APPLICATION OF UNIVARIATE AND MULTIVARIATE PROCESS CONTROL PROCEDURES IN INDUSTRY

Mali Abdollahian* H. Abachi+ and S. Nahavandi++

*Department of Statistics and Operations Research
RMIT University, Melbourne, Australia
+Department of Electrical & Computer Systems Eng,
Monash University, Melbourne, Australia
++School of Engineering Technology, Deakin University, Geelong Australia

ABSTRACT: *Traditional statistical process control charts used to monitor key process variables are based on the assumption that measurements are independent and identically distributed about a target value. In practice they are not and often are actually correlated. Reliance on univariate charts can lead to misleading conclusions. This paper addresses the methods for improving the quality of industrial products using T^2 Multivariate Quality Control charts. In practice, one of the main problems in implementing the T^2 multivariate process control chart is that it only identifies the sample that causes the out-of-control situation. However, T^2 control chart is unable to identify the quality characteristic(s) that caused the out-of-control signal which is regarded as a major disadvantage in the implementation of this control chart. In this paper, Murphy's method is implemented which not only identifies the sample, but also selects and prioritise the out-of-control variables using the T^2 control procedures. The results are compared with the performance of the individual charts. The practical example using the data provided by ESSAR Steel Limited, clearly shows the superiority of multivariate control charts over the univariate charts.*

Keywords: *univariate control charts, $\bar{X}$ charts, R charts, Exponentially Weighted Moving Average (EWMA) charts, Average Run Length (ARL), Hotelling's T^2 procedure, D-test algorithm*

INTRODUCTION

Quality assurance and quality improvement have always been an integral part of virtually all products and services. In terms of manufacturing and cost principles, a major area of concern for any product is related to quality and possibilities for quality enhancement.

This paper addresses the modern practice of statistical quality control in a steel plant. It provides a comprehensive coverage of the subject form the principles to the usage of both univariate and multivariate control charts.

The study was carried out based on data collected from ESSAR Steel Limited, India over the period January 1996 to February 1996. The process industry monitors the quality of the reformed gas stream. The main components of the reformed gas are hydrogen, methane and carbon dioxide and these gas components maintained at their required specification levels determined the overall operations of the reduction process. The gas samples were collected 3 times a day, giving 59 samples of size 3. Here the ratio of hydrogen to carbon monoxide was monitored separately to observe whether the respective compositions were maintained at controlled specifications or not.

The paper discusses the use of univariate control charts, with the most commonly used variable control charts, the $\bar{X}$, the R and the EWMA charts. Since in a process industry where both sudden movements and sudden drifts occur, there is value in examining both the average and the moving average charts. Here the ARLs for both $\bar{X}$ and EWMA charts were computed in order to recommend the steel plant a suitable technique for monitoring their process.

Finally, the paper puts its focus on multivariate control charts with special emphasis on Murphy's (1987) method to select and prioritize the out-of-control variables using the Hotelling's T^2 procedure (Hotelling (1947)). The results are compared with the performance of the individual charts.

DEALING WITH THE DATA

An important assumption underlying statistical quality control is that their interpretation is based on normal distribution of the process output. In practical situations it is generally seen that the key variables do show departures from normality. But monitoring techniques of a statistical process control requires the key variables to approximately follow normal distribution. Researchers have found methods to monitor the process with non-normal data. The study here discusses all techniques considering the assumption of normality.

When the normal probability plots were designed on all the four gas components, the gas samples showed departures from normality. In other words, the data were highly skewed. Transformations such as the logs, square roots, negative reciprocals and negative reciprocal roots could not transform the data to normality. Stronger transformations like the Box-Cox transformations that deal with highly skewed data well also could not achieve normality. The reason for this could be the presence of large outliers in the data set. Davies (1995) provided a method to solve the problem of non-normality. He suggested that calculating the median and throwing away everything that is more than 5.2 absolute deviations from the median would solve the problem. The issue of how many observations can be set aside without compromising the statistical analysis is a difficult one with no obvious solution. However, we could examine the methods of insightful data analysis, and use these as a guide to both what is possible and what is necessary.

Tukey (1991), Stigler (1977) and Rocke et al. (1982) have suggested that the best analyses require deletion of a greater proportion of data. Stigler examined many historical data sets and concluded, "outliers are present in small quantities, but a small amount of trimming (no more than 10 percent) may be the best way of dealing with them". Stigler's 10 percent trimming actually meant discarding 10 percent of the highest observations and 10 percent of the lowest observations, so his recommendations amounts to not more than 20 percent of the total observations. Considering this and the method suggested by Davies it was required to throw away 13% of hydrogen, 15% of carbon dioxide, 19% of methane and 14% of the ratio of hydrogen to carbon monoxide of process data. After having thrown away the required proportion of contaminated data, the gas components were checked for normality. It was observed that all gas components showed to approach normality. The missing observations were treated as missing and were then used for monitoring the process using variable control charts.

UNIVARIATE CONTROL CHARTS AND PROCESS CAPABILITY

The $\overline{X}$ and R charts were designed on hydrogen, methane, carbon dioxide and the ratio hydrogen to carbon monoxide. The R charts for the respective gases were examined for out-of-control signals. It was observed that all gases except the ratio of hydrogen to carbon monoxide showed inherent variation to be in control and that there were no assignable causes disturbing the ranges. Then the $\overline{X}$ charts for hydrogen, methane and carbon dioxide were examined for out-of-control signals. Each of these charts showed out of control signals. The extreme subgroup numbers were identified, removed and new charts were designed based on the new sample list. This procedure was continued until all charts showed in control situations. It was concluded that all gases showed shifts in the process mean.

Once the process parameters were found to be in control, it was necessary to determine whether the process was capable or not. A process capability analysis was performed on all the three gases. Process capability analysis is judged by comparing process performance with the process requirements. Since meeting the specification limits is one of the basic requirements, capability analyses usually involves specification limits in their calculations. The judgement is based on an index called the process capability index, denoted as C_p. If $C_p=1$, then the process parameters is said to be marginally capable of meeting its specification limits. A C_p that exceeds 1.33 is considered to be good and is commonly used as a goal to achieve process capability. On the other hand, if C_p is less than 1.0 it implies that the process is not capable of meeting specifications. Table 1 gives the C_p values obtained for the process parameters considered. It is shown that hydrogen and carbon dioxide met their specifications but the flow rates for methane gas was operating at an unacceptable level.

Table 1: *Process capability index for the process parameters*

Process Parameter	Specification Limits	Capability Index (C_p)
Hydrogen	54.89 ± 4.00	1.72
Methane	1.50 ± 0.50	0.89
Carbon dioxide	2.50 ± 0.50	1.14

In addition to $\overline{X}$ and R charts, the EWMA charts were constructed. It is believed that this chart tends to provide quicker responses to smaller shifts in the process average than the $\overline{X}$ charts do, because each point on a EWMA chart incorporates information from all the subgroups, not just from the current subgroup. EWMA charts were designed on weighting constant, $\lambda = 0.20$ for all the gases. Experience have shown that λ 's in the range 0.10 to 0.30, generally give good results. Different values in the specified range were tried but $\lambda = 0.20$ showed better inherent variations. It was observed that the EWMA charts had responded to the shifts in similar fashions to that of the alternate $\overline{X}$ charts.

To examine the relative size of the shifts, ARLs were calculated for both charts. It was found that the $\overline{X}$ charts had smaller ARLs than the EWMA charts. Table 2 below gives the ARLs obtained for both the $\overline{X}$ charts and the EWMA charts (Crowder (1987)). Looking at table 2 it can be concluded that $\overline{X}$ charts had detected the shifts earlier than the EWMA charts. In other words, the process shifts in the Steel Plant were relatively large.

Table 2: *ARLs for* $\overline{X}$ and EWMA charts

Process Parameter	Mean	Xbar chart	EWMA chart
Hydrogen	59.41	3.15	11.71
Methane	1.63	3.20	9.20
Carbon dioxide	2.40	3.10	12.84

Having monitored the process using univariate charts it was considered necessary to extend the study to multivariate techniques. Rather it is important for the company to monitor their process parameters using multivariate techniques since their process involves more than one parameter to be monitored and requires for their levels to be simultaneously under control.

MULTIVARIATE CONTROL CHARTS

Most of the process industries with modern data acquisition equipment and on-line computers have several streams of correlated data output that can be monitored simultaneously. The task of meaningfully dealing with this class of problem falls into the area of multivariate quality control problems. Therefore many problems in modern industrial quality control will involve a vector of measurements of several quality characteristics, rather than only a single characteristic. One analysis approach would be to ignore the correlation between the characteristics and monitor the process using separate charts for each characteristic. This approach often results in a prohibitively large number of control charts. Moreover, monitoring these correlated characteristics independently can be misleading due to the masking of certain effects that become important in combination. Better sensitivity can be obtained by using multivariate methods that exploit the correlations.

The multivariate control charts though it achieved importance over $\overline{X}$ charts, its interpretation was difficult. Jackson (1980) suggested keeping principal component plots and individual value plots as well, but the collection of plots proved very difficult to interpret. A better method designed by Murphy (1987) put away the problems of interpreting the multivariate control charts. He suggested that an approach based on discrimination, in which the variables are partitioned into subset thought to causing the problem and then calculating the difference between the T^2 statistics based on full and subset variables. The principal advantage of using the T^2 statistic is its proper reflection of the correlation structure of the populations. Not allowing for correlation in the variables is the main weakness in the present practice of using independent charts to track each of the p quality characteristics individually.

Murphy's method enables the person monitoring the process to identify which of their process parameters caused the signal first, second and so on. In other words, his method gives us an order of selection of the out of control variables. He designed an out-of-control variable selection algorithm known as the D-test algorithm. A C program was written by authors to implement his algorithm.

RESULTS

The results shown here are based on process data sets with complete observations. In other words, samples with missing observations were removed and the analysis was carried out with data restricted to 36 samples. Table 3 below gives the observations made when the quality characteristics were monitored individually.

Table 3: *Observations made on individual charts*

Sample	On individual charts $\bar{x}_1$	$\bar{x}_2$	$\bar{x}_3$
1	NS	NS	NS
2	NS	NS	NS
3	NS	NS	NS
4	NS	NS	NS
5	NS	NS	NS
6	NS	NS	NS
7	NS	NS	NS
8	NS	NS	NS
9	NS	NS	NS
10	NS	NS	NS
11	NS	NS	NS
12	NS	NS	NS
13	NS	NS	NS
14	NS	NS	NS
15	NS	NS	NS
16	NS	NS	NS
17	NS	NS	NS
18	NS	NS	NS
19	NS	NS	NS
20	NS	NS	NS
21	**S**	**S**	NS
22	NS	NS	NS
23	NS	NS	NS
24	NS	NS	NS
25	NS	NS	**S**
26	NS	NS	**S**
27	NS	NS	NS
28	NS	NS	NS
29	NS	NS	**S**
30	NS	NS	NS
31	NS	NS	NS
32	NS	NS	NS
33	NS	NS	NS
34	NS	NS	NS
35	NS	NS	NS
36	NS	NS	NS

x_1 = hydrogen; x_2 = methane; x_3 = carbon dioxide; S = Significant; NS = Not significant

D-test results

Table 4 below gives the results obtained on using the D-test algorithm on the process data. With p =3, the in control population $\Pi_0 \sim N_3(\mu_0, \Sigma)$ and that the observed sample means $\bar{x} = (\bar{x}_1, \bar{x}_2, \bar{x}_3) \sim N_3(\mu, \Sigma/n)$ with n = 3. We set

$$\mu_0 = \begin{pmatrix} 59.4110 \\ 1.6759 \\ 2.4073 \end{pmatrix} ; \Sigma / n = \begin{pmatrix} 0.172 & 0.008 & -0.005 \\ 0.008 & 0.022 & 0.008 \\ -0.005 & 0.008 & 0.013 \end{pmatrix}$$

giving moderate correlation $(\rho_{23} = 0.459)$ and low correlations $(\rho_{12} = 0.129)$ and $(\rho_{13} = -0.099)$. The cut-off chosen for the testing of hypothesis is $K = 7.814 = \chi^2_{3, 0.05}$.

Table 4: *Out of control quality characteristics selected using the D-test algorithm*

Sample	$T_3^2(\bar{x})$	On T^2 chart	$T_2^2(\bar{x}_i)$						Test of			Order of selection of variables
			$\bar{x}_1$	$\bar{x}_2$	$\bar{x}_3$	$D_2(3)$	$D_2(2)$	$D_2(1)$	$D_2(3)$	$D_2(2)$	$D_2(1)$	
1	6.735	NS	-	-	-	-	-	-	-	-	-	-
2	3.934	NS	-	-	-	-	-	-	-	-	-	-
3	17.543	**S**	6.483	8.787	3.214	14.329	8.756	11.057	S	S	S	*
4	12.851	**S**	0.982	11.061	0.411	12.441	1.251	11.869	S	NS	S	(2)
5	20.422	**S**	5.943	13.202	0.051	20.371	7.219	14.479	S	S	S	*
6	12.906	**S**	2.673	8.787	0.051	12.855	0.442	10.233	S	NS	S	(2)
7	2.422	NS	-	-	-	-	-	-	-	-	-	-
8	9.465	**S**	1.942	0.026	4.320	5.145	9.439	7.523	S	S	S	(3)
9	9.719	**S**	5.561	1.956	3.214	6 505	7.763	4.158	S	S	NS	(1)
10	9.221	**S**	2.673	1.387	5.631	3.590	7.834	6.548	NS	S	S	(3)
11	11.882	**S**	6.637	0.083	3.214	8.668	11.799	5.545	S	S	NS	(1)
12	2.578	NS	-	-	-	-	-	-	-	-	-	-
13	7.332	NS	-	-	-	-	-	-	-	-	-	-
14	8.707	**S**	1.328	1.633	1.470	7.237	7.074	7.739	S	S	S	*
15	2.601	NS	-	-	-	-	-	-	-	-	-	-
16	2.403	NS	-	-	-	-	-	-	-	-	-	-
17	4.853	NS	-	-	-	-	-	-	-	-	-	-
18	2.402	NS	-	-	-	-	-	-	-	-	-	-
19	6.552	NS	-	-	-	-	-	-	-	-	-	-
20	8.229	**S**	4.260	1.633	1.470	6.759	6.596	3.969	S	S	NS	(1)
21	24.998	**S**	15.944	8.049	0.859	24.131	16.948	9.054	S	S	S	*
22	4.271	NS	-	-	-	-	-	-	-	-	-	-
23	3.252	NS	-	-	-	-	-	-	-	-	-	-
24	2.866	NS	-	-	-	-	-	-	-	-	-	-
25	14.117	**S**	0.982	3.792	13.611	0.506	10.325	13.135	NS	S	S	(3)
26	19.944	**S**	7.609	3.792	13.611	6.397	16.152	12.335	S	S	S	*
27	14.683	**S**	5.942	5.707	6.434	8.244	8.976	8.740	S	S	S	*
28	0.893	NS	-	-	-	-	-	-	-	-	-	-
29	13.854	**S**	1.390	3.792	11.584	2.270	10.062	12.464	NS	S	S	(3)
30	11.884	**S**	2.760	1.633	7.825	3.913	10.251	9.124	NS	S	S	(3)
31	6.455	NS	-	-	-	-	-	-	-	-	-	-
32	9.954	**S**	4.595	0.371	3.831	6.123	9.583	5.359	S	S	NS	(1)
33	13.390	**S**	8.682	0.532	0.649	12.471	12.858	4.708	S	S	NS	(1)
34	1.344	NS	-	-	-	-	-	-	-	-	-	-
35	2.116	NS	-	-	-	-	-	-	-	-	-	-
36	8.262	**S**	3.928	1.956	4.320	3.942	6.306	4.334	NS	S	NS	(3,1)

x_1 = hydrogen; x_2 = methane; x_3 = carbon dioxide; S = Significant; NS = Not significant; * - all variables require attention

DISCUSSION

Comparing tables 3 and 4, we see that the number of samples to have indicated an out of control signal on the individual charts is comparatively less than the number flagged on the multivariate T^2 charts. The reason could be that these correlated characteristics when monitored independently masked the interaction effects and caused it to mislead in indicating which variable or subset of variables caused the out of control signal.

Samples 4, 6, 8-11, 20, 21, 25, 29, 30, 32 and 33 in table 4, have their differences not significant thus implying that only those individual variables require attention. But in the case of sample 36 we observe that quality characteristics (1 and 3) require attention. This anomaly has been caused by low correlation between $\bar{x}_1$ and $\bar{x}_3$, although we should have expected a signal for quality characteristic 2 and 3 which have moderate correlation. To identify the reason for this anomaly a check was carried out on the process operations at that time. It was found that there were multiple upsets in the process operations. In any process industry multiple upsets are uncommon. It is also not uncommon for two quality characteristics which are not correlated to be out-of-control, in this case $\bar{x}_1$ and $\bar{x}_3$. Carbon dioxide went out of control due to contamination of the catalyst stream with water. On the other hand, hydrogen went out of control due to increased temperature in the furnace caused by the reduction in the flow of cooling water into furnace inter-cooler, which in turn was caused by the tripping of the cooling water pump.

CONCLUSION

In this paper a basis has been presented for improving achievable quality in multi stream processes by applying the multivariate control chart in place of the univariate approach. It has been demonstrated that the univariate charts can in some cases give a misleading indication that a process is in control, when the multivariate approach would have flagged a problem. The implementation of the D-test algorithm enables the manufacturers to identify and prioritise the quality characteristics that caused the out of control signals. There are considerable potential benefits to be gained by manufacturers in upgrading quality control procedures through the use of multivariate control techniques.

REFERENCES

Crowder, S.V. (1987), *"A Simple Method of Studying Run-Length Distributions of Exponentially Weighted Moving Average Charts"*. Technometrics, vol 29, pp. 401-407.

Davies, P.L. (1995), *"Data Features".* Statistica Neerlandica, vol. 49, pp. 185-245.

Hotelling, H. (1947), *"Multivariate Quality Control"*, in Eisenhart, Hatsay and Wallins (Eds). Techniques of Statistical Analysis, McGraw-Hill, New York.

Jackson, J.E. (1980*), "Principal Components and Factor Analysis: Part 1 - Principal Components"*. Journal of Quality Technology, vol. 12, pp. 210-213.

Murphy, B.J. (1987), *"Selecting Out of Control Variables with* T^2 *Multivariate Quality Control Procedure"*. The Statistician, vol. 36, pp. 571-583.

Rocke, D.M., Downs, G.W. and Rocke, A.J. (1982), *"Are Robust Estimators Really Necessary?"* Technometrics, vol. 24, pp. 95-101.

Stigler, S. (1977), *"Do Robust Estimators Work with Real Data?"* Annals of Satistics, vol. 2, pp. 1055-1098.

Tukey, J.W. (1991), *Exploratory Data Analysis*, Reading, Mass: Addison Wesley.