Improved Performance of Fault Detection Based on Selection of the Optimal Number of Principal Components

LI Yuan[1] TANG Xiao-Chu[1]

Abstract This paper presents a new method for selecting the number of principal components (PCs) in fault detection based on principal component analysis (PCA). On the basis of the proposed fault signal-to-noise ratio (SNR), the optimal number of PCs can be determined. SNR indicates the relationship between the sensitivity of fault detection and the number of PCs. By maximizing the fault SNR, the optimal number of PCs can be selected and the performance of fault detection can be improved. This method is applied to Tennessee Eastman process (TEP) and compared with the cumulative percent variance (CPV) method. The simulation results demonstrate its good Performance.

Key words Fault detection, fault signal-to-noise ratio (SNR), sensitivity, principal component (PC)

Process monitoring including fault detection and diagnosis based on multivariate statistical process control (MSPC) has been rapidly developed in recent decades[1−3]. One of the most popular MSPC methods is based on principal component analysis (PCA), which builds a PCA model using data measured during normal operating conditions. In this methodology, it is necessary to determine the dimension of the PCA model, that is, the number of principle components (PCs).

There are many methods for selecting the number of PCs[4−6], such as cumulative percent variance (CPV)[7−8], cross validation[9], variance of reconstruction error (VRE)[10], and so on. In the CPV method, the number of PCs is determined when the PCA model can express the main information of original data. The cross-validation method uses part of the training samples for model construction, and the remaining samples are compared with the predicted model. When the prediction residual sum of squares becomes smaller than the residual sum of squares of the previous model, the new component is added to the model. The VRE method presents that the corresponding PCs are deemed as the optimal PCs when the error of fault reconstruction comes to minimum.

In most of the traditional methods, the determination of the number of PCs seems to be based on the subjective idea, such as CPV, and none of these methods consider the performance of fault detection. In this paper, we introduce a new index to select the number of PCs; the new index is termed fault signal-to-noise ratio (SNR). Fault SNR reflects the relationship between the number of PCs and fault detection sensitivity, which measures the performance of fault detection[11]. In PCA method, SPE statistic and T^2 statistic are usually used for fault detection. It is well known that SPE statistic and T^2 statistic have different performances of fault detection for a certain fault, and both statistics show different behaviors with different numbers of PCs. Therefore, we define fault SNR for different statistics and easily determine the number of PCs that gives maximum sensitivity by examining fault SNR. In the definition of fault SNR, we consider not only fault detection sensitivity but also fault class. So, the proposed method based on fault SNR further improves the fault detection performance. This method is applied to Tennessee Eastman process (TEP)[12] for fault detection of sensor fault, and then compared with CPV method and the simulation results show the advantages of the proposed method.

This study is organized as follows. In Section 1, some background knowledge on PCA is discussed. In Section 2 the proposed method for selecting the number of PCs based on fault SNR is introduced. Finally, the simulation based on TEP is given compared with CPV method to illustrate the good performance of the proposed method. The final section is conclusion.

1 Principal component analysis

PCA is a statistical technique that transforms a correlated original data to uncorrelated data set that represents most of the information in original data. Let $X^0 \in \mathbf{R}^{n\times m}$ denote the original data matrix with n samples and m variables. In PCA method, the original data is first scaled to a matrix X with zero mean and unit variance. Then, based on a singular value decomposition (SVD) algorithm, the matrix X can be decomposed as follows:

$$X = TP^{\mathrm{T}} + E \tag{1}$$

where $T \in \mathbf{R}^{n\times l}$ and $P \in \mathbf{R}^{m\times l}$ are the score matrix and the loading matrix, respectively. The PCA transforms the original set of m variables to $a = l$ principal components. PCA can be regarded as a classical linear dimension reduction technique, and the number of PCs is commonly determined based on CPV method. The CPV method is introduced as follows:

$$\frac{\sum\limits_{i=1}^{a} \lambda_i}{\sum\limits_{i=1}^{m} \lambda_i} \times 100\% \geq 85\% \tag{2}$$

where λ_i is the variance of score vector. When CPV is larger than 85 %, the corresponding number of PCs is determined.

In PCA-based fault detection, statistics and their control limits should be established to determine whether a process is in control or not. Common statistics include SPE statistic, which indicates the degree of deviation of each sample from the model, and T^2 statistic, which reflects on variations with PCA model. The two statistics and corresponding control limits are given as follows:

Received October 28, 2008; in revised form April 13, 2009
Supported by National Natural Science Foundation of China (60774070), and Science Foundation of Liaoning Education Department (20060669, 2004D041)
1. College of Information Engineering, Shenyang Institute of Chemical Technology, Shenyang 110142, P. R. China
DOI: 10.3724/SP.J.1004.2009.01550

$$SPE = \|x_{\rm new}^{\rm T} - x_{\rm new}^{\rm T} P_a P_a^{\rm T}\|^2 = \|x_{\rm new}^{\rm T} \tilde{P}_a \tilde{P}_a^{\rm T}\|^2 \qquad (3)$$

where $x_{\rm new}$ is a new observed vector to be detected. P_a is a $m \times a$ matrix that consists of the first a columns of the loading matrix P and $\tilde{P}_a$ is a $m \times (m-a)$ matrix that consists of the remaining columns of P. The control limit is calculated[13].

$$SPE_{\rm limit} = \theta_1 \left[\frac{C_\alpha \sqrt{2\theta_2 h_0^2}}{\theta_1} + 1 + \frac{\theta_2 h_0 (h_0 - 1)}{\theta_1^2} \right]^{\frac{1}{h_0}} \qquad (4)$$

$$\theta_i = \sum_{j=a+1}^{n} \lambda_j^i (i = 1, 2, 3) \qquad (5)$$

$$h_0 = 1 - \frac{2\theta_1 \theta_3}{3\theta_2^2} \qquad (6)$$

where C_α is normal distribution value with level of significance α.

The T^2 statistic is calculated as

$$T^2 = x_{\rm new}^{\rm T} P_a S^{-1} P_a^{\rm T} x_{\rm new} \qquad (7)$$

where S is a diagonal matrix, which is the estimated covariance matrix of principal component scores. The control limit is calculated[14].

$$T_{a,n,\alpha}^2 = \frac{a(n-1)}{n-a} F_{a,n-1,\alpha} \qquad (8)$$

where $F_{a,n-1,\alpha}$ is an F-distribution with degree of freedom a and $n-1$ and with level of significance α. If the SPE statistic shows an unexpected large value, it means that the data points go into the residual subspace. This means that correlations between sensors that are observed during normal operating conditions are lost. When T^2 statistic shows unexpected values, it indicates deviation from normal values in PC subspace.

2 Selecting the optimal number of PCs based on fault SNR

2.1 Definition of fault SNR

If observed sensor encounters a fault $f\xi_i$, the sensor output is described as follows:

$$x = x^* + f\xi_i \qquad (9)$$

where x^* is an observation of sensor without the fault, f is a scalar value that indicates the magnitude of fault, and ξ_i is a fault direction vector that describes the effect of fault. Based on (9), the SPE statistic can be given as follows

$$\begin{aligned} SPE = & \|(x^* + f\xi_i)^{\rm T} - (x^* + f\xi_i)^{\rm T} P_a P_a^{\rm T}\|^2 = \\ & \|(x^* + f\xi_i)^{\rm T} \tilde{P}_a \tilde{P}_a^{\rm T}\|^2 = \|f\xi_i^{\rm T} \tilde{P}_a \tilde{P}_a^{\rm T}\|^2 \end{aligned} \qquad (10)$$

In (10), we allow $x^* = 0$ and $f = 1$. It is due to x^* that is normalized to zero mean and unit variance. Equation (10) shows that SPE is the squared norm of the projection of $f\xi_i$ onto the residual subspace, and it depends on the number of PCs and the direction of fault. Based on the facts described above, the fault SNR for SPE can be defined as follows:

$$SNR_{SPE} = \frac{\|\xi_i^{\rm T} \tilde{P}_a \tilde{P}_a^{\rm T}\|^2}{SPE_{\rm limit}} = \frac{\|\tilde{\xi_i}\|^2}{SPE_{\rm limit}} \qquad (11)$$

where $\tilde{\xi_i}$ is the projection of ξ_i onto the residual subspace. SNR_{SPE} is the ratio of SPE statistic to corresponding control limit, and this ratio can be regarded as the sensitivity. Therefore, fault SNR reflects on the sensitivity of fault detection. When $x^* = 0$, the fault SNR means the ratio of a projection of fault onto residual subspace to the control limit. So, it is reasonable to consider the numerator as a signal and control limit as fault SNR noise.

For fault detection using T^2-statistic, it is calculated as follows:

$$T^2 = (x^* + f\xi_i)^{\rm T} P_a S^{-1} P_a^{\rm T} (x^* + f\xi_i) \qquad (12)$$

The fault SNR can be defined in a similar way as follows:

$$\begin{aligned} SNR_{T^2} = & \frac{\|(x^* + f\xi_i)^{\rm T} P_a S^{-1} P_a^{\rm T} (x^* + f\xi_i)\|^2}{T_{\rm limit}^2} \\ & \frac{\|(\xi_i^{\rm T} P_a S^{-1} P_a^{\rm T} \xi_i)}{T_{\rm limit}^2} \end{aligned} \qquad (13)$$

By maximizing the fault SNR, the optimal number of PCs giving maximum sensitivity is determined. Equations (11) and (13) show that fault SNR depends on the number of PCs and fault direction. Therefore, the number of PCs that maximizes the fault SNR is related with fault direction. This indicates that fault detection based on only one PCA can result in error detection and misdetection if process exhibits different faults. However, determination of the number of PCs based on fault SNR considers not only the fault detection sensitivity but also the fault direction. A different statistic has different optimal number of PCs, which can improve the performance of fault detection.

2.2 Sensor fault

From the definition of fault SNR, it depends upon the kind of fault and the number of PCs. If we want to determine the number of PCs, we should firstly determine the fault direction. For sensor not included in control loops, the fault direction is determined as follows: if the fault occurs in the i-th sensor, the fault direction is given by (the i-th component is 1, and the others are 0).

$$\xi_i = [0, \cdots, 0, 1, 0, \cdots, 0] \qquad (14)$$

3 Application studies for Tennessee Eastman process

3.1 Process description

TEP simulator has been widely used as a bed of continuous process for optimization strategies, monitoring, and diagnosis. This process consists of five major units: a reactor, a product condenser, a separator, a recycle compressor, and a product stripper. TEP has 12 manipulated variables, 22 continuous process measurements, and 19 composition measurements. In this paper, we selected 16 variables from 22 continuous measured variables, which are listed in Table 1.

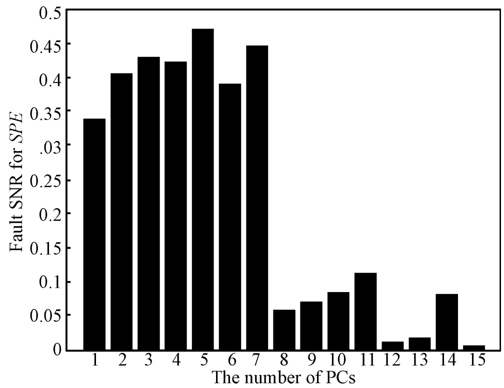
Table 1 Process variables for fault detection and diagnosis

Variable	Description	Variable	Description
1	A feed	9	Product separator temperature
2	D feed	10	Product separator pressure
3	E feed	11	Product separator underflow
4	A and C feed	12	Stripper pressure
5	Recycle flow	13	Stripper temperature
6	Reactor feed rate	14	Stripper steam flow
7	Reactor temperature	15	Reactor cooling water outlet temperature
8	Purge rate	16	Separator cooling water outlet temperature

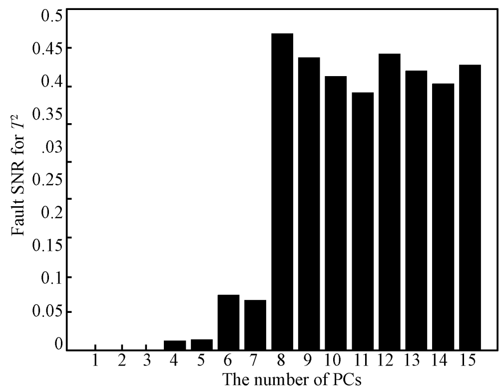
3.2 Selecting the number of PCs based on fault SNR

In this paper, the proposed method that selects the number of PCs based on fault SNR is illustrated using TEP simulation data. The simulation results compared with the CPV method demonstrate the superiority of the proposed method and improved performance in fault detection. The normal training data has 500 observations, and sampling interval is 3 min. Three sensor faults are introduced here: the reactor feed rate sensor fault, product separator pressure sensor fault, and separator cooling water outlet temperature, denoted by F1 $\sim$ F3, respectively. The fault occurs at the 201st sample (603rd min).

First, we determine the number of PCs based on CPV method. The result is given in Fig. 1, where we can see when the number of PCs is 10, the CPV is larger than 0.85. So, the number of PCs is 10 according to the CPV method. Then, we calculated the fault SNR for F1. The fault SNR for SPE statistic and T^2 statistic is shown in Fig. 2. Fig. 2 shows that the number of PCs that maximize the fault SNR for SPE statistic is 5, and the number of PCs that maximize the fault SNR for T^2 statistic is 8. According to the proposed method described in Section 2, the number of PCs that maximize the fault SNR can give maximum sensitivity of fault detection. Figs. 3 and 4 give plots of each statistic normalized to its control limits with different numbers of PCs. To conserve space, this paper gives all plots of statistic normalized to its control limit with $190 \sim 210$ samples. The plot illustrates that the number of PCs determined on the basis of fault SNR gives the maximum fault detection sensitivity.



(a) Fault SNR for SPE statistic



(b) Fault SNR for T^2 statistic

Fig. 2 Fault SNR for fault F1

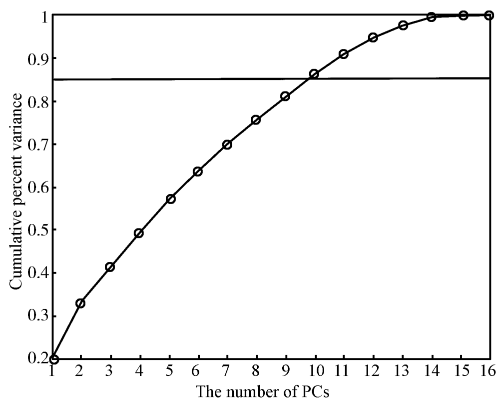
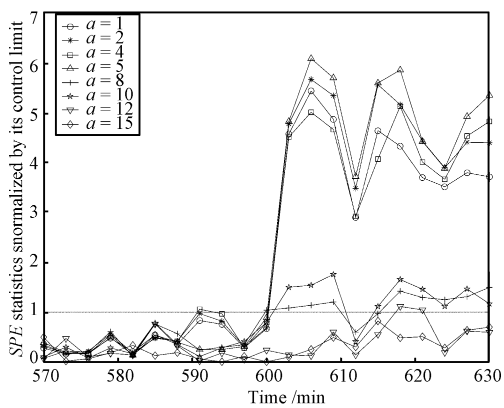


Fig. 1 The cumulative percent variance with the number of PCs



Figs. 3 SPE statistics normalized by its control limit for F1

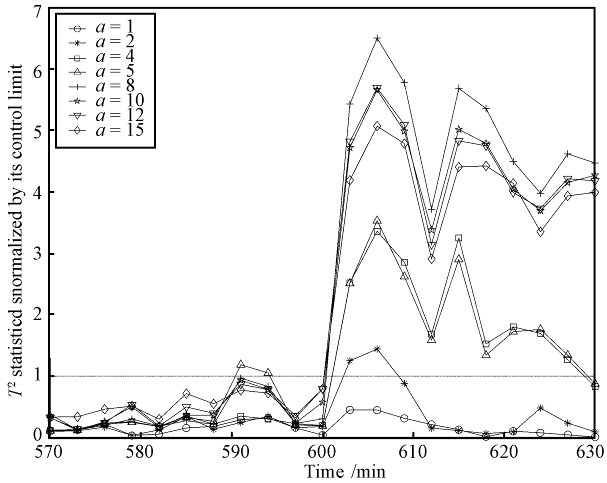
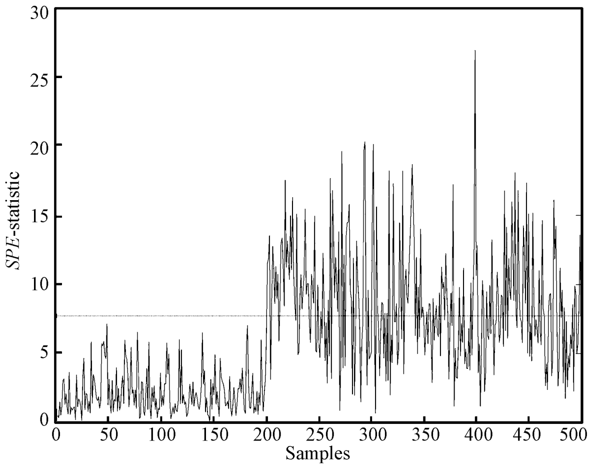
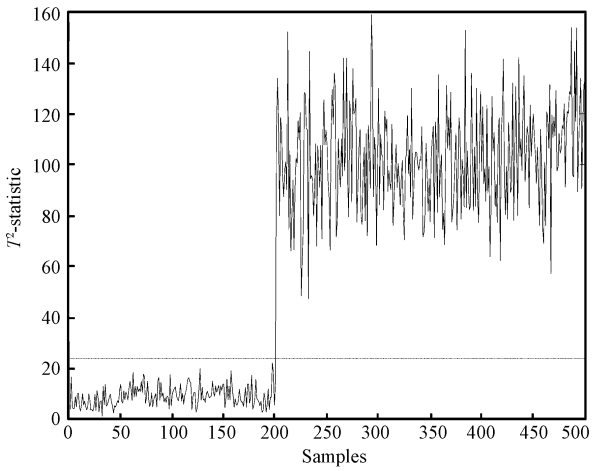


Fig. 4 T^2 statistic normalized to its control limit for F1



(a) SPE statistic plot

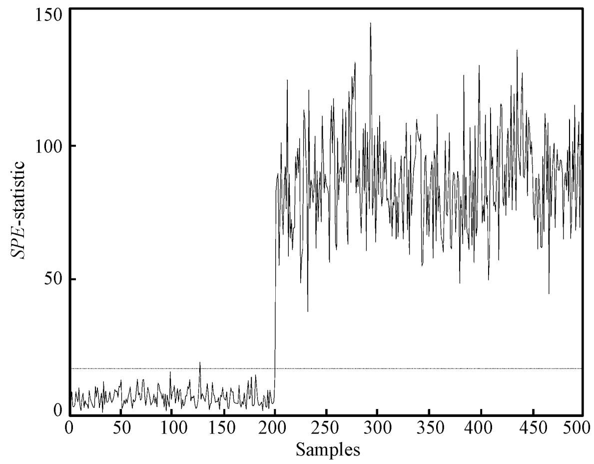


(b) T^2 statistic plot

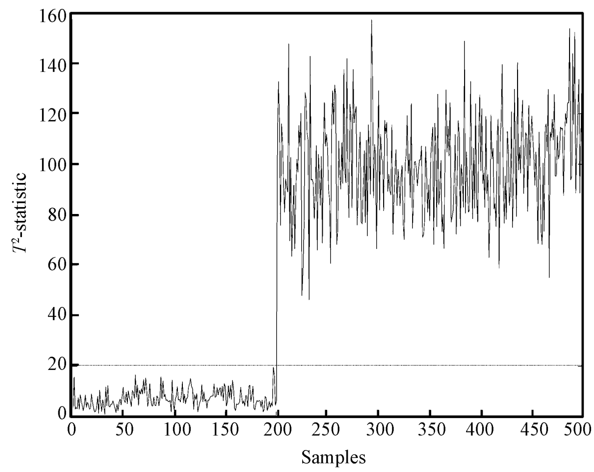
Fig. 5 Statistic plots with PCA model determined by the CPV method for F1

In Figs. 3 and 4, the value of straight line is 1 which represents the normal operating condition. When the value of statistic normalized by its control limit is larger than 1, it illustrates that the sensor fault is detected. A larger degree of deviation from control limit means the fault detection is more sensitive. Therefore, for SPE statistic; when the number of PCs is 12 and 15, fault F1 cannot be detected. When the number of PCs is 5, statistic normalized to its control limit deviate from the normal condition with the largest value, which indicate that when the number of PCs is 5, fault detection using SPE statistic is the most sensitive. In a similar way, Fig. 4 shows that when the number of PCs is 8, fault detection using T^2 statistic is the most sensitive for fault F1. Based on the above facts, the determination of the number of PCs based on fault SNR is effective.

In Section 2, we described that the selection of the number of PCs based on fault SNR considers not only the fault detection sensitivity but also the kind of fault. However, selection method based on CPV only considers normal operation data and never considers the performance of fault detection[15−16]. Figs. 5 and 6 give statistic plots with PCA model determined by fault SNR and CPV method, respectively.



(a) SPE statistic plot



(b) T^2 statistic plot

Fig. 6 Statistic plots with PCA model determined by fault SNR for F1

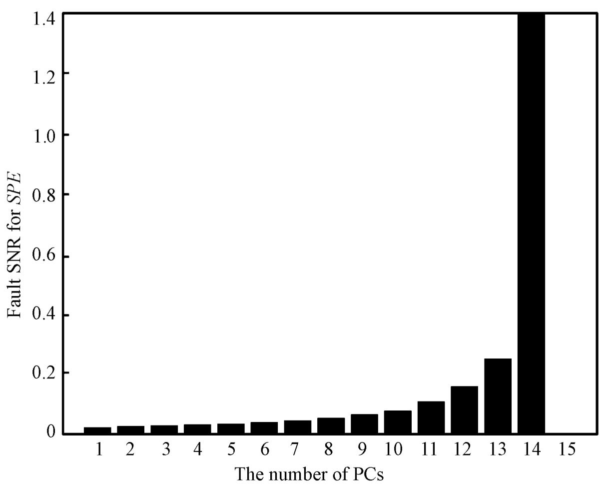
From Figs. 5 and 6, we can see that T^2 statistic plot with the PCA model determined by the CPV and fault SNR have desired fault detection results. However, for SPE

statistic, the performance of fault detection is poor with PCA model based on CPV method. However, the number of PCs determined by fault SNR gives good performance of fault detection. Therefore, from Figs. 3 $\sim$ 6, we can conclude that selection of the number of PCs by maximizing fault SNR considers the fault detection sensitivity and the kind of fault so that the number of PCs determined gives good performance of fault detection.

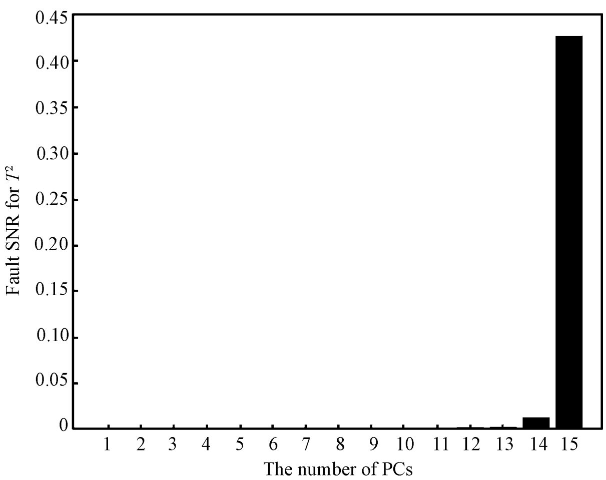
Figs. 7 and 8 give the fault SNR for faults F2 and F3, respectively. From Fig. 7, we can see that the number of PCs that gives maximum sensitivity is 14 for SPE statistic and 15 for T^2 statistic. Fig. 8 shows that for fault F3, when the number of PCs is 12, the PCA model gives maximum fault detection sensitivity using SPE statistic is 12, and for T^2 statistic, the number of PCs that gives good performance is 15. From Figs. 7 and 8, we can see that different faults have different PCA models and different statistics have different optimal number of PCs. Based on the above advantages, the selection of the number of PCs using fault SNR gives good performance. Figs. 9 $\sim$ 12 show the plots of statistic normalized to their control limit for faults F2 and F3. As shown in these figures, PCA models with the numbers of PCs selected by the proposed method show the optimum sensitivity for each fault.

For fault F2, we can clearly see from Figs. 9 and 10 that the number of PCs that give the maximum sensitivity is 14 based on SPE statistic monitoring, and for T^2 statistic, the optimal number of PCs is 15. The results that Figs. 9 and 10 show are consistent with the results based on the fault SNR. For fault F3, the plots of statistic normalized by their control limit in Figs. 11 and 12 also give the consistent results based on fault SNR. Therefore, based on fault SNR, the number of PCs for optimum sensitivity can be determined without any prior information of faults. The performance of PCA model based on fault SNR is shown, compared with the CPV method in Figs. 13 $\sim$ 16.

Figs. 13 and 14 give the statistic plots with PCA model determined by CPV and fault SNR, respectively. From these figures, we can see that for fault F2, no matter what statistic is used either SPE or T^2, the PCA model determined by fault SNR has better performance than that of the CPV method. For fault F3 in Figs. 15 and 16, PCA model based on CPV method is not accurate for SPE statistic. However, the PCA model determined by fault SNR gives a good performance of fault detection. Therefore, all the simulation results given above demonstrated that the number of PCs is optimum by maximizing fault SNR.

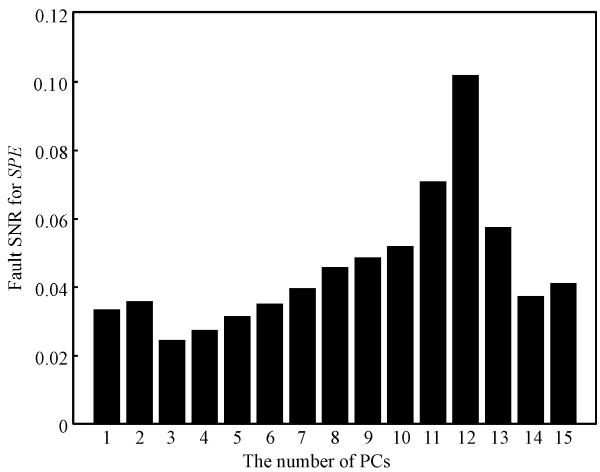


(a) Fault SNR for SPE statistic

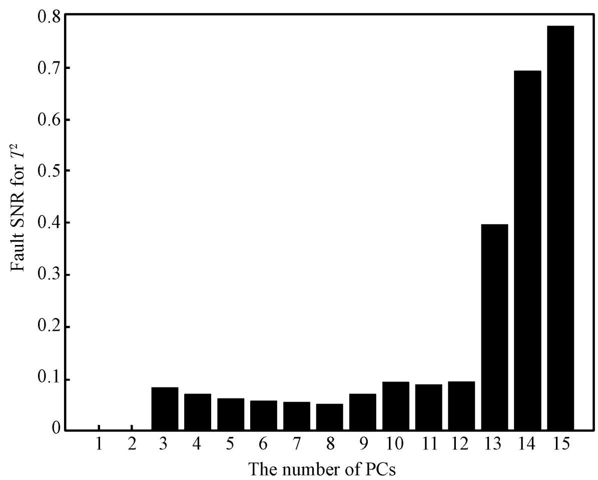


(b) Fault SNR for T^2 statistic

Fig. 7 Fault SNR for fault F 2



(a) Fault SNR for SPE statistic



(b) Fault SNR for T^2 statistic

Fig. 8 Fault SNR for fault F3

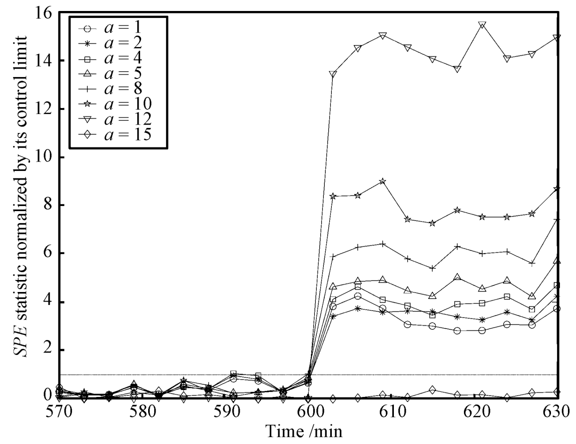


Fig. 9 SPE statistic normalized by its control limit for F2

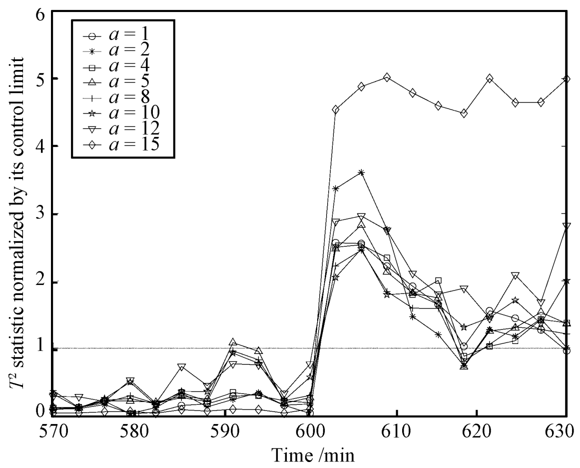


Fig. 10 T^2 statistic normalized by its control limit for F2

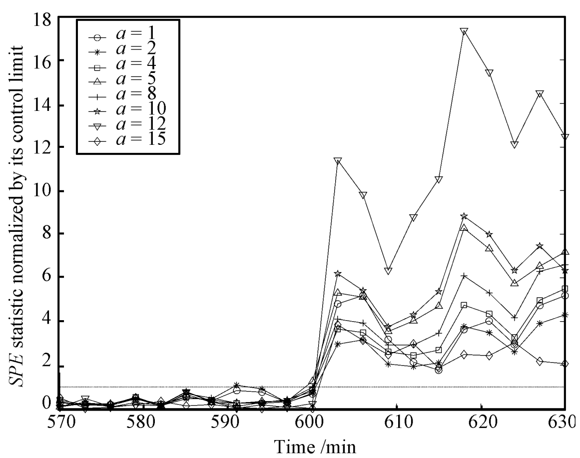


Fig. 11 SPE statistic normalized by its control limit for F3

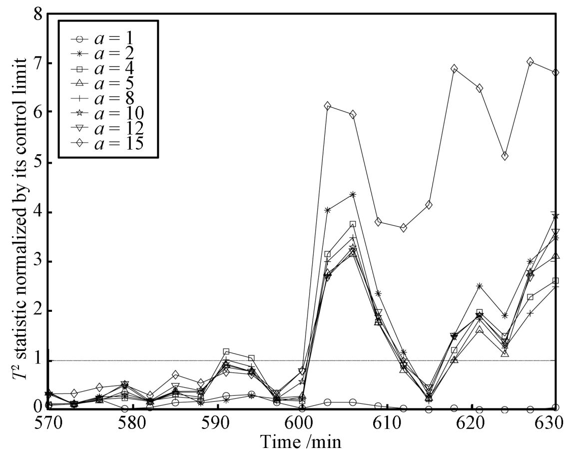
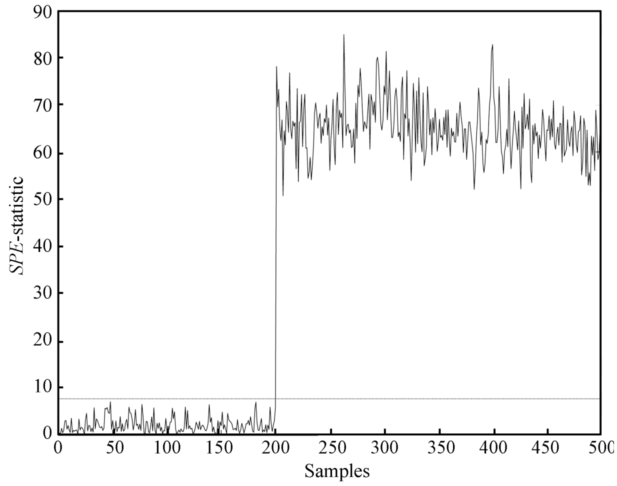
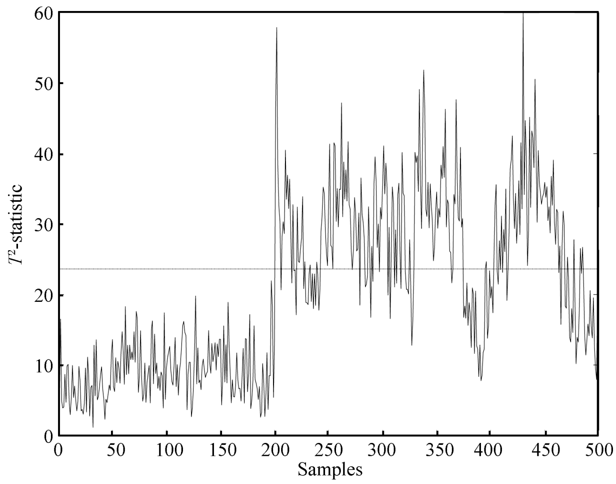


Fig. 12 T^2 statistic normalized by its control limit for F3

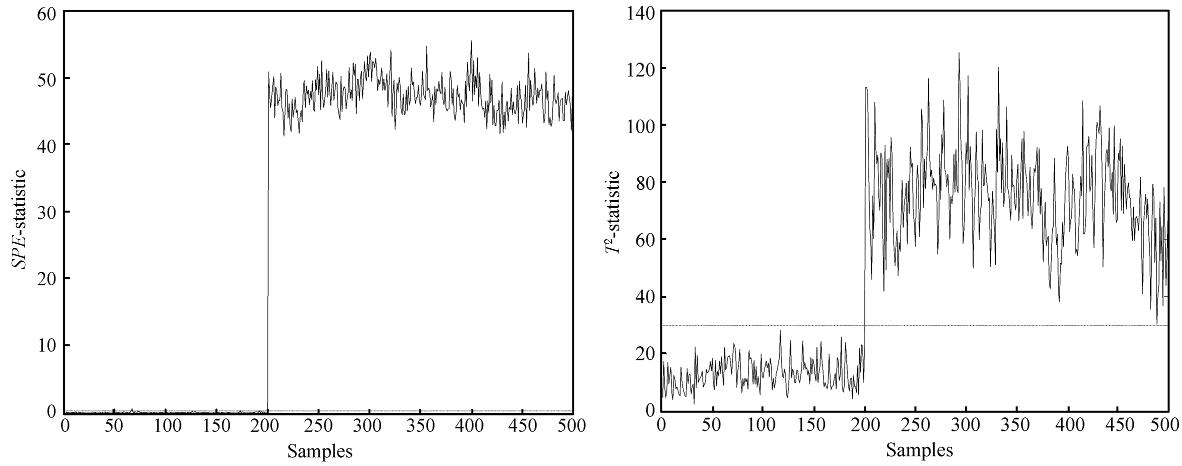


(a) SPE statistic plot



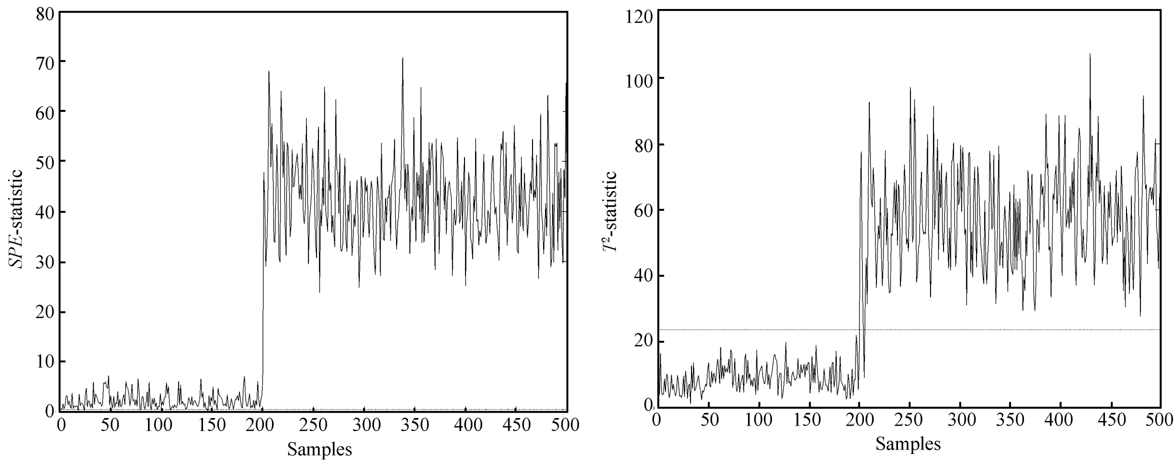
(b) T^2 statistic plot

Fig. 13 Statistic plots with PCA model determined by the CPV method for F2



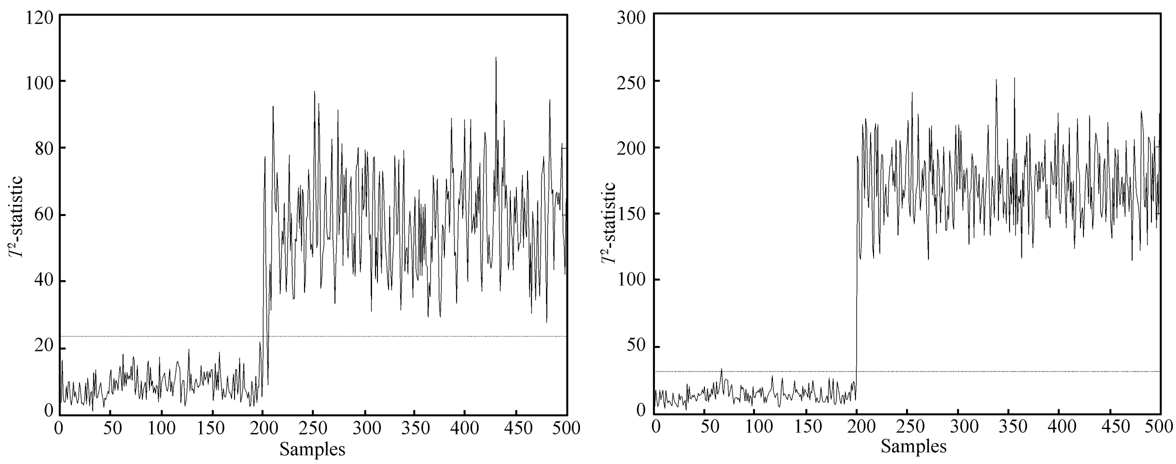
(a) SPE statistic plot (b) T^2 statistic plot

Fig. 14 Statistic plots with PCA model determined by fault SNR for F2



(a) SPE statistic plot (b) T^2 statistic plot

Fig. 15 Statistic plots with PCA model determined by the CPV method for F3



(a) SPE statistic plot (b) T^2 statistic plot

Fig. 16 Statistic plots with PCA model determined by fault SNR for F3

4 Conclusion

There are many known methods for selecting the number of PCs, but none of these previous methods consider the performance of fault detection. In this paper, fault SNR is presented as an index of fault detection ability. The number of PCs that maximizes the fault SNR to give optimal sensitivity is considered. Examining the fault SNR, the optimum number of PCs can be easily determined for sensor faults. Plots of statistic normalized by its control limit (which can be regarded as the sensitivity) are given to demonstrate that the number of PCs determined by the proposed method is the most optimum and thus improves sensitivity to fault detection. Furthermore, the proposed method is compared with the CPV method, and the results show that the selection of the number of PCs based on presented method gives the superior performance of fault detection for different kinds of sensor faults.

References

1 Zhou D H, Ye Yin-Zhong. *Modern Fault Diagnosis and Tolerance Control*. Beijing: Tsinghua University Press, 2000

2 Qin S J. Statistical process monitoring: basics and beyond. *Journal of Chemometrics*, 2003, **17**(8-9): 480−502

3 Kresta J V, MacGregor J F, Marlin T E. Multivariate statistical monitoring of process operating performance. *Canadian Journal of Chemical Engineering*, 2009, **69**(1): 35−47

4 Jackson J E. *A User's Guide to Principal Components*. New York: John Wiley and Sons, 1991

5 Zwick W R, Velicer W F. Comparison of five rules for determining the number of components to retain. *Psychological Bulletin*, 1986, **99**(3): 432−442

6 Himes D M, Storer R H, Georgakis C. Determination of the number of principal components for disturbance detection and isolation. In: Proceedings of American Control Conference. Washington D. C., USA: IEEE, 1994. 1279−1283

7 Malinowski E R. *Factor Analysis in Chemistry*. New York: Wiley-Interscience, 1991. 328−334

8 Li W H, Yue H H, Valle-Cervantes S, Qin S J. Recursive PCA for adaptive process monitoring. *Journal of Process Control*, 2000, **10**(5): 471−486

9 Wold S. Cross-validatory estimation of the number of components in factor and principal components models. *Technometrics*, 1978, **20**(4): 397−405

10 Dunia R, Qin S J. Joint diagnosis of process and sensor faults using principal component analysis. *Control Engineering Practice*, 1998, **6**(4): 457−469

11 Tang X C, Li Y. Selection of the number of principal components based on fault signal-to-noise ratio. In: Proceedings of the 7th World Congress on Intelligent Control and Automation. Washington D. C., USA: IEEE, 2008. 3637−3642

12 Downs J J, Vogel E F. A plant-wide industrial process control problem. *Computers and Chemical Engineering*, 1993, **17**(3): 245−255

13 Jackson J E, Mudholkar G. Control procedures for residuals associated with principal component analysis. *Technometrics*, 1979, **21**(3): 341−349

14 MacGregor J F, Jaeckle C, Kiparissides C, Koutoudi M. Process monitoring and diagnosis by multiblock PLS methods. *American Institute of Chemical Engineers Journal*, 1994, **40**(5): 826−838

15 Li Y, Zhou D H, Xie Z. Fault detection and isolation based on abnormal sub-regions using the improved PCA. *Journal of Chemical Engineering of Japan*, 2004, **37**(4): 514−522

16 Li Y, Zhou D H, Xie Z. Faulty sensor detection and reconstruction for a PVC making process. *Chinese Journal of Chemical Engineering*, 2004, **12**(2): 227−233

LI Yuan Professor at the College of Information Engineering, Shenyang Institute of Chemical Technology. She received her Ph. D. degree from the College of Information Science and Engineering, Northeastern University in 2004. Her research interest covers statistical process control and process fault detection and diagnosis. Corresponding author of this paper.
E-mail: li-yuan@mail.tsinghua.edu.cn

TANG Xiao-Chu Received her master degree from the College of Information Engineering, Shenyang Institute of Chemical Technology. Her research interest covers statistical process control and process fault detection and diagnosis.
E-mail: tangxc0420@126.com