Markov Chain Method for Monitoring Mean Vector in Multivariate Cumulative Sum Control Chart

O. Ikpotokin
Department of Mathematics and Statistics, Ambrose Alli University, Ekpoma, Nigeria.
E-mail: ikpotokinosayomore@yahoo.co.uk

Abstract

In many quality control settings, the manufacturing process may have two or more correlated variables. The usual practice has been to maintain a separate (univariate) chart for each characteristic. Unfortunately, this could result in some false (out-of-control) alarms when the characteristics are highly correlated. Therefore, the purpose of this work is to apply MCUSUM scheme to simultaneously monitor the quality characteristics that can identify a change in mean vector of steel manufacturing process, machining process and detergent production process. Results obtained by Markov Chain procedure gave the desired in-control Average Run Length as 250, 200, 200 *and the decision limit* (h) *as* 3.97, 4.86, 6.64 *respectively. While the graphical results showed that the* 3*rd*, 11*th and* 6*th samples respectively are the point at which out-of-control signal set in. Hence, the ability of the MCUSUM chart to detect small to moderate shift in the mean vector was demonstrated.*

1 Introduction

The quality of any manufactured product is to meet customer expectations. Product that meets the required need is one produced by a process capable of operating with little variability around the target of the products quality characteristics. In real life situations, processes often operate in the in-control state for relatively long periods of time. After a while, assignable causes will occur at random resulting in a shift to an out-of-control state where a larger proportion of the process output does not conform to requirements, Montgomery (2009). The control charting procedure recognizes as a fact that, in any production process, a certain amount of variability will always be present. This variability can be either due to chance or due to assignable causes. The form of variability that is of more interest in control charting procedures is the assignable cause of variation caused by either internal or external factors such as machine setting, operator error, raw material or other factors that can be controlled. The presence of this type of variability represents an unacceptable level of process performance and results in an out-of-control state. When this happen, the cause should be identified and eliminated from the process. Control charts are the most widely used Statistical Process Control (SPC) tools to reveal abnormal variations of monitored measurements, as well as to locate their assignable causes, Bersimis *et al.* (2007), Yu and Xi (2009).

In many quality control settings, the manufacturing process may have two or more correlated variables. The usual practice has been to maintain a separate (univariate) chart for each characteristic. Unfortunately, this could result in some false (out-of-control) alarms when the characteristics are highly correlated. Various types of multivariate control charts have been proposed to take advantage of the relationships among the variables being monitored. These are the Hotellings T2, Multivariate Exponentially-weighted moving Average (MEWMAS) and the multivariate cumulative sum (MCUSUM) charts, Hotellings (1947), Crosier (1988), Crowder (1989), Pignatiello and Runger (1990), Lowry *et al.* (1992), Smiley and Keoagile (2005).

Estimation of multivariate statistical process control charts could be achieved based on the underlying distributional assumptions with known parameters and absence of outliers in the data, Hawkins (1991), Runger *et al.* (1995). Multivariate statistical methods which provide simultaneous scouting of several variables are needed for monitoring and diagnosis purposes in modern manufacturing systems. The problem associated with use of multivariate control charts is that; it is computationally demanding to estimate the covariance matrix for high-dimensional variables. This problem shall be addressed with the use of computer software packages like MATLAB. Therefore, the purpose of this work is to apply Markov Chain Method to simultaneously monitor the quality characteristics that identify change in mean vector of steel manufacturing process, machining process and detergent production process. This work also presents a graphical approach to identify the source of signal variability from the multivariate cumulative sum control (MCUSUM) chart, and as well examine the number of points for the process to be in control (ARL).

2 Multivariate Cumulative Sum Control Chart for Mean Vector

One of the charts developed in an effort to supplement the Shewhart chart is the CUSUM control chart which was first introduced by Page (1954). This chart has been widely used to monitor the quality of continuous manufacturing processes. This technique plots the cumulative sums of deviations of the sample values from a target value against time. An important feature of the CUSUM chart is that, it incorporates all the information in the sequence of sample values. This makes the CUSUM chart more sensitive to even smaller shifts in the process mean, Marquardt (1984). Assuming that there is a sequence of independent and identically distributed multivariate normal random variables $X_1, X_2, \cdots$, where $X_1 = (X_{11}, \cdots, X_{1P})$ is a p 1 vector of observations. The first $X_1, X_2, \cdots, X_{(m-1)}$ vectors, have a good distribution function F_G, but the next $X_m, X_{(m+1)}, \cdots$, have a different distribution, F_B showing shift in the mean vector. And if the production process shifts at an unknown time m, the objective is to detect the shift as well as the time of the shift. The CUSUM procedure signals when the shift in the mean vector has occurred as soon as:

$$S_i = \max\left(0, s_{i-1} + \log\frac{f_B(x_i)}{f_G(x_i)}\right) > L \tag{1}$$

where is the CUSUM chart for detecting a shift in the mean vector of a multivariate normal process, and are densities corresponding to F_G and F_B, respectively and L is a constant that determines the operating characteristics of the CUSUM procedures, Healy (1987), Lucas and Crosier (2000), Smiley and Keoagile (2005). Assuming that X_i comes from a multivariate normal distribution with either a good mean μ_G, when the process is in control, or bad mean μ_B, when the process is out of control ($\mu = \mu + \delta$) and with a known common covariance matrix Σ. If for each independent normal random variable, X_i will measure p quality characteristics, a vector of size $p \times 1$ is formed and a covariance matrix of order $p \times p$ is also formed. For the multivariate normal distribution, the CUSUM chart is developed through the likelihood ratio given as:

$$\begin{aligned}\frac{f_B(x_i)}{f_G(x_i)} &= \frac{(2\pi)^{-p/2}|\Sigma|^{-1/2}\exp(-0.5(X_i-\mu_B)'\Sigma^{-1}(X_i-\mu_B))}{(2\pi)^{-p/2}|\Sigma|^{-1/2}\exp(-0.5(X_i-\mu_G)'\Sigma^{-1}(X_i-\mu_G))} \\ &= \frac{\exp(-0.5(X_i-\mu_B)'\Sigma^{-1}(X_i-\mu_B))}{\exp(-0.5(X_i-\mu_G)'\Sigma^{-1}(X_i-\mu_G))}\end{aligned} \tag{2}$$

Taking natural logarithms, we obtain

$$\log\frac{f_B(x_i)}{f_G(x_i)} = (\mu_B-\mu_G)'\Sigma^{-1}X_i - 0.5(\mu_B+\mu_G)'\Sigma^{-1}(\mu_B-\mu_B) \tag{3}$$

The CUSUM for the multivariate process is computed by substituting Equation (3) into Equation (2) and dividing both sides of Equation (3) by a constant gives the CUSUM procedure for the multivariate process as:

$$S_i = \max(S_{i-1} + a'X_i - k, 0) \tag{4}$$

where

$$a' = \frac{(\mu_B-\mu_G)'\Sigma^{-1}}{\left[(\mu_B-\mu_G)'\Sigma^{-1}(\mu_B-\mu_G)\right]^{1/2}} \tag{5}$$

$$k = 0.5\frac{(\mu_B-\mu_G)'\Sigma^{-1}(\mu_B-\mu_G)}{\left[(\mu_B-\mu_G)'\Sigma^{-1}(\mu_B-\mu_G)\right]^{1/2}} \tag{6}$$

and h is control limit. Now the random variable $a'X_i$ has a univariate normal distribution. Define the non-centrality parameter as:

$$G = \left[(\mu_B - \mu_G)'\Sigma^{-1}(\mu_B - \mu_G)\right]^{1/2} \tag{7}$$

and

$$Z_i = a'(X_i - \mu_G) \tag{8}$$

The CUSUM chart for detecting a shift in the mean vector of a multivariate normal process may be written as:

$$C_i = \max(0, C_{i-1} + a'Z_i - k) > h \tag{9}$$

The function Z_i therefore has a standard univariate normal distribution when X_i has mean equal to μ_G. If the mean shift to μ_B , then $a'(X_i - \mu_G)$ has a univariate normal distribution with mean 0 and variance 1.

2.1 Markov Chain Method for Computing Average Run Length (ARL)

Average Run Length (ARL) is the key index for comparing the performance of statistical control charts. There are three ways to calculate the ARL of control charts: Markov chain model, integral equation, and simulation, Brook and Evans (1972), Champ and Rigdon (1991). In this work, the Markov chain method for evaluating the ARL control chart was used. The main idea of Markov chain method is to approximate monitoring statistics to a Markov chain with finite states, and regard intervals of statistics as states of the Markov Chain, then gain one-step transition probability matrix. For different control charts, differences in Markov chain method lie in one-step transition probability matrix in the form of

$$S_n = \sum_{i=1}^{n}(X_i - k) \tag{10}$$

where S_n is a one-step transition probability matrix, is observed variable, k is reference value. If considered discrete case, that is, X_i, and k are positive integers, then S_n can only have integral values such as $0, 1, 2, , h$, where h is control limit. If $S_n = i$, it is call process in state E_i. If $S_n \geq h$, call it absorbing state. The initial state is assumed to be E_0. After dividing state space, one can easily gain one-step transition probability according to the distribution of observed variable, X.

$$\begin{aligned}
p_{ij} &= P\{S_{n+1} \in E_i | S_n \in E_i\} = P\{S_n + X_{n+1} - k = j | S_n = i\} \\
&= P\{X_{n+1} - k = j - i\}, i \neq h, j \neq h, h \neq 0 \\
p_{i0} &= \{X \leq k - i\}, \\
p_{ih} &= \{X \geq k + h - i\}, \\
p_{hj} &= 0, j = 0, 1 \cdots, h - 1, \\
p_{hh} &= 1,
\end{aligned} \tag{11}$$

When h, k and X, are given, let $p_r = P\{X - k = r\}$, $F_r = P\{X - k \leq r\}$ then onestep transition probability matrix has the following form:

$$P = \begin{bmatrix} F_0 & p_1 & \cdots & p_{k-1} & 1 - F_{k-1} \\ F_{-1} & p_2 & \cdots & p_{k-2} & 1 - F_{k-2} \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ F_{1-k} & p_{2-k} & \cdots & p_0 & 1 - F_0 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix} \tag{12}$$

where P is a square matrix of $h + 1$ dimension, whose last column represents transition probability from state E_i to absorbing state E_h, and last row represents transition probability from absorbing state E_h, to transiting state E_i. Because interest is in its first h rows when calculate ARL, it can be written in the form of partitioned matrix:

$$P = \begin{pmatrix} R & (I - R)1 \\ 0^T & 1 \end{pmatrix} \tag{13}$$

where R is a matrix after deleting the last column and row of P, I is an identity matrix of n, and 1 is a column vector whose elements are all 1. According to properties of Markov Chain, m-step transition probability matrix should be,

$$P_m = P^m = \begin{pmatrix} R^m & (I - R^m)1 \\ 0^T & 1 \end{pmatrix} \tag{14}$$

Let T_i denote the number of steps process need form the first transition to absorbing state E_h when starting at E_i, and $(T = T_0, T_1, \cdots, T(h-1))'$. Then ARL is just the ith component of $E(T)$ when the initial state of statistics is E_i. To calculate those corresponding ARLs, for $r = 1, 2, \cdots$ define:

$$F_r = (P\{T_0 \le r\}, P\{T_1 \le r\}, \cdots, P\{T_{h-1} \le r\}) \tag{15}$$
$$L_r = (P\{T_0 = r\}, P\{T_1 = r\}, \cdots, P\{T_{h-1} = r\}) \tag{16}$$

And according to properties of Markov Chain, one can gain (when $r = 1, 2, \cdots$)

$$\begin{aligned} F_r &= (I - R^r)1 \\ L_r &= RL_{r-1} - R^{r-1}(I - R)1 \end{aligned} \tag{17}$$

According to the definition of expectation,

$$E(T_i) = \sum_{m=1}^{\infty} mP\{T_i = m\} = \sum_{m=1}^{\infty} P\{T_i \ge m\} \tag{18}$$

So that ARL is,

$$ARL = E(T) = (I - R)^{-1}1 \tag{19}$$

Equation (19) reveals that the distributional function of run has a form very similar to geometrical distribution. For the CUSUM control chart, the one-step transition probability,

$$p_{ij} = P\{S_{n+1} \in E_i | S_n \in E_i\} \tag{20}$$

can be calculated approximately as following,

$$p_{i0} = P\{S_n \in I_0 | S_{n-1} = i\omega\} = P\{X_n \le k - i\omega + \omega/2\} \tag{21}$$
$$\begin{aligned} p_{ij} &= P\{S_n \in I_j | S_{n-1} = i\omega\} \\ &= P\{(j-1)\omega - \omega/2 < X_n - k \le (j-1)\omega + \omega/2\},\ 1 \le j \le t-1, \end{aligned} \tag{22}$$
$$p_{it} = P\{S_n \in I_t | S_{n-1} = i\omega\} = P\{X_n - k > (t-i)\omega - \omega/2\} \tag{23}$$

where, $\omega = 2h/(2t-1)$.
Note that, the conditional probability is gained under the assumption that point falling in some interval is set to midpoint value of this interval, then with Equation (19) ARL can be calculated. Obviously, as the number of states t increases, the above method will have a higher accuracy. In usual cases, t is set to 50 or larger. Of course, the number of states depends on the size of control limit of control chart, and as the number of states increases, the time of computing also increases.

The MCUSUM statistic used in this work has the Markov chain property that used a transition matrix of size 100 in our computation but due to space limitation, we provide programs for transition matrix of size 10 in Appendix 1, and assume the mean shift by and the standard deviation shifts by . Using MATLAB Version 7.4.0.287 (R2007a), ARL for the charts were computed. A MATLAB program is also provided for obtaining the MCUSUM charts in Appendix 2. For ARL in-control, the Markov chain is approximated as follows:

$$p_{ij} = P\{j\omega + \frac{\omega}{2} + k, m, s\} - P\{j\omega - \frac{\omega}{2} + k, m, s\}, 0 < j \le t \tag{24}$$

$$p_{i0} = P\{\frac{\omega}{2} + k, m, s\}, j = 0 \tag{25}$$

$$P = \begin{bmatrix} F_0 & p_1 & \cdots & p_{t-1} \\ F_{-1} & p_2 & \cdots & p_{t-1} \\ \cdots & \cdots & \cdots & \cdots \\ F_{1-t} & p_{2-t} & \cdots & p_0 \end{bmatrix} \tag{26}$$

Then ARL can be calculated as

$$ARL = X'(1-P)^{-1}X \tag{27}$$

where $X = [1, 0, \cdots, 0]^T$

The ARL depends on the standard univariate CUSUM chart parameters; the reference value k and decision interval h, together with the multivariate non-centrality parameter D defined in Equation (7). At a given value of h, the decision interval, k is calculated for the given level of shift in the mean vector and/or covariance matrix intended to be quickly detected and use the (h, k) combination to calculate the ARL. To guard against a particular shift in the mean vector and/or covariance matrix, the reference value is computed as $k = 0.5D$. The shift in the covariance matrix is denoted by b as specified earlier and the shift in the mean vector is denoted by δ . Here the mean vector shifts is assume by the same amount for all components in a unit and that all variances shift proportionally, therefore the correlations between variables remain the same.

2.2 Charting Procedures

The charting procedure of a Max-MCUSUM chart is similar to that of the Max-CUSUM chart. If a point plots below the decision interval, the process is said to be in statistical control. An out-of-control signal is issued if any point plots above the decision interval. The following procedure is used to build the CUSUM chart:

1. Specify the following parameters; p and the in-control or target value of the mean vector μ_G the bad mean vector μ_B and the in-control or target value of the covariance matrix Σ.

2. If μ_B is not known, use the sample mean vector X which is a p-dimensional vector of sample means. In the same manner, if the population covariance matrix is unknown, we use the sample covariance matrix S to estimate the population covariance matrix.

3. For each sample, compute Z_i.

4. To detect specified changes in the process mean vector and covariance matrix, choose an optimal (h, k) combination and calculate (i) the cumulative sums, C_i^+, C_i^- (ii) If $C_i \geq h$ plot $C+$. This shows a shift in the process mean vector.

5. Investigate the cause(s) of the shift for each out-of-control point in the chart and carry out the remedial measures needed to bring the process back into an in-control state.

3 Application of Multivariate CUSUM Control Chart

The multivariate cumulative sum control chart proposed by Smiley and Keoagile (2005) is applied to real data obtained from Smiley and Keoagile (2005), Alves (2009) and Oyeyemi (2011) to generate the graph for process control affected by two, three and four quality characteristics respectively.

Case with P = 2: The data for two-quality characteristics Smiley and Keoagile (2005) were gotten from a steel manufacturing process that measured the Brinnel hardness (X) and the tensile strength (Y) for 30 samples which are divided into six subgroups shown in Table 1. The target value of the mean vector is given as:

$$\mu_o = \begin{bmatrix} 175 \\ 55 \end{bmatrix}$$

It is assumed that the covariance matrix does not shift, therefore, the covariance matrix, $S = \begin{bmatrix} 332.13 & 69.26 \\ 69.26 & 29.97 \end{bmatrix}$ and mean vector, $\bar{\bar{X}} = \begin{bmatrix} 174.67 \\ 51.67 \end{bmatrix}$ used to monitor the process were estimated from the sample. To construct the MCUSUM chart for detecting a shift in the process mean vector, estimate μ_{circ} with $\bar{\bar{X}}$ and Σ with S. Substituting these estimators into Equation (7) gives D = 0.827 and the reference value $k = 0.5 * D = 0.414$. A MATLAB program for computing and H for the calculated reference value k is shown in Appendix 1. Also a MATLAB program for computing the cumulative sum control chart is where R is a matrix after deleting the last column and row of P, I is an identity matrix of n, and 1 is a column vector whose elements are all 1. According to properties of Markov Chain, m-step transition probability matrix should be,

$$P_m = P^m = \begin{pmatrix} R^m & (I - R^m)1 \\ 0^T & 1 \end{pmatrix} \tag{14}$$

Let T_i denote the number of steps process need form the first transition to absorbing state E_h when starting at E_i, and $(T = T_0, T_1, \cdots, T_(h-1))'$. Then ARL is just the ith component of $E(T)$ when the initial state of statistics is E_i. To calculate those corresponding ARLs, for $r = 1, 2, \cdots$ define:

$$F_r = (P\{T_0 \le r\}, P\{T_1 \le r\}, \cdots, P\{T_{h-1} \le r\}) \tag{15}$$

$$L_r = (P\{T_0 = r\}, P\{T_1 = r\}, \cdots, P\{T_{h-1} = r\}) \tag{16}$$

And according to properties of Markov Chain, one can gain (when $r = 1, 2, \cdots$)

$$\begin{aligned} F_r &= (I - R^r)1 \\ L_r &= RL_{r-1} - R^{r-1}(I - R)1 \end{aligned} \tag{17}$$

According to the definition of expectation,

$$E(T_i) = \sum_{m=1}^{\infty} mP\{T_i = m\} = \sum_{m=1}^{\infty} P\{T_i \ge m\} \tag{18}$$

So that ARL is,

$$ARL = E(T) = (I - R)^{-1}1 \tag{19}$$

Equation (19) reveals that the distributional function of run has a form very similar to geometrical distribution. For the CUSUM control chart, the one-step transition probability,

$$p_{ij} = P\{S_{n+1} \in E_i | S_n \in E_i\} \tag{20}$$

can be calculated approximately as following,

$$p_{i0} = P\{S_n \in I_0 | S_{n-1} = i\omega\} = P\{X_n \le k - i\omega + \omega/2\} \tag{21}$$

$$\begin{aligned} p_{ij} &= P\{S_n \in I_j | S_{n-1} = i\omega\} \\ &= P\{(j-1)\omega - \omega/2 < X_n - k \le (j-1)\omega + \omega/2\},\ 1 \le j \le t-1, \end{aligned} \tag{22}$$

$$p_{it} = P\{S_n \in I_t | S_{n-1} = i\omega\} = P\{X_n - k > (t - i)\omega - \omega/2\} \tag{23}$$

where, $\omega = 2h/(2t-1)$.

Note that, the conditional probability is gained under the assumption that point falling in some interval is set to midpoint value of this interval, then with Equation (19) ARL can be calculated. Obviously, as the number of states t increases, the above method will have a higher accuracy. In usual cases, t is set to 50 or larger. Of course, the number of states depends on the size of control limit of control chart, and as the number of states increases, the time of computing also increases.

The MCUSUM statistic used in this work has the Markov chain property that used a transition matrix of size 100 in our computation but due to space limitation, we provide programs for transition matrix of size 10 in Appendix 1, and assume the mean shift by and the standard deviation shifts by . Using MATLAB Version 7.4.0.287 (R2007a), ARL for the charts were computed. A MATLAB program is also provided for obtaining the MCUSUM charts in Appendix 2. For ARL in-control, the Markov chain is approximated as follows:

$$p_{ij} = P\{j\omega + \frac{\omega}{2} + k, m, s\} - P\{j\omega - \frac{\omega}{2} + k, m, s\}, 0 < j \le t \tag{24}$$

$$p_{i0} = P\{\frac{\omega}{2} + k, m, s\}, j = 0 \tag{25}$$

$$P = \begin{bmatrix} F_0 & p_1 & \cdots & p_{t-1} \\ F_{-1} & p_2 & \cdots & p_{t-1} \\ \cdots & \cdots & \cdots & \cdots \\ F_{1-t} & p_{2-t} & \cdots & p_0 \end{bmatrix} \tag{26}$$

Then ARL can be calculated as

$$ARL = X'(1 - P)^{-1}X \quad (27)$$

where $X = [1, 0, \cdots, 0]^T$

The ARL depends on the standard univariate CUSUM chart parameters; the reference value k and decision interval h, together with the multivariate non-centrality parameter D defined in Equation (7). At a given value of h, the decision interval, k is calculated for the given level of shift in the mean vector and/or covariance matrix intended to be quickly detected and use the (h, k) combination to calculate the ARL. To guard against a particular shift in the mean vector and/or covariance matrix, the reference value is computed as $k = 0.5D$. The shift in the covariance matrix is denoted by b as specified earlier and the shift in the mean vector is denoted by δ . Here the mean vector shifts is assume by the same amount for all components in a unit and that all variances shift proportionally, therefore the correlations between variables remain the same.

2.2 Charting Procedures

The charting procedure of a Max-MCUSUM chart is similar to that of the Max-CUSUM chart. If a point plots below the decision interval, the process is said to be in statistical control. An out-of-control signal is issued if any point plots above the decision interval. The following procedure is used to build the CUSUM chart:

1. Specify the following parameters; p and the in-control or target value of the mean vector μ_G the bad mean vector μ_B and the in-control or target value of the covariance matrix Σ.
2. If μ_B is not known, use the sample mean vector X which is a p-dimensional vector of sample means. In the same manner, if the population covariance matrix is unknown, we use the sample covariance matrix S to estimate the population covariance matrix.
3. For each sample, compute Z_i.
4. To detect specified changes in the process mean vector and covariance matrix, choose an optimal (h, k) combination and calculate (i) the cumulative sums, C_i^+, C_i^- (ii) If $C_i \geq h$ plot $C+$. This shows a shift in the process mean vector.
5. Investigate the cause(s) of the shift for each out-of-control point in the chart and carry out the remedial measures needed to bring the process back into an in-control state.

3 Application of Multivariate CUSUM Control Chart

The multivariate cumulative sum control chart proposed by Smiley and Keoagile (2005) is applied to real data obtained from Smiley and Keoagile (2005), Alves (2009) and Oyeyemi (2011) to generate the graph for process control affected by two, three and four quality characteristics respectively.

Case with P = 2: The data for two-quality characteristics Smiley and Keoagile (2005) were gotten from a steel manufacturing process that measured the Brinnel hardness (X) and the tensile strength (Y) for 30 samples which are divided into six subgroups shown in Table 1. The target value of the mean vector is given as:

$$\mu_o = \begin{bmatrix} 175 \\ 55 \end{bmatrix}$$

It is assumed that the covariance matrix does not shift, therefore, the covariance matrix, $S = \begin{bmatrix} 332.13 & 69.26 \\ 69.26 & 29.97 \end{bmatrix}$ and mean vector, $\bar{\bar{X}} = \begin{bmatrix} 174.67 \\ 51.67 \end{bmatrix}$ used to monitor the process were estimated from the sample. To construct the MCUSUM chart for detecting a shift in the process mean vector, estimate μ_{circ} with $\bar{\bar{X}}$ and Σ with S. Substituting these estimators into Equation (7) gives D = 0.827 and the reference value $k = 0.5 * D = 0.414$. A MATLAB program for computing and H for the calculated reference value k is shown in Appendix 1. Also a MATLAB program for computing the cumulative sum control chart is shown in Appendix 2. Fig.1 shows Multivariate Cumulative control chart that simultaneously monitor the quality characteristics X (Brinnel hardness) and Y (tensile strength) of steel manufacturing process.

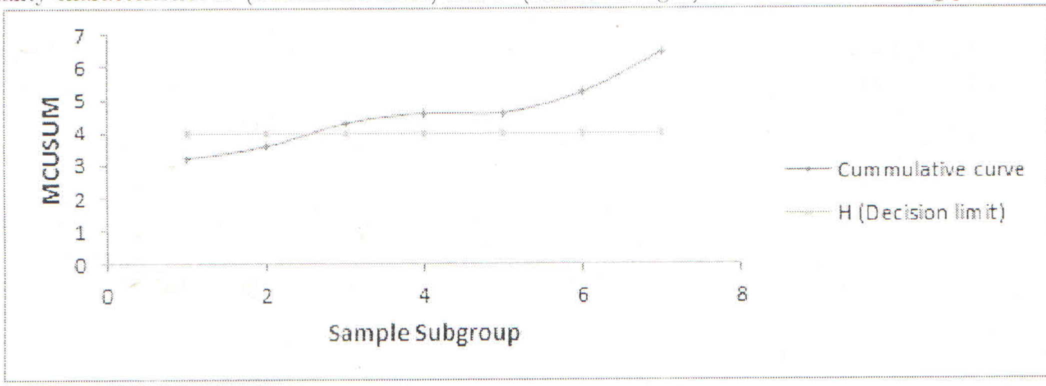

Fig.1. MCUSUM chart generated from Brinnel hardness and tensile strength data.

Case with P = 3: The data for three-quality characteristics are related to machining process of engine block Alves (2009). In this process, the quality characteristics to monitor simultaneously are X_1: positional hole for the X coordinate; Y_1 : positional hole for the Y coordinate; and : distance between the centers of holes 1 and 2 for 31 samples shown in Table 2. The in-control value of the mean vector is given as:

$$\mu_o = \begin{bmatrix} 5 \\ 103.25 \\ 194.27 \end{bmatrix}$$

It is assumed that the covariance matrix does not shift, therefore, the covariance matrix, $S = \begin{bmatrix} 0.00030928 & -0.000128501 & -0.000020346 \\ -0.000128501 & 0.000169902 & 0.000028765 \\ -0.000020346 & 0.000028765 & 0.000008291 \end{bmatrix}$ and mean vector, $\bar{\bar{X}} = \begin{bmatrix} 5.01 \\ 103.26 \\ 194.27 \end{bmatrix}$ used to monitor the process were calculated from the sample. To construct the MCUSUM chart for detecting a shift in the process mean vector, we estimate μ_1 with $\bar{\bar{X}}$ and Σ with S. Substituting these estimators into Equation (7) gives $D = 1.5301$, and the reference value $k = 0.5 * D = 0.7651$. A MATLAB program for computing ARL_o and H for the calculated reference value k is shown in Appendix 1. Also a MATLAB program for computing the cumulative sum control chart is shown in Appendix 2. Fig.2 shows Multivariate Cumulative control chart that simultaneously monitor the quality characteristics X_1, Y_1 and D_{12} of the machining process.

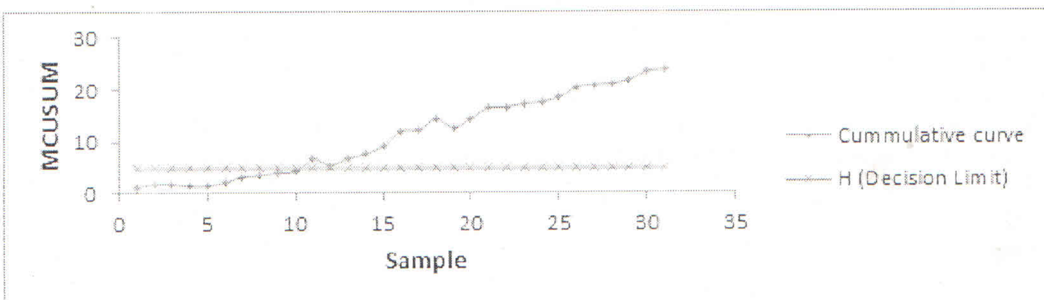

Fig.2. MCUSUM chart generated from the process data of engine block.

Case with p = 4: The data for four-quality characteristics are from a detergent manufacturing company Oyeyemi (2011). Table 3, shows the four quality characteristics that affect the production process for 35 samples. They include X_1 (active detergent), X_2 (moisture content), X_3 (bulk density) and X_4 (ph level). The target value of the mean vector is given as:

$$\mu_o = \begin{bmatrix} 24.45 \\ 3.5 \\ 305 \\ 10.5 \end{bmatrix}$$

It is assumed that the covariance matrix does not shift, therefore, the covariance matrix, $S = \begin{bmatrix} 1.3021 & -0.1217 & 1.5289 & -0.0222 \\ -0.1217 & 0.1066 & -1.5065 & -0.0418 \\ 1.5289 & -1.5065 & 147.4957 & 0.0783 \\ -0.0222 & -0.0418 & 0.0783 & 0.0865 \end{bmatrix}$ and mean vector, $\bar{\bar{X}} = \begin{bmatrix} 23.3377 \\ 3.4197 \\ 309.4537 \\ 10.4520 \end{bmatrix}$ used to monitor the process were calculated from the sample.

To construct the MCUSUM chart for detecting a shift in the process mean vector, we estimate μ_1 with and Σ with S. Substituting these estimators into Equation (7) gives $D = 1.2984$. The reference value $k = 0.5 * D = 0.6492$. A MATLAB program for computing and H for the calculated reference value k is shown in Appendix 1. Also a MATLAB program for computing the cumulative sum control chart is shown in Appendix 2. Fig.3 shows Multivariate Cumulative control chart that simultaneously monitor the quality characteristics X_1, X_2, X_3 and X_4 of the production process.

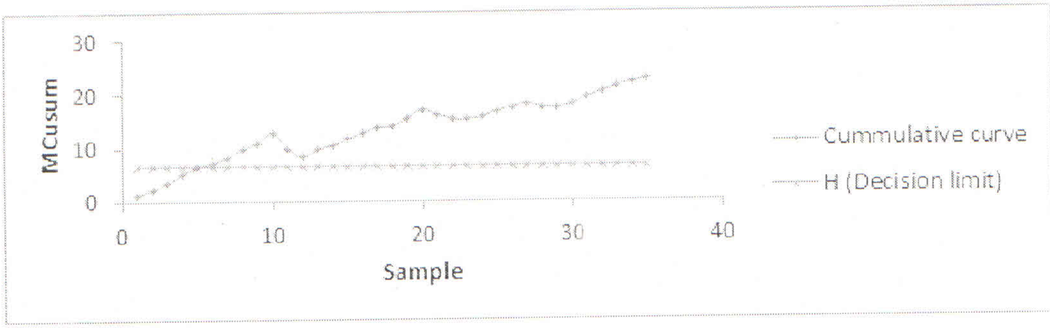

Fig.3. MCUSUM chart generated from detergent manufacturing data.

3.1 Discussion of Results

Case with p = 2: The in-control Average Run Length is obtained as $ARL_o = 250$, i.e. false alarm rate $\alpha = 0.05$, and the decision limit $H = 3.97$ by Markov Chain procedure. It is observed that the MCUSUM chart signaled a shift in the mean vector at the 3rd sample. The cause of out-of control in the manufacturing process can either be quality characteristic X or Y. Therefore, readjustment of any the parameters responsible should be carried out to keep the process continuously in-control. **Case with p = 3**: In-control Average Run Length is obtained as $ARL_{circ} = 200$, i.e. false alarm rate $\alpha = 0.05$ and the decision limit $H = 4.86$ by Markov Chain procedure. It is observed that the MCUSUM chart signaled a shift in the mean vector at the 11th sample. The cause of out-of control in the process can either be quality characteristic $X1$ or $Y1$ or $D12$. Therefore, readjustment of any the parameters responsible for the out-of -control should be carried out to keep the process continuously in-control. **Case with p = 4**: In-control Average Run Length is obtained $ARL_o =$, i.e. false alarm rate $\alpha = 0.05$ and the decision limit $H = 6.64$ by Markov Chain procedure. It is observed that the MCUSUM chart signaled a change in the mean vector at the 6th sample. The cause of out-of control in the process can either be quality characteristic X_1 or X_2 or X_3 or X_4. Therefore, readjustment of any the parameters responsible for the out-of -control should be carried out to keep the process continuously in-control.

4 Conclusion

Control charts are valuable tools for detecting an out-of-control process. The effective simultaneous monitoring of many quality characteristics of a production process often depends on statistical tool that are becoming more and more specific. In modern industries, there is a need to track more than one quality characteristics. If separate univariate charts are used, the overall probability of false alarm may be inflated since correlation between variables is ignored. In such cases, multivariate cumulative control charts should be considered. This work will enable the engineers and quality control management to develop a scheme that can easily identify the sample point at which a process signals an out-of-control. Corrective measures and adjustment can be easily implemented since the cause of variability can be identified.

Table 1: : Brinnel Hardness and Tensile Strength Data

	Subgroup					
Samples	1	2	3	4	5	6
	1 143	6 178	11 175	16 182	21 160	26 195
	34.2	51.5	57.3	57.2	45.5	58
	2 200	7 162	12 187	17 177	22 183	27 134
	57	45.9	58.5	50.6	53.6	45.7
	3 160	8 215	13 187	18 204	23 179	28 187
	47.5	59.1	58.2	55.1	51.2	42
	4 181	9 161	14 186	19 178	24 194	29 135
	53.4	48.4	57	50.9	57.7	40.5
	5 148	10 141	15 172	20 196	25 181	30 159
	47.8	47.3	49.4	57.9	55.6	58

Source: Smiley and Keoagile (2005)

Table 2: : Machining Process Data

Sample	X_1	Y_1	D_{12}	Sample	X_1	Y_1	D_{12}
1	5.019	103.25	194.274	17	4.998	103.259	194.274
2	5.01	103.257	194.274	18	5.033	103.27	194.274
3	5.007	103.241	194.273	19	5.022	103.245	194.26
4	5.009	103.233	194.271	20	5.019	103.265	194.276
5	5.02	103.239	194.27	21	5.002	103.281	194.279
6	5.018	103.247	194.273	22	5.01	103.26	194.271
7	5.019	103.256	194.273	23	4.992	103.277	194.276
8	5.019	103.244	194.271	24	4.996	103.269	194.274
9	5.019	103.243	194.272	25	5.004	103.27	194.275
10	5.024	103.241	194.27	26	5.04	103.245	194.273
11	5.03	103.254	194.277	27	4.994	103.275	194.275
12	4.977	103.258	194.273	28	4.968	103.275	194.28
13	5.014	103.266	194.275	29	5.026	103.252	194.27
14	5.014	103.258	194.273	30	4.993	103.291	194.279
15	5.01	103.258	194.27	31	5.002	103.248	194.275
16	5.042	103.252	194.277				

Source: Alves (2009)

Table 3: : Detergent Production Process Data

Sample	Detergent	Moisture	Density	Ph Level	Sample	Detergent	Moisture	Density	Ph Level
1	22.23	3.65	315.45	10.47	19	23.16	2.97	318.79	10.63
2	22.27	3.9	295.05	10.03	20	23.31	2.96	319	10.6
3	22.34	3.67	289.64	10.24	21	25.41	3.22	305.6	10.43
4	21.49	3.53	314.67	10.7	22	25.21	3.35	306.8	10.48
5	22.26	3.43	325.58	10.7	23	24.69	3.24	303.08	10.29
6	23.23	3.46	334.35	10.66	24	23.99	3.29	303.97	10.35
7	23.58	3.22	322.86	10.41	25	23.61	3.16	301.34	10.29
8	23.44	3.12	324.2	10.41	26	23.57	3.28	301.16	10.4
9	23.35	3.24	329.04	10.43	27	23.38	3.57	306.78	10.25
10	23.73	3.01	330.92	10.16	28	24.06	3.92	297.73	10.1
11	25.81	3.75	300.94	10.8	29	24.01	3.49	308.99	10.47
12	25.27	3.28	305.63	10.86	30	23.81	3.47	336.54	10.32
13	22.53	3.2	300.87	10.72	31	22.59	3.57	309.78	10.32
14	22.05	3.23	300.15	11.21	32	20.44	4.71	291.88	10.01
15	22.34	3.16	299.69	10.94	33	23.51	3.54	312.13	10.05
16	22.12	3.32	301.42	10.94	34	23.67	3.5	307.6	10.06
17	22.7	3.23	296.7	10.61	35	24.01	3.57	310.55	10.01
18	23.65	3.48	302	10.47					

Source: Oyeyemi (2011)

References

Alves, C.C. (2009). The Method of integral Equation with Gaussian Quadrature to optmize the parameters of the Multivariate sum control chart. Doctorate Thesis, UFSC, Florianopolis.

Bersimis, S., Psarakis, S. and Panaretos, J. (2007). Multivariate statistical process control charts: an overview. *Quality and Reliability Engineering International*, Vol. 23, Issue 5, pg. 517543.

Brook, D. and Evans, D. A. (1972). An Approach to the Probability Distribution of CUSUM Run Length. *Biometrilca*, Vol.59, pg. 539-549.

Champ, C. W. and Rigdon, S. E. (1991). A Comparison of the Markov Chain and the Integral Equation Approaches for Evaluating the Run Length Distribution of Quality Control Charts. *Communications in Statistics-Simulation*, Vol. 20, pg. 191-204.

Crosier, R. B. (1988). Multivariate Generalizations of Cumulative Sum Quality-Control Schemes. *Technometrics*, Vol. 30, pg. 291-303.

Crowder, S. V. (1989). Design of Exponentially Weighted Moving Average Scheme. *Journal of Quality Technology*, Vol. 21, pg 155-162.

Hawkins, D. M. (1991). A Multivariate quality control based on regression-adjustment variables. *Technometrics*, Vol. 33, pg. 61-75.

Healy, J. D. (1987). A Note on Multivariate CUSUM Procedures. *Technometrics*, Vol. 29, pg. 409-412.

Hotelling, H. (1947). Multivariate Quality Control Techniques of Statistical Analysis, McGraw-Hill, New York, pg. 113-184.

Lowry, C. A., Woodall, W. H., Champ, C. W. and Rigdon, S. E. (1992). A Multivariate exponential weighted moving average control chart. *Technometrics*, Vol. 34, pg. 46 53.

Lucas, J. M. and Crosier, R. B. (2000). Fast Initial Response for CUSUM Quality Control Scheme: Give Your CUSUM a Head Start. *Technometrics*, Vol. 42, pg. 702-707.

Marquardt, D. W. (1984). New Technical and Educational Directions for Managing Product Quality. *The American Statistician*, Vol. 38, pg. 8-14.

Montgomery, D. C. (2009). Introduction to Statistical Quality Control 6th edition. John Wiley and Sons, Inc., New York.

Oyeyemi, G. M. (2011). Principal Component Chart for Multivariate Statistical Process Control. *The Online Journal of Science and Technology (TOJSST)*, Vol. 1 Issue 2, pg. 22-31.

Page, E. S. (1954). Continuous Inspection Schemes. *Biometrika*, Vol. 4, pg. 100-115.

Pignatiello, J.J. and Runger, G.C. (1990). Comparisons of Multivariate CUSUM Charts. *Journal of Quality Technology*, Vol. 22 Issue 3 pg.173-186.

Runger, G. C., Willemain, T. R. and Prabhu, S. (1995). Average Run Lengths for CUSUM Control Charts Applied to Residuals. *Communications in Statistics-Theory and Methods*, Vol. 24, pg. 273-282.

Smiley, W.C. and Keoagile, T. (2005). Multivariate Max-CUSUM Chart. *Quality Technology & Quantitative Management*, Vol. 2 Issue 2, pg. 221-235.

Yu, J. and Xi, L. (2009). A Neural Network Ensemble-based Model for On-Line Monitoring and Diagnosis of Out-of-Control Signals in Multivariate Manufacturing Processes. *Expert Systems with Applications*, Vol. 369 Issue 1, pg. 909921.

Appendix 1

$$h = 3.392;$$
$$t = 10;$$
$$a = 0;$$
$$d = 1;$$
$$b = 1;$$
$$m = a * d;$$
$$s = b * d;$$
$$k = 0.5; \%a * d/2$$
$$w = 2 * h/(2 * t - 1);$$

$$f0 = normcdf(w/2 + k, m, s);$$
$$f1 = normcdf(-w + w/2 + k, m, s);$$
$$f2 = normcdf(-2 * w + w/2 + k, m, s);$$
$$f3 = normcdf(-3 * w + w/2 + k, m, s);$$
$$f4 = normcdf(-4 * w + w/2 + k, m, s);$$
$$f5 = normcdf(-5 * w + w/2 + k, m, s);$$
$$f6 = normcdf(-6 * w + w/2 + k, m, s);$$
$$f7 = normcdf(-7 * w + w/2 + k, m, s);$$
$$f8 = normcdf(-8 * w + w/2 + k, m, s);$$
$$f9 = normcdf(-9 * w + w/2 + k, m, s);$$
$$p0 = normcdf(w/2 + k, m, s) - normcdf(-w/2 + k, m, s);$$
$$p1 = normcdf(w + w/2 + k, m, s) - normcdf(w - w/2 + k, m, s);$$
$$p2 = normcdf(2 * w + w/2 + k, m, s) - normcdf(2 * w - w/2 + k, m, s);$$
$$p3 = normcdf(3 * w + w/2 + k, m, s) - normcdf(3 * w - w/2 + k, m, s);$$
$$p4 = normcdf(4 * w + w/2 + k, m, s) - normcdf(4 * w - w/2 + k, m, s);$$
$$p5 = normcdf(5 * w + w/2 + k, m, s) - normcdf(5 * w - w/2 + k, m, s);$$
$$p6 = normcdf(6 * w + w/2 + k, m, s) - normcdf(6 * w - w/2 + k, m, s);$$
$$p7 = normcdf(7 * w + w/2 + k, m, s) - normcdf(7 * w - w/2 + k, m, s);$$
$$p8 = normcdf(8 * w + w/2 + k, m, s) - normcdf(8 * w - w/2 + k, m, s);$$
$$p9 = normcdf(9 * w + w/2 + k, m, s) - normcdf(9 * w - w/2 + k, m, s);$$
$$pm1 = normcdf(-w + w/2 + k, m, s) - normcdf(-w - w/2 + k, m, s);$$
$$pm2 = normcdf(-2 * w + w/2 + k, m, s) - normcdf(-2 * w - w/2 + k, m, s);$$
$$pm3 = normcdf(-3 * w + w/2 + k, m, s) - normcdf(-3 * w - w/2 + k, m, s);$$
$$pm4 = normcdf(-4 * w + w/2 + k, m, s) - normcdf(-4 * w - w/2 + k, m, s);$$
$$pm5 = normcdf(-5 * w + w/2 + k, m, s) - normcdf(-5 * w - w/2 + k, m, s);$$
$$pm6 = normcdf(-6 * w + w/2 + k, m, s) - normcdf(-6 * w - w/2 + k, m, s);$$
$$pm7 = normcdf(-7 * w + w/2 + k, m, s) - normcdf(-7 * w - w/2 + k, m, s);$$
$$pm8 = normcdf(-3 * w + w/2 + k, m, s) - normcdf(-8 * w - w/2 + k, m, s);$$

$$
\begin{aligned}
P = [&f0p1p2p3p4p5p6p7p8p9\\
&f1p0p1p2p3p4p5p6p7p8\\
&f2pm1p0p1p2p3p4p5p6p7\\
&f3pm2pm1p0p1p2p3p4p5p6\\
&f4pm3pm2pm1p0p1p2p3p4p5\\
&f5pm4pm3pm2pm1p0p1p2p3p4\\
&f6pm5pm4pm3pm2pm1p0p1p2p3\\
&f7pm6pm5pm4pm3pm2pm1p0p1p2\\
&f8pm7pm6pm5pm4pm3pm2pm1p0p1\\
&f9pm8pm7pm6pm5pm4pm3pm2pm1p0];
\end{aligned}
$$

$$
\begin{aligned}
A &= eye(10, 10);\\
B &= [1; 1; 1; 1; 1; 1; 1; 1; 1; 1];\\
X &= [1\,0\,0\,0\,0\,0\,0\,0\,0\,0];\\
ARL &= X * inv(A - P) * B
\end{aligned}
$$

Appendix 2

$$
\begin{aligned}
A &= input(A);\\
B &= input(B);\\
C &= cov(A, B);\\
u &= [175, 55]';\\
s &= [332.1369.26; 69.2629.97];\\
x &= [174.67, 51.67]';\\
D &= sqrt((x - u)' * inv(s) * (x - u));\\
k &= 0.5 * D;\\
a1 &= (x - u)' * inv(s)/D; \%aprime\\
k2 &= 2 + 0.5 * D^2;\\
n &= 30;\\
C1 &= 0;\\
C2 &= 0;\\
S1 &= 0;\\
fori &= 1 : n\\
Z &= a1 * ([A(i)B(i)]' - u);\\
C1 &= max(0, C1 + Z - 0.5 * D);\\
C2 &= max(0, C2 - Z - 0.5 * D);\\
&end
\end{aligned}
$$