Fault Identification using Kernel Principle Component Analysis

Mohamed A. Bin Shams*

Department of Chemical Engineering, University of Bahrain, P.O Box 32038, Isa Town, Kingdom of Bahrain (Tel: +97317684844;email: mshams@eng.uob.bh)

Abstract: In this paper, a new fault identification procedure based on Kernel Principal Component Analysis (KPCA) is proposed. In contrast to the linear PCA, a key disadvantage of the KPCA is that it cannot be used directly for fault identification, e.g. using contribution plots. Therefore, a fault identification strategy based on the power series approximation of the kernel functions is proposed. To demonstrate the functionality of the proposed algorithm, a nonlinear system is used. The proposed method is found effective in identifying the fault relevant variables for the individual and simultaneous occurrence of faults.

Keywords: Fault identification, KPCA, Multivariate statistics, Process safety, Fault detection

1. INTRODUCTION

The prompt detection and precise diagnosis of faults become a main requirement for any enterprise for safe, optimal and profitable operation (Isermann, 2006; Venkatasubramanian, 2003). For that reason, the problem of fault detection and diagnosis (FDD) for industrial processes has received considerable attention during the last two decades (Yoon and MacGregor, 2004; Chiang and Braatz, 2001; Kourti and MacGregor, 1995; Ku *et al*., 1995). Although of the nonlinear characteristics of most of the chemical and biochemical processes, the majority of the fault detection and diagnosis methods are of linear nature e.g. PCA/PLS. However, linear based fault detections method can show degraded performance when the process exhibits strong nonlinear correlations between its variables. As a result, different nonlinear fault detection and diagnosis method have been proposed (Cheng *et al*., 2010; Choi *et al*., 2005; Lee *et al*., 2004; Dong and McAvoy, 1996; Kramer, 1991; Maulud *et al*., 2006). For PCA based detection and diagnosis, two spaces are defined, namely, the score and the residual spaces (Qin, 2003). The nonlinearity issue can be interpreted differently from detection or diagnosis perspectives. From detection perspective, it is important to capture the nonlinear cross correlations between the variables used for monitoring. These nonlinear steady state correlations have to be interpreted correctly within the calibration model. Any exceeding of the normal limits of the variables constrained by the identified nonlinear correlations will be detected within the score space. Any break of these correlations will be detected in the residual space. Hence, whether the fault's variation is within or outside the identified score space, the precise interpretation of these nonlinear correlations is crucial for accurate detection. On the other hand, nonlinearity issue in the diagnosis problem manifest itself through the nature of the classifier separate the different classes. If the different classes can be separated using linear classifier, then the problem is linear, otherwise, it is nonlinear. Among the different nonlinear fault detection and diagnosis algorithms, Kernel Principal Component Analysis (KPCA) is found most attractive (Scholkopf *et al*., 1998). KPCA combines the linear PCA with the kernel trick so that the nonlinearity is implicitly accounted for. Following the detection of a fault, identifying the variables correlated to the occurred fault is of great significant. Contribution plot is the main used diagnosis tool when fault historical data is not available (Miller *et al*., 1993). Cho *et al*., (2005) proposed a gradient based method to calculate the contribution of the variables to the Hotelling's T^2 and squared prediction errors Q statistics when a fault occurred. They treated the problem of fault identification as the variable selection problem in pattern recognition. In the latter, sensitivities of a predefined criterion e.g. misclassification, is calculated to select a subset of input variables (Rakotomamonjy, 2003). Accordingly, the identification strategy of Cho *et al*., (2005) is based on estimating the sensitivities of the monitored statistics i.e. T^2 and Q with respect to each monitored variables. Although it is solely KPCA based identification strategy available in the literature, a key disadvantage of their method is the need to analytically calculate the gradient for each statistic with respect to each variable, which becomes prohibitive when the number of variables increases or when the used kernel function becomes more involved. To exploit the powerful of KPCA to handle nonlinearity and to overcome the aforementioned limitation associated with the KPCA fault identification procedure, we propose a simple fault identification strategy that approximates the kernel function using power series. Following, the approximation, the proposed strategy can be used in a similar manner as the linear PCA based contribution plots. In addition, the proposed methods can be easily generalized to any kernel function using an appropriate series expansion.
An overview of the use of linear PCA for detection and diagnosis is given in section 2. An overview of the kernel principle component analysis (KPCA) for fault detection is given in section 3. The proposed KPCA based identification

strategy is presented in section 4. Section 5 illustrates the performance of the proposed identification procedure using a nonlinear case study. Conclusions are given in section 6.

2. FAULT DETECTION AND IDENTIFICATION USING LINEAR PCA

2.1 Fault detection using linear PCA

For a process with n measurement variables, one alternative is to use n univariate control charts to monitor the process. As mentioned above, to simplify the presentation of information, a second alternative consists of using Principal Component Analysis (PCA) model to produce T^2 and Q charts for monitoring the n variables simultaneously. PCA involves the computation of loadings and scores using the covariance matrix of the data $X \in R^{m\times n}$; where n is the number of variables and m is the total number of samples. If the original variables are correlated, it is possible to summarize most of the variability present in the n variables space in terms of a lower p dimensional subspace (p<<n) where p represents the number of the principal components. If only two principal components are found, two dimensional score plots are used (i.e. T_1 versus T_2). For more than two principal components, Hotelling's T^2 and Q statistics are usually used to monitor the process. The T^2 statistic, based on the first p PCs, is defined as

$$T^2 = \sum_{i=1}^{p} \frac{t_i^2}{\lambda_i} \tag{1}$$

where λ_i is the i-th eigenvalue of the covariance matrix of the original data matrix. Confidence limit for T^2 at significance level (1-α) are related to the F-distribution as follows:

$$T^2_{m,a} = \frac{(m-1)a}{(m-a)} F_{a,m-a} \tag{2}$$

where $F_{a,m-a}$ is the upper 100α% critical point of the F-distribution with a and (m-a) degrees of freedom. Monitoring the process variables by the T^2 values based on the first a principal components is often not sufficient since it only help detecting whether or not the variation is within the plane defined by the first a principal components which generally capture the steady state correlations corresponding to normal operation of the process. In the occurrence of new events, not represented by the calibration data set used to identify the reference model, additional principal components may become significant and the new observation vector x_i will move off the calibrated plane. Such new events can be detected by computing the squared prediction error or Q statistic. Let denote the i-th multivariate observation vector whose corresponding score is $t_i = x_i \times P$. The prediction from the PCA model for x_i is given by:

$$Q = e_i e_i^T \tag{3}$$

Accordingly Q can be viewed as a measure of plant-model mismatch. The confidence limits for Q are given by Jackson (1991). This test suggests the existence of abnormal condition when $Q > Q_\alpha$, where Q_α is defined as follows:

$$Q_\alpha = \Theta_1 \left(1 + \frac{c_\alpha h_0 \sqrt{2\Theta_2}}{\Theta_1} + \frac{\Theta_2 h_0 (h_0 - 1)}{\Theta_1^2} \right)^{1/h_0} \tag{4}$$

$$\Theta_i = \sum_{j=p+1}^{h} \lambda_j^i \; ; \; i = 1,2,3 \tag{5}$$

c_α are the confidence limits for the (1-α) percentile in a standard normal distribution. These confidence limits are calculated based on the assumptions that the measurements are time independent and multivariate normally distributed.

2.2. Fault isolation using linear PCA

Contribution plots have been used as main tools for fault isolation when fault historical data is not available (Miller *et al*., 1993; Qin, 2003). Two contribution plots are usually used to identify those variables affected by the presence of faults, namely, the contribution plots related to either T^2 or Q statistics. The total contribution of variable j to the Q statistic at each sampling instant i is given by:

$$Cont_{ij} = e_{ij}^2 \tag{6}$$

On the other hand, the contribution of variable j to the T^2 statistic for principal components at each sampling instance i is given by:

$$Cont_{ig} = \sum_{k=1}^{a} \left(\begin{array}{l} \dfrac{p_{jk}^2}{\lambda_k} x_{ig}^2 + \dfrac{2 p_{jk}}{\lambda_k} \left(\sum\limits_{\substack{r=1 \\ r \neq j}}^{n} p_{rk} x_{ir} \right) x_{ig} + \\ \dfrac{1}{\lambda_k} \left(\sum\limits_{\substack{r=1 \\ r \neq j}}^{n} p_{rk} x_{ir} \right)^2 \end{array} \right) \tag{7}$$

where p_{jk} is the jk element of the loading matrix. To obtain the total contribution of the variables g within a specific time period, the corresponding $Cont_{ig}$ is summed over the required time window. As shown in equation (7) the contribution of the g variable to the T^2 statistic consists of three terms. The first term includes solely variable g and the second term contains a cross-product between variable g and the rest of variables. The last term does not contain x_{ig}, so it does not contribute to the $Cont_{ig}$ calculation.

3. KERNEL PRINCIPAL COMPONENT ANALYSIS

3.1 Fault detection using KPCA

Compared to the various nonlinear PCA algorithms, the one proposed by Scholkopf *et al*., (1998) has the advantage of being a straightforward generalization of the linear PCA explained in section 2.1. Suppose $X \in R^{m\times n}$ and $x_i \in R^n$ be the sample vector at instant i. The idea of all kernel based algorithms is to map the sample vector x_i in the low dimension space to an arbitrary large dimension space F using a nonlinear mapping Φ, that is, $\Phi : R^n \longrightarrow F$; such that the nonlinearly correlated variables in the measurement space R^n can be linearly correlated in the feature space F.
The large dimension space F is known as the feature space as opposed to the original measurements or input space. Similar to linear PCA, KPCA uses the covariance matrix in the

feature space F as the building block to obtain the loading matrix. The covariance matrix in the mapped space is given as:

$$C = \frac{1}{m}\sum_{i=1}^{m}\Phi(x_i)\Phi(x_i)^T \tag{8}$$

The loading matrix is the eigenvectors of the covariance matrix C, that is

$$\lambda v = Cv = \frac{1}{m}\sum_{i=1}^{m}\left(\Phi(x_i).v\right)\Phi(x_i) \tag{9}$$

?

where (.) is the dot product, λ is the eigenvalue and v is the eigenvector of the covariance matrix C. From Equation (9), it can be seen that the solution v can be regarded as a linear combination of the mapped measurement vector $\Phi(x_i)$, that is

$$v = \sum_{i=1}^{m}\alpha_i\Phi(x_i) \tag{10}$$

To avoid the explicit nonlinear mapping, the kernel trick is used (Scholkopf *et al.*, 1998). Different kernel functions that satisfied Mercer's theorem can be used to achieve implicit nonlinear mapping. Examples of the kernel functions are the radial basis kernel, the polynomial kernel and the sigmoid kernel and are given respectively as:

$$k(x,y) = \exp(-\|x-y\|^2/c) \tag{11}$$

$$k(x,y) = \langle x,y\rangle^d \tag{12}$$

$$k(x,y) = \tanh(\beta_0\langle x,y\rangle + \beta_1) \tag{13}$$

where d, β_0, β_1, and c are specified by the user. A specific choice among these functions implicitly specifies the nonlinear mapping Φ and the feature space F. Using the kernel trick, $k_{ik} = \Phi(x_i).\Phi(x_k)$; the eigenvalue problem in equation (9) can be rewritten as

$$\lambda\alpha = (1/m)K\alpha \tag{14}$$

where $\alpha = [\alpha_1\,\alpha_2 \ldots \alpha_m]^T$ and $K \in R^{m\times m}$ is the Gram matrix with elements k_{ik}. Since it is difficult to centre the data in the feature space, the Gram matrix can be mean centred instead. The centred Gram matrix can be given as $K_{centred} = K - KE - EK + EKE$ where $E \in R^{m\times m}$ and each element has the value 1/m. Similar to the linear PCA, the score value $t_{k,l}$ of the k-th measurement vector projected on the l-th principal component is given by

$$t_{k,l} = \sum_{i=1}^{m}\alpha_i^l k(x_k, x_i) \tag{15}$$

where α_i^l is the i-th element of the l-th eigenvector calculated from equation (14) and l=1,2,…p, where p is the total number of principal components retained and $k(x_k, x_i)$ is the k_i element of the Gram matrix. The elements $t_{k,l}$ are obtained per equation (15) and the process is monitored using equation (1) with a threshold given by equation (2). On the other hand, the residual space can be monitored using the square prediction error or Q which is given by (Lee *et al.*, 2004):

$$Q = \left\|\Phi(x) - \hat{\Phi}(x)\right\| = \sum_{j=1}^{m} t_j^2 - \sum_{j=1}^{p} t_j^2 \tag{16}$$

The threshold of the above statistic is estimated using equations (4) and (5).

4. FAULT IDENTIFICATION STRATEGY FOR KPCA

As opposed to the linear PCA, the direct use of the contribution plots, i.e. equations (6) and (7) is not straightforward. This is mainly due to the difficulty associated with inverting kernel functions e.g. equations (11), (12) and (13). Suppose that radial basis function, equation (11), is used to form Gram matrix $K_{centred}$; and the contributions of the n variables of the sample vector x_i to the monitoring statistics are sought. An appropriate mathematical series can be used to approximate the used kernel function. For example, suppose that radial basis function is chosen. The radial basis function can be approximated using a power series expansion. In particular, equation (11) can be written as:

$$k(x_i, y_k) = \exp\left(\frac{-x_i^T x_i + 2x_i y_k - y_k^T y_k}{c}\right) \tag{17}$$

Since the last term, i.e. $y_k^T y_k$, does not contain x_i, hence, can be omitted from the contribution calculation. Therefore, equation (17) can be given as:

$$k(x_i, y_k) = \exp(-(x_1^2 + \ldots + x_n^2) + 2(y_1x_1 + \ldots + y_nx_n)) \tag{18}$$

The y_k elements are belong to the training set and will be denoted as $c_1,\ldots c_n$, since they are constants and not contributing to the current sample vector. Furthermore, the argument of the exponential function is a scalar quantity (z), hence it can be written as $k\,(x_i,y_k)$=exp(-z). The latter can be expanded as a power series:

$$e^{-z} = \sum_{n=0}^{\infty}\frac{(-z)^n}{n!} \tag{19}$$

Figure 1 depicts the approximation with n=1, 2,...7. As can be seen from figure 1, the exponential function can be approximated, to certain accuracy, using n terms of equation (19).

This in turn will allow the interpretation of the variable contributions to the monitoring statistics in similar way as done with the linear PCA. By expanding equation (17) using equation (19), the elements of the Gram matrix can be written explicitly as a function of each variable x_g. For example, for n=0,...3, the Gram matrix element can be given by equation (20). The generalization to higher n is straightforward. Hence, an explicit contribution of the different variables of the sample vector x_t to the monitoring statistics i.e. T^2 and Q can be obtained in a similar manner as the linear PCA based contribution plots as in equations (6) and (7). In particular, for T^2 statistics, the total contribution of the g variable over the monitoring window 1 to m is given by equation (21).

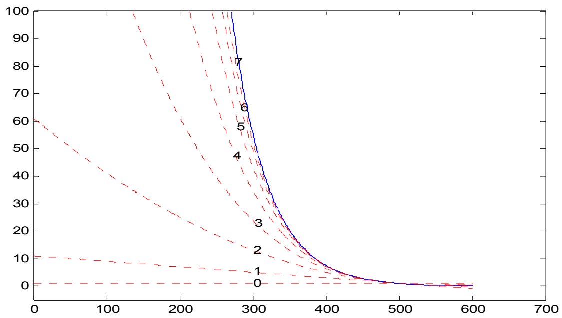

Fig.1 The power series expansion of the radial basis function for n=1,2…,7. The solid line represents the analytical function.

$$k(x_i, y_k) = \left(\begin{array}{l} 2c_g + (2c_g - 1)2\sum_{\substack{u=1 \\ u \neq g}}^{n} c_g x_u^2 - \\ \sum_{\substack{u=1 \\ u \neq g}}^{n} x_u^2 + ((2c_g - 1) \\ (2\sum_{\substack{u=1 \\ u \neq g}}^{n} c_g x_u^2 - \sum_{\substack{u=1 \\ u \neq g}}^{n} x_u^2)^2 / 2) - 1 \end{array} \right) x_g^2 + \left(\begin{array}{l} ((2c_g - 1)^2 / 2) + \sum_{\substack{u=1 \\ u \neq g}}^{n} x_u^2 + ((2c_g - 1)^2 \\ (2\sum_{\substack{u=1 \\ u \neq g}}^{n} c_g x_u^2 - \sum_{\substack{u=1 \\ u \neq g}}^{n} x_u^2)^2 / 2) \end{array} \right) x_g^4 + \left(\frac{(2c_g - 1)^3}{6} \right) x_g^6 \tag{20}$$

$$T_g^2 = \sum_{i=1}^{m} \left[\begin{array}{l} \left(\sum_{j=1}^{a} \alpha_{ij} \left(\sum_{q=1}^{m} \frac{\alpha_{qj}}{\lambda_j} (k(x_t, x_q) - \frac{1}{m} \sum_{z=1}^{m} k(x_t, x_z)) \right) \right) . \\ \left(k(x_t, x_i) - \frac{1}{m} \sum_{z=1}^{m} k(x_t, x_z) \right) \end{array} \right] \tag{21}$$

where $k(x_t,x_i)$ is given by equation (19) and α_{qj}, λ_j are the qj-th element of the loading matrix and the j-th eigenvalue, respectively. The derivation of equation (21) is straightforward using equations (15) and (1). The mapped data are centred and all terms that have not contained x_g are omitted. Using equation (16), the contribution of the variable g to the Q statistic is similarly given as:

$$Q_g = \sum_{i=1}^{m} \left[\begin{array}{l} \left(\sum_{j=1}^{m} \alpha_{ij} \left(\sum_{q=1}^{m} \frac{\alpha_{qj}}{\lambda_j} (k(x_t, x_q) - \frac{1}{m} \sum_{z=1}^{m} k(x_t, x_z)) \right) \right) \\ \left(k(x_t, x_i) - \frac{1}{m} \sum_{z=1}^{m} k(x_t, x_z) \right) \end{array} \right] - \sum_{i=1}^{m} \left[\begin{array}{l} \left(\sum_{j=1}^{a} \alpha_{ij} \left(\sum_{q=1}^{m} \frac{\alpha_{qj}}{\lambda_j} (k(x_t, x_q) - \frac{1}{m} \sum_{z=1}^{m} k(x_t, x_z)) \right) \right) \\ \left(k(x_t, x_i) - \frac{1}{m} \sum_{z=1}^{m} k(x_t, x_z) \right) \end{array} \right] \tag{22}$$

4. CASE STUDY

To investigate the efficiency of the proposed identification strategy, the nonlinear system given by Dong and McAvoy (1996) will be used and is given by:

$$x_1 = t + e_1 \tag{23}$$
$$x_2 = t^2 - 3t + e_2 \tag{24}$$
$$x_3 = -t^3 + 3t^2 + e_3 \tag{25}$$

where e_1,e_2 and e_3 are N(0,0.01) and t is uniformly distributed in the rang [0.01,2]. A total of 300 points are generated. The fault onsets at sample k=101. Three faults are considered: (1) A step change of -0.4 in x_2; (2) Linear increase in x_1 from sample 101 to 270 according to 0.01(k-100), where k is the sample number; (3) A step change of -1.5 starts at sample k=101 in x_2 and a linear increase in x_1 from sample 101 to 170 according to 0.01(k-100), occur simultaneously. Following the successful detections, the identification of the variables contributing to the occurred abnormality is an important task. Figure 2 depicts the detection results for fault (1) using the T^2 and Q statistic. Both statistics fails to observe the occurrence of fault (1). It is expected then that the contribution plots will perform poorly as confirmed by figure 3. As can be seen figure 3, an ambiguous conclusion can be obtained from the PCA based contribution plots. On the on the hand, figure 4 shows the detection results in response to fault (1) using KPCA. Although the Q statistic performs better than T^2 statistic in observing the occurrence of fault (1), both statistics show better response compared to their linear PCA counterpart. From the relatively improved response of the KPCA based T^2 compared to the linear PCA based T^2 in observing fault (1), it can be concluded that the score space is better represented by the KPCA than the linear PCA. In addition, figure 5 precisely identifies x_2 as the main root cause of the occurrence abnormality. Figure 6 represents the linear PCA detection's result in response to fault (2), that is, a ramp like change in x_1 in the time frame k=101 to 270. Although the Q statistic detects the occurrence of abnormality, the T^2 fails to observe any changes. Consequently, the associated contribution plots, i.e. figure 7, can not indicate the problem's root cause. However, figure 8 and figure 9 show the superiority in detecting and identifying the fault (2) relevant variable i.e. x_1, respectively. Often more than one fault can occur simultaneously. The detection and the identification of the fault's responsible variables are more challenging in the simultaneous case. The simultaneous occurrence situation is simulated using fault (3), that is, a step change in x_2 of magnitude -1.5 combines with a ramp like change in x_1. Figure 10 and figure 11 show the deficiency of linear PCA in observing this fault. Although the Q statistic detect fault (3), the contribution plots cannot clearly isolate the root causes. On the other hand, the KPCA shows a superior performance in observing the occurrence of fault (3), as shown in figure 12. Figure 13 depict the contributions of the variables to the monitored statistics. Although the difference in the contributions of x_1 and x_2 to fault (3), the contribution plot based on the Q statistic is able to identify the two variables, that is, x_1 and x_2 as the main root causes of the occurred abnormality.

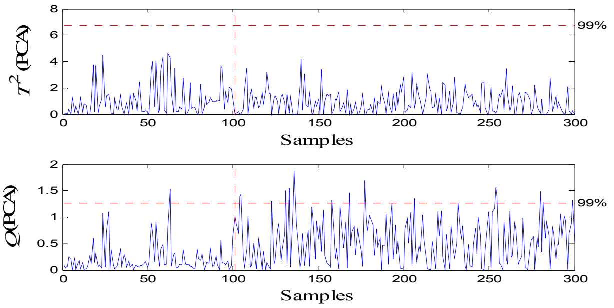

Fig.2 The detection of fault (1): A step change in x_2 (-0.4) using T^2 and Q based PCA.

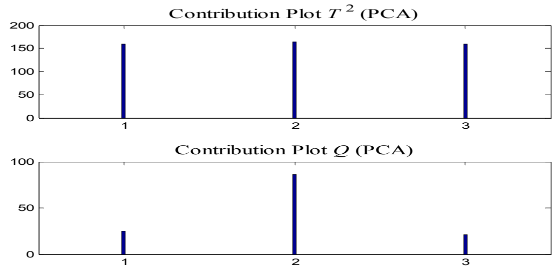

Fig.3 The contribution plots for fault (1): A step change in x_2 using linear PCA

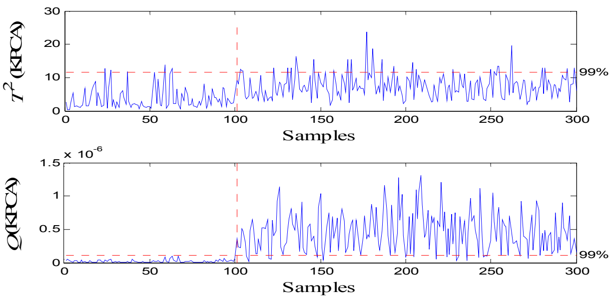

Fig.4 The detection of fault (1): A step change in x_2 (-0.4) using T^2 and Q based KPCA.

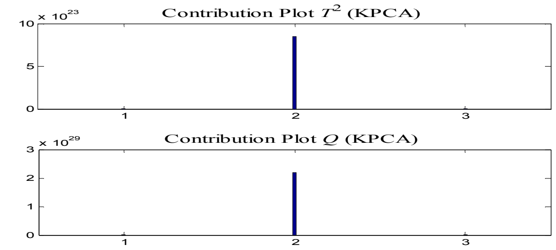

Fig.5 The contribution plots for fault (1): A step change in x_2 using KPCA. The contribution plot correctly identifies the root cause of the fault, i.e. x_2.

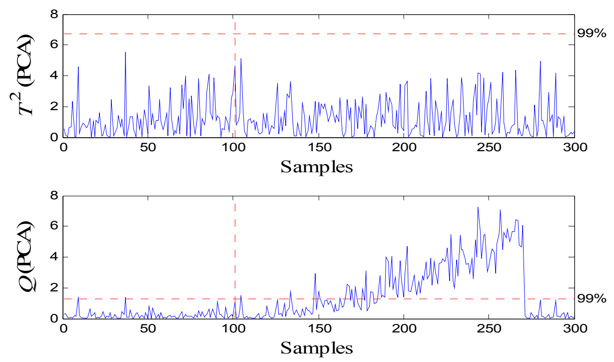

Fig.6 The detection of fault (2): A ramp like change in x_1 using T^2 and Q based PCA. Although the Q statistic detects this fault T^2 fails to detect it.

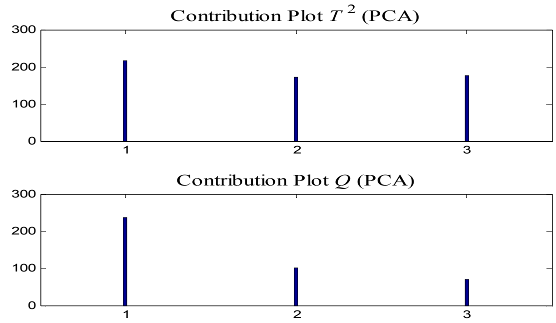

Fig.7 The contribution plots for fault (2): A ramp like change in x_1 using linear PCA. The isolation of the faulty variable is ambiguous.

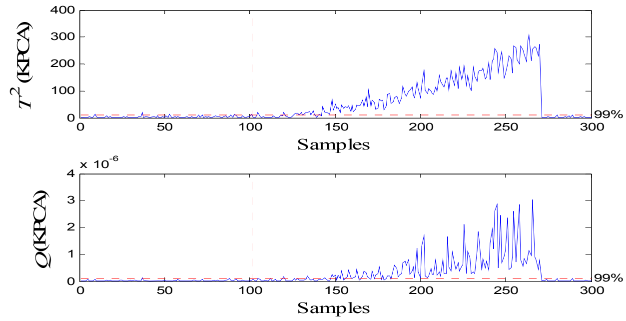

Fig.8 The detection of fault (2): A ramp like change in x_1 using T^2 and Q based KPCA. Both statistics are successful in observing this fault.

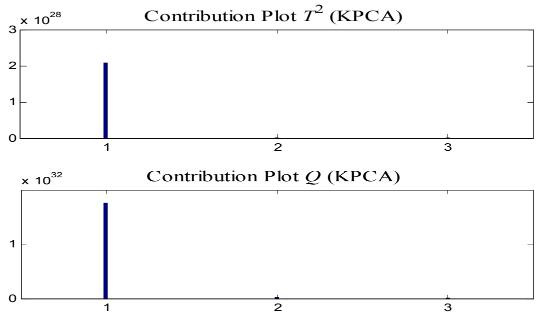

Fig.9 The contribution plots for fault (2): A ramp like change in x_1 using KPCA. The contributions plots unambiguously locate x_1 as the root cause.

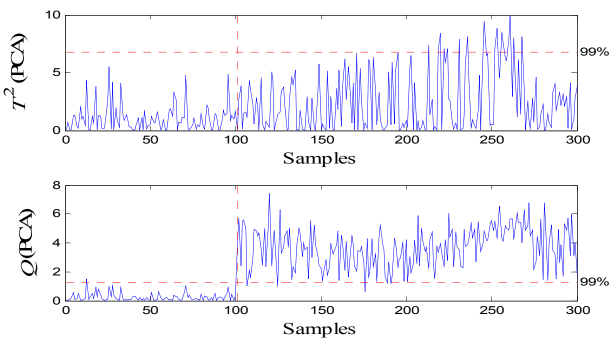

Fig.10 The detection of fault (3): A step change in x_2 (-1.5) and a ramp like change in x_1 using T^2 and Q based PCA. Although the Q statistic detects this fault, T^2 fails to detect it.

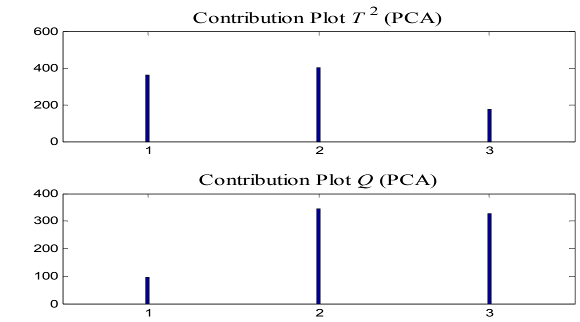

Fig.11 The contribution plots for fault (3): A step change in x_2 (-1.5) and a ramp like change in x_1 using T^2 and Q based PCA. The isolation of the faulty variable is ambiguous.

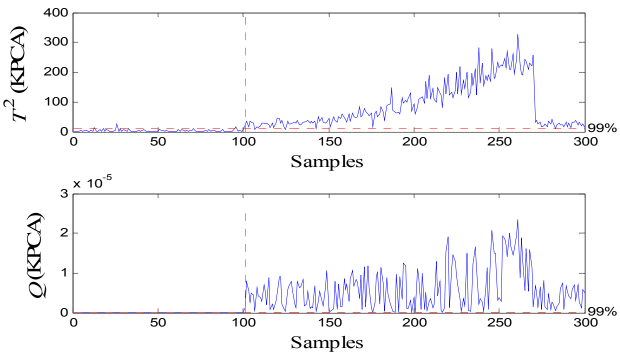

Fig.12 The detection of fault (3): A step change in x_2 (-1.5) and a ramp like change in x_1 using T^2 and Q based KPCA. Both statistics detect the occurrence of the fault.

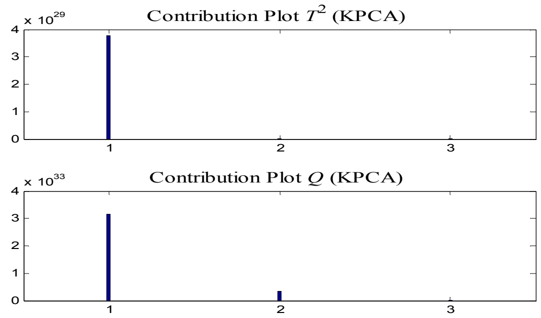

Fig.13 The contribution plots for fault (3): A step change in x_2 (-1.5) and a ramp like change in x_1. Although T^2 contribution plot locates x_1 as a possible root cause, the Q contribution identifies both, x_1 and x_2 as the root causes.

5. CONCLUSION

A new identification strategy using kernel principal component analysis (KPCA) is proposed. The proposed KPCA based identification strategy is based on the power series approximation of the kernel functions. The proposed strategy allows the use of the contribution plots in way similar to the linear PCA. Nevertheless, the proposed identification method show superior performance in identifying the main root cause in the individual as well as the simultaneous occurrence of faults. The computation complexity of the proposed KPCA identification strategy increases as the problem size increases. This is mainly due to the Gram matrix calculations. A computationally effective algorithm of the proposed identification strategy suitable for larger industrial scale problem is currently under investigation.

REFERENCES

Cheng, C.Y., Hsu, C.C., Chen, M.C. (2010). Adaptive kernel principle component analysis for monitoring small disturbances. *Industrial and Chemistry Research*, 49, 2254-2262

Chiang, L.H., Russel, E.L., Braatz, R.D. (2001). *Fault detection and diagnosis in industrial systems*. Springer-Verlag, London

Cho, J.H. Lee, J.M., Choi, S.W., Lee, D., Lee, I.B. (2005). Fault identification for process monitoring using kernel principal component analysis. *Chemical Engineering Science*, 60, 279-288

Choi, S.W., Lee, C., Lee, J.M., Park, J.H., Lee, I.B. (2005). Fault detection and identification for nonlinear processes based on kernel PCA. *Chemometrics and Intelligent Laboratory Systems, 75, 55-67*

Dong, D., McAvoy, T.J. (1996). Nonlinear principal component analysis based on principal curves and neural networks. *Computer and Chemical Engineering* 20, 65-78

Isermann, R. (2006) *Fault diagnosis systems*. Springer, Berlin

Kramer, M.A. (1991). Nonlinear principal component analysis using autoassociative neural networks. *AIChE Journal*, 37, 233-243

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

Lee, J.M., Yoo, C.K., Choi, S.W., Vanrolleghem, P.A., Lee, I.B. (2004). Nonlinear process monitoring using kernel principal component analysis. *Chemical Engineering Science*, 59, 223-234

MacGregor, J.F., Kourti, T. (1995). Statistical process control of multivariate processes. *Control Engineering Practice*, 3, 403 414

Maulud, A.,Wang, D., Romagnoli,J. (2006). A multi-scale orthogonal nonlinear strategy for multi-variate statistical process monitoring. *Journal of Process Control*, 16, 671-683

Miller, P., Swanson, R.E., Heckler, C.F. (1998). Contribution plots: the missing link in multivariate quality control. *Applied Mathematics and Computer Science*, 8, 775-792

Qin, S.J. (2003). Statistical process monitoring: basics and beyond. *Journal of Chemometrics*, 54, 480-502

Rakotomamonjy, A. (2003). Variable selection using SVM based criteria. *Journal of Machine Learning Research*, 3, 1357-1370

Scholkopf, B., Smola, A.J., Muller, K. (1998). Nonlinear component analysis as a kernel eigenvalue problem. *Neural Computation*, 10, 1299-1399

Venkatasubramanian, V., Rengaswamy, R., Kavuri, N., Yin, K. (2003). A review of process fault detection and diagnosis Part III: Process history based method. *Computer and Chemical Engineering* 27, 327-346

Yoon, S., MacGregor, J.F. (2004). Principal component analysis of multiscale data for process monitoring and fault diagnosis. *AIChE Journal*, 50, 2891-2903