

This is the author's final, peer-reviewed manuscript as accepted for publication. The publisher-formatted version may be available through the publisher's web site or your institution's library.

A visualization decision support tool for multivariate SPC diagnosis using marginal CUSUM glyphs

Shing I. Chang and Shih-Hsiung Chou

How to cite this manuscript

If you make reference to this version of the manuscript, use the following information:

Chang, S. I., & Chou, S. (2010). A visualization decision support tool for multivariate SPC diagnosis using marginal CUSUM glyphs. Retrieved from <http://krex.ksu.edu>

Published Version Information

Citation: Chang, S. I., & Chou, S. (2010). A visualization decision support tool for multivariate SPC diagnosis using marginal CUSUM glyphs. *Quality Engineering*, 22(3), 182-198.

Copyright: © Taylor & Francis Group, LLC

Digital Object Identifier (DOI): doi:10.1080/08982111003724937

Publisher's Link: <http://www.tandfonline.com/doi/pdf/10.1080/08982111003724937>

This item was retrieved from the K-State Research Exchange (K-REx), the institutional repository of Kansas State University. K-REx is available at <http://krex.ksu.edu>

A Visualization Decision Support Tool for Multivariate SPC Diagnosis using Marginal CUSUM Glyphs

Shing I Chang and Shih-Hsiung Chou*

Quality Engineering Laboratory
Department of Industrial and Manufacturing Systems Engineering
Kansas State University, Manhattan, KS 66502-5101

Abstract

A marginal CUSUM glyph is proposed to visualize and decompose out-of-control signals over time. The proposed visualization tool, consisting of a two-sided CUSUM and star glyphs, is capable of indicating when and which variables contributing to the cause. Complementing traditional multivariate control charts after indicating that the process under monitoring is out of control, the proposed method provides a visualization tool with additional diagnostic information. In addition, the proposed tool is capable of handling responses with high dimensions. Extensive simulation results conducted for up to 20 dimensions provide a user guideline of how to implement the proposed methodology.

1. Introduction

Multivariate control charts such as Hotelling T^2 are often used for a process of multiple quality characteristics. The look of a multivariate control chart often mimics that of a univariate control chart such as \bar{X} chart consisting of three horizontal lines enclosing dots connected by straight line segments. Unlike a univariate process, an out-of-control point on a multivariate control chart does not reveal which quality characteristics contribute to the problem without further analysis. A visualization tool is needed for users to make diagnostic related decisions such as when the out-of-control conditions start to take place as well as which quality characteristics are responsible. Although multiple univariate control charts, one for each quality characteristic, can be implemented as additional visual aids, this approach does not work when the number of quality characteristics is large. Not only is it difficult to control the overall type I error but scanning many univariate control charts simultaneously won't provide effective visual aid.

Many high-dimensional data visualization techniques have been developed to explore or to present datasets with multiple dimensions. Chen *et al.* (2008) summarized five tools visualizing high-dimensional datasets, they are the mosaic plots, scatter plot matrix, parallel coordinate plot, trellis displays, and star glyphs. The *Mosaic Plots*, proposed by Hartigan and Kleiner (1984), provide contingency tables used to display the relationship among two or more categorical variables. Mosaic plots consist of groups of rectangles whose sizes are corresponding to values in contingency tables. Users can interpret the mosaic plot by looking at its sizes and the positions of the rectangles. Figure (a) shows an example of the mosaic plot for a 2009 automobile dataset (collected from Automotive.com <http://www.automotive.com/index.html>). The mosaic plot displays the number of cars by country and

* Contacting author, cls3415@ksu.edu

prices under the aspects of horse power (hp) and miles per gallon (mpg). Although the mosaic plot is a well-known visualization tool to multivariate datasets, it can only cope with categorical datasets, meaning that data has to be put into categories. It is difficult to apply a mosaic-plot approach to monitor a statistic process when quality characteristics are quantitative. In the car example, horse power is categorized into two classes – 0-350 and 350-850. In addition, a mosaic plot can only handle a small number of factors, e.g. four factors are considered in the car example in Figure (a). A mosaic plot will not be able to handle problems with much larger number of factors.

A *Scatter Plot Matrix* is another visualization tool to view all the pair-wise scatter plots into a matrix. Each scatter diagram displays the relationship between any two-variable pairs. The scatter plot matrix can be potentially used for process monitoring. Figure (b) displays an example of scatter plot for a three-dimensional dataset. The dataset was generated from the multivariate normal distribution with one-step mean shifted, i.e. $\mu=(5,10,20)$ change to $\mu=(7,12,22)$ and $\Sigma=\begin{bmatrix} 1 & 0 & 0.3 \\ 0 & 1 & 0 \\ 0.3 & 0 & 1 \end{bmatrix}$, where μ is a vector of mean and Σ is a variance-covariance matrix. Examining Figure 1 (b), users can easily identify that var1 and var3 are corrected, but it is very difficult to distinguish there is a mean shift by looking at this plot. Montgomery (2008) gave an example of using principal component analysis and scatter plot with 95% confidence ellipse to show out-of-control points. Although the scatter plot matrix is a good visualization tool to display the pair-wise relationships in a snap shot, it has some limitations. First, a scatter plot presents all data points in a snapshot. Users only can see some points plot outside an ellipse but the time sequence of the plotting points is lost. Second, a scatter plot cannot handle hyper-dimensional data, such as a 50-dimensional or 100-dimensional data set. In fact, 20 dimensions will post a challenge for pair-wise displays. There will be 20 x 20 boxes needed to be displayed simultaneously.

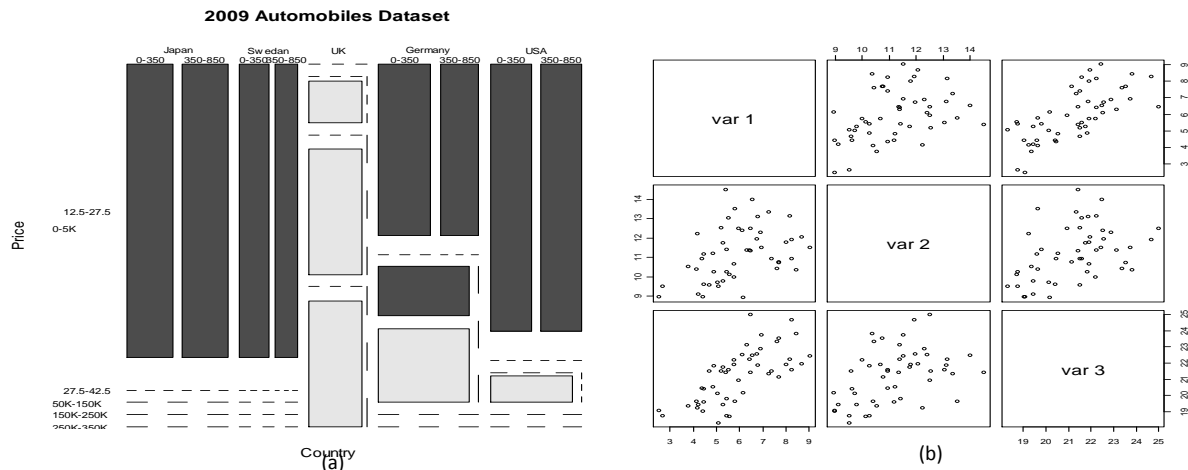


Figure (a) The Example of the Mosaicplot for 2009 Automobile Dataset; (b) The Example of the Scatter Plot Matrix for the Three-Dimensional Dataset.

The third multivariate visualization tool is the *Parallel Coordinate Plot*. A parallel coordinate plot contains coordinate axes (vertical or horizontal) in parallel that can accommodate many variables at a time in the same plant. The number of parallel axes is corresponding to the number of variables. For each parallel axes, the bottom of line is the minimum value of the variable, while the maximum value of the variable is at the end of the line. A particular observation is represented by a line that connects the

vertical line at the specific height, which is the proportion of value of the observation to length of vertical line. Figure (a) is a simple example of the parallel coordinate plot. The dataset consists of a three-dimensional data with five observations generated from the multivariate normal distribution. Although a parallel coordinate plot can handle high-dimensional data, too many observations on the same plot tends to lose its effectiveness. As shown in Figure (b), it is difficult to make out the changes made on an individual response among 100 observations with 20 dimensions each. Therefore, it is hard to tell if any out-of-control signal occurs.

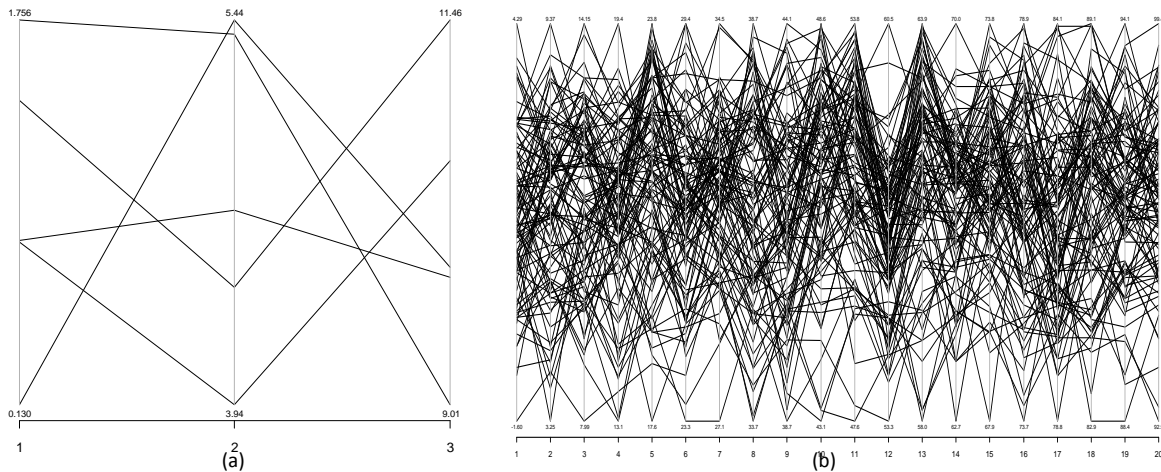


Figure (a) An Simple Example of 5 observations, 3-dimensional dataset; (b) Example of 100 observations, 20-dimensional dataset with one step mean shift.

Although there were many techniques to visualize the multivariate data about its quantities or relationship, it is difficult to apply these techniques to visualize data for the purpose of statistical process control for multivariate processes. On the other hand, multivariate control charts based on Hotelling's T^2 statistic (Jackson, 1985), MCUSUM (Woodall and Ncube, 1985; Crosier, 1988; Pignatiello and Runger, 1990), MEWMA (Lowry, *et al.*, 1992) cannot effectively indicate which variable(s) contribute to an out-of-control signal.

Hawkins (1991) proposed a procedure called regression adjustment. Essentially, the regression adjustment is a set of univariate control charts of the residuals from the regression of each variable on all others. He further used regression adjustment to each variable to analyze out-of-control-caused variables. Since each variable contains at least one chart, it needs a large number of charts as many as the number of variables for high-dimensional datasets. Another method that indicates variables which caused out-of-control is based on decomposition of the Hotelling's T^2 statistic. Mason, Tracy and Young (1995) decomposed T^2 to interpret signals. Suppose $p=3$, there are $3!=6$ combinations of decompositions of one T^2 value. For example $T^2 = T_1^2 + T_{2,1}^2 + T_{3,1,2}^2$. It would lead to a very large number of combinations when p is large. Furthermore, Runger, Alt, and Montgomery (1996) used a similar concept to decompose an out-of-control signal. They used $d_i = T^2 - T_{(i)}^2$ as an indicator to show the contribution of the i^{th} variable. T^2 is the current value of statistic and $T_{(i)}^2$ is the value of the statistic for all process variables except the

i^{th} one. Then, d_i is a major contributor to T^2 if d_i is large. Note that decomposition is usually applied to the latest T^2 value when the process is deemed out of control.

Although these methods introduced above can help diagnose **which** variables contribute to an out-of-control signal, they do not offer the information **where** or **when** the responsible variables went out of control. The reason is that decomposition is applied on the latest point that causes out-of-control signal. In this study, we introduce a novel technique that combines CUSUM control charts and glyphs to provide a visualization decision support tool to indicate **when and which** variable(s) are out of control. As shown in Figure and 4 in the background section, a trellis displays and star glyphs are very effective visualization tool for multivariate datasets. Features of these visualization approaches are applied to the proposed method in this study.

The rest of this paper is organized as follows. Section 2 introduces backgrounds of the trellis displays and star glyphs as well as the standardized two-side CUSUM and the multivariate CUSUM. Section 3 is the proposed method using marginal CUSUM Glyphs followed by the experimental design of the proposed simulation study. Next, the computational results of simulations are showing in section 4. Section 5 consists of conclusions and future studies.

2. Background

This section summarizes the trellis displays and star glyphs. In addition, the multivariate control charts adopted in the proposed visualization tool are also introduced in this section. Although both Exponential Weighted Moving Average (EWMA) and Cumulative Sum (CUSUM) control charts are effective detection and diagnosis tools for univariate processes, Hawkins (1993) stated that the CUSUM chart is a little better. In addition, Lowry *et al.* (1992) and Lowry and Montgomery (1995) both concluded that the multivariate EWMA can be as good as a multivariate CUSUM chart to detect small mean shifts. Thus, the proposed method will adopt two CUSUM charts: standardized two-side CUSUM (SD2CUSUM) and Multivariate CUSUM (MCUSUM). To investigate if any out-of-control signal occurred, users first run MCUSUM as a preliminary step of process monitoring so that the overall type I error is under control. If the MCUSUM provide an out-of-control signal, users can show all observations that are before the out-of-control signal in marginal CUSUM glyphs on the trellis display.

2.1 Trellis Displays

A trellis display is a lattice-like arrangement to lay out plots into rows, columns, and pages on multiple panels. The plots on the panels can be histogram, kernel density plot, theoretical quantile plot, two-sample quantile plot, stripchart, bar plot, scatter plot, parallel coordinate plot, etc (Sarkar, 2008). Each panel contains a subset of the data graphed by plots. For example, Yates (1935) studied a split-plot experiment. The structure of the oats data set is including 72 rows and 4 columns, and its attributes are including six blocks (I, II, III, IV, V, and VI), three varieties (Victory, Golden Rain, and Marvellous), four concentrations of nitrogen, and yield. Figure shows the trellis display of the Oats data with block I and II. The yield of oats is plotted against concentration of nitrogen for three varieties of oats and two blocks. In this paper, the idea of lattice-like arrangement of the trellis display is applying to the proposed multivariate visualization tool. The block assignment will be replaced by the time sequence while the

star glyphs will be adopted to represent each multivariate observation on panels. The star glyphs are introduced in next section.

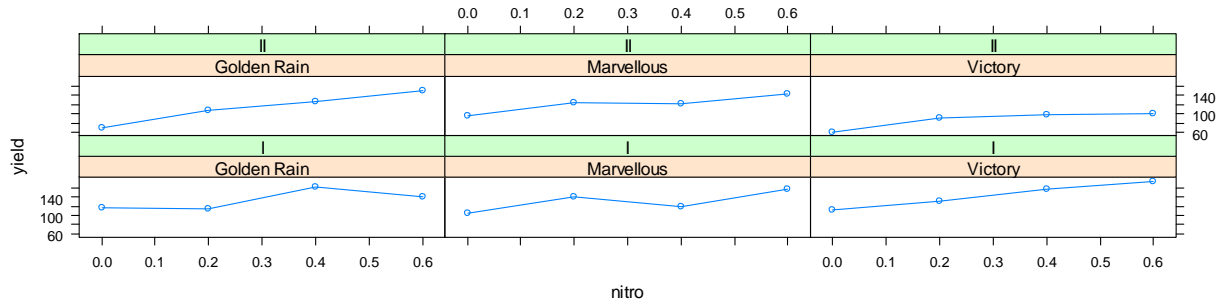


Figure A Trellis displays of the Oats data.

2.2 Star Glyphs

Star glyphs (also called star plot) display a multivariate data set in a geometrical shape such as hexagon for six dimensional data. In a star glyph, each individual graphic represents an observation. A star contains n spikes that radiate from the center with even space. The angle between each spike is equal to $360/n$ degree. Each spike has a value on it with the same proportion of the variable for that observation, which means observations must be standardized before stars are constructed. At the end of the point of each spike is usually connected to each other by line segments. To demonstrate the star glyph, five cars with seven attributes are used as an example. The seven attributes are: miles per gallon (mpg), weight of lbs (weight), horsepower (hp), engine displacement in cubic inches (disp.), number of cylinders (cyl.), torque of Newton-Meter (Torque), and price in US dollar (Price).

Figure shows five examples of star glyphs. The graphic (a) to (e) are five different sedan vehicles with seven attributes for star glyphs demonstration, while graphic (f) shows the basis of variable assignment in which each spike represents one attribute. The longer the spike the large the values for that attribute. The car “Saab 9-3 2.0T” shown in Figure (b) has high mpg, less weight, low horse power, small engine displacement, four cylinders, small torque, and lowest price compared to the other cars. A side-by-side comparison with Bentley demonstrates why the mpg of Bentley seems to disappear. The Saab has 29 mpg while the Bentley only has 14 mpg. However, the sticker prices are the opposite: Saab \$30,360 vs. Bentley \$224,990. The function of star glyphs called “stars” in the R language (<http://www.r-project.org/>) in its graphics package was used to generate these star glyphs in Figure .

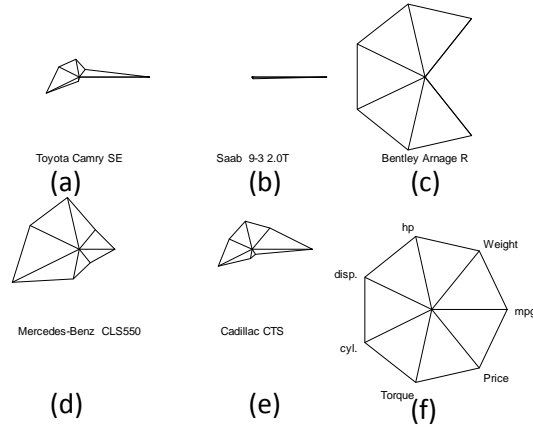


Figure . Examples of Star glyphs where (f) Represents the Legend

2.3 Standardized Two-Side Univariate CUSUM (SD2CUSUM)

The standardized two-side univariate CUSUM (SD2CUSUM) essentially is a tabular CUSUM with the standardized observations. This control chart is mainly designed for a process with only one quality characteristic. To run this procedure, it is necessary to obtain a phase I data set, so that mean μ_0 and standard deviation σ can be estimated. The data collected during a phase I process is often deemed a nominal representation of the underlying process for future observations to follow. Basically, the tabular CUSUM accumulates the differences between observations and the target. Two types of tabular CUSUM statistics are one for positive mean shifts and the other for negative mean shifts. Specifically, the positive part deals with the observations that are above the target, while the negative part copes with those below the target. Let x_i be the one dimensional i^{th} observation on the process. Note that the statistic of the negative part is a positive number as well. However, when these two statistics are plotted on the same chart, the CUSUM statistic for the negative shifts are plotted below the center line as showed in Figure . The statistics of the tabular CUSUM is showing below (Hawkins, 1993).

$$C_i^+ = \max[0, x_i - (\mu_0 + K) + C_{i-1}^+] \quad (1)$$

$$C_i^- = \max[0, (\mu_0 - K) - x_i + C_{i-1}^-] \quad (2)$$

where the C^+ and C^- represent the positive and negative parts of tabular CUSUM. The starting values are $C_0^+ = C_0^- = 0$. K is called the reference value (or called the allowance or slack value). Usually, K is about half of the absolute difference between the target value μ_0 and the out-of-control value of mean μ_1 to be detected, i.e.,

$$K = \frac{|\mu_1 - \mu_0|}{2} \quad (3)$$

K can also be expressed in standard deviation form in terms of $\delta = |\mu_1 - \mu_0|/\sigma$, which provides the magnitude of the shift to be detected. And K then becomes $\frac{\delta}{2} \sigma$ or $K = k \sigma$, $k = \frac{\delta}{2}$.

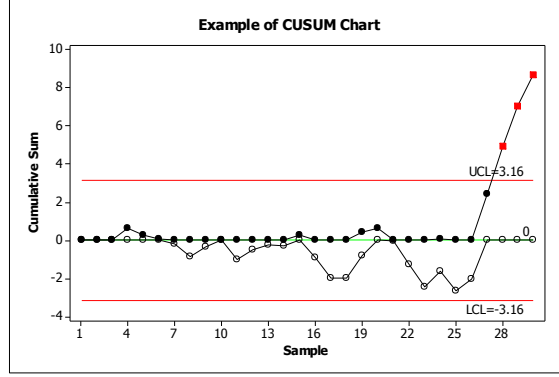


Figure . An Example of the CUSUM Chart. The first 25 dataset was generated from $N(10,1^2)$, while the rest 5 dataset was from $N(12,1^2)$.

The following steps in equations (4) to (6) are taken to standardize the original data set. Specifically, the i^{th} observation is standardized as

$$y_i = \frac{x_i - \mu_0}{\sigma} \quad (4)$$

where σ is from the phase I estimate or known, and the equation (5) and (6) become

$$C_{p,i} = \max[0, y_i - k + C_{p,i-1}] \quad (5)$$

$$C_{n,i} = \max[0, -k - y_i + C_{n,i-1}] \quad (6)$$

where $C_{p,i}$ is the i^{th} observation statistic of standardized two-side CUSUM for positive mean shifts, while $C_{n,i}$ is the i^{th} observation statistic for negative mean shifts.

To decide whether the observation is out-of-control or not, H is the critical value for an out-of-control signal. If either C_i^+ or C_i^- is larger than H , it is considered that the process is out of control. Montgomery (2008) suggested that H be 5 times the process standard deviation σ , i.e, $H=h\sigma$, where $h=5$. Since the process has been standardized, the detection criterion can be revised to whether $C_{p,i}$ or $C_{n,i}$ is larger than h or not.

Furthermore, since the CUSUM procedure is based on the accumulated difference between observations and the target, a delay phenomenon will affect when an out-of-control signal is detected. For example, if a process has a mean shift at x_i , but because of this phenomenon, the system will not give an alarm until x_j , $i < j$. Montgomery (2008) provided an example of finding the last-in-control observation, and in this paper, a pseudocode of tabular CUSUM is implemented in Figure 6, where N^+ is the counter that records the number of consecutive periods that the upper-side CUSUM $C_{p,i}$ values are larger than zero, while N^- is the counter for the lower-side CUSUM $C_{n,i}$. And the last-in-control position l is equal to $o-N^+$ or $o-N^-$, where o is the stopping counter number when the first out-of-control signal takes place during the phase II monitoring.


```

Procedure of Tabular CUSUM
set  $i=1, C_{p,0}$  and  $C_{n,0}$ 
while  $C_{p,i}$  and  $C_{n,i}$  smaller than  $h$  do
  calculate  $C_{p,i}$  and  $C_{n,i}$  by using (5) and (6).
  if  $C_{p,i} > 0$  then  $N^+ = N^+ + 1$ 
  else  $N^+ = 0$ 
  if  $C_{n,i} > 0$  then  $N^- = N^- + 1$ 
  else  $N^- = 0$ 
   $i++$ 
end while
out-of-control  $o=i$ 
last-in-control  $l = o - N^+$  or  $o - N^-$ 
print  $C_{p,i}, C_{n,i}, o$ , and  $l$  for all  $i$ .

```

Figure Pseudocode of Procedure of Tabular CUSUM

2.4 The Multivariate CUSUM Chart

For processes with multiple quality characteristics, a multivariate CUSUM should be implemented. Multivariate control charts are often recommended over running multiple univariate control charts simultaneously because the overall type I error can be controlled to a desired level. Crosier (1988) proposed one of the better multivariate CUSUM schemes in term of the ARL performance called vector-valued CUSUM. The main idea of this multivariate CUSUM (MCUSUM) is similar to that of the univariate CUSUM, except that MCUSUM multiplies the statistic by a weight.

Let X_i be i^{th} observation on the process. X_i is from a multivariate normal distribution with mean T and variance-covariance Σ which are known or can be estimated from a phase I data set. The statistics of MCUSUM is as below.

$$C_i = [(S_{i-1} + X_i - T)' \Sigma^{-1} (S_{i-1} + X_i - T)]^{1/2} \quad (7)$$

then

$$\begin{cases} S_i = 0 & \text{if } C_i \leq k \\ S_i = (S_{i-1} + X_i - T)(1 - \frac{k}{C_i}) & \text{if } C_i > k \end{cases} \quad (8)$$

where $S_0=0$ and $k>0$

An out-of-control signal is generated when

$$y_i = [S_i' \Sigma^{-1} S_i]^{1/2} > h \quad (9)$$

where k is a reference value and h is a decision interval for the process.

The chosen value of k and h values for in-control ARL of 200 and 500 can be found in Crosier's study (1988). In this study, we reorganize their result in Table 1.

Table . Chosen value of k and h with different kinds of dimension p under two in-control ARLs.

p	In-Control ARL's of 200		In-Control ARL's of 500	
	h	k	h	k
2	5.50	0.5	6.65	0.5
5	9.46	0.5	10.9	0.5
10	14.9	0.5	17.2	0.5
20	24.7	0.5	28.0	0.5

3. The Proposed Marginal CUSUM Glyphs

The proposed framework first relies on the MCUSUM described in the previous section to detect any process shift so that the overall type I error is desirable. If the type I error is excessive, too many false alarms will be triggered. Operators may become placid and ignore a true process shift that may prove to be costly in the future. Therefore, in this study, the proposed method uses the multivariate control chart as the primary detection method, while the marginal univariate control chart described in section 2.3 is used as the diagnostic tool once the multivariate control chart signals an out-of-control observation.

The proposed diagnostic tool, the marginal CUSUM glyphs, is presented in this section. The proposed method, integrating marginal CUSUM into one glyph, is a visualization tool for a decision support system dealing with multivariate quality characteristics. Specifically, glyphs plotted over time are organized in panels of trellis. Every row contains two time series glyphs, i.e., C_{pi} and C_{ni} . C_{pi} is a glyph of i^{th} observation for the upper (positive)-side CUSUM statistic, while C_{ni} is a glyph of the same observation for the lower (negative)-side CUSUM statistic. Each vertex in the glyph represents a response that shares equal space around 360 degrees. When a CUSUM statistic becomes large, the corresponding vertex would grow in length. On the other hand, an in-control process would have all vertices close to the center "dot." Finally, a circle with the radius h encircles the center dot, and this circle represents a control limit for each glyph. If a few responses are responsible for an out-of-control MCUSUM signal, their marginal CUSUM would show abnormality in the glyph over time leading to the out-of-control moment, and those spikes would exceed the circle. The following steps describe the main procedure of the proposed marginal CUSUM glyphs.

Step 1: Run MCUSUM until it provides an out-of-control signal at location or time τ assuming one sample is taken for each sample period.

Step 2: Run SD2CUSUM procedure to generate $C_{p,i}$, $C_{n,i}$ and record any out-of-control o , and last-in-control location l for each dimension. The out-of-control position for the i^{th} dimension o_i is the location that exceeds the control limit h , while the last-in-control location l_i is equal to $o_i - N^+$ or $o_i - N^-$ for dimension i where N^+ and N^- are defined in section 2.3.

Step 3: Construct star glyphs starting from sample periods 1 to τ for both positive and negative glyphs. Each glyph consists of a circle with radius of h and spikes with length of $C_{p,i}$ or $C_{n,i}$ radiating from the center of a circle. We choose to start the spikes from 3 o'clock position and move the spikes in counterclockwise direction with identical angle between any adjacent spikes.

Step 4: Print out out-of-control o and last-in-control l for each dimension, and display the marginal CUSUM glyphs. Out-of-control signals indicate *which* variables contribute statistics of MCUSUM, while last-in-control locations show *when* these variables become out-of-control.

With a stable and in-control process, a glyph on the trellis panel should present a “dot” with some tiny spike inside of a circle. However, users might not see some small change by looking at a “dot”, for a tiny spike might be difficult to distinguish. Instead of a “dot”, users can add a constant c on both $C_{p,i}$ and $C_{n,i}$ to expend a “dot” into a polygon shape. The new statistics showing in equation (10) and (11) denoted $CC_{p,i}$ and $CC_{n,i}$ would replace the original $C_{p,i}$ and $C_{n,i}$. Also the same c is required to be added to the radius h to maintain the same scale of threshold as shown in equation (12). This new limit is denoted as H .

$$CC_{p,i} = c + \max[0, y_i - k + C_{p,i-1}] \quad (10)$$

$$CC_{n,i} = c + \max[0, -k - y_i + C_{n,i-1}] \quad (11)$$

$$H=c+h \quad (12)$$

3.1 An Example to Demonstrate the Use of the Proposed Method

An example of 5-dimensional, 20 observations with one sigma mean shift example is presented in this section. The dataset is shown in column of x_1 to x_5 in Table 2. It was generated from the multivariate normal distribution. The first 10 observations was simulated as in-control data from a multinormal distribution with mean vector $\mu_0=(5,10,15,20,25)$ and variance-covariance matrix shown in equation (13), while the last 10 observations represented out-of-control data from $\mu_1=(6,10,16,20,26)$, i.e. a small mean shift at the first, third and fifth variables, and its variance-covariance structure same as that of the first 10 observations.

$$\Sigma = \begin{bmatrix} 1 & 0.3 & 0.3 & 0.3 & 0.3 \\ 0.3 & 1 & 0.3 & 0.3 & 0.3 \\ 0.3 & 0.3 & 1 & 0.3 & 0.3 \\ 0.3 & 0.3 & 0.3 & 1 & 0.3 \\ 0.3 & 0.3 & 0.3 & 0.3 & 1 \end{bmatrix} \quad (13)$$

A Scatter plot of this example is shown in Figure , in which the dots represent the first 10 observations, while the squares are for the last ten. The ellipses are 95% confidence contours based on the normal assumption. The solid-line ellipses are for the first 10 observations, while the dash-line ellipses are for the last 10. Note that, in reality, one would not know in advance when such a shift takes place or if it takes place at all. Even with dots and squares it is difficult to identify the process status.

At the beginning of the proposed method, the MCUSUM is applied. After using the equations (7), (8), and (9) with $h=9.46$ and $k=0.5$, the statistic of MCUSUM is showing in the column of C in Table 2 and its control chart is showing in Figure . The out-of-control signal occurred starting at 15th observation.

At this point, the diagnostic process begins. The first step is to standardize every dimension by using equation (4) where $\sigma=1$ and $\mu_0=(5,10,15,20,25)$. Next, the procedure of two-side CUSUM is applied via equations (5) and (6) with $h=5$ and $k=0.5$. The marginal CUSUM statistics are shown in Table .

To Determine Which Variables Contribute to the Out-of-control Signal

The data of $C_{p,i}$ and $C_{n,i}$ where $i=1$ to 15 in Table are used to construct the proposed glyphs. After applying the traditional star glyphs described in section 2.2 for each $C_{p,i}$ and $C_{n,i}$, the solid-line circle with radius= h is also added on each of glyph. In Figure , the proposed glyphs are shown. The graph of “basic” is an elementary glyph of a scale. Note that the star glyphs are not visible when the process is in control, where the values of $C_{p,i}$ and $C_{n,i}$ are close to zero. Therefore, a constant c is added to $C_{p,i}, C_{n,i}$ and h according to equations (10) to (12). In this case, the glyphs with $c=3$ are shown in Figure . There are three spikes exceeding the circle, between CC_{p14} and CC_{p20} indicating that the means of variables $x_1, x_3,$ and x_5 have shifted. In addition, there is no spike exceeding the circle in any of C_n or CC_n . One can conclude that no negative mean shift occurred.

To Determine When Shifts Took Place

In Table , the 14th observation of C_p statistic in the first dimension is 5.83, which is larger than $h=5$. Thus, it is deemed that the first dimension has shift at 14th observation on positive-side CUSUM. Moreover, since the correspond $N_{p,1}$ of 14th observation is 6, the last-in-control observation is $14-6=8^{\text{th}}$ observation according to the definition of last-in-control in section 2.3. Using the same philosophy, the third and the fifth dimensions are found to shift at 16th and 17th, respectively. And their last-in-control locations are at 10th observation for the third dimension, and 11th observation for the fifth dimension. And there is no shift indicated in negative-side CUSUM. The proposed method indicates that variables $x_1, x_3,$ and x_5 make contribution to the out-of-control signal at location 8th, 10th, and 11th, respectively.

Table . 5-dimensional, 20 observations dataset.

	X_1	X_2	X_3	X_4	X_5	C
1	4.79858	10.0034	15.0335	19.2843	25.0273	0.463891
2	4.85333	9.1170	14.8924	18.3470	21.8418	2.607937
3	4.32499	8.9170	16.2335	19.6594	25.6710	1.775256
4	5.23184	10.2377	14.3568	18.8797	25.0120	2.239928
5	6.76232	8.4736	14.8295	18.7632	25.3518	3.221392
6	3.93670	10.0544	14.8085	18.3438	26.4359	2.147643
7	4.98181	8.9128	12.3676	20.7162	24.5627	2.440458
8	4.57178	12.2499	15.3191	20.6362	26.2305	2.130893
9	5.85859	9.7962	14.5142	20.2430	24.0448	0.772875
10	6.95341	9.3181	15.1976	20.6888	24.6941	1.788469
11	7.15771	8.9142	15.6287	19.4212	25.4814	3.296222
12	6.23149	9.7179	16.4945	19.4940	26.2335	5.596017
13	6.16967	9.5291	17.2668	19.7801	26.9197	7.212117
14	6.45825	10.9076	15.4995	19.0572	25.7781	8.722549
15	7.39763	10.4058	15.5910	21.2973	25.9004	11.45267
16	5.26227	10.2907	17.4474	21.4488	25.6842	13.99988
17	5.42656	12.8401	18.6317	19.5853	25.1729	15.96414
18	7.06759	10.2345	16.2082	19.4704	26.4889	15.93851
19	7.20849	10.2192	13.8677	19.0505	26.8322	16.17303
20	6.00803	10.2438	15.9425	20.9734	26.3481	18.32671

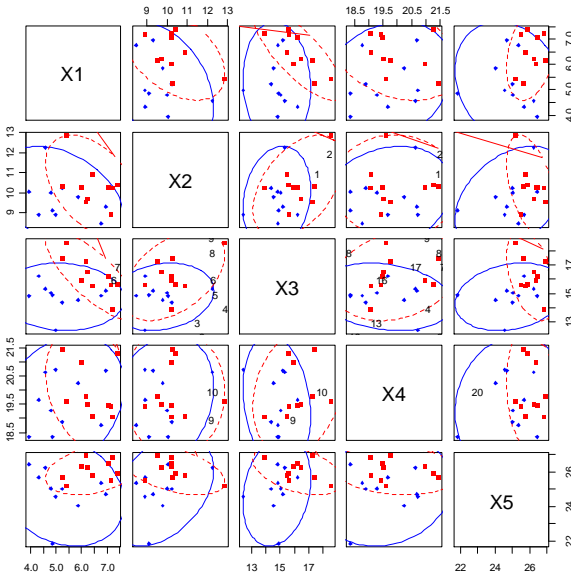


Figure . The scatter plot of the given example.

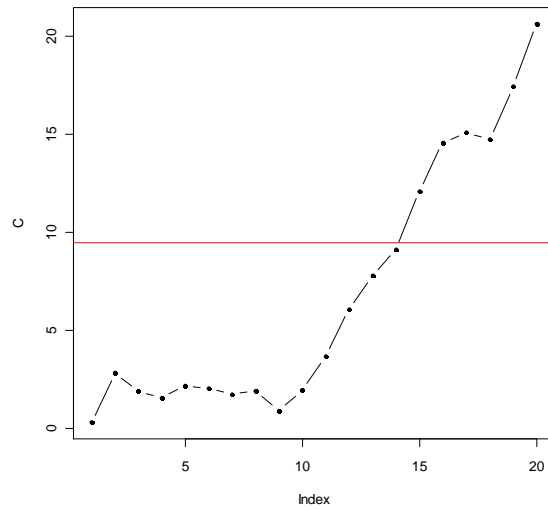


Figure . MCUSUM control chart of the given example.

Table . The result of two-side standardized CUSUM for the given example.

(a) SD2CUSUM for Positive Mean Shifts

i	$C_{p,i,1}$	$N_{p,i,1}$	$C_{p,i,2}$	$N_{p,i,2}$	$C_{p,i,3}$	$N_{p,i,3}$	$C_{p,i,4}$	$N_{p,i,4}$	$C_{p,i,5}$	$N_{p,i,5}$
1	0.00	0	0.00	0	0.00	0	0.00	0	0.00	0
2	0.00	0	0.00	0	0.00	0	0.00	0	0.00	0
3	0.00	0	0.00	0	0.73	1	0.00	0	0.17	1
4	0.00	0	0.00	0	0.00	0	0.00	0	0.00	0
5	1.26	1	0.00	0	0.00	0	0.00	0	0.00	0
6	0.00	0	0.00	0	0.00	0	0.00	0	0.94	1
7	0.00	0	0.00	0	0.00	0	0.22	1	0.00	0
8	0.00	0	1.75	1	0.00	0	0.35	2	0.73	1
9	0.36	1	1.05	2	0.00	0	0.10	3	0.00	0
10	1.81	2	0.00	0	0.00	0	0.28	4	0.00	0
11	3.47	3	0.00	0	0.13	1	0.00	0	0.00	0
12	4.20	4	0.00	0	1.12	2	0.00	0	0.73	1
13	4.87	5	0.00	0	2.89	3	0.00	0	2.15	2
14	5.83	6	0.41	1	2.89	4	0.00	0	2.43	3
15	7.73	7	0.31	2	2.98	5	0.80	1	2.83	4
16	8.90	8	0.00	0	4.81	6	0.78	2	5.53	5
17	9.06	9	0.00	0	6.64	7	0.19	3	6.75	6
18	9.35	10	0.00	0	7.00	8	0.00	0	6.44	7
19	8.93	11	0.00	0	6.12	9	0.00	0	7.75	8
20	11.18	12	0.87	1	7.04	10	0.16	1	9.56	9

(b) SD2CUSUM for Negative Mean Shifts

i	$C_{n,i,1}$	$N_{n,i,1}$	$C_{n,i,2}$	$N_{n,i,2}$	$C_{n,i,3}$	$N_{n,i,3}$	$C_{n,i,4}$	$N_{n,i,4}$	$C_{n,i,5}$	$N_{n,i,5}$
1	0.00	0	0.00	0	0.00	0	0.22	1	0.00	0
2	0.00	0	0.38	1	0.00	0	1.37	2	2.66	1
3	0.18	1	0.97	2	0.00	0	1.21	3	1.49	2
4	0.00	0	0.23	3	0.14	1	1.83	4	0.98	3
5	0.00	0	1.25	4	0.00	0	2.57	5	0.12	4
6	0.56	1	0.70	5	0.00	0	3.72	6	0.00	0
7	0.08	2	1.29	6	2.13	1	2.51	7	0.00	0
8	0.01	3	0.00	0	1.31	2	1.37	8	0.00	0
9	0.00	0	0.00	0	1.30	3	0.63	9	0.46	1
10	0.00	0	0.18	1	0.60	4	0.00	0	0.26	2
11	0.00	0	0.77	2	0.00	0	0.08	1	0.00	0
12	0.00	0	0.55	3	0.00	0	0.08	2	0.00	0
13	0.00	0	0.52	4	0.00	0	0.00	0	0.00	0
14	0.00	0	0.00	0	0.00	0	0.44	1	0.00	0
15	0.00	0	0.00	0	0.00	0	0.00	0	0.00	0
16	0.00	0	0.00	0	0.00	0	0.00	0	0.00	0
17	0.00	0	0.00	0	0.00	0	0.00	0	0.00	0
18	0.00	0	0.00	0	0.00	0	0.01	1	0.00	0
19	0.00	0	0.00	0	0.00	0	0.00	0	0.00	0
20	0.00	0	0.00	0	0.00	0	0.00	0	0.00	0

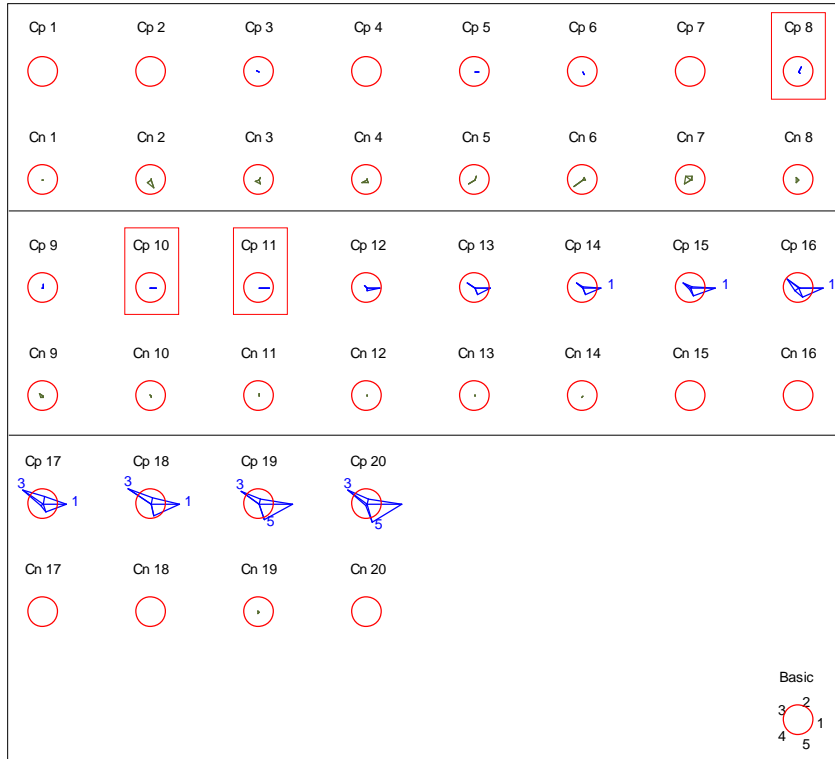


Figure . The CMCUSUM glyphs of the given example with $c=0$.



Figure . The CMCUSUM glyphs of the given example with $c=3$.

4. A Simulation Study

To examine the capability of the proposed method, a simulation study is planned to explore magnitude of h , number of dimensions, number of dimensions shifted, and type of correlation-coefficient structures. Figure shows the detailed experimental plan of the proposed study. The design is to test the capability of the proposed method at different types of value of h in terms of absolute value of deviation between true shift location and calculated shift location. Each factor combination is repeated 1000 times. Each run contains 100 observations with 1σ mean shifts taking place at 31st observation for all planned dimensions. For example, considering a five-dimensional, 100 observations dataset with three of five dimensions having a step mean shift, the simulation data set would have the first 30 observations generated from the multivariate normal distribution with μ_0 , while the last 70 observations are from μ_1 . The means of the non-shifted dimensions stay at the same magnitudes. The variance shift is not considered in this study.

Magnitude of h: h=3, 4, 5, 6, 7, 8
Number of Dimension: dim=3, 5, 10, 20
Number of Dimensions Shifted: 1. Small: 1/3 of number of dimensions shifted—3(1); 5(2); 10(3); 20(5) 2. Medium: half of number of dimensions shifted—3(2); 5(3); 10(5); 20(10) 3. Large: All of number of dimensions shifted—3(3); 5(5); 10(10); 20(20) Note: 3(2) means two out of three dimensions are shifted.
Types of Correlation-Coefficient Structure: $\rho=0, 0.5, 0.9$
Input Data Series Type: Multivariate Normal Distribution

Figure . Experimental design of the proposed method at different types of h .

Magnitude of h

Since h is the major criterion of setting the control limit of CUSUM, it is necessary to test the proposed method under different magnitudes of h . The smaller the value of h the more quickly the SD2CUSUM detects a shift. However, the type I error would increase. On the other hand, if h has been increased, the type II error would increase. In this paper, six types of value of h are applied to, *i.e.*, $h=3, 4, 5, 6, 7, \text{ and } 8$.

Number of Dimensions

The number of dimensions is one of the factors considered in implementing multivariate control charts, especially in visualization of multivariate dataset. In this study, four dimensions are considered. They are 3, 5, 10, and 20. This factor allows us to study how effective the proposed method is when the dimension increases.

Number of Dimensions Shifted

There are three different levels of number of dimensions shifted in this study: small, medium, and large. For small level, 1/3 of number of dimensions shifted. For example, if one dimension has 1σ

mean shift in a three-dimensional dataset, this case is in the small level category. That is, the first 30 observations follow multivariate normal distribution with $\mu_0=(5, 10, 15)$, while the last 70 observations are generated from multivariate normal distribution with $\mu_1=(6, 10, 15)$. The notation of this case is 3(1). In addition, the other small number of dimension shifted cases has 2, 3, and 5 dimensions shifted for 5, 10, and 20 dimensions, respectively. For the medium level, half of number of dimensions shifted, while all of number of dimensions has shifted for large number of dimensions shifted cases. Note that the goal of this factor is to study the effect of how wide-spread mean shifts are to the responses rather than the magnitude of the shifts.

Types of Correlation-Coefficient Structure

Different types of variance-covariance structures are also considered in this simulation. Without the loss of generality, three types of correlation-coefficient matrices are used to study no, medium and high correlations among the responses. The correlation-coefficient structure for the simulated data that generated from multivariate normal distribution is the following.

$$\Sigma = \begin{bmatrix} 1 & \dots & \rho \\ \vdots & \ddots & \vdots \\ \rho & \dots & 1 \end{bmatrix}, \text{ where } \rho=0, 0.5, \text{ and } 0.9 \quad (11)$$

4.1 Simulation Results and Discussions

This section shows the computational results of the proposed simulation study using 1,000 runs for each combination – 216,000 runs in total. The performance of the proposed method is measured by four statistics: the **correct identification percentage**, the **average absolute value of deviation**, **type I error rate (false alarm rate)**, and **type II error rate**. These four statistics will be applied to the last-in-control signals. The first one is the bigger the better, while the rest of which are the smaller the better. The correct identification percentage indicates how accurate the system is to identify which responses are responsible for the out-of-control signal. The average absolute value of deviation measures the distance between true location of a mean shift and its calculated location among all dimensions that exhibit mean shifts. Specifically, it is calculated the follows:

$$\text{Average absolute value of deviation} = \frac{\sum_{i=1}^m |l_i - t_i|}{m} \quad (12)$$

where l_i is the calculated location of i^{th} dimension of dataset, t_i is the true location of i^{th} dimension, and m is the number of dimensions that are identified correct out of control. Under the measure of those out-of-control dimensions, the type I error rate, or the false alarm rate, is the probability when the proposed method indicates a dimension is out of control while, in truth, it is not. The type II error rate, on the other hand, is the probability when it fails to identify a shifted dimension. The desired result for the correct identification percentage is 100%, while that of the other three statistics is the closer to zero the better.

The suggestion of h for the proposed method provides a guideline for implementing the proposed method and will be discussed. The simulation results are shown in Figure to 14, and their corresponding tables are available from the authors upon request. For example, Figure (a) indicates the simulation results of small number of dimensions shifted with $\rho=0$. Each sub-chart consists of four performance measures – correct identification percentages (correctness), average absolute value of deviation(dev), type I error rate (in %) (type I), and type II error rate (in %)(type II). Specifically Figure (a) summarizes the simulation result of small number of dimensions shifted under different h with $\rho=0$, while Figure (b) and 12(c) are for $\rho=0.5$ and $\rho=0.9$ respectively. Next level down, Figure (a-1) to (a-4) depict four performance criteria respectively. For example in Figure (a-1), correct identification percentage plots against the dimension. Although $h=3$ and 4 have the highest correct identification percentages, they also have high deviation and type I error rate. Therefore, **$h=5$** is a more balanced control limit for this case because it has larger correct identification percentages and lower type II error rate than those of $h=6, 7$, and 8. Although values of deviation and type I error rate at $h=6, 7$, and 8 are lower than those of $h=5$, they are close to each other.

In the case of small number of dimensions shifted with $\rho=0.5$ as shown in Figure (b), $h=4, 5$, and 6 have the top three highest percentages of correct identification. The control limit $h=4$, however, have higher values of deviation and type I error rate than those of $h=5$ and $h=6$. The differences between $h=5$ and 6 for all criteria and dimensions are small. Thus, **$h= 5$ or 6** is the recommended control limit for this case if the correct identification percentage and accuracy of where a process shift takes place are important. Note that $h=7$ and 8 are chosen due to their poor correct identification percentages 88.3% and 78.1% as oppose to those(98.3% and 95.9%) of $h= 5$ and 6.

Finally the case of $\rho=0.9$ with small number of dimensions shifted is considered, $h=5, 6, 7$, and 8 have larger correct identification percentages than that of $h=3$ and 4. Moreover, $h=7$ and 8 have very small value of deviation and type I error rate shown in Figure (c-3). However, Figure (c-4) indicates high type II error rate in 3 and 5 dimensional datasets of $h=7$ or 8 may not be acceptable. Therefore, it is suggested that **$h=5$ or 6** is used as the control limit in this case. In summary, **$h=5$** is a good choice for all dimensions and covariance structure when the number of dimensions exhibiting shifts is small.

In the case of the medium number of dimensions shifted, half of response variables contain mean shifted. The simulation results are shown in Figure . In Figure (a-1), **$h=5$** performs a better result than any others for all ρ because it has high correctness rate, small deviation, type I error rate, and type II error rate for last-in-control signal.

The final situation is when all dimensions shift. Since all mean levels of dimensions have shifted, there is no type I error occurred in this case. The computational results are shown in Figure (a-1) to (c-4). All of them show that $h=5$ has the best choice among h because of its balanced performance in terms of acceptable correct identification percentages, small deviation and type II error rate. Thus, **$h=5$** is a recommended control limit for the proposed method when large number of dimensions shifted happened. From the above analysis, **$h=5$** seems to provide the best balanced performance for all dimensions, covariance structure, and number of dimensions exhibiting shifts.

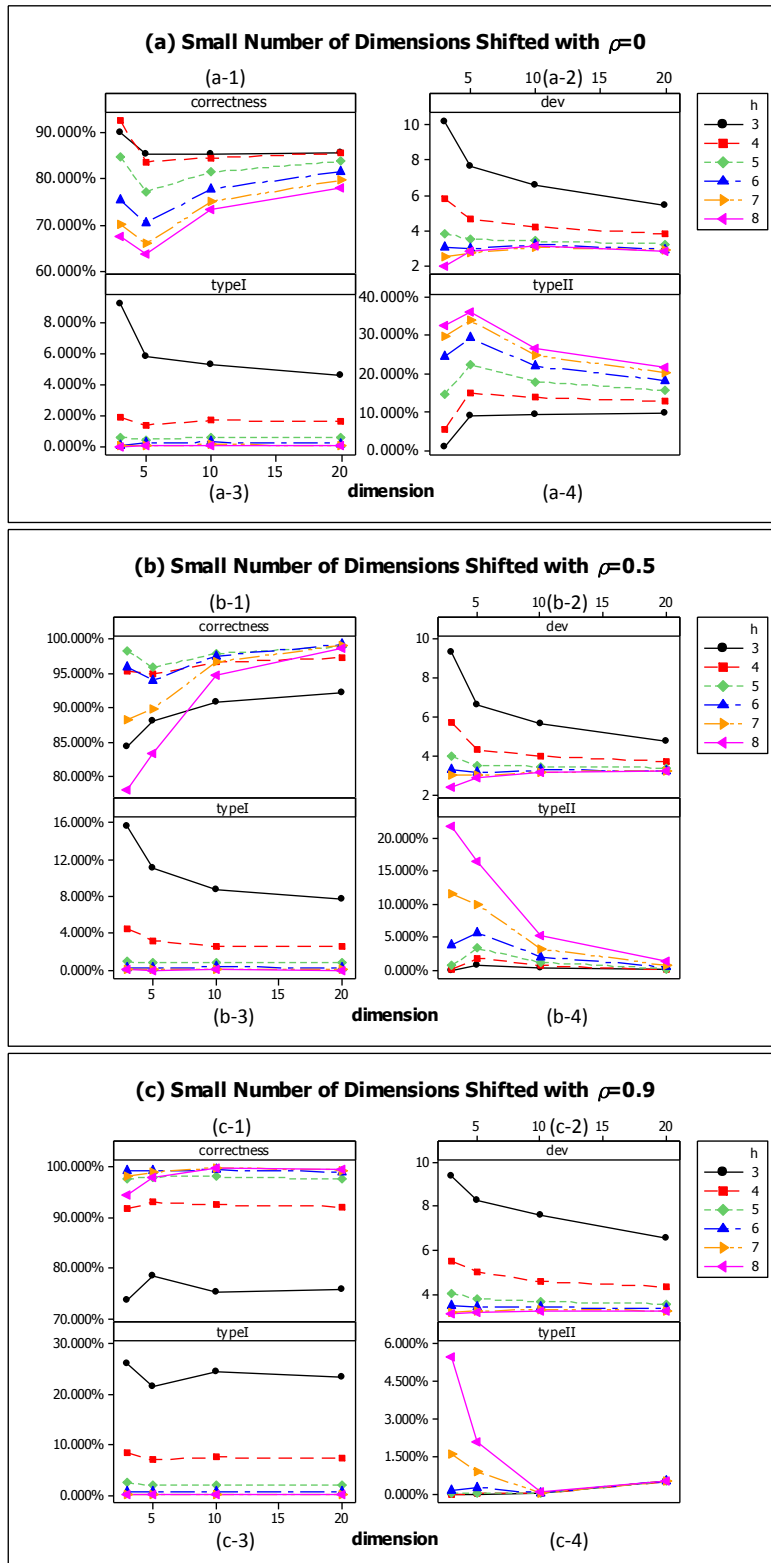


Figure .Performance statistics for small level of number of dimensions shifted with (a) $\rho=0$, (b) $\rho=0.5$, and (c) $\rho=0.9$ under different h . The y axis of plots (1) to (4) are correct classification percentage, average absolute value of deviation, type I error rate in percentage, and type II error rate in percentage, respectively. The x axis represents the numbers dimension for last-in-control location.

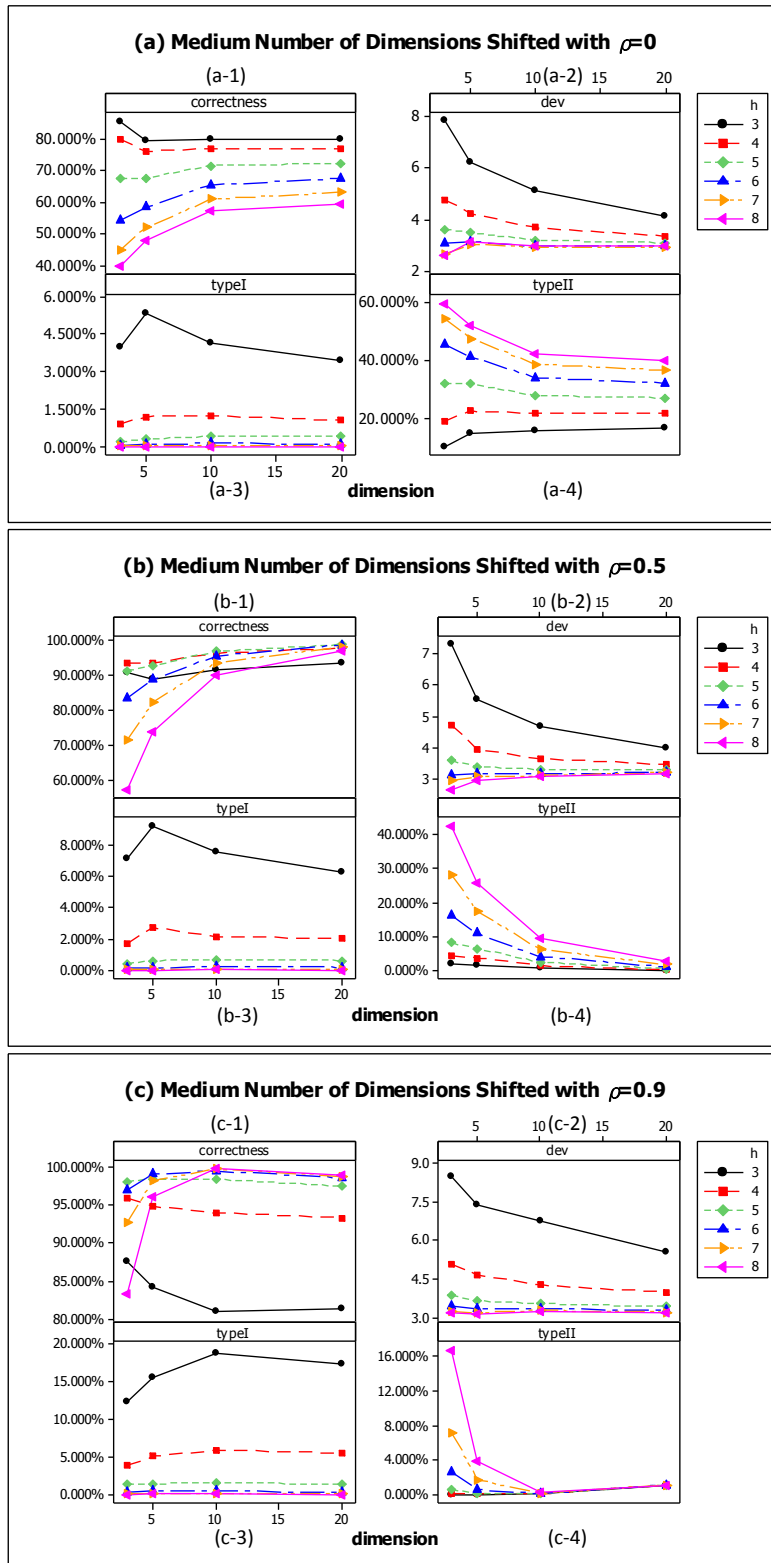


Figure . Performance statistics for medium level of number of dimensions shifted with (a) $\rho=0$, (b) $\rho=0.5$, and (c) $\rho=0.9$ under different h . The y axis of plots (1) to (4) are correct classification percentage, average absolute value of deviation, type I error rate in percentage, and type II error rate in percentage, respectively. The x axis represents the numbers dimension for last-in-control location.

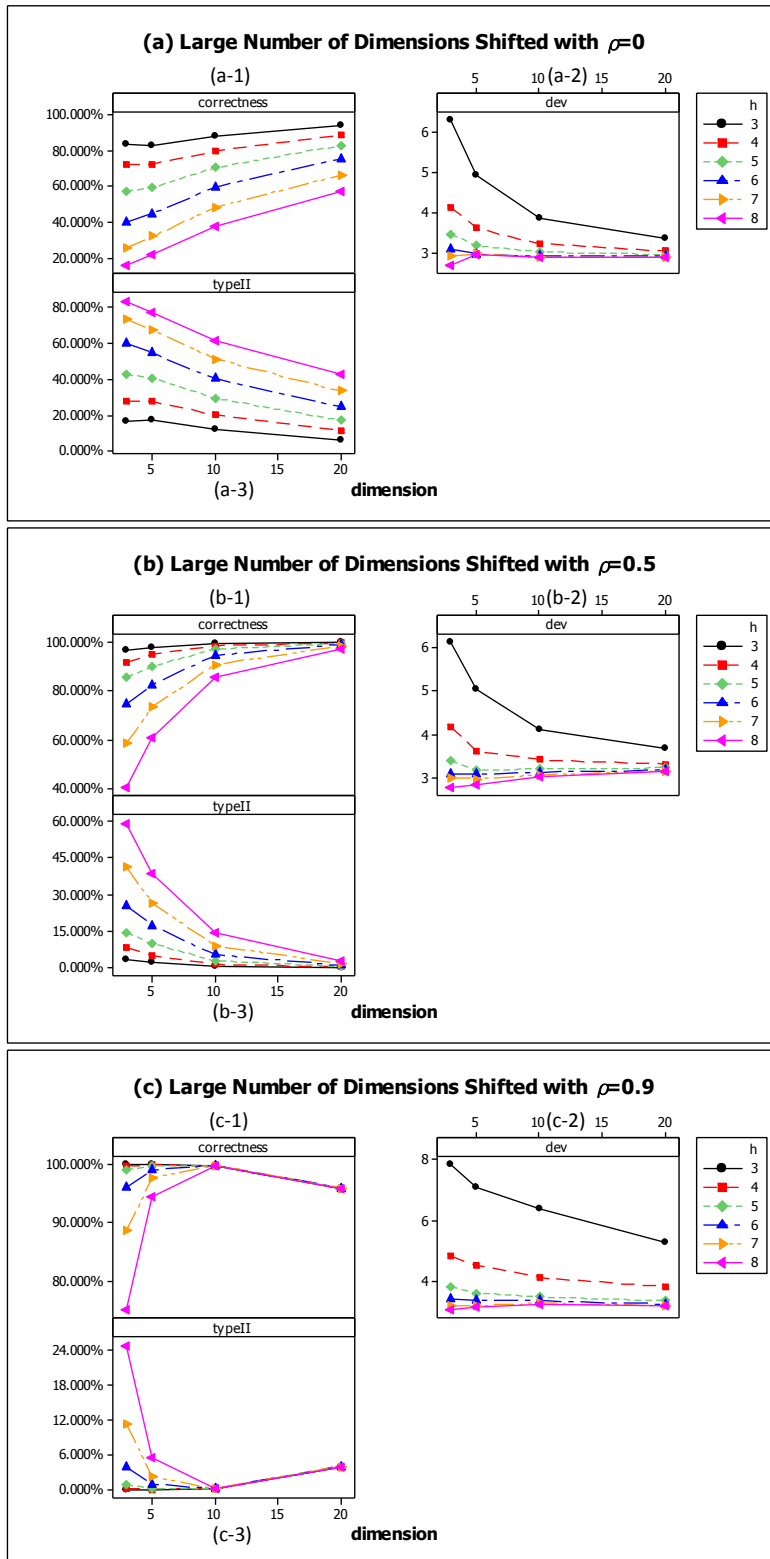


Figure . Performance statistics for large level of number of dimensions shifted with (a) $\rho=0$, (b) $\rho=0.5$, and (c) $\rho=0.9$ under different h . The y axis of plots (1) to (3) are correct classification percentage, average absolute value of deviation, and type II error rate in percentage, respectively. The x axis represents the numbers dimension for last-in-control location.

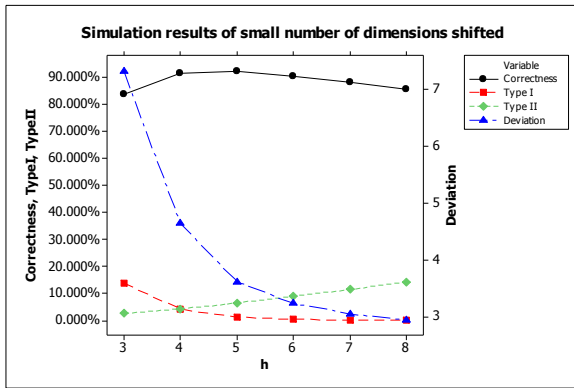


Figure . Simulation results of small number of dimensions shifted.

Table . Simulation results of small number of dimensions shifted.

h	Correct ident. %	Type I Err. Rate (%)	Type II Err. Rate (%)	Avg. abs. value of Deviation
3	83.733%	13.698%	2.569%	7.328
4	91.650%	4.127%	4.223%	4.645
5	92.446%	1.145%	6.409%	3.620
6	90.725%	0.349%	8.926%	3.238
7	88.421%	0.099%	11.480%	3.056
8	85.760%	0.030%	14.210%	2.943

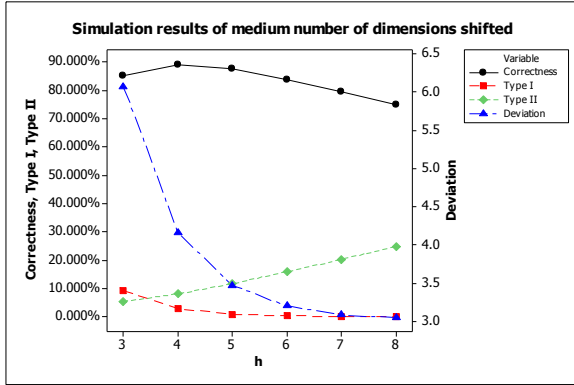


Figure . Simulation results of medium number of dimensions shifted.

Table . Simulation results of medium number of dimensions shifted.

h	Correct ident. %	Type I Err. Rate (%)	Type II Err. Rate (%)	Avg. abs. value of Deviation
3	85.394%	9.323%	5.282%	6.082
4	89.114%	2.811%	8.075%	4.164
5	87.591%	0.788%	11.620%	3.469
6	83.887%	0.223%	15.890%	3.202
7	79.725%	0.064%	20.211%	3.084
8	75.119%	0.020%	24.861%	3.046

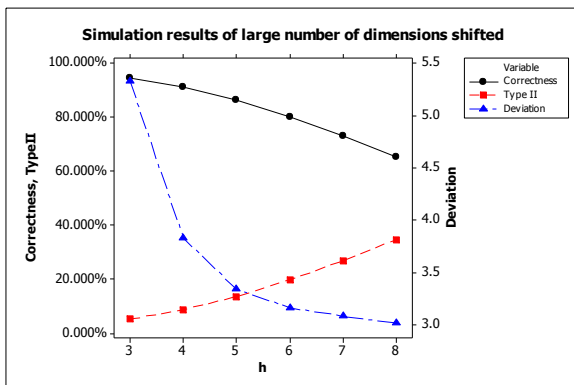


Figure . Simulation results of large number of dimensions shifted.

Table . Simulation results of large number of dimensions shifted.

h	Correct ident. %	Type II Err. Rate (%)	Avg. abs. value of Deviation
3	94.884%	5.116%	5.332
4	91.264%	8.736%	3.832
5	86.474%	13.526%	3.336
6	80.167%	19.833%	3.159
7	73.063%	26.937%	3.080
8	65.346%	34.654%	3.010

Table to 6 present the simulation results in another perspective by averaging performance statistics over all covariance structures and dimensions. For example, the correct identification percentages of $h=3$ is an average value of different types of ρ and different numbers of dimensions when $h=3$. Figure to 17 correspond to Table to 6. An h should be chosen so that the correct identification percentage is as large as possible while the rest of the performance statistics are as small as possible. Note that y axis on the left in Figure to 17 are percentages for criteria correct identification percentage, type I error rate in percentage, and type II error rate in percentage while that on the right-hand side is for the deviation from the true change point.

Figure to 17 provide an overall guideline for choosing h that is determined by three criteria – correct identification percentage, type II error rate, and average absolute deviation. An ideal choice of h provides a high correct identification percentage, a low type II error rate, and a small deviation from the true out-of-control location. However, users may not know what kind of underlying covariance structure among responses or number of responses that may have shifted in practice. These three figures suggest that $h=5$ and 6 provide the best overall balanced performance because the line segments of these three h values are very close together. Note that $h=5$ was recommended earlier from the analyses based on Figure to 14. The overall comparisons here provide further confirmation. For example, if a user chooses $h=5$ to a situation where medium number of dimensions may be likely to shift, it will provide 87.6% correct identification percentage, 0.79% type I error rate, 11.6% type II error rate, and 3.5 of its average deviation from the true spot. This user should contemplate whether $h=6$ is a better choice because the criterion values of this new choice are 83.9% correct identification percentage, 0.22% type I error, 15.9% type II error, and 3.2 as its average deviation. Do the gains in type I error (-0.57%) and average deviation (-0.3) outweigh the loss in correct identification (-3.7%) and type II error rate (+4.3%)?

Because there are multiple criteria present in this problem, a solution is generally reached by considering the trade off of all criteria. Eventually a compromised solution is generated. This type of problems often referred to as the multi-criteria decision making is beyond the scope of this work. Please refer to Triantaphyllou (2000) for more details.

5. Conclusions and Future Study

This study provides a visualization tool for making quality-related decision when the number of dimensions is large such as 20. The proposed method, the marginal CUSUM glyphs, can help users not only decompose the traditional multivariate control chart, such as, Hotelling's T^2 , multivariate EWMA, or multivariate CUSUM, but also visualize the original dataset. Unlike traditional multivariate control charts that only indicate an out-of-control condition has occurred, the proposed method provides the information which variable are responsible as well as where or when a variable is out of control. From the simulation results, the proposed method is capable of detecting mean shifts. In this study, according to the simulation results, $h=5$ and 6 are recommended to be applied to all situations regardless of the covariance structure among responses or how many of them exhibiting shifts. The tables and figures

based on the simulation results also provide a user guideline for choosing other values of h via the correct classification percentage, magnitude of deviation, and false alarm rate and type II error rate.

For future research, the following issues can be studied. First, the effect of outliers should be considered. Outliers can make the statistics of SD2CUSUM increase dramatically to cause unnecessary out-of-control signals. To deal with this issue, filtering the dataset may be necessary. Techniques of outlier identification for high dimensional dataset include MVE/MCD (Rousseeuw, 1985; Rousseeuw and van Driessen, 1999), OGK estimator (Maronna and zamar, 2002), and PCOut (Filzmoser *et al.*, 2008).

In this study, the highest number of dimension is 20. It is not a trivia task to expand the proposed method to response dimension over 100 or more. The angles between spikes for 50 and 100 dimensions are 7.2 and 3.6 respectively. The more spikes are within a circle in a star glyph the more clutter they cause. When a process is in control, cluttering is not an issue because all spikes are within the circles of control limits. Users do not need distinguish individual dimension. However, when there are many spikes that exceed control limit circles, it depends on how close these spikes are located. If they are equally spaced or far apart, the results of display are usually satisfactory. On the other hand, if all out-of-control spikes are adjacent to each other, users need to consult with the corresponding diagnostic report to make out which spikes exceed the circles. Another way to address high dimension problem is to reduce the number of dimensions, such as using principle component analysis (PCA). However, once PCA is applied to the original dataset, the original domain is transferred to a much reduced domain. It is hard to identify which original dimensions contribute to an out-of-control signal.

Finally, the situation in which the number of dimensions is larger than the number of observation is challenging. Neither can the proposed method nor traditional multivariate control charts cope with this situation because the estimate of the variance-covariance structure becomes a nontrivial task. Boyles (1996) and Chang and Ho (2001) have provided some ground work in this area.

References

- Boyles, R. A. (1996). Multivariate Process Analysis with Lattice Data. *Technometrics*, 38(1): 37-49.
- Chang, S. I and E. S. Ho (2001). Multivariate Statistical Process Control for Inspection Data from Coordinate Measuring Machines. *International Journal of Industrial Engineering – Theory, Applications and Practices*, 8(4): 347-358.
- Chen, C., Härdle, W., & Unwin, A. (2008). *Handbook of data visualization (springer handbooks of computational statistics)*. New York, NY: Springer.
- Crosier, R. B. (1988). Multivariate generalizations of cumulative sum quality-control schemes. *Technometrics*, 30(3): 291-303.
- Filzmoser, P., Maronna, R., & Werner, M. (2008). Outlier identification in high dimensions. *Computational Statistics and Data Analysis*, 52(3): 1694-711.
- Hartigan, J. A., & Kleiner, B. (1984). A mosaic of television ratings. *The American Statistician*, 38(1): 32-35.
- Hawkins, D. M. (1991). Multivariate quality control based on regression-adjusted variables. *Technometrics*, 33(1): 61-75.
- Hawkins, D. M. (1993). Cumulative sum control charting: An underutilized SPC tool. *Quality Engineering*, 5(3): 463-477.
- Jackson, J. E. (1985). Multivariate quality control. *Communications in Statistics - Theory and Methods*, 14(11): 2657-2688.
- Lowry, C. A., Woodall, W. H., Champ, C. W., & Rigdon, S. E. (1992). A multivariate exponentially weighted moving average control chart. *Technometrics*, 34(1): 46-53.
- Lowry, C. A. & Montgomery, D. C. (1995). A Review of Multivariate Control Charts. *IIE Transactions*, 27(6): 800-810.
- Maronna, R. A. & Zamar, R. H. (2002). Robust estimates of location and dispersion for high-dimensional datasets. *Technometrics*, 44(4): 307-317.
- Mason, R. L., Tracy, N. D., and Young, J. C. (1995). Decomposition of T2 for multivariate control chart interpretation. *Journal of Quality Technology*, 27(2): 99-108.
- Montgomery, D. C. (2008). *Introduction to statistical quality control*. New York, NY: John Wiley & Sons.

Pignatiello, J. J., Jr., & Runger, G. C. (1990). Comparisons of multivariate CUSUM charts. *Journal of Quality Technology*, 22(3): 173-86.

The R project for statistical computing. Star (Spider/Radar) Plots and Segment Diagrams.
<http://wiki.r-project.org/rwiki/doku.php?id=rdoc:graphics:stars>

Runger, G. C., Alt, F. B. and Montgomery, D. C. (1996). Contributors to a Multivariate Statistical Process Control Signal. *Communications in Statistics—Theory and Methods*. 25(10): 2203-2213.

Rousseeuw, P. J. (1985). Multivariate estimation with high breakdown point. *Mathematical Statistics and Applications*, 8, 283-297.

Rousseeuw, P. J., van Driessen, K. (1999). A fast algorithm for the minimum covariance determinant estimator. *Technometrics*, 41(3): 212-223.

Sarkar, D. (2008). *Lattice: Multivariate data visualization with R (use R)*, 1st ed., New York, NY: Springer.

Triantaphyllou, E. (2000), *Multi-Criteria Decision Making Methods: A Comparative Study*. Boston, MA: Kluwer Academic Publishers.

Yates, F. (1935). Complex Experiments. *Journal of the Royal Statistical Society (Supplement)*, 2: 181-247.

Woodall, W. H., & Ncube, M. M. (1985). MULTIVARIATE CUSUM QUALITY-CONTROL PROCEDURES. *Technometrics*, 27(3), 285-292.