorithm is proposed. The proposed PPA-based fault detection and diagnosis strategy is as follows.

For online process monitoring, two new monitoring statistics T_{PPA}^2 and Q_{PPA} are derived in the PPA setting to measure the variability in the PPC subspace and residual subspace, respectively. When a new observation $x_{new} \in \Re^d$ becomes available, it can be projected onto the corresponding PPC subspace and residual subspace using the learned model parameters in the training phase. More specifically, following the sequential decomposition steps of PPA in (4) and (5), the p-th PPC score α_p^{new} is given by

$$\alpha_p^{new} = e_p^{\mathrm{T}} x_{p-1}^{new} \tag{15}$$

$$x_p^{new} = \mathbf{E}_p^{\mathrm{T}} x_{p-1}^{new} - \hat{m}_p = \mathbf{E}_p^{\mathrm{T}} x_{p-1}^{new} - W_p v_p \tag{16}$$

where e_p, v_p, $\mathbf{E}_p$, and W_p denote a group of learned model parameters of the PPA model from the normal training data. Suppose that the number of retained significant PPCs in the PPA representation is ℓ. In most applications, they only require the first few ℓ significant PPCs. Hence, it is simple and efficient to calculate the PPCs using the sequential algorithm. Let $\boldsymbol{\alpha}_{new} = [\alpha_1^{new}, \alpha_2^{new}, \cdots, \alpha_\ell^{new}]^{\mathrm{T}} \in \Re^\ell$ denote the score vector which corresponds to the first ℓ significant PPCs. The T_{PPA}^2 statistic is defined as follows:

$$T_{PPA}^2 = \boldsymbol{\alpha}_{new}^{\mathrm{T}} \Lambda_{PPA}^{-1} \boldsymbol{\alpha}_{new} \tag{17}$$

where $\Lambda_{PPA} \in \Re^{\ell\times\ell}$ denotes the diagonal matrix, and its diagonal elements are the variances of the retained PPC scores. The Q_{PPA} statistic is defined as follows:

$$Q_{PPA} = \|x_{new} - \hat{x}_{new}\|^2 \tag{18}$$

where $\hat{x}_{new}$ denotes the reconstructed observation vector given the first ℓ PPCs. $\hat{x}_{new}$ is acquired via recursively conducting the following transformation:

$$\hat{x}_{p-1}^{new} = \begin{pmatrix} e_k & \mathbf{E}_p \end{pmatrix} \begin{pmatrix} \alpha_p^{new} \\ \hat{x}_p^{new} + W_p v_p \end{pmatrix} \tag{19}$$

where $p = \ell : -1 : 1$, $\hat{x}_\ell^{new} = \mathbf{0}_{(d-\ell)\times 1}$, and $\hat{x}_{new} = \hat{x}_0^{new}$.

The threshold for the T_{PPA}^2 statistic can be calculated as follows [11]:

$$CL_{T_{PPA}^2} = \frac{(N-1)\ell}{(N-\ell)} F_\alpha(\ell, N-\ell) \tag{20}$$

where $F_\alpha(\ell, N-\ell)$ denotes an F distribution with the degree of freedoms ℓ and $N-\ell$ at significance level α.

The threshold for the Q_{PPA} statistic can be calculated as follows [26]:

$$CL_{Q_{PPA}} = g\chi_{h,\alpha}^2; \quad g = \frac{\bar{v}}{2\bar{\mu}}, \ h = \frac{2\bar{\mu}^2}{\bar{v}} \tag{21}$$

where $\chi_{h,\alpha}^2$ denotes a chi-square distribution with the degree of freedom h at significance level α; $\bar{\mu}$ and $\bar{v}$ denote the estimated mean and variance of Q_{PPA} from the normal operating condition data. The process is identified as abnormal if either T_{PPA}^2 or Q_{PPA} statistic goes beyond its corresponding threshold.

Once an abnormal event is detected, it is critical to identify the possible causes of this abnormal event in order to execute corrective actions. Contribution plots method is a well-known diagnostic tool for providing valuable information to determine which process variables responsible for the fault [26].

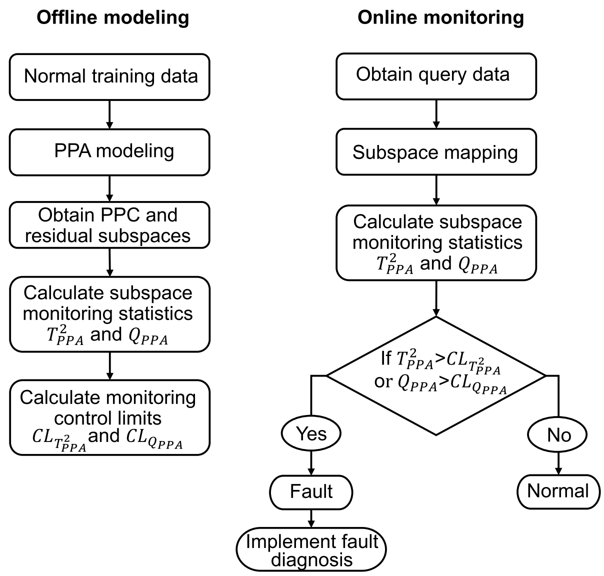

FIGURE 2. **Flowchart of the proposed PPA-based monitoring system.**

Due to its transparency, simplicity, and interpretability, contribution plots method has been widely used in various industrial processes [28], [29]. In this work, two new types of the PPA-based contribution plots called $C_i^{Q_{PPA}}$ and $C_i^{T_{PPA}^2}$ are proposed to measure the contribution of each process variable to the fault as follows.

When the Q_{PPA} statistic of a new observation is above the confidence limits, the contribution of the i-th process variable on the Q_{PPA} statistic is calculated as follows:

$$C_i^{Q_{PPA}} = \tilde{x}_{new,\,i}^2 = \left(x_{new,\,i} - \hat{x}_{new,\,i}\right)^2 \tag{22}$$

where $\tilde{x}_{new,\,i}$ is the residual of the new observation of the i-th process variable. A large value of $C_i^{Q_{PPA}}$ indicates that the i-th process variable has a significant contribution to the current fault.

Similarly, the contribution of the i-th variable on the T_{PPA}^2 statistic can be calculated as follows:

$$C_i^{T_{PPA}^2} = \sum_{p=1}^{\ell} \left(e_p^{\mathrm{T}} x_{p-1}^{new} \left(\Lambda_{PPA}^{(p,p)} \right)^{-\frac{1}{2}} \right)^2 \tag{23}$$

where $x_0^{new} = \Xi_i x^{new}$ and Ξ_i is a d-by-d square matrix in which the (i, i)-th element is one and all the other elements are zero. $\Lambda_{PPA}^{(p,p)}$ denotes the p-th diagonal element of Λ_{PPA}. The larger of $C_i^{T_{PPA}^2}$, the higher of the contribution of the i-th process variable to the current fault.

The detailed steps to perform the proposed PPA-based fault detection and diagnosis method are summarized in the following two parts. The flowchart of the proposed PPA-based monitoring technique is given in Fig. 2.

Offline modeling:

1) Acquire the process data X under normal operating condition, and then scale it to zero mean and unit variance.

2) Construct the PPA model on the data matrix to get the PPC subspace and residual subspace.
3) Compute the monitoring statistics Q_{PPA} and T^2_{PPA} in the residual subspace and PPC subspace, respectively.
4) Compute the corresponding thresholds $CL_{Q_{PPA}}$ and $CL_{T^2_{PPA}}$ for the Q_{PPA} and T^2_{PPA} statistics.
5) Store the learned model parameters e_p, v_p, $\mathbf{E}_p$, and W_p.

Online monitoring:

1) Obtain a new observation x_{new}, and then scale it using the mean and variance of the modeling data.
2) Map x_{new} into the residual subspace and PPC subspace using the learned model parameters, and then compute the monitoring statistics Q_{PPA} and T^2_{PPA}.
3) Compare Q_{PPA} and T^2_{PPA} with the corresponding thresholds $CL_{Q_{PPA}}$ and $CL_{T^2_{PPA}}$, respectively. If either Q_{PPA} or T^2_{PPA} exceeds the corresponding threshold, the process is considered as faulty and the fault diagnosis step is performed to identify the root cause.

IV. CASE STUDIES

In this section, the efficiency and advantages of the proposed PPA-based fault detection and diagnosis method were illustrated through its application to a nonlinear numerical example and a benchmark of Tennessee Eastman process. The application results were compared with those of the standard PCA-based and KPCA-based monitoring methods. The calculations were conducted in the software environment of Windows XP and MATLAB R2018a.

A. NONLINEAR NUMERICAL EXAMPLE

A nonlinear simulation example is given to illustrate the effectiveness of the proposed PPA-based method in feature extraction, fault detection, and fault diagnosis. The nonlinear system is described by the following equations:

$$\begin{aligned} x_1 &= u^2 + 0.7\sin(2\pi u) + \varepsilon_1 \\ x_2 &= u + \varepsilon_2 \\ x_3 &= u^3 + u + 1 + \varepsilon_3 \end{aligned} \tag{24}$$

where u is a random number uniformly distributed in the interval $[-1, 1]$ and ε_i is the zero-mean white noise with a standard deviation of 0.01.

First, the feature extraction capacity of PPA for nonlinear process data was investigated. To construct the PCA and PPA models, 300 samples were generated. Fig. 3 displays the feature extraction results using PCA and PPA, where the green dots denote the original data and the red asterisks denote the one-factor representations (or reconstructions). KPCA was not presented since the intractable kernel transformation makes it practically not possible to implement direct reverse projection. The polynomial degree was set to 15 in the PPA model. As shown in Fig. 3(a), PCA failed to identify the underlying structure of the data and gave a straight line fit of the data, which is obviously inappropriate for the data. By contrast, PPA succeeded in identifying the underlying nonlinear structure of the data and gave a more accurate curvilinear fit of the data, as shown in Fig. 3(b). Consequently, PPA is more efficient in describing the nonlinear data than PCA.

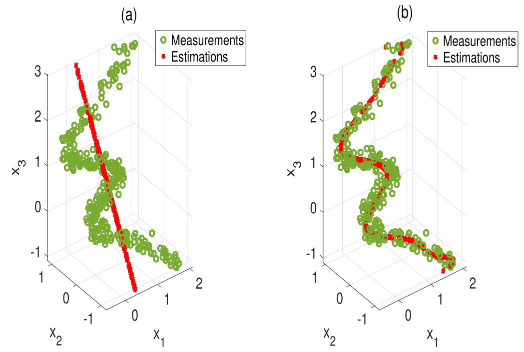

FIGURE 3. Feature extraction results using (a) PCA and (b) PPA.

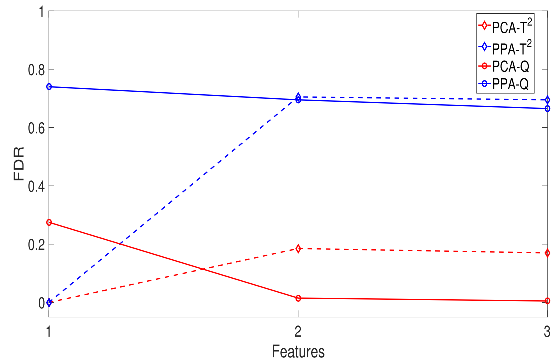

FIGURE 4. Fault detection rates of PCA-based and PPA-based methods with respect to the number of retained features.

To show the fault detection performance of the proposed PPA-based method, a ramp fault was induced by adding $0.01(n - 100)$ to x_1 from sample 101 to sample 300, where n is the sample number. Fig. 4 shows the fault detection rates (FDR) of PCA-based and PPA-based methods with different numbers of features retained in each model. Here, the number of retained features represents the number of retained principal components (PCs) and principal polynomial components (PPCs) in PCA-based and PPA-based monitoring models, respectively. The higher of FDR, the better performance of the corresponding monitoring statistic provides. The confidence limit was specified as 99% for each method. From Fig. 4, it can be clearly seen that the proposed PPA-based method provided a higher FDR than the PCA-based method with respect to the same number of retained features. More interestingly, using one feature is enough in the proposed PPA-based method (PPA-Q statistic) to outperform the monitoring results of the PCA-based method using any number of features. Fig. 5 shows the detailed monitoring results of PCA-Q and PPA-Q using one feature. From Fig. 5(a), it can be clearly observed that the PCA-Q statistic cannot detect the fault efficiently since the majority of fault samples are below the threshold. In comparison, the monitoring results of the proposed PPA-Q statistic

TABLE 1. Fault detection rates of different methods.

Statistics	PCs=1	PCs=2	PCs=3	Statistics	PPCs=1	PPCs=2	PPCs=3	Statistics	KPCs=4
PCA-T^2	0	0.19	0.17	PPA-T^2	0	0.71	0.70	KPCA-T^2	0.15
PCA-Q	0.28	0.02	0.01	PPA-Q	0.74	0.70	0.67	KPCA-Q	0.49

TABLE 2. Process variables used for process monitoring in TE process.

No.	Variables	No.	Variables
X1	A feed (stream 1)	X18	Stripper temperature
X2	D feed (stream 2)	X19	Stripper steam flow
X3	E feed (stream 3)	X20	Compressor work
X4	A and C feed (stream 4)	X21	Reactor cooling water outlet temperature
X5	Recycle flow (stream 8)	X22	Separator cooling water outlet temperature
X6	Reactor feed rate (stream 6)	X23	D feed flow valve
X7	Reactor pressure	X24	E feed flow valve
X8	Reactor level	X25	A feed flow valve
X9	Reactor temperature	X26	A and C feed flow valve
X10	Purge rate (stream 9)	X27	Compressor recycle valve
X11	Product separator temperature	X28	Purge valve
X12	Product separator level	X29	Separator pot liquid flow valve
X13	Product separator pressure	X30	Stripper liquid product flow valve
X14	Product separator underflow (stream 10)	X31	Stripper steam valve
X15	Stripper level	X32	Reactor cooling water flow
X16	Stripper pressure	X33	Condenser cooling water flow
X17	Stripper underflow (stream 11)		

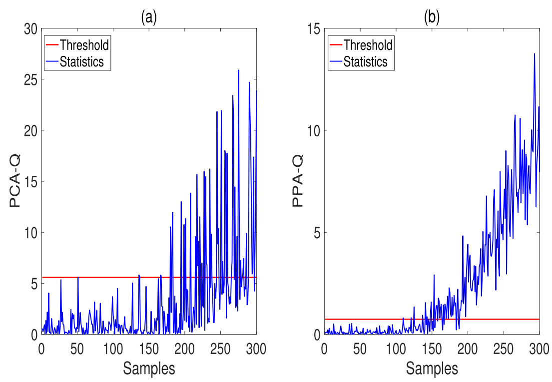

FIGURE 5. Process monitoring results of (a) PCA-Q and (b) PPA-Q.

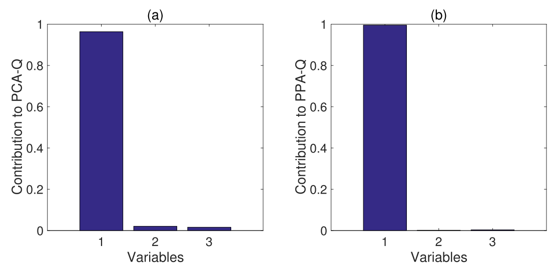

FIGURE 6. Variables contribution to (a) PCA-Q and (b) PPA-Q.

are much better than those of PCA-Q, as shown in Fig. 5(b). In addition, the proposed PPA-Q statistic can detect the fault earlier than the PCA-Q statistic. It is also noteworthy that the threshold of the PPA-Q statistic is smaller than in PCA-Q statistic, which is consistent with the theoretical principle of improved reconstruction errors in PPA. After the fault was first detected, the variable contribution plots corresponding to the PCA-Q and PPA-Q statistics are provided to identify the variables responsible for the out-of-control situation, as depicted in Fig. 6. As shown in Fig. 6, both the PCA-Q and PPA-Q contribution statistics clearly identified that the variable responsible for the out-of-control situation is x_1, which agrees well with that of the induced fault. Moreover, it is interesting to notice that x_1 has a higher contribution to the proposed PPA-Q statistic as compared to PCA-Q, as shown in Fig. 6. It indicates that the proposed PPA-Q contribution statistic is more sensitive than the PCA-Q contribution statistic. All of these results demonstrate the feasibility and superiority of the proposed PPA-based method in monitoring the nonlinear process than the PCA-based method. In order to further verify that the proposed PPA-based method is also superior to other nonlinear methods in monitoring performance, the standard KPCA-based method was performed in the same data, and its FDR is summarized in Table 1. The number of kernel principal components (KPCs) retained in KPCA-based method was 4 and the kernel parameter is chosen as $5m$ (m is the dimension of input space), which is suggested as in [17] and [30]. As shown in Table 1, the KPCA-Q statistic is more sensitive than its corresponding KPCA-T^2 statistic. The FDR of KPCA-Q is higher than that of the PCA-based method, whereas it is lower than that of the proposed PPA-based method. As a result, the proposed PPA-based method achieved the highest FDR among the three methods, as shown in Table 1.

B. TE BENCHMARK PROCESS

The Tennessee Eastman (TE) process is an industrial benchmark for testing the efficiencies of fault detection and diagnosis methods in process system engineering [10], [31]–[34]. Fig. 7 shows the flowchart of the TE process, which comprises five major units: a reactor, a condenser, a vapor-liquid separator, a recycle compressor, and a product

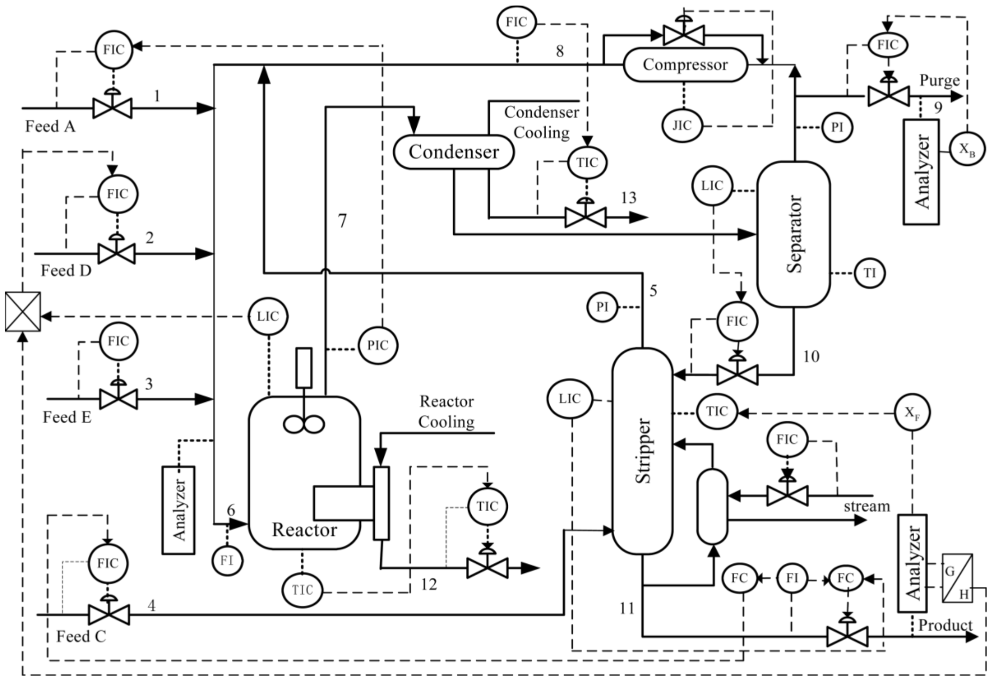

FIGURE 7. The flow diagram of the Tennessee Eastman process.

stripper. Four raw materials of A, C, D, and E are fed into the reactor in which the liquid products G and H are formed. The process includes 12 manipulated variables, 22 process measurement variables, and 19 composition variables. Since the composition variables are measured with dead time and time delays, they are not used for building the online monitoring models. In addition, the agitation speed variable is excluded since it is not manipulated. Thus, 22 process measurement variables XMEAS (1-22) and 11 manipulated variables XMV (1-11) were finally used to construct the monitoring models. Table 2 shows the selected process variables. The training data containing 500 observations collected under normal operating condition was used to build the model. The testing data contains a set of 21 different process faults, which were introduced into the process at sample 161. That is, the process ran normally in the first 160 samples, and then the faults occurred from sample 161 to the end. A detailed description of the faults is listed in Table 3. In Table 3, Faults 3, 9, and 15 are small faults, which have little effect on the overall process behavior due to feedback control. Each testing dataset consists of 960 observations. The sampling interval of manipulated and measured variables is 3 min. All the training and testing datasets are available at http://web.mit.edu/braatzgroup/links.html.

The superiority of the proposed PPA-based method was investigated for fault detection of the 21 faults of the TE process. The number of retained PCs and KPCs in the PCA-based and KPCA-based method was respectively 14 and 22, which can explain most of the process data information. The number of PPCs retained in the proposed

TABLE 3. Induced process faults in TE process.

No.	Process variable	Type
IDV(1)	A/C feed ratio, B composition constant	Step
IDV(2)	B composition, A/C ration constant	Step
IDV(3)	D feed temperature	Step
IDV(4)	Reactor cooling water inlet temperature	Step
IDV(5)	Condenser cooling water inlet temperature	Step
IDV(6)	A feed loss	Step
IDV(7)	C header pressure loss-reduced availability	Step
IDV(8)	A, B, and C feed composition	Random variation
IDV(9)	D feed temperature	Random variation
IDV(10)	C feed temperature	Random variation
IDV(11)	Reactor cooling water inlet temperature	Random variation
IDV(12)	Condenser cooling water inlet temperature	Random variation
IDV(13)	Reaction kinetics	Slow drift
IDV(14)	Reactor cooling water valve	Sticking
IDV(15)	Condenser cooling water valve	Sticking
IDV(16)	Unknown	Unknown
IDV(17)	Unknown	Unknown
IDV(18)	Unknown	Unknown
IDV(19)	Unknown	Unknown
IDV(20)	Unknown	Unknown
IDV(21)	The valve fixed at steady state position	Constant position

PPA-based method was 4 and the polynomial degree was set to 4, which are determined by cross-validation. The confidence level was specified as 99%. The quantitative fault detection results of PCA-based, KPCA-based, and PPA-based methods are summarized in Table 4, where the FDRs of each monitoring statistic for all 21 faults were calculated. The highest FDR is highlighted in bold. Table 4 reveals that the proposed PPA-Q statistic provides the best monitoring results in most of the faults of the TE process as compared to the other monitoring statistics. For the purpose of illustration, two types of representative faults, Faults 15 and 20, are taken as the examples to show the fault detection performance

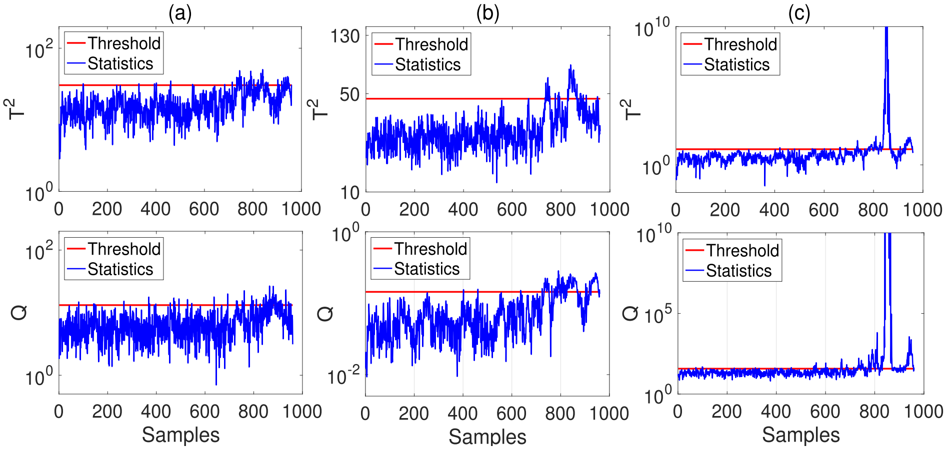

FIGURE 8. Process monitoring results of fault 15 by (a) PCA, (b) KPCA, and (c) PPA.

TABLE 4. Fault detection rates for all 21 faults by different methods in TE process.

No.	PCA-T^2	PCA-Q	KPCA-T^2	KPCA-Q	PPA-T^2	PPA-Q
IDV(1)	0.99	**1**	**1**	**1**	0.99	**1**
IDV(2)	0.98	**0.99**	**0.99**	**0.99**	0.97	**0.99**
IDV(3)	0.06	0.06	0.03	0.11	0.12	**0.20**
IDV(4)	0.32	**1**	**1**	0.80	0.16	0.99
IDV(5)	0.28	0.29	0.26	0.32	0.31	**0.40**
IDV(6)	0.99	**1**	**1**	**1**	0.99	**1**
IDV(7)	**1**	**1**	**1**	**1**	0.48	**1**
IDV(8)	0.97	0.96	0.98	0.98	0.94	**1**
IDV(9)	0.05	0.05	0.01	0.09	0.12	**0.18**
IDV(10)	0.46	0.46	0.37	0.62	0.50	**0.63**
IDV(11)	0.49	**0.79**	0.70	0.67	0.36	**0.79**
IDV(12)	**0.99**	0.96	**0.99**	**0.99**	0.96	**0.99**
IDV(13)	0.94	0.95	0.95	0.95	0.95	**0.96**
IDV(14)	**1**	**1**	**1**	**1**	0.82	**1**
IDV(15)	0.08	0.09	0.08	0.17	0.16	**0.24**
IDV(16)	0.31	0.47	0.21	0.57	0.43	**0.60**
IDV(17)	0.80	**0.96**	**0.96**	0.89	0.78	**0.96**
IDV(18)	0.90	**0.91**	0.90	0.90	0.89	**0.91**
IDV(19)	0.15	0.29	0.04	0.14	0.04	**0.39**
IDV(20)	0.43	0.60	0.53	0.61	0.43	**0.65**
IDV(21)	0.38	**0.58**	0.43	0.44	0.34	0.51

of the proposed PPA-based method. Fault 15 is a typical small fault, which is associated with the sticking of the condenser cooling water valve. As discussed in [10], traditional statistics are hard to detect this type of fault. Fig. 8 shows the process monitoring charts of Fault 15 using PCA-based, KPCA-based, and PPA-based methods. Compared to PCA-based and KPCA-based methods, the proposed PPA-based method is more sensitive to this fault, since the changes of PPA-T^2 and PPA-Q statistics are much more significant than those of PCA-based and KPCA-based statistics. Moreover, the proposed PPA-Q statistic can detect the fault earlier than the other monitoring statistics as shown in Fig. 8. The FDR of the proposed PPA-Q statistic for Fault 15 is the highest among all the monitoring statistics as shown in Table 4. In addition, the improvement of fault detection performance of the proposed PPA-based method for other small faults (i.e. Faults 3 and 9) is also significant as shown in Table 4. For Fault 20, the monitoring results using PCA-based, KPCA-based, and PPA-based methods are presented in Fig. 9. Fault 20 is a type of unknown fault in the TE process [10]. From Fig. 9, it can be clearly observed that the proposed PPA-based method is more sensitive than the standard PCA-based and KPCA-based methods, since the changes of PPA-T^2 and PPA-Q statistics are much more significant than those of PCA-based and KPCA-based statistics. The FDR of the proposed PPA-Q statistic for Fault 20 is higher than the other monitoring statistics as shown in Table 4. Both the quantitative comparison given in Table 4 and the monitoring charts of Figs. 8-9 demonstrated that the proposed PPA-based method performed better for most of the faults of the TE process than the other monitoring statistics.

For fault diagnosis, Faults 4 and 10 were used to test the proposed PPA-based contribution plots technique. Its application performance was compared with the standard PCA-based contribution plots. KPCA-based contribution plots was not presented since the intractable kernel transformation makes it practically not possible to calculate variable contribution via direct reverse projection [17]. Fault 4 involves a step change in the reactor cooling water inlet temperature. As illustrated in [10], the most distinct impact of Fault 4 is to cause a step change in the reactor cooling water flow rate (X32). Meanwhile, Fault 4 results in a sudden increase in the reactor temperature (X9) when the fault occurs. Fig. 10 shows the comparison of the normalized variable contributions using PCA-based and PPA-based diagnosis methods for Fault 4. It can be observed that variable 9 (X9) and variable 32 (X32) were identified as the most closely related variables to Fault 4 by the PCA-T^2 contribution statistic, as shown in Fig. 10(a). However, the PCA-Q contribution statistic failed to identify the most closely related process variables to Fault 4 as shown in Fig. 10(b). In comparison, the proposed PPA-T^2 and PPA-Q contribution statistics correctly identified variable 9 (X9) and variable 32 (X32) as the most closely related to Fault 4, as shown in Figs. 10(c) and 10(d). Furthermore, it is interesting to notice that variables 9 and 32 combined have a higher contribution to the proposed PPA-T^2 statistic (91%) as

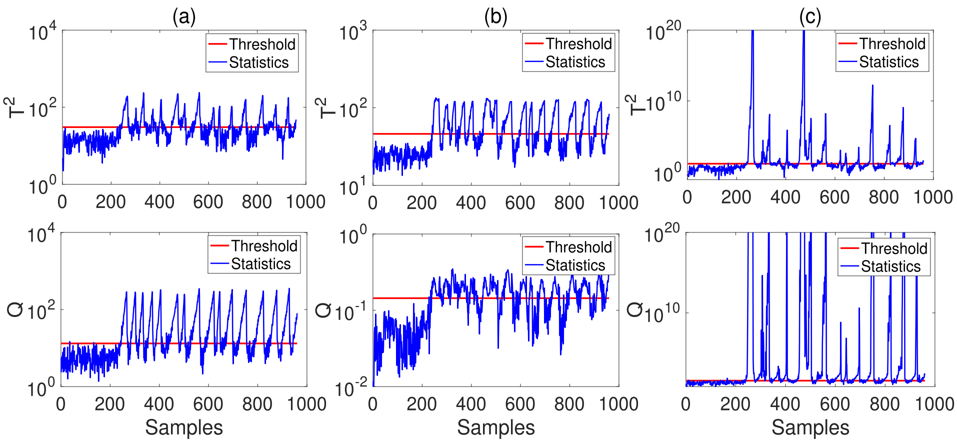

FIGURE 9. Process monitoring results of fault 20 by (a) PCA, (b) KPCA, and (c) PPA.

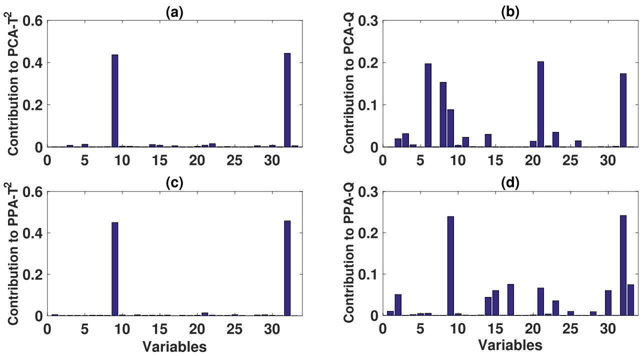

FIGURE 10. Variables contribution of fault 4 to (a) PCA-T^2, (b) PCA-Q, (c) PPA-T^2, and (d) PPA-Q.

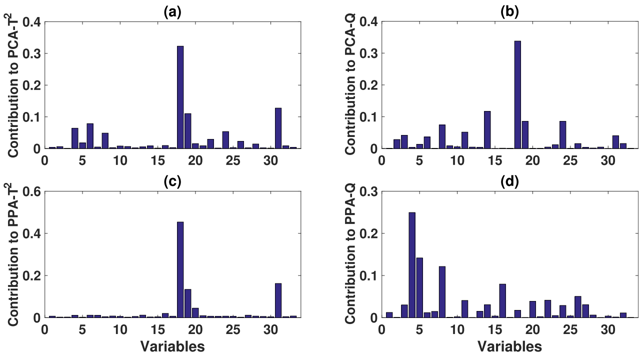

FIGURE 11. Variables contribution of fault 10 to (a) PCA-T^2, (b) PCA-Q, (c) PPA-T^2, and (d) PPA-Q.

compared to PCA-T^2 (88%). Similarly, variables 9 and 32 jointly contribute 48% to the proposed PPA-Q statistic whereas it is only 26% for the case of PCA-Q statistic. To further demonstrate the identification performance of the proposed PPA-based contribution statistics, they were applied to Fault 10. Fault 10 involves an abnormal variation in the C feed temperature (X4). Once the fault is introduced, it directly affects the downstream unit of product stripper and thus results in an abnormal change in the stripper temperature (X18) [28]. The variable contributions to PCA-based and PPA-based statistics for Fault 10 are shown in Fig. 11. As shown in Figs. 11(a) and 11(b), variable 18 (X18) was identified as the most likely related variable to Fault 10 by the PCA-T^2 and PCA-Q contribution statistics. However, both the PCA-T^2 and PCA-Q contribution statistics failed to locate the faulty variable 4 (X4). By comparison, the proposed PPA-T^2 and PPA-Q contribution statistics jointly succeeded in identifying variable 4 (X4) and variable 18 (X18) as the most likely related variables to Fault 10, as shown in Figs. 11(c) and 11(d). The identification results of Fault 10 have further consolidated the feasibility of the proposed PPA-based diagnosis statistics. Consequently, the proposed PPA-based diagnosis statistics performed better than the PCA-based diagnosis statistics in identifying the most closely related variables to the faults.

V. CONCLUSIONS

In this paper, a novel fault detection and diagnosis technique based on principal polynomial analysis is developed. The proposed PPA-based method is a promising nonlinear monitoring technique, which has the following advantages: (1) compared to the PCA-based methods, PPA is more effective in handling process nonlinearity; (2) unlike other nonlinear approaches such as kernel-based and neural networks-based methods, PPA has the appealing features of straightforward out-of-sample extension, volume-preservation, and invertibility; (3) PPA implements a simple and efficient sequential learning algorithm to extract the nonlinear features, and thus it does not require to specify the number of reduced dimensionality prior to modeling as compared to the nonlinear mainfold learning methods. The effectiveness and advantages of the proposed PPA-based monitoring method were verified through its applications to a nonlinear numerical example and a benchmark of Tennessee Eastman process. The application results have demonstrated that the proposed PPA-based monitoring method performed better than the standard PCA-based and KPCA-based methods. The future work is to further implement the proposed PPA-based approach in other nonlinear industrial processes.

REFERENCES

[1] Z. Geng and Q. Zhu, "Multiscale nonlinear principal component analysis (NLPCA) and its application for chemical process monitoring," *Ind. Eng. Chem. Res.*, vol. 44, no. 10, pp. 3585–3593, May 2005.

[2] Q. Jiang and X. Yan, "Just-in-time reorganized PCA integrated with SVDD for chemical process monitoring," *AIChE J.*, vol. 60, no. 3, pp. 949–965, Mar. 2014.

[3] X. Zhang and M. Kano, "Soft-sensor reliability evaluation and y-analyzer fault identification with applications to vinyl acetate monomer (VAM) benchmark process," presented at the 56th Annu. Conf. Soc. Instrum. Control Eng. Jpn. (SICE), 2017.

[4] L. Luo, S. Bao, J. Mao, and D. Tang, "Monitoring batch processes using sparse parallel factor decomposition," *Ind. Eng. Chem. Res.*, vol. 56, no. 44, pp. 12682–12692, Oct. 2017.

[5] Y. Li and X. Zhang, "Variable moving windows based non-Gaussian dissimilarity analysis technique for batch processes fault detection and diagnosis," *Can. J. Chem. Eng.*, vol. 93, no. 4, pp. 689–707, Apr. 2015.

[6] X. Zhang, Y. Li, and M. Kano, "Quality prediction in complex batch processes with just-in-time learning model based on non-Gaussian dissimilarity measure," *Ind. Eng. Chem. Res.*, vol. 54, no. 31, pp. 7694–7705, Aug. 2015.

[7] M. Kano and Y. Nakagawa, "Data-based process monitoring, process control, and quality improvement: Recent developments and applications in steel industry," *Comput. Chem. Eng.*, vol. 32, nos. 1–2, pp. 12–24, 2008.

[8] Y. Pan, C. Yang, R. An, and Y. Sun, "Robust principal component pursuit for fault detection in a blast furnace process," *Ind. Eng. Chem. Res.*, vol. 57, no. 1, pp. 283–291, Dec. 2017.

[9] Y. Li and X. Zhang, "Diffusion maps based k-nearest-neighbor rule technique for semiconductor manufacturing process fault detection," *Chemometrics Intell. Lab. Syst.*, vol. 136, pp. 47–57, Aug. 2014.

[10] E. L. Russell, L. H. Chiang, and R. D. Braatz, "Principal component analysis," in *Data-Driven Methods for Fault Detection and Diagnosis in Chemical Processes*. London, U.K.: Springer, 2012, pp. 33–52.

[11] S. J. Qin, "Survey on data-driven industrial process monitoring and diagnosis," *Annu. Rev. Control*, vol. 36, no. 2, pp. 220–234, Dec. 2012.

[12] Z. Ge, Z. Song, and F. Gao, "Review of recent research on data-based process monitoring," *Ind. Eng. Chem. Res.*, vol. 52, no. 10, pp. 3543–3562, 2013.

[13] Z. Ge, Z. Song, S. X. Ding, and B. Huang, "Data mining and analytics in the process industry: The role of machine learning," *IEEE Access*, vol. 5, pp. 20590–20616, 2017.

[14] M. A. Kramer, "Nonlinear principal component analysis using autoassociative neural networks," *AIChE J.*, vol. 37, no. 2, pp. 233–243, Feb. 1991.

[15] W. W. Hsieh, "Nonlinear principal component analysis by neural networks," *Tellus A*, vol. 53, no. 5, pp. 599–615, Oct. 2001.

[16] R. Rosipal, M. Girolami, L. J. Trejo, and A. Cichocki, "Kernel PCA for feature extraction and de-noising in nonlinear regression," *Neural. Comput. Appl.*, vol. 10, no. 3, pp. 231–243, 2001.

[17] J.-M. Lee, C. K. Yoo, S. W. Choi, P. A. Vanrolleghem, and I.-B. Lee, "Nonlinear process monitoring using kernel principal component analysis," *Chem. Eng. Sci.*, vol. 59, no. 1, pp. 223–234, Jan. 2004.

[18] M. Belkin and P. Niyogi, "Laplacian eigenmaps for dimensionality reduction and data representation," *Neural Comput.*, vol. 15, no. 6, pp. 1373–1396, 2003.

[19] V. Laparra, S. Jiménez, D. Tuia, G. Camps-Valls, and J. Malo, "Principal polynomial analysis," *Int. J. Neural Syst.*, vol. 24, no. 7, p. 1440007, Nov. 2014.

[20] S. T. Roweis and L. K. Saul, "Nonlinear dimensionality reduction by locally linear embedding," *Science*, vol. 290, no. 5500, pp. 2323–2326, Dec. 2000.

[21] J. B. Tenenbaum, V. de Silva, and J. C. Langford, "A global geometric framework for nonlinear dimensionality reduction," *Science*, vol. 290, no. 5500, pp. 2319–2323, Dec. 2000.

[22] R. R. Coifman and S. Lafon, "Diffusion maps," *Appl. Comput. Harmon. Anal.*, vol. 21, no. 1, pp. 5–30, Jul. 2006.

[23] M. Belkin and P. Niyogi, "Laplacian eigenmaps and spectral techniques for embedding and clustering," in *Proc. Adv. Neural Inf. Process. Syst.*, 2002, pp. 585–591.

[24] S. Wold, K. Esbensen, and P. Geladi, "Principal component analysis," *Chemometrics Intell. Lab. Syst.*, vol. 2, nos. 1–3, pp. 37–52, 1987.

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

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

[27] V. Laparra, D. Tuia, S. Jiménez, G. Camps-Valls, and J. Malo, "Principal polynomial analysis for remote sensing data processing," presented at the Geosci. Remote Sens. Symp. (IGARSS), 2011.

[28] H. Yu, F. Khan, and V. Garaniya, "A sparse PCA for nonlinear fault diagnosis and robust feature discovery of industrial processes," *AIChE J.*, vol. 62, no. 5, pp. 1494–1513, May 2016.

[29] S. Gajjar, M. Kulahci, and A. Palazoglu, "Real-time fault detection and diagnosis using sparse principal component analysis," *J. Process Control*, vol. 67, pp. 112–128, Jul. 2017.

[30] 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.

[31] 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.

[32] Z. Ge and Z. Song, "Process monitoring based on independent component analysis-principal component analysis (ICA-PCA) and similarity factors," *Ind. Eng. Chem. Res.*, vol. 46, no. 7, pp. 2054–2063, Mar. 2007.

[33] 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, 2012.

[34] L. Jiang, Z. Song, Z. Ge, and J. Chen, "Robust self-supervised model and its application for fault detection," *Ind. Eng. Chem. Res.*, vol. 56, no. 26, pp. 7503–7515, Jun. 2017.

XINMIN ZHANG received the B.S. degree from Qingdao Agricultural University, China, in 2012, and the M.S. degree from the Shenyang University of Chemical Technology, China, in 2015. He is currently pursuing the Ph.D. degree in system science with Kyoto University, Japan. His research interests include fault detection and diagnosis, process data analysis, process control, and virtual sensing technology with application to industrial processes.

MANABU KANO received the bachelor's, master's, and Ph.D. degrees from the Department of Chemical Engineering, Kyoto University, in 1992, 1994, and 1999, respectively. From 1999 to 2000, he was a Visiting Scholar with The Ohio State University, USA. Since 1994, he has been an Instructor with Kyoto University, where he has been a Professor of systems science since 2012. He has conducted researches in collaboration with various industries, such as semiconductor, chemical, steel, and pharmaceutical. His research interests include process data analysis for quality improvement, process monitoring, and process control. He was a recipient of several awards, including the Best Paper Award, the Technology Award, the Pioneer Technology Award from the Society of Instrument and Control Engineers, the Instrumentation, Control and System Engineering Research Award from the Iron and Steel Institute of Japan, and the Research Award for Young Investigators from the Society of Chemical Engineers, Japan.

YUAN LI received the B.S. and M.S. degrees from the Institute of Chemical Technology, China, in 1986 and 2001, respectively, and the Ph.D. degree from Northeastern University, China, in 2004. She is currently a Professor with the Shenyang University of Chemical Technology, China. Her research interests include fault detection and diagnosis, statistical process control, and data-driven process monitoring and applications in chemical processes.