Comparison of Different Fault Detection Statistics Detectability in PCA

Tian Qiu
Key Laboratory of Measurement and Control for Industry Process
North China Electric Power University
Beijing, China
Email: qiutian@ncepu.edu.cn

Xiaojing Bai
Key Laboratory of Measurement and Control for Industry Process
North China Electric Power University
Beijing, China
Email: baixiaojing1987@163.com

***Abstract*-Detectability of the sensor fault detection system is the basic criteria for selecting of different fault detection statistics. PVR and CVR in MPCA are compared with T_H^2. The complementary among them is analyzed qualitatively. Results of power plant data simulation show that PVR and CVR in MPCA can improve sensor fault detection, but T_H^2 is still better for some sensors. Combination of PVR, CVR and T_H^2 can improve sensor fault detectability of PCA-based fault detection system.**

Keywords-process monitoring, fault detection, Principal component analysis

I. INTRODUCTION

Process monitoring in power plants need operators to monitor a considerable amount of process variables. Current methods usually set the upper and lower limit of variables to realize simple monitor. This kind of methods cannot change the upper and lower limit according to the operation condition automaticly, so the faults with little magnitude cannot be detected, which is detrimental to early fault detection and cannot satisfy the increasingly strict monitoring requirement. Multivariable analysis methods build process models with the correlation among process variables. These kind of methods has better performance in fault detection, when the correlation among process variables has changed [1]. In multivariable analysis methods, Principal Component Analysis based sensor fault detection has been researched deeply and applied into different industrial process monitoring [2-12].

SPE (Square Prediction Error) and Hotelling T^2 are important indice in PCA for fault detection. Dunia et al [2] proposed a whole set PCA method for sensor fault detection, diagnosis and reconstruction using SPE. Wang et al [3] analyzed the performance of T^2 for fault detection based on PCA, and proposed the method of fault diagnosis and reconstruction based on T^2. Qin [4] summarized fault detecion based on PCA. SPE statitic has different detectability for different sensors [5], Wang et al [6] divided SPE into two statistics PV (Principal-component-related Variable) and CV (Common Variable), and proposed an improvement for classic PCA method to improve the fault detectability.

Hawkins T_H^2 was considered to be similar with SPE. They have different fault detectability for different sensors in fault detetion [7]. It is necessary to investiage the fault detectability among Hawkins T_H^2, PVR and CVR in improved PCA method, for improved PCA is consided to have been improved the fault detectability of PCA method.

II. PCA THEORY

Let $x \in R^m$ denote a vector of m sensors. Assume $X \in R^{N\times m}$ is composed by N samples of each sensor, and each row represent a sample. According to PCA theory, x can be divided into 2 parts:

$x = \hat{x} + \tilde{x}$

$\hat{x}$ and $\tilde{x}$ represent the modeling part and residual part separately.

x can be projected to the PCS(Priciple Component Subspace) and then projected back, the result part is defined as $\hat{x}$:

$\hat{x} = PP^T x$

$P \in R^{m\times l}$ is the loading matrix in PCA, which is composed by the lf first eigenvector of the samples's covariant matrix. Sample vectors has more variation at these directions, which shows that sample variants have more correlation at these directions. $l \leq 1$ is the dimension of the PCS, the chosen of l depend on the correlation among sample variants. If l is assigned to m, $PP^T = I$, sample vector has no information loss in PCA.

xf can be projected to the RS(Residual Subspace) and then projected back, the result part is defined as $f\tilde{x}$:

$x = (I - PP^T)x$

$(I - PP^T)$ is the residual matrix in PCA, which is composed by the Residual eigenvector of the samples's covariant matrix. Sample vectors have less variance at these directions, which shows that sapmle variants have less correlation at these directions. And the variance at these directions is often consided as noise.

Let consider the situation in which only one sensor has malfunction. When the sensor has fault, the correlation among sample variants is changing. Original PCA model doesn't fit, as $28\tilde{x}$ is increasing, which means the projection of sample vector to RS is increasing. SPE can be used to detect this kind of fault:

$SPE = \|\tilde{x}\|^2 < \delta_{SPE}$

The calculation of δ_{SPE} is described in paper [13]. When SPE is bigger than the control limit, some sensor is considered to have fault, on the other hand, every sensor in good condition.

Wang et al [6] proposed an improved PCA method, in which two indice are proposed using multiple correlation among sensors and the principal components. PVR (PV Residuals) is defined by the first s process variables (Principal-component-related Variable, PV), which have remarkable correlation with principalm components; $C28VR$ is defined by the rest process variables (Common Variable, CV). These two indice can be calculated as:

$$PVR = x_s(I - P_s P_s^T)x_s^T$$

$$CVR = x_{m-s}(I - P_{m-s}P_{m-s}^T)x_{m-s}^T$$

Subscript $s2$ and $m-s$ in xf and P represent PV and CV part in data vector x and loading matrix P seperately. The control limit of PVR and CVR can be calculated with the description in paper [6].

Hawkins' T_H^2 is defined as [14]:

$$T_H^2 = \sum_{i=l+1}^{m} \frac{t_i^2}{\lambda_i}$$

$\tilde{\Lambda} = diag(\lambda_{l+1}, \ldots, \lambda_m)$ is the diagonal matrix composed with the last $m-l$ eigenvalue of the samples's covariance matix. The control limit can be calculated accroding to the description in paper [14].

III. COMPARISON BETWEEN T_H^2 AND PVR/CVR

Feng et al [5] pointed out that SPE has different fault detectability with different sensors. T_H^2 and SPE are indices measuring the projections of process variables in the residual subspace, which means T_H^2 and SPE can be compared directly in the residual subspace. The essence of improved PCA method is to decompose SPE into two indice [6], so PV and CV indice should have different fault detectability too. For PV and CV indice have different variable subspace, compared with T_H^2 , they cannot be compared directly, so the analysis has to be qualitative in the following part to compare PVR and CVR with T_H^2 and SPE.

PV and CV are the SPE in partial variable space, these two indice should have similar properties as SPE, so the difference between SPE and T_H^2 is discussed here mainly. SPE and T_H^2 have different fault detectability with different sensors, because their control limits δ_{SPE} and δ_H have some kind of conservation [7].

The simulation example in paper [4] is applied here for qualitative analysis:

$$\begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{bmatrix} = \begin{bmatrix} 0.3873 & 0.1190 \\ -0.1291 & 0.2379 \\ 0.9037 & -0.1530 \\ 0.1291 & 0.9518 \end{bmatrix} \begin{bmatrix} s_1 \\ s_2 \end{bmatrix} + noise$$

x_i is the measured variable, s_1 and s_2 are zero-mean random sequences with standard deviations of 1, noise is the white noise, whose standard deviations are 0.1, 0.8, 0.1, 0.4 for $x_1, \ldots, x_4$ seperately.

1000 samples are generated with the above equation. The PCA model is built with the samples, the dimension of principal subspace is 2. According to MPCA method, x_1 and x_3 are the PV, the rest variables are CV. Calculate the indices with 99% confidence coefficient, the result in shown in Figure 1.

Blue points in Figure 1 represent the projection of data vector to the residual subspace. In the residual subspace, with the coordinates of t_3 and t_4, the control limit of SPE is a circle, while the control limit of T_H^2 is a ellipse. Figure 1 (a) shows that the control limit of SPE and T_H^2 can envelope sample data, and T_H^2's ellipse shape of the control limit is better than SPE's to fit the sample data. Figure 1 (b) shows the projection direction of different sensor's fault, the directions of x_1 and x_3 are closer to t_4 coordinate, the directions of x_2 and x_4 are closer to t_3 coordinate. And only fault in x_2 can cross the control limit of SPE first, T_H^2 can detect other sensors' fault better than SPE. This result is consistent with the analysis of Feng et al in paper [5].

Figure 1 (c) and (d) shows that the control limit of PVR evelopes the sample data set well, while the control limit of CVR leaves a few sample datas outside of the limit, which indicates that MPCA maybe too strict to keep false alarm rate as small as declared by the confidence coefficient.

The conservation SPE statstic represents process data have different divergent in different direction in RS, while the control limit of δ_{SPE} is the same value at different direction. Improved PCA method decomposes SPE into 2 indice, which decrease the conservation of SPE in some degree, but PVR and CVR still have similar conservation as SPE. Improved PCA method should be able to improve the fault Detectability, but T_H^2 cannot be replaced.

IV. POWER PLANT DATA SIMULATION

Thermal process data is retrieved from a 215MW power plant to build PCA model, and sensor fault is simulated for fault detection testing.

According to the correlation analysis of the data, 10 variables are selected, such as feedwater pressure, low-pressure cylinder exhausting pressure, main steam flow rate, pressure behind feedwater valve, pressure before feedwater valve, reheater steam temperature, circulating water temperature I, circulating water temperature II, superheater spray flow rate, reheater spray flow rate.

Two sets of process data are retrieved when thermal process is in stable condition. The first data set which has 700 samples is used as sample data for modeling. The second data set which has 60 samples was added with fault for verifying the fault detectability of different indices. The fault was added from the 10th sample in the second data set.

The first data set was used to build PCA model, 3 principal components was seleted for their contribution rate is 84%.

According to the improved PCA method, PV variables are selected for the multiple correlation coefficient is bigger than 0.85 and the absolute value of the correlation coefficient is bigger than 0.1. $\{x_1, x_2, x_6, x_7, x_8\}$ is selected as PV variables, the rest sensors are considered as CV variables. The threshold value of different fault detection indices is listed in Table.1 (NOTE: PVR and CVR can only be calculated to different sensors, so $|f_k|_{PVR}$v represents the thresholds of fault detection for both PVR and CVR).

Compared with SPE, the fault detectability of improved PCA method has been improved as shown in Table.1, except x_6. To the most sensors, $\{x_1, x_2, x_4, x_5, x_7, x_8\}$, T_H^2's fault detectable threshold is smaller than SPE, PVR and CVR, which means that T_H^2 has better fault detectability than the other indices. Bias faults, 0.4 and 1.1, are introduced into x_7, as condition I and II. Fault detection results are shown in Figure 2. According to improved PCA method, x_7 belongs to high multiple correlation coefficient sensors, PVR should be used for fault detecion.

Figure 2 shows: SPE can only detect a few fault to bigger bias fault; PVR, CVR and T_H^2 detect the bigger bias fault just at the sample where bias fault has been introduced; SPE, PVR and CVR indices can not detect the smaller bias fault; T_H^2 can not only detect bigger bias fault, but also give sustaining alarm for smaller bias fault. This proves that SPE, PVR, CVR and T_H^2 do have different detectaility for different sensors, which means these indices should be combined to detect fault in process monitoring.

V. SUMMARIES

The fault detectabilities of SPE, PVR / CVR, T_H^2 indices are analyzed qualitatively and quantitatively. According to the qualitative analysis, SPE, PVR/CVR and T_H^2 indices have different fault detectability to different sensors. According to the quantitative analysis, PVR/CVR have better fault detectibility than SPE, while T_H^2 has better result in partial sensors than PVR/CVR, which means T_H^2 should be combined in the fault detection system based on PCA. The qualitative analysis of PVR and CVR shows that the control limit of CVR is too small to envelope most sample datas in it. This indicates MPCA-based fault detection system has large false alarm rate, which need further reaserach to confirm.

ACKNOLEDGEMENT

This work was supported by "The Fundamental Research Funds for the Central Universities"

REFERENCES

[1] Kourti T.; MacGregor J. F.. Process Analysis, Monitoring and Diagnosis, Using Multivariate Projection Methods. Chemometrics and Intelligent Laboratory Systems [J], 1995, 28 (1), PP:3–21.

[2] Dunia R.; Qin S. J.; Edgar T. F. et al. Identification of Faulty Sensors Using Principal Component Analysis. AIChE Journal [J], 1996, 42 (10), PP:2797–2812.

[3] Wang H.; Song Z.; Li P.. Fault Detection Behavior and Performance Analysis of Principal Component Analysis Based Process Monitoring Methods. Ind Eng Chem Res [J], 2002, 41 (10), PP:2455–2464.

[4] Qin S. J.. Statistical Process Monitoring: Basics and Beyond. Journal of Chemometrics [J], 2003, 17(8–9), PP:480–520.

[5] Feng X.; Yang, X.; Xu, Y.. Squared Prediction Error Analysis of Multivariate Statistical Process Control. Journal of Tsinghua University (Science and Technology) [J], 1999, 39(7), PP:41–45.

[6] Wang H.; Song Z.; Li P.. Improved PCA with Application to Process Monitoring and Fault Diagnosis. Journal of Chemical Industry and Engineering (China) [J], 2001, 52(6), PP:471–475.

[7] Qiu T.; Ding Y.; Wu Z.. Sensor Fault Detection Statistics Based on Principal Component Analysis. Journal of Tsinghua University (Science and Technology) [J], 2006, 46(8), PP:1147–1450.

[8] Huang X.; Niu Z.; Niu Y. et al. Behavior Analysis of Sensor Fault Detection Using PCA Approach. Ind Eng Chem Res [J], 2003, 16 (4), PP:419–423.

[9] Huang X.; Liu J.; Niu Y.. Application of PCA-based Fault Detection in Thermal Boiler Processes. Power Engineering [J], 2004, 24(4), PP:542–547.

[10] Dunia R.; Qin S. J.; Edgar T. F. et al. Use of Principal Component Analysis for Sensor Fault Identification. Computers and Chemical Engineering [J], 1996, 20(Suppl.), PP:S713–S718.

[11] Wang H.; Song Z.; Li P.. Study on the Fault Detectability of Principal Component Analysis. Chinese Journal of Scientific Instrument [J], 2002, 23(3), PP:232–240.

[12] Wang H.; Song Z.; Wang H.. Fault Detection Behavior Analysis of PCA-based Process Monitoring Approach. Journal of Chemical Industry and Engineering (China) [J], 2002, 53 (3), PP:297–301.

[13] Jackson J. E.; Mudholkar G. S.. Control Procedures for Residuals Associated with Principal Component Analysis. Technometrics [J], 1979, 21(3), PP:341–349.

[14] Jackson J. E.. A User's Guide to Principal Components [M], 2003, John Wiley & Sons Inc.

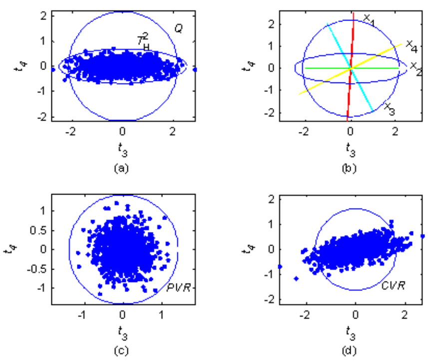

Figure 1 Comparison among the control limits of SPE, PVR, CVR and T_H^2

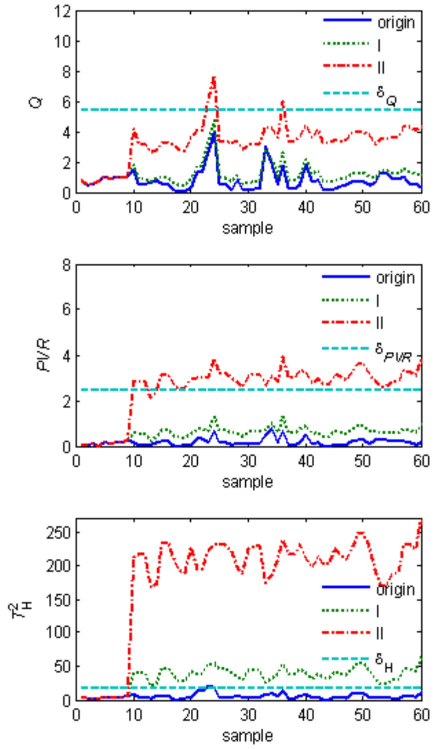

Figure 2 Fault detection result for bias fault

TABLE I. THRESHOLDS UNDER NECESSARY CONDITION FOR FAULT DETECTION

sensors	$\lvert f_k \rvert_{SPE}$	$\lvert f_k \rvert_{PVR}$	$\lvert f_k \rvert_{H}$
x_1	2.2526	1.1947	0.9386
x_2	4.8559	3.2442	0.4499
x_3	66.3125	41.1624	41.2026
x_4	0.5476	0.3352	0.3307
x_5	1.5447	0.8181	0.5609
x_6	35.2096	77.1744	39.5725
x_7	8.6739	4.6090	0.4296
x_8	8.2347	4.3656	0.6236
x_9	33.0671	21.3747	38.563
x_{10}	2.3499	1.6309	3.3966