FAULT DETECTION AND DIAGNOSIS, FDD VIA IMPROVED UNIVARIATE STATISTICAL PROCESS CONTROL CHARTS, USPC

Noorlisa Harun*, Kamarul Asri Ibrahim

Process Control and Safety Laboratory, Department of Chemical Engineering,
Universiti Teknologi Malaysia, 81310 Skudai, Johor.
*Corresponding author. Email: noorlisa@kuktem.edu.my

ABSTRACT

A new approach for detecting and diagnosing fault via correlation technique is introduced in this study. The correlation coefficient is determined using multivariate analysis technique, Partial Correlation Analysis (PCorrA). Individual charting technique such as Shewhart, Exponential Weight Moving Average (EWMA), and Moving Average and Moving Range (MAMR) charts are used to facilitate the Fault Detection and Diagnosis (FDD). A precut multi component distillation is used as the case study in this work. Based on the result from this study Shewhart control chart gives the best performance with the highest FDD efficiency.

Keywords: Fault Detection and Diagnosis (FDD), Shewhart Xhart, Exponential Weight Moving Average (EWMA) Chart, Moving Average and Moving Range (MAMR) Chart, Partial Correlation Analysis (PcorrA).

1 INTRODUCTION

Malfunction of plant equipment, instrumentation and degradation in process operation increase the operating costs of any process industries. More serious are a gross accident such as explosion. Even major catastrophes and disasters from chemical plant failures may be infrequent, minor accidents are very common, occurring on a day to day basis, resulting in many occupational injuries, illnesses, and costing the society billions of dollars. Venkat, et al., (2003) mentioned that the petrochemical industry annually losses approximately $20 billion due to poor management in abnormal detection events. Chen, et al., (2004) also highlighted that the US-based petrochemical industry could save up to $10 billion annually if abnormal process behavior could be detected, diagnosed and appropriately dealt with. Therefore, effective monitoring strategy for early fault detection and diagnosis is very important not only from a safety and cost viewpoints, but also for the maintenance of yield and the product quality in a process.

Statistical Process Control, SPC is an alternative approach in chemical processes to detect and diagnose fault. The major benefits of this approach are that there is no need for a fundamental or causal model of the system. In chemical processes, data based approaches rather than model-based approaches have been widely used for process monitoring, because it is often difficult to develop detailed physical models (Manabu et al., 2000). SPC only requires a good database of normal historical data, and the models are quickly and easily built from this.

SPC chart is the most technically sophisticated tool to monitor the performance of any given process. The function of this control chart is to compare the current state of the process against Normal Operating Condition, NOC. The NOC condition exists when the process or product variables remain close to their desired values or in statistical control. In contrast, the Out of Control, OC occurs when fault appears in the process. In general, fault is deviations from the normal operating behavior in the plant that are not due disturbance and set point changes in the process. Fault detection is to determine the occurrence of an abnormal event in a process, and that of fault diagnosis is to identify its reason or sources.

Traditional SPC methods assume that process data is statistically independent and stationary (Nong et al., 2000) and ignoring the cross correlation between the variables. This can lead to faulty interpretation during process monitoring. To overcome this limitation, a multivariate analysis approach is applied in the USPC realm procedure to detect and diagnose the faulty condition. Multivariate analysis method that is Partial Correlation Analysis, PCorrA is used to develop the control limits of USPC charts. Ibrahim (1997) has introduced PCorrA method to be applied in chemical process data to develop Multivariate Statistical Process Control (MSPC) scheme known as Active SPC.

2 DATA GENERATION

Figure 1 shows the schematic diagram of dynamic simulated distillation column developed by Mak and Kamarul (2003) that is used in this case study. The monitoring purpose of this column is to maintain the composition of oleic acid and linoleic acid at the range of 0.134 to 0.135 mole fraction and 0.024 to 0.025 mole fraction respectively. This column is used to generate two sets of data i.e NOC data and OC data.

SPC variable is categorized into two i.e quality variables and process variables. The quality variables acted as an indicator variables to show that the process in the state of statistical control or in the state of out of control. If any of the points of these variables are fall out of control limits, it shows that the fault situation is taken place. On the other hand, the process variables used to find the causes of the fault situation. Table 1 shows the list of quality variables and process variables used in this study.

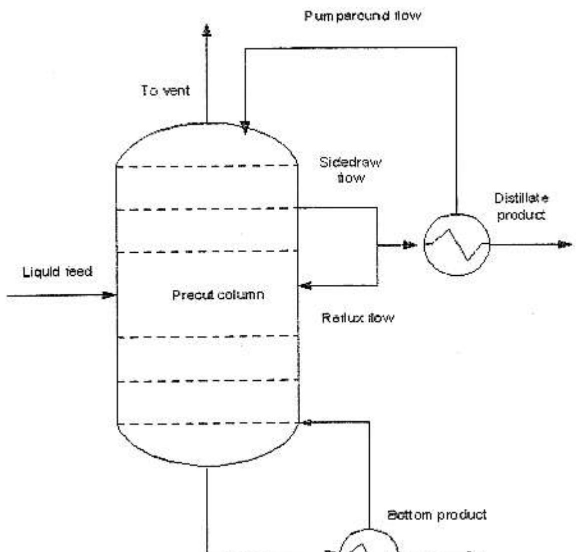

FIGURE 1. Schematic diagram of distillation column

TABLE 1. SPC variables

Quality variable	Process variable
$xC8$	F
$xC9$	Tf
	P
	Re
	Q_r

where F = feed flowrate, T_f = feed temperature, P = pumparound flowrate, Re = reflux flowrate, Q_r = reboiler duty, $xC8$ = oleic acid composition, and $xC9$ = linoleic acid composition.

NOC data that consist of quality variables and process variables were generated and arranged in the matrix form, **X** when the process is in statistical control or quality variables remain close to their desired values. The matrix data, **X** with m observations on p variables can be written as,

$$\mathbf{X} = \begin{bmatrix} x_{11} & x_{12} & \cdots & x_{1p} \\ x_{21} & x_{22} & \cdots & x_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ x_{m1} & x_{m2} & \cdots & x_{mp} \end{bmatrix} \tag{1}$$

NOC data are very important in SPC methodology since it is used to predict the future behavior of the process. Some noises were imbedded into the process variables using Matlab simulator to create random process data with normally distributed. After NOC step done, faulty condition was introduced in the process by inserting deviations in the process variables and OC data was collected during this condition.

Both NOC and OC data are standardized before further analysis since the variables have different units and wide range of data measurements. Each variable is adjusted to zero mean by subtracting off the original mean of each column and adjusted to unit variance by dividing each column by its standard deviation. After the standardization, each variable have equal weights with zero mean and one standard deviation (N (0, 1)). The linear relationship between quality variables and process variables is developed using multivariate analysis techniques, PCorrA during normal process operation. This relationship is interpreted in terms of correlation coefficient, C_{ik} which is used to diagnose the cause of the fault during the OC situation.

3 PARTIAL CORRELATION ANALYSIS, PCORRA

Partial correlation coefficient is defined as a correlation of quality variable, x_k and process variable, x_i when the effects of other process variable(s) have been removed from x_k and x_i. If the two variables of interest are $x_{k,1}$ and $x_{i,1}$ and the controlled variables are x_{i2}, x_{i3} ... x_{in}, then the corresponding partial correlation coefficient is

$$c_{xk1,xi1|xi2,xi3\ldots xin} = \frac{c_{xk1,xi1|xi3,xi4\ldots,xin} - c_{xk1,xi2|xi3,xi4\ldots,xin}\, c_{xi1,xi2|xi3,xi4\ldots,xin}}{\sqrt{1-c^2_{xk1,xi2|xi3,xi4\ldots,xin}}\sqrt{1-c^2_{xi1,xi2|xi3,xi4\ldots,xin}}} \tag{2}$$

As shown mathematically above, PCorrA is done by separating the group of process variables into subgroup in which one or more variables are held constant before determining the correlation among the other variables.

4 INDIVIDUAL SPC CHARTS

SPC chart is used to monitor the performance of any given process. There are two types of causes, which contribute to the existing of faults in the process. Chance, or common causes are small random changes in the process that cannot be avoided. Variation of this type is only removable by making a change in the existing process. Assignable causes, on the other hand, are large variation in the process that can be identified as having specified cause. Assignable causes are causes that are not part of the process on a regular basis. This type of variation arises because of specific circumstances. Sources of variation can be found in the process itself, the material used, the operator's actions, or the environment.

Control charts approach is based on the assumption that a process subject to common cause variation will remain in a state of statistical control under which process remain close to target which is known as NOC data for this study. By monitoring the performance of a process over time, OC events known as assignable cause can be detected as soon as they occur. If the causes for such events can be diagnosed and the problem can be corrected, the process is driven back to its normal operation. Individual Shewhart, Exponential moving average (EWMA), and Moving Average (MA) and Moving Range (MR) are USPC charts. They are used for individual data. Correlation coefficient from the multivariate analysis technique is used to relate the quality variables with the process variables. This correlation coefficient is used to translate the control limits of USPC charts from quality variables into process variables, which is used to perform fault diagnosis for the process operation. The OC situations will be considered whenever a point fall outside the control limits with 99.73% confidence limits for all these charts.

5 IMPLEMENTATION OF PCORRA IN USPC CHART

PCorrA method is used to determine the correlation coefficient between quality variable and process variables. Let x_k is the quality variable and x_i is the process variable. The relationship between standardized quality variable, x_k^s and standardized process variable, x_i^s can be written as,

$$x_k^s = C_{ik} x_i^s \tag{3}$$

where $x_k^s = (x_k - \bar{x}_k)/s_k$, $x_i^s = (x_i - \bar{x}_i)/s_i$. s is the standard deviation, while $\bar{x}_k$ and $\bar{x}_i$ is quality variable mean and process variable mean respectively. The control limit for quality variable in general is

$$LCL < x_k^s < UCL \tag{4}$$

where UCL and LCL is upper control limit and lower control limit respectively. Substitute equation 3 into equation 4 and rearrange the equation 4. The control limits for corresponding process variable is,

$$LCL/C_{ik} < x_i^s < UCL/C_{ik} \tag{5}$$

Equation 5 is used to calculate the control limits for process variables. These limits are calculated based on the NOC data. Table 2 shows both of the limits for all control charts as FDD tools in this study.

TABLE 2. Control limits for quality variable and process variable

Control chart	Quality variable control limit	Process variable control limit
Shewhart Individual	$UCL = 3s, LCL = -3s$	$UCL = 3s/C_{ik}$, $LCL = -3s/C_{ik}$
Shewhart Range	$UCL=D'_{.001}\bar{R}$, $LCL=D'_{.999}\bar{R}$	$UCL=D'_{.001}\bar{R}/C_{ik}$, $LCL=D'_{.999}\bar{R}/C_{ik}$
EWMA	$UCL=+Ls\sqrt{\frac{\lambda}{2-\lambda}[1-(1-\lambda)^{2i}]}$	$UCL=+(Ls\sqrt{\frac{\lambda}{2-\lambda}[1-(1-\lambda)^{2i}]})/C_{ik}$
	$UCL=-Ls\sqrt{\frac{\lambda}{2-\lambda}[1-(1-\lambda)^{2i}]}$	$UCL=-(Ls\sqrt{\frac{\lambda}{2-\lambda}[1-(1-\lambda)^{2i}]})/C_{ik}$
MA	$UCL=+A_2\bar{R}$, $LCL=-A_2\bar{R}$	$UCL=+A_2\bar{R}/C_{ik}$, $LCL=-A_2\bar{R}/C_{ik}$
MR	$UCL=D'_{.001}\bar{R}$, $LCL=D'_{.999}\bar{R}$	$UCL=D'_{.001}\bar{R}/C_{ik}$, $LCL=D'_{.999}\bar{R}/C_{ik}$

where $\bar{R}$ = average of range; λ = weighing factor; A_2, $D'_{.001}$, $D'_{.999}$ = constants, L = width of the control limit.

Table 3 shows the equation to determine statistical data for each control charts.

TABLE 3. Statistical data for control charts

Control chart	Quality variable	Process variable
Shewhart Individual	$x = x_i$	$y = y_i$
Shewhart Range	$R_i = max\ [x_{i-1+1}] - min\ [x_{i-1+1}]$	$R_i = max\ [y_{i-1+1}] - min\ [y_{i-1+1}]$
EWMA	$z_i = \lambda x_i + (1-\lambda) z_{i-1}$	$z_i = \lambda y_i + (1-\lambda) z_{i-1}$
MA	$MA_i = (x_i + x_{i-1} + \ldots + x_{i-w+1})/w$	$MA_i = (y_i + y_{i-1} + \ldots + y_{i-w+1})/w$
MR	$MR_i = max\ [x_{i-w+1}] - min\ [x_{i-w+1}]$	$MR_i = max\ [y_{i-w+1}] - min\ [y_{i-w+1}]$

6 FDD EFFICIENCY USING DIFFERENT CONTROL CHARTS

Both quality variable and process variable were monitored continuously using three types of control charts. The efficiency of the FDD method using Shewhart, EWMA and MAMR chart is evaluated based on two aspects i.e the successful of SPC chart to detect the fault and the successful of SPC chart to identify the correct process variable as fault cause for each fault situation. The efficiency of fault detection, $\eta_{Fdetect}$ and the efficiency of the fault diagnosis, $\eta_{FDiagnose}$ is determined using the following equation,

$$\eta_{Fdetect} = [\text{Number of faults detected / Total of faults generated in the process}] \times 100$$
$$\eta_{FDiagnose} = [\text{Number of faults diagnosed / Total of diagnosed fault}] \times 100$$

The overall performance is,

$$\eta_{FDD} = \eta_{Fdetect} \times \eta_{Fdiagnose} \times 100\%$$

80 fault locations consist of 50 single fault and 30 multiple faults were introduced into the process. Figure 2 and figure 3 show the efficiency of FDD on quality variables, oleic acid, C8 and linoleic acid,C9 respectively using different SPC charts. Shewhart chart give 100% performance in FDD, which is better than EWMA and MAMR. Shewhart chart is plotted using 100% current data but MAMR statistic is calculated using window size of four, which consist of 25% current data and 75% previous data. This caused detection delay using MAMR. EWMA statistic exponential weighted average of all prior data, including the most recent data. The weighted average depends on weighting factor, λ. Small λ will give less weight to current data and more weight to previous data and vice versa. In this study, $\lambda = 0.4$ is used. This give the EWMA statistic consist of 40% current data 60% previous data. This shows that FDD efficiency decreases as the percentage of previous data involved in calculates the statistics value for each charts increases. False alarm (action is taken due to signal but in fact the process does not change at all) rate using MAMR is about 10% for C8 and 7% for C9. This is higher than EWMA, which has 3% false alarm rate for C8.

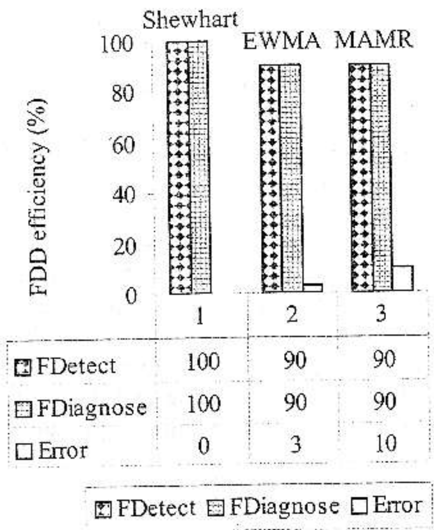

FIGURE 2. Oleic acid FDD effieciency

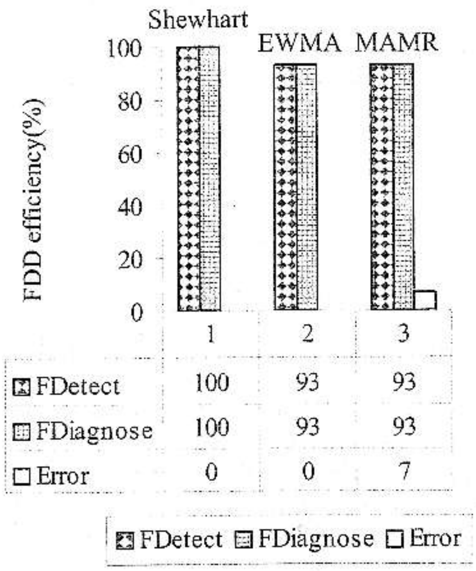

FIGURE 3. Linoleic acid FDD efficiency

Figure 4 shows the fault detection efficiency using Shewhart, EWMA and MAMR in three different regions using single fault data. The OC data that is greater than ±3s is divided into three regions. Region 1, region 2, and region 3 refer to mean ±4s, mean ±5s and over mean ± 5s respectively. Both Shewhart chart and EWMA chart can detect OC data in region 1 but Shewhart give better performance than EWMA. MAMR chart can detect deviation, which is greater than mean ±4s. The result obtained shows that Shewhart chart is more efficient to detect small shift in the process.

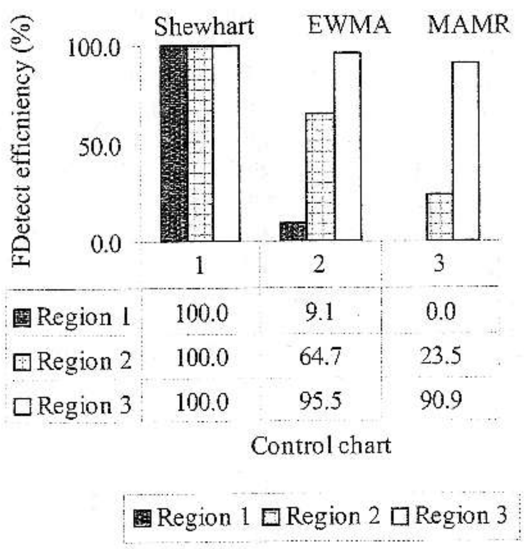

FIGURE 4. Fault detection efficiency in different region

7 CONCLUSIONS

Improved USPC chart for FDD using PCorrA technique, which is used to develop correlation between quality variables and process variables have been presented. Process monitoring for fault diagnosis can be done using process variables with the implementation of correlation coefficient. Performance of Shewhart chart is the best compared to EWMA and MAMR in detecting and diagnosing faults. FDD result using EWMA is better than MAMR because high false alarm rate using MAMR shows the risk of taking wrong action on the process.

REFERENCES

Chen, Q., Kruger, U., Meronk, M. and Leung, A.Y.T. (2004). "Synthesis of T^2 and Q statistic for process monitoring." *Control Engineering Practice*, 12, 745-755.

Ibrahim, K.I. (1997). "Application of Partial Correlation Analysis in Active Statistical Process Control." Proceedings of Regional Symposium of Chemical Engineering, 434-438, Johor.

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

Manabu K., Koji N., Shinji H., Iori H., Hiromu, O., Ramon S., and Bhavik B. (2000). "Comparison of statistical process monitoring methods: application to the Eastman challenge problem." *Computers and Chemical Engineering*, 24, 175-181.

Nong Y., Qiang C., Syed Masum E., and Kyutae N. (2000). "Chi-square Statistical Profiling for Anomaly Detection." Proceedings of the 2000 IEEE Workshop on Information Assurance and Security United States Military Academy, West Point, NY, 187-193.

Venkat Venkatasubramanian, Raghunathan Rengaswamy, Kewen Yin, Surya N. Kavuri (2003). "A review of process fault detection and diagnosis Part I: Quantitative model-based methods, *Computers and Chemical Engineering*, 27, 293-311.