Korean J. Chem. Eng., 34(8), 2135-2146 (2017)
DOI: 10.1007/s11814-017-0119-9

pISSN: 0256-1115
eISSN: 1975-7220

Sparse probabilistic principal component analysis model for plant-wide process monitoring

Jing Zeng, Kangling Liu, Weiping Huang, and Jun Liang†

State Key Lab of Industrial Control Technology, Institute of Industrial Control Technology,
Zhejiang University, Hangzhou, 310027, P. R. China
(*Received 5 December 2016 • accepted 25 April 2017*)

Abstract−In the industrial monitoring process, probabilistic principal component analysis (PPCA) is a popular algorithm for reducing the dimension. However, the principal components (PCs) are not easy to interpret and its preserved number cannot be determined automatically. In this paper, we propose a sparse PPCA (SPPCA) to improve the interpretability by adding a prior and introducing sparsification to the loading matrix of PPCA. An expectation-maximization (EM) algorithm is used to obtain the parameters of the probabilistic formulation, and the dimensionality of the latent variable space can be automatically determined during the iterative process. With the sparse representation, a process monitoring strategy is then developed with the construction of several partial PPCA models. Case studies of SPPCA to a numerical case and Tennessee Eastman (TE) benchmark process demonstrate its feasibility and efficiency.

Keywords: Sparse Probabilistic Principal Component Analysis, Plant-wide Process Monitoring, Multi-block, Partial Models, Fault Detection

INTRODUCTION

In modern industrial process modeling and monitoring, data-based statistical process monitoring methods have gained much attention. Among them, principal component analysis (PCA) may be one of the most widely used method in the past two decades [1-6]. PCA conducts a linear projection of the observed variables from a high dimension to a low dimension in a deterministic perspective. However, in practical application, process data is always contaminated by noises. To address this issue, a probabilistic framework of PCA, namely PPCA, is proposed by Tipping and Bishop [7] and used for process monitoring by Kim and Lee [8]. PPCA and its extensions [9-11] make it possible to catch the noise information simultaneously in industrial application.

Both PCA and PPCA capture the largest information in the first few principal components (PCs), and each PC is a linear projection of all the original variables. However, these two method have two limitations: (i) The linear combination of all measured variables makes it difficult to interpret and explain the physical meaning of the PC; (ii) The number of the reserved PCs needs to be known beforehand, but it may be problematic if there is not enough data for cross validation.

To deal with the interpretation problem, a specific norm can be adopted as a constraint on the loading vectors to gain a few nonzero elements. Thus, the PCs are much easier to interpret. Vines [12] considered to restrict the loading vectors to take values from a simple set of integers such as 0, 1, and −1. Jelliffe et al. [13] proposed a method named SCoLASSO which selects nonzero elements by adding an L1 norm regularizer on the loading vectors in the linear regression as an extension of PCA. Zou et al. [14] introduced a sparse PCA (SPCA) which adds both L1 and L2 regularization terms to the PCA least squares error objective as a penalty in a regression framework. Xie et al. [15] proposed a shrinking PCA with an L1 norm and used it for fault detection and isolation.

The same as PCA, sparse PCA also studies in a deterministic perspective and can be extended to a probabilistic model to deal with the contamination of the noises. Besides, the probabilistic model can overcome the limitation of sparse PCA that the sparsity of the loading matrix cannot be determined automatically. The probabilistic formulation of sparse PCA can handle these problems in model selection. The study in probabilistic models for sparse classification was proposed by Tipping [16] by introducing a probabilistic Bayesian learning framework for obtaining the sparse weight in Relevance Vector Machine (RVM). Sigg and Buhmann [17] achieved sparsification by adding a L1 norm constraint on the loading matrix during the expectation maximization (EM) process. While in Cawley et al.'s [18] work, a Laplace prior was introduced to obtain a sparse multinomial logistic regression method. Archambeau and Bach [19] extended it by the imposing of appropriate prior distributions to a generative model. Guan and Dy [20] further gave a sparse version of PPCA by the applying of a two-level hierarchical decomposition with a Laplace distribution, an inverse-Gaussian prior and a Jeffrey's prior, which is utilized in image processing and human face recognition. Koyejo et al. [22] studied the distribution of latent variables with sparse support, while Khanna et al. [22] further considered a sparse submodular probabilistic PCA. Latouche et al. [23] addressed a generative model with a spike-and-slab-like prior and Bouveyron et al. [24] extended it to a globally sparse perspective for PPCA.

However, the Laplace prior is not a conjugate prior for the Gauss-

†To whom correspondence should be addressed.
E-mail: jliang@iipc.zju.edu.cn
Copyright by The Korean Institute of Chemical Engineers.

ian distribution, and the computation will be complex with the utilization of two-level hierarchical decomposition. Besides, it is still a problem as the retained number of the PCs cannot be decided automatically. Inspired by the sparse Bayesian Learning [16] and Bayesian regularization [11] in PPCA, we propose another probabilistic formulation of sparse PCA, namely sparse probabilistic PCA (SPPCA). To achieve the sparsification, we introduce a Gaussian prior to the elements of the loading vectors, and solve the formulation of the model by the application of the EM algorithm. By the introduction of the hyperparameters, SPPCA can automatically control the sparsity of the loading matrix and the dimension of the latent variable space at the same time.

To our best knowledge, the sparse probabilistic PCA has not been used for process monitoring. In this paper, we further extend the SPPCA method into the process monitoring area by dividing the input network into several sub-blocks with the obtaining of the sparse loading matrix. For small-scale process, it is efficient to use a global model for fault prediction assuming that all the measurements are sent to a central location. However, if we obtain the measurements from a plant-wide process having many variables, the process monitoring will become a complicated problem and the results can be hard to interpret. Multiblock approach can reduce the complexity of process analysis and give a better detection performance [26-28]. However, most of these methods require process knowledge for block division, which is not always available, especially in complex chemical processes. The proposed SPPCA method can be constructed without a knowledge of the process. The sparsification of loading vectors decomposes the input network into potentially overlapping blocks and allows the natural definition of a set of partial monitoring models which are sensitive to specific process faults.

The rest of this paper is organized as follows. In section 2, we review the principle of the traditional PPCA monitoring method and discuss its deficiency. Then we propose our SPPCA method in section 3 and provide a detailed demonstration. This is followed by the use of SPPCA in process monitoring in the next section. In section 5, the studies on a numerical case and the Tennessee Eastman (TE) benchmark case are provided to evaluate the efficiency of the proposed method. Finally, a summary of the paper is made.

PROCESS MONITORING BASED ON PPCA MODEL

PPCA method was first proposed by Tipping and Bishop [7] as a maximum likelihood solution to a latent variable model, and it was used for process monitoring by Kim and Lee [8]. This section presents a summary of the PPCA method.

The formulation of PPCA method to each data sample can be given as:

$$x=Wt+e \tag{1}$$

where $x\in R^d$ is the process measured variable, $t\in R^q$ is the latent variable, $W\in R^{d\times q}$ is the loading matrix, $e\in R^d$ is a noise term with zero mean and variance $\sigma^2 I_d$. Both the latent variable t and the noise e are assumed to be isotropic and follows Gaussian distributions: $p(t)=N(0, I_q)$ and $p(e)=N(0, \sigma^2 I_d)$. Therefore, the conditional distribution of x given t is also Gaussian, $p(x|t)=N(Wt, \sigma^2 I_d)$. The marginal distribution of measured variable x can be calculated as follow:

$$p(x|W, \sigma^2)=\int p(x|t, W, \sigma^2)p(t)dt \tag{2}$$

Given a process dataset $X=\{x_n\}$, n=1, …, N with N data samples, then the parameters W and σ^2 can be determined by computing the maximum-likelihood estimator with the model in Eq. (1):

$$L(W, \sigma^2)=\ln\prod_{n=1}^{N} p(x_n|W, \sigma^2) \tag{3}$$

The PPCA model reserves q PCs to capture the largest information of the process dataset X and guarantees minimal information loss and reconstruction error. However, in PPCA model, the elements of loading matrix are almost nonzero and each PC is a linear combination of all the original measurements. This makes it difficult to give a physical interpretation to each PCs. Besides, the dimension of the latent space q is assumed to be known beforehand or needed to be calculated by cross validation. If there is not enough data for cross validation, it will become problematic to determine the dimensionality of the latent space. In the next section, a sparse PPCA method is proposed with a Bayesian regularization method which can automatically determine the latent space dimensionality.

SPARSE PPCA (SPPCA) MODEL CALIBRATION

In this section, the SPPCA model based monitoring method is proposed and demonstrated, including the introduction of the sparsity to the loading matrix and the utilization of EM algorithm in sparse modeling of PPCA.

1. Model Specification

The dimensionality and the sparsity of the model can be determined by imposing additional constraints on the parameters, which can be handled by a Bayesian method. Depending on the PPCA formulation in Eq. (1), we introduce a prior distribution over the loading matrix W by adopting a less complex function, the traditional Gaussian distribution [16]:

$$p(W|\alpha)=\prod_{i=1}^{d}\prod_{j=1}^{q} N(w_{ij}|0, \alpha_{ij}^{-1}) \tag{4}$$

where w_{ij} is the element of the loading matrix W on the *i*th row and *j*th column; q is the dimension of the latent space and is initialized to its maximum value q=d; α is a d×q hyperparameter matrix with $\alpha=\{\alpha_{ij}\}$, i=1, …, d, j=1, …, q, and its hyperpriors are defined over it.

$$p(\alpha)=\prod_{i=1}^{d}\prod_{j=1}^{q} Gamma(\alpha_{ij}|a, b) \tag{5}$$

where

$$Gamma(\alpha|a, b)=\Gamma(a)^{-1} b^a \alpha^{a-1} e^{-b\alpha}$$

with $\Gamma(a)=\int_0^\infty t^{a-1}e^{-t}dt$ is the gamma function. By marginalizing out the hyperparameter a_{ij}, we can get the prior of w_{ij}:

$$p(w_{ij})=\int p(w_{ij}|\alpha_{ij})p(\alpha_{ij})d\alpha_{ij}=\frac{b^a\Gamma(a+1/2)}{(2\pi)^{1/2}\Gamma(a)}(b+w_{ij}^2/2)^{-(a+1/2)} \tag{6}$$

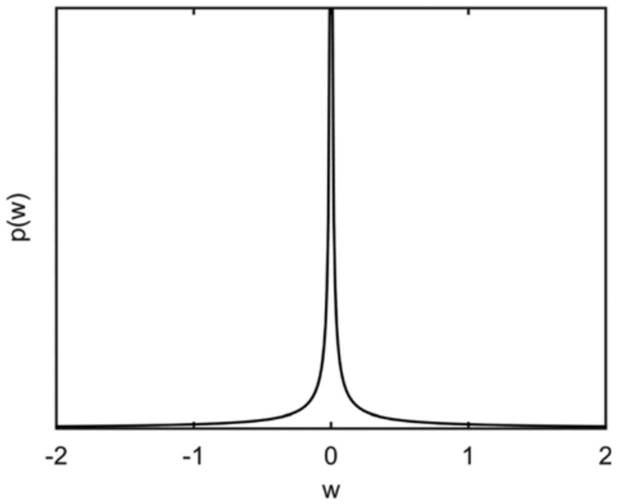

Fig. 1. The improper prior.

where $\Gamma(\cdot)$ a gamma function as it is defined above. However, by setting parameters a=b=0, we can get a uniform hyperprior (over a logarithmic scale). Then the improper prior $p(w_{ij})\propto 1/|w_{ij}|$ is got, whose distribution is shown in Fig. 1. Observe that this prior is sharply peaked at zero and gives high probabilities to values close to zero, which leads to the sparsification of the loading matrix.

The adoption of the hyperparameter matrix introduces sparsity to the loading matrix W. During the calculation, the Gaussian prior has advantage over Laplacian prior for it does not need to calculate the parameter of the hyperprior, and EM algorithm can be used since the prior is conjugated for a Gaussian distribution. Moreover if all the elements in the jth column of W are zero, that is $w_{ij}=0$, i=1, …, d, the corresponding column $W_{\cdot j}$ can be removed from the latent space loading matrix.

2. EM Algorithm for SPPCA

After adopting a prior distribution over loading matrix W, the parameter set is renewed as $\theta=\{W, \sigma^2, \alpha\}$. The optimal values of the parameters can be determined by using the EM algorithm. Then given the process data set $X\in R^{d\times N}$, the complete log-likelihood $L(\theta)$ is defined as [29]:

$$L(\theta)=\ln p(\theta|X)=\ln\frac{p(X|W)p(W)}{\int p(X, W)dW} \\ =L(W, \sigma^2)+\ln p(W)+const \tag{7}$$

where "*const*" represents a constant value in the function. With the distribution of each term the log-likelihood can be further defined as:

$$L(\theta)=\sum_{n=1}^{N}\Big(-\frac{d}{2}\ln\sigma^2-\frac{1}{2\sigma^2}(x_n-Wt_n)^T(x_n-Wt_n)\Big) \\ +\sum_{i=1}^{d}\sum_{j=1}^{q}\Big(\frac{1}{2}\ln\alpha_{ij}-\frac{\alpha_{ij}}{2}w_{ij}^2\Big)+\eta \tag{8}$$

In Eq. (8), η signifies θ independent terms in the function and can be omitted for simplicity. Now the EM algorithm can be performed as follows. In the E step, the parameter set θ_{old}, obtained in the previous M step, is given. The expectation of log-likelihood $L(\theta)$ is computed with respect to the conditional joint distribution of the latent variables, and it can be defined as:

$$Q(\theta|\theta_{old})=E(L(\theta)|X, \theta_{old}) \tag{9}$$

Then in the M step, the parameters are updated to maximize Eq. (9).

In the E step, the posterior distribution for the unobserved latent variables is evaluated, given the observed variables and the current values of the parameters.

$$p(t_n|x_n, \theta_{old})=\frac{p(x_n|t_n, \theta_{old})p(t_n|\theta_{old})}{p(x_n|\theta_{old})} \tag{10}$$

where $p(x_n|t_n, \theta_{old})=N(Wt_n, \sigma^2 I_d)$ and $p(t_n+\theta_{old})=N(0, I_q)$. For x and t are joint-distributed, the expected sufficient statistics of the latent variables can be calculated as:

$$var(t_n|x_n)=var(t_n)-cov(t_n, x_n)\,var(x_n)^{-1}cov(x_n, t_n)=(\sigma^{-2}W^TW+I_d)^{-1} \tag{11}$$

$$E(t_n|x_n)=E(t_n)+cov(t_n, x_n)\,var(x_n)^{-1}[x_n-E(x_n)] \\ =\sigma^{-2}(\sigma^{-2}W^TW+I_d)^{-1}W^Tx_n \tag{12}$$

By combining Eq. (11) and Eq. (12), we can get:

$$E(t_nt_n^T|x_n)=cov(t_n)+E(t_n)E^T(t_n) \tag{13}$$

Using $E(t_n|x_n)$ and $E(t_nt_n^T|x_n)$ obtained in the E step, the M step adjusts the model parameters by maximizing Eq. (9). Each parameter is updated in its turn while the others are fixed. First, considering the loading matrix W, the optimal problem of Eq. (9) can be decomposed into d linear least-squares problems. Note that $W_{i\cdot}$ (i=1, …, d) is the ith row of W. Taking the derivative of each subproblem with respect to $W_{i\cdot}$ and setting it to zero gives:

$$W_{i\cdot, new}=\sigma^{-2}\Bigg[\sum_{n=1}^{N}x_{ni}E^T(t_n|x_n)\Bigg]A_i\Bigg\{\sigma^{-2}\Bigg[\sum_{n=1}^{N}E(t_nt_n^T|x_n)\Bigg]A_i+I_q\Bigg\}^{-1} \tag{14}$$

where x_{ni} is the ith element in the nth row of X; $A_i=diag(1/\alpha_{ij}$, j=1, …, q) is a diagonal matrix. Similar to the calculation of W, take the deviation of Eq. (9) with respect to σ^2 and set it to zero, then we get:

$$\sigma_{new}^2=\frac{\sum_{n=1}^{N}\{x_n^Tx_n-2x_n^TWE(t_n|x_n)+Tr[W^TWE(t_nt_n^T|x_n)]\}}{Nd} \tag{15}$$

where $Tr(\cdot)$ is an operator for trace value calculation. Unlike in SPCA where the parameter α needs to be decided before hand, in SPPCA we can automatically adjust the hyperparameter through a type II maximum likelihood [30]. Then we get:

$$\alpha_{ij, new}=\frac{1}{w_{ij}^2} \tag{16}$$

The EM algorithm for adjusting the parameters of SPPCA is completed. The whole procedure can be summarized as follows: first, we initialize the parameters, then we update the parameters with the E step and M step, which are recursively repeated until the convergence criterion is satisfied. Specifically, after each M step, all elements in each column of α are checked in order to decide whether the corresponding column of W is removed or not. Finally, the optimal values of W, σ^2 and α are determined. In this paper, we initialize W using conventional PCA and set σ^2 to 1.

Till now, the SPPCA model is constructed. Compared with the traditional PPCA model, the proposed SPPCA method can automatically determine the effective number of the PCs. What's more, the method gives a sparse loading matrix, which may make it convenient to interpret each PC.

PROCESS MONITORING BASED ON SPPCA

So far, the SPPCA model has been constructed and the parameters have been determined. In this section, two statistics, T^2 statistic and SPE statistic, are applied for fault detection and a process monitoring strategy is proposed.

1. Monitoring Statistics

With the SPPCA algorithm proposed above, we obtained sparse loading vectors W with q sparse components $W=[w_1, w_2, \ldots, w_q]$. Each component defines a block, and the non-zero elements of w_j ($j=1, 2, \ldots, q$) define the variables of the jth block.

For each block, given a measured variable x, let us denote x^j as the process measured variable for the jth block. A separate PPCA partial model is constructed as follows:

$$x^j=P_j t_j+e_j \tag{17}$$

where t_j, P_j and e_j, respectively, stand for the latent variable, the loading matrix and the noise of the jth partial model.

Meanwhile, a global PPCA model is constructed based on the measure variable x:

$$x=P_g t_g+e_g \tag{18}$$

where t_g, P_g and e_g, respectively, stand for the latent variable, the loading matrix and the noise of the global model.

Given a new data sample x_{new}, it is first projected by the global model and the partial models, respectively, and the latent variables can be estimated as:

$$t_{g,\,new}=P_g^T(P_gP_g^T+\sigma_g^2 I_d)^{-1}x_{new}=Q_g x_{new} \tag{19}$$

$$t_{j,\,new}=P_j^T(P_jP_j^T+\sigma_j^2 I)^{-1}x_{new}^j=Q_j x_{new}^j \tag{20}$$

where $Q_g=P_g^T(P_gP_g^T+\sigma_g^2 I)^{-1}$ and $Q_j=P_j^T(P_jP_j^T+\sigma_j^j I)^{-1}$. And the error terms of the new sample can be estimated as:

$$e_{g,\,new}=x_{new}-P_g t_{g,\,new}=[I_d-P_gP_g^T(P_gP_g^T+\sigma_g^2 I_d)^{-1}]x_{new} \tag{21}$$

$$e_{j,\,new}=x_{new}^j-P_j t_{j,\,new}=[I-P_jP_j^T(P_jP_j^T+\sigma_j^2 I)^{-1}]x_{new}^j \tag{22}$$

Therefore, a total of q+1 pairs of T^2 and SPE statistics are calculated and compared to their corresponding confidence limits, respectively, as:

$$T_{g,\,new}^2=t_{g,\,new}^T(\mathrm{var}(t_{g,\,new}))^{-1}t_{g,\,new}\le T_{g,\,lim}^2 \tag{23}$$

$$SPE_{g,\,new}=e_{g,\,new}^T e_{g,\,new}\le SPE_{g,\,lim} \tag{24}$$

$$T_{j,\,new}^2=t_{j,\,new}^T(\mathrm{var}(t_{j,\,new}))^{-1}t_{j,\,new}\le T_{j,\,lim}^2 \tag{25}$$

$$SPE_{j,\,new}=e_{j,\,new}^T e_{j,\,new}\le SPE_{j,\,lim} \tag{26}$$

where $\mathrm{var}(t_{g,new})=Q_g(P_gP_g^T+\sigma_g^2 I_d)Q_g^T$ and $\mathrm{var}(t_{j,new})=Q_j(P_jP_j^T+\sigma_j^2 I)Q_j^T$ are the estimated variance of the global and partial latent variables, respectively. The confidence limit of T^2 and SPE statistics can be estimated using kernel density estimation (KDE) [29-31] with a Gaussian kernel function and a confidence level of γ using the training data.

2. Monitoring Strategy

In the above analysis, we proposed a sparse probabilistic method and decomposed the global model into q partial PPCA models based on the sparse loading matrix. The process monitoring strategy can be summarized as follows:

(i) Obtain the normal process dataset $X=\{x_n\}$, $n=1, \ldots, N$ and perform the SPPCA method on it; obtain the sparse loading matrix W with q sparse components $W=[w_1, w_2, \ldots, w_q]$;

(ii) The original input network is decomposed into q potentially overlapping partial blocks, which are obtained based on a small number of nonzero elements in the corresponding sparse loading vector w_j;

(iii) Perform PPCA in each block and the original network to get q partial PPCA models and a global PPCA model; calculate the confidence limits of the T^2 and SPE statistics for the global model and each partial model;

(iv) Given a new data sample x_{new}, calculate T^2 and SPE statistics for the q partial models and the global model; if any of these q+1 sets of statistics violate the confidence limits, this measurement is reported to be out of control and such sample is a faulty sample.

For the sparsification of the loading matrix, the strongest correlations among variables are preserved. The partial models usually consist of fewer variables than the global PPCA model, which makes it easy to interpret the relationship between different variables. Besides, by inspecting which partial model gives an out-of-control report, one can analyze the influence of a fault and find out its root cause preliminarily.

CASE STUDIES

In this section, two case studies are provided to evaluate the efficiency and the reliability of the proposed method. The first one is a numerical case with ten variables which verifies the proposed SPPCA model can give a sparse solution. The other one is an industrial TE process monitoring case to evaluate the performance of the proposed method.

1. Numerical Case

This numerical case consist of ten measurement variables, which are driven by two latent variables. The two latent variables are Gaussian distributed and equally important. The first four measurement variables are related to the first latent variable t_1, while the next four measurement variables are the observations of the second latent variable t_2. The relationships between them are given by:

$$x=At+e$$
$$A=\begin{bmatrix}1&1&1&1&0&0&0&0&0&0\\0&0&0&0&1&1&1&1&0&0\end{bmatrix}^T \tag{27}$$
$$t=\begin{bmatrix}t_1\\t_2\end{bmatrix}\approx N\left\{\begin{bmatrix}0\\0\end{bmatrix},\begin{bmatrix}1&0\\0&0.98\end{bmatrix}\right\}$$

where $x\in R^{10}$ is the measurement variables; $t\in R^2$ is the latent variables; A is the loading matrix; e is a zero-mean isotropic Gaussian process noise with variance $\Sigma_e=diag\{[0.09\ 0.09\ 0.09\ 0.09\ 0.25\ 0.25\ 0.25\ 0.25\ 0.16\ 0.16]\}$. To build the SPPCA model and PPCA model, 200 samples are generated as the reference data set and used for model construction. To compare the monitoring efficiency of the proposed method and the traditional PPCA model, a fault is simulated by adding a small step bias of 1.2 to the first variable after the 50th sample.

First, an SPPCA model and a PPCA model are built based on the reference data. To clearly show the sparsification ability of the proposed method, the three-dimensional exhibitions of the loading matrices for both SPPCA and PPCA methods are shown in Fig. 2(a) and Fig. 2(b), respectively. As shown in Fig. 2(a), two PCs are automatically reserved in the SPPCA model and they are only composed of the associated measurement variables of the latent variables t_1 and t_2, respectively. This matches the loading matrix A in Eq. (27) and demonstrates the sparsification ability of our proposed SPPCA model. Besides, since the first latent variable t_1 has a larger variance, PC1 is composed of the first four measurement variables, while PC2 is associated with the next four variables whose relative latent variable has a smaller variance. The other elements in the loading matrix are all zero and the noise variables are removed from the loading matrix. In contrast, the number of retained PCs in PPCA model is hard to determine from the result shown in Fig. 2(b). And each PC is composed of all the measurement variables, which makes it difficult to give an interpretation to the PC.

Next, following the monitoring strategy proposed in section 4, ten measured variables are decomposed into two blocks based on the two retained PCs in the SPPCA model, and two partial PPCA models are constructed with variable sets $\{x_1, x_2, x_3, x_4\}$ and $\{x_5, x_6, x_7, x_8\}$, respectively, based on the reference data. Besides, the global PPCA model, which was built before using traditional PPCA method, retains four PCs that explains over 90% of the total variance in the reference data set. To evaluate the monitoring performance of the proposed method, the generated fault case is tested. Fig. 3 shows the monitoring results of the global PPCA model and the partial PPCA model based on the first PC in SPPCA. The confidence levels of both monitoring statistics are set as 99%, as the dashed line shows in the figure. The other partial PPCA could hardly detect the bias in the first variable and its monitoring result is not shown here. It can be seen that the simulated fault can be detected by both models, but the performance of the partial model is outstanding. In Fig. 3(a), after the fault was introduced, only 13.91% present of the portion of the SPE values were over the control limit

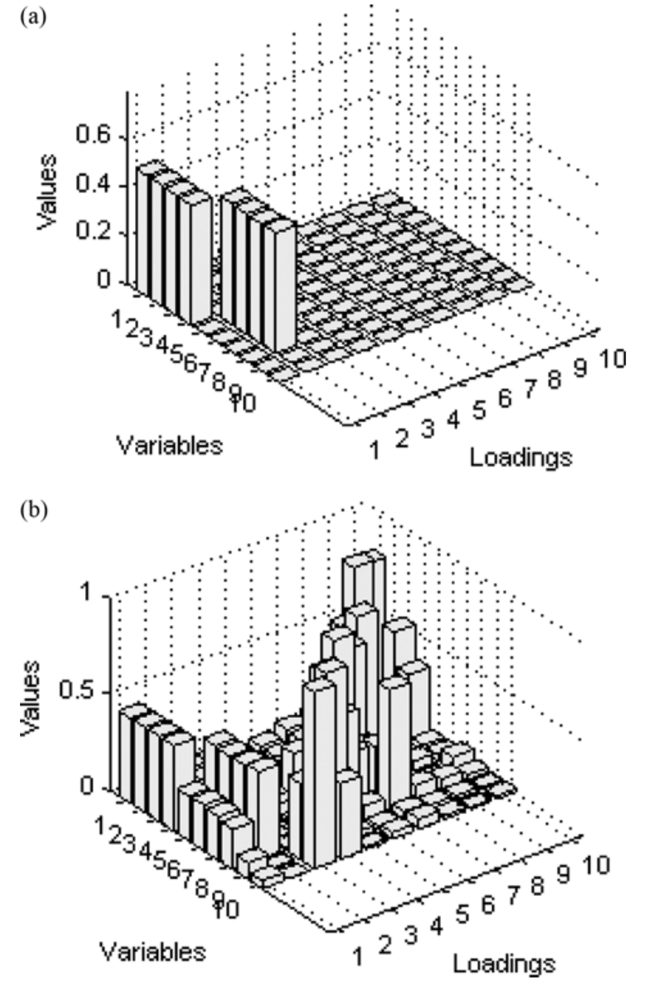

Fig. 2. Three-dimensional plots of the absolute element values of the loading matrix in: (a) SPPCA (b) PPCA.

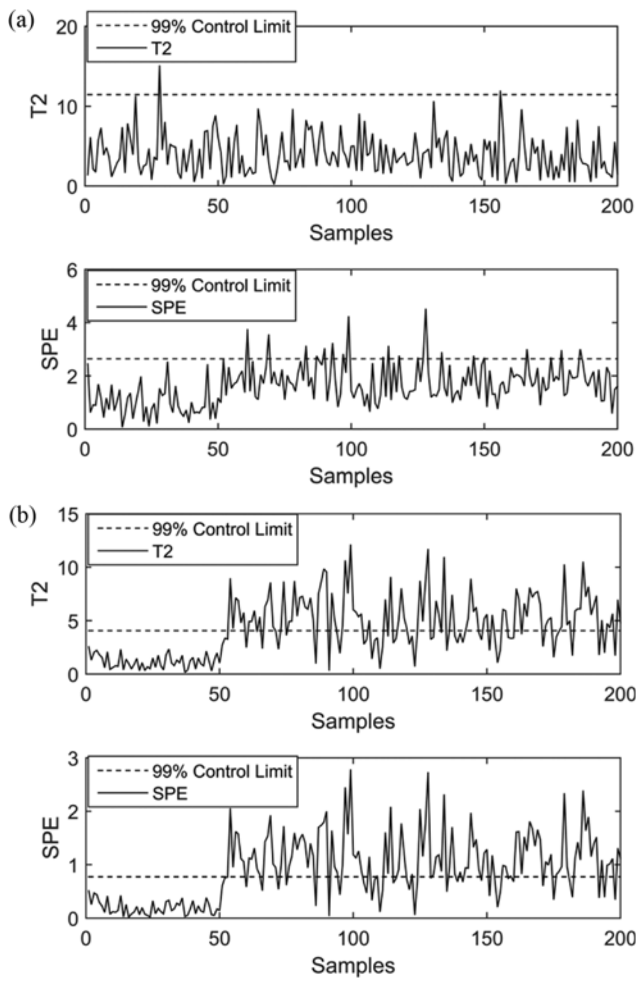

Fig. 3. Monitoring results for the fault by: (a) PPCA (b) SPPCA.

in the traditional PPCA model. While in Fig. 3(b), the detection rate of T^2 and SPE statistics was 63.58% and 72.85%, respectively in this partial model. Although, in our simulated works, the detection rates of both models rise if we choose another more significant bias, Fig. 3 demonstrates the proposed partial models have a better performance when the bias of measurement variable is not that large. Besides we can preliminarily figure out that the root cause is in the variable set $\{x_1, x_2, x_3, x_4\}$.

This case confirms that compared with the traditional PPCA method the proposed SPPCA method can improve the monitoring performance. The proposed method improves the interpretability of the constructed model and automatically determines the retained number of PCs during the model construction. The efficiency and the reliability is further elaborated in the next section with the TE process.

2. TE Process

The proposed SPPCA monitoring approach is evaluated on the TE process data set in this section. TE process is a widely used simu-

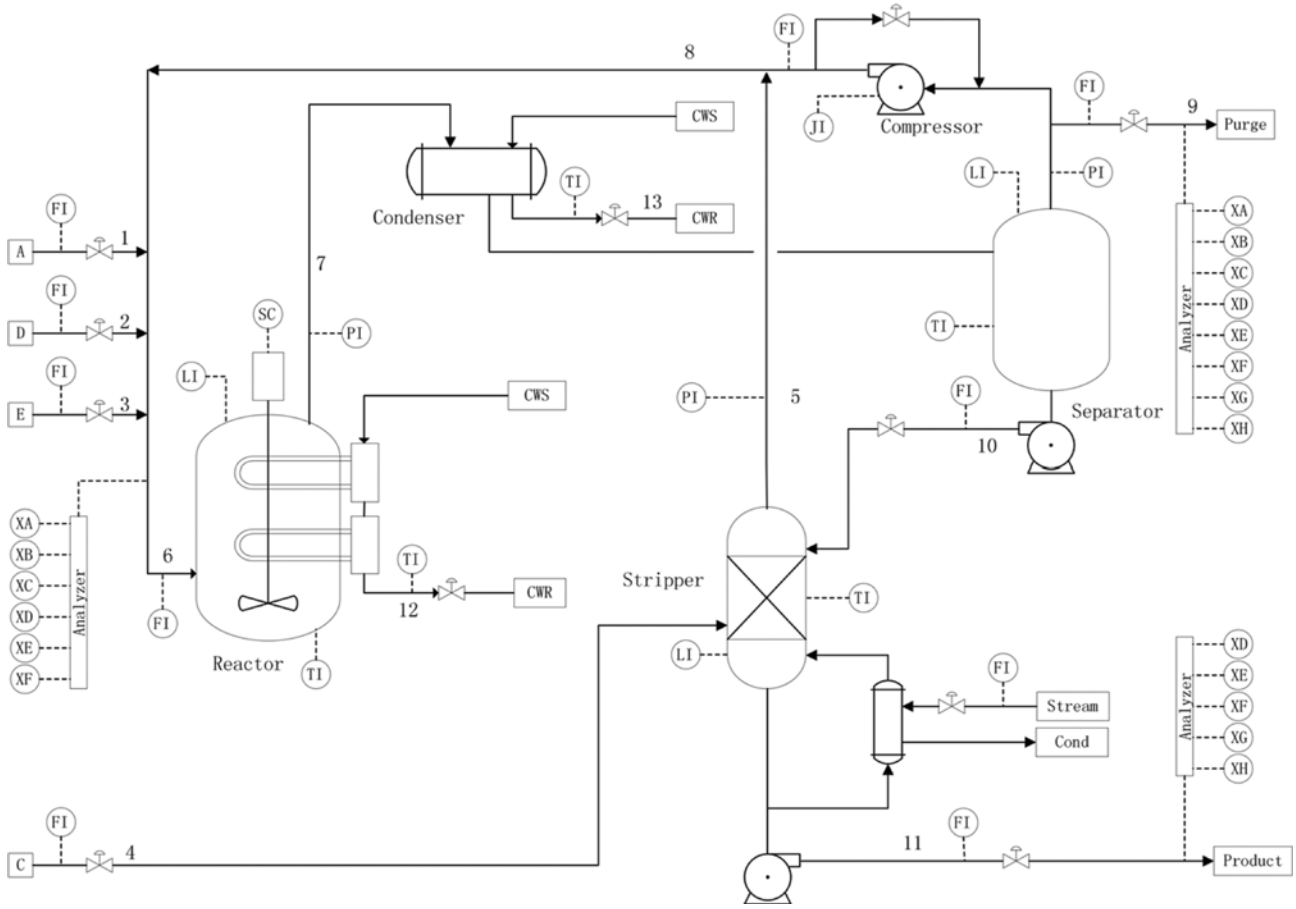

Fig. 4. Diagram of TE process.

Table 1. Monitoring variables in the TE process

Xmeas no.	Measurement variables	Xmeas no.	Measurement variables	Xmv no.	Manipulated variables
1	A feed	12	Product separator level	1	Feed D flow valve
2	D feed	13	Product separator pressure	2	Feed E flow valve
3	E feed	14	Product separator underflow	3	Feed A flow valve
4	Total feed	15	Stripper level	4	Total feed flow valve
5	Recycle flow	16	Stripper pressure	5	Compressor recycle valve
6	Reactor feed rate	17	Stripper underflow	6	Purge valve
7	Reactor pressure	18	Stripper temp.	7	Separator pot liquid flow valve
8	Reactor level	19	Stripper steam flow	8	Stripper liquid product flow valve
9	Reactor temp.	20	Compressor work	9	Stripper steam valve
10	Purge rate	21	RCW outlet temp.	10	RCW flow
11	Product separator temp.	22	SCW outlet temp.	11	CCW flow

lated industrial process for testing the performance of different control and monitoring strategies, and it has become a benchmark process for different process control and monitoring approaches [34]. The TE process produces two products and a byproduct from four reactants, and it involves five major unit operation units: the reactor, the product condenser, a vapor-liquid separator, a recycle compressor and a product stripper. Fig. 4 shows a diagram of the process.

The process has a total of 53 variables, including 22 continuous process measurements, 19 sampled process measurements and 11 process manipulated variables. In this section, the research only focuses on the 22 continuous process measurements and 11 manipulated variables, which are listed in Table 1. Note that RCW is the abbreviation of reactor cooling water, CCW is abbreviated for condenser cooling water, SCW stands for separator cooling water. The remaining 19 sampled process measurements are ignored.

To develop the monitoring model, a normal data set containing N=960 data samples was generated. And 21 types of faulty dataset, including 15 known fault condition and 6 unknown fault conditions, were generated. While faults 3, 9, 15 are very subtle faults that are hard to detect, they are not considered in the evaluation. The other 18 faults are described in Table 2. Each fault was monitored for 960 sample intervals. During the generations, the operating conditions were kept constant and started with no fault, and the faults were introduced into the faulty data sets at the 161th sample.

The normal SPPCA monitoring model was constructed based on the normal data set using the SPPCA approach proposed in section 3. In this experiment, we initialized the loading matrix W with the conventional PCA algorithm, and set δ^2 and all the elements in the hyperparameter matrix α to be 1. We set our condition for convergence tolerance to be 10^{-3}. With the proposed method, the number of the retained PCs and the sparsity of the loading matrix can be decided automatically. The loading matrix gained

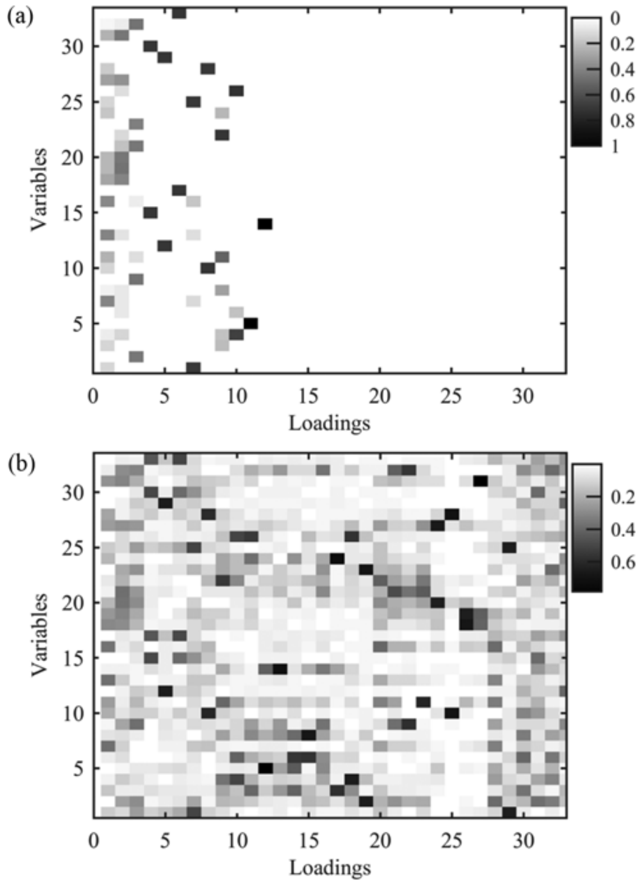

Fig. 5. Loading matrix of (a) SPPCA (b) PPCA.

from the SPPCA model is shown in Fig. 5(a). We can see that a total of 12 sparse components are preserved, while the other 21

Table 2. TE process fault description

Fault no.	Description	Type
1	A/C feed ratio, B composition constant (stream4)	Step change
2	A/C feed ratio, B composition constant (stream4)	Step change
4	RCW inlet temperature	Step change
5	CCW inlet temperature	Step change
6	A feed loss (stream1)	Step change
7	C header pressure loss-reduced availability (stream4)	Step change
8	A, B, C feed composition (stream4)	Random variation
10	C feed temperature (stream2)	Random variation
11	RCW inlet temperature	Random variation
12	CCW inlet temperature	Random variation
13	reaction kinetics	Random variation
14	RCW valve	Random variation
16	Unknown	Unknown
17	Unknown	Unknown
18	Unknown	Unknown
19	Unknown	Unknown
20	Unknown	Unknown
21	Valve position constant (stream 4)	Constant position

Table 3. Partial models constructed by SPPCA

Partial model (PM) no.	Variables
1	*xmeas*7, *xmeas*11, *xmeas*13, *xmeas*16, *xmeas*18, *xmeas*19, *xmv*5, *xmv*9
2	*xmeas*4, *xmeas*18, *xmeas*19, *xmeas*20, *xmeas*21, *xmeas*22, *xmv*5, *xmv*9
3	*xmeas*2, *xmeas*9, *xmeas*11, *xmeas*16, *xmeas*21, *xmv*1, *xmv*10
4	*xmeas*15, *xmv*8
5	*xmeas*12, *xmv*7
6	*xmeas*17, *xmv*11
7	*xmeas*1, *xmeas*7, *xmeas*13, *xmeas*16, *xmv*3
8	*xmeas*10, *xmv*6
9	*xmeas*3, *xmeas*4, *xmeas*8, *xmeas*11, *xmeas*22, *xmv*2
10	*xmeas*4, *xmeas*6, *xmv*4
11	*xmeas*5
12	*xmeas*14

columns are eliminated for the elements in their corresponding columns of the hyperparameter matrix are all infinite. For comparison, a PPCA model is also constructed and the nonzero elements in its loading matrix are shown in Fig. 5(b). From which we can see that each PC is nearly the composition of all the variables. And we cannot determine the number of the preserved PCs from the figure. It is clear that SPPCA has a better performance on the sparsification of the loading matrix.

Using the SPPCA method, a sparse representation of the loading matrix was achieved. As the strategy proposed in section 4, the input network can be decomposed into 12 overlapping sub-blocks, which are used for constructing the partial PPCA models for the process monitoring. The values of the loading element indicate the degree of relationship between the corresponding variable and PC [14]. Considering the variables' correlation with the PC, we give a further selection to reduce the redundancy and improve the conciseness and efficiency of the partial model. If the number of nonezero elements in a loading vector was more than 8, we preserve the first eight most correlative variables with the first-eight-largest elements in the block to build the corresponding partial model [35]. Table 3 gives the resulting variable distribution of each partial PPCA model. Note that PM is an abbreviation of partial model, *xmeas* stands for the continuous process measurements and *xmv* denotes the process manipulated variables. The first and the second partial model were built on the first eight most correlative variables, while the number of the remaining models' variables were less than eight. Besides, a global PPCA model was also constructed, which retained 14 principal components explaining over 85% of the total variance.

Till now, we have 12 partial models which are easy to interpret. As shown in Table 3, the partial models are closely related to the pro-

Table 4. Monitoring results of the TE process

Fault no.	PPCA		SPPCA	
	T^2	SPE	T^2	SPE
1	0.9900	0.9975	0.9988	0.9988
2	0.9825	0.9600	0.9863	0.9738
4	0.1810	0.9988	0.9988	0.9988
5	0.2347	0.2459	0.9988	0.9988
6	0.9900	0.9988	0.9988	0.9988
7	0.9988	0.9988	0.9988	0.9988
8	0.9663	0.8664	0.9813	0.9838
10	0.2971	0.2996	0.8702	0.8539
11	0.3845	0.7640	0.8914	0.8614
12	0.9825	0.9076	0.9988	1.0000
13	0.9351	0.9513	0.9588	0.9650
14	0.9900	0.9988	0.9988	0.9988
16	0.1361	0.3283	0.9376	0.9376
17	0.7615	0.9588	0.9713	0.9738
18	0.8914	0.9014	0.9151	0.9201
19	0.0861	0.1823	0.9563	0.9700
20	0.3171	0.5194	0.8826	0.9001
21	0.3845	0.4919	0.6167	0.6380

Table 5. Monitoring results of partial PPCA models

Fault no.	Detected by partial PPCA models
1	**1**, **2**, 3, 6, **7**, 8, **9**, **10**
2	**1**, **2**, **3**, **6**, 7, **8**, **9**, **10**
4	3
5	1, 2, 3, **6**, 7, 8, 9, 10
6	**1**, **2**, **3**, **6**, **7**, **8**, **9**, **10**
7	1, 2, 3, 6, 7, 8, 9, **10**
8	**1**, **2**, **3**, 6, **7**, **8**, **9**, 10
10	**1**, **2**, 3, 6, 7, 8, 9, 10
11	2, **3**, 9
12	**1**, **2**, **3**, **6**, **7**, 8, **9**, 10
13	**1**, **2**, **3**, **6**, **7**, 8, **9**, 10
14	**2**, **3**, 7
16	**1**, **2**, 3, 6, 7, 9, 10
17	1, **2**, **3**, 6, 9
18	**1**, **2**, **3**, **6**, **7**, **8**, **9**, **10**
19	**1**, 2, 3, 7, 11
20	1, 2, 3, 6, 7, 8, 9, 10
21	1, 2, 3, 6, 7, 9

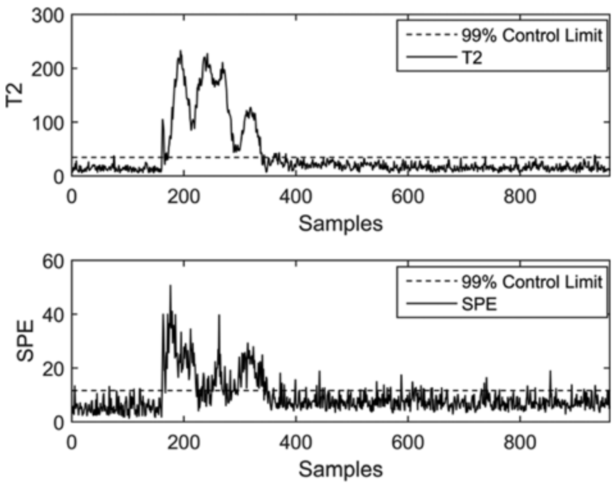

Fig. 6. Monitoring result of PPCA model for Fault 5.

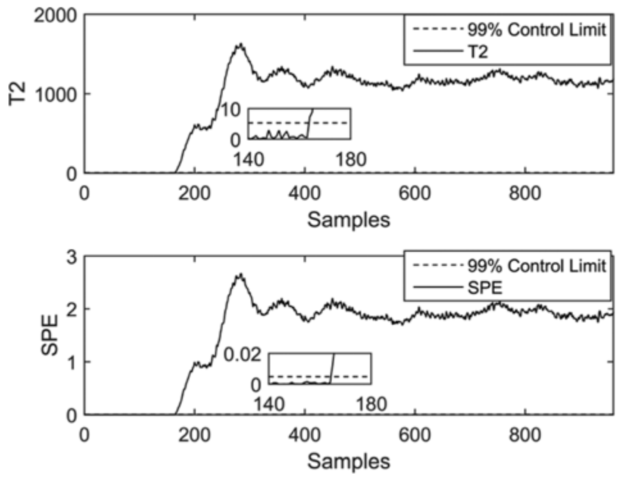

Fig. 7. Monitoring result of PM6 for Fault 5.

cess mechanism or the controller effects. For example, PM1 exhibits the connection of the pressure between different units. PM5 shows that the product separator level (*xmeas*12) is affected by the separator pot liquid flow valve (*xmv*7) as a result of feedback control. And PM6 shows that the condenser cooling water flow (*xmv*11) has great influence on the stripper underflow (*xmeas*17).

For the purpose of fault detection, the proposed strategy was implemented on the 18 faults, and the confidence level of both T^2 and SPE statistics was set as 99%. Both T^2 and SPE statistics of 12 partial models and the global model are calculated for each sample. If any of these 13 statistics violate the control limit, the corresponding sample is reported to be faulty. For comparison, the traditional PPCA model was also used for detection, which, the same as the global PPCA model, retained 14 principal components explaining over 85% of the total variance. The monitoring results of these two method are illustrated in Table 4, and the detection rate of both T^2 and SPE statistics is calculated based on the reported number of faulty samples and the total number. It is shown

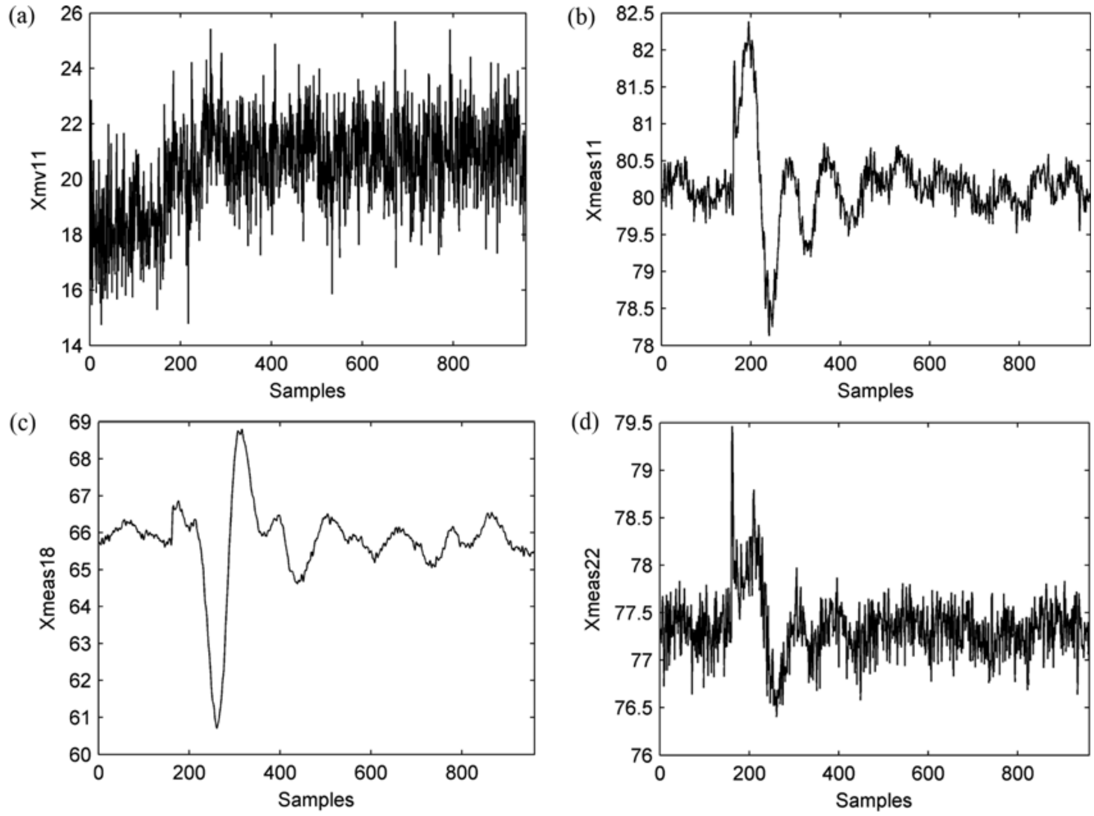

Fig. 8. Data characteristics of four variables in Fault 5: (a) *xmv*11 (b) *xmeas*11 (c) *xmeas*18 (d) *xmeas*22.

that proposed strategy outperforms the conventional PPCA method in most of the cases, and all of its detection rates are over 50% in both T^2 and SPE statistics. Especially, the detection rates of the proposed method for Fault 5, Fault 10, Fault 11, Fault 16, Fault 19 and Fault 20 are much higher than those of PPCA.

Detailed information of the detection results of each partial model is shown in Table 5. Note that the italic numbers highlight that the corresponding partial models had a good performance on the faults. It is that for each partial model at least one of its statistics detection rate is over 80%, or at least one of its T^2 and SPE statistics keeps violating the confidence limit. It can be figured out from this table that different partial models have different performance on different faults. Some faults can only be detected by few partial models, such as Fault 4, for these faults affect few variables. Some faults have plant-wide effect on many variables and can be detected by most of partial models with different performance, such as fault 1 and fault 2. With Table 5, we can give a simple isolation to each fault, since each partial model is sensitive to a specific process fault. Following the discussion in Ge and Song's paper [36], to further evaluate the monitoring performance of the SPPCA method, two typical process faults (fault 5 and fault 10) are chosen and discussed in the following text.

Fault 5 is a step change of the condenser cooling water inlet temperature, and it causes a step change in the flow rate of the outlet stream of the condenser, which can affect the temperature in the separator as well as the cooling water outlet temperature of the separator.

The monitoring results of PPCA are shown in Fig. 6. Fault 5 can be easily detected at the 161th sample, but it was soon compensated by the control system and its SPE values went beyond the confidence limit at 349th sample. It seems like the fault has been corrected in about 10h. However, with the proposed monitoring strategy, we can get a totally different judgment. In Fig. 7, we can figure out that Fault 5 keeps affecting the process from the monitoring results of PM6. Both T^2 and SPE values keep violating the confidence limit even after 349th sample without any going down. And the detection rates of both T^2 and SPE are over 99%. Based on the variables in PM6, we can see the difference between these two monitoring results happens because of the adjustment of the manipulated variable *xvm*11 (CCW flow), whose performance is shown in Fig. 8(a). For the temperature of the condenser cooling water inlet temperature kept above its normal value, the value of *xvm*11 increased after fault 5 happened and stabilized around a point higher than its normal value. Meanwhile, the other measured variables, such as *xmeas*11 (product separator temperature), *xmeas*18 (stripper temperature), and *xmeas*22 (SCW outlet temperature), whose performance is shown in Fig. 8(b), (c), and (d), affected by the fault were pulled back to their normal values. Thus, the compensation of the manipulated variable *xvm*11 makes the values of SPE and T^2 go back to normal and gives a low detection rate of PPCA. This may mislead the operator to believe that the fault is corrected. In contrast, this fault is easily detected and illustrated by PM6, and the high detection rate of PM6 may keep informing the operator of the existence of the fault.

Fault 10 is a random variation of feed C temperature in stream 4, which may cause a change to the condition of stripper and condenser. The monitoring results of PPCA and partial models are shown in Fig. 9 and Fig. 10, respectively. Fault 10 can be detected by PPCA model with a low detection rate (0.2971 for T^2 statistic

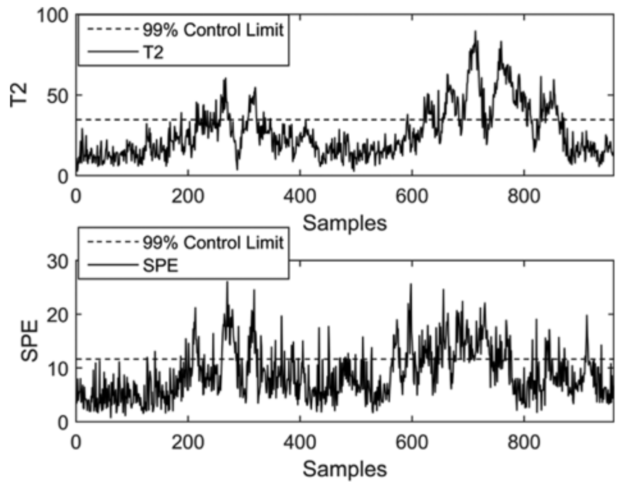

Fig. 9. Monitoring result of PPCA model for Fault 10.

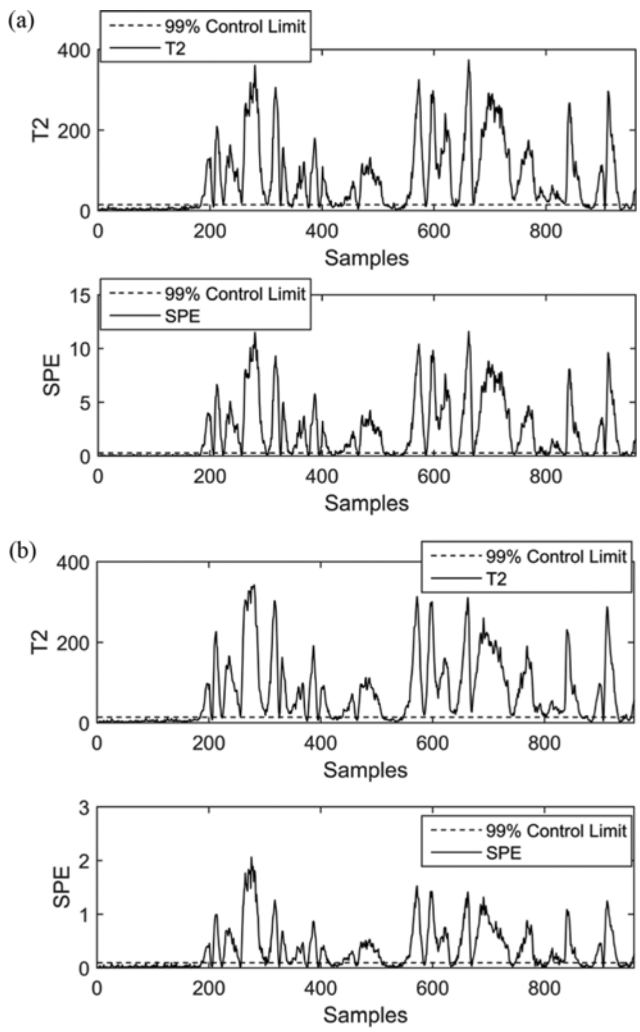

Fig. 10. Monitoring result of PMs for Fault 10: (a) PM1 (b) PM2.

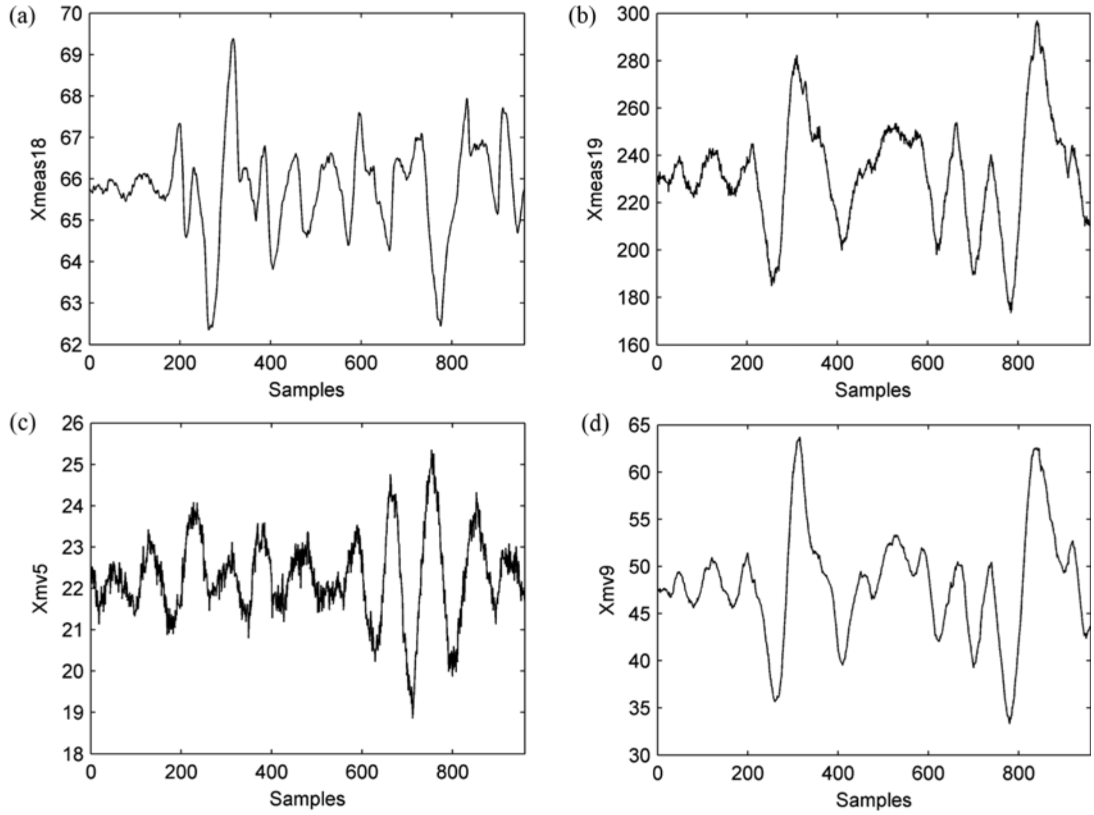

Fig. 11. Data characteristics of four variables in Fault 10: (a) *xmeas*18 (b) *xmeas*19 (c) *xmv*5 (d) *xmv*9.

and 0.2996 for SPE statistic). Besides, considering T^2 statistic, PPCA can hardly detect the fault between 350th sample and 600th sample. On the other hand, from the monitoring results of PM1 and PM2 shown in Fig. 10(a) and (b), respectively, it is obvious that these two partial models have a much better performance on fault 10 than the PPCA model, and over 70% present of the portions of T^2 and SPE values are violating the control limits in both partial models. The detection rate is 0.8414 for T^2 statistic and 0.7241 for SPE statistic in PM1, the detection rate is 0.8227 for T^2 statistic and 0.7865 for SPE statistic in PM2. Besides, it can be found in Table 3 that these two partial models share the four same variables, *xmeas*18 (stripper temperature), *xmeas*19 (stripper steam flow), *xmv*5 (compressor recycle valve) and *xmv*9 (stripper steam valve). This can be explain as that the variation of feed C temperature in stream 4 influenced the temperature in the stripper *xmeas18*, which affected the pressure in the stripper and the compressor recycle valve *xmv*5 was manipulated for compensation. The stripper steam valve *xmv*9 was also manipulated to adjust the temperature in the stripper, which influenced the stripper steam flow *xmeas*19. The performance of these four variables is shown in Fig. 11, and their values give large fluctuation after the fault added into the process. Thus, these four variables can be regarded as the root cause of the fault.

CONCLUSIONS

A sparse model has been proposed for the traditional PPCA method. The sparsification of the model is achieved by adding a Gaussian prior to the loading matrix W, and it gives a better interpretability of the PCs compared to the traditional PPCA method. The parameters are obtained by an EM algorithm, and the sparsity of the loading matrix and the dimensionality of the latent variable space can be automatically determined during the iterative process. Then a monitoring strategy has been proposed based on the loading matrix by constructing partial PPCA model in each sub-block. This strategy acquires no process knowledge and is evaluated to be more efficient than the traditional PPCA on the basis of the monitoring results on a numerical case and the TE benchmark process. Note that this method does a little work on the fault isolation. Our future work will focus on the combination of the monitoring results from different partial models. Based on the benefits of probabilistic model, a mixture model of sparse probabilistic PCA can also be developed for multi-mode application.

ACKNOWLEDGEMENT

This study was supported by the National Natural Science Foundation of China (U1664264, U1509203, 61174114).

NOMENCLATURE

a, b : parameters of the hyperprior
d : dimension of the process measured variable
e : noise variable $e \in R^d$

e_g : noise of the global model
e_j : noise of the *j*th partial model
I, I_d, I_q : identity matrix $I_d \in R^{d\times d}$, $I_q \in R^{q\times q}$
N : number of samples in dataset
P_j : loading matrix of the *j*th partial model
P_g : loading matrix of the global model
q : number of the reserved principal components
t : latent variable $t \in R^q$
t_g : latent variable of the global model
t_j : latent variable of the *j*th partial model
w_{ij} : element of the loading matrix W on the *i*th row and *j*th column
w_j : the *j*th loading vector in the loading matrix W
W : loading matrix $W \in R^{d\times q}$
$W_{i\cdot}$: the *i*th row of the loading matrix W
$W_{\cdot j}$: the *j*th column of the loading matrix W
x : Process measured variable $x \in R^d$
x^j : process measured variable for the *j*th block
x_{ni} : the *i*th element in the *n*th row of X
X : process dataset $X=\{x_n\} \in R^{d\times N}$, n=1, …, N
α : hyperparameter matrix $\alpha=\{\alpha_{ij}\} \in R^{d\times q}$, i=1, …, d, j=1, …, q
η : θ independent terms in the function
θ : the parameter set of the proposed method
σ^2 : the variance of the noise

REFERENCES

1. L. H. Chiang, E. L. Russell and R. D. Braatz, *Meas. Sci. Technol.*, **12**, 1745 (2001).
2. S. J Qin, *J. Chemom.*, **17**, 480 (2003).
3. Y. Zhan and S. J. Qin, *AIChE J.*, **54**, 3207 (2008).
4. X. Wang, U. Kruger, G. W. Irwin, G. McCullough and N. McDowell, *IEEE Trans. Control Syst. Technol.*, **16**, 122 (2008).
5. Y. Zhang and C. Ma, *Chem. Eng. Sci.*, **66**, 64 (2011).
6. Q. C. Jiang and X. F. *Korean J. Chem. Eng.*, **31**, 1935 (2014).
7. M. E. Tipping and C. M. Bishop, *J. R. Stat. Soc. Ser. B Statistical Methodol.*, **61**, 611 (1999).
8. D. Kim and I.-B. Lee, *Chemom. Intell. Lab. Syst.*, **67**, 109 (2003).
9. S. W. Choi, J. H. Park and I.-B. Lee, *Comput. Chem. Eng.*, **28**, 1377 (2004).
10. T. Chen and Y. Sun, *Control Eng. Pract.*, **17**, 469 (2009).
11. Z. Q. Ge and Z. H. Song, *AIChE J.*, **56**, 2838 (2010).
12. S. K. Vines, *J. R. Stat. Soc. Ser. C-Applied Stat.*, **49**, 441 (2000).
13. I. T. Jolliffe, N. T. Trendafilov and M. Uddin, *J. Comput. Graph. Stat.*, **12**, 531 (2003).
14. H. Zou, T. Hastie and R. Tibshirani, *J. Comput. Graph. Stat.*, **15**, 265 (2006).
15. L. Xie, X. Lin and J. Zeng, *Ind. Eng. Chem. Res.*, **52**, 17475 (2013).
16. M. E. Tipping, *J. Mach. Learn. Res.*, **1**, 211 (2001).
17. C. D. Sigg and J. M. Buhmann, *Proc. 25th Int. Conf. Mach. Learn. - ICML '08.*, 960 (2008).
18. G. Cawley, N. Talbot and M. Girolami, *NIPS*, **19**, 209 (2007).
19. C. Archambeau and F. R. Bach, *NIPS*, 1 (2008).
20. Y. Guan and J. G. Dy, *AISTATS*, **5** 185 (2009).
21. O. Koyejo, J. Ghosh, R. Khanna and R. A. Poldrack, *NIPS*, 676 (2014).
22. R. Khanna, J. Ghosh, R. Poldrack and O. O. Koyejo, *AISTATS*, **38**, 453 (2015).
23. P. Latouche, P. A. Mattei, C. Bouveyron and J. Chiquet, *J. Multivariate Anal.*, **146**, 177 (2014).
24. C. Bouveyron, P. Latouche and P. A. Mattei, *Bayesian variable selection for globally sparse probabilistic PCA*, Technical Report, HAL-01310409, Universite Paris Descartes (2016).
25. S. J. Qin, S. Valle and M. J. Piovoso, *J. Chemometr.*, **15**, 715 (2001).
26. S. W. Choi and I. B. Lee, *J. Process Contr.*, **15**, 295 (2005).
27. Y. Zhang, H. Zhou, S. J. Qin and T. Chai, *IEEE T. Ind. Inform.*, **6**, 3 (2010).
28. B. Wang, Q. C. Jiang and X. F. Yan, *Korean J. Chem. Eng.*, **31**, 930 (2014).
29. C. M. Bishop, *NIPS*, **11**, 382 (1998).
30. C. M. Bishop, *Springer-Verlag*, New York (2006).
31. E. B. Martin and A. J. Morris, *J. Process Contr.*, **6**, 349 (1996).
32. Q. Chen, R. J. Wynne, P. Goulding and D. Sandoz, *Control Eng. Pract.*, **8**, 531 (2000).
33. Q. Chen and U. Kruger and A. T. Y. Leung, *Control Eng. Pract.*, **12**, 267 (2004).
34. J. J. Downs and E. F. Vogel, *Comput. Chem. Eng.*, **17**, 245 (1993).
35. M. Grbovic, W. C. Li, P. Xu, A. K. Usadi, L. M. Song and S. Vucetic, *J. Process Contr.*, **22**, 738 (2012).
36. Z. Ge and Z. Song, *Ind. Eng. Chem. Res.*, **52**, 1947 (2013).