Journal of Applied Sciences

ISSN 1812-5654

Journal of Applied Sciences 11 (13): 2400-2405, 2011
ISSN 1812-5654 / DOI: 10.3923/jas.2011.2400.2405

Fault Detection and Diagnosis for Gas Density Monitoring using Multivariate Statistical Process Control

N.S. Che Din, N.A.F. Abdul Samad and S. Y. Chin
Faculty of Chemical and Natural Resources Engineering, Universiti Malaysia Pahang, Kuantan, Pahang, Malaysia

Abstract: Malfunction of plant equipment, instrumentation and degradation in process operation increase the operating cost of any chemical process industries. Thus, modern chemical industries need to operate as fault free as possible because faults that present in a process increase the operating cost due to the increase in waste generation and products with undesired specifications. Effective monitoring strategy for early fault detection and diagnosis is very important not only from a safety and cost viewpoint, but also for the maintenance of yield and the product quality in a process as well. Therefore, an efficient fault detection and diagnosis algorithm needs to be developed to detect faults that are present in a process and pinpoint the cause of these detected faults. Multivariate analysis technique i.e., Principal Component Analysis (PCA) and Partial Correlation analysis (PCorrA) are used to determine the correlation coefficients between the process variables and quality variables while control chart with the calculated correlation coefficients are used to facilitate the Fault Detection and Diagnosis (FDD) algorithm. A procedure for FDD has been described in this study and the proposed method is demonstrated on an Air Flow Pressure Temperature (AFPT) control system pilot plant. Results show that method based on PCA and PCorrA was able to detect the pre-designed faults successfully and identify variables which cause the faults.

Key words: Correlation coefficient, partial correlation analysis, shewhart control chart

INTRODUCTION

Modern chemical industries are facing a lot of challenges either from internal or external. This industry have to keep sustainable production and within the quality specification for the production. The whole production process has to operate at the minimum production waste, minimum consumption of utilities and minimum cost of re-work and re-processing. Therefore, for the successful operation of any process, it is important to perform an effective process monitoring to detect and diagnose any process upsets, equipment malfunctions, or other special events as soon as possible after their occurrence and then remove the factors causing those events.

Those events or also known as fault can be caused by temporary or permanent physical changes in the system. Obviously, process monitoring is essential to ensure that the plants operate safely and economically while meeting environmental standards (Seborg *et al*., 2004).

However, in most processes, there are more than one measurement processes to monitor (Yang and Trewn, 2004) and it is increasingly difficult to determine the root cause of defects if multiple process variables exhibit faults or process deviations at the same moment in time. Moreover, most processes are highly correlated, particularly for assembly operations and chemical processes (Chen *et al*., 2001).

Due to the multivariate, dynamic and nonlinear nature of chemical processes, the efforts toward process monitoring and fault detection and diagnosis continue. In this study, the Multivariate Statistical Process Control (MSPC) monitoring techniques was applied to monitor the process variability of the pilot-scale Air Flow Temperature Pressure (AFPT) unit used primarily for gas density reading. MSPC is a method that is able to extract the desired information from the very large volumes of data by carrying out data reduction without losing the original information. Many industrial processes involved a set of process variables as well as quality variables, which are highly correlated. If one of the variable change, it will effect the other correlated variables. Thus, ignoring the cross-correlation between the variable can lead to misinterpretation of the process behavior.

Principal Component Analysis (PCA) is the MSPC technique that has been utilized in this study to determine the correlation coefficient between the process variables

Corresponding Author: N.S. Che Din, Faculty of Chemical and Natural Resources Engineering, Universiti Malaysia Pahang, Kuantan, Pahang, Malaysia Tel: +6019-9880570

and quality variables. PCA is capable of utilizing massive amounts of data and compress the information in this data down into low dimensional latent variable spaces in which monitoring of the process and interpreting the results are much easier (Kourti *et al*., 1996).

The developed correlation coefficient will be used together with conventional Shewhart Control Chart and Exponentially Weighted Moving Average (EWMA) Control Chart as FDD tools. The control charts were used for monitoring the overall process, showing large amounts of variability dominating the process. These periods of increased process variability were further analyzed for detecting their source of the fault.

MATERIALS AND METHODS

Process modeling and data generation: Gas density control system has been utilized in various chemical industries such as oil and gas, petrochemical and power industries. Generally, density is one of the thermophysical properties in which it is very much dependant on the pressure and temperature of the given process under consideration and it need to be set precisely (Chen *et al*., 1998). Therefore, it is intended of this study to develop a monitoring system to control the given process based on the density setting value by which it relates all the available measurements from the process as comprehensively as possible through MSPC.

The first step in developing the FDD algorithm is the selection of quality variables of interest and the corresponding key process variables that are related with the selected quality variables of interest. The next section will explain how the selection of quality variables and key process variables are carried out. The employed procedures in formulating the MSPC FDD algorithm for this research are presented here.

Selection of variables: The most important part in obtaining an accurate correlation between the key process variables and the quality variables of interest is the data generation. This research was completely utilized the original AFPT control system, which contains three control loops i.e., flow, temperature and pressure control loop. The monitoring purpose of this system is to detect fault in determining the gas density measurement and to ensure that the density reading is within specification. Figure 1 shows the block diagram of an Air Flow Pressure Temperature (AFPT) control training system that was used in this study (Juwari *et al*., 2007).

The quality variable of interest in this research is the density of the air in the tank. After the selection of quality variable of interest, the key process variables that are related with the selected quality variable will be chosen from a list of available (measured) process variables. From the thermodynamic properties, density is very much dependant on pressure and temperature. Thus, the process variables of interest in this system are pressure and temperature. Once the quality variable of interest and key process variables were selected, the system is ready to generate the required data for the development of the proposed FDD algorithm. Those data were collected from the DCS attached to the system.

The quality variable acted as indicator variable to show that the process is in the state of statistical control or out of control. On the other hand, the process variables are used to find the causes of the fault detected.

Two sets of process operating data were generated from the system. The Nominal Operating Condition (NOC) data are a set of data in which, the selected key process

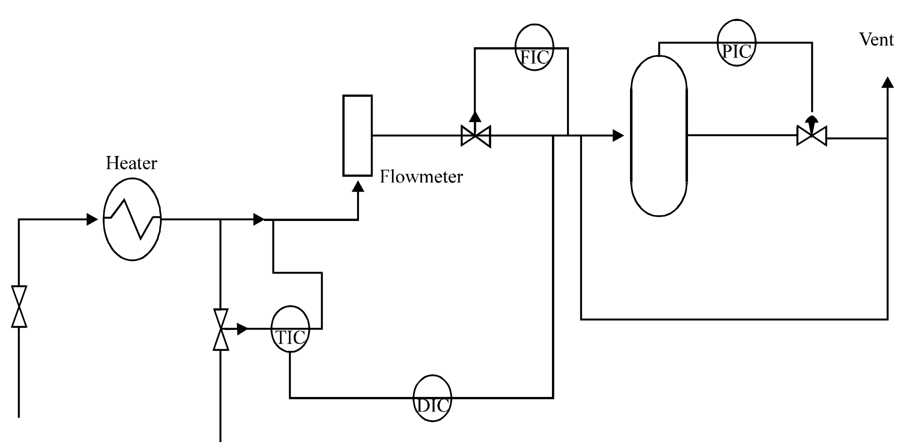

Fig. 1: Block diagram of the AFPT control training system

variables which have values within statistical control limits of the statistical control chart. This NOC data that consist of quality variables and key process variables was then arranged in the matrix form, X with m observations on p variables. The matrix X can be written as:

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

NOC data is very important in MSPC 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. While the second set of data is the Out of Control (OC) data set. After NOC step done, faulty condition was introduces in to process by inserting deviations in the process variables and the 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 weight with zero mean and one standard deviation (N (0, 1)).

Before any data can be sampled from the system, the process sampling time, T_{MSPC} is an important parameter that needs to be determined first. This process sampling time is for sampling data to be used in derivation of correlation coefficients between the selected key process variables and the quality variable of interest and it's different from the controller sampling time T_{APC} which is used to sample data for control purposes. T_{MSPC} was determining by autocorrelation plot. Autocorrelation is the correlation between successive data in a time series data. In a time series data, autocorrelation plays an important role in affecting the variation of the data. In order to find the true correlations between variables, the autocorrelation within the data series of a variable need to be omitted through the selection of a suitable process sampling time, T_{MSPC}.

The method in obtaining the NOC data set is shown in Fig. 2. The steps involving the analysis NOC data for correlation using PCA and PCorrA and the building of the control limits for the statistical control charts will be discussed in detail in next section.

Derivation of correlation coefficient: The FDD algorithm will be built based on the development of a correlation

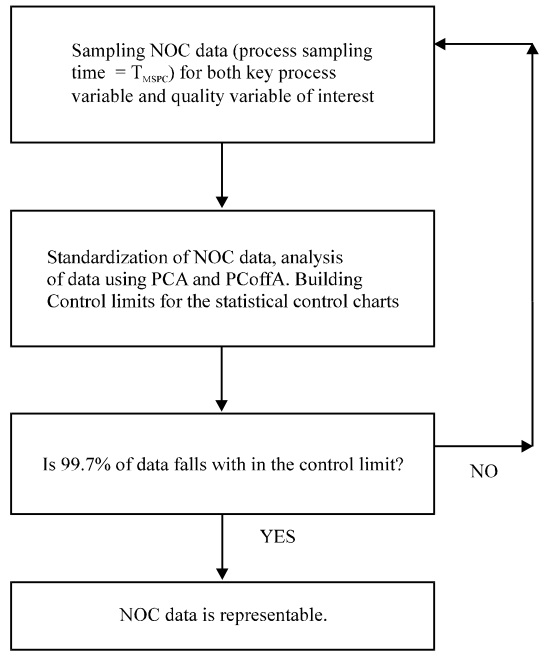

Fig. 2: Procedure in obtaining the NOC data set

coefficient, C_{ik} between the quality variables of interest and the selected key process variables. This correlation coefficient will provide linear relationship between the quality variables of interest and the selected key process variables. A selected key process variable is related to a quality variable of interest by the following equation:

$$x_i = \frac{y_i}{C_{ik}} \quad (2)$$

Where:

y_i = Quality variable
x_i = Process variable
C_{ik} = Correlation coefficient between y_i and x_i

There are two methods used in this paper to derive the correlation coefficient, C_{ik}, from the data matrix. The correlation coefficient will be determined through Principal Component Analysis (PCA) and Partial Correlation Analysis (PCorrA).

Correlation coefficient derivation using principal component analysis: PCA is a useful tool in MSPC for handling abundant of information from process measurements. The advantages of this method rely on data reduction and information extraction. The method of PCA is a linear transformation of the original variables into a new set of variables which is orthogonal to each other (Bezergianni and Kalogianni, 2008).

A set of original variables $x_1, x_2, \ldots, x_p$ is transform to a set of new variable $PC_1, PC_2, \ldots, PC_p$. The mathematical representations that describe the transformation of a data matrix, X (m,p) consist of m observations on p variables is shown as follow:

$$PC_{(p,m)} = V^{T}_{(p,p)} X_{(p,m)} \quad (3)$$

V is the eigenvector matrix, which consists of eigenvector v_1, v_2, ...,v_p. Singular Value Decomposition (SVD) technique is used to decompose data matrix, X (m, p) into a product of the eigenvectors of XX^T, the eigenvectors of X^TX and a function of their eigenvalue. The fundamental identity of SVD is shown by the following equation:

$$X_{(m,p)} = U_{(m,p)} L^{1/2}_{(p,p)} V^{T}_{(p,p)} \quad (4)$$

The diagonal elements of $L_{(p,p)}$, $\lambda_1, \lambda_2, \ldots, \lambda_p$ are called eigenvalues of X while the columns vector of $U_{(m,p)}$, $u_1, u_2, \ldots, u_p$ and the columns vector of $V_{(p,p)}$, $v_1, v_2, \ldots, v_p$ are called eigenvectors of X and both eigenvector are orthonormal. Matrices of U, V and $L^{1/2}$ have the following properties:

$$U^T U = UU^T = I \quad (5)$$

$$V^T V = VV^T = I \quad (6)$$

$$\left(L^{1/2}\right)\left(L^{1/2}\right)^T = \left(L^{1/2}\right)^T \left(L^{1/2}\right) = L \quad (7)$$

The correlation between the variable k, x_k and variable i, x_i if j variables are involved can be written as the following:

$$c_{ik} = \sum_{j=1}^{n} v_{ij} v_{kj} \lambda_j \quad (8)$$

Where:
n = Number of retained eigenvectors
v_k = Eigenvectors of process variable
v_i = Eigenvectors of quality variable

The number of retained eigenvector can be determined using Scree Plot (Ralston *et al.*, 2001).

Correlation coefficient derivation using partial correlation analysis: PCorrA correlation coefficient is defined as a correlation of quality variable, x_i and process variable, x_k when the effects of other process variables(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 $x_{k2}, x_{k3}, \ldots, x_{kn}$ then the corresponding partial correlation coefficient is:

$$C_{ik_{12}} = \frac{r_{12(4,\ldots,j-2)} - r_{13(4,\ldots,j-2)} r_{23(4,\ldots,j-2)}}{\left(1 - r^2_{13(4,\ldots,j-2)}\right)^{1/2} \left(1 - r^2_{23(4,\ldots,j-2)}\right)^{1/2}} \quad (9)$$

Where:
r_{12} = Correlation between variable 1 and 2
$r_{12.3}$ = Correlation between variable 1 and 2 after the effect of variable 3
$r_{12.(3,4,\ldots,j-2)}$ = Correlation between variable 1 and 3 after the effect of j-2 variables

Constructing statistical control charts: After the derivation of the correlation coefficient through PCA, the control limits for the statistical control charts will be built. The statistical control charts used in this research are Shewhart Control Chart and Exponential Weight Moving Average. The control limits for these charts will be based on the calculated correlation coefficient.

The statistical control chart was used to determine whether the process in a state of statistical control or in a state of out of statistical control using a set of historical data.

Generally, there are three characteristics that must have in every control chart. The first one is the center line that represents the mean value for the in statistical control process. And there were two other horizontal lines called the upper control limits (UCL) and the lower control limit (LCL). These control limits are chosen so that almost all of the data points will fall within these limits as long as the process remains in a state of statistical control.

In this study, control chart 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 its target known as NOC data. By monitoring the process over time, any OC events can be detected as soon as they occurred. If the causes for those events can be diagnosed and the problem can be corrected, the process is driven back to its normal operation.

The calculated correlation coefficient through PCA was used to relate the relationship between the quality variable of interest and key process variables. Let x_k is the quality variable and x_i is the process variable. The relationship between the standardized quality variable, x^s_k and standardized process variable, x^s_i can be written as:

$$x^s_k = C_{ik} x^s_i \quad (10)$$

$$x^s_k = \left(x_k - \bar{x}_k\right) / s_k \quad (11)$$

$$x^s_i = \left(x_i - \bar{x}_i\right) / s_i \quad (12)$$

Table 1: Control limit for the SPC chart

Control chart	Control limit: Quality variable	Process variable
Shewhart	UCL = 3s LCL = -3s	UCL = $3s/C_{ik}$ LCL = $-3s/C_{ik}$
EWMA	$UCL = L_s\sqrt{\frac{\lambda}{2-\lambda}\left[1-(1-\lambda)^{2i}\right]}$ $UCL = -L_s\sqrt{\frac{\lambda}{2-\lambda}\left[1-(1-\lambda)^{2i}\right]}$	$LCL = \left(L_s\sqrt{\frac{\lambda}{2-\lambda}\left[1-(1-\lambda)^{2i}\right]}\right)/C_{ik}$ $LCL = -\left(L_s\sqrt{\frac{\lambda}{2-\lambda}\left[1-(1-\lambda)^{2i}\right]}\right)/C_{ik}$

Table 2: SPC chart statistic for FDD development

Control chart	Chart statistic
Shewhart	$x=x_i$
EWMA	$z_i = \lambda x_i + (1-\lambda)\, z_{i-1}$

λ = Weighting factor

Equation 10 and 11 shows the standard deviation and $\bar{x}_k$ and $\bar{x}_i$ is mean of the quality variable and process variable respectively. The control limit for quality variable is general is:

$$LCL < x_k^s > UCL \tag{13}$$

Substitute Eq. 9 into Eq. 12 and rearrange the Eq. 10. Equation 12 also can be use to determine the control limit for corresponding process variable. These control limits were calculated based on the standardized NOC data, which represent the desired process operation. The control limit was determined so that only 0.27% of the entire samples fall outside the control limit.

The control limit for quality variable and process variable for each SPC chart (Harun, 2005) is accordingly represented in Table 1.

SPC chart using quality variables were used to detect the faulty condition in the process. Samples were taken over time and values of statistics are plotted. The statistical data is calculated using equation in Table 2.

The process is considered to be in statistical control state if the plotted statistic of the quality variables falls within the control limits. While control chart based on process variables data were used to diagnose the cause of the faults.

RESULTS AND DISCUSSION

Fault detection and diagnosis (FDD) efficiency using SPC chart:

$$\eta_{Fdetect} = \frac{\text{Number of fault detected}}{\text{Total of faults generated in the process}} \times 100 \tag{14}$$

$$\eta_{FDiagnose} = \frac{\text{Number of fault diagnosed}}{\text{Total of diagnosed fault}} \times 100 \tag{15}$$

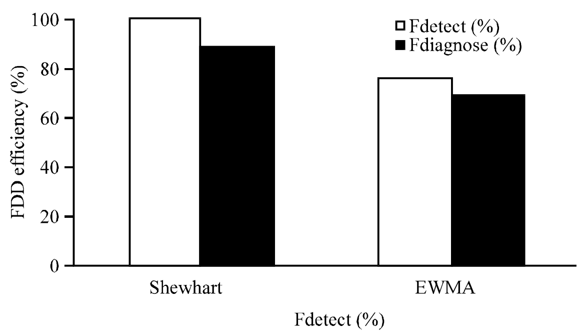

Fig. 3: Density FDD efficiency

Fifty fault locations consist of 30 single fault and 20 multiple faults were introduced into the process. Figure 3 shows the efficiency of FDD on the quality variable, which is density using different type of control chart. Shewhart chart give 100% performance in fault detection which is better than EWMA chart. Shewhart chart used 100% current data but 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, $\lambda = 0.4$ were used. This give the EWMA statistic consist of 40% current data 60% previous data. This caused EWMA is more efficient in detecting shift in the process since EWMA statistic gives more weight to current data.

Figure 4 show the fault detection using Shewhart and EWMA in three different regions based on single fault data. The OC data that is greater than ±3sec is divided into three regions. Region 1, 2 and 3 refer to mean±4 sec, Mean±5 sec and over Mean±5 sec, respectively. Refer to Fig. 4, both Shewhart chart and EWMA can detect OC data in region 1 but Shewhart (100% fault detection efficiency) give better performance than EWMA (9.1% fault detection efficiency).

Based on fault diagnosis result, the smallest deviation (0.24%) from the target in the process variables, while contribute to the out of control situation in the

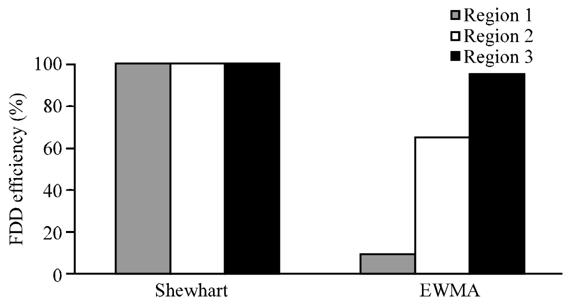

Fig. 4: Density FDD efficiency based on different region

quality variable, can be identified by Shewhart while EWMA can diagnose for 0.58% deviation in the process. The result obtained shows that Shewhart chart was able to detect small shift in the process while EWMA chart can be used to detect moderate shift in the process.

CONCLUSION

Early fault detection and fault diagnoses are essential in maintaining product quality and reduce the cost as well as for safety reason in most modern chemical industries. PCA and PCorrA techniques were used to develop the correlation coefficient between quality variable and process variable in order to formulate the FDD algorithm. One major advantage of the correlation coefficients method is process monitoring for fault diagnosis can be done using process variables.

Process fault diagnosis can be done in straightforward manner. The simplicity of the presentation and interpretation of the control chart based on multivariate analysis technique makes these charts useful to the plant engineers and operator to identify the out of statistical control in the process. Performance of Shewhart chart which used 100% current data is the best compared to EWMA in detecting and diagnosing the faults. EWMA chart incorporate previous data in calculating the control limits.

ACKNOWLEDGMENTS

Special thanks to Fundamental Research Grant Scheme (FRGS) RDU070101 for financial support in this present study.

REFERENCES

Bezergianni, S. and A. Kalogianni, 2008. Application of principal component analysis for monitoring and disturbance detection of a hydrotreating process. Ind. Eng. Chem. Res., 47: 6972-6982.

Chen, G., T.J. McAvoy and J. Piovoso, 1998. A multivariate statistical controller for on-line quality improvement. J. Process Control, 8: 139-149.

Chen, K.H., D.S. Boning and R.E. Welch, 2001. Multivariate statistical process control and signature analysis using eigenfactor detection methods. Proceedings of the 23rd Symposium on the Interface of Computer Science and Statistics, June 2001, Costa Mesa, CA, pp: 1-21.

Harun, N.B., 2005. Fault detection and diagnosis via improved multivariate statistical process control. MSc. Thesis, Department of Chemical Engineering and Natural Resources, Universiti Teknologi, Malaysia.

Juwari, B.B.A.A. and M.Y.M. Yunus, 2007. Soft gas densitometer using semi-empirical model. Proceedings of the 21st Symposium of Malaysian Chemical Engineers, Dec. 12-14. UPM, Serdang, Selangor, Malaysia.

Kourti, T., J. Lee and J.F. MacGregor, 1996. Experiences with industrial applications of projection methods for multivariate statistical process control. Comput. Chem. Eng., 20: 745-750.

Ralston, P., G. DePuy and H. Graham, 2001. Computer based monitoring and fault diagnosis: A chemical process case study. ISA Trans., 40: 85-98.

Seborg, D.E., T.F. Edgar and D.A. Mellichamp, 2004. Process Dynamics and Control. 2nd Edn., John Wiley and Sons, Inc., New Jersey, New York.

Yang, K. and J. Trewn, 2004. Multivariate Statistical Methods in Quality Management. 1st Edn., McGraw Hill Professional, UK., pp: 288.