APPLICATION OF PARTIAL CORRELATION IN ACTIVE STATISTICAL PROCESS CONTROL

Kamarul Asri Ibrahim
Jabatan Kejuruteraan Kimia
Universiti Teknologi Malaysia

ABSTRACT

Most of the research in Statistical Process Control (SPC) has been focused on the charting techniques which are employed to monitor the process. Unfortunately, little attention is focused on the importance of bringing the process in control automatically via these charting techniques. By utilising Automatic Process Control (APC) concepts it is possible to devise a method whereby the process is monitored and automatically controlled via SPC procedures. The proposed method avoids the time series or dynamic model identification problem, by focusing only on statistical input-output relationship via Partial Correlation Analysis (PCorrA). The advantage of this method lies in the ability of PCorrA to quantify the relationship between quality variables and input variables. A discussion of how to implement PCorrA in SPC is given. Finally the results of applying this strategy to a simulated reversible reaction process are shown and discussed.

Keyword

Statistical Process Control, Shewhart Chart, Partial Correlation Analysis.

INTRODUCTION

Statistical Process Control (SPC) and Automatic Process Control (APC) have different origins. The former was developed in the manufacturing industry, while the later was nurtured in the process industry. Although both strategies come from different backgrounds, the common objective is to reduce the deviation of controlled outputs from target values. SPC accomplishes this by detecting changes when assignable causes occur. On the other hand, APC counteracts the effects of process disturbances so that output quality is always on target. SPC does not control the process, but rather performs a monitoring function and signals when control is needed in the form of identification and removal of the root causes. While APC does not remove the root or assignable causes, it uses continuous adjustments to keep process variables on target. To gain the advantage of both methods, there is an increasing trend towards developing hybrid schemes. (Montgomery *et al.*, 1994; Thompson and Twig, 1994, Tucker *et al.*, 1993; Vander Weil *et al.*, 1992; Box and Kramer, 1992; MacGregor, 1988). Several workers attempt to integrate this concepts to overcome the difficulty of adopting SPC strategy in continuous chemical processes (Vander Weil and Vardeman, 1992, Efthimiadu *et al.*, 1993, MacGregor, 1988).

The area of SPC emphasis in this paper is that of off-line and on-line quality control. The principles involved are similar to those used in the feedforward APC schemes. Since the strategy provides anticipatory correction to prevent out-of-control situations, we call this *active* SPC. It departs from the traditional SPC rule of monitoring product or quality variables by proposing an alternative monitoring and manipulation strategies which emphasis on the deviation in inputs. If the inputs wander away from their respective control limits, it is an indication that the quality variable would also be out of statistical control.

SPC AND CONTROL CHARTS

The two main objective of implementing SPC are: (a) to verify and confirm only inherent or common cause occurred in the system. (b) to monitor, track and eliminate deviations that is unlikely to be due to chance. These are achieved by employing control charts such as Shewhart charts, Cumulative Sum (CUSUM) charts, and Exponential Moving Average (EWMA) control charts to detect *common* cause and *assignable* cause. The term *common* cause is refer as a behaviour because of fault in the process. Nothing can be done accept to alter the process. On the other hand, *assignable* cause is a term for those temporary deviations from target. Thus, when only common cause occurs in the system, the process is said to be in a state of control. On the other hand, when assignable causes inhabit the system, out of control situation occurred and elimination of this behaviour is required.

In order to identify the state of the process mentioned above, the process data must follow a normal distribution curve. This implies that the probability of data values falling in a certain range can be predicted. A set of data is said normally distributed when 99.7% of the data values fall within ± three standard deviations from the distribution mean ($\mu \pm 3\sigma$). Within this region of ($\mu \pm 3\sigma$), the process is in state of control, while beyond this, the process is out of control. By the provision of this "three sigma" limit lines, the application of Shewhart Control charts will allow the process to have stable random variations about the target value.

To be alert, SPC charts will always be continuously performing statistical hypothesis tests. The null hypothesis H_0 for a normal distribution process is that the true process mean (μ) will equal the target process mean (μ_0). The alternative hypothesis, H_1, is that μ does not equal to μ_0. The interpretation of H_o is that the process is behaving well and should be left alone, whereby H_1 means there is a problem and actions should be taken.

Any statistical hypothesis test will carry two types of risks: the producer's risk (α) and the consumer's risk (β). The producer's risk is the chance of making a type I error, i.e., rejecting the null hypothesis when the process output really follows the distribution of $N(\mu_0, \sigma^2)$. This results in taking action due to a signal given by an extreme observation when infact the process does not change at all. This phenomena is also called manufacturer's risk or false alarm.

The second type of risk, the consumer's risk is the chance of making type II error. It means accepting H_o when the process output does not follow the distribution $N(\mu_0, \sigma^2)$. It is the chance of not detecting that the process mean has deviated from the target value. Since the β risk implies that the consumer has the chance of accepting bad or unacceptable products, it is termed consumer's risk.

Two kind of charts are used in this work: Shewhart chart with action line and Shewhart chart with both action and warning lines. It consists of a time plot of data with control limits centred about the target, which was based on previous historical data. The control limits are set at ± 3 standard deviations from the mean. The process is out of control when the current observation falls outside the control limits. To make it more sensitive Shewhart control chart has been modified with supplemental run rule i.e., two out of three observations in a row beyond two standard deviations as out of control.

This contribution describes an alternative approach which possess the advantage of both traditional SPC and APC schemes. The technique is designed to keeps the process output under statistical control automatically without unnecessary manipulations of the inputs. Central to this strategy is the use of Partial Correlation Analysis.

PARTIAL CORRELATION ANALYSIS

Partial Correlation is a multivariable technique that quantifies the relative influence of independent variables $\mathbf{x_1}$ and $\mathbf{x_2}$ on variations in dependent variable $\mathbf{y}$, and whether these separate influences are statistically significant. Consider a (px1) random vector $\mathbf{x}$ which is distributed $N(\mathbf{x}:\mu, \Sigma)$, where Σ is a positive definite covariance matrix. Partition of $\mathbf{x}$, μ and Σ are shown below:

$$x = \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} \qquad \mu = \begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix} \qquad \Sigma = \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix} \tag{1}$$

where $\mathbf{x_1}$ and μ_1 have sizes (qx1) and Σ_{11} the covariance matrix of $\mathbf{x_1}$ has size (qxq). The next step is to calculate the partial covariance matrix $\Sigma_{11.2}$ which is the covariance matrix of $\mathbf{x_1}$ after the dependence of $\mathbf{x_2}$ has been removed and is given by:

$$\Sigma_{11.2} = \Sigma_{11} - \Sigma_{12}\,\Sigma_{22}^{-1}\Sigma_{21} \tag{2}$$

The partial correlation coefficient of x_i and x_j (which are in $\mathbf{x_1}$) given $\mathbf{x_2}$ = constant is denoted by $r_{ij.(q+1,\dots,p)}$ and define by :

$$r_{ij.(q+1,\dots,p)} = \frac{\sigma_{ij.(q+1,\dots,p)}}{\sqrt{\sigma_{ii.(q+1,\dots,p)}\,\sigma_{jj.(q+1,\dots,p)}}} \tag{3}$$

where $\sigma_{ij.(q+1,\dots,p)}$ is the ij th element of $\Sigma_{11.2}$. The function of the partial correlation coefficient is to measure the linear association between two variables when both have been adjusted for their linear association with the remaining variables in the data set. Thus, compared to ordinary correlation analysis, PCorrA provides a better insight into the relationship between input and output variables from a multivariate data set. As shown mathematically above, this is done by separating the population into sub populations in which one or more random variables are held constant before determining the correlation among the other random variables.

Active Statistical Process Control

As being mentioned before, APC strategies will continuously making attempt to remove the effects of any disturbance on the output by adjusting the manipulated variables. In contrast the traditional SPC only give indications on when action should be taken, namely when the quality variable exceed some specified limit on the control chart. It is however, possible to devise a method whereby the process is monitored and automatically controlled via SPC procedure. The new strategy avoids the model identification problem by focusing attention on statistical input-output relationship via PCorrA. Input or manipulated variables are being monitored in addition to product quality. If these inputs can be kept within their respective control limits, then the quality variable will also be maintained within its control limits.

Traditionally the SPC approaches impose limits on the output of interest, e.g. control limits at $\mu \pm 3\sigma$. Based on partial correlation between input and outputs, limits on the output can be translated to corresponding limits on inputs. When partial correlation has been calculated based on the standardised variables, the relationship between standardised quality variable x_j^s and input variable x_i^s can be written as:

$$x_j^s = C_{ij} x_i^s \tag{4}$$

where $x_k^s = \left(x_k - \mu_k\right) / \sigma_k$, $C_{ij} = r_{ij.(q+1,\ldots,p)}$ is the partial correlation coefficient between x_j^s and x_i^s. When $\mu_0 \pm 3\sigma$ control limits are used, then since x_j^s is standardised, the limits becomes:

$$-3 < x_j^s < 3 \tag{5}$$

If we translate the corresponding limits to input variables x_i:

$$-3/C_{ij} < x_i^s < 3/C_{ij} \tag{6}$$

When the variables are not in standardised form Equations (5) can be rewritten as:

$$x_j = \sigma_j + (x_i - \mu_i)\, C_{ij}\, \sigma_j / \sigma_i \tag{7}$$

$$\mu_i - 3\,\sigma_j < x_j < \mu_i + 3\,\sigma_j \tag{8}$$

then the control limits for the input variables become :

$$(-3/C_{ij})\, \sigma_i + \mu_i < x_i < (3/C_{ij})\, \sigma_i + \mu_i \tag{9}$$

In certain situation some of the input variables are not measured using on line measurement. To measure concentration continuously, one needs to install gas chromatography (GC) on-line and this procedure may not be feasible due to economic constraint. In other situation the input variable cannot be manipulated on-line. This work assume that all the input variables can be measured on-line, but some of it variables cannot be manipulated. In this case an alternative manipulative input variable, x_m, which can affect the quality variable x_j must be determined. If the monitored variable x_i exceeds the limits specified, an appropriate adjustment will be make in the input manipulated variable x_m so that the quality variable x_j is still within target though x_i is out of control. The new equation can be expressed by using x_m as follows:

$$x_j = \mu_j + (x_m - \mu_m)\, C_{mj}\, \sigma_j / \sigma_m \tag{10}$$

The predicted value of x_j from the monitor value of x_i is given in Equation (11). Thus the difference between this value and a target limit l_i will be:

$$\Delta x_j = C_{ij}\,(x_i - l_i)\; \sigma_j / \sigma_i \tag{11}$$

The amount that x_m has to be adjusted to compensate for the predicted deviation Δx_j is obtained by combining Eqs. (10) and (11):

$$\Delta x_m = \frac{\Delta x_j \sigma_m}{C_{mj}} = \frac{\left(x_i - l_i\right) C_{ij} \sigma_m}{C_{mj} \sigma_i} \tag{12}$$

For on-line control purposes the value of C_{ij} can be exponentially smooth to reduce any abnormal changes due to the out of control situations. The exponential smoothing coefficient $C_{ij}(t)$, calculated at time t will be :

$$\tilde{C}_{ij}(t) = (1 - \lambda)\, \tilde{C}_{ij}(t-1) + \lambda\, C_{ij}(t) \tag{13}$$

where $\tilde{C}_{ij}(t)$ is the exponentially smoothed value of $C_{ij}(t)$ at time t while λ is the smoothing constant. $\tilde{C}_{ij}(t)$ is then used in place of $C_{ij}(t)$ in the equation above.

Control chart for Shewhart Chart with both action and warning lines, can be developed in similar manner. When there is deviation in input variables from the respective limits, it indicate that the quality variable will suffer out of control situation. If the monitor input variable is also the manipulated variable, adjustment will be made when out of control situation occurred so that the manipulated variable is back on target.

IMPLEMENTATION ASPECTS

Given the above relationships or control laws, process monitoring and correction can thus be automated, yielding on line or 'active' SPC strategies. In monitoring the important variables, the charting techniques employed are familiar to the Shewhart type charts. There is no need to resort to multivariate charts because the inputs and outputs of the process have effectively been decomposed to smaller independent sub-systems. Since the probable causes of process deviations have been predetermined, on-line SPC reduces the need for expensive and time consuming experimentation after the incidence of out of control. Moreover, abnormal variations in the input variables can be corrected before they affect output quality. Control is therefore achieved in an anticipatory manner. An example of performance achievable using this Active SPC strategy are shown in figure 1 and 2.

The proposed on-line SPC methodology was applied on to a non linear simulation of a continuous stir tank reactor (CSTR). Two reactant A and B were fed in excess to the CSTR to produce C. The process is reversible via second order exothermic reaction. The energy generate by the process is absorbed by heat transfer through a cooling jacket. The input variables considered therefore are concentration of Ain, concentration of Bin; the temperature of reactants Tin; the temperature of cooling medium Tjin; the flowrate of reactants F and flowrate of cooling medium Fj. The output variables are the temperature of output mass Tout; cooling medium output temperature Tjout; concentration of Aout and Bout; and lastly Cout the concentration of product C. Cout is chosen as the quality variable which has to be kept under statistical control.

This work will compare two different mode of operation, whether on-line or off-line, two type of charts, Shewhart Action or Shewhart Action with Warning and two ways of adjusting manipulated variables, manipulate all or manipulate some.

For manipulating some variables, it was assumed that the input concentrations of the reactants and the input temperature of the cooling jacket cannot be manipulated on-line. Appropriate adjustment in Tin could compensate the variations in Ain and Bin and thus keep Cout in control. Similarly, appropriate adjustment in Fj could compensate for variations in Tjin.

The simulation is implemented by collecting 50 samples of data from the process. The mean and standard deviation of all variables were calculated and treated as constant for subsequent analysis. Off-line partial correlation will next be applied. Based on this value the control limits line is determined for the rest of analysis in the off-line case. For on-line situation a window of 50 samples is updated at each 5 sampling instances. For the first 50 samples the off-line control limit was used. The new control rule then be updated and calculated based on equation (13) with smoothing constant $\lambda = 0.3$.

DISCUSSION

Figure 3 shows the percentage of out of control (NOC), controlled (NUC) and false alarm (NFA) associated with the quality variables for 500 observations. It compares two different mode of operation whether off-line or on-line, two type of charts, Shewhart with Action lines or Shewhart with Action and Warning lines and two ways of adjusting manipulated variables, manipulate all or manipulate some. It is difficult to discern which configuration gave the best control perfromance from fig 3. However, it can seen that configurations utilising both action and warning lines (code: x2x) provided better control in terms of higher NUCs than the corresponding configuration using only action lines (code: x1x). This is expected as the former is more sensitive as discussed previously. Figure 3 also shows that a high percentage of points under control is normally accompanied by a high rate of false alarms, i.e. increased producer's risk, which is undesirable.

In judging which of the active SPC configuration provided the best performance, a compromise is therefore necessary to balance reduced consumer risk against increased producer risk. In this investigation, the following heuristically derived *index of performance* (IP) was used to identify the configuration that provided the best control:

$$IP = 1 - 0.5\left(\frac{\%\,NOC + \%\,NFA}{\%\,NUC + \%\,NFA}\right) \qquad (14)$$

This index penalises the percentage out-of-control points as well as the percentage of false alarms. It is scaled so that the best control strategy, i.e. no out-of-control points and no false alarms, would have an IP equal to one. The worst case, which corresponds to all points being out-of-control (i.e NOC=100, NUC=0) and 100 percent false alarms, would lead to an IP of zero. The IPs for the various configurations tested are plotted in Fig. 4. It shows that the best performance was obtained using a Shewhart chart with action and warning lines with partial correlations calculated on-line, manipulating only Tin and Fj.(code 222).

CONCLUSION

In this contribution, a procedure that can provide for automatic SPC has been described. Traditional charts are still used for monitoring purposes. In addition to the quality variable, those inputs identified using partial correlation analysis that have the potential for causing process upsets are also monitored. The limits for the monitoring charts and, more significantly, the manipulation rules to keep the process under statistical control arise naturally as part of the analysis. Unlike traditional SPC strategies, control is therefore achieved in an anticipatory manner. Since the potential causes of process deviations have been pre-determined, abnormal variations in the input variables can be corrected automatically before they affect output quality. The need for expensive and time consuming experimentation after out-of-control incidences is therefore reduced. In conclusion the method look promising and the utilisation of different type of control charts and multivariate techniques are the subjects of current investigation.

REFERENCES

[1] Box, G.E.P., and Kramer, T. (1992), " Statistical Process Monitoring and Feedback Adjustment - A Discussion," Technometrics, 34, 251 - 267.

[2] Box, G.E.P. (1993)," Quality Improvement-The New Industrial Revolution," International Statistical Review, 61, 3-19.

[3] Deming, W.E. (1982), " Out of Crisis," Cambridge University Press, Massachusetts.

[4] Efthimiadu, I., Tham, M.T. and Willis, M.J (1993), " Engineering Control and Product Quality Assurance in the Process Industries, " Proceeding of Quality and its Applications 1993. University of Newcastle Upon Tyne. 593-598.

[5] Juran, J.M., Gryna, F.M. Jr., Bingham, R.S.Jr., (1974), " Quality Control Handbook," 3rd. edition McGraw-Hill, New York.

[6] Thompson, M., and Twigg, P.M. (1994), " Statistical Process Control Methods In The Supervision and Selective Application of PID Control," IEE Conference Control 1994, Warwick, United Kingdom, March 21-24, 612 - 617.

[7] MacGregor, J.F. (1988), "On-Line Statistical Process Control," Chemical Engineering Progress, October, 21-31.

[8] Montgomery, D.C., Bert Keats, J., Runger, G.C., and Messina, W.S. (1994), "Integrating Statistical Process Control and Engineering Process Control," Journal of Quality Technology, 26 (2), 79-87.

[9] Tucker, W.T., Faltin, and Tucker, W.T., Faltin. (1993), " Algorithmic Statistical Process Control: An Elaboration," Technometrics, 35, 363 - 375.

[10] Vander Weil, S.A., Tucker, W.T., Faltin, F.W., and Doganaksoy, N. (1992), " Algorithmic Statistical Process Control: Concepts and Applications," Technometrics, 34, 286 - 297.

Graph Nomenclature:

First Digit:	(1) On-line mode	(2) Off-line mode
Second Digit:	(1) Shewhart Action chart	(2) Shewhart Action and Warning Chart
Third Digit:	(1) Manipulate all input variables	(2) Manipulate Tin and Fj

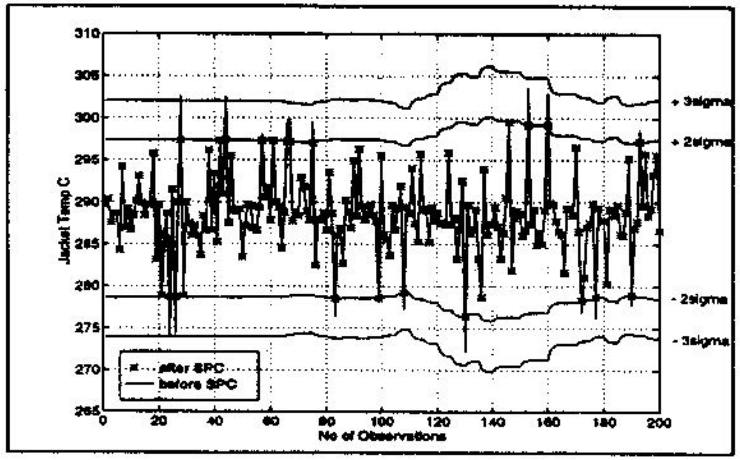

Fig. 1: Jacket Temperature using On-line Shewhart chart with action and warning lines.

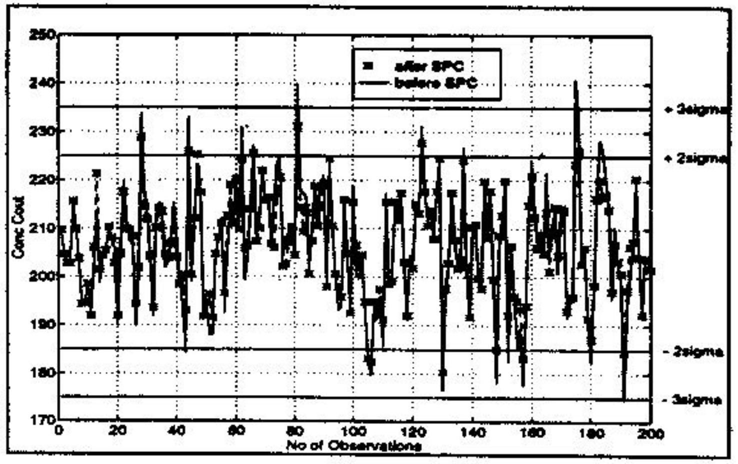

Fig 2 : Output quality control using On-line Shewhart chart with action and warning lines, manipulating all the variables.

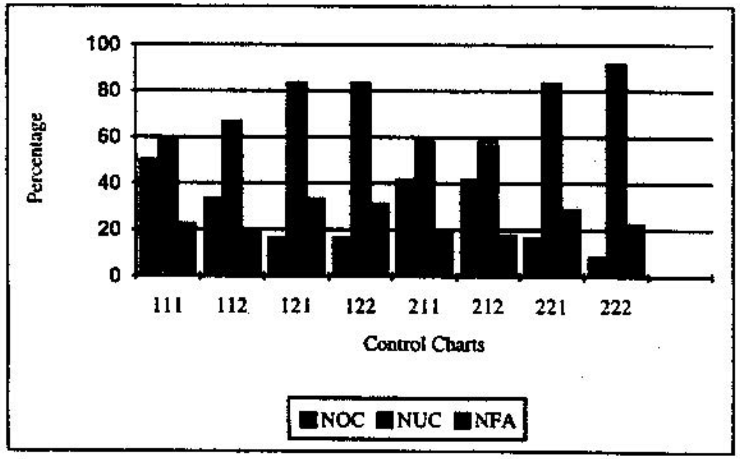

Fig. 3 : Performances of several type of control charts

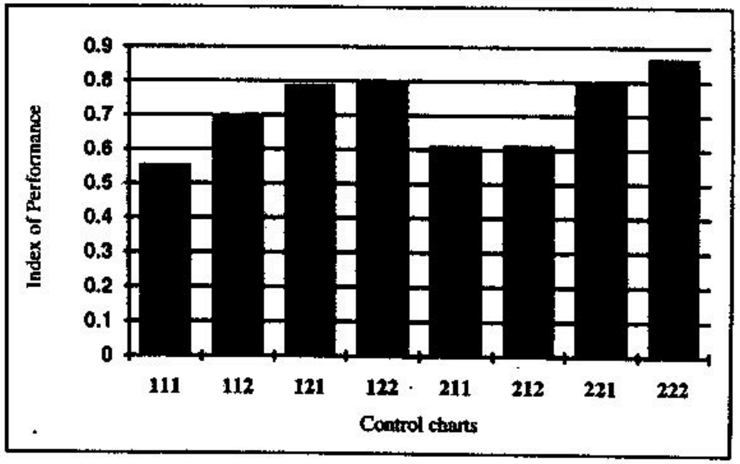

Fig. 4 :Index of Performance for several type of control charts.