AN INTELLIGENT PCA APPROACH FOR ON-LINE FAULT ISOLATION

Liu J. [*]*, Lim K.W.* [**]*, Srinivasan R.* [*]*, Doan X.T.* [*]
**Institute of Chemical and Engineering Sciences, Singapore 627833*
***Agency for Science, Technology and Research, Singapore 138671*

Abstract: This paper describes a real-time on-line fault detection and diagnosis approach using PCA (Principal Component Analysis) and KBS (Knowledge-Based system). Then this paper applies the approach to a refinery process simulation that produces cyclohexane using benzene and hydrogen. The result shows that both sensor faults and process faults can be quickly detected and sensor faults can be clearly differentiated from process faults by using the intelligent PCA approach.

Keywords: Process monitoring, Fault isolation, PCA, Process control

1. INTRODUCTION

Modern chemical plants are susceptible to various failures resulting in abnormal situation as they are highly complex and integrated, processing large volumes of materials and operating at extremes of pressure and temperature (Dash and Venkatasubramanian, 2001). However, it is often difficult to completely rely on operators to cope with such abnormal situations (Venkatasubramanian, *et al*. 2003a; Liu *et al*., 2003, Srinivasan *et al.* 2004). Industrial statistics show that about 70% of the industrial accidents are caused by human errors (Venkatasubramanian, *et al*. 2003a).

In most chemical plants, distributed control systems (DCS) are used to simultaneously control thousands of process variables such as temperature and pressure (Hamaguchi et al. 2003). Now DCS are able to read, store, present and process a large amount of real-time plant data. Operators may have neither the time, nor often, the expertise to effectively deal with such a large amount of data. Multivariate statistical approaches and principal component analysis (PCA) in particular, have been widely investigated to deal with this challenging problem (Qin 2003; Lane *et al*. 2003; Chiang *et al*. 2003, Wang et al. 2002). However, they do not possess 'fingerprint' or 'signature' properties for diagnosis, which makes the fault isolation difficult (Venkatasubramanian, *et al.* 2003b). Some techniques such as contribution charts have been proposed for fault isolation purpose (Liu *et al.* 2005). However, contribution charts can only point out the process variable dominating the error space whereas it can not differentiate whether the fault is due to sensor faults so as to isolate faulty sensors or due to process faults where much attention and further diagnosis is required.

This paper proposes an intelligent PCA method, which is a combination of PCA-monitoring approach and expert-systems approach, to differentiate sensor faults from process faults.

Then the method is applied to a refinery process simulation that produces cyclohexane by using benzene and hydrogen. Both sensor faults and process faults are considered. The real-time online simulation result shows that both sensor faults and process faults can be quickly detected and the intelligent PCA method can clearly differentiate sensor faults from process faults.

The rest of this paper is organized as follows. Section 2 gives a theoretic review of conventional PCA. Section 3 describes methodology of the intelligent PCA method for real-time on-line process monitoring and fault isolation. Section 4 applies this approach to a refinery process simulation. Section 5 gives some discussions and Section 6 gives some conclusions.

2. PRINCIPAL COMPONENT ANALYSIS

Consider a data matrix **X** ($n*m$) representing n observations of m variables, PCA transforms it as following.

$$\mathbf{X} = \mathbf{T}\mathbf{P}^{\mathrm{T}} = \sum_{i=1}^{m} \mathbf{t}_i \mathbf{p}_i^{\mathrm{T}} \quad (1)$$

Where $\mathbf{p}_i$ is called principal component loading vector and

$$\mathbf{P}^{\mathrm{T}}\mathbf{P} = \mathbf{I} \quad (2)$$

$\boldsymbol{t}_i$ is called principal component score vector and

$$\mathbf{T}^{\mathrm{T}}\mathbf{T} = \mathbf{\Lambda} = \mathrm{diag}(\lambda_1, \lambda_2, \cdots, \lambda_m) \quad (3)$$

Where λ_i (i=1, 2, … , m) is eigenvalue of $\mathbf{X}^{\mathrm{T}}\mathbf{X}$ and

$$\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_m \quad (4)$$

Normally, first k (k<<m) principal components capture most of the variance, while less important components, which mostly describe noise information in the data, can be abandoned without loss of significant information. By doing so, matrix **X** can be reconstructed as

$$\mathbf{X} = \hat{\mathbf{X}} + \mathbf{E} = \sum_{i=1}^{k} \mathbf{t}_i \mathbf{p}_i^{\mathrm{T}} + \mathbf{E} = \mathbf{T}_k \mathbf{P}_k^{\mathrm{T}} + \mathbf{E} \quad (5)$$

The principal component score vectors $\mathbf{t}_1$, … , $\boldsymbol{t}_k$, span a lower dimensional subspace used for further analysis.

PCA can be used in real-time online process monitoring and fault diagnosis to check shift of operating point, sensor failure, process fault and plant disturbance. This is done via statistical hypothesis tests on two indices, Hotelling's T^2 and Q statistics, respectively in the principal component subspace and residual space. T^2 statistic is a scaled squared 2-norm of an observation vector **x** from its mean, while Q statistic, also known as squared prediction error (SPE), is mathematically the total sum of residual prediction errors.

3. AN INTELLIGENT PCA APPROACH FOR MONITORING AND FAULT ISOLATION

The algorithm for on-line real time process monitoring and fault diagnosis using PCA consists of the following steps

1. Build a PCA model using historic plant data representing normal operating condition or step response plant test data within normal operation range, and then check T^2 and SPE for real-time samples on violation of control limits (Liu *et al.* 2005).
2. If SPE or T^2 exceeds the control limit, a fault is detected and an alarm is set, indicating that the process is in abnormal status. When both SPE and T^2 exceed the control limits, go to step 3 for fault isolation
3. If one process variable is dominating the SPE (such as contributing more than 60%), build a new PCA model excluding this process variable from historic data or plant test data and check SPE and T^2 .
4. If the real-time sample is still out of both control limits using the new PCA model, it means that other variables also contribute significantly to the fault. Hence it's a process fault and it's an important alarm to the operator.

5. If the real-time sample is within both SPE and T^2 control limits using the new PCA model, it can be determined that there is a sensor fault for this variable as only this process variable contributes significantly to the fault. In this case, the fault is not an important alarm to the operator.
6. If there is a sensor fault, the original PCA model will be not used for monitoring until the real-time sample goes within both SPE and T^2 limits again using the original PCA model, which means that the sensor fault has been corrected. During the sensor fault, the new PCA model excluding the sensor-faulty variable is used for process monitoring.

4. CASE STUDY

A refinery process case study is used to illustrate the approach proposed in this paper. The process flow sheet is shown in Figure 1. It is the reaction section of a cyclohexane plant, including 2 reactors and 1 steam drum. The objective of the process is to use benzene and H2 to produce cyclohexane. As it is an exothermic reaction, the heat generated is utilized to produce steam.

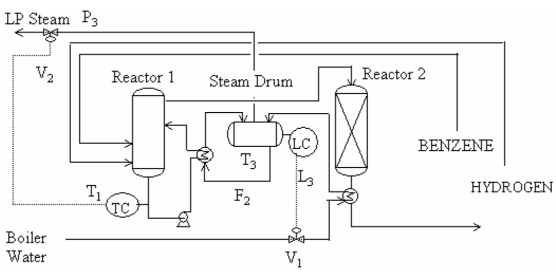


Figure 1 Flow sheet of the refinery process

The first step of the algorithm is to build a PCA model for normal operating condition. As there is no enough historic data representing normal operating condition, we do plant test via simulation. Plant step test signal is chosen as following: set-point of L3 is changed from 65 to 64 at second 1000 and back to 65 at second 2000. Step response of this process variable is shown in figure 2. After plant test, collect plant data, scale it and build a PCA model and then use this PCA model for real-time on-line process monitoring and fault isolation.

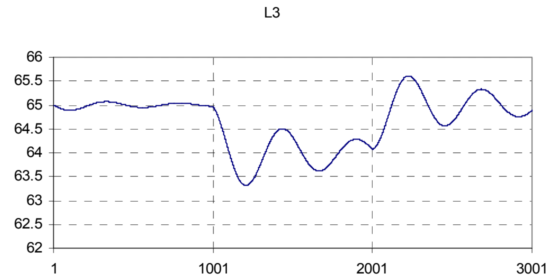


Figure 2 Step response of plant test

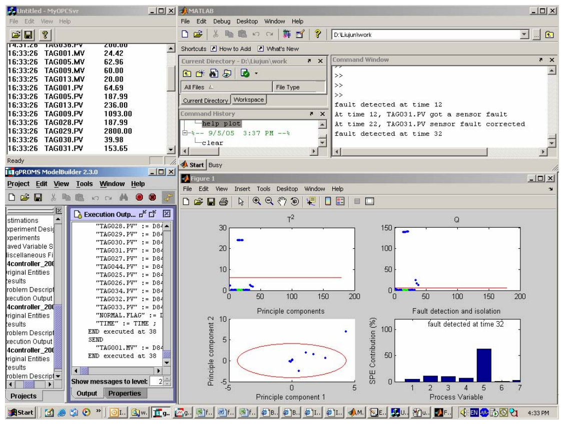


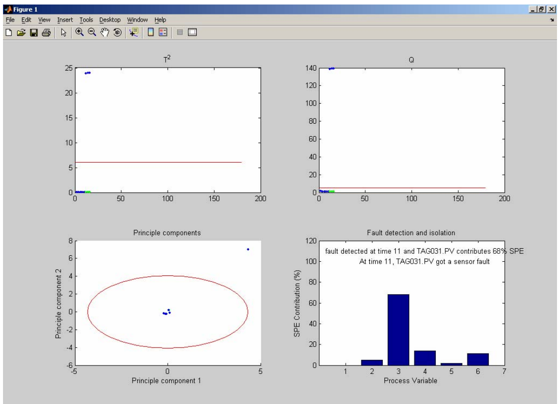
Figure 3 Real-time simulation and PCA monitoring

The real-time process simulation and PCA monitoring is shown in Figure 3. The process simulation is done in gPROMS and it sends process data to an OPC (OLE for Process Control) server every second. However, the real-time PCA monitoring is done in MATLAB roughly every 2 seconds as 1 second seems not enough for the application to access the process data, do the calculation and display the result. Both sensor fault and process fault are considered. A drift in the measurement sensor of the steam drum temperature (T3) is introduced between second 10 and second 20, and between second 100 and second 110. Boiler water pressure loss is introduced between second 30 and second

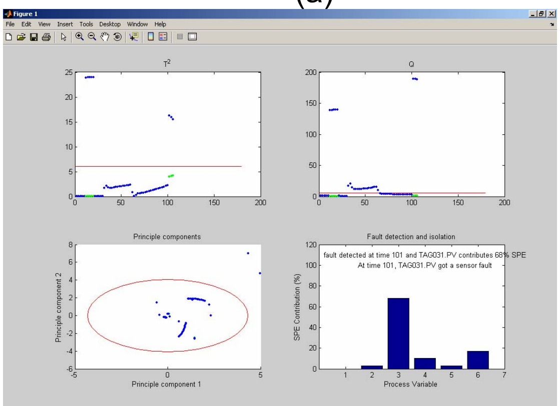
60. Pump failure is introduced between second 160 and second 170.

Figure 4 (a) shows that a fault is detected at second 11. At the same time, both T^2 limit and SPE limit are exceeded. Fault isolation task is started. As process variable T3 (TAG031.PV) contributes 68% of SPE, we build a new PCA model excluding this variable and monitor T^2 and SPE using this new PCA model. From figure 4 (a) it can be seen that the process is within both T^2 and SPE control limits (green dots in the figure) when using the new PCA model while the process is out of both control limits using the original PCA model (blue dots). It means that only this variable has contributed to this fault. Considering that a process fault will affect various process variables, it can be determined that it's a sensor fault. As shown in figure 4 (a), the algorithm not only detected the fault quickly, but also successfully isolated this fault as a sensor fault. Real-time simulation result shows that correction of the sensor fault can also be quickly detected (Correction is detected at second 21 while it occurs at second 20). After correction of sensor failure, the original PCA model is used for real-time monitoring again. Figure 4(b) shows that the sensor fault between second 100 and 110 is also quickly detected and isolated.

Figure 5 shows that a fault is detected at second 161. At the same time, both T^2 limit and SPE limit are exceeded. Fault isolation task is started. As process variable F2 (TAG044.PV) contributes 68% of SPE, we build a new PCA model excluding F2 and monitor T^2 and SPE using the new PCA model. From figure 5 it can be seen that the process is still out of both T^2 and SPE control limits (green dots in the figure) when using the new PCA model. It means that except for the dominating variable, other process variables also contribute significantly to the fault. Therefore it is a process fault. Although it does not pinpoint the exact root cause of the fault, it does show that the dominating variable is the liquid flow entering the boiler (heat exchanger) of reactor 1. An experienced operator will be able to find out the real fault (pump failure) easily according to the above information. Therefore, the proposed intelligent PCA approach can quickly detect both sensor faults and process faults. It can also clearly differentiate sensor faults from process faults.



(a)



(b)

Figure 4 Sensor faults immediately detected and isolated

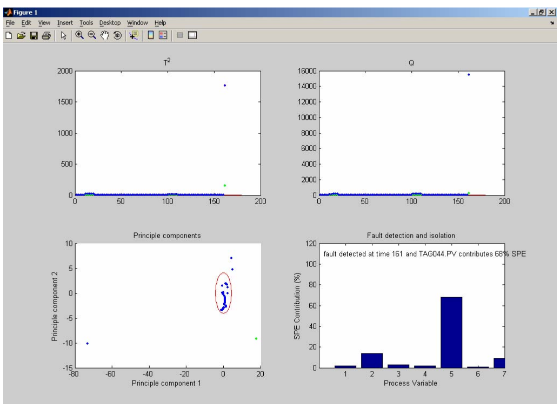


Figure 5 Pump failure immediately detected and dominating variable isolated

Figure 6 shows that a fault is detected at time 31. However, T^2 chart shows that the T^2 control limit is not exceeded, which means that the real-time sample is quite close to the mean under normal operation although correlation between the process variables has been broken. At time 65, the approach detected that the process is within both control limits and displayed normal operation in fault detection and isolation chart. The operator does not need to be worried about this fault any more as the fault has been corrected and it does not result in significant change in operating states of the process.

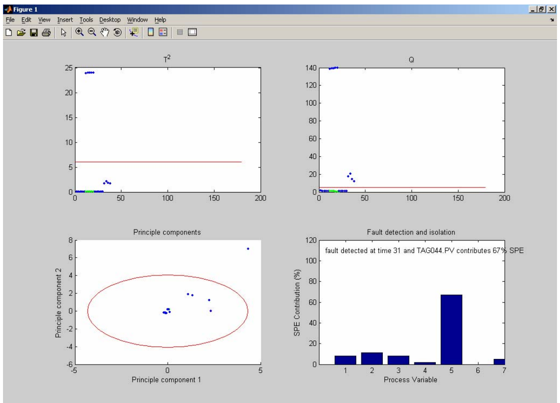


Figure 6 Boiler water pressure loss detected

5. DISCUSSIONS

Both T^2 chart and Q chart (SPE) can be used for fault detection. From Figure 4, 5 and 6, it can be seen that Q chart (SPE) is more sensitive than T^2 chart to check for possible sensor faults or process faults where correlations among the variables no longer hold. Therefore, Q chart is more suitable for fast detection of both sensor faults and process faults or early detection of small faults that will gradually become significant fault such as failure of critical equipment, while T^2 is more suitable for detecting changes in operating states of the process. In this paper, if any control limit of T^2 or SPE is exceeded, a fault is detected. This is to ensure fast detection of a fault. However the diagnosis or isolation task is started only after both control limits are violated. This is to obtain enough process information for diagnosis and to avoid diagnosis error.

To build a right PCA model for normal operating condition, sufficient historic plant data representing normal operating condition is necessary. When such data is not available, plant test within normal operation range can be conducted to collect the required data. Some guidelines for choosing the step test signals are as follows.

- Choose the magnitude of the step test signal so that the response can cover the normal operation range as more as possible.
- Choose suitable time length of the step test signal so as to cover enough process dynamic information (for example, two times of the response time period)

6. CONCLUSIONS

An intelligent PCA approach for real-time on-line process monitoring and fault isolation is proposed. This approach is applied to a refinery process simulation.

Real-time simulation results show that both sensor faults and process faults can be quickly detected with the proposed approach. The approach can also clearly differentiate process faults from sensor faults so that the operator will know whether this fault is important or not as normally sensor faults are not as important to the operator as process faults.

REFERENCES

Chiang L.H., R.J. Pell and M.B. Seasholtz (2003). Exploring process data with the use of robust outlier detection algorithms. *Journal of Process Control* **13**, 437-449.

Dash, S. and V. Venkatasubramanian (2001). Offline safety analysis for online fault diagnosis. In: *4th IFAC workshop on on-line fault detection and supervision in the chemical process industries*, 179-184. Jejudo Island, Korea.

Hamaguchi, T., Y. Hashimoto, T. Itoh, A. Yoneya and Y. Togari (2003). Abnormal situation correction based on controller reconfiguration. *Journal of Process Control* **13**, 169 – 175.

Lane, S., E. B. Martin, A. J. Morris, and P. Gower (2003). Application of exponentially weighted principal component analysis for the monitoring of a polymer film manufacturing process. *Transactions of the Institute of Measurement and Control* **25**, 17 – 35.

Liu, J., K. W. Lim, W. K. Ho, K. C. Tan , R. Srinivasan and A. Tay (2003), The Intelligent Alarm Management System. *IEEE Software*, March/April issue.

Liu, J., K. W. Lim, R. Srinivasan. On-Line Process Monitoring and Fault Isolation Using PCA. *Proceedings of the 2005 IEEE International Symposium on Intelligent Control*. Limassol, Cyprus, June 27-29, 2005, 658-661

Qin, S. J. (2003). Statistical process monitoring: basics and beyond. *Journal of Chemometrics* **17**, 480 – 502.

Srinivasan R., C. Wang, W.K. Ho and K.W. Lim (2004). Dynamic principal component analysis based methodology for clustering process states in agile chemical plants. *Ind. Eng. Chem. Res.* **43**, 2123 – 2139.

Venkatasubramanian, V., R. Rengaswamy, K. Yin and S. N. Kavuri (2003a). A review of process fault detection and diagnosis part I: quantitative model-based methods. *Computers & Chemical Engineering* **27**, 293-311.

Venkatasubramanian, V., R. Rengaswamy, K. Yin and S. N. Kavuri (2003b). A review of process fault detection and diagnosis part III: process history based methods. *Computers & Chemical Engineering* **27**, 327-346.