Research Article

Quality and Reliability Engineering International

(wileyonlinelibrary.com) DOI: 10.1002/qre.1448

Published online 19 September 2012 in Wiley Online Library

Robust Multivariate Control Chart for Outlier Detection Using Hierarchical Cluster Tree in SW2

Shu-Kai S. Fan,[a*†] Hui-Kuo Huang[b] and Yuan-Jung Chang[b]

The goal of this paper is to develop a new multivariate control chart that can effectively detect potential outlier(s) in multidimensional data while keeping the masking and swamping effects under control. The hierarchical clustering tree plays a central role in the proposed control chart, in an attempt to improve the Sullivan and Woodall's second method, known as the SW2 method. Historical multivariate datasets taken from the literature are used as the benchmarks to illustrate the performance of the proposed control charts in comparison to nine existing methods for outlier detection. The two criteria, the masking and swamping rates, are used as yardsticks for the evaluation purpose. An additional simulation study by means of Monte Carlo experiments further verifies that the proposed control chart that incorporates the hierarchical clustering tree performs much better in outlier detection and swamping prevention than the original SW2 and minimum volume ellipsoid methods. Copyright © 2012 John Wiley & Sons, Ltd.

Keywords: outlier detection; multivariate control chart; masking effect; swamping effect; hierarchical cluster tree

1. Introduction

Outlier detection in the dataset plays an important role in many practical applications. The dataset may contain outliers with abnormal values irregularly larger or smaller than the normal observations. Outliers may cause misjudgment on analyses in such applications as, for example, regression, analysis of variance, statistical process control (SPC) and profile monitoring. Such points may contain some insightful information on the abnormal behavior of the system concealed in the dataset. Thus, outlier detection is an important task prior to processing data analysis. Historical data often suffer from transcription errors and problems in data quality. These errors make historical data prone to outliers. Hawkins[1] provided a definition of an outlier; he mentioned that 'an observation which deviates so much from other observations as to arouse suspicions that it was generated by a different mechanism'. In some cases, the outliers can be detected by residuals that are considerably larger in absolute value than the others; say, three or four standard deviations from the mean (Montgomery *et al.*[2]).

The same idea is also considered in control charts. In SPC, control charts, such as Shewhart charts, are powerful tools for determining whether or not a process is in a state of statistical control. The Shewhart charts were developed by Walter A. Shewhart in the early 1920s. In multivariate SPC, Hotelling[3] worked in multivariate control charts and applied his procedure, called Hotelling's T^2, to sample bombsight data during World War II. The main objective of a multivariate control chart is to detect the presence of assignable causes of variation for multivariate quality characteristics. In particular, for a retrospective phase I analysis of a historical dataset, the objective is twofold: (i) to identify shifts in the mean vector that might distort the estimation of the in-control mean vector and variance-covariance matrix; (ii) to identify and eliminate multivariate outliers. The purpose of seeking an in-control subset of historical dataset is to estimate in-control parameters for use in a phase II analysis (Williams *et al.*[4]).

In one-dimensional data, outlier detection is simple as an extreme observation detected by residual larger in absolute value than the others. In bivariate case, the suspected outlier may be spotted without undue difficulty when the data are represented in a scatter plot. In multivariate data, when the dataset has more than two dimensions, outlier detection is not feasible by these intuitive methods. Hotelling's T^2 and principal component analysis (PCA) are the two most widely recognized multivariate methods applied to the control chart in high dimensional dataset. Relevant research regarding multivariate control charts will be briefly discussed shortly.

[a]Department of Industrial Engineering and Management, National Taipei University of Technology, 1, Sec. 3, Chung-Hsiao E. Road, Taipei City, Taiwan 106, ROC
[b]Department of Industrial Engineering and Management, Yuan Ze University, 135, Yuan-Tung Rd., Chung-Li City, Taoyuan County, Taiwan 320, ROC
*Correspondence to: Shu-Kai S. Fan, Department of Industrial Engineering and Management, National Taipei University of Technology, 1, Sec. 3, Chung-Hsiao E. Road, Taipei City, Taiwan 106, ROC.
†E-mail: morrisfan@ntut.edu.tw

Both Hotelling's T^2 and PCA have drawn considerable attention in the past a few decades. When a few outliers are hidden in the dataset, Hotelling's T^2 and PCA might detect the outliers well. Nevertheless, when several outliers are hidden in the dataset, Hotelling's T^2 and PCA might fail to detect them. In such cases, the analysis may lead to the biased estimation of the unknown parameters. To circumvent such situations, several methods have been discussed to adjust the Hotelling's T^2 and PCA procedures to make them insensitive to outliers. These methods include the trimming method by improving Hotelling's T^2 (Langenberg and Iglewicz[5]; Rocke[6]; Rock[7]), the successive difference method (Sullivan and Woodall[8] called herein as the Sullivan and Woodall first approach, henceforth referred to as SW1) and the stalactite plot (Atkinson and Mulira,[9] referred to as the Sullivan and Woodall second approach, henceforth referred to as SW2). These methods are frequently assessed by two metrics: detection rate (rate in which an outlier is correctly classified as an outlier) and swamping rate (rate in which an inlier is misclassified as an outlier). The aforementioned methods can actually improve Hotelling's T^2 but are still not effective in detecting numerous outliers being present in the data. Rousseeuw[10] and Rousseeuw and Leroy[11] proposed two algorithms that can deal with the issue of detecting a large number of outliers, the minimum volume ellipsoid (MVE) method and the minimum covariance determinant (MCD) method. However, both methods demand enormously heavy computation. Many discussions pertaining to the approximation of MVE and MCD have been made in the past two decades.

This paper aims to remedy the drawbacks of the existing methods. The drawbacks can be alleviated by developing new techniques to simultaneously improve the estimation robustness and also reduce the heavy computation effort incurred. Two important issues that should be taken into account are as follows: (i) reducing computation effort may lead to ineffective detection for a large number of outliers existing in the dataset (as the trimming method, SW1 and SW2); (ii) effective detection for a large number of outliers frequently requires heavy computation (such as MVE and MCD). Most research in this area is focused on the computation time of the second issue. The researchers have proposed different approaches to approximate the MVE and MCD with much improved computation time, at least better than the original version. However, comparing to the trimming method, those approaches are still time-consuming in a practical sense. The number of iterations largely affects the accuracy of the approximation. If the number of iteration is not sufficient, the approximation might not be stable in estimation accuracy.

The rest of the paper is as follows. Section 2 gives the literature review, including the definition of outliers and several multivariate control charts that will be investigated for the experimental study. In Section 3, the shortcoming of SW2 is first investigated, followed by the concept of the unsupervised hierarchical clustering method that will be incorporated into SW2. Section 4 discusses the source of the historical datasets first, and then the performance of the proposed method is assessed by comparing its computational results to those of the existing methods reviewed in Section 2. An additional experimental study via Monte Carlo simulation is conducted to verify the effectiveness of the proposed method in comparison to the original SW2 and MVE methods. Section 5 summarizes the findings of the paper.

2. Literature review of outlier detection methods

In this section, several noted multivariate control charts designed for outlier detection will be reviewed briefly and used as benchmarks to assess the performance of the proposed method. The control chart methods include Hotelling's T^2, trimming T^2, SW1, SW2, MVE, MCD, PCA and robust estimation for PCA.

2.1. The Hotelling's T^2 control chart

The most well-known multivariate control chart for monitoring the process variation is the Hotelling's T^2 control chart. The Hotelling's T^2 statistic measures the Mahalanobis distance of the p-dimensional observation vector $\mathbf{X}_i$ from the sample mean vector. The general form of the statistic is given by

$$T_i^2 = (\mathbf{x}_i - \overline{\mathbf{x}})S^{-1}(\mathbf{x}_i - \overline{\mathbf{x}})', \quad i = 1, 2, \cdots, m, \tag{1}$$

where

$$\overline{\mathbf{x}} = \frac{1}{m}\sum_{i=1}^{m}\mathbf{x}_i \text{ and } S = \frac{1}{m-1}\sum_{i=1}^{m}(\mathbf{x}_i - \overline{\mathbf{x}})(\mathbf{x}_i - \overline{\mathbf{x}})',$$

where m is the sample size of observation vectors. Tracy *et al.*[12] considered a process which was in the start-up stage with individual observations. They presented an exact method based on the beta distribution, and its upper control limit (UCL) was given by

$$UCL = \frac{(m-1)^2}{m} Beta\left(1 - \frac{\alpha}{2}; \frac{p}{2}, \frac{m-p-1}{2}\right). \tag{2}$$

In their work, the control limit and the results were obtained by approximating the beta distribution with appropriate F and chi-square distributions.

2.2. Trimming T^2 control chart

Trimming has been successfully applied in univariate control charts (Langenberg and Iglewicz[5]; Rocke[6]; Rock[7]). However, the trimming technique for the multivariate control chart is not practically straightforward. The idea is to compute the Mahalanobis distance for each observation vector, then delete points with the largest distances and recompute the sample mean vector and the sample covariance matrix with the remaining observation vectors. This approach works well when few outliers exist in the data, but suffers from the masking effect. Rousseeuw and Leroy[11] and Vargas[13] suggested a better alternative for trimming. Their methods resemble the original one, but delete a fixed percentage of the number of observation vectors. Rocke[6] suggested that 25% of the observations should been excluded; Stefatos and Hamza[14] excluded three observations in three examined examples. The trimming technique was shown to be efficient only when a small amount of outliers were present in the data.

2.3. SW1

Sullivan and Woodall[8] employed the vector difference between successive observations to estimate the in-control covariance matrix of the process. The definition is described as follows. Let $\mathbf{V}_i$ be the difference vector between successive observation vectors,

$$\mathbf{v}_i = \mathbf{x}_{i+1} - \mathbf{x}_i \quad i = 1, 2, \ldots, m-1. \tag{3}$$

The estimator of the covariance matrix is defined by

$$S_{SW1} = \frac{1}{2(m-1)} \sum_{i=1}^{m-1} \mathbf{v}_i \mathbf{v}'_i. \tag{4}$$

The approach based on $\overline{\mathbf{X}}$ and $\mathbf{S}_{SW1}$ is called SW1. This method is not effective in detecting a large number of outliers (Vargas[13]).

Williams *et al.*[4] studied how to accurately determine the UCL for the T^2 control chart based on successive differences of multivariate individual observations. The exact distribution for the resulting T^2 statistics had not been thoroughly investigated when the successive difference estimator was used. They demonstrated several useful properties of the T^2 statistics based on the successive difference estimator and gave a more accurate approximate distribution for calculating the UCL for individual observation vectors. For small sample sizes, their method can still find an appropriate UCL for phase I analysis as long as $m > p$. However, if $m < p$, the covariance matrix becomes a singular matrix, and the inverse of covariance does not exist. They also gave useful information for the number of samples required for the SW1 T^2 chart. When $m > p^2 + 3p$, using the chi-square approximate will generate satisfactory results for the SW1 T^2 control chart. The corresponding control limit for SW1 isIf $m > p^2 + 3p$, they suggested using an alternative approach of dividing each T^2 statistic by its true maximum value, thus constraining it to fall between zero and one so that a beta distribution may be a more accurate approximate distribution.

2.4. SW2

Sullivan and Woodall[8] brought up the second approach modified from the stalactite plot by Atkinson and Mulira.[9] The stalactite plot produces a visual display requiring insightful interpretation. It can be modified so as to provide a numerical threshold for an out-of-control signal. The Atkinson and Mulira's multiple-step method starts by randomly selecting $p+1$ observations to calculate the sample mean vector and the covariance matrix. The resulting estimates are used to calculate the m Mahalanobis distances. Then, the $p+2$ observation vectors with the smallest distances are selected to calculate new estimates of the sample mean vector and the covariance matrix. The procedure continues adding one observation vector each time until a fixed amount of observation vectors are included in the final subset. During the selection procedure, the outliers tend to stand out with a large Mahalanobis distance up to the point where the subset includes all observations except the outliers. Sullivan and Woodall[8] provided a decision rule such that the method terminates before the final step is reached. In their research, for the case where $m = 30$, they recommended the forward search until 25 observation vectors are included in the subset; that is, to use approximately 83% of the entire dataset for the estimates. Last, this subset is used to estimate the sample mean vector and the sample covariance matrix. Note again that this technique remains vulnerable to data that contains a large number of outliers and also depends on the robustness of its initial random sample (Vargas[13]), so SW2 might also be not effective in detecting a large number of outliers as SW1.

2.5. MVE

The MVE was the first high breakdown robust estimation method of multivariate location and dispersion advocated by Rousseeuw[10] and Rousseeuw and Leroy.[11] The MVE covers at least half of the m observation vectors to construct robust estimators. Subsets of size h are called half-sets because h is often chosen to be just more than half of the m data points. Rousseeuw and Zomeren[15] used the integer value of $h = \frac{m+p+1}{2}$ for the MVE estimator to achieve the highest breakdown point possible. The location estimator is the mean vector of the smallest ellipsoid of a particular half-set, and the dispersion estimator is the covariance matrix of those observations within the ellipsoid.

The estimators tend to be unbiased when the observation vectors obey multivariate normal distribution (Vargas[13]). The advantage of the MVE estimator is that it has a breakdown point of approximate 50% (Lopuhaä and Rousseeuw[16]). However, it requires heavy computation and may not be realistically calculable in some practical circumstances. For an $m \times p$ data matrix, it needs to compute the volumes of $\frac{m!}{h!(m-h)!}$ ellipsoids and to select the ellipsoid with the minimum volume. For instance, for a dataset with the observations $m = 30$, the variables $p = 3$, and the integer value $h = \frac{30+3+1}{2} = 17$, there are $\frac{30!}{17!13!} = 119,759,850$ ellipsoids (Hadi[17] and Jensen *et al.*[18]).

To cope with the heavy computation of the MVE, several algorithms have been suggested for approximation. One of the algorithms is the resampling algorithm proposed by Rousseeuw and Leroy.[11] The resampling algorithm repeats the following procedure n times as n subsamples are considered:

1. Take a fixed number of random subsamples J with $(p+1)$ observation vectors, then calculate the sample mean vector and the sample covariance matrix as

$$\bar{\mathbf{x}}_J = \frac{1}{p+1}\sum_{i\in J}\mathbf{x}_i \text{ and } S_J = \frac{1}{p}\sum_{i\in J}\left(\mathbf{x}_i - \bar{\mathbf{x}}_J\right)\left(\mathbf{x}_i - \bar{\mathbf{x}}_J\right)'. \tag{6}$$

2. The corresponding ellipsoid is then inflated or deflated to contain exactly h observation vectors, which corresponds to computing the Mahalanobis distance d_J^2 and the ellipsoid volume V_J (Rousseeuw and Zomeren[15]) as below

$$d_J^2(i) = \left(\mathbf{x}_i - \bar{\mathbf{x}}_J\right)' S_J^{-1}\left(\mathbf{x}_i - \bar{\mathbf{x}}_J\right) \text{ and } V_J = \left(m_J^2\right)^p \det(S_J), \tag{7}$$

where m_J^2 is the h^{th}-order statistic of the $d_J^2(i)$. Usually, $h = \left\lfloor \frac{m+p+1}{2} \right\rfloor$, where $\lfloor a \rfloor$ represents the largest integer not greater than a.

3. Keep the subsample J^* for which V_J has the minimum volume across all the n replications. The MVE estimators are

$$\bar{\mathbf{x}}_{MVE} = \bar{\mathbf{x}}_{J_*} \text{ and } S_{MVE} = c_{m,p}^2\left(\chi_{p,0.5}^2\right)^{-1} m_J^2 S_{J_*}, \tag{8}$$

where $c_{m,p}^2$ is a correction factor for small samples, and $\chi_{p,0.5}^2$ is the median of the chi-squared distribution with p degrees of freedom. Rousseeuw and Zomeren[15] found a reasonable small-sample correction factor via the simulation study as given by:

$$c_{m,p}^2 = \left(1 + \frac{15}{m-p}\right)^2. \tag{9}$$

4. Repeating the MVE algorithm with random subsamples of $n = 30{,}000$ could be readily handled in a short period of time with a standard contemporary personal computer (Rousseeuw and van Aelst[19]).

2.6. MCD

The MCD is one of the affine equivariant and highly robust estimators of multivariate location and dispersion proposed by Rousseeuw[10] and Rousseeuw and Leroy.[11] The MCD also covers at least half of the observation vectors m to construct robust estimators and uses the same size of half-sets as MVE. The integer value of $h = \frac{m+p+1}{2}$ is used to achieve the high breakdown points possible for the MCD estimators. The minimum value of the determinant of the covariance matrix by the half-set is returned. The location and dispersion estimators are the sample mean vector and the covariance matrix of the points that are from the half-set. The MCD estimates have the same maximum breakdown point with MVE. In addition, the MCD estimates can be very computationally difficult to obtain as the numbers of samples m and dimensions p increase. As a result, the approximate methods and algorithms to obtain MVE estimates can also be used to obtain the MCD estimates.

The common hybrid algorithm based on the MCD proposed by Rousseeuw and van Driessen[20] is called the FAST-MCD algorithm. The FAST-MCD method can handle large datasets within a reasonable amount of time. For smaller datasets (all with $m \leqslant 75$ and $p \leqslant 5$), the FAST-MCD algorithm generated estimates that were equivalent to the exact MCD estimates. They mentioned that the MCD has several advantages over the MVE, but the MCD has been rarely used due to its computation difficulty in large datasets. They constructed the FAST-MCD algorithm whose computation time is actually much faster than any MVE algorithm, and the new algorithm can deal with a sample size in tens of thousands. The basic idea behind FAST-MCD uses a partitioning of the dataset to avoid doing all the calculations in the entire data, and the techniques are called 'selective iteration' and 'nested extension'. They presented two important applications with large sample sizes to show that FAST-MCD renders more accurate results than existing algorithms and is faster in computation time by orders of magnitude.

2.7. PCA

PCA is an explanatory technique whose objective is to reduce the dimensionality of a dataset while retaining as much information as possible to the inherent variability within the data (Stefatos and Hamza[14]). Principal components are linear transformations of the original dataset, and the components are uncorrelated and ordered so that the first few components contain nearly as much information as in the original dataset (Jolliffe[21]). As discussed in Gnanadesikan and Kettenring,[22] the outliers detectable from a plot of the first few principal components inflate with variance and covariance. If an outlier causes a large dispersion, then it must be extreme on those variables and thus detectable by visualizing the scatter plots of one or more original variable. On the contrary, the last few principal components represent linear transformations of the original variables with the minimum variance. These components are sensitive to the observations that are inconsistent with the correlation structure of the data but are not outliers with respect to the original variables.

In the principal component control chart, the principal component on correlation is often used, and each principal component is expressed as a linear combination of original standardized variables. PCA has two main advantages over other multivariate SPC techniques. First, the principal components are uncorrelated, and second, only a few components are required in order to capture most of the variance (Yang and Trewn[23]). The control limits for the principal component control chart are easily computed because the variance of the component is equal to the eigenvalue λ. If the standardized principal component score is divided by $\sqrt{\lambda}$, then the standardized principal component score should follow a standard normal distribution. The corresponding three-sigma control limit is simply ± 3 (UCL = 3, LCL = −3), and the centerline is equal to zero.

3. Proposed multivariate control chart

In SW2, the first step is to randomly pick a subset for estimating the sample mean vector and the sample covariance matrix. If the selected subset contains any outlier, the estimates will be biasedly inflated, and then the performance on outlier detection will be heavily influenced by the masking effect. In other words, the procedures may fail to detect multiple outliers due to the masking effect. Meanwhile, labeling “good” data as outliers is commonly referred to as the swamping effect. Since the parameters are incorrectly estimated at the outset, the parameter estimation in the subsequent iterations will lead to mistakes, thus causing the masking and/or swamping effects. As mentioned by Vargas,[13] SW2 is efficient when a small amount of outliers are hidden in the data. Additionally, the initial sample seriously influences the SW2 method. In this light, in order to alleviate the masking and swamping effects at the same time, the proposal is to screen the original dataset for potential outliers prior to the SW2 method. If the potential outliers can be correctly filtered from the original dataset, an improvement on the performance of the original SW2 method is anticipated even if a number of outliers are present in the dataset. Here, we focus on using clustering approaches as the initial outlier filter. The reason is twofold: first, the mean vector and the variance-covariance matrix of the dataset under study will be iteratively estimated in the SW2 method, so we do not intend to double our effort in estimating these parameters. Second, the procedures of clustering approaches are typically more computationally efficient than control charting approaches.

In an attempt to cull the outliers from the dataset, this paper proposes using hierarchical clustering procedures for preliminary screening. Hierarchical clustering tends to be one of the widely recognized algorithms that identify some groupings or clusterings of the objects under investigation that best represent certain empirically measured relations of similarity (Webb[24]). Its popularity may be attributed to the fact that clustering methods presume very little in the way of data characteristics or *a priori* environment knowledge (Murtagh[25]). A dichotomous clustering, dividing data into inliers and outliers, will suffice for the conditions stated herein. A hierarchical cluster tree is a nested set of partitions represented by a tree diagram, frequently referred to as the dendrogram. Sectioning a tree at a particular level produces a partition into several disjoint groups. An agglomerative algorithm starts with m subclusters each containing a single data point, and at each stage merges the two most similar groups to form a new cluster, thus reducing the number of clusters by one. The algorithm halts until all the data fall within a single cluster. In order to make the algorithm more concrete, the meaning about the two most similar clusters should be defined. The similarity is calculated by a distance matrix. There are several distance functions calculating the distance between two clusters. Choosing different distance functions may result in different tree structures. The commonly used distance functions include the single link method, complete link method, centroid distance, median distance, group average link method and sum of squares method.

In this paper, the single link method is selected for outlier screening prior to the SW2 method. The single link method is deemed one of the conventional methods of cluster analysis. Single link clustering defines the distance between two clusters as the minimum Euclidean distance between their members. It is called ‘single link’ because it says clusters are close if they have even a single pair of close points, a single ‘link’. It is of particular importance to note that the covariance structure between variables is ignored while the distance between any two clusters of observation vectors is evaluated. The single link method can handle quite complicated cluster shapes. In the application of outlier detection, we only aim to separate the dataset into two clusters, one for the inliers and the other for outliers. The major purpose of the preliminary filter is to exclude most severe potential outliers from the dataset that will be used for sampling in the SW2 method. The chaining effect conveyed by the single link method shows an advantage of safeguarding the SW2 method against the swamping effect. Hence, the single link method is considered a logical linkage choice for this situation. The authors also conducted an earlier screening experiment to test seven different link methods for potential outlier detection and found that the single link method shows great promise in this application. For interested readers, the experimental results are available upon request from the first author. Assume that two objects a and b belong to the same single link cluster at level d. If there exists a chain of intermediate objects $i_1, \ldots, i_{m-1}$, linking them generates all the distances

$$d_{i_k, i_{k+1}} < d \text{ for } k = 0, \ldots, m-1, \tag{10}$$

Where $i_0 = a$ and $i_m = b$. The distance between two groups A and B is the distance between their closest members, i.e.

$$d_{AB} = \min_{i \in A, j \in B} d_{ij}. \tag{11}$$

The hierarchical cluster tree by means of the single link method is incorporated into SW2, referred to as HSW2 henceforth. The hierarchical cluster tree is utilized to classify the dataset into two groups, the normal dataset and the outlying dataset. The outlying dataset is eliminated and then the SW2 method is implemented. The flow chart of HSW2 is shown in Figure 1. Following the SW2 method, a subset containing $p + 1$ observation vectors is randomly sampled from the screened group generated by using the

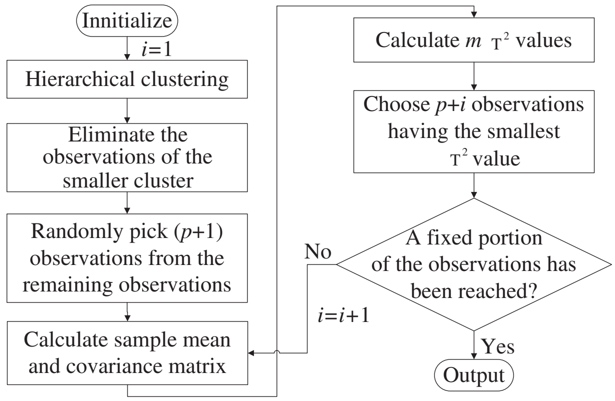

Figure 1. The flow chart of HSW2

hierarchical cluster tree to calculate the sample mean vector and the covariance matrix. Subsequently, the T^2 statistic of the current dataset is computed to choose the minimum p +2 observation vectors. The remaining steps go on as the SW2 method mentioned previously. In what follows, two experimental results are presented to illustrate the proposed method, one based on the four well-known datasets taken from the literature and the other based on Monte Carlo simulations.

4. Computational experience

In this section, four multivariate historical datasets will be used as the test-bed to assess the outlier detection capability among various methods. They include Hotelling's T^2 (Hotelling[4]), trimming T^2 (Rousseeuw and Leroy,[11] with trimming rate equal to 15% given by Stefatos and Hamza[14]), SW1 (Sullivan and Woodall[8]), SW2 (Sullivan and Woodall,[8] with 85% of the entire observations given by Stefatos and Hamza[14]), MVE (Rousseeuw,[10] with 30,000 iterations suggested by Rousseeuw and van Aelst[19]), MCD (Rousseeuw,[10] with 500 iterations suggested by Rousseeuw and van Driessen[20]), PCA, PCAMVE, PCAMCD (Stefatos and Hamza[14]). Two measures, the detection and swamping rates, are used as the yardstick for the performance evaluation. The detection rate is computed as the ratio of the number of correctly classified outliers by the methods to the total number of the outliers in the dataset. The swamping rate is computed as the ratio of the number of misclassified inliers by the methods to the total number of the inliers in the dataset. The computational results of these ten methods are reported by taking the average of 1000 simulation runs. To further verify the effectiveness of the proposed method, a Monte Carlo simulation study of the HSW2 method is conducted as compared to the original SW2 and MVE methods in a later part of Section 4.

4.1. *Woodmod dataset*

The wood gravity data was used in Draper and Smith[26] to determine the influence of anatomical factors on wood specific gravity. It was revisited and modified by Rousseeuw and Leroy[11] called woodmod which contained 20 observations. Each observation vector has five variables as:

x_1: number of fibers per square millimeter in Springwood
x_2: number of fibers per square millimeter in Summerwood
x_3: fraction of Springwood
x_4: fraction of light absorption by Springwood
x_5: fraction of light absorption by Summerwood.

Rousseeuw and Leroy,[11] Atkinson and Mulira,[9] Sebert *et al.*,[27] Bartkowiak[28] and Kim and Krzanowski[29] had analyzed the woodmod dataset and reported observations 4, 6, 8 and 19 as outliers, while Alqallaf *et al.*[30] and Stefatos and Hamza[14] viewed observations 7, 11 and 16 as additional faults. To make a fair comparison in this paper, only observations 4, 6, 8 and 19 are considered outliers here. The computational results of the ten methods for the woodmod dataset are presented in Table I.

In the woodmod dataset, the real outlier ratio is 20% (4 out of 20). Based on the detection rate results, the MCD and PCAMCD methods take the lead of having a perfect detection rate (100%). The proposed HSW2 method is ranked the second with a detection rate of 98.90%, indicating only 44 out of 4000 instances of outlier detection were masked. The original SW2 only renders a low detection rate of 23.20%; the original T^2 and PCA methods cannot even detect any outlier. Regarding the swamping rate results, except the original T^2 and PCA methods, the HSW2 method yields the lowest swamping rate of 0.20% among the compared methods, meaning that only 32 out of 16,000 instances of inlier evaluation are mistakenly deemed as outlier. The MVE and PCAMVE methods are ranked the second with a swamping rate around 3.25%, followed by the original SW2 method with 14.09%. The remaining methods suffer severely from the swamping effect, having a swamping rate up to 18.75%, corresponding to 3000 inliers that are incorrectly identified as outlier. As can be evidently seen from Table I, the proposed HSW2 method greatly improves the performances of the original SW2 method on both the detection and

Table I. The computational results of the 10 methods for the woodmod dataset

Methods	T^2	Trimming	SW1	SW2	HSW2
Detection rate	0.00%	0.00%	0.00%	23.20%	98.90%
Swamping rate	0.00%	18.75%	18.75%	14.09%	0.20%
Methods	MVE	MCD	PCA	PCAMVE	PCAMCD
Detection rate	74.40%	100.00%	0.00%	74.40%	100.00%
Swamping rate	3.26%	18.75%	0.00%	3.25%	18.75%

swamping rates for the woodmod dataset. In terms of these two measures, the HSW2 method, overall, has an edge over the other nine methods.

To gain more insight into the comparison results, Figure 2 shows the control charts of the discussed methods from a single simulation. Note that, to simplify matters, the result of the original SW2 method is suppressed, and the SW2 result shown here corresponds to the proposed HSW2 method. Clearly from the figure, the original T^2, SW1 and PCA methods are adversely subjected to the masking effect; the trimming and SW1 methods are strongly subjected to the swamping effect. The MVE, MCD, PCAMVE and PCAMCD methods swamped one (observation 11) and two (observations 11 and 16) inliers, respectively.

4.2. Phosphorus content dataset

An investigation of the source from which corn plants obtain their phosphorus was carried out by chemically determining the concentrations of inorganic and organic phosphorus in the soils. The phosphorus content of the corn grown in these soils was measured. Eighteen soil samples were used in the experiment from Snedecor and Cochran.[31] Each observation contains three variables as follows:

x_1: inorganic phosphorus in soil
x_2: organic phosphorus in soil
y: plant phosphorus.

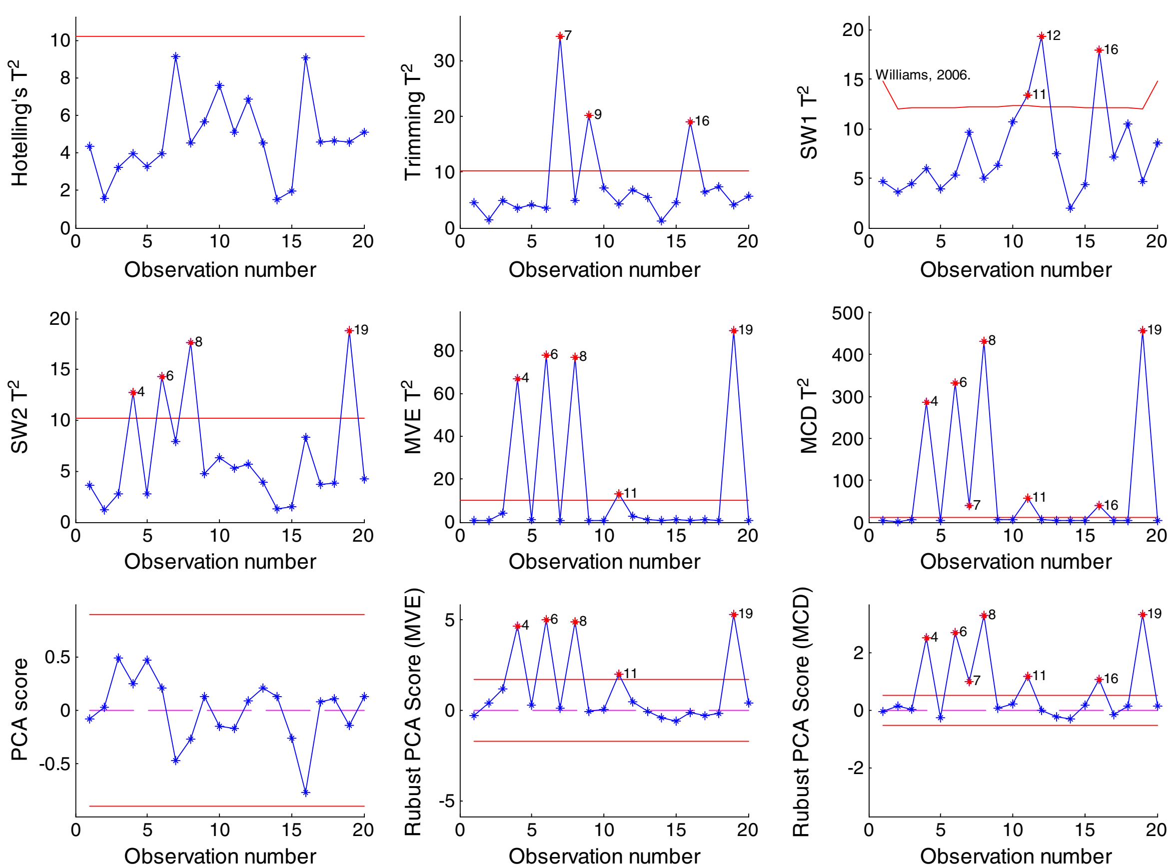

Figure 2. The control charts of the studied methods for the woodmod dataset

Prescott[32] and Lund[33] investigated that observation 17 is claimed an outlier; Rousseeuw and Zomeren[34] viewed observation 10 also an outlier. In this paper, only observation 17 is considered as an outlier. The computational results of the ten methods for the phosphorus content dataset are displayed in Table II.

In dataset phosphorus content, all the methods except PCA and PCAMVE can detect observation 17 as an outlier. In regard to the swamping rate performance, the T^2 and MVE methods deliver zero swamp rates. The trimming, SW1, SW2 and HSW2 methods produce about 5%–12% swamping rates, implying that 850–2040 instances of inliers have been misjudged as outliers. The MCD and PCAMCD methods appear to be seriously affected by the swamping effect. The HSW2 method provides a lightly better performance than the SW2 method in the swamping rate. In essence, there exist three high leverage inliers (observations 1, 6 and 10) that have strong influence on the SW2 and HSW2 methods.

As before, Figure 3 shows the control charts of the nine methods obtained via a single simulation. In the cases of the PCA and PCAMVE methods, the outlier is concealed. The trimming, SW1, HSW2, MCD and PCA MCD methods swamped some inliers. The original Hotelling's T^2 and MVE methods detected the real outlier, observation 17.

4.3. *Hawkins–Bradu–Kass dataset*

This is an artificial dataset constructed by Hawkins *et al.*[35] The dataset consists of 75 observations in four dimensions, with one response and three explanatory variables. It provides a good example of the masking effect. The first 14 observations are set up as outliers that are

Table II. The computational results of the 10 methods for the phosphorus content dataset

Methods	T^2	Trimming	SW1	SW2	HSW2
Detection rate	100.00%	100.00%	100.00%	100.00%	100.00%
Swamping rate	0.00%	5.88%	11.76%	7.77%	7.52%
Methods	MVE	MCD	PCA	PCAMVE	PCAMCD
Detection rate	100.00%	100.00%	0.00%	0.00%	100.00%
Swamping rate	0.00%	35.29%	0.00%	0.00%	23.53%

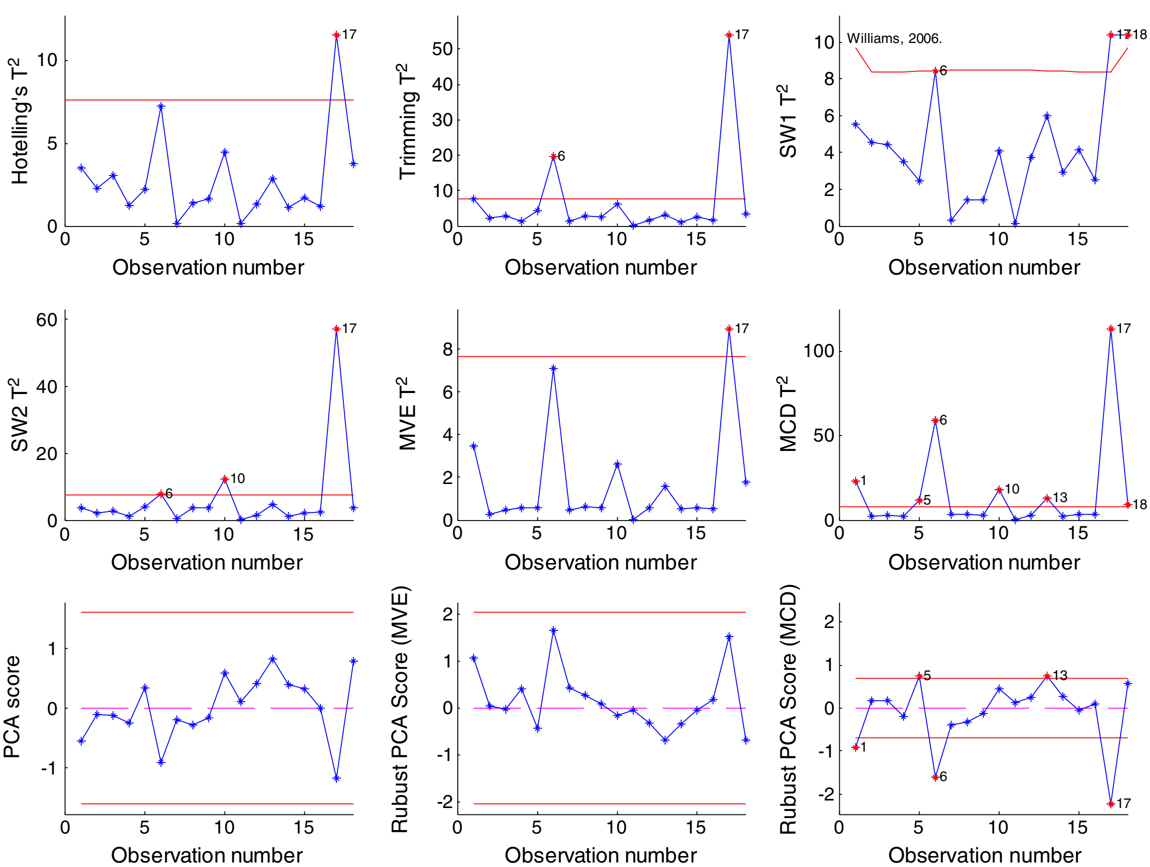

Figure 3. The control charts of the studied methods for the phosphorus content dataset

created into two groups: 1–10 and 11–14. The first group is considered as bad leverage points; the second group is considered as good leverage points. In other words, their explanatory variables are outlying, but the corresponding response variable fits the model well. They had tried many different methods and discovered that the outliers were masked and the four leverage inliers are swamped. Rousseeuw and Leroy,[11] Rousseeuw and Zomeren,[15] Bartkowiak[28] and Kim and Krzanowski[29] reported that observations 1–10 are outliers by regression methods. However, Hadi,[17] Atkinson and Mulira,[9] Sebert *et al.*,[27] and Marchette and Solka[36] stressed that the first three explanatory variables are specially constructed such that the first 14 observations are outliers. In this article, only the three explanatory variables are considered, so observations 1–14 are deemed outliers. The computational results of the ten methods for the Hawkins–Bradu–Kass dataset are displayed in Table III.

Clearly, from the table, the hierarchical clustering mechanism does not aid the original SW2 method in that the SW2 method has achieved perfect performances on both measures. Similar to the previous two datasets, the SW1, MCD and PCAMCD methods yielded significant swamping rates. In this particular situation where multiple outliers are present, the T^2 and PCA methods performed poorly in the detection rate. Recall that, in the woodmod dataset where multiple outliers also exist, these two methods failed to deliver satisfactory detection rates.

Again, Figure 4 illustrates the control charts of the studied methods for the Hawkins–Bradu–Kass dataset. The original Hotelling's T^2 and PCA detected only two outliers (observations 2 and 4) and one outlier (observation 4), respectively. The SW1, MCD and PCAMCD methods constructed stringent control charts that caused a number of swamped inliers. Basically, the trimming, HSW2 and MVE methods worked best for the Hawkins–Bradu–Kass dataset.

Table III. The computational results of the 10 methods for the Hawkins–Bradu–Kass dataset

Methods	T^2	Trimming	SW1	SW2	HSW2
Detection rate	14.29%	100.00%	100.00%	100.00%	100.00%
Swamping rate	0.00%	0.00%	22.95%	0.00%	0.00%
Methods	MVE	MCD	PCA	PCAMVE	PCAMCD
Detection rate	100.00%	100.00%	7.14%	93.37%	40.24%
Swamping rate	0.30%	28.70%	0.00%	0.28%	10.68%

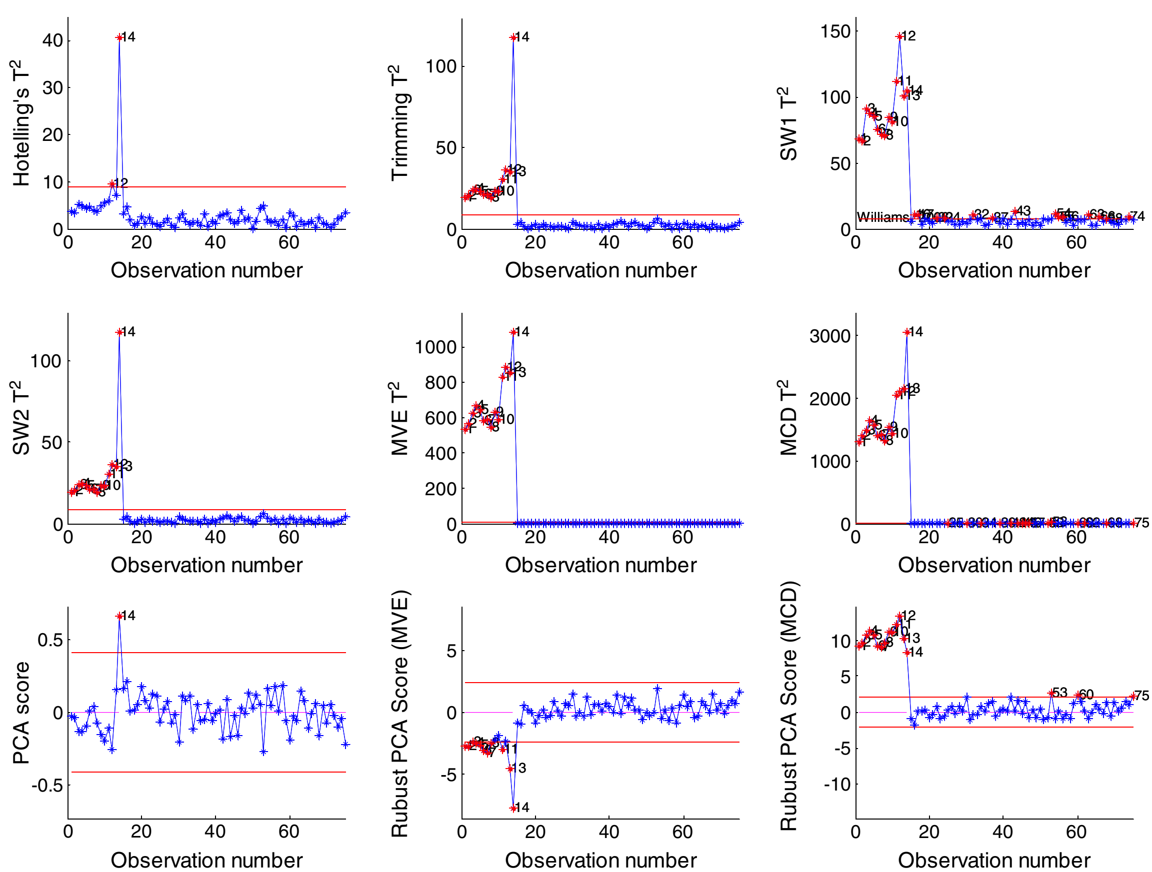

Figure 4. The control charts of the studied methods for the Hawkins–Bradu–Kass dataset

4.4. Dilemma dataset (esoteric example)

The last test is to consider the dilemma dataset that was explained by Hocking and Pendleton.[37] It contains 26 observations with three explanatory variables and one response variable. The dataset contains three outliers (observations 11, 17 and 18) and an influential point (observation 24), all of which are regarded as outliers. Hawkins, Bradu and Kass[36] also drew the same conclusion concerning the outliers of this dataset. To be consistent with the Hawkins–Bradu–Kass Dataset, in this study, the three explanatory variables are only considered with four outliers. The computational results of the 10 compared methods are listed in Table IV. The original T^2, SW1 and PCA methods only achieved a detection rate of 25%; the remaining methods generated a detection rate of 75%. Concerning the swamping rate results, except the three methods previously mentioned, the MVE and PCAMVE methods take the lead with a zero swamping rate. The SW2 and HSW2 methods are ranked the second with a swamping rate of 4.55%, followed by the PCAMCD and MCD methods that have swamping rates of 22.73% and 36.36%, respectively. Based on a single simulation, Figure 5 exhibits the control charts of the compared methods for the dilemma dataset. It is obvious from the figure that all the methods cannot detect the outlier of observation 17 and the inlier of observation 13 is most likely to be swamped.

5. Further simulation study

As evidenced by the computational results in terms of four classic datasets, the proposed HSW2 method improves the original version in the former two datasets and produces at least equally good performances as the original version in the latter two datasets. In the mean

Table IV. The computational results of the 10 methods for the dilemma dataset

Methods	T^2	Trimming	SW1	SW2	HSW2
Detection rate	25.00%	75.00%	25.00%	75.00%	75.00%
Swamping rate	0.00%	4.55%	0.00%	4.55%	4.55%
Methods	MVE	MCD	PCA	PCAMVE	PCAMCD
Detection rate	75.00%	75.00%	25.00%	75.00%	75.00%
Swamping rate	0.00%	36.36%	0.00%	0.00%	22.73%

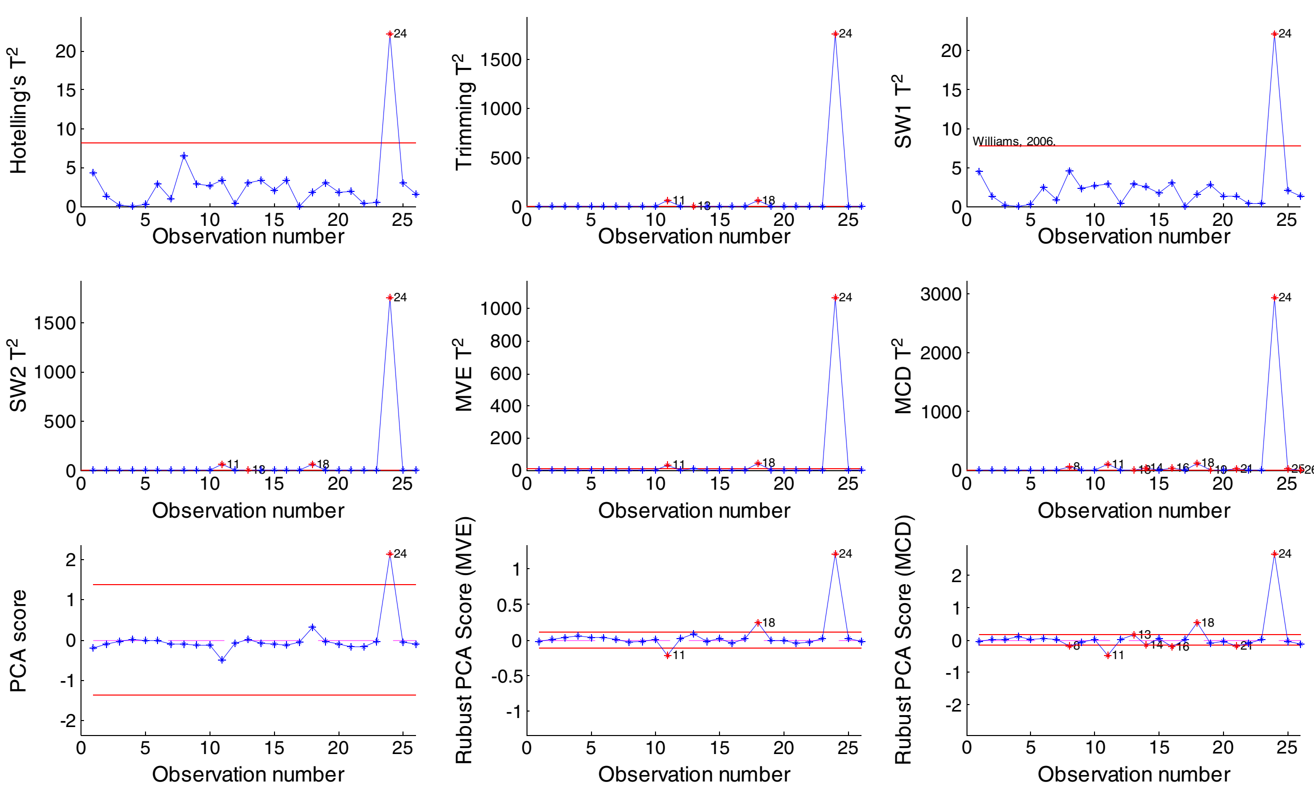

Figure 5. The control charts of the studied methods for the dilemma dataset

time, the MVE method also shows extremely competitive performances in both measures as compared to the SW2 and HSW2 methods. To further verify that the proposed method is really effective in enhancing the detection rate and simultaneously avoiding a high swamping rate, a Monte Carlo simulation study that emulates the phase I analysis in SPC to detect outliers is conducted. These outliers are assumed to be single excursions that may have resulted from assignable causes of short duration. Multivariate random datasets, containing a few outliers, will be independently generated from normal distributions. The severity level of outliers in each simulated dataset is given according to the designated number of dimensions where the corresponding coordinate values are outlying. Note again that the two measures, the masking and swamping rates, are used for assessment purposes. In each simulation cycle, 30 independent datasets with 20 observations each in five dimensions are artificially generated to compare three methods, SW2, HSW2 and MVE. The outlier ratio is fixed at 20%, indicating four outliers generated to be deliberately mixed in each dataset. The detection and swamping rates are averaged over 100 simulation cycles. In every simulation cycle, different random number seeds are used in order to produce fair computational results.

In each dataset, the 16 inliers are sampled within ± 3 standard deviations from the standard normal distribution, and the outlier is sampled from a uniform distribution within the interval of shift level $\pm$ a half standard deviation). For example, if the shift level is set at 3.5, then the sampling interval is [3, 4]. The probability of an observation appearing in this interval from the standard normal distribution is about 0.1318%. As the observation so rarely happens, it should be considered as an outlier. The simulation study considers the cases with shift levels, ranging from 2 to 5 standard deviations, to observe the changes of the detection and swamping rates; five different numbers of outlying dimensions are considered. It is also important to note that, in each dataset of 20 observations, four outliers are randomly chosen; given a fixed number of outlying dimensions, these variables receiving the shift are fixed throughout the dataset. This setup seems a reasonable way for outliers to enter the process where the outliers are unreliably measured due to process instability.

Tables 5–9 tabulates the simulation results of the three compared methods as the four outliers of each dataset are contaminated by one, two, three, four and five dimensions, respectively. In each table, the shift levels from 2 to 5 are considered. The standard deviations of the detection and swamping rates computed over 100 simulation cycles are also recorded. Clearly from these tables, the detection rate increases as the shift level increases; the swamping rate decreases as the shift level increases. Likewise, the detection rate increases as the severity level of outliers increases; the swamping rate decreases as the severity level of outliers increases. For situations that will more realistically happen in practice, the cases of shift levels greater than or equal to 3.5 in Tables 5–9 are more appropriate for being called outliers than the other cases. In the discussions that follow, the emphasis will be put on the cases of shift levels greater than or equal to 3.5.

It can be seen from Table V that the proposed HSW2 method overwhelmingly outperforms the other two methods in both the detection and swamping rates. The proposed HSW2 method yields the detection rate from 44.17% up to 80.43%, from 29.77% to 38.25% for the SW2 method, and from 17.83% to 40.37% for the MVE method. The proposed HSW2 method delivers very good performance in the swamping rate, from 8.22% down to 2.35%. The SW2 method seems unfazed by the shift level changes to deliver a quite stable swamping rate between 10% and 11%. The MVE method also performs well in the swamping rate, from 7.04% down to 4.15%. Table VI reports the simulation results with two dimensions of outliers. By contrast to the results shown in Table V, the proposed HSW2 method presents far more competitive performances than the other two methods. The HSW2 method achieves the detection rate up to 96.42% and the swamping rate down to 1.47% for the shift level of 5. Looking at the standard deviation statistics of both measures from Tables VI and VII, it has been discovered that the HSW2 and MVE methods produce more stable (with less variability) performance rates than the SW2 method.

Tables 7–9 show the simulation results as the severity level of outliers become more aggravated than Tables V and VI Those are three, four and five dimensions of the outliers, respectively. The detection rates of the HSW2 method range from 96.71% to 99.61%, and the swamping rates are between 0.28% and 1.87%. The SW2 and MVE methods are still not competitive with the HSW2 method in both

Table V. The Monte Carlo simulation results as one dimension of the outliers is outlying

Outlier		SW2		HSW2		MVE	
Shift	Dim.	Detection	Std (%)	Detection	Std (%)	Detection	Std (%)
2	1	21.64%	16.60	21.77%	16.08	13.98%	8.08
2.5	1	22.33%	17.50	23.82%	15.77	18.20%	9.79
3	1	25.30%	20.33	36.94%	17.37	19.19%	12.58
3.5	1	29.77%	22.17	44.17%	17.43	17.83%	10.41
4	1	31.94%	23.52	52.06%	18.59	27.63%	16.33
4.5	1	33.92%	28.55	68.56%	16.51	37.36%	16.92
5	1	38.25%	31.21	80.43%	17.23	40.37%	16.08
Shift	Dim.	Swamping	Std (%)	Swamping	Std (%)	Swamping	Std (%)
2	1	13.48%	4.40	13.26%	4.24	5.74%	3.74
2.5	1	13.32%	4.78	13.02%	4.19	9.50%	4.15
3	1	12.73%	5.07	10.32%	4.39	7.29%	3.76
3.5	1	11.32%	5.47	8.22%	4.32	7.04%	3.23
4	1	11.40%	5.43	6.84%	4.30	5.61%	3.07
4.5	1	11.31%	6.14	4.54%	3.62	5.15%	3.77
5	1	10.08%	6.18	2.35%	3.34	4.14%	2.31

Table VI. The Monte Carlo simulation results as two dimensions of the outliers are outlying

Outlier		SW2		HSW2		MVE	
Shift	Dim.	Detection	Std (%)	Detection	Std(%)	Detection	Std (%)
2	2	16.49%	17.90	21.35%	16.91	7.61%	7.01
2.5	2	23.76%	25.84	51.44%	20.20	30.22%	16.28
3	2	28.95%	28.36	55.25%	21.85	30.91%	16.06
3.5	2	32.93%	32.53	77.08%	16.20	41.17%	11.95
4	2	41.03%	35.21	89.19%	14.36	47.28%	23.74
4.5	2	41.00%	38.03	91.70%	15.69	58.96%	20.48
5	2	38.42%	39.96	96.42%	12.30	79.02%	15.35
Shift	Dim.	Swamping	Std (%)	Swamping	Std (%)	Swamping	Std (%)
2	2	14.77%	5.22	13.26%	5.00	9.01%	5.74
2.5	2	13.58%	6.03	7.83%	4.91	6.51%	3.41
3	2	12.27%	6.27	7.03%	4.73	6.55%	4.59
3.5	2	11.73%	6.47	3.60%	3.06	6.25%	3.47
4	2	9.82%	6.49	1.59%	2.34	6.18%	3.02
4.5	2	10.60%	6.97	2.16%	2.51	6.43%	4.69
5	2	10.83%	7.22	1.47%	1.92	3.07%	1.70

Table VII. The Monte Carlo simulation results as three dimensions of the outliers are outlying

Outlier		SW2		HSW2		MVE	
Shift	Dim.	Detection	Std (%)	Detection	Std (%)	Detection	Std (%)
2	3	24.72%	26.15	49.45%	20.41	11.34%	8.96
2.5	3	28.60%	32.42	73.93%	19.71	27.74%	16.91
3	3	25.09%	32.86	73.29%	17.75	27.84%	12.15
3.5	3	35.15%	40.66	97.34%	12.51	53.58%	17.24
4	3	34.15%	40.50	97.13%	7.83	59.31%	17.20
4.5	3	33.19%	41.51	97.19%	9.12	63.81%	22.43
5	3	34.28%	41.15	97.84%	8.48	70.23%	12.33
Shift	Dim.	Swamping	Std (%)	Swamping	Std (%)	Swamping	Std (%)
2	3	12.94%	6.36	7.58%	4.88	6.99%	4.73
2.5	3	13.03%	7.03	4.08%	4.10	8.70%	4.68
3	3	13.59%	7.25	4.18%	4.01	7.15%	3.35
3.5	3	11.86%	7.33	1.87%	2.04	6.55%	3.30
4	3	12.09%	7.47	1.25%	1.41	6.75%	4.24
4.5	3	12.33%	7.62	1.56%	1.53	5.56%	3.52
5	3	12.15%	7.66	1.44%	1.33	5.33%	2.76

measures. As can be obviously seen from Tables 7–9 for the shift levels greater than 3, the HSW2 methods are able to produce much more stable performance rates (with less variability) than the other two methods.

In short, the computational evidence illustrated in Section 4 clearly demonstrates that the proposal to use a hierarchical clustering scheme prior to the SW2 method is shown to be very effective in improving the detection capability of the original SW2 method while keeping a low swamping rate. However, the best-practice stopping criterion for the proposed method still remains unanswered and deserves further investigation. Another important facet worth mentioning at this closing stage is that, from a computation point of view, the proposed HSW2 method, on average, works 60 to 100 times faster than the MVE method while detecting potential outliers. It is also of particular importance to mention that all the experimental results in Section 4 are obtained with Microsoft Windows XP Service Pack 3 system on an Intel Core i7 CPU 860@2.80 GHz, 3.0 GB DDR3-RAM platforms.

6. Conclusions

In many practical situations, it is frequently necessary to simultaneously monitor multiple quality characteristics at the same time. Identifying potential outliers or influential observations in multivariate datasets becomes an increasingly important task before

Table VIII. The Monte Carlo simulation results as four dimensions of the outliers are outlying

Outlier		SW2		HSW2		MVE	
Shift	Dim.	Detection	Std (%)	Detection	Std (%)	Detection	Std (%)
2	4	19.33%	27.77	57.33%	22.66	15.83%	10.69
2.5	4	24.07%	37.12	83.20%	18.96	37.29%	16.14
3	4	28.13%	39.85	92.11%	13.63	45.04%	18.78
3.5	4	27.37%	41.21	96.71%	10.47	56.24%	19.94
4	4	27.35%	40.43	96.79%	8.57	68.87%	18.83
4.5	4	27.62%	42.48	99.13%	6.81	59.73%	23.15
5	4	28.94%	42.85	99.26%	4.27	70.53%	14.13
Shift	Dim.	Swamping	Std (%)	Swamping	Std (%)	Swamping	Std (%)
2	4	14.91%	6.73	6.33%	5.20	12.04%	4.67
2.5	4	14.09%	7.42	3.20%	3.35	9.25%	4.64
3	4	13.84%	7.93	1.88%	2.23	10.16%	4.13
3.5	4	13.36%	7.94	0.69%	1.74	7.73%	5.03
4	4	13.58%	7.89	1.01%	1.57	5.24%	3.55
4.5	4	14.00%	7.95	1.39%	1.13	7.01%	3.90
5	4	13.69%	8.44	0.74%	0.75	5.69%	2.81

Table IX. The Monte Carlo simulation results as five dimensions of the outliers are outlying

Outlier		SW2		HSW2		MVE	
Shift	Dim.	Detection	Std (%)	Detection	Std (%)	Detection	Std (%)
2	5	16.03%	34.01	74.88%	23.66	9.98%	7.14
2.5	5	24.28%	42.63	96.37%	13.27	21.11%	16.09
3	5	22.62%	40.94	90.40%	16.21	18.88%	8.70
3.5	5	24.25%	42.83	98.30%	9.39	35.17%	21.72
4	5	24.08%	42.57	98.80%	7.89	46.27%	16.17
4.5	5	23.28%	41.83	98.20%	7.83	51.03%	21.70
5	5	24.56%	42.91	99.61%	3.32	63.70%	20.11
Shift	Dim.	Swamping	Std (%)	Swamping	Std (%)	Swamping	Std (%)
2	5	16.63%	7.84	3.73%	4.99	16.65%	4.83
2.5	5	14.87%	8.58	0.99%	2.63	13.94%	7.20
3	5	15.94%	8.29	3.03%	3.22	15.47%	4.70
3.5	5	14.98%	8.46	1.12%	1.62	11.90%	5.75
4	5	15.33%	8.60	1.07%	1.51	10.68%	4.79
4.5	5	15.39%	8.00	1.79%	1.45	11.79%	5.70
5	5	14.81%	8.73	0.28%	0.60	7.44%	4.01

proceeding to more advanced statistical analyses. Multivariate control charts are the most common statistical tool used in such circumstances.

Some classic multivariate control charts, such as Hotelling's T^2, cannot detect multiple potential outliers effectively, while some methods, such as MVE and MCD, require tremendous computation efforts. For this reason, it is desirable to design robust control charts for multiple outlier situations that also maintain a nice balance between detection capability and computation effort. Toward this end, we modify the Sullivan and Woodall's second approach (as known as SW2) by introducing a hierarchical clustering scheme as a preliminary filter for potential outliers before implementing the SW2 method. Thus, the proposed new method is termed HSW2.

To illustrate and compare the proposed method to other well-known methods, four noted multivariate datasets are taken from the literature. The experimental results show that the proposed HSW2 method proves to be a robust and effective control chart as compared to the original SW2 method and the other eight methods. In particular, the HSW2 method significantly improves the original version on both the detection and swamping rates. On the other hand, the computation time of the HSW2 method is far less than the MVE method. For instance, in the woodmod dataset, the computation time is 0.079 s for the HSW2 method and 5.93 s for the MVE method. A Monte Carlo simulation study where the variables are assumed uncorrelated each other was conducted to further verify the robustness of the HSW2 method in comparison to the SW2 and MVE methods. As evidenced by the Monte Carlo simulation results, the HSW2 method overwhelmingly outperforms the other two methods.

Acknowledgements

The authors are grateful to the handling editor, Dr. Douglas Montgomery, and an anonymous referee for providing insightful comments and constructive suggestions that greatly improve an earlier draft of the paper. This paper is partly funded by National Science Council (Taiwan) Grant No. NSC 100-2221-E-027-115-MY3.

References

1. Hawkins DM. *Identification of Outliers*. Chapman and Hall: New York, 1980.
2. Montgomery DC, Peck EA, Vining GG. *Introduction to Linear Regression Analysis*, 3rd ed. John Wiley & Sons: New York, 2001.
3. Hotelling H. Multivariate Quality Control. *Techniques of Statistical Analysis*, Eisenhart, Hastay, and Wallis (eds.), McGraw-Hill: New York 1947.
4. Williams JD, Woodall WH, Birch JB, Sullivan JH. Distribution of Hotelling's T2 Statistic Based on the Successive Differences Estimator. *Journal of Quality Technology* 2006; **38**:217–229
5. Langenberg P, Iglewicz B. Trimmed Mean. *Journal of Quality Technology* 1986; **18**:152–161.
6. Rocke DM. Robust Control Charts. *Technometrics* 1989; **31**:173–184.
7. Rocke DM. robust control charts. *Journal of the Royal Statistical Society*. Series D 1992; **41**:97–104.
8. Sullivan JH, Woodall WH. A Comparison of Multivariate Control Charts for Individual Observations. *Journal of Quality Technology* 1996; **28**:398–408.
9. Atkinson AC, Mulira HM. The Stalactite Plot for the Detection of Multivariate Outliers. *Statistics and Computing* 1993; **3**:27–35.
10. Rousseeuw PJ. Least Median of Squares Regression. *Journal of the American Statistical Association* 1984; **79**:871–880.
11. Rousseeuw PJ, Leroy A. *Robust Regression and Outlier Detection*, John Wiley & Sons: New York, 1987.
12. Tracy ND, Young JC, Mason RL. Multivariate Control Charts for Individual Observations. *Journal of Quality Technology* 1992; **24**:88–95.
13. Vargas JA. Robust Estimation in Multivariate Control Charts for Individual Observations. *Journal of Quality Technology* 2003; **35**:367–376.
14. Stefatos G, Ben Hamza A. Fault Detection using Robust Multivariate Control Chart. *Expert Systems with Applications* 2009; **36**:5888–5894.
15. Rousseeuw PJ, Zomeren BC. Unmasking Multivariate Outliers and Leverage Points. *Journal of the American Statistical Association* 1990; **85**, 633–639.
16. Lopuhaä HP, Rousseeuw PJ. Breakdown Points of Affine Equivariant Estimators of Multivariate Location and Covariance Matrices. *The Annals of Statistics* 1991; **19**:229–248.
17. Hadi AS. Identifying Multiple Outliers in Multivariate Data. *Journal of the Royal Statistical Society* 1992; **54**:761–771.
18. Jensen WA, Birch JB, Woodall WH. High Breakdown Estimation Methods for Phase I Multivariate Control Charts. *Quality and Reliability Engineering International* 2007; **23**:615–629.
19. Rousseeuw PJ, van Aelst S. Minimum Volume Ellipsoid. *Wiley Interdisciplinary Reviews: Computational Statistics* 2009; **1**:71–82.
20. Rousseeuw PJ, van Driessen K. A Fast Algorithm for the Minimum Covariance Determinant Estimator. *Technometrics* 1999; **41**:212–223.
21. Jolliffe IT. *Principal component analysis*. New York: Springer, 1986.
22. Gnanadesikan R, Kettenring JR. Robust Estimates, Residuals, and Outlier Detection with Multiresponse Data. *Biometrics* 1972; **28**:81–124.
23. Yang K, Trewn J. *Multivariate Statistical Methods in Quality Management*. McGraw Hill Professional, 2004.
24. Webb AR. *Statistical Pattern Recognition*, 2nd Edition, QinetiQ Ltd.: Malvern, UK, 2003.
25. Murtagh F. A Survey of Recent Advances in Hierarchical Clustering Algorithms. *The Computer Journal* 1983; **26**:354–359.
26. Draper NR, Smith H. *Applied Regression Analysis*, John Wiley & Sons: New York, 1966
27. Sebert DM, Montgomery DC, Rollier DA. A Clustering Algorithm for Identifying Multiple Outliers in Linear Regression. *Computational Statistics & Data Analysis* 1998; **27**:461–484.
28. Bartkowiak A. Robust Mahalanobis Distances Obtained Using the 'Multout' and 'Fast-mcd' Methods. *Biocybernetics and Biomedical Engineering* 2005; **25**:7–21.
29. Kim S, Krzanowski, WJ. Detecting Multiple Outliers in Linear Regression Using a Cluster Method Combined with Graphical Visualization. *Computational Statistics* 2007; **22**:109–119.
30. Alqallaf FA, Konis KP, Martin RD, Zamar RH. Scalable Robust Covariance and Correlation Estimates for Data Mining. *Proceedings of the eighth ACM SIGKDD international conference on Knowledge discovery and data mining*, 2002; 14–23.
31. Snedecor GW, Cochran WG. *Statistical Methods*, 6th Edition. the Iowa State University Press: Ames, Iowa, 1967.
32. Prescott P. An Approximate Test for Outliers in Linear Models. *Technometrics* 1975; **17**:129–132.
33. Lund RE. Tables for an Approximate Test for Outliers in Linear Models. *Technometrics* 1975; **17**:473–476.
34. Rousseeuw PJ, Zomeren BC. A Comparison of Some Quick Algorithms for Robust Regression. *Computational Statistics & Data Analysis* 1992; **14**:107–116.
35. Hawkins DM, Bradu D, Kass GV. Location of Several Outliers in Multiple-Regression Data Using Elemental Sets. *Technometrics* 1984; **26**:197–208.
36. Marchette DJ, Solka JL. Using Data Images for Outlier Detection. *Computational Statistics & Data Analysis* 2003; **43**:541–552.
37. Hocking RR, Pendleton OJ. The Regression Dilemma. *Communications in Statistics – Theory and Methods* 1983; **12**:497–527.

Authors' biographies

Shu-Kai S. Fan is a Professor of Industrial Engineering and Management at National Taipei University of Technology, Taiwan, Republic of China. He received his Bachelor degree in Industrial Engineering from Tunghai University, Master degree in Industrial Engineering from University of Pittsburgh and Doctor degree in industrial engineering from the University of Texas-Arlington. His research interests include areas of applied statistics, applied mathematical programming, response surface methodology, profile monitoring, advanced process control, multi-objective evolutionary algorithms, and image registration. He is the recipient of 2008 Quality Award given by Chinese Society for Quality (CSQ) and Outstanding Professor Award (Science, Technology and Humanity Category) given by Far Eastern Y.Z. Hsu Science and Technology Memorial Foundation. Dr. Fan serves as an associate editor of Engineering Optimization.

Hui-Kuo Huang received his Master degree in Industrial Engineering from Yuan-Ze University, Taiwan, Republic of China. Now he works as a senior engineer at Macronix International Co., Ltd., Taiwan, Republic of China. His research interests include multivariate data analysis and advanced process control.

Yuan-Jung Chang is a doctoral student of Industrial Engineering and Management at Yuan-Ze University, Taiwan, Republic of China. His research interests include areas of Run-to-Run control, advanced process control, fault detection and classification, and profile monitoring.