A Simulation Based Study on the Effectiveness of Bootstrap-Based T^2 Control Chart for Normal Processes

Venu Balijireddy [1], Vijaya Babu Vommi [2]

[1] *Department of Mechanical Engineering, Raghu Institute of Technology, Dakamarri, Visakhapatnam – 531162, AP – INDIA.*
[2] *Department of Mechanical Engineering, Andhra University College of Engineering (A), Visakhapatnam – 530003, AP – INDIA.*

ABSTRACT ---- *Control charts have been used effectively for years to monitor processes and to detect and correct the abnormal behaviour. Systematic use of a control chart is an excellent way to reduce the variability. Univariate control charts are used when a single quality characteristic is of importance, otherwise a multivariable control chart is suggested. A popular multivariate control chart for variable quality characteristics is a Hotelling's T^2- Control chart. The main limitation for T^2- control chart is that it is applicable only when the process under consideration follows a multivariate normal distribution. Poovich et al. (2011) suggested a bootstrap based T^2- control chart for normal as well as non-normal processes. In the present study, the performance of a bootstrap based multivariate T^2 - control chart has been studied for its effectiveness in monitoring a normal process. A simulated process with two quality characteristics has been considered and the performance of conventional T^2 control chart and the boot-strap based T^2 control chart is compared with different sample sizes. It is observed that the performance of bootstrap based T^2 control chart is superior in detecting the shifts in the mean compared to that of conventional T^2 control chart.*

Keywords --- Hotelling's T^2 control chart, Bootstrap based T^2 chart, Shift in mean

I. INTRODUCTION

The main purpose of a control chart is the detection of an out – of – control signal so that process quality can be maintained and production of defective products prevented. Univariate control charts have been devised to monitor the quality of a single process variable; multivariate control charts monitor a number of process variables simultaneously. When many charts are maintained, there is a not-so-small probability that at least one chart will emit an out-of-control message due to chance alone. Meltzer and Storer (1993) described a scenario in which approximately 200 control charts for individual variables were used, and false signals occurred so frequently that both the production shop and the engineering personnel lost confidence in the charts. If 200 separate charts were to be used and points were plotted one at a time, the probability of having a false signal from at least one chart at a particular point in time is $1 – (1\text{-}0.0027)^{200} = 0.42$, if the quality characteristics were independent, normality was assumed, and 3-sigma limits were used. Therefore, this might be used as an approximation if the correlations between the quality characteristics were quite small, since the actual probability cannot be determined analytically.

Multivariate process measurement benefits from the use of inherent multivariate methods rather than a collection of univariate charting methods applied to the individual components. The development of multivariate control charts originates from the work by Hotelling (1947). Recent works have focused mostly on developing control charts for monitoring small changes in the process mean. Woodall and Ncube (1985), Healy (1987), Crosier (1988), Pignatiello and Runger(1990) and Hawkins (1991, 1993) for accounts of multivariate cumulative SUM (MCUSUM) control charts and Lowry et al.(1992), Runger and Prabhu (1996) and Linderman and Love (2000) for accounts of multivariate exponentially weighted moving average (MEWMA) control charts. Qiu and Hawkins (2001, 2003) proposed a rank-based multivariate CUSUM procedure to detect a shift in the process mean. Other recent works focus on developing procedures for monitoring the process variability. Alt and Bedewi (1986), Tang and Barnett (1996a,b), Liu (1995), Chan and Zhang (2001), Yeh et al. (2003, 2004, 2005) and Hawkins and Maboudou-Tchao (2008) for example. Generally, the process mean and variance may change simultaneously during the monitoring period.

Hawkins and Maboudou-Tchao (2008) considered a combination of the MEWMA chart and the multivariate exponentially weighted moving covariance matrix (MEC) chart, which is called the MAC chart here. Alt (1985) gave a review of multivariate quality control charts and pointed out that an important area worth further research was to develop a single control chart for the simultaneous monitoring of both the process location and dispersion.

In the present study, the main focus is on comparing the performance of Hotelling's T^2 control chart (Hotelling, 1947) with that of bootstrap T^2 chart. The T^2 statistic measures the distance between an observation and the scaled-mean estimated from the in-control data. Assuming that the observed process data follow a multivariate normal distribution, the control limit of a T^2 control chart is proportional to the percentile of an F distribution (Mason and Young, 2002).

The remainder of this article is organised as follows. Section 2, describes the multivariate Hotelling's T^2 control chart and the procedure to calculate the boot strap based T^2 control chart. Section 3 presents the comparison of the performance of both Hotelling's and bootstrap based T^2 control on detecting the shifts for a simulated normal multivariate process. Section 4 provides the summary and conclusions.

II. MULTIVARIATE CONTROL CHARTS

There are many situations in which the **simultaneous monitoring** or control of two or more related quality-characteristics is necessary. For example, suppose that a bearing has both an inner diameter (x_1) and an outer diameter (x_2) that together determine the usefulness of the part. Suppose that x_1 and x_2 have independent normal distributions. Because both quality characteristics are measurements, they could be monitored by applying the usual $\bar{x}$ chart to each characteristic.

Monitoring these two quality characteristics independently can be very misleading. The probability that either $\bar{x}_1$ or $\bar{x}_2$ exceeds three-sigma control limits is 0.0027. However, the joint probability that both variables exceed their control limits simultaneously when they are both in control is (0.0027) (0.0027) = 0.00000729, which is considerably smaller than 0.0027. Furthermore, the probability that both $\bar{x}_1$ and $\bar{x}_2$ will simultaneously plot inside the control limits when the process is really in control is (0.9973) (0.9973) = 0.99460729. Therefore, the use of two independent $\bar{x}$ charts has distorted the simultaneous monitoring of $\bar{x}_1$ and $\bar{x}_2$, in that the type I error and the probability of a point correctly plotting in control are not equal to their advertised levels for the individual control charts. This distortion in the process-monitoring procedure increases as the number of quality characteristics increases. In general, if there are p statistically independent quality characteristics for a particular product and if an $\bar{x}$ chart with P {type I error} = α is maintained on each, then the true probability of Type I error for the joint control procedure is

$$\alpha' = 1 - (1 - \alpha)^{p} \quad (1)$$

And the probability that all p means will simultaneously plot inside their control limits when the process is in control is p{all p means plot in control}= $(1 - \alpha)^p$ (2)

Clearly, the distortion in the joint control procedure can be severe, even for moderate values of p. Furthermore, if the p quality characteristics are not independent, which usually would be the case if they relate to the same product, then equations (1) and (2) do not hold, and we have no easy way even to measure the distortion in the joint control procedure.

Hence, the quality characteristics should not be charted independently. The appropriate distribution when the p quality characteristics are jointly distributed as normal is the multivariate normal distribution. The density function of the multivariate normal distribution is given as under:

$$f(x) = \frac{1}{(2\pi)^{p/2}|\Sigma|^{1/2}} e^{-\frac{1}{2}(x-\mu)'\Sigma^{-1}(x-\mu)} \quad (3)$$

Where $-\infty < x_j < \infty$, j = 1, 2, ………, p.

Σ -- covariance matrix

μ -- mean of the normal distribution

Most popular control chart based on multivariate normal distribution is a Hotelling's T^2 control chart. Hotelling's T^2 control charts have been widely used to monitor multivariate processes with individual observations (Hotelling, 1947). Suppose that a dataset contains n observations and each observation is characterised by p process variables. Assume that the dataset follows a multivariate normal distribution with an unknown μ and a covariance matrixΣ.

The Hotelling's T^2 statistics can be calculated by the following equation:

$$T^2 = (x - \bar{x})^T S^{-1} (x - \bar{x}) \quad (4)$$

where $\bar{x}$ is a sample mean vector and S is a sample covariance matrix from an in control process. The control limits of T^2 can be computed by using the procedures that will be discussed in the subsequent sections.

III. BOOTSTRAP-BASED T^2 CONTROL CHART

In order to compare the performance of the Hotelling's T^2 control chart with that of boot-strap control chart, the procedure proposed by Poovich et al (2011) to calculate control limits is explained as under:

1) Compute the T^2statistics with n observations from an in-control dataset.

2) Let $T_1^{2(i)}$, $T_2^{2(i)}$,...., $T_n^{2(i)}$ be a set of n T^2values from i^{th} bootstrap sample (i =1, ... , B) randomly drawn from the initial T^2statistics with replacement. In general, B is the large number (e.g., B > 1,000).

3) In each of B bootstrap samples, determine the 100 · $(1 - \alpha)^{th}$ percentile value given a users - specified value αwith a range between 0 and 1.

4) Determine the control limit by taking an average of B 100 · $(1 - \alpha)_{th}$ percentile values ($\bar{T}^2_{100(1-\alpha)}$). Note that statistics other than the average can be used (e.g., median).

5) Use the established control limit to monitor a new observation. That is, if the monitoring statistic of a new observation exceeds $\bar{T}^2_{100(1-\alpha)}$, we declare that specific observation as out of control.

The boot strap T^2control charts and Hotelling's T^2 control charts are developed using Matlab programs. The multivariate normal data set with two quality characteristics is obtained by simulation and the control limits for monitoring the process are obtained with a data set consisting of 100 subgroups. The control limits are established for both the charts with the procedure described above. A shift in the mean values is introduced and the data set with 1000 subgroups following multivariate normal distribution has been generated with the new mean. The performance of the control charts is compared by recognizing their ability to detect the assignable causes by indicating the points plotting outside the control limits. Simulation experiments are conducted with different sample sizes and the performances are compared.

Case 1: Sample size n = 2

Process variables, p= 2

No of subgroups, m=100

The control limits obtained for both types of the control charts are as follows:

UCL for boot strap control chart = 6.5231

UCL for Hotelling's T^2 control chart = 9.9561

Figure 1 shows the control limits and the in-control data plotted on both the control charts. After introducing the shift in the mean, the data of 1000 subgroups are plotted in both types of control chart. Figure 2.1 shows the performance of the both control charts in detecting the shifts. Similarly, Figure 2.2 shows the performance of both charts after introducing the shift in the second run. Likewise, for the ten runs, Figure 2.1 through Figure 2.6 show the performance of both the charts after the shift is introduced.

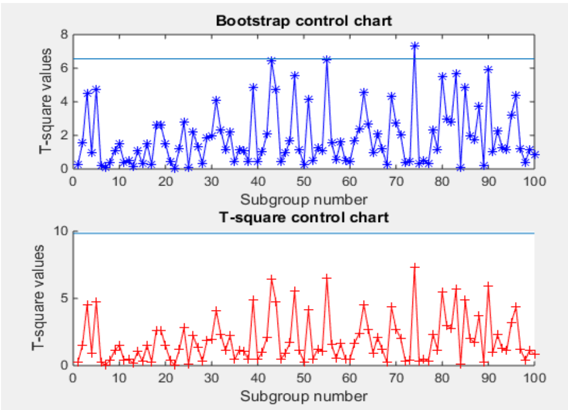


Figure 1: T-square control chart and Bootstrap based T-square control chart for sample size, n=2

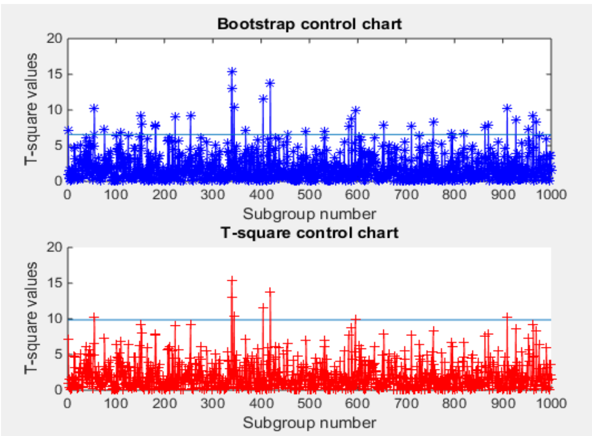


Figure 2.1: Comparison of the control charts with shift in the mean for the first run

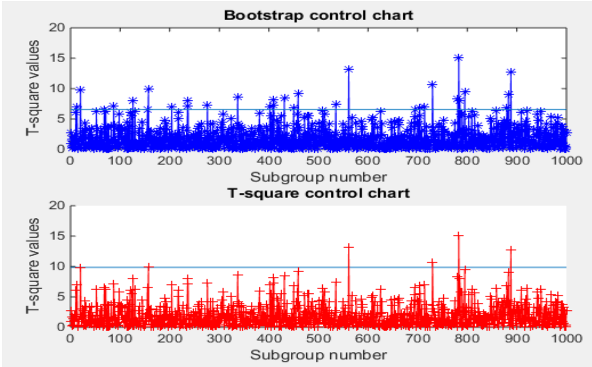


Figure 2.2: Comparison of the control charts with shift in the mean for the second run

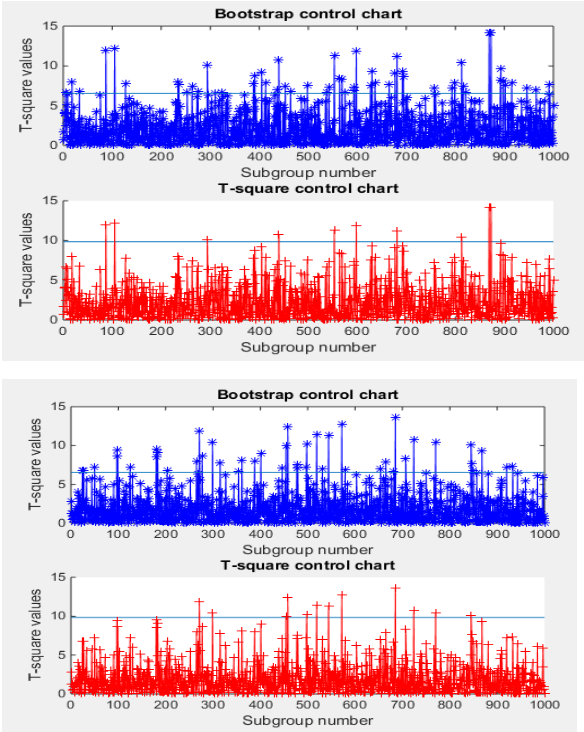
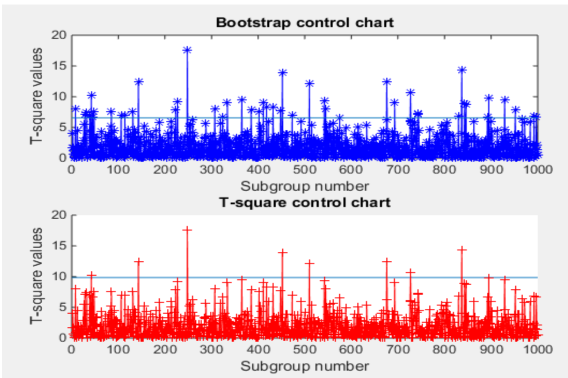


Figure 2.3 Comparison of the control charts with shift in the mean for the third and fourth runs



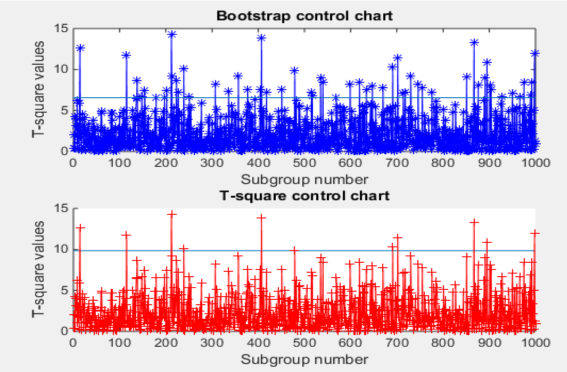


Figure 2.4: Comparison of the control charts with shift in the mean for the fifth and sixth runs

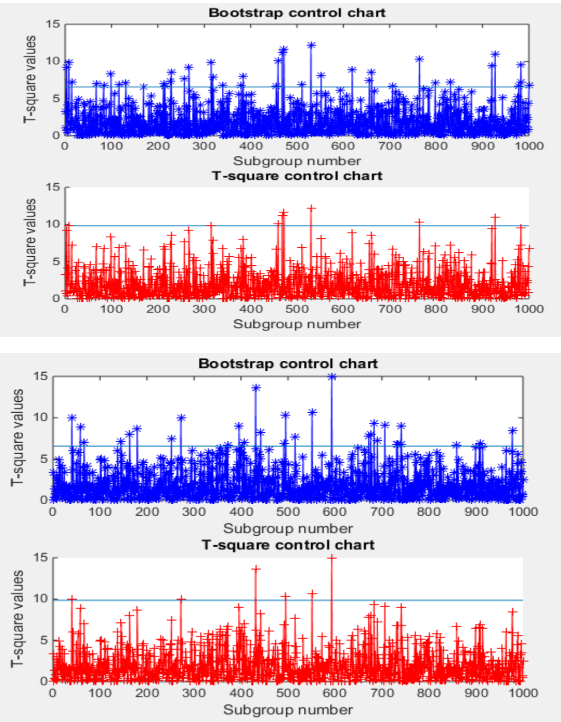
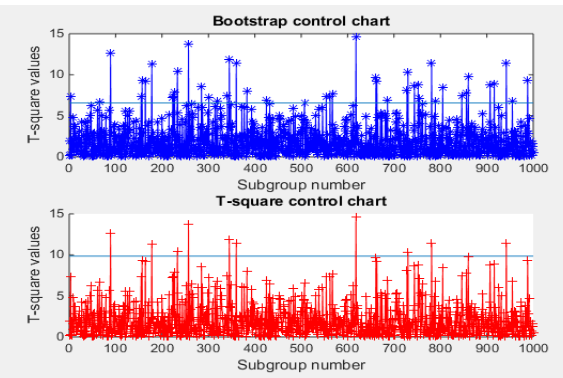


Figure 2.5: Comparison of the control charts with shift in the mean for the seventh and eighth runs



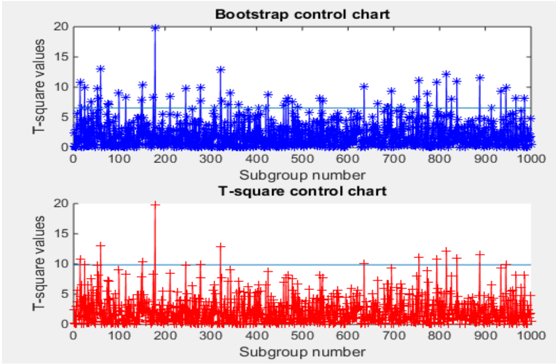


Figure 2.6: Comparison of the control charts with shift in the mean for the ninth and tenth runs

The out of control points indicated by the control charts serve as the basis for measuring the performance of the respective control chart. This is because of the intentional change in the mean value of the process. Table 3.1 shows the comparison of the performances for both the charts with a sample size of n = 2. It can be observed from the Table 3.1 that in all the 10 runs, bootstrap control chart was successful more number of times in detecting the shift. On average, the power of the control chart is observed to be 0.431 where as the Hotelling's T ^{2} control chart showed a power of 0.092.

Table 3.1: Comparison of effectiveness of the charts for sample size, n=2

S.No	No of Sample Points Plotting Outside the Control Limits of	
	Bootstrap-Based T^2 Chart	Hotelling's T^2 Chart
1	34	8
2	30	5
3	55	10
4	44	12
5	46	8
6	56	11
7	39	8
8	30	6
9	43	10
10	53	14
Power of Bootstrap-Based T^2 Chart		Power of Hotelling's T^2 Chart
0.431		0.092

By following the same procedure used for n = 2, the results of simulations with sample sizes of n = 4, 6, 8, 10 and 15 are tabulated in Table 3.2 through Table 3.6.

Table 3.2: Comparison of effectiveness of the charts for sample size, n=4

S.No	No of Sample Points Plotting Outside the Control Limits of	
	Bootstrap-Based T^2 Chart	Hotelling's T^2 Chart
1	20	5
2	24	13

3	15	4
4	19	9
5	20	10
6	19	7
7	26	10
8	17	6
9	28	12
10	19	9
Power of Bootstrap-Based T^2 Chart		Power of Hotelling's T^2 Chart
0.247		0.085

Table 3.3: Comparison of effectiveness of the charts for sample size, n=6

S.No	No of Sample Points Plotting Outside the Control Limits of	
	Bootstrap-Based T^2 Chart	Hotelling's T^2 Chart
1	33	5
2	36	14
3	52	17
4	36	11
5	35	8
6	38	8
7	38	13
8	24	5
9	27	3
10	30	10
Power of Bootstrap-Based T^2 Chart		Power of Hotelling's T^2 Chart
0.349		0.094

Table 3.4: Comparison of effectiveness of the charts for sample size, n=8

S.No	No of Sample Points Plotting Outside the Control Limits of	
	Bootstrap-Based T^2 Chart	Hotelling's T^2 Chart
1	22	8
2	19	4
3	17	8
4	20	8
5	23	12
6	23	8
7	18	8
8	25	9
9	24	7
10	25	11
Power of Bootstrap-Based T^2 Chart		Power of Hotelling's T^2 Chart
0.216		0.083

Table 3.5: Comparison of effectiveness of the charts for sample size, n=10

S.No	No of Sample Points Plotting Outside the Control Limits of	
	Bootstrap-Based T^2 Chart	Hotelling's T^2 Chart
1	49	13
2	42	17
3	31	13
4	47	12
5	45	10
6	28	3
7	43	10
8	38	9
9	49	11
10	39	8
Power of Bootstrap-Based T^2 Chart		Power of Hotelling's T^2 Chart
0.411		0.106

Table 3.6: Comparison of effectiveness of the charts for sample size, n=15

S.No	No of Sample Points Plotting Outside the Control Limits of	
	Bootstrap-Based T^2 Chart	Hotelling's T^2 Chart
1	19	9
2	31	16
3	17	10
4	15	8
5	18	12
6	15	6
7	18	8
8	22	14
9	20	11
10	30	17
Power of Bootstrap-Based T^2 Chart		Power of Hotelling's T^2 Chart
0.205		0.111

From above results, it is clear that the probability of finding assignable causes or the power of the chart is more for a Bootstrap-based T^2 control chart compared to Hotelling's T^2 control chart.

IV. SUMMARY AND CONCLUSIONS

In the present study, the performance of Hotelling's T^2 control chart with that of a bootstrap based T^2 control chart is studied. A process with two quality characteristics following a multivariate normal distribution has been simulated and the control limits for both the charts have been calculated when the process is in control. A shift in the process mean has been introduced, and the process is monitored with both the types of control charts. The chart which has the higher probability of recognizing the shift is observed with different sample sizes. In all the cases, it has been observed that the bootstrap based control chart outperformed the Hotelling's T^2 control chart. The main limitation for implementing the T^2 control chart is that the process must follow a multivariate normal distribution. The present study reveals the fact that even though the process follows a normal distribution, the performance of Hotelling's T^2 control chart is inferior to bootstrap based T^2 control chart. Since the construction of bootstrap control chart do not rely on any statistical distributionbe is easily understood by personnel without a strong statistical , even though the process does not follow normal distribution it can be safely applied. Also, the construction of bootstrap control chart background.

REFERENCES

[1] Alt and Bedewi (1986), *Monitoring process variability.*

[2] Babak Abbasi, Montserrat Guillen, *Bootstrap control charts in monitoring value at risk in insurance,* An article published on-line in 2013.

[3] Chan and Zhang (2001), *Monitoring process variability.*

[4] Crosier (1988), *Multivariate cumulative SUM (MCUSUM) control charts.*

[5] Eugene L. Grant and Richard S. Leavenworth, *Statistical Quality Control,* 7th edition, Tata McGraw - Hill, 2006.

[6] Hawkins (1991, 1993), *Multivariate cumulative SUM (MCUSUM) control charts.*

[7] Hawkins and Maboudou - Tchao (2008), *Monitoring process variability.*

[8] Healy (1987), *Multivariate cumulative SUM (MCUSUM) control charts.*

[9] Hotelling, H. *Multivariate Quality Control,* In: Eisenhart, C., Hastay, M.W., Wills, W.A., eds., Techniques of Statistical Analysis, New York: McGraw - Hill, pp. 111-184, 1947.

[10] J.N. Pan, *A study of multivariate pre-control charts* An article published on-line on 2nd October 2009. An article published on-line on 5th January 2005. George Stefatos , A. Ben Hamza, Fault detection using robust multivariate control chart, An article published on-line in 2008.

[11] Kazuya Nishimura , Shun Matsuura, Hideo Suzuki, *Multivariate EWMA control chart based on a variable Selection using AIC for multivariate statistical process Monitoring,* An article published on-line on 20th January 2015.

[12] Linderman and Love (2000*), Multivariate exponentially weighted moving average (MEWMA) control charts.*

[13] Liu (1995), *Monitoring process variability.*

[14] Lowry et al.(1992), *Multivariate exponentially weighted moving average (MEWMA) control charts.*

[15] Montgomery, D.C. (2000*) Statistical Quality Control – A modern introduction,* Wiley - India.

[16] Pignatiello and Runger(1990), *Multivariate cumulative SUM (MCUSUM) control charts.*

[17] Polansky, A.M., *A general frame work for constructing control charts,* Quality and Reliability Engineering International 21: 633 – 653, 2005.

[18] Poovich Phaladiganon, Seoung Bumkim, Victoria C.P. Chen, Jun-Geol Baek and Sun - Kyoung Park, *Multivariate Control Charts,* Communications in Statistics - Simulation and Computation, An article published on-line on 16th March 2011.

[19] Qiu and Hawkins (2001 and 2003), *Rank-based multivariate CUSUM procedure.*

[20] Richard A. Johnson, R.A. and Dean W. Wichern *Applied Multivariate Statistical Analysis,* Pearson/Prentice Hall, Sixth Edition, 2008.

[21] Rudra Pratap, *Matlab*, Oxford University Press, 2010.

[22] Runger and Prabhu (1996), *Multivariate exponentially weighted moving average (MEWMA) control charts.*

[23] Tang and Barnett (1996a,b), *Monitoring process variability.*

[24] Thomas P. Ryan, *Statistical Methods for Quality Improvement,* John Wiley & Sons, Inc., 2000.

[25] Woodall and Ncube (1985*), Multivariate cumulative SUM (MCUSUM) control charts.*

[26] Woodall, W.H., *Controversies and contradictions in statistical process control, Journal Quality Technology* 32(4): 341-350, 2000.

[27] Woodall, W.H., Montgomery D.C., *Research issues and ideas in statistical process control,* Journal of Quality Technology 31(4): 376-386, 1999.

[28] Yeh et al. (2003, 2004, 2005), *Monitoring process variability.*