

## [ED02] Fault detection and diagnosis using Multivariate Statistical Process Control (MSPC)

**Mak Weng Yee, Kamarul Asri Ibrahim**

Faculty of Chemical Engineering and Natural Resources Engineering (FKKSA), Universiti Teknologi Malaysia, 81310 Skudai, Johor Bahru, Johor.

### **Introduction**

Currently, chemical plants face numerous challenges like stringent requirements are needed on the desired final product quality, utilization of a lot of energy, must be environmentally friendly and fulfill safety requirements. High operation cost is needed in order for chemical plants to overcome the stated challenges. Any faults that are present in a chemical process will yield higher operation cost on the plant due to increase in production of waste, re-work, re-processing and consumption of utilities. Therefore, accurate process fault detection and diagnosis (FDD) on a chemical process at an early stage is important to reduce the cost of operation due to present of faults.

The important task of detecting and diagnosing abnormal process behavior (faults) has led to the evolution of a range of statistically based condition monitoring approaches (Treasure *et al.*, 2004). These approaches are collectively referred to as Multivariate Statistical Process Control (MSPC) and have gained attention over the past decades noticeable by the large number of publications in this area (MacGregor and Kourti, 1995). Application of MSPC as a fault detection tool in previous works was based on two conventional control chart: Hotelling's  $T^2$  Statistic control chart and Square Prediction Error Statistic control chart (SPE) (Wachs and Lewin, 1999). These two control charts have shown good fault detection performance for simulated model unit operations (Wachs and Lewin, 1999). MSPC using the two stated conventional control charts is a very powerful tool for fault detection but its main limitation lies in the ability to isolate or diagnose the actual causes of the detected faults. The main fault diagnosis tool used together with the two control charts is the Contribution Plots (CP) (Wachs and Lewin, 1999). Although CP is used to diagnose the cause of the detected faults, they tend to be noisy and ambiguous. These plots also do not have confidence

limit/control limit, thus making it difficult to determine whether a situation is normal or abnormal (Yoon and MacGregor, 2000).

The present fault diagnosis tool using CP has limited usage in diagnosing causes of detected faults. Faults that have effect propagated into other variables are hard to be isolated using CP. In enhancing the fault isolation ability of MSPC and overcoming the ambiguity of CP, fault signatures have been proposed. Faults from process data are collected and fault signatures are developed using Principal Component Analysis (PCA). Any new detected faults will exhibit certain fault signature and this signature will be compared to the database of fault signatures developed earlier on. Good results were obtained for the application of the proposed method (Yoon and MacGregor, 2001). Although the fault signature method shown better fault diagnosis ability compared to the previous Contribution Plots, there are several weaknesses of the former method. The fault signature database needs to be as comprehensive as possible to cover all possible faults in a process and great amount of computer calculation is needed in diagnosing a fault for highly multivariable processes. The present work focuses on overcoming the ambiguity nature of fault isolation using MSPC through contribution plots and also the need for big database of faults signatures by introducing fault diagnosis using correlation coefficients of process variables and quality variables. The proposed FDD method in this paper is an extension of fault detection using correlation coefficients (Mak and Kamarul, 2003).

Correlation coefficients between key process variables and quality variables of interest are used as fault detection and diagnosis tools. These coefficients are developed from nominal operating condition (NOC) data using multivariate projection techniques such as PCA and Partial Correlation Analysis (PCorrA). PCorrA has been applied in many applications (Ding and

Nancy, 2000) and hardly been used in MSPC as a method for determining correlation between variables. The developed correlation coefficients will be used together with conventional Shewhart Control Chart and Range Control Chart as FDD tools. The proposed method is applied to a simulated industrial column model (Wong, 2003).

### Methodology

#### Process modeling and data generation

The most important part in obtaining an accurate correlation between the process variables and quality variables is the data mining section. In this research, data is obtained from a simulation model. A distillation column from a Palm Oil Fractionation Plant is selected as the case study. The model of this column is developed based on the model from literature with slight modifications to suit the present work (Wong, 2003). Figure 1 shows the distillation column with the key variables of the process. From the column model, two sets of process operating data were generated. For NOC data, some noises with zero mean were imbedded into the simulation program. The noises considered are small random change in selected key variables such as feed flow rate, feed temperature, reboiler duty, cooler duty, reflux flow rate and pumparound flow rate. While for Out-of-Control (OC) data, some large changes (significant faults) and moderate changes (insignificant faults) were purposely added into the process model as faults. These faults represent valve faults, sensor faults and controller faults. The description of each type of fault is described in Table 1. The feed flow rate and feed temperature to the study column are assumed to be fixed. Any abnormal changes of the value of these two variables are due to faults as shown in Table 1 and not due to common cause variation (NOC). The generated NOC and OC data are mean-centered and variance scaled. The NOC data will be subjected to analysis using PCA and PCorrA for deriving the correlation coefficients between the selected process variables with the selected quality variables. The two quality variables of interest in this research are the oleic acid mole fraction,  $x_8$ , and linoleic acid mole fraction,  $x_9$ , in the bottom flow rate. The objective of the

proposed FDD tools is to maintain the value of these two variables at their steady-state value through detection and diagnosis of faults present in the process.

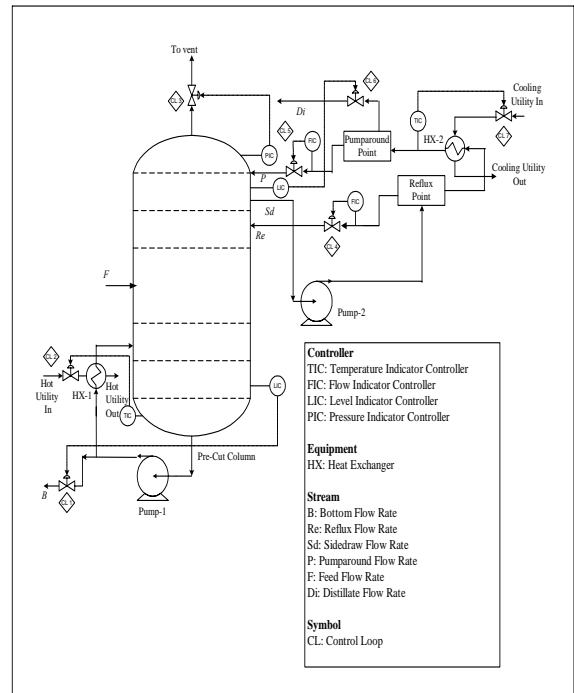


FIGURE 1 Distillation column model

#### Derivation of correlation coefficients

After the NOC data are obtained, the correlation coefficients between the selected key process variables and the quality variables of interest are determined using PCA and PCorrA. Method for obtaining correlation coefficients between the variables,  $C_{ik}$ , using PCA was based on previous PCA work (Lam and Kamarul, 2002). Correlation coefficients using PCA are calculated as in Equation 1.

$$C_{ik} = \sum_{j=1}^n v_{ij} v_{kj} \lambda_j$$

(Eq.1)

Where:

$v_{ij}, v_{kj}$  = eigenvectors obtained from process data using PCA

$\lambda_j$  = eigenvalue obtained from process data using PCA

TABLE 1 Fault Descriptions

Sensor Fault	Valve Fault	Controller Fault
<p>-For open loop variables, only the value of the variable changes abnormally. For closed loop variables, only the value of the disturbance (D) OR the manipulated variable (MV) OR the control variable (CV) changes abnormally.</p>	<p>-For open loop variables, only the value of the variable changes abnormally. For closed loop variables, both manipulated variable (MV) AND control variable (CV) changes abnormally together.</p>	<p>-For closed loop variables, the value of manipulated variable (MV) AND control variable (CV) changes abnormally together.</p>

PCorrA determines the correlation between two variables while allowing the effect of other correlated variables on these two variables. For calculating correlation coefficient,  $C_{ik}$ , for variable 1 and 2 using PCorrA after allowing the effect of  $j-2$  variables is as shown in Equation 2 (Cliff and Ord, 1973).

$$C_{ik_{12}} = \frac{r_{12(4,\dots,j-2)} - r_{13(4,\dots,j-2)}r_{23(4,\dots,j-2)}}{(1 - r_{13(4,\dots,j-2)}^2)^{1/2} (1 - r_{23(4,\dots,j-2)}^2)^{1/2}} \quad (\text{Eq.2})$$

Where:

- $r_{12}$  = correlation between variable 1 and 2
- $r_{12.3}$  = partial correlation between variable 1 and 2 after the effect of variable 3
- $r_{12.(3,4,\dots,j-1)}$  = partial correlation between variable 1 and 2 after the effect of  $j-2$  variables

**Development of FDD Tools**

$C_{ik}$  relates a process variable,  $x_i$ , with a quality variable,  $y_i$ , in the following way:

$$x_i = \frac{y_i}{C_{ik}} \quad (\text{Eq.3})$$

For conventional Shewhart Control Chart, the Upper Control Limit (UCL), Center Line (CL) and Lower Control Limit (LCL) for mean-centered and variance-scaled variables are +3, 0 and -3 respectively (McNeese and Klein, 1991). Using the information from Equation 3, the UCL, CL and LCL for quality variables and process variables will be +3, 0 and -3 and  $+3/C_{ik}$ , 0 and  $-3/C_{ik}$  respectively. After the NOC control charts are established, they are used for fault detection of the OC data.

The UCL, CL and LCL for conventional Range Control Chart for mean-centered and variance-scaled variables are mean of the range values,  $R_{mean}$  multiplied by a constant,  $d_2$ ,  $R_{mean}$  and 0 respectively (McNeese and

Klein, 1991). The constant,  $d_2$ , is determined by the number of subgroup used in calculating the range values. In the present work,  $d_2$  is 3.267 for a subgroup,  $n = 2$  (McNeese and Klein, 1991).  $R_{mean}$  is determined as shown in Equation 4.

$$R_{mean} = \frac{\sum_{i=1}^n R_i}{n} \quad (\text{Eq.4})$$

Where:

$R_i$  =  $i$ -th Range value  
 $R_{mean}$  = mean of the range values  
 $n$  = number of range values

For the present work, the UCL, CL and LCL for the Range Control Chart of quality variables will be of the conventional Range Control Chart. For the selected process variables, the UCL, CL and LCL will be  $(R_{mean} * d_2)/C_{ik}$ ,  $(R_{mean})/C_{ik}$  and 0 respectively.

The major assumption in the proposed method is that all key process variables are measured. The process variables that are major contributors to the variation of the process are included into the correlation analysis. In this way, the behavior of the process will be well represented by the correlation determined from the selected key process variables and the developed fault detection and diagnosis method will suit the dynamic behavior of the process. From Figure 1, the study column is installed with several control loops to ensure the stable operation of the column. Any common cause changes in the column either through load problem (disturbance changes) or servo problem (set point changes) will be taken care of through these controllers. The causal cause changes of interest in this work are those involving abnormal changes in the values of the variables of the process not through the two mentioned problems rather through faults in sensors, valves or even controllers. For NOC data, only common cause variation is present in the process. While for OC data, the observed causal cause variation is caused by faulty operation of the process sensors, valves and controllers.

When a process variable changed from its normal steady-state value, the variable of that control chart will be checked whether it is a closed loop variable or open loop variable. A fault signal is observed only when either the Range Control Chart or Shewhart Control

Chart of one or more quality variable show value that exceeds its control limit AND one or more process variable observed a value out of its control limit either in its Shewhart Control Chart or Range Control Chart. For open loop variable, the fault will be of sensor fault or valve fault as pre-designed while fault for closed loop variable can be of valve fault, sensor fault or controller fault. The cause variable(s) of each detected fault is diagnosed by checking the control charts of the process variables. Process variables that show value exceeding its control limit (either in Shewhart Control Chart or Range Control Chart) are diagnosed as the cause of the observed fault. To determine which type of fault is detected, the method used is as the previous paragraph.

### Results and Discussions

Figure 2 shows an example of the fault detection and diagnosis using the proposed method based on PCA. For the PCorrA method, a similar plot of graphs will be observed as well. Due to space limitation, only the Shewhart Control Chart for the 6 selected key process variables (feed flow rate ( $Lf$ ), feed temperature ( $Tf$ ), reflux flow rate ( $Re$ ), pumparound flow rate ( $P$ ), reboiler duty ( $Qr$ ) and bottom column temperature ( $Tbot$ )) and quality variable 1 (oleic acid mole fraction in the bottom flow rate ( $x_8$ )) were shown in Figure 2. Similar results will also be observed through the Range Control Chart of these variables. The performance of PCA and PCorrA in detecting the faults and diagnosis the cause of each detected fault is shown in Figure 3 and Figure 4.

Both methods based on PCA and PCorrA were able to diagnose the cause of each fault detected. Out of the 17 faults in the fault data, 13 faults (both single fault and multiple faults) were successfully detected by the PCA method (Using data reduction with 95% of the variation of the original data retained). The 4 faults that were not detected by the PCA method were insignificant faults (moderate changes in the values of the process variables). The method based on PCorrA performed better than the PCA method by successfully detecting all the 17 pre-designed faults (both significant faults and insignificant faults). The PCorrA method performed better because the correlation coefficients developed by this method are closer to the actual value of the correlation coefficients representing the

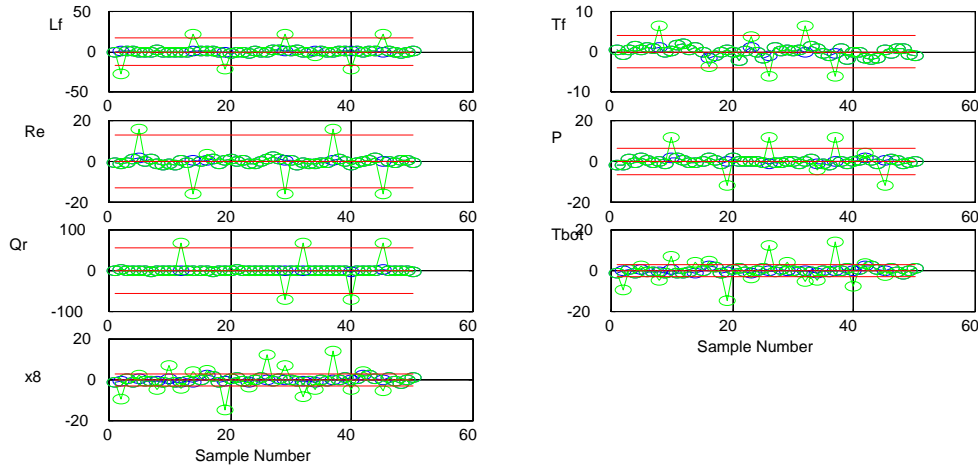


FIGURE 2 Example of Fault Detection and Diagnosis based on PCA

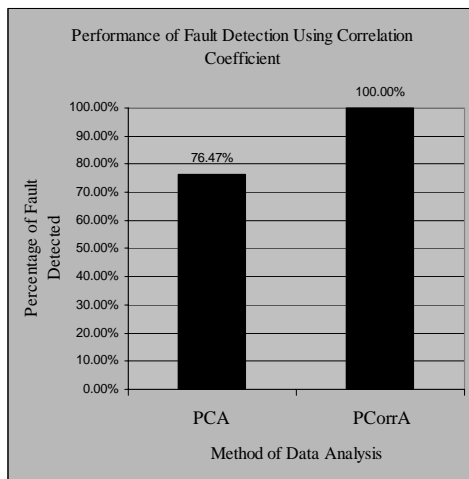


FIGURE 3 Performance of fault detection using correlation coefficient

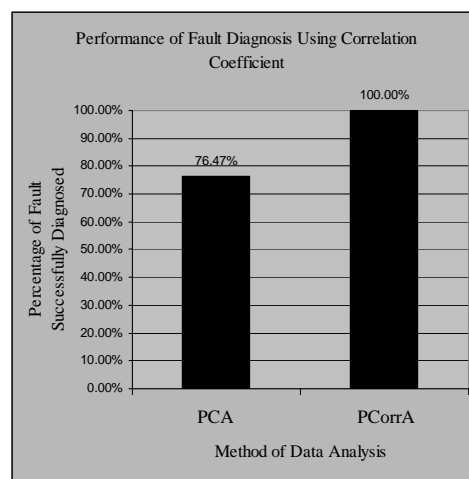


FIGURE 4 Performance of fault diagnosis using correlation coefficient

correlation between the selected process variables with the quality variables of interest. This is because the PCorrA method sets other selected process variables at constant values when calculating the correlation between a selected process variable with a quality variable. The PCA method calculates the cross-correlation between variables (interaction between variables) when determining the correlation coefficients between the process variables and quality variables. However, the PCorrA method was superior in determining the correlation between variables judging from the observed fault detection and diagnosis results of the study column.

One major advantage of the correlation coefficients method is the simplicity in determining the fault cause(s) of a detected

fault. The control charts of the selected process variables will trigger alarm if any of them exhibit value out of their control limits and the charts that triggers an alarm will be determined as the root causes of the detected fault. Furthermore, the availability of control limits in these control charts will shed away any ambiguities of whether a change in value of the selected process variables are due to common cause (NOC) or causal cause (OC). For online process monitoring, the data that are used for calculating the correlation coefficients can be updated with dynamic data to take account into the changes of the process due to change in raw material, fouling in heat exchangers and other changes in the process parameters. This area can be further researched and are a research problem for future work. The application of the developed

FDD tools on a multiple unit operation case study is also a research work for the future.

### Conclusion

An approach for fault detection and diagnosis using correlation coefficients based on PCorrA and PCA was presented. The performance of the approach was studied on an industrial distillation column. The results show that the fault detection and diagnosis method using cross correlation coefficient was able to detect the faults and diagnose the fault cause of each detected fault (both single fault cause and multiple fault causes). Although both methods based on PCA and PCorrA were successful in diagnosing the cause of each fault detected, PCorrA managed to detect all the pre-designed faults (both significant faults and insignificant faults) while PCA only managed to detect the significant fault. This is due to the fact that PCorrA determines the correlation between two variables after taken account into the effect of other variables that are correlated with the two variables of interest. Therefore, the correlation coefficients developed using the PCorrA method was better in representing the correlation between the selected process variables and the quality variables of interest.

### Acknowledgements

The present work was funded by the National Science Fellowship (NSF) scholarship. Special thanks to Mr. Wong Teck Siang for providing the basic model for the case study of this work. Mr. Lam Hon Loong's help in providing the information on the derivation of the correlation coefficient via eigenvector-eigenvalue approach for the PCA method is gratefully acknowledged.

### References

- Cliff, A.D. and Ord, J.K. (1973). Spatial Autocorrelations. London: Pion.
- Ding, A.C. and Nancy, P.L. (2000). Partial Correlation of Fuzzy Sets. *Fuzzy Sets and Systems* Vol. 110: 209-215.
- Lam, H.L. and Kamarul, A.I. (2002). Improved Multivariate Statistical Process Control for Chemical Process Fault Detection and Diagnosis via Cross Variable Correlation Coefficient Approach. *Regional*

*Symposium on Chemical Engineering, RSCE/SOMChem 2002, Hotel Hilton, Petaling Jaya, Selangor, Malaysia, 28-30 October 2002.* Vol. 2: 1637-1644.

MacGregor, J.F. and Kourti, T. (1995). Statistical Process Control of Multivariate Processes. *Control Engineering Practice* Vol. 3 (3): 403-414.

Mak, W.Y. and Kamarul, A.I. (2003). Fault Detection for Distillation Column Using Multivariate Statistical Process Control (MSPC). *Proceedings of The 17<sup>th</sup> Symposium of Malaysian Chemical Engineers (SOMChE 2003), 29-30 December 2003, Hotel Copthorne, Penang, Malaysia:* 92-96.

McNeese, W.H. and Klein, R.A. (1991). Statistical Methods for the Process Industries. USA: ASQC Quality Press.

Treasure, R.J., Kruger, U. and Cooper, J.E. (2004). Dynamic Multivariate Statistical Process Control Using Subspace Identification. *Journal of Process Control* Vol. 14: 279-292.

Wachs, A. and Lewin, D.R. (1999). Improved PCA Methods for Process Disturbance and Failure Identification. *AIChE Journal* Vol. 45: 1688-1700.

Yoon, S.K. and MacGregor, J.F. (2000). Statistical and Causal Model-Based Approach to Fault Detection and Isolation. *AIChE Journal* Vol. 46 (9): 1813-1824.

Yoon, S.K. and MacGregor, J.F. (2001). Fault Diagnosis with Multivariate Statistical Models Part I: Using Steady-state Fault Signatures. *Journal of Process Control* Vol. 11 (4): 387-400.

Wong, T.S. (2003). Advanced Process Control of A Fatty Acid Distillation Column Using Artificial Neural Networks. Unpublished Masters Thesis (Universiti Teknologi Malaysia), Malaysia.