A MULTIVARATE QUALITY CONTROL PROCEDURE IN MULTISTAGE PRODUCTION SYSTEMS

S. T. A. Niaki and B. Moein-Zadeh[1]

Department of Industrial Engineering
Sharif University of Technology
Tehran, Iran

Abstract In this paper a multivariate quality control procedure is offered in which several correlated stages are present in production systems and in each stage there are several correlated control variable. Using Hotelling's T^2 statistic, first, each stage is tested for being out of control. Then out of control variables are selected using Murphy's method. The remainder of the research involves evaluating each stage, when the present and the previous stages are both out of control. In this regard, regression of the present stages' T^2 statistic on prior stages' T^2 statistics is considered and a residual test between the present stage estimated and observed T^2 statistics is made.

Key Words Statistical Process Control, Hotelling's T^2 Chart, Mutivariate Control Charts, Regression Residuals, Discriminant Analysis

چکیده دراین مقاله روشی جهت کنترل کیفیت چند متغیره در صنایعی که تولید در آنها درچند مرحله وابسته به هم انجام می شود و در هر مرحله چند پارامتر کیفی وابسته به هم وجود دارد، ارائه شده است. الگوریتم پیشنهادی در چهار فاز صورت می پذیرد. در فاز اول مشاهداتی از یک سیستم تحت کنترل به دست می آید و براساس آن و با استفاده از آماره هتلینگ (Hotelling's T^2) حد بالای کنترل جهت تحت کنترل بودن هریک از مرحل تولید سیستم در آینده محاسبه می شود. آنگاه در فاز دوم با استفاده از نمودارهای کنترل فاز اول، تحت کنترل بودن هریک از مراحل مورد بررسی قرار می گیرد. در فاز سوم با استفاده از مفهوم تبعیض گذاری (Discriminant Analysis) متغیر(های) عامل انحراف، در صورتی که هریک از مراحل تولیدی خارج از کنترل باشند، شناسایی می گردند. درصورتی که دو مرحله متوالی در حالت خارج از کنترل قرار گیرند، در فاز چهارم مرحله ای که عامل انحراف بوده است، مشخص می شود. جهت این کار از تکنیک رگرسیون و آزمون باقیمانده مدلی که از آماره T^2 مرحله جاری و مرحله قبل به دست آمده است، استفاده می شود. الگوریتم پیشنهادی جهت مشاهدات شبه تصادفی تولید شده ازیک توزیع نرمال چند متغیره با پارامترهای مختلف که در آن عمداً بعضی از پارامترها تغییر داده شد، به کار گرفته شد و نتیجه رضایت بخش بوده است.

INTRODUCTION

In industrial quality control, when two or more quality characteristics are present and have to be monitored and controlled simultaneously, the usual practice has been to maintain a separate (univariate) chart for each characteristic. In this situation the quality must be measured by the joint levels of several variables. Nevertheless, only the properties of univariate cumulative sum (CUSUM) procedures have been discussed in the literature [1-5]. Unforunately, due to existing correlation between characteristics, the use of separate control charts leads to large errors in statistical process control [6-8]. In this situation, on the one hand each of the several correlated variables of the process can be shown to be in-control by its respective control charts; even though, the process is practically out of control in terms of the joint control region of the variables. On the other hand, while one or more control charts for individual variables may show out of control signal, the whole process may be actually in control.

Traditionally the control charts of Hotelling's T^2 [6], which summarize several variables into one, have been used for multivariate processes. There are, however, practical pitfalls in using T^2 statistic. While

[1]Dr. Niaki is an Assistant Professor of Industrial Engineering at Sharif University of Technology. Mr. Moein-Zadeh is a Doctoral Student at University of Cincinnati, USA.

it provides evidence that the process is out of control, it does not show what has been changed. Thus, further diagnostic detective work is needed. This problem has been addressed in the literature by several authors such as Hanwkins [9], Doganaksoy et al. [10], Murphy [11], Healy [12], to name a few. Also, a single multivariate CUSUM (MCUSUM) has been the method of controlling the multivariate process. For instance Woodall and Ncube [13] proposed an MCUSUM procedure in which some methods are given to approximate parameters of the distribution of the minimum run length for univariate CUSUM charts. However, they assumed the variance-covariance matrix of the random vector consisting of the measurments' mean to be known.

Furthermore some production systems possess a mechanism that leads to characteristics having a natural ordering, in which if any characteristic undergoes a parameter shift, it may affect some or all of the following characteristics, but none of the preceding characteristics. This problem is typified by serial value-added manufacturing processes. For example, in a detergent making chemical plant where the material passes through successvie steps, the quality of the product at the end of any step depends not only on the performance of that step, but also on how well the preceding steps have performed. Therefore, if a particular step goes out of control, it can affect the product quality both in that step and in some or all of the following steps.

Based on the above-mentioned parts, one can divide multivariate quality control problems in two main categories: **1)** in every quality test, where there are several correlated quality characteristics, the purpose is to detect where the process has gone "out-of-control" and also which of the variables is the cause of deterioration, and **2)** in production lines we have several stages, and quality characteristics of each stage depends on the quality of the prior stage.

As mentioned before, the most common method for detecting an out of control signal is application of Hotelling T^2 Statistic, which is given in many current quality control text books [14]. In the cause selecting methods (finding which variable(s) is the cause of deterioration), a number of researchers, have utilized various procedures, among which principal components analysis, discriminant analysis, and regression analysis have been quite influential play the most important role. Using discriminant analysis, Murphy [11] has developed a statistic and an algorithm, that can be applied whenever the population parameters (μ, Σ) are available. Tracy, Young, and Mason [15] discuss the modifications necessary for the situation in which the parameters (μ, Σ) are not exactly known. Jackson [16] and Chang [17] applied the principal components, and Hawkins [9] proposed a method based on regression to solve the cause selecting problems. Some other charts have been developed but none of then provides a statistical test for selecting the variables which caused the deterioration [18-22].

In the second category, applying regression to purify the quality of the current stage from the effect of the prior stage, Zhang [23] has studied a situation in auto industries, where there are several stages in the production system and the quality in each stage depends on the quality of the prior stage. However, he has only considered one parameter in each stage.

RESEARCH OBJECTIVES

In this research, manufacturing systems are considered that have several stages in production and several correlated quality characteristics are present in each stage. The purpose of this research is monitoring a shift in the process mean only. In this regard, a statistic and an algorithm are developed for the cause selecting problem in which the population parameters are not known and are to be estimated. Also applying the Hotelling T^2 statistic and regression procedure a criterion is developed to eliminate the effect of the prior stage from the current stage's quality characteristics.

RESEARCH DEVELOPMENT

It is assumed that the process possesses a multivariate normal distribution with unknown parameters (μ, Σ). The proposed algorithm consists of four phases, in all of which the estimated value of (μ, Σ) is used by the collected rational data from an under-control process. The four phases of the algorithm are:

Phase 0: Collecting data for the Model Setup.
Phase 1: Detecting departures in each stage.
Phase 2: Detecting which variables have caused the 'out-of-control' signal.
Phase 3: Considering the effect of the previous stage on the quality characteristics of the current stage and separating it from the resulted output.

Phase 0: Collecting data for the Model Setup

Consider a process that has several stages, and in each stage there are p_s (s = 1,2,...p) parameters (quality characteristics). In general, assuming the process is under-control, in the setup phase, for parameter **i** in each stage of the process, $\mathbf{k_{is}}$ groups of sample data are collected so that each group has $\mathbf{n_s}$ observations on stage **s**. For the sake of simplicity, assume that **s** = 2, $\mathbf{k_{is}}$'s are equal to **k**, and that the $\mathbf{n_s}$'s are equal to **n**. Now for each stage, let $\mathbf{X_{ijm}}$ be the observation of the **m**-th sample of parameter **i** in the jth group. Define the sample mean of parameter **i** in the jth group as:

$$\overline{X}_{ij}=\frac{1}{n}\sum_{m=1}^{n} X_{ijm} \quad ; \; j = 1,\ldots,k, \quad i = 1,2,\ldots,p \tag{1}$$

and the overall sample mean of the ith parameter to be:

$$\overline{\overline{X}}_{i}=\frac{1}{k}\sum_{J=1}^{k} \overline{X}_{ij} \quad ; \quad i = 1,2,\ldots,p \tag{2}$$

Then the sample mean vector of the parameters will be:

$$\overline{\overline{X}} = \left| \overline{\overline{X}}_1, \overline{\overline{X}}_2, \ldots, \overline{\overline{X}}_p \right|^T \tag{3}$$

which has an estimated variance-covariance matrix (**S**) whose element of row i and column j is calculated as:

$$S_{ij} = S_{ji} = \frac{\sum_{m=1}^{k} (\overline{X}_{im} - \overline{\overline{X}}_i)(\overline{X}_{jm} - \overline{\overline{X}}_j)}{(k-1)} \quad ; \; i,j = 1,2,\ldots,p \tag{4}$$

Matrix **S** is symmetric, at least positive semi-definite, and has an inverse, $\mathbf{S}^{-1}$. Now, to make sure that the setup phase has been successfully completed, for each of the k groups we calculate T_j^2 as [14].

$$T_j^2= n\,(\overline{X^{(j)}} - \overline{\overline{X}})^T S^{-1}(\overline{X^{(j)}} - \overline{\overline{X}}) \tag{5}$$

and compare it against the Upper Control Limit (UCL) as: [24]

$$UCL= \left(\frac{knp - kp - np + p}{kn - k - p + 1}\right) F_{\alpha(P,\, kn-k-p+1)} \tag{6}$$

where $\overline{\mathbf{X}^{(j)}}$ denotes a vector with **p** elements that contains the group averages for each of the **p** parameters. If T_j^2 in all of the **k** groups does not exceed the UCL, one can regress T^2 values of stage s+1 on the T^2 values of stage s and estimate the regression parameters "a" and "b" and the sample variance of T_j^2's($\mathbf{S_d}$) for future use. At this point the phase 0 is completed and we go on to phase 1 starts. Otherwise, if any of the T_j^2 values exceeds the UCL, the corresponding group would be investigated. Specifically, it is necessary to determine which parameters are causing the out-of-control signal. This is done by applying the Bonferroni's intervals and construct **p** intervals for each group that has produced an out-of-control message on the multivariate chart. Thus, for the jth group the interval for the ith parameter would be:

$$\overline{\overline{X}}_i - t_{\frac{\alpha}{2p},\,k(n-1)} S_{pi} \sqrt{\frac{k-1}{nk}} \leq \overline{X_i}\,(j) \leq \overline{\overline{X}}_i + t_{\frac{\alpha}{2p},\,k(n-1)} S_{pi} \sqrt{\frac{k-1}{nk}} \qquad (7)$$

where $\mathbf{S_{pi}}$ denotes the square root of the sample variance for the ith parameter, and the other components of the above equation are as calculated in phase 0. If the above equation is not satisfied for the ith parameter, the values of that parameter would then be investigated for the jth group. If an assignable cause is detected, the entire group would be eliminated (for all of the **p** parameters) and the UCL should be recalculated [24]. Also, T^2 values, regression parameters and $\mathbf{S_d}$ must be re-computed. At this point, the calculation of phase 0 is complete and we go on to the next phase starts.

Phase 1*: *Detecting departures in each stage

In this phase we use Hotelling T^2 statistics to test the following hypothesis in each stage of production process [14]:

$H_0: \mu = \mu_0$
$H_1: \mu \neq \mu_0$

where μ is the mean vector of dimension p consisting of the parameter means of the system under consideration. The k subgroups would be tested by computing k values of T_j^2 in each stage as:

$$T_{js}^2 = n\,(\overline{X}_{js} - \overline{\overline{X}}_s)^T\, S_s^{-1} (\overline{X}_{js} - \overline{\overline{X}}_s)\ ,\quad s = 1,2\ ,\ j = 1,\ldots,k \qquad (8)$$

and comparing them with the upper control limit (UCL) of this step as [14]:

$$UCL = \left[\frac{p(k-a+1)\,(n-1)}{(k-a)\,n-k+a-p+1}\right] F_{\alpha,\,[p,\,(k-a)\,n-k+a-p+1]} \qquad (9)$$

where "**a**" denotes the number of original groups that were eliminated in phase 0 of the algorithm. Note that even when **a**=0, the UCL in the setup phase is different from the UCL in phase 1. This comes from the fact that a different distribution theory is involved, because, the sets of groups in phase 1 are assumed to be independent of the sets of groups in the setup phase, and it is desirable that the UCL in phase 1 to be tighter than the UCL in phase zero [24].

In each stage, if any of the T_j^2 values exceeds the UCL, then an "an out-of-control signal" is detected for group j. At this point phase 2 of the algorithm is performed to find which of the parameters are causing deterioriation. Also if both of the two subsequent stages are out of control, phase 3 of the algorithm will be activated to see whether the later stage is out of control or the former stage has been out of control and caused an out-of-control signal for the current stage.

Phase 2*: *Detecting the variable(s) causeing an out-of-control signals

Using discriminant analysis, Murphy [11] has developed an algorithm to test which parameters are cause an out-of-control signal. He assumed that the vector mean and variance-convariance matrix, both were known. However, in most situations in practice these quantities are unknown and must be estimated. In this phase the discriminant analysis by Murphy was used except that in the present test the population parameters are assumed to be unknown and must be estimated.

Suppose the mean vector and the variance-convariance matrix are estimated as in phase zero of the algorithm. Then assuming that the underlying p dimensional process is normally distributed, $\overline{\overline{X}}$ will be distributed as a multinormal distribution with the mean of μ and the variance of Σ[i.e. $N_p(\mu,\Sigma)$]. Also, for **n>p**, S will have a Wishart distribution with parameters Σ and **n**[i.e., $W_p(\Sigma, n)$] [25], and if vector Y is defined as:

$$Y=\sqrt{\frac{nk}{k+1}}\,(\overline{X}_j-\overline{\overline{X}}) \qquad (10)$$

then Y and S will be independent of each other [14].

Now suppose that the first r (r = 1,2,…,p) quality characteristics are the cause of detoriation. To check this, let partition, the parameter mean space (μ) into two mutually exclusive subspaces of dimensions r and s such that $r+s=p$. That is:

$$\mu^T=(\mu_1^{(r)},\mu_2^{(s)})^T \quad , \ r+s=p$$

Also let partition the transformed estimated mean space into:

$$Y^T=(Y_1^{(r)},Y_2^{(s)})^T$$

Suppose further that the given null hypothesis is H_0: $\mu=0$ against the alternative hypothesis of H_1: $\mu_2^{(s)}=0$. Equivalently if the parameter mean and variance space are defined as:

$$\Omega_0=\{(\mu,\Sigma);\mu=0,\Sigma>0\}$$

and

$$\Omega=\{(\mu,\Sigma);\mu_2^{(s)}=0,\Sigma>0\}$$

Then the test of hypothesis becomes:

H_0: $(\mu,\Sigma)\in\Omega_0$
H_1: $(\mu,\Sigma)\in\Omega$

When **Y** and **S** are independent, the likehood-ratio (λ) for testing the above hypothesis is given by [26]

$$\lambda=\left(\frac{1+Y_2^{(s)T}S_{22}^{-1}Y_2^{(s)}}{1+Y^TS^{-1}Y}\right)^{n/2} \qquad (11)$$

Now if L is defined as:

$$L=\frac{1+Y^TS^{-1}Y}{1+Y_2^{(s)T}S_{22}^{-1}Y_2^{(s)}} \qquad (12)$$

Srivastava & Worsley [26] have shown that under H_0 the distribution of $\frac{(n-p+1)(L-1)}{r}$ is a central F distribution with r and n-p+1 degrees of freedom. Thus, if $\frac{(n-p+1)(L-1)}{r}$ is greater than or equal to $F_{\alpha(r,n-p+1)}$, then the null hypothesis is rejected, and it cannot be rejected that the first r parameters are not the cause of deterioration.

Based on the above explanation, the following algorithm is developed to detect the parameters that have caused the "out-of-control" signal

Cause Selecting Algorithm:

Step I: Calculate L_1(I) for all subsets that have {(1), (p-1)} elements. Choose

$L_{max}(I)=Max_i(L_I(I))$. If $\frac{(n-p+1)\,L_{max}(i)-1]}{r}\le F_{\alpha(r,\,n-p+1)}$, go to step II. Otherwise, stop.

Only the ith parameter requires attention.

Step II: Calculate L_2(i, j) for all subsets that have {(2), (p-2)} elements. Choose

$L_{max}(i,j)=Max_{i,j}(L_2(i,j))$, If $\frac{(n-p+1)\,L_{max}(i,j)-1]}{r}\le F_{\alpha(r,\,n-p+1)}$, go to step III. Otherwise, stop. Only jth and ith parameters require attention.

Step III: Similar to step II, but test $L_{max}(i,j,k)$.

……

……

Step p: If the final $L_{p-(p-1)}(I,\ldots)$ test is NS (not significant), all p parameters require attention.

Note that the number of iterations in the algorithm is 2^p-1, which increases dramatically as the number of parameters in each stage increases. However, in practical situations, p is restricted to values around 3 or 4 and this does not cause any difficulty.

The above algorithm was applied with a significance level of $\alpha=0.05$ to a set of data generated from a multivariate normal distribution with different parameters. The algorithm responded effectively as the values of some parameters consciously changed.

Phase 3: Detecting the out of control stage

In order to analyze dependent production system, i.e., systems in which the quality of the current stage, depends to some extent on the quality of the preceding stages, Zhang [23] proposed two kinds of quality: **(1)** Total quality: contributed by the current stage of the operation and all of the preceding stages, and **(2)** Specific quality resulting from the current operation itself, not including the influence of the preceding stages, hence it is only a part of total quality. The fundamental thinking is to compare these two kinds of quality at any operation. The key point, as Zhang demonstrated, is how to measure or judge these two kinds of quality. He proposed two kinds of control charts: the Shewhart chart for the total quality and the cause-selecting (cs) chart for the specific quality.

In this phase a cause-selecting chart divides the assignable cause further into the controllable part and the uncontrollable part. For example, the effect of the preceding stage may be taken as an uncontrollable assignable cause and those assignalbe causes specifically belonging to the current stage itself may be taken as the controllable assignable causes.

Suppose, in general, the quality index to be controlled, y_i, has a normal distribution with a mean of μ_i and variance of σ_i^2 and there are some causes (chance cause and the uncontrollable assignable cause) such as x_i that the parameters of y_i are under the effects of the uncontrollable assignable cause by the following functions:

$$\mu_i = F(x_i) \quad ; \quad \sigma_i^2 = G(x_i)$$

These functions may be found by regression. Thus the observed value of the quality index has a family of normal distribution and the following transformation may be proposed to transform y_i into the cause selecting values y_{csi} with a standard normal distribution:

$$y_{csi} = \frac{y_i - \mu_i}{\sigma_i} \tag{13}$$

where its parameters are independent of the suffix, i, under large samples if the estimated values of μ and σ are used. Now, whenever there is a controllable assignable cause, a cause selecting chart can be used to indicate any abnormality. In special cases where $G(x_i)$ is constant, Zhang [23] obtained the control limits of this chart.

Based on the idea given in the preceding paragraphs, suppose the quality index, the value of Hotelling T^2 of the current stage in each subgroup, has a linear relation with Hotelling T^2 of the previous stage as:

$$T^2_{current} = \alpha + \beta T^2_{prior} + \varepsilon \tag{14}$$

where ε has a normal distribution with a mean of zero and a constant variance. Using the set up data, the estimated regression parameters (a,b) in phase 0 of the algorithm were calculated. As a result, for each current stage there will be:

$$\hat{T}^2_{current} = a + b\, T^2_{prior} \tag{15}$$

Then if T_D is defined as:

$$T_D = \frac{(\hat{T}^2_{current} - T^2_{current})}{S_d} \tag{16}$$

where S_d is the sample variance of $T^2_{current}$ in the setup phase, the absolute values of T_D greater than or equal

TABLE 1. Decision Criteria for a 2-Stage System.

Case	Prior Stage	Current Stage	Decision
1	In Control	In Control	Both stages are In Control
2	In Control	Out of Control	Only current stage is Out of Control
3	Out of Control	In Control	Only Prior stage is Out of Control
4	Out of Control	Out of Control	If the value of T_D is large enough, both stages are Out of Control: Else only prior stage is Out of Control and we have a "not controllable cause" in the current stage

to $t_{a/2,k-a-1}$ can show the current stage to be an out of control stage ("a" denotes the number of eliminated subgroups is phase zero of the algorithm). Otherwise, the large value of $T^2_{current}$ is under the effect of the process input that comes from the previous stages.

In summary, the situations and the dercision criteria in a 2-stage manufacturing system are given in Table 1.

From Table 1 it maybe seen that considering cases 3 and 4 is not necessary. However, in many manufacturing systems, the production process is continuous and the quality tests are time consuming. In this case, before the results of the tests are known, the product is being processed in the next stage. As a result, while the current stage is in or out of control, the previous stage might also be out of control.

A Computer Software for Application of the Algorithm

A computer software has been developed for the application of the algorithm. This software, with database, computational, and graphical features, is named MVQCS(Multi .Variate Quality Control Sofware) and is programmed on the Visual Basic (VB) Programming Software. The reasons for the selection of VB are that, not only it runs under Windows, and therefore takes advantages of graphical capabilities supported in Windows, but also can interface with MS Excell 5.0 to do the database and computational tasks intended in the MVQCS.

To apply the software on a multivariate multistage manufacturing system, the user must take two basic steps. In step one, which is called the ***setup*** step, the user is questioned to type information regarding the number of stages, names of the stages, number of parameters in each stage, and number of samples in each group of data. Then the historical data from an under control system is entered and at the end the software does the estimating of the vector mean, the variance-covariance matrix and the regression parameters. Step two, which is called the ***run*** step, is performed in three different programs. First in ***runtime entry*** the information regarding the values of the parameters corresponding to a specific group of data in a specific stage is entered. At this point the stage is tested for being in or out of control. If it is out of control, then in the ***cause selecting*** program, the parameter that caused the out of control signal is detected. Finally, if two subsequent stages are both out of control, in the ***multistage control*** program it is decided that which one of the stages is out of control.

Numerical Example

Consider a two-stage manufacturing system with two parameters in each stage. Sample data for 20 subgroups, each containing 3-sample value of parameters number one and two in stage one (X_{1_i}) are shown in Table 2. These data along with the hand calculations for estimating the vector mean, the variance-covariance matrix of the vector mean, and the T^2 values are taken exactly from the text written by Ryan [14]. Also the sample values of the parameters in stage 2 (X_{2_i}) are generated using data from stage one by:

$$X_{2_i} = 2X_{1_i} + Y, \quad Y \approx \text{Normal}\ (0,\ 1) \tag{17}$$

Table 2 shows part of the data. Note that the sample values of the parameters in stage 2 are correlated with the ones in stage one and the i.i.d. standard normal

TABLE 2. The Setup Data.

Stage	Group	Parametere	Sample	Value	Stage	Group	Parameter	Sample	Value
1	1	1	1	72	2	1	1	1	144.846
1	1	1	2	84	2	1	1	2	169.799
1	1	1	3	79	2	1	1	3	159.403
1	1	2	1	23	2	1	2	1	46.1874
1	1	2	2	30	2	1	2	2	59.7094
1	1	2	3	28	2	1	2	3	55.7264
1	2	1	1	56	2	2	1	1	113.028
1	2	1	2	87	2	2	1	2	175.071
1	2	1	3	33	2	2	1	3	62.0270
1	2	2	1	14	2	2	2	1	28.8599
1	2	2	2	31	2	2	2	2	63.1718
1	2	2	3	8	2	2	2	3	16.2314
1	3	1	1	55	2	3	1	1	109.064
1	3	1	2	73	2	3	1	2	145.118
1	3	1	3	22	2	3	1	3	43.7498
1	3	2	1	13	2	3	2	1	27.1893
1	3	2	2	22	2	3	2	2	44.0765
1	3	2	3	6	2	3	2	3	12.0819
1	2	2	1	35	2	2	2	1	69.1943
1	2	2	2	38	2	2	2	2	76.5110
1	2	2	3	41	2	2	2	3	82.2136

variates are generated using:

$$Y_1 = \sqrt{-2\,Ln\,(U_1)}\;Cos\;(2\,\pi\,U_2)$$

$$Y_2 = \sqrt{-2\,Ln\,(U_1)}\;Sin\;(2\,\pi\,U_2) \tag{18}$$

where U_1 and U_2 are two i.i.d. uniformly distributed variate on (0,1).

The data from Table 2 were entered and the setup calculations were made by the MVQCS. Then the values of the T^2 were estimated for both stages (those of stage 1 were exactly the same as the ones estimated in the text). Figures 1 and 2 show the results from the setup phase. From the figures, it can be seen that group number 10 in both stages is out of control. This group were eliminated from the setup calculations. Hence, based on the other 19 groups the vector mean and the variance-covariance matrix for each stage were estimated and are shown in Tables 3, 4 and 5. These values will be used later in the other steps of the algorithm.

In order to test the algorithm, to the value of parameter 1 of the group 4 in the first stage, a value equal to the 3s were deliberately added, without altering the value of parameter 2 in this stage or the value of parameter 1 in stage 2. Applying the computer software all of the three parts of the algorithm responded effectively and this change was detected and the cause was only parameter number one. Figure 3 shows the values of the statistic for group number one to twenty in the setup phase. From the Figure it is clear that subgroup number 4 is completely out-of-control. Also Figure 4 shows the values of the statistics for the 20 subgroups in the phase 2 of the algorithm for this change. Although the data are the same, but the UCL in this part of the algorithm is greater than

TABLE 3. Parameter Means Estimated from the Calculations.

Stage	Parameter	Parameter Mean
1	1	60.375
1	2	18.4875
2	1	121.1966
2	2	37.13693

TABLE 4. The Variance-Covariance Matrix of the First Stage.

Stage 1	Parameter 1	Parameter 2
Parameter 1	222.0333	103.1167
Parameter 2	103.1167	56.57917

TABLE 5. The Variance-Covariance Matrix of the 2nd Stage.

Stage 2	Parameter 1	Parameter 2
Parameter 1	963.1572	420.0222
Parameter 2	420.0222	221.3535

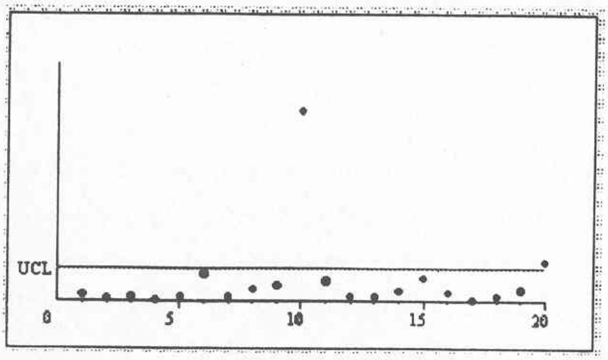

Figure 1. T^2 values of the setup phase in the first stage.

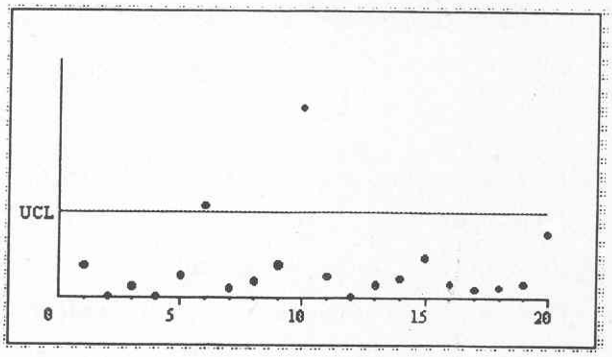

Figure 2. T^2 values of the setup phase in the second stage.

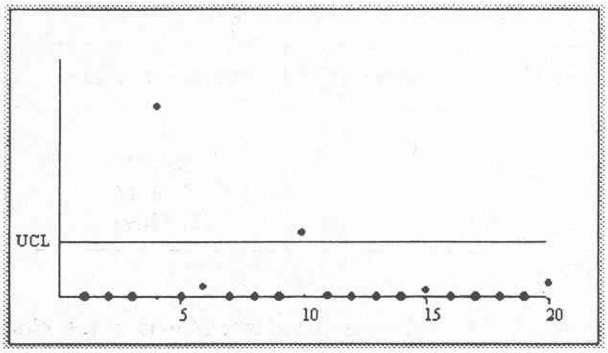

Figure 3. Values of the statistics for the first type of change in the first stage.

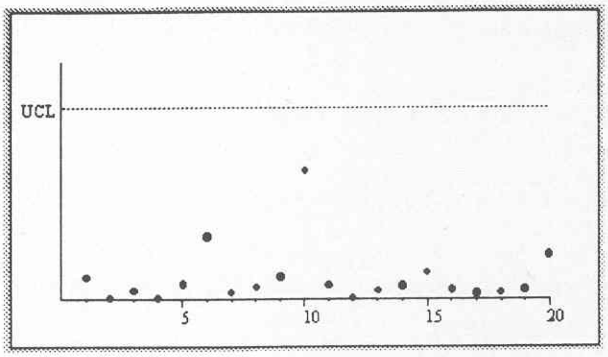

Figure 4. Values of the statistics for the first type of change in the second stage.

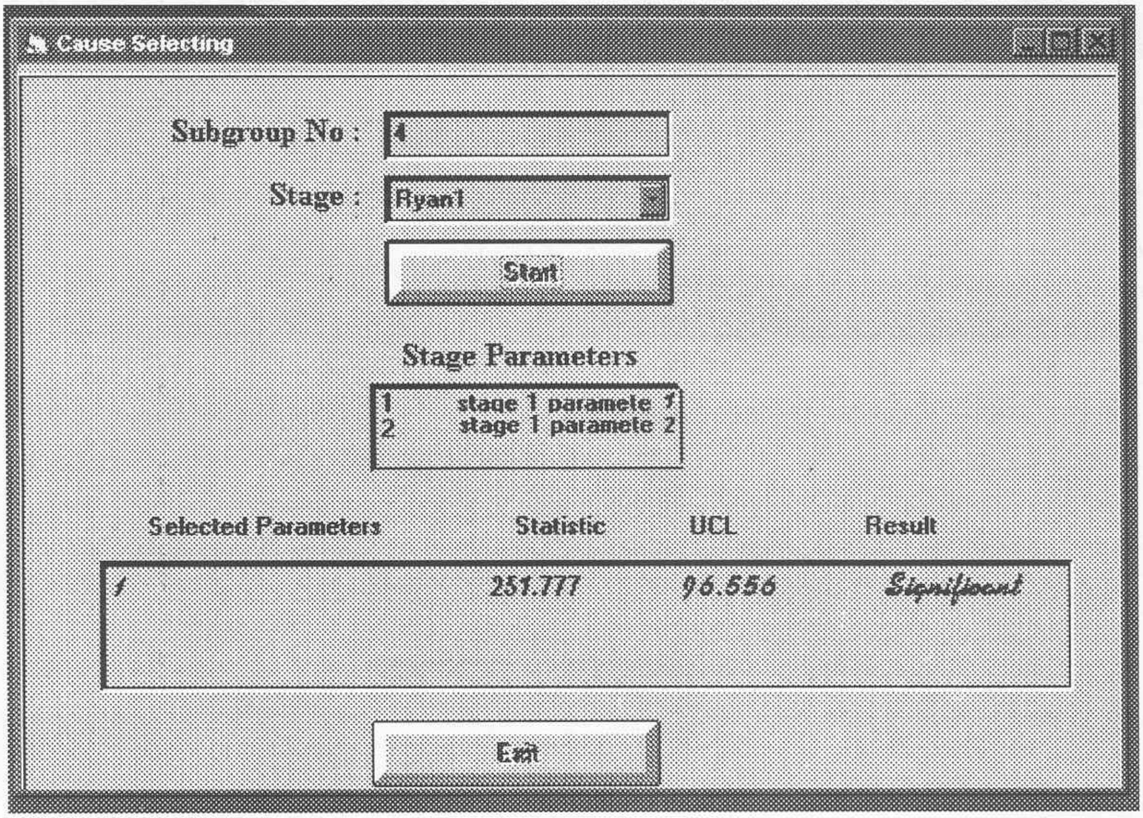

Figure 5. Results of the cause-selecting algorithm.

the UCL in the setup phase, and this is expected as it was stated previously. Besides, Figure 5 shows the results of the cause-selecting algorithm.

Furthermore, to each value of parameter number one of the group 5 in the first stage the value of 2s were added and a value of 1s were subtracted from the values of parameter number 2 in this stage. In this case the phase 1 of the algorithm responded effectively and both parameters were detected as the cause of deterioration. Figure 6 shows the results. From the figure, it can be seen that group number 5 is completely out-of-control.

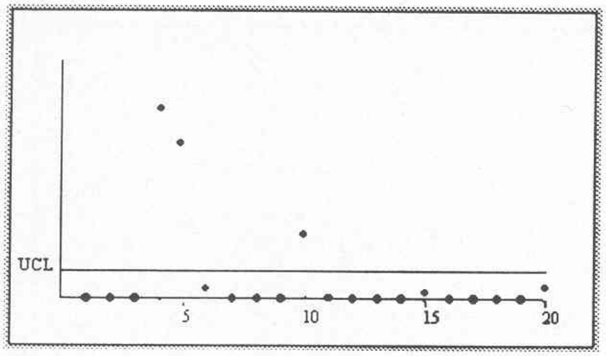

Figure 6. Values of the statistics for the second type of change in the first stage.

Application of the Algorithm to the Real Data:

At this point the algorithm was applied to the real data from a detergent-making company which is one of the five major companies in Iran in this filed. Although this company has a well-equipped quality control laboratory, but never was a statistical quality control procedure used. The process in this company has seven major steps in each of which there are some quality characteristics to be monitored. For example, in the Solfunation process four correlated parameters are of high importance and are monitored; free oil percentage, color, acidity percent and the acidity number. The standard range for these parameters are (9%-15%), (65-70), (96%-98%), and (1.1-2.1) respectively. The quality control section of the company samples the produced acid in every hour of the operation. Then they record the result of the tests in a personal computer using Quatro-Pro computer software. Note that the acidity number is calculated from the acidity percent by a deterministic mathematical formula (assuming fixed purity of acid). Hence while the acidity percent being considered as a quality characteristic, the acidity number can not be considered as a separate parameter, since all of its variations is considered in the acidity percent.

The data for the four months of the operation in 1373 (Aban, Azar, Dey and Bahman) of this company were considered in this project. These data then were re-structured in a form suitable for the software. The data for the first 19 days of Dey were considered as the data for the set-up phase and entered in the designed software. The results of the company's test on the above three parameters in each day were considered as a subgroup. Then the algorithm was run on these data. Figure 7 shows an interesting and surprising result. Almost all of the subgroup means are showing to be out-of-control, even though these subgroups were used in the setup phase. First it was thought that the algorithm did not work well. At this step, to further test the software, multinormal random variates were generated. The results, which are discussed in the next part, show that the algorithm did work well. Then the population from which the data were gathered was tested for normality. To do this we applied the Excell 5.0 software and plotted the observations. Figure 8 shows the result. From the figure it can be seen that the observations almost come from a bimodal normal distribution, and since the whole algorithm is based on the assumption of a multi-normal distribution, the algorithm did not work well. Further investigations on the data were made at this point and it was found that the data from the month of Dey were collected from a different plant of the company, while we used those data in the setup phase and tested the other plant of the company. This process shows how important the data in the setup phase are.

Then the data from the month of Mehr was considered as the setup data. This time using the plot, the data w ere tested for normality. Figure 9, 10, and 11 show the result for the three quality characteristics. Again Figures 10 and 11 show the pattern in a multimodal normal distribution.

Figure 12 shows the results obtained from the

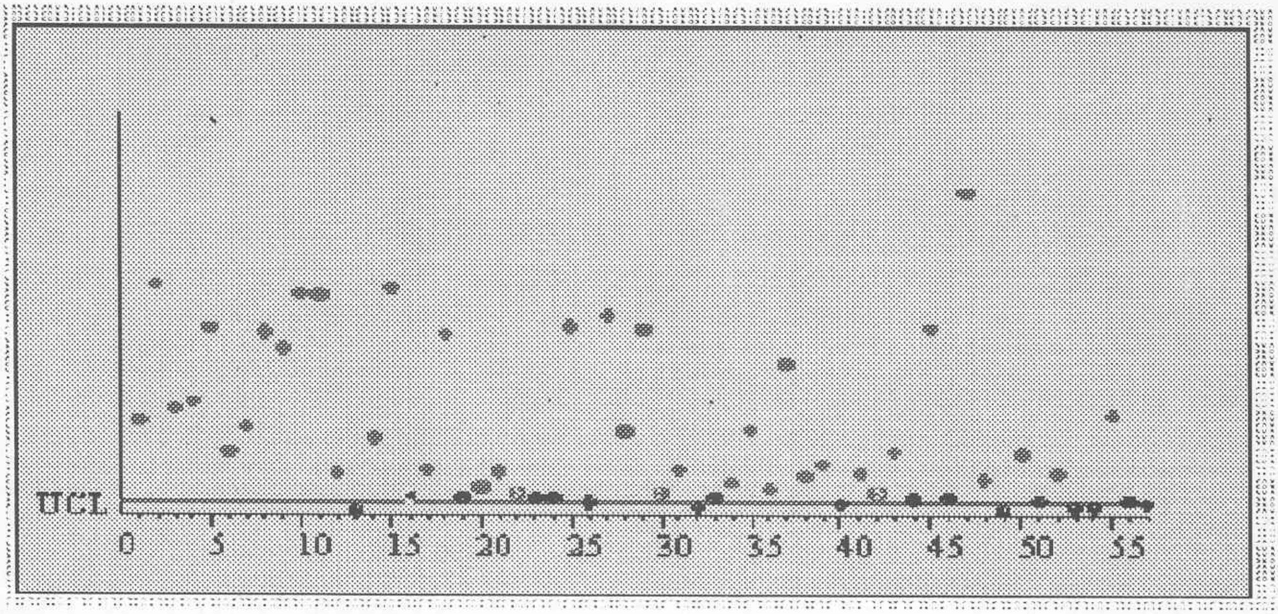

Figure 7. Values of statistics calculated based on the company's test.

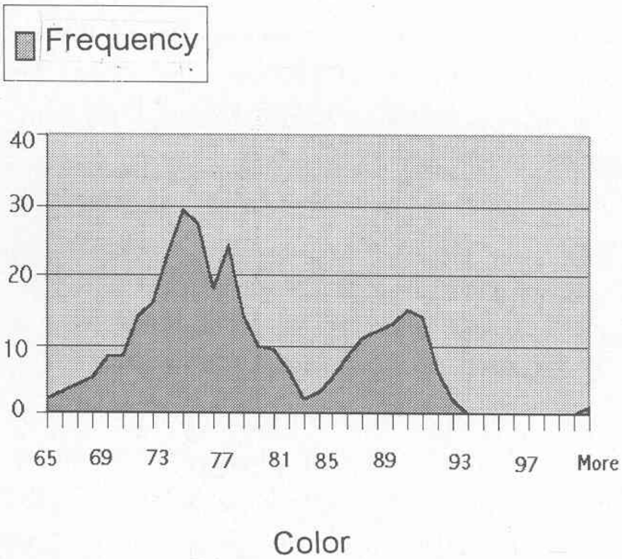

Figure 8. Color plotted observations during month of Dey.

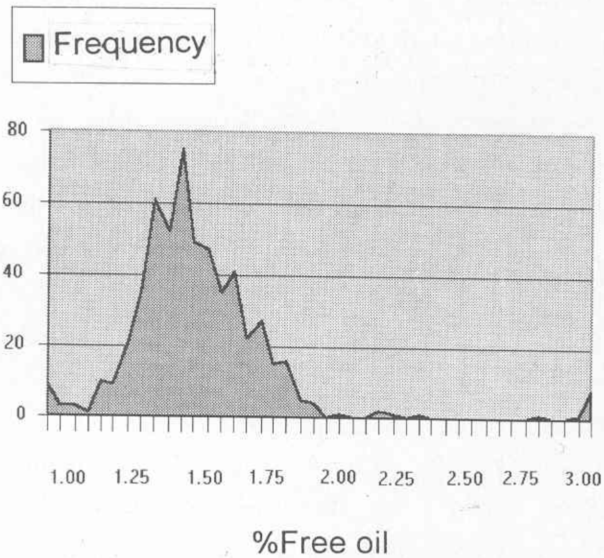

Figure 9. Free oil percentage plotted observations during month of Mehr.

software on these data. From the figure it is clear that most of the subgroups are out-of-control. Further investigation revealed that in the period of data collection there were two different laboratories to produce the acid and hence there were two different kinds of acid production. This investigation caused the company to reform its quality control strategy and from then on they decided to test the row materials. However, since there was not an effective way to separate the acid from the different productions, and hence there were no available real data, the multi-stage part of the algorithm was not applied to this process.

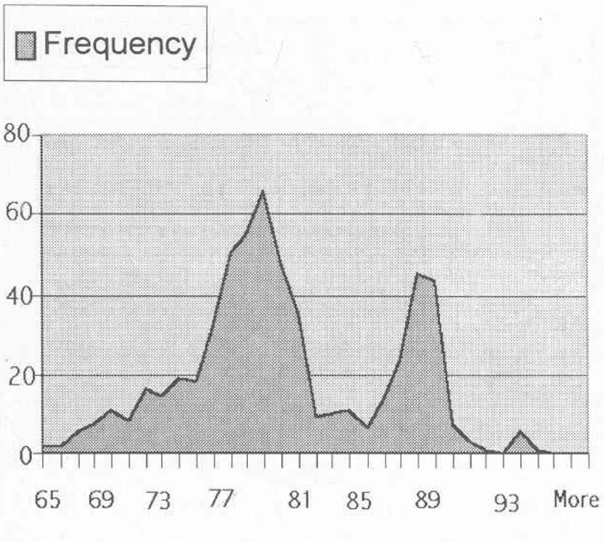

Figure 10. Color plotted observations during month of Mehr.

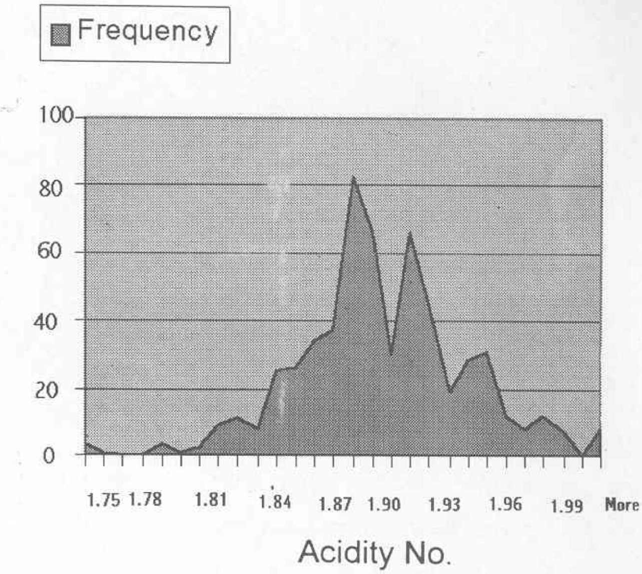

Figure 11. Acidity number plotted observations during month of Mehr.

Application of the Algorithm to Simulated Data: This part of research was originated from the facts stated in the previous part, where we found some difficulties in the application of the algorithm to the real data. At that point it was suspected that the algorithm might have some pitfalls. Hence, simulated data were used.

In order to test the algorithm, one hundred groups of multinormal variate (four variables), each group containing five samples were generated using the following parameters: [27].

$$\mu = [50, 125, 170, 20]^T \; ; \; \Sigma = \begin{bmatrix} 10.0 & 15.0 & 7.5 & 5.0 \\ 15.0 & 50.0 & 4.0 & 4.0 \\ 7.5 & 4.0 & 50.0 & 10.0 \\ 5.0 & 4.0 & 10.0 & 10.0 \end{bmatrix}$$

The generated data were used as the data for the setup phase and the following population parameter were estimated at this phase:

$$\bar{\bar{X}} = [49.66, 124.75, 169.75, 19.81]^T$$

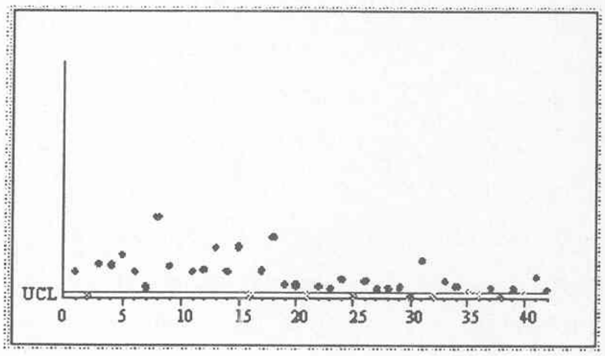

Figure 12. Values of the statistics calculated for the observations in month of Mehr.

$$S = \begin{bmatrix} 9.86 & 14.72 & 6.78 & 5.89 \\ 14.72 & 48.04 & 2.50 & 6.85 \\ 6.98 & 2.50 & 48.27 & 10.44 \\ 5.88 & 6.85 & 10.44 & 11.40 \end{bmatrix}$$

Figure 13 shows the results of the setup phase. Except subgroup number 3, 40, and 56 the remaining subgroups are in-control.

Subgroup number 56 is deleted from further consideration and the final calculation is made with

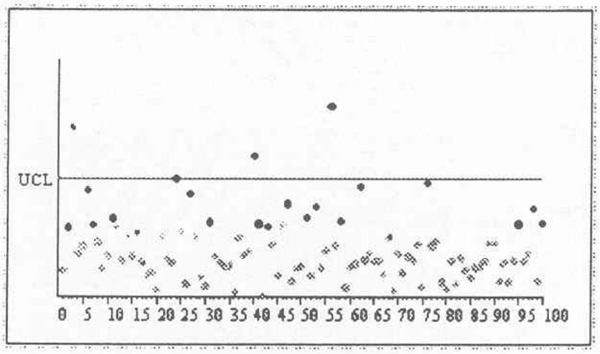

Figure 13. T^2 values of the setup phase for the generated data.

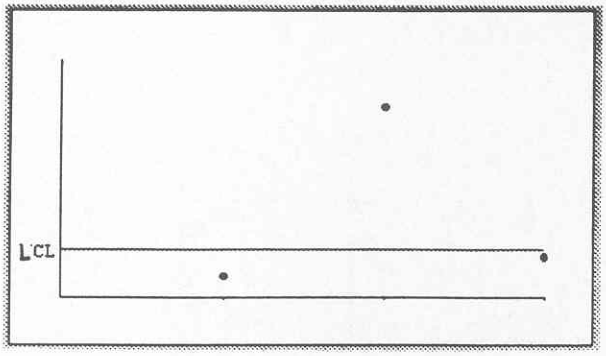

Figure 14. Values of the statistics for the manipulated data.

the information of the other subgroups and the following parameters were estimated as this point:

$$\bar{\bar{X}} = [49.68, 124.85, 169.79, 19.84]^T$$

$$S = \begin{bmatrix} 9.87 & 14.80 & 7.04 & 5.83 \\ 14.80 & 48.28 & 2.74 & 6.97 \\ 7.04 & 2.74 & 47.89 & 10.27 \\ 5.83 & 6.97 & 10.27 & 11.27 \end{bmatrix}$$

To determine the power of the algorithm, the data in phase 1 of the algorithm was manipulated in three different ways. In the first way, a value of 2σ was added to the value of the second parameter in each sample. In the second way, a value of 2σ was added to the value of the first parameter and a value of 1σ was subtracted from the second parameter in each sample. Finally in the third way, a value of 1σ was added to the value of every parameter in each sample. The results using 95% confidence are shown in Figure 14. From the figure it can be seen that the first and the third subgroups are both out-of-control but the second subgroup shows to be in control. To determine the cause of deterioration (phase 2), using 95% confidence, the cause selecting algorithm worked very well and in all of the subgroups detected the parameters which caused the out-of-control signal. Figures 15, 16, and 17 show the results.

It should be mentioned that in the second subgroup while the T^2 Statistics could not distinguish the out-of-control situation, the cause selecting algorithm of this project found the process to be out-of-control and determined the cause of the deterioration.

SUMMARY AND CONCLUSION

In idustrial quality control literature, the multi variate problem and the multistage problem, together, have not been previously investigated. In this paper, assuming that the process, which is to be monitored, possesses a multi variate normal distribution, we developed a multi variate quality control procedure in which there were several correlated stages in production systems and in each stage there were several correlated control variables. In this regard, the Hotelling's T^2 Statistic first has been applied to test the out-of-control stage. Then, based on the Murphy's method, an algorithm was developed to detect the out-of-control variables in the out-of-control stage. At the end, for the multistage processes, applying the regression concept, a new procedure has been proposed to evaluate the performance of the present and the previous stage, when both stages are showing out-of-control signals.

To examine the effectiveness of the proposed method, a computer software then has been developed. This computer software is called MVQCS and is

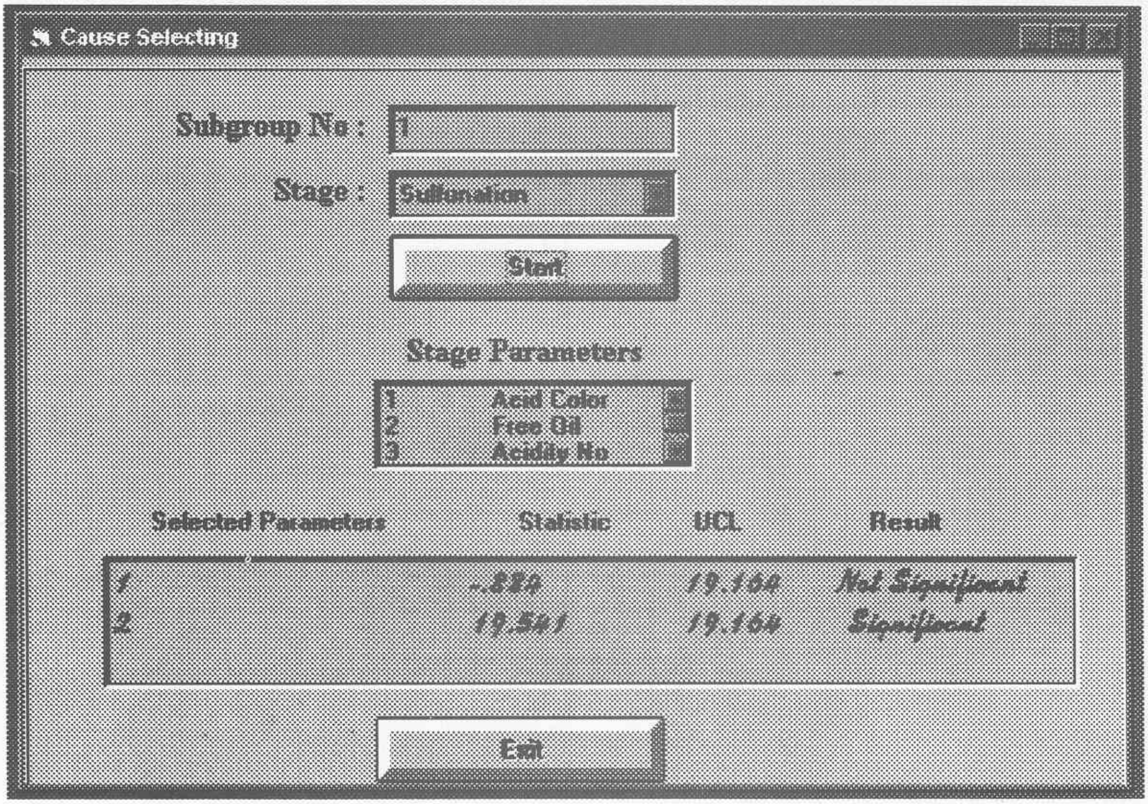

Figure 15. Results of the cause-selecting algorithm for subgroup 1.

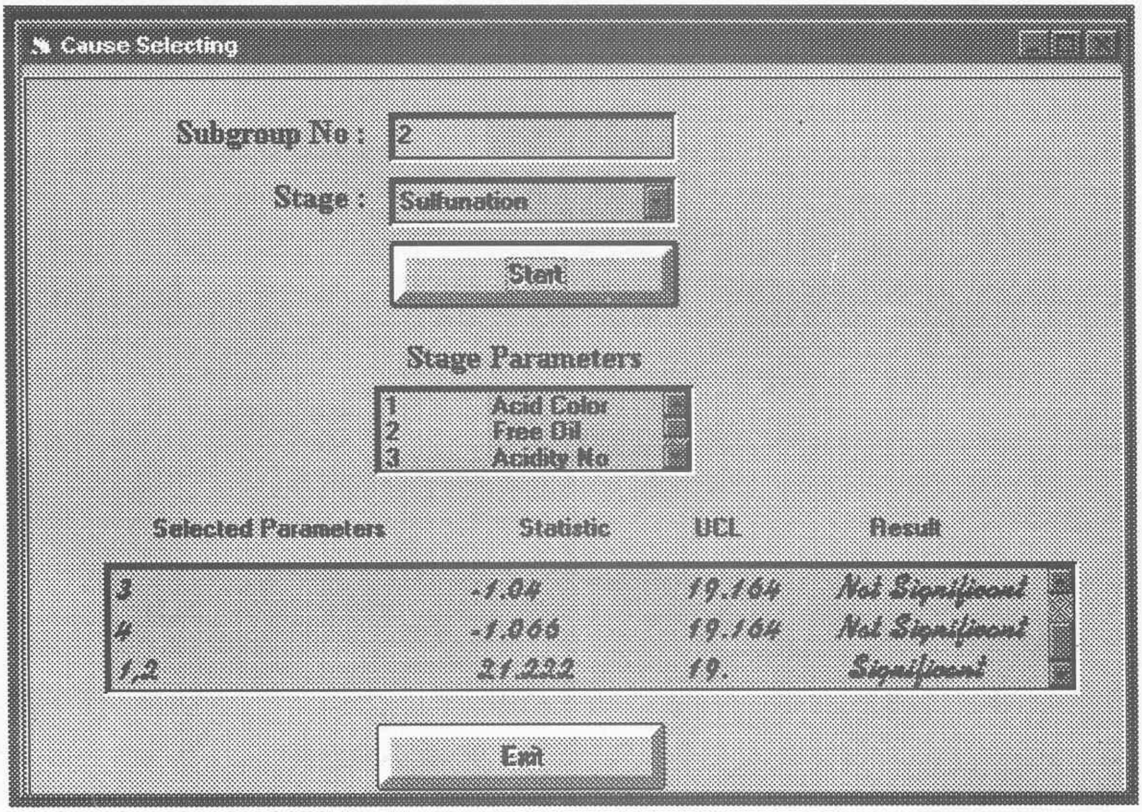

Figure 16. Results of the cause-selecting algorithm for subgroup 2.

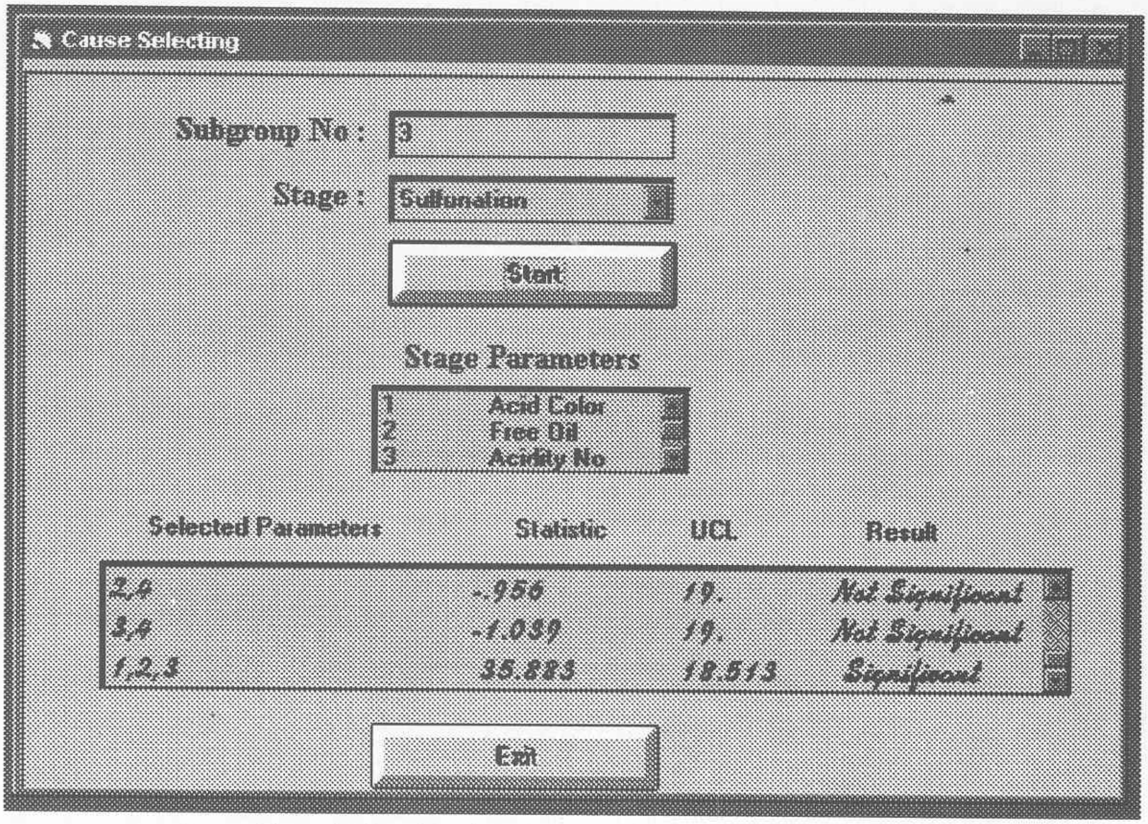

Figure 17. Results of the case-selecting algorithm for subgroup 3.

programmed on the Visual Basic Programming Language. Then, the proposed method was applied in three different situations. In the first situation, we used a numerical example along with its calculations (Ryan) and then we applied the algorithm to the example and reached the same results. In the second category, we applied the algorithm to the real data and unfortunately due to the lack of proper information could not run the phase three of the algorithm. Finally in the third situation we used simulated data. Results show that while sometimes the well-known T^2 statistic could not find an existing out-of-control situation, the proposed algorithm not only found the signal but also determined the cause of deterioration.

REFERENCES

1. W. D. Ewan and K. Kemp. "Sampling Inspection of Continuous Processes with no Autocorrelation Between Successive Results," *Biometrika*, Vol. 47, (1960).
2. D. Brook and D. A. Evans "An Approach to the Probability Distribution of CUSUM Run Length,"*Biometrika*, Vol. 59, (1972).
3. J. M. Lucas and R. B. Crosier "Combines Shewhart-CUSUM Quality Control Schemes," *Journal of Quality Technology*, Vol. 14, (1982).
4. W. H. Woodall, "The Distribution of Run Length of One-sided CUSUM Procedures for Continuous Random Variables," *Technometrics*, Vol. 26, (1983).
5. W. H. Woodall, "On the Markov Chain Approach to the Two-sided CUSUM Procedure," *Technometrics*, Vol. 26, (1984).
6. H. Hotelling, "Multivariate Quality Control" *Techniques of Statistical Analysis*, (C. Elsenhart, M. W., Hastay, & W. A. Wallis eds.), McGraw-Hill, New York, (1974).
7. J. E. Jackson, "Quality Control Methods for Several Related Variables," *Technometrics*, Vol. 1, (1959).
8. D. M. Hawkins, "Multivariate Quality Control Based on

Regression-Adjusted Variables," *Technometrics*, Vol. 33, No. 1, (1991).

9. D. M. Hawkins, "Regression Adjusted for Variables in Multivariate Quality Control," *Journal of Quality Technology*, Vol. 25, No. 3, (1993).

10. N. Doganaksoy, F. W. Faltin and W. T. Tucker "Identification of Out-of-Control Quality Characteristics in a Multivariate Manufacturing Environment," *Communications in Statistics-Theory and Methods*, Vol. 20, (1991).

11. B. J. Murphy, "Selecting Out of Control Variables with the T^2 Multivariate Quality Control Procedure," *The Statistician*, Vol. 36, (1987).

12. J. D. Healy, "A Note on Multivariate CUSUM Procedures," *Technometrics*, Vol. 29, No. 4, (1987).

13. W. H. Woodall, and M. M. Ncube "Multivariate CUSUM Quality Control Procedures, *Technometrics*, Vol. 27, No. 3. (1985).

14. T. P. Ryan, *Statistical Methods for Quality Improvement*. John Wiley & Sons., New York, (1989).

15. N. D. Tracy, J. C. Young and R. L. Mason "Multivariate Control Charts for Individual Observations", *Journal of Quality Technology*, Vol. 24, No. 2, (1992).

16. J. E. Jackson, "Principal Component and Factor Analysis", *Journal of Quality Technology*, Vol. 12, No. 4, (1989).

17. T. Chang, "Statistical Control of Correlated Variables". *ASQC Quality Congress Transactions*, Milwaukee, (1991).

18. E. Anderson, "A Semigraphical Method for the Analysis of Complex Problem," *Technometrics*, Vol. 2, (1960).

19. D. F. Andrews, "Plots of High-Dimensional Data," *Biometrics*, Vol. 28, (1972).

20. H. Chernoff, "The Use of Faces to Represent Points in K-Dimensional Space Graphically," *Journal of American Statistical Association*, Vol. 74, (1973).

21. J. S. Chamber, W. S. Cleveland, B. Kleiner and P. A. Tukey "Graphical Methods for Data Analysis," Wadsworth Publishing Co., Inc., Belmont, CA, (1983).

22. D. M. Scott, "Displaying Multivariate Data Using Polyplots," *Journal of Quality Technology*, Vol. 19, No. 2, (1987).

23. G. Zhang, "A New Diagnosis Theory with Two Kinds of Quality," *Total Quality Management*, Vol. 1, No. 2, (1990).

24. F. B. Alt, "Multivariate Quality Control, "Encyclopedia of Statistical Sciences, Vol. 6 (S. Kotz and N. Johnson, eds.), John Wiley & Sons, New York, (1985).

25. T. W. Anderson, "*An Introduction to Multivariate Statistical Analysis*," 2nd. ed., John Wiley & Sons, New York, (1984).

26. M. S. Srivastava and K. J. Worsley "Likelihood Ratio Tests for a Change in the Multivariate Normal Mean," *Journal of the American Statistical Association*, Vol. 81, (1986).

27. G. S., Fishman, "Principles of Discrete Event Simulation", John Wiley, (1978).