Comparisons of Multivariate *CUSUM* Charts

JOSEPH J. PIGNATIELLO, JR.

The University of Iowa, Iowa City, Iowa 52242

GEORGE C. RUNGER

International Business Machines, Inc., Tucson, Arizona 85744

We consider several distinct approaches for controlling the mean of a multivariate normal process including two new and distinct multivariate *CUSUM* charts, several multiple univariate *CUSUM* charts, and a Shewhart χ^2 control chart. The performances of these charts are compared by estimating the average run lengths. A Markov chain is used to evaluate the average run length performance of one of the charts while Monte Carlo simulation is used to evaluate the other multivariate schemes. The *ARL* performance of the multiple univariate scheme is shown to be dependent upon the manner in which the process mean shifts whereas one of the multivariate *CUSUM* charts provides stable ARL performance over a diverse set of off-target conditions. The average run length data are presented.

Introduction

CUMULATIVE sum (*CUSUM*) charts are often used instead of standard Shewhart charts when detection of small changes in a process parameter is important. For comparable average run lengths (*ARLs*) when the process is on-target, *CUSUM* charts can be designed to give shorter *ARLs* than Shewhart charts for detecting certain small changes in process parameters. As demonstrated by Champ and Woodall (1987), the superiority of the *CUSUM* chart over the Shewhart chart also holds when the Shewhart chart is augmented with runs rules. Thus, it is only natural to investigate whether the shorter *ARL*'s for the univariate case can be extended to the multivariate case.

We note that our concern in this paper is with monitoring the mean of a multivariate normal process. We use the term "on-target" to indicate that the process is in-control with respect to its mean. Likewise, we use the term "off-target" to indicate that the mean of the multivariate normal process has shifted. We will assume that the process dispersion is stable and is not prone to changes. Since the phrase "out of control" can include more than a shift in the mean, we prefer the "off-target" term for our purposes.

In this study we use the Markov chain approach described in Brook and Evans (1972) and a Monte Carlo simulation to compare the performance of several schemes for monitoring a multivariate normal process. Two multivariate cumulative sum schemes are introduced. We refer to these multivariate *CUSUM* charts as multivariate *CUSUM* #1 (*MC*1) and multivariate *CUSUM* #2 (*MC*2). Their average run lengths will be compared to the average run lengths of multiple univariate *CUSUM* charts as developed by Woodall and Ncube (1985) and multivariate Shewhart χ^2 charts.

Naturally, multivariate processes and procedures are more complicated than univariate ones. Due to the vector and matrix algebra that is inevitably required, computerized implementations of the multivariate procedures are almost always needed. Although we will compare the *ARL* performances of several multivariate procedures here, we leave it for the reader to assess the potential trade-offs between the increase in mathematics and computer savvy that is

Dr. Pignatiello is an Associate Professor. He holds appointments in the Department of Industrial and Management Engineering and the Department of Statistics. He is a Member of ASQC.

Dr. Runger is a Senior Engineer at the IBM General Products Division. He can be reached at 5723 N. Via Umbrosa, Tucson, AZ 85715. He is a Member of ASQC.

inevitably required and the *ARL* improvements that can result with the multivariate procedures.

The multivariate *CUSUM* chart proposed by Crosier (1988) is similar to our *MC*1 chart, yet distinctly different. Crosier's multivariate *CUSUM* chart is somewhat more complicated than our *MC*1 chart, but it does have a similar *ARL* performance. Crosier also presents results on another chart that is almost equivalent to *MC*2 and finds (as we do) that its *ARL* performance is poor. A multivariate *CUSUM* approach discussed by Alwan (1986) has similarities to the familiar univariate *CUSUM V*-mask procedure. We shall not discuss these procedures further here, however.

The following sections contain a review of some control chart procedures for monitoring a multivariate normal process. This is followed by a discussion of "directional invariance," the generation and interpretation of signals from a control scheme and the development of the two multivariate *CUSUM* control chart schemes. In a subsequent section, we discuss the calibration of the various schemes so that a fair *ARL* comparison can be made. We then present the simulation results and consider a modified multiple univariate *CUSUM* scheme.

Standard Control Chart Methods for Multivariate Normal Processes

In this section we briefly review the control chart procedures that were investigated and our modeling assumptions. We also describe the development of two proposed multivariate *CUSUM* charts.

For successive samples, multivariate control chart techniques used for controlling the mean of a multivariate normal process can be interpreted as repeated tests of significance of the form

$$H_0:\quad \mu = \mu_0 \qquad\qquad (1)$$
$$H_1:\quad \mu \neq \mu_0$$

where μ represents a multivariate normal process mean whose true value is unknown and μ_0 is the target value for the parameter. For simplicity, we will assume that $\mu_0 = 0$, although the more general case can be handled easily by translation.

We let $\mathbf{X}_t = (x_{1,t}, x_{2,t}, \cdots, x_{p,t})'$ denote the $p \times 1$ vector of quality characteristic measurements made on a part from a multivariate normal process where $x_{j,t}$ is the observation on variate j at time t. We assume that the successive $\mathbf{X}_t$ are independent and identically distributed multivariate normal random vectors with known and constant covariance matrix Σ. That is, the $\mathbf{X}_t$ are *iid* $N_p(\mu, \Sigma)$. Without loss of generality, we will let $\mathbf{X}_t$ denote the (sample) mean vector at time t and we will let Σ denote its covariance matrix.

The Shewhart χ^2 Chart

To test the hypothesis in (1), it is well known that the null hypothesis should be rejected at time t if $\chi^2 > \chi^2_{p;\alpha}$ where

$$\chi^2 = (\mathbf{X}_t - \mu_0)'\Sigma^{-1}(\mathbf{X}_t - \mu_0) \qquad (2)$$

and $\chi^2_{p;\alpha}$ is the upper 100α percentage point of the χ^2 distribution with p degrees of freedom. The noncentrality parameter associated with χ^2 is

$$\lambda^2(\mu) = (\mu - \mu_0)'\Sigma^{-1}(\mu - \mu_0). \qquad (3)$$

We note that $\lambda(\mu)$, the square root of the noncentrality parameter, is often used to represent a measure of the distance of μ from μ_0. This measure of distance is also called the "Mahalanobis distance" by Morrison (1976, page 237) and the "statistical distance" by Johnson and Wichern (1988). We note that the "straight line" or "Euclidean" distance assumes an identity covariance matrix instead. We will henceforth use the word "distance" to mean the square root of the noncentrality parameter in (3).

A χ^2 control chart operates by plotting χ^2 on a chart with an appropriate *UCL*. If a point plots above the upper control limit, the process mean is deemed to be out of control and the assignable causes of the variation are sought. The average run length (*ARL*) of this control scheme can be calculated as $1/P$ where P denotes the probability that χ^2 exceeds the *UCL*. The on-target value of P is determined from the probability that χ^2 exceeds the *UCL* under the (central) χ^2_p distribution while the off-target value of P is the probability that χ^2 exceeds the *UCL* under the noncentral χ^2_p distribution. These *ARL* results are provided in Table 1 for selected values of $\lambda(\mu)$.

The Multiple Univariate *CUSUM* Scheme

Woodall and Ncube (1985) describe how a p-dimensional multivariate normal process can be monitored by using p two-sided univariate *CUSUM* charts. The j^{th} two-sided univariate *CUSUM* is operated by forming the cumulative sums

$$S_{j,t} = \max(0, S_{j,t-1} + x_{jt} - k_j)$$
$$T_{j,t} = \min(0, T_{j,t-1} + x_{jt} + k_j)$$

where $S_{j,0} \geq 0$, $T_{j,0} \leq 0$, $k_j > 0$, and x_{jt} is the sample mean at time t for variate j.

The j^{th} two-sided chart signals that the corresponding process mean has shifted when either $S_{j,t} > h_j$ or $T_{j,t} < -h_j$ for some *CUSUM* control chart parameters k_j

and h_j. The multiple univariate *CUSUM* scheme signals an off-target condition when *any* of the p two-sided schemes produces an off-target signal. Therefore, the on-target average run length of the multiple univariate scheme is *less* than the average run length of any one of the univariate *CUSUM* charts.

Directional Invariance

It is well known that the χ^2 chart (as well as the Hotelling T^2 chart discussed in Kramer and Jensen [1969a, 1969b] and Jackson [1959, 1981]) is *directionally invariant*. That is, the *ARL* performance of the χ^2 chart is determined solely by the *distance* of the off-target mean from the on-target mean and not by the particular *direction* (or, location) of that mean [where "distance" is defined as the square root of the noncentrality parameter in (3)].

For example, suppose that there are $p = 2$ dimensions, the target mean is $(0, 0)'$ and the covariance matrix is the 2 by 2 identity matrix. Then, because of directional invariance, the χ^2 chart has the property that the *ARL* is the same for any vector that is the same distance from the target. Thus, for example, shifts from the target value for the mean to $(1, 0)'$, $(0, -1)'$, $(0.5\sqrt{2}, 0.5\sqrt{2})'$, and $(-0.5\sqrt{2}, -0.5\sqrt{2})'$ all have the same *ARL*. Now, if the covariance matrix is

$$\Sigma = \begin{bmatrix} 1 & \rho \\ \rho & 1 \end{bmatrix}$$

then the *ARL* is the same at all means μ for which

$$\lambda^2(\mu) = (\mu - \mu_0)'\Sigma^{-1}(\mu - \mu_0)$$

are equidistant from μ_0 for some fixed value of $\lambda^2(\mu)$. Thus, for $\mu_0 = 0$ and $\rho = 0.5$, the means $(1.0, 0.0)'$, $(\sqrt{2.5}, \sqrt{2.5})'$, $(-1.0, 0.0)'$, $(0.0, 1.0)'$, $(0.0, -1.0)'$, and $(-\sqrt{2.5}, -\sqrt{2.5})'$ are all equidistant from $\mu_0 = 0$ in the sense of (3). We also note that, for $\rho = 0.5$, the means $(1.0, 1.0)'$ and $(-1.0, 1.0)'$ are located at distances of 1.3333 and 4.0000, respectively, from $\mu_0 = 0$ even though they are the same Euclidean distance from $\mu_0 = 0$.

Other charts also have this same directional invariance property. In the univariate case, a two-sided control chart is directionally invariant if a simultaneous change of signs of all the data provides the same decision. (This is also true for non-zero target means, except that instead of a change in the signs of the data, the change would be a reflection of all the data about this target.) This is true of *symmetric* two-sided Shewhart $\bar{X}$ charts and *symmetric* two-sided *CUSUM* charts. With these charts, increases and decreases of the same magnitude in the process mean are detected equally well in the sense that they have the same *ARL*.

For a multivariate process, however, a multiple univariate scheme is not directionally invariant. Univariate charts are specifically designed for the purpose of detecting shifts in the process parameters along their respective axes. Using the example above with the identity covariance matrix and applying it to a multiple univariate scheme, the *ARLs* for the multiple univariate scheme will be different for shifts in the mean from $(0, 0)'$ to means along the axes, such as shifts to $(1, 0)'$ and $(0, 1)'$, than for off-target means that are 45 degrees from an axis, that is, means that are of the form $(\pm 0.5\sqrt{2}, \pm 0.5\sqrt{2})'$, even though those off-target means are the same *distance* from the target.

Multiple univariate charts can be used to monitor multivariate processes, of course, but such schemes must be "aimed" in some particular direction, such as along the axes of the quality characteristics themselves or aimed in the direction of the principal components. If a set of univariate charts are aimed along the axes of the individual quality characteristics, multiple univariate charts will generally do a good job of detecting shifts along these axes. In fact, Healy (1987) has suggested that to detect a shift in one particular direction, a univariate *CUSUM* chart aimed in that one direction will give the best *ARL* performance. However, if the process mean should shift in a different or unanticipated direction, such as in a direction along a principal component axis, then multiple univariate charts aimed along the axes of the original quality characteristics will not be as effective in detecting that shift. Woodall and Ncube (1985) recommend using univariate charts on principal components depending on the type of shift in the mean vector that is considered to be important to detect. However, such charts would then be less effective in detecting shifts along the axes of the original quality characteristics.

If both the individual quality characteristics and the principal components are important to monitor, then twice as many two-sided univariate *CUSUM* charts would be required. The problem with having many univariate *CUSUM* charts is that to control the false alarm rate (that is, the on-target *ARL*), control limits (or, decision intervals h) must be widened further. Thus, when there is a genuine shift in a process parameter, it is much more difficult for the univariate chart aimed in that direction to detect it. This aspect of the scheme's performance will be investigated below. We will also look at the *ARL* performance of the multiple univariate scheme when the original variables are correlated.

what also a T.S. model

Unlike non-directionally-invariant charts, a directionally-invariant chart does not lose sensitivity in detecting shifts in the process parameter when multiple directions of a shift of a given magnitude in the parameter are important. It will be seen later that when many directions of off-target shifts are important, the directionally-invariant chart can be more sensitive than a multiple univariate scheme in detecting shifts in the process parameter since the multiple univariate scheme is sensitive to both the magnitude and the direction of the shift. The directional sensitivity of the multiple univariate scheme will be shown to be even more pronounced when the original variables are correlated.

One possible disadvantage to using a multivariate, directionally-invariant chart is that it may not always be clear as to what caused the chart to signal an off-target condition. However, as in the case of using χ^2 charts, looking at individual measurements along with the principal components can provide an insight. We discuss this issue further in the next section.

Generation and Interpretation of Signals From a Control Scheme

When comparing multivariate control schemes, there are two performance aspects that should be discussed. One performance aspect concerns the question of how quickly the scheme generates a signal when an actual change in the process has occurred. Clearly, the quicker a scheme responds to a real change in a process parameter, the better. A control scheme that can quickly detect real process changes while not being overly sensitive to "false alarms" is desired.

Below, we compare several different multivariate process control schemes that have the same robustness to false alarms. This is accomplished by selecting control chart parameters such that the on-target *ARL* is the same for each scheme. Thus, the relative performance of various schemes can easily be compared by the various off-target *ARL* performances.

The second important performance aspect that arises when comparing multivariate control schemes concerns the interpretation of the signal from the multivariate control scheme. Once such a signal is generated, the question that arises is "why?" In practice, a process control engineer would want to find an assignable cause for the signal and to adjust the process control variables that will bring the process back on target. Although it can be argued that signals from a multivariate control scheme may not be trivial to interpret, a similar case could also be argued for interpreting signals from a multiple univariate scheme. Also, it may not be possible to provide corrective action on a single variable without affecting one or more of the other variables.

Since the ability to partition or isolate problems and to target specific solutions may be limited, one should use all of the available information to evaluate the process and identify appropriate corrective actions. Such information would include the relationships (correlations) between the variables. The model of a single variable being monitored and corrected (or, adjusted) in isolation is not always adequate.

For example, consider a process involving a plastic injection molding operation. Suppose that there are $p = 3$ quality characteristics of interest: the length, width, and height of a simple manufactured plastic part. If three univariate charts are being used to monitor the process, it is possible for such a scheme to issue one, two, or three signals of an off-target process. Whenever a control chart scheme signals an off-target condition, the process control engineer must determine which of several different actions to take. One possible action is to adjust the injection pressure. A low pressure would tend to produce product dimensions that were all below their target values, however, cavity geometries and local temperature gradients may increase the sensitivity of one dimension to this process drift. Other factors, including temperature changes, raw material differences, and ambient environmental conditions, all in combination with pressure fluctuations, may change the degree of sensitivity.

As an example of the interpretation complications that can exist, suppose that only the width dimension has exceeded its control limits. The process control engineer must check whether the other dimensions are on-target or are also on the low side. The engineer must consider the relationships between all of the variables. Thus, looking at the univariate components individually may not be a sufficient analysis of the signal being generated. To properly interpret the signal from such a scheme, all of the univariate control charts, as well as the principal components for the process, must be considered jointly. It is only after such an analysis is performed that the process can be properly rectified.

The diagnosis, as well as making the adjustment, is complicated. Increasing the pressure alone may not be appropriate if a combination of temperature and pressure drifts has caused the change in the process. The original process control variables settings may require some modifications if unforeseen variability has been introduced from, say, raw material changes by a supplier or other variability not accounted for in the original experimental designs when the process was implemented. A correction cannot be implemented without affecting all of the variables. The joint relationship of all the variables must be considered in order to choose the most appropriate corrective action.

We propose the use of *one* chart, the *MC*1 chart (described below), for purposes of *monitoring* a process. The *MC*1 chart can effectively replace this battery of control procedures to provide enhanced detection of process shifts in both anticipated and unanticipated directions. To assist in the interpretation of signals from the process, the *MC*1 chart could also be supplemented with a multiple univariate control scheme. Alternatively, a multiple univariate control scheme could be supplemented with the *MC*1 chart. The issue of interpretation of the signal from the chart, of course, still remains for both multiple univariate and multivariate charts. We will restrict our attention in the remainder of this discussion to the problem of quick detection of a shift in an unknown direction in the mean of a multivariate normal process.

Multivariate *CUSUM* Charts

In this section we introduce two multivariate *CUSUM* charting procedures. These procedures make direct use of the covariance matrix. Both of these multivariate *CUSUM* procedures are based on quadratic forms of the mean vector. The difference between the two multivariate *CUSUM* procedures discussed here centers on the point at which the accumulation (i.e., the sum) is made. Multivariate *CUSUM* #1 accumulates the **X** vectors *before* producing the quadratic forms while multivariate *CUSUM* #2 calculates a quadratic form for each **X** and then accumulates *those* quadratic forms. A procedure similar to *MC*1, described below, was first proposed by Pignatiello and Kasunic (1984). The development of the multivariate *CUSUM* charts is outlined in the following sections.

Multivariate *CUSUM* #1

To introduce the first multivariate *CUSUM* scheme, we consider the multivariate sum

$$\mathbf{C}_t = \sum_{i=t-n_t+1}^{t} (\mathbf{X}_i - \boldsymbol{\mu}_0)$$

where n_t is formally defined in (5) and can be interpreted as the number of subgroups since the most recent renewal (i.e., zero value) of the *CUSUM*. Since $\frac{1}{n_t}\mathbf{C}_t$ may be written as

$$\frac{1}{n_t}\mathbf{C}_t = \left(\frac{1}{n_t}\sum_{i=t-n_t+1}^{t} \mathbf{X}_i\right) - \boldsymbol{\mu}_0$$

the vector $\frac{1}{n_t}\mathbf{C}_t$ represents the difference between the accumulated sample average and the target value for the mean. Consequently, at time t, the multivariate process mean can be estimated to be $\frac{1}{n_t}\mathbf{C}_t + \boldsymbol{\mu}_0$. The norm of $\mathbf{C}_t$

$$\|\mathbf{C}_t\| = \sqrt{\mathbf{C}_t'\boldsymbol{\Sigma}^{-1}\mathbf{C}_t}$$

is seen as a measure of the distance of our estimate of the mean of the process from the target mean for the process. A multivariate control chart can be constructed by defining *MC*1 as

$$MC1_t = \max\{\|\mathbf{C}_t\| - kn_t, 0\} \tag{4}$$

and

$$n_t = \begin{cases} n_{t-1} + 1 & \text{if } \; MC1_{t-1} > 0 \\ 1 & \text{if otherwise} \end{cases} \tag{5}$$

where the choice of the reference value $k > 0$ is discussed below. The *MC*1 chart operates by plotting $MC1_t$ on a control chart with an upper control limit of UCL_1. If $MC1_t$ exceeds UCL_1 then the process is deemed to be off-target.

The *ARL* performance of the *MC*1 scheme cannot be modeled as a simple stationary Markov chain as described in Brook and Evans (1972). For this reason, a Monte Carlo simulation was employed to evaluate the *ARL* performance of this scheme.

The appendix outlines a proof showing the *ARL* performance of the *MC*1 chart depends only on the noncentrality parameter.

The multivariate *CUSUM* chart independently proposed and developed by Crosier (1988) is similar to, yet distinctively different from, our *MC*1 *CUSUM* procedure. Crosier "contracts" or "shrinks" each of his vectors of cumulative sums toward the zero vector by multiplying the cumulative sum by a scalar. The norm of the contracted cumulative sum is then compared with an upper control limit. It turns out that the *ARL* performance of Crosier's multivariate cumulative sum chart is similar to that of our *MC*1 chart. Further details are given in Crosier (1988).

Multivariate *CUSUM* Chart #2

Rather than basing a multivariate *CUSUM* statistic on the square of the distance of the accumulated sample average from μ_0, one could consider the square of the distance of each sample mean from μ_0 and then accumulate *those* squared distances. Hence, as an alternative to *MC*1, one could consider the square of the distance D_t^2 of the t^{th} sample mean from the target value of μ_0 where

$$D_t^2 = (\mathbf{X}_t - \mu_0)'\Sigma^{-1}(\mathbf{X}_t - \mu_0)$$

has a χ^2 distribution with p degrees of freedom when the process is on-target and a non-central χ^2 distribution when the process is off-target. A one-sided univariate *CUSUM* can now be formed as

$$MC2_t = \max\{0, MC2_{t-1} + D_t^2 - k\}$$

with $MC2_0 = 0$. The choice of the reference value k is discussed below. To use this multivariate *CUSUM*, one would declare the process to be off-target if $MC2_t$ exceeds an upper control limit H_2. (A multivariate *CUSUM* chart could also be constructed in terms of D_t. We have analyzed such a chart and found the difference in performance between these two charts to be negligible.)

Choice of the Reference Value k

In the (upper) one-sided, univariate *CUSUM* procedure the reference value k is often taken to be the average of the expected values of the process mean under H_0: $\mu = \mu_0$ and H_1: $\mu = \mu_1$ where μ_0 represents the on-target state and μ_1 represents a specified, unacceptable off-target state. The choice of a value for k follows from the derivation of the *CUSUM* from Wald's sequential probability ratio test. Although the (non-) central χ^2 distribution of the observations for the (off-) on-target state of *MC*2 is not symmetric, this same approach can be used. That is, the k used in *MC*2 is

$$p + 0.5\lambda^2(\mu_1).$$

Since the form of *MC*1 is different from the other *CUSUM* charts, the choice of k cannot be derived analogously. Instead, we chose k to be half of the distance of μ_1 from μ_0 (where, again, μ_1 is a specified off-target state). That is, we took $k = 0.5\lambda(\mu_1)$. Note that for both multivariate *CUSUM* charts, the value for k depends only on the magnitude of the distance of μ_1 from μ_0. It may be possible to improve the performance of the *MC*1 chart at selected off-target conditions with alternate choices for k.

Calibration of the Control Charts

To compare the performance of the various control chart schemes, several multivariate normal processes of three different dimensions (p = 2, 3, and 10) were considered. These multivariate normal processes were simulated and various control schemes were applied to these processes. The *ARL* performance of the various control schemes were then compared. The performance of the *MC*2 scheme, however, was determined using the Markov chain approach of Brook and Evans (1972) and, consequently, was not investigated by simulation.

A p-dimensional multivariate normal process was simulated by generating pseudo-random variates from a multivariate normal distribution whose mean was μ and whose covariance matrix was a $p \times p$ identity matrix. It can be shown that to monitor a linear combination of the original variables from a multivariate normal process, either of our directionally-invariant multivariate control charts can be transformed into a chart that monitors a linear combination of the principal components of the process. Since the principal components of the process can be scaled such that their covariance matrix is an identity matrix, our use of the identity covariance matrix is not a limitation. Since the multiple univariate scheme is not directionally invariant, we also consider the bivariate case for which independence of the variables is not assumed. It will be seen that, for equidistant off-target shifts, the *ARL* performance of the multiple univariate scheme can vary substantially. We also consider the case of using multiple univariate charts aimed along the principal components.

By manipulating the value of μ, an on-target ($\mu = 0$) or an off-target ($\mu \neq 0$) process can then be simulated. The only off-target conditions for the simulated processes that were considered were off-target means for the process. These off-target process means were modeled as being sudden and constant shifts in the process. That is, "slow drifts" in the mean were not investigated.

For a given distance $\lambda(\mu)$ of the process mean μ from the target value μ_0, off-target means of two forms were considered separately. Off-target means of the form

$$\mu = (\delta, 0, 0, \cdots, 0)$$

represent shifts in the process mean along one of the axes of the original variables. Off-target means of the form