STATISTICS PATTERN ANALYSIS BASED FAULT DETECTION AND DIAGNOSIS

Hector J. Galicia[a], Q. Peter He[b,*] and Jin Wang[a,*]
[a] Department of Chemical Engineering, Auburn University, Auburn, AL 36849
[b] Department of Chemical Engineering, Tuskegee University, Tuskegee, AL 36088

Abstract

Statistics pattern analysis (SPA) is a new multivariate statistical monitoring framework proposed by the authors recently. It addresses some challenges that cannot be readily addressed by the commonly used multivariate statistical methods such as principal component analysis (PCA) in monitoring batch processes in the semiconductor industry. It was later extended to the monitoring of continuous processes using a moving window based approach. In this work, we explore the potential of SPA in fault diagnosis. Specifically, we derive variable contributions based on the fault detection indices to generate contribution plots for fault diagnosis. The superior performance of the proposed method is demonstrated using the challenging Tennessee Eastman process (TEP), and compared with the commonly used contribution plots based on PCA and dynamic PCA (DPCA).

Keywords

Fault diagnosis, fault detection, process monitoring, multivariate statistical method, statistics pattern analysis, contribution plot.

Introduction

The increasing demand for safer and more reliable systems in modern process operations leads to the rapid development of process monitoring techniques. By promptly detecting process upsets, equipment malfunctions, and other special events, online process monitoring not only plays an important role in ensuring process safety, but also improves process efficiency and product quality. With a large amount of variables measured and stored automatically by distributed control systems (DCS), multivariate statistical monitoring methods have become increasingly common in process industry. Specifically, PCA-based process monitoring methods have gained wide application in chemical and petrochemical industries. PCA-based monitoring methods can easily handle high dimensional, noisy and highly correlated data generated from industrial processes, and provide superior performance compared to univariate methods. In addition, PCA-based process monitoring methods are attractive because they only require a good historical data set of normal operation, which are easily obtainable for the computer-controlled industrial processes.

Although PCA-based monitoring methods have been successful in many applications, there are cases where they do not perform well. Two of the possible reasons are given below. First, PCA only considers the mean and variance-covariance of the process data, and lacks the capability of providing higher-order representation for non-Gaussian data. Second, the control limits of Hotelling's T^2 and the squared prediction error (SPE) charts are developed based on the assumption that the latent variables follow a multivariate Gaussian distribution. Therefore when the latent variables are non-Gaussian distributed due to

* Corresponding authors
Email addresses: hjg0002@tigermail.auburn.edu (Hector J. Galicia), qhe@tuskegee.edu (Q. Peter He), wang@auburn.edu (Jin Wang)

process nonlinearity or other reasons, using Hotelling's T^2 and SPE may be misleading (Lee et al., 2006, Martin and Morris, 1996). To address the above mentioned challenges presented in industrial processes that cannot be readily addressed by PCA, several alternative approaches have been developed (Kano et al., 2002, Kano et al., 2003, Lee et al., 2006, Lee et al., 2004, Luo et al., 1999). Recently, we proposed a new multivariate statistical process monitoring framework, named Statistics Pattern Analysis (SPA) (Wang and He, 2010, He and Wang, 2011). The major difference between the PCA-based and SPA-based fault detection methods is that PCA monitors process variables while SPA monitors the statistics of the process variables. In other words, SPA examines the variance-covariance of the process variables statistics (e.g. mean, variance, autocorrelation, cross-correlation etc.) to perform fault detection. In SPA, different statistics that capture the different characteristics of the process can be selected to build the model for normal process operation, and various higher-order statistics can be utilized explicitly. Fault detection methods derived based on the SPA framework have been shown to provide superior monitoring performance for several batch and continuous processes compared to the PCA and dynamic PCA based methods (Wang and He, 2010, He and Wang, 2011).

In this work, the potential of the SPA framework for fault diagnosis of continuous processes is explored. Specifically, we derive variable contributions from the fault detection indices generated by SPA; then construct contribution plots to perform fault diagnosis. Because the SPA based fault detection method for continuous processes is a window-based approach, the SPA based contribution plot is an averaged contribution plot for all samples in the window. The challenging Tennessee Eastman Process (TEP) is used to demonstrate the performance of the SPA-based fault detection and diagnosis method, which is compared with PCA and dynamic PCA (DPCA) methods.

Fault detection in SPA

In this section, we briefly review the SPA-based fault detection method for continuous processes. As shown in Figure 1, two steps are involved in the SPA-based monitoring for continuous processes: statistics pattern (SP) generation and dissimilarity quantification. For a continuous process, an SP is a collection of various statistics calculated from a window (or a segment) of the process measurements. These statistics capture the characteristics of each individual variable (such as mean, variance, and skewness), the interactions among different variables (such as correlation), as well as process dynamics (such as autocorrelation and cross-correlation). Note that, for different processes, different statistics can be selected to capture the dominant process characteristics such as dynamics and nonlinearity. After the SPs are computed from the training data, the dissimilarities among the training SPs are quantified to determine the upper control limit of the detection index. In this work, we apply PCA to quantify the dissimilarities among the training SPs and define two detection indices similar to Hotelling's T^2 and SPE. When a single or a block of new measurements becomes available for fault detection, the window is shift forward by one or multiple samples and a new SP is computed; then its dissimilarity to the training SP's is quantified and compared with the threshold to perform fault detection. To distinguish the SPA-based fault detection indices from the traditional PCA-based fault detection indices, we use D_{p} and D_{r} to denote the T^2 and SPE in the SPA framework. The process is considered normal if the dissimilarity indices are below the thresholds, i.e., $D_{\mathrm{p}} \le T_{\alpha}^2$ and $D_{\mathrm{r}} \le \delta_{\alpha}^2$, where T_{α}^2 and δ_{α}^2 denote the upper control limits for dissimilarity index in the principal component subspace (PCS) and residual subspace (RS) with a significance level α. It is worth noting that PCA is just one way to determine the similarities or dissimilarities among different samples; other methods can be implemented to obtain distance-based or angle-based similarity indices (Singhal and Seborg, 2002). The details of the statistics pattern generation and dissimilarity quantification can be found in (Wang and He, 2010).

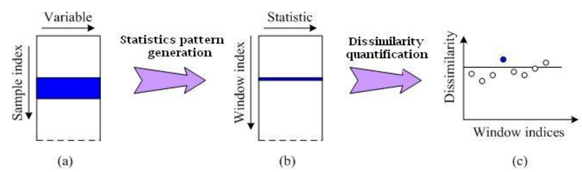

Figure 1: Schematic of the window-based SPA method for monitoring continuous processes: (a) original process data; (b) computed statistics pattern; (c) fault detection. A block or a window of process variables shaded in (a) is used to generate an SP shaded in (b), which is then used to generate a point or dissimilarity measure shaded in (c) to perform fault detection.

Fault diagnosis in SPA

After a fault is detected by one or more fault detection indices exceeding their control limits, it is desirable to perform fault diagnosis to identify the root cause of the fault. For PCA-based fault detection methods, contribution plot is the most commonly applied fault diagnosis method, which is based on the assumption that the variables with the largest contributions to the fault detection index are most likely the faulty variables. Because in the SPA-based fault detection method we apply PCA to quantify the dissimilarity among different statistics patterns, here we construct the contribution plots for D_P and D_r to perform fault diagnosis. Let $\mathbf{X} \in \Re^{m \times n}$ denote the SP matrix with n samples (rows) and m statistics (columns). After autoscaling to zero mean and unit variance, the matrix $\mathbf{X}$ is decomposed as follows

$$\mathbf{X} = \mathbf{T}\mathbf{P}^{\mathrm{T}} + \tilde{\mathbf{X}} = \mathbf{T}\mathbf{P}^{\mathrm{T}} + \tilde{\mathbf{T}}\tilde{\mathbf{P}}^{\mathrm{T}} = \left[\mathbf{T}\tilde{\mathbf{T}}\right]\left[\mathbf{P}\tilde{\mathbf{P}}\right]^{\mathrm{T}} \tag{1}$$

where $\mathbf{T} \in \Re^{n \times l}$ and $\mathbf{P} \in \Re^{m \times l}$ are the score and loading matrices, respectively; l is the number of principal components.

For fault detection on a new sample vector $\mathbf{x}$, two indices are used: D_p and D_r as defined below,

$$D_p = \left(\mathbf{x}^{\mathsf{T}} \mathbf{P} \boldsymbol{\Lambda}^{-1} \mathbf{P}^{\mathsf{T}} \mathbf{x}\right) \tag{2}$$

$$D_r = \|\tilde{\mathbf{x}}\|^2 = \left(\mathbf{x}^{\mathsf{T}} \tilde{\mathbf{P}} \tilde{\mathbf{P}}^{\mathsf{T}} \mathbf{x}\right) \tag{3}$$

where $\boldsymbol{\Lambda}$ is the diagonal matrix of the l largest eigenvalues of $\mathbf{X}\mathbf{X}^{\mathsf{T}}$. The D_p statistic is a measure of the process variation in the principal component subspace; while the D_r statistic indicates how well each sample conforms to the model.

The contribution of the i^{th} statistics to D_p and D_r are the following,

$$C_i^p = \left(\boldsymbol{\xi}_{\mathrm{i}}^{\mathsf{T}} \left(\mathbf{P} \boldsymbol{\Lambda}^{-1} \mathbf{P}^{\mathsf{T}}\right)^{1/2} \mathbf{x}\right)^2 \tag{4}$$

$$C_i^r = \left(\boldsymbol{\xi}_{\mathrm{i}}^{\mathsf{T}} \tilde{\mathbf{P}} \tilde{\mathbf{P}}^{\mathsf{T}} \mathbf{x}\right)^2 \tag{5}$$

where $\boldsymbol{\xi}_i$ is the i^{th} column of the m-dimensional identity matrix.

Case study

In this section, the performance of the SPA-based fault detection and diagnosis method is compared with the traditional PCA and DPCA methods using the Tennessee Eastman Process (TEP) (Downs and Vogel, 1993). The TEP process simulator has been widely used by the process monitoring community as a realistic example to compare various approaches (Kano et al., 2001, Ku et al., 1995, Russell et al., 2000). The simulator was originally developed by Downs and Vogel (1993) and different control strategies were implemented in different modified versions (Banerjee and Arkun, 1995, Chiang et al., 2001, Lyman and Georgakis, 1995, Ricker, 1996). In this work, we use Ricker's simulator to generate normal and faulty data. The simulator can simulate normal process operation together with 20 faulty conditions. Following (Chiang et al., 2001), the data are collected at a sampling interval of 3 min. The process data include 11 manipulated variables, 22 continuous process measurements, and 19 composition measurements which are sampled less frequently. Similar to Lee at al. (2006), 22 continuous process measurements and 9 manipulated variables listed in Table 1 are used to monitor the process. The other two manipulated variables were fixed to constants in Ricker's decentralized control scheme, and therefore are not included in the monitoring. Details about the process variables can be found in (Downs and Vogel, 1993, Ricker, 1996). We use 800 hr of normal data for training, 50 hr normal data for false alarm rate calculation, and each fault consists of 50 hr samples with the first 10 hr being normal.

Table 1: Monitored variables in TEP

no.	Variable description	no.	Variable description
	Process Measurements		
1	A feed (stream 1)	12	Product separator level
2	D feed (stream 2)	13	Product separator pressure.
3	E feed (stream 3)	14	Product separator underflow (stream 10)
4	A and C feed (stream 4)	15	Stripper level
5	Recycle flow(stream 8)	16	Stripper pressure
6	Recycle feed rate (stream 6)	17	Stripper underflow (stream 11)
7	Reactor pressure	18	Stripper temperature
8	Reactor level	19	Stripper steam flow
9	Reactor temperature	20	Compressor work
10	Purge rate (stream 9)	21	Reactor cooling water outlet temperature
11	Product separator temperature	22	Separator cooling water temperature
	Manipulated Variables		
23	D feed flow (stream 2)	28	Separator pot liquid flow (stream 10)
24	E feed flow (stream 3)	29	Stripper liquid product flow (stream 11)
25	A feed flow (stream 1)	30	Reactor cooling water flow
26	A and C feed flow (stream 4)	31	Condenser cooling water flow
27	Purge valve (stream 9)		

The settings of different methods are listed in Table 2, which are similar to the settings in Wang and He (2010). The empirical method is used to determine the upper control limits so that different methods can be compared based on the same confidence level (Wang and He, 2010).

Table 2: Settings of different fault detection methods

Methods	Variables	PCs	lags	Window width	Window shifting step
PCA	31	9	-	-	-
DPCA	93	20	2	-	-
SPA	257	6	2	100	50

● **Fault detection**: The detection rates of the three methods for all faults are listed in Table 3. It is worth noting that the results in Table 3 are different from the results in Wang and He (2010). This is due to the fact that different data sets were used. In Wang and He (2010), the data set was generated by Chiang et al. (2001) where the plant-wide control structure recommended in Lyman and Georgakis (1995) was implemented to generate the closed loop normal and faulty process data. While in the current work, the TEP simulator with decentralized control system developed by Ricker (1996) was implemented to generate the closed loop normal and faulty process data. By visually comparing the two control strategies under normal operation condition, we noticed that the process variation under the decentralized control is much smaller than that under the plant-wide control. Therefore, the normal process model with the decentralized control defines a tighter region of normal operation than that with the plant-wide control, which makes the model more sensitive and therefore more effective to fault detection. This is true no matter what fault detection is used. As a result, we see that although similar settings were used, the fault detection rates of all the three methods are higher for almost all faults in this work than those in Wang and He (2010), with PCA and DPCA improved the most.

From Table 3, we see that all three methods are effective in detecting most of the faults. It is worth noting that faults 3, 9 and 15 have been suggested to be difficult to detect when the plant-wide control strategy is implemented. In this work, we find that these faults are also difficult to be detected by any of the three methods when the decentralized control strategy is implemented. Therefore, these three faults are not listed in Table 3. In addition, we find that fault 16 cannot be detected by any of the three methods either when the decentralized control strategy is implemented. After visually inspecting the four difficult cases, we believe the reason for these faults not being detected is that the disturbances were completely rejected without introducing noticeable changes to any process variables. For the rest of the 16 faults, PCA and DPCA have difficulty in detecting faults 5 and 12 with detection rates lower than 50% in most of the cases, which are depicted in Figure 2. On the other hand, SPA based method is able to detect all 16 faults with all detection rates higher than 90%. It is worth noting that although detection delay is usually associated with window-based approaches, there is only minor detection delay associated with SPA method. This is because the detection limit is much tighter in SPA, since the variance of the variable statistics is only $1/n$ of the process variable (n is the window width). In addition, for second or higher-order statistics, the contribution from the faulty data to the variable statistics is not linear. Therefore, as shown in Table 3, the detection rate is almost same for the faults that are easily detected. For the faults that cannot be detected by PCA or DPCA methods, such as fault 18, SPA is more effective. To ensure fair comparison, the false alarm rates of all the three methods are listed in Table 4. The results indicate that the thresholds determined empirically with 99% confidence level are reasonable and consistent because the false alarm rates are not far from 1% for all the methods.

● **Fault diagnosis**: In this subsection, the proposed SPA based fault diagnosis approach is applied to TEP to identify the root cause of each fault. As a comparison, PCA and DPCA are also applied to diagnose the faults. Due to page limit, the diagnoses of all faults are not discussed in detail except the three illustrative examples given below. We do want to mention that, in general, all the three methods are effective in pinpointing the major fault-contributing process variables for most of the faults.

Table 3: Fault detection rates (percentage) of PCA, DPCA, and SPA for TEP

	PCA		DPCA		SPA	
fault	T^2	SPE	T^2	SPE	D_p	D_r
1	99.9	100.0	100.0	100.0	99.3	99.5
2	99.5	99.3	99.6	99.4	99.0	98.8
4	100.0	100.0	100.0	100.0	99.8	99.9
5	3.6	0.6	3.6	1.5	0.0	93.4
6	100.0	100.0	100.0	100.0	98.6	99.3
7	100.0	100.0	100.0	100.0	99.9	99.9
8	100.0	100.0	100.0	100.0	99.9	99.9
10	83.7	93.1	93.0	96.2	98.1	98.5
11	94.9	98.2	98.9	99.8	99.0	99.5
12	44.2	14.6	59.1	39.1	93.8	94.6
13	98.8	98.8	99.0	99.1	97.9	98.3
14	99.1	100.0	100.0	100.0	99.5	99.8
16	4.4	0.5	4.1	0.4	0.0	0.0
17	97.3	99.3	99.3	99.6	98.4	99.0
18	71.3	76.5	73.0	85.7	93.6	96.0
19	98.0	99.0	99.5	99.0	96.5	98.9
20	98.9	98.6	99.0	99.0	97.9	98.5

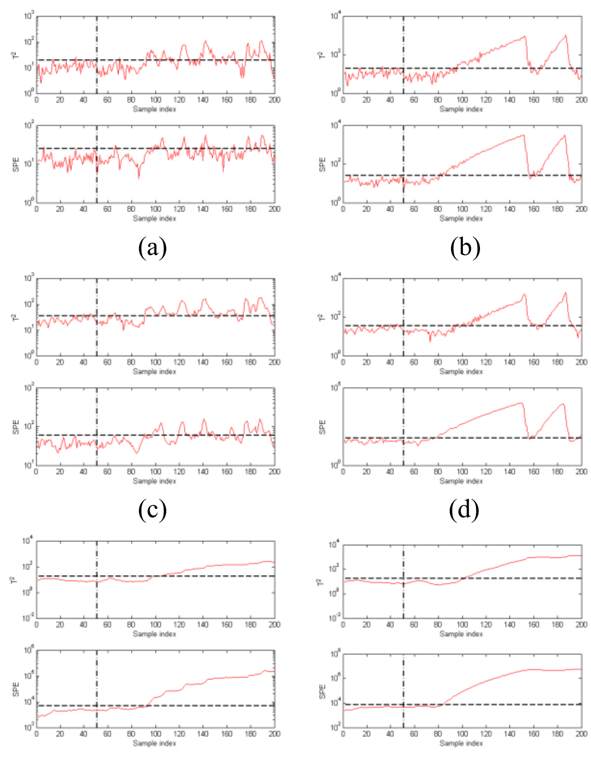

Figure 2: Detection performance of PCA, DPCA and SPA for fault 12 ((a) PCA; (c) DPCA; (e) SPA), fault 18 ((b) PCA; (d) DPCA, (f) SPA). The dash-dotted lines indicate the fault onsets and dashed lines the index thresholds.

For the contribution plots based on SPA, we use two subplots to show the contributions from variable mean and standard deviation. The contributions from other statistics such as auto- and cross-correlations are small and are not shown due to limited space.

The example of fault 10 is shown in Figure 3 where random variation is introduced to C feed (steam 4 to the stripper) temperature. Figure 3 shows that all the three methods correctly identify that the variable 18 (stripper temperature) is the major fault-contributing variable. Compared to PCA and DPCA based contribution plots, the SPA based contribution plots provide extra information by identifying that it was the variance, not the mean, of the variable 18 that contribute to this fault.

Table 4: False alarms rate (percentage) of PCA, DPCA, and SPA for the TEP.

Data	PCA		DPCA		SPA	
	T^2	SPE	T^2	SPE	D_p	D_r
training	2.1	0.4	2.4	0.7	0.1	0.1
Testing	1.5	0.3	1.4	0.7	0	2.6

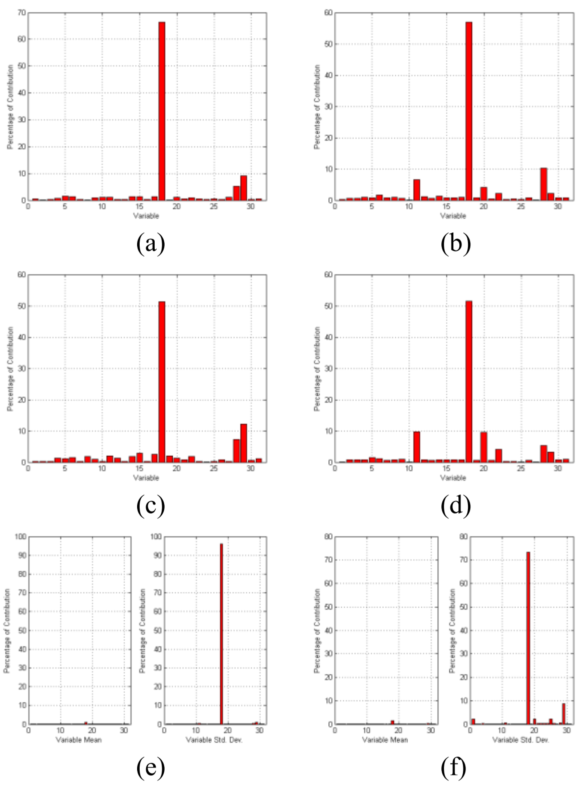

Figure 3: Diagnosis of fault 10. (a) PCA-T^2; (b) PCA-SPE; (c) DPCA-T^2; (d) DPCA-SPE; (e) SPA-D_p; (f) SPA-D_r.

There are cases where SPA does a better job in fault detection than PCA and DPCA. In these cases, SPA also does a better job in fault diagnosis. One such example is fault 5 shown in Figure 4 where a step change occurred in condenser cooling water inlet temperature. Table 3 shows that PCA and DPCA cannot detect this fault. Therefore, it is expected that they cannot diagnose the fault either, which is verified by Figure 4 (a)-(d) where a range of process variables across different units all contribute to this fault. On the other hand, SPA was able to isolate the fault root cause to the manipulated variable 31, the condenser cooling water flow rate. Since the cooling water inlet temperature is not measured, the manipulated variable 31 is the most related variable. In addition, SPA was able to indicate that a mean change, not variance, in variable 31 is the root cause, which is directly related to the step change in the condenser cooling water inlet temperature.

In the final example of fault 12 (random variation in condenser cooling water inlet temperature) shown in Figure 5, we want to illustrate that a successful fault detection does not necessarily lead to a correct fault diagnosis. From Table 3, we see that all the three methods were able to detect the fault. However, from Figure 5, we see that PCA and DPCA were not able to isolate the root cause of the fault. SPA was able to relate the fault to the separator temperature (variable 11) right after the condenser. Again, since the cooling water temperature or the condenser temperature was not measured and this random variation would not trigger the manipulated cooling water flow rate to change, the downstream separator right after the condenser is the closest to the actual fault location. In other words, the random variation in condenser cooling water inlet temperature (not measured) will immediately lead to random variation in the separator temperature (measured, variable 11). It is also worth noting that SAP correctly identified that the standard deviation, not the mean, of variable 11 is the major fault contributor.

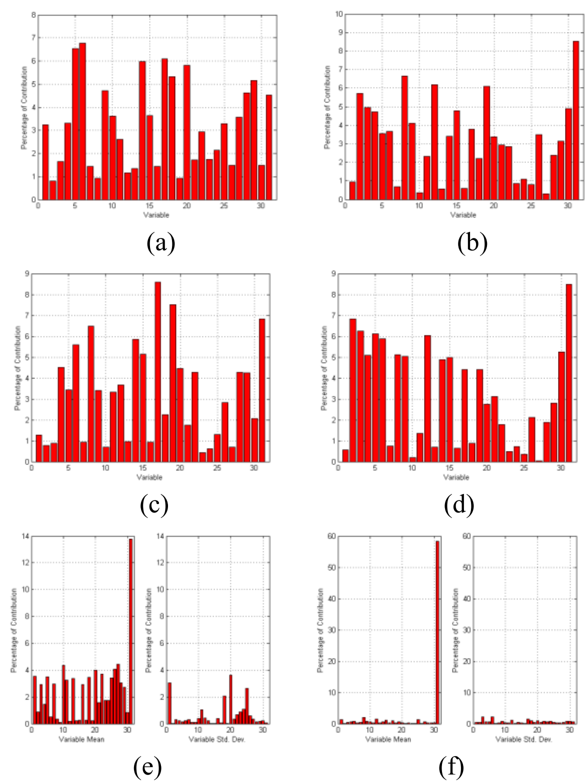

Figure 4: Diagnosis of fault 5. (a) PCA-T^2; (b) PCA-SPE; (c) DPCA-T^2; (d) DPCA-SPE; (e) SPA-D_p; (f) SPA-D_r.

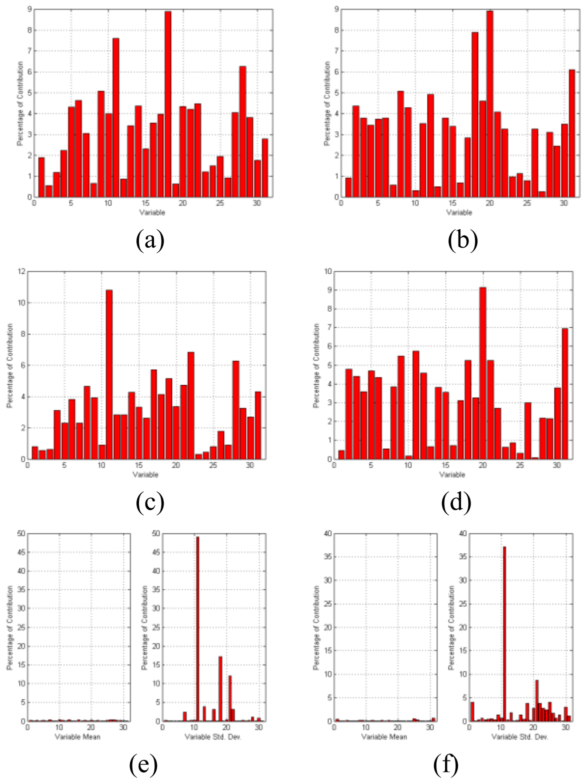

Figure 5: Diagnosis of fault 12. (a) PCA-T^2; (b) PCA-SPE; (c) DPCA-T^2; (d) DPCA-SPE; (e) SPA-D_p; (f) SPA-D_r.

Conclusions

In this work, we explore the potential of our recently proposed SPA framework in fault diagnosis for continuous process. Specifically, the variable contributions are derived based on the fault detection indices used in SPA to generate contribution plots for fault diagnosis. We use the Tennessee Eastman process to evaluate the performance of the developed fault detection and diagnosis method based on the SPA framework, and compare with the PCA and DPCA based methods. The case study shows that in general all three methods work well in detecting and diagnosing most of faults. However, for some faults that are difficult to detect and/or diagnose by PCA and DPCA based methods, SPA based method provides superior performance. In addition, because the SPA-based method breaks down the contribution of a fault to different variable statistics, they provide extra information in addition to identifying the major fault-contributing variable(s). For example, the SPA-based contribution plots tell us whether the fault is due to a change in variable mean or variance. It also should be noted that in general SPA requires more training data to build a reliable model due to the computation of variable statistics. However, this is not a big issue because most modern processes are equipped with DCS systems and therefore are data rich.

Acknowledgments

The authors thank NSF for financial support under Grants CBET-0853983 (JW) and CBET-0853748 (QPH).

References

Banerjee, A. and Arkun, Y. (1995). Control configuration design applied to the Tennessee Eastman plant-wide control problem. *Computers and Chemical Engineering 19(4): 453-480.*

Chiang, L.H., Russell, E.L., Braatz, R.D. Fault Detection and Diagnosis in Industrial Systems. *Springer*: London, 2001.

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

He, Q. P. and Wang, J. (2011). Statistics pattern analysis: A new process monitoring framework and its application to semiconductor batch processes. *AIChE J. 57: 107-121.*

Kano, M., Hasebe, S., Hashimoto, H. (2002). Statistical process monitoring based on dissimilarity of process data. *AIChE Journal 48(6): 1231-1240.*

Kano, M., Tanaka, S., Hasebe, S., Hashimoto, H. (2003). Monitoring independent components for fault detection. *AIChE Journal 49(4): 969-976.*

Ku, W., Storer, R. H. Georgakis, C. (1995). Disturbance detection and isolation by dynamic principal component analysis. *Chemometrics and Intelligent Laboratory Systems 30(1): 179-196.*

Lee, J.M., Yoo, C., Lee, I.B. (2004). Statistical process monitoring with independent component analysis. *Journal of Process Control 14(5): 467-485.*

Lee, J.M., Qin, S.J., Lee, I.B. (2006). Fault detection and diagnosis based on modified independent component analysis. AIChE Journal 52(10): 3501-3514.

Luo, R., M. Misra, M., Himmenlblau, D.M. (1999). Sensor fault detection via multiscale dnalysis and dynamic PCA. *Industrial and Engineering Chemistry Research 38(4): 1489-1495.*

Lyman, P. R. and Georgakis, C. (1995). Plant-wide control of the Tennessee Eastman problem. *Computers and Chemical Engineering 19(3): 321-331.*

Martin, E. B. and Morris, A. J. (1996). Non-parametric confidence bounds for process performance monitoring charts. *Journal of Process Control 6(6): 349-358.*

Ricker, N. L. (1996). Decentralized control of the Tennessee Eastman Challenge Process. *Journal of Process Control 6(4): 205-221.*

Russell, E. L., Chiang L. H., Braatz, R. D. (2000). Fault detection in industrial processes using canonical variate analysis and dynamic principal component analysis. *Chemometrics and Intelligent Laboratory Systems 51(1): 81-93.*

Singhal, A. and Seborg, D. E. (2002). Pattern matching in multivariate time series databases using a moving-window approach. *Industrial and Engineering Chemistry Research 41(16): 3822-3838.*

Wang, J. and He, Q. P. (2010). Multivariate statistical process monitoring based on Statistics Pattern Analysis. *Industrial and Engineering Chemistry Research 49 (17): 7858–7869.*