

## A STATISTICAL CONCEPT IN DETERMINATION OF THRESHOLD VALUE FOR FUTURE DIAGNOSIS IN MTS: AN ALTERNATIVE TO TAGUCHI'S LOSS FUNCTION APPROACH

Prasun Das<sup>1)</sup>  
Shubhabrata Datta<sup>2)</sup>

1) Indian Statistical Institute  
SQC & OR Division  
203 B. T. Road  
Kolkata 700108, India  
2) School of Materials Science  
& Engineering  
Bengal Engineering and  
Science University, Shibpur,  
Howrah 711 103, India

**Abstract:** Mahalanobis-Taguchi System (MTS) is a pattern information technology, which has been used in different diagnostic applications to make quantitative decisions by constructing a multivariate measurement scale using data analytic methods, without any assumption regarding statistical distribution. It uses a threshold value computed through quadratic loss function approach for the future diagnosis. However, the procedure for determining the threshold is lacking statistical explanation and primarily dependent on the domain expertise, if available, from the respective field. This study makes an alternate attempt to determine a threshold value using the property of multivariate statistical distributions related to Mahalanobis D<sub>2</sub>, used for MTS, and the critical values of the distributions for different levels of significance. The findings of the study can be used to eliminate the subjectivity of MTS in compliance to its data analytic philosophy.

**Keywords:** Threshold, Functional limit, Quadratic loss function, Mahalanobis D<sub>2</sub>, Multivariate upper control limit

### 1. INTRODUCTION

In most of the multi-dimensional systems, the objective is to make a decision based on several characteristics or variables. In MTS approach, Mahalanobis distance (MD), a multivariate measure (Mahalanobis, 1936) is used to measure the degree of abnormality in multi-dimensional patterns and the principles of Taguchi methods are used to evaluate the accuracy of predictions based on the scale constructed (Taguchi and Jugulum 2000). After construction of the measurement scale, also called the Mahalanobis Space, future diagnosis is performed based on a limiting value, known as 'Threshold'. Taguchi computes threshold value based on Quadratic Loss Function (QLF) approach developed.

This method of determining the threshold has already been criticized for its lack of clarity with certain terms used in the method, and also for 'subjective' nature of it (Woodall et al. 2003). Although, MTS is independent of any assumption regarding statistical distribution, certain aspect of MTS can be statistically proved for the data being multivariate normal. The work developed by Tracy et al. 1992 in the area of multivariate control chart for individual observation in the start-up stage of a process, utilizes the properties of exact and approximate distribution of the charting statistic to develop the control limits.

In this paper, we try to develop an alternate to the 'threshold' computed using QLF approach of Taguchi, for the future diagnosis in MTS, by providing a UCL as in multivariate control charts through an example with steel rolling mill data. In Section 2, an overview of MTS

is provided. In Section 3, procedure for computing threshold using QLF approach, is discussed briefly along with its shortcomings. Section 4 is devoted to the concept of multivariate control chart for individual observations along with description of relationship between control chart and MTS philosophy. Finally, in Section 5, we illustrate the effectiveness of UCL obtained from exact or approximate distribution of Hotelling's T<sup>2</sup>, as a possible alternate to the 'threshold' value computed using QLF approach, to use in MTS. We close in Section 6 with conclusive comments.

### 2. MAHALANOBIS-TAGUCHI SYSTEM (MTS)

The main objective of MTS is to make accurate predictions in the multidimensional systems by constructing a measurement scale. In the MTS, the Mahalanobis space (reference group) is obtained using the standardized variables of healthy or normal data. The Mahalanobis space (MS) can be used to discriminate between normal and abnormal data. Once this MS is established, the number of variables or attributes is reduced using orthogonal array (OA) and signal-to-noise (S/N) ratio by evaluating the contribution of each attribute.

The different stages of MTS method are summarized below:

#### Steps of MTS:

##### i) Construction of a measurement scale with Mahalanobis space as reference;

This is done using the data from a normal group

and calculating their MD's, whose value should be close to 1.

To construct the scale, MD of normal data are to be found out using the formula:

Let,  $X_j^i = i$ -th tuple of normal group for  $j$ -th variable;  $i = 1, \dots, k, j = 1, \dots, p$

$Z_j^i =$  standardised  $i$ -th tuple of normal group for  $j$ -th variable; where

$$Z_j^i = \frac{X_j^i - \hat{\mu}_j}{\hat{\sigma}_j}$$

$$\hat{\mu}_j = \text{mean of } j\text{-th variable} = \frac{\sum_{i=1}^k X_j^i}{k}$$

$$\hat{\sigma}_j = \text{standard deviation of } j\text{-th variable} = \frac{1}{k-1} \sqrt{\sum_{i=1}^k (X_j^i - \hat{\mu}_j)^2}$$

If  $Z(i) = (Z_1^i, Z_2^i, \dots, Z_p^i)$ ;  $i = 1, \dots, k$  and  $C_{pp} =$  Correlation matrix for  $p$  standardised variables; then, the Mahalanobis distance (MD) calculated for the  $i$ -th normal group observation (tuple) in a sample of size  $k$  with  $p$  variables is given by

$$MD(i) = D_i^2 = Z(i)^T C^{-1} Z(i)$$

In MTS, the scaled MD is used as  $MD(i) = D_i^2 = [Z(i)^T C^{-1} Z(i)]/p$ , i.e. dividing with the number of variables  $p$ .

**ii) Validation of measurement scale;**

Here, 'abnormal' data points are used to calculate the MD's whose value should be considerably greater than one.

**iii) Identification of useful variables (developing stage);**

The original variables are considered as factors of an OA and S/N ratio for each run is calculated. The variables with the positive gains are chosen as useful, and used thereafter.

Calculation of S/N ratio: when the true levels of severity of abnormal are not known, larger the better S/N ratios are used using the formula,

$$S/N \text{ ratio} = \eta_q = -10 \log_{10} \left[ \frac{1}{t} \sum_{i=1}^t \left( \frac{1}{D_i^2} \right) \right]$$

Where,  $\eta_q$  is S/N ratio for  $q^{\text{th}}$  run of the OA

$t =$  number of abnormal conditions and,  $D_i^2 =$

corresponding MD's,  $i=1$  (1)t:

**iv) Future diagnosis with useful variables;**

While monitoring the conditions of variables, the MD of each observation compared to a threshold. If the value of MD is lower than the threshold it is considered as normal; abnormal otherwise. Threshold has a great impact on the accuracy of the diagnosis process. Incorrect thresholds will result in many false alarms and thereby incurring a huge loss. So, the threshold is to be calculated so as to minimize the loss and hence is done using Quadratic Loss Function (QLF) approach (Taguchi, 1986).

**Statistical Background of the Measurement Scale in MTS**

Taguchi and Jugulum 2000 stated that the expected value of  $MD_i$ 's for the normal items is unity. This

approximation is probably based on  $\chi^2$  distribution with  $p$  degrees of freedom. This is the probability distribution of  $pMD_i$ , provided the sampling is from multivariate normal distribution with known mean and dispersion matrix. However, under the assumption of multivariate normality, Tracy et al. 1992, reported that the marginal distribution of  $MD_i$  is related to a  $\beta$  distribution and has a mean of  $(k-1)/k$ , not unity, where  $k$  is the number of normal observations. Finally, it can be shown using matrix algebra that the average value of MD for  $k$  normal items is always exactly  $(k-1)/k$ , (Woodall et al. 2003).

**3. DETERMINATION OF THRESHOLD IN MTS**

In multivariate systems, determination of a threshold is very important to carry out the diagnosis process effectively. In case of MTS, the threshold can be thought of as the safety factor below which an individual is considered as 'normal', like, a patient is healthy or a manufactured product is acceptable. The threshold is useful for future diagnosis while monitoring the conditions. It has a great impact on the accuracy of the diagnosis process. Incorrect thresholds will result in many false alarms and thereby incurring a huge loss. To minimize the total loss, value of threshold is calculated using QLF approach. In MTS, QLF for the case of smaller-the better characteristic is applicable since the target value of MD is zero.

Let,  $\Lambda_0$  be the functional limit, which is the critical distance. This value corresponds to the distance when a patient dies or when there is a manufacturing product having functional failure. Let,  $A_0$  be the loss associated with  $\Lambda_0$ . Then the loss function equation can be obtained as,

$$\text{Loss} = (A_0/\Lambda_0^2).MD$$

Let us assume that the cost of further diagnosis of the system is  $A$  units. Let,  $\Lambda$  be the distance corresponding to  $A$ . Then  $\Lambda$  can be obtained by,

$$\Lambda = \Lambda_0 \sqrt{\frac{A}{A_0}}$$

Therefore, the threshold,  $T$ , is given by,  $T = \Lambda^2$ .

For future diagnosis, the decision rule can be stated as:

If the value of  $MD$  is equal to  $T$ , then the observation may be 'normal' or 'abnormal'; if  $MD$  is below  $T$ , then the observation is 'normal'; and if  $MD$  is above  $T$ , then it is 'abnormal' (Taguchi et al. 2001; Taguchi and Jugulum 2002).

**Shortcomings of QLF Approach**

The terms used to calculate the threshold are hardly well defined and subjective in nature. The choice of the functional limit is very crucial and dependent on the domain knowledge expertise. The procedure of

determining the threshold has been exemplified in a medical case study, where the threshold is the value, which, if exceeded by a MD value, will determine that the concerned patient should have a more complete check-up (Taguchi and Jugulum 2000). Here,  $\Lambda_0$  is the ‘mid-value’ of the MD of a patient group having the subjective symptoms. But, the implication of the ‘subjective symptoms’ is not well defined (Woodall et al. 2003). Thus, determination of threshold seems to be largely dependent on the expert’s opinion of the respective field. In essence, the procedure for obtaining the threshold, inevitable for the future diagnosis, is highly subjective in nature and contradicts in this respect the philosophy of MTS, which demands to be data analytic.

**4. MULTIVARIATE CONTROL LIMITS FOR INDIVIDUAL OBSERVATIONS**

The quality of the output of a production process is often measured by the joint level of several correlated characteristics and monitored using control charts. There are two distinct phases in constructing control charts. The first phase is termed as start-up stage of the process as the purpose is to obtain a set of data (a reference sample) to establish the control limits for the monitoring purposes. The goal of the first stage is to establish the statistical control and find accurate control limits for stage two. The second phase consists of using the control chart to maintain the control, i.e., detecting any departure from the process standards.

Now, to construct a multivariate control chart based on Hotelling’s  $T^2$  statistic, for observation  $X_i$  the charting statistic used is,

$$Q_i = (X_i - \bar{X})' S^{-1} (X_i - \bar{X}).$$

where,

$X_i = (X_{i1}, X_{i2}, \dots, X_{ip})$  follows  $N_p(\mu_p, \Sigma_{p \times p})$ ,  $\bar{X}$  and  $S$  are estimated sample mean vector and sample covariance matrix

**Case-1.**

If  $\bar{X}$  and  $S$  are assumed to be the true values of  $\mu$  and  $\Sigma$ , respectively, then the statistic  $Q_i \sim \chi^2_p$ . The lower and upper control limits in this case are

$$LCL = \chi^2(1-\alpha/2; p) \dots\dots\dots(1)$$

$$UCL = \chi^2(\alpha/2; p)$$

where,  $\chi^2(\alpha/2; p)$  is the  $(1 - \alpha)$  percentile point of the  $\chi^2$  distribution with  $p$  degrees of freedom (Seber 1984).

**Case-2.**

If the  $i^{th}$  observation  $X_i$  is independent of both  $\bar{X}$  and  $S$ , then  $Q_i$  multiplied by a constant follows an  $F_{p,m-p}$  distribution. The control limits are calculated as

$$LCL = \frac{p(m-1)(m+1)}{m(m-p)} F(1-\alpha/2; p, m-p) \dots(2)$$

$$UCL = \frac{p(m-1)(m+1)}{m(m-p)} F(\alpha/2; p, m-p)$$

where,  $F(\alpha, p, m-p)$  is the  $(1 - \alpha)$  percentile point of the  $F$  distribution with  $p$  and  $m-p$  degrees of freedom.

**Case-3.**

During analysis of past data in the start-up phase, the limits described in the above two cases are not valid. In this situation, the exact distribution of  $Q_i$  is derived as follows (Gnanadesikan and Kettenring 1972).

$$Q_i \sim \frac{(m-1)^2}{m} B(p/2, \frac{m-p-1}{2})$$

Subsequently, the limits are given by,

$$LCL = \frac{(m-1)^2}{m} B(1-\alpha/2, p/2, \frac{m-p-1}{2}) \dots(3)$$

$$UCL = \frac{(m-1)^2}{m} B(\alpha/2, p/2, \frac{m-p-1}{2})$$

where,  $B(\alpha, p/2, m-p-1/2)$  is the  $(1 - \alpha)$  percentile point of the  $\beta$  distribution with parameters  $p/2$  and  $(m-p-1)/2$ .

Note that since any shift in the mean leads to an increase in the statistic  $Q_i$ , therefore the LCL can be ignored.

**Relationship between Control Chart and MTS**

In case of construction of multivariate control chart, the first phase deals with testing of the process for its in-control situation based on analysis of past individual or subgroup observations. In the event of control situation, the distribution of the process (mean vector, variance-covariance matrix) is estimated. For MTS, a measurement scale with Mahalanobis space as reference is constructed from a ‘normal’ group of observations based on calculation of MD’s. To establish the validity of measurement scale for future diagnosis, MDs are calculated for ‘abnormal’ events to test whether each value becomes considerably greater than one. Now in the second phase, in case of control charts, any departure from the process standards with respect to shifts in mean vector or changes in the covariance matrix is detected when current process is called for control. In case of MTS, the MD of each new observation of any individual is compared to a threshold, computed on the basis of QLF approach to identify the abnormality requiring diagnosis. As it appears there are plenty of similarities and difference in the two concepts as well. We have been interested to look at both and hybridize the two systems accordingly.

Now, the measurement of the MTS scale  $D_i^2$ , as defined previously, is related to  $Q_i$  as follows,

$$Q_i = p \times D_i^2$$

From Section 2 it is observed that the probability distribution of  $pMD_i$  is based on  $\chi^2$  distribution with  $p$  degrees of freedom and also the marginal distribution of  $MD_i$  is related to a  $\beta$  distribution with certain assumptions. All these relationships help us to develop our interest in working with the control limits of the

stated distributions for determination of threshold in MTS philosophy. Since, the “abnormal” observations in MTS do not belong to a different population (as in case of out-of-control points in a multivariate control chart, where out-of-control situations correspond to either of change of mean or the change of variance, signifying the formation of a different population), but are some instantaneous ill-mannered outcome of a supposedly well-behaved system. Hence, it is decided to use only the UCL to discriminate between ‘normal’ and ‘abnormal’ observations.

## 5. ILLUSTRATIVE EXAMPLE

### 5.1 Initial Database

The data comes from a process of manufacturing of steel plates, which can be broadly divided into three stages viz. iron making, steel making and plate making. After the plates are manufactured, samples are taken for test plate analysis (TPA) with respect to their mechanical properties. Depending on the conformance to the specifications of these sample plates, the corresponding heat for steel making is certified as “OK” or “Diverted”.

The diverted plates are sold in the local market at a reduced price. There are a total of 67 observations coming from TPA of the plate rolling mill with the values of the chemical compositions, i.e. 5 variables (viz., C, Mn, S, P, Si), as a result of Ladle Sample Analysis. Among them 41 are from “OK” heat and 26 from “Diverted”. For MTS, all 41 ‘OK’ data are used as data from ‘normal’ group and diverted data are considered as ‘abnormal’ observations. For generalization of study, it is decided to simulate data based on original observations of ‘normal’ group and use each simulated dataset to construct the measurement scale. The multivariate normal simulation technique is used with the mean vector and dispersion matrix of the original data set of 41 observations. Each simulated set contains 41 data and 10 such sets are generated.

### 5.2 Application of MTS

Table 1. Experimental Plan, S/N Ratios and Gains

Exptl. Run	C%	Mn%	S%	P%	Si%	S/N ratio
1	P	P	P	P	P	2.1295
2	P	P	A	P	P	2.0395
3	P	P	P	A	P	1.9800
4	P	P	A	A	P	1.5114
5	P	P	P	P	A	1.7012
6	P	P	A	P	A	2.0493
7	P	P	P	A	A	0.4356
8	P	P	A	A	A	0.3792
Gain	12.226	12.226	0.267	3.613	3.095	--

To start with, all the 41 “OK” data are considered to form the ‘normal’ or ‘healthy’ group and as the first step to MTS the MD’s are calculated to construct the measurement scale for Mahalanobis space (MS). The average of MD’s of MS is found to be 0.9756, close to one as expected.

To validate the measurement scale, same mean vector and correlation matrix obtained from the normal group are used to calculate the MDs of abnormal data. The average MD of abnormal data is found to be 3.4413, clearly greater than the average MD of the normal group, signifying the effectiveness of the discrimination ability of the scale. However, some observations with MD values close to 1 or even less than that indicate that the scale cannot distinguish between normal and abnormal in those cases.

Then using the orthogonal array (OA) and the “larger-the-better” S/N ratio, the significance of each variable can be tested and any insignificant variable, if found, can be dropped to obtain improved MD values. This is the prescribed method for selection of useful subset of variables. However, keeping in mind the importance of the presence of ‘C’ and ‘Mn’ in C-Mn steel, all the other three chemistry are tested for their influence in the steel composition, a 2<sup>3</sup> full factorial experiment is planned.

The two levels of the three chemistry are fixed as ‘present (P)’ and ‘absent (A)’. ‘C’ and ‘Mn’ are always kept at ‘present (P)’ as explained. So, altogether 8 trials are designed and for each trial, based on the combination of ‘present’ and ‘absent’ level, MD values ( $D_i^2$ ) are calculated using the abnormal observations, followed by the calculation of S/N ratio. Larger-the-better S/N ratio is calculated for each run as response and for each variable the average gain is calculated which is the effect of each attribute. Variables with positive gain will be considered as significant and those with negative gains will be dropped from the analysis hoping that their exclusion would improve the performance of the measurement scale. The experimental layout, S/N ratios for all 8 trials and the gain for each variable are given below;

As indicated by the above table, there is no negative gain resulting in taking decision for including all the five variables. Thus, it shows that all the five chemistry parameters have some influences on the final steel product.

For generalization, the measurement scale

constructed from each simulated set is tested using the original abnormal observations and their MD values are recorded for each simulation. 10 such simulation results along with the average MD's for 26 abnormal data are tabulated in *Table 2*.

Table 2. Mahalanobis Distance ('Abnormal' Group) Using Simulated Data

Sl .	MD <sub>1</sub>	MD <sub>2</sub>	MD <sub>3</sub>	MD <sub>4</sub>	MD <sub>5</sub>	MD <sub>6</sub>	MD <sub>7</sub>	MD <sub>8</sub>	MD <sub>9</sub>	MD <sub>10</sub>	Avg. (MD)
1	3.6349	6.4348	6.8375	6.2907	6.5558	7.0626	4.9747	6.142	3.6035	6.4047	5.79412
2	1.0845	1.7716	1.7873	1.1271	1.3173	2.032	1.4633	1.313	1.0381	1.6682	<b>1.46024</b>
3	2.6828	4.3612	4.4866	3.6755	4.0104	4.8849	3.5699	4.0196	2.4621	4.2188	3.83718
4	4.5044	8.6501	8.912	9.924	9.7481	8.5787	7.2078	9.5233	5.8283	9.6147	8.24914
5	1.3209	1.764	1.8679	1.3481	1.8416	1.9828	1.8642	1.5334	1.0712	1.9744	<b>1.65685</b>
6	4.4766	7.5508	8.7606	6.3318	6.8715	7.4415	5.3319	4.2706	4.0288	6.419	6.14831
7	8.5457	12.91	13.742	12.699	11.148	13.516	9.8337	9.7668	7.8984	11.549	11.1609
8	3.3522	6.3352	7.2883	4.3676	5.4911	5.9647	4.2174	3.6155	3.3509	5.363	4.93459
9	6.7287	10.068	9.5175	7.8983	6.7534	10.028	7.4099	8.4018	5.5111	9.1339	8.14506
10	2.0262	3.5943	3.3564	3.2886	3.3157	3.7078	3.1395	4.151	2.1134	3.7034	3.23963
11	2.5584	4.6309	5.0866	3.3325	3.9208	4.703	3.181	2.6156	2.3179	3.8204	3.61671
12	3.3917	5.5594	5.4872	2.7901	3.8963	5.1679	3.6589	2.7426	2.7777	3.7693	3.92411
13	2.6215	3.5679	3.9724	3.671	3.5596	4.1459	3.3405	3.2813	2.4177	3.7632	3.4341
14	3.4803	6.6176	8.1387	7.2123	7.2969	6.7698	5.7013	5.444	4.4126	7.5194	6.25929
15	2.6443	3.9446	5.1596	4.4639	5.7566	4.1198	3.4452	3.7782	2.5604	4.8719	4.07445
16	4.7039	7.5255	7.2301	3.4867	5.1828	6.6258	4.8596	3.6463	3.7722	4.7524	5.17853
17	2.8697	4.3571	4.7609	4.2633	4.1188	4.9457	3.77	3.6751	2.6536	4.3275	3.97417
18	2.7313	3.3096	3.486	1.8993	2.6363	3.5237	2.4553	2.0595	2.019	2.5337	2.66537
<b>19</b>	<b>0.1546</b>	<b>0.2584</b>	<b>0.2229</b>	<b>0.1440</b>	<b>0.1809</b>	<b>0.4045</b>	<b>0.2275</b>	<b>0.1513</b>	<b>0.1973</b>	<b>0.3056</b>	<b>0.22471</b>
20	1.8183	3.4604	3.4885	3.2026	3.1376	3.7555	2.6724	3.0717	2.0008	3.4324	3.00402
21	4.8572	8.0694	8.5296	7.6813	7.1503	8.6158	6.1545	6.4000	4.5133	7.4568	6.94282
22	1.3617	1.1726	1.3492	0.7828	0.9279	1.7967	1.389	1.0397	1.1295	1.3574	<b>1.23064</b>
23	2.4027	3.2377	3.3432	2.1586	2.4933	3.5785	2.3411	1.6668	2.1934	2.466	2.58813
24	6.0972	9.8179	10.274	8.5321	7.99	10.274	7.2198	6.9129	5.2982	8.4418	8.08579
25	2.5316	4.4354	3.9897	2.154	3.3754	4.1085	3.2686	3.0059	2.2371	3.416	3.25222
26	1.7182	3.2942	3.3661	1.561	2.823	2.666	1.9325	2.2436	1.4866	2.8669	2.39581

Clearly, the mean of the average MD values, 4.4414, is greater than the MD of 'normal' group, showing the validity of the measurement scale. To calculate the threshold value, QLF approach is taken as demonstrated in Section 3. Here, the functional limit, where the plate is considered as diverted, is taken as  $\Lambda_0 = 2.5MD$ . Corresponding loss ( $A_0$ ), which is the cost of misclassifying a 'OK' heat as 'Diverted', is Rs. 3060/-. Now, the cost associated with further processing of a diverted plate, i.e. the difference in processing costs of the two grades, is A

= Rs. 2520/-. Let,  $\Lambda$  be the MD value corresponding to A. Then  $\Lambda$  is given by,

$$\Lambda = (A / A_0) * \Lambda_0$$

We take  $\Lambda$  as the threshold value. In testing the accuracy of the scale, if MD of an observation is greater than  $\Lambda$  it will be considered as abnormal and normal if less than  $\Lambda$ . Here,  $\Lambda$  comes to be 2.059. The diagnosis process is carried out using this threshold following the rule depicted in Section 3, and the result of the classification for both original and simulated data is tabulated along with the MD values.

Table 3: Comparison of threshold based diagnosis using original and simulated data

Class original	MD(original)	Threshold=2.059	MD(simulated)	Threshold=2.059
0	1.5218	0	1.5218	0
0	0.75132	0	0.75132	0
0	0.55125	0	0.55125	0
0	<b>2.0635</b>	<b>1</b>	<b>2.0635</b>	<b>1</b>
0	0.78129	0	0.78129	0
0	1.5269	0	1.5269	0
0	0.93079	0	0.93079	0
0	1.1844	0	1.1844	0
0	1.6331	0	1.6331	0
0	0.68426	0	0.68426	0
0	0.98715	0	0.98715	0
0	0.79348	0	0.79348	0
0	0.97543	0	0.97543	0
0	0.20876	0	0.20876	0
0	0.2186	0	0.2186	0
0	1.6651	0	1.6651	0
0	1.0504	0	1.0504	0
0	0.073949	0	0.073949	0
0	0.57905	0	0.57905	0
0	0.86502	0	0.86502	0
0	0.39067	0	0.39067	0
0	1.5915	0	1.5915	0
0	1.4827	0	1.4827	0
0	1.0338	0	1.0338	0
0	1.5298	0	1.5298	0
0	0.4967	0	0.4967	0
0	0.91932	0	0.91932	0
0	1.7045	0	1.7045	0
0	1.4546	0	1.4546	0
0	0.18302	0	0.18302	0
0	0.2036	0	0.2036	0
0	1.6053	0	1.6053	0
0	0.59889	0	0.59889	0
0	0.041977	0	0.041977	0
0	0.79563	0	0.79563	0
0	0.54205	0	0.54205	0
0	0.61074	0	0.61074	0
0	1.7358	0	1.7358	0
0	1.3918	0	1.3918	0
0	1.6279	0	1.6279	0
0	1.0141	0	1.0141	0
1	4.2335	1	5.79412	1
1	<b>1.0807</b>	<b>0</b>	<b>1.46024</b>	<b>0</b>

1	2.8431	1	3.83718	1
1	6.2682	1	8.24914	1
1	<b>1.353</b>	<b>0</b>	<b>1.65685</b>	<b>0</b>
1	4.8309	1	6.14831	1
1	8.4833	1	11.16086	1
1	4.0166	1	4.93459	1
1	6.2705	1	8.14506	1
1	2.2822	1	3.23963	1
1	2.7958	1	3.61671	1
1	3.2391	1	3.92411	1
1	2.5406	1	3.4341	1
1	4.9326	1	6.25929	1
1	3.1769	1	4.07445	1
1	4.3934	1	5.17853	1
1	2.9077	1	3.97417	1
1	2.2742	1	2.66537	1
1	<b>0.1508</b>	<b>0</b>	<b>0.224708</b>	<b>0</b>
1	2.1817	1	3.00402	1
1	5.1283	1	6.94282	1
1	<b>1.0125</b>	<b>0</b>	<b>1.230643</b>	<b>0</b>
1	2.1067	1	2.58813	1
1	6.1075	1	8.08579	1
1	2.786	1	3.25222	1
1	2.0785	1	2.39581	1
<b>No. of misclassification</b>		<b>5</b>		<b>5</b>

Here, class ‘0’ represents normal observation and ‘1’ implies abnormality. The MDs corresponding to misclassification are highlighted. It is clear that, misclassifications regarding abnormal observations are exactly same for the study using original ‘normal’ data and simulated normal observations.

These findings have encouraged us to use the concept of multivariate control limits for individual observations to substitute the QLF approach of threshold described in MTS. The formulae for UCL is used as threshold, i.e., if  $5 \times MD$  of an observation is

less than or equal to UCL, its class will be ‘0’ or ‘normal’, and if the  $5 \times MD$  is greater than UCL, it will of class ‘1’ or ‘abnormal’. As a selection of UCL, all the three distributional limits, viz.,  $\chi^2$ ,  $F$  and  $\beta$  distributions have been used as trial, where  $\chi^2$  and  $F$  used as approximates and  $\beta$  as exact. The value of  $\alpha$  has been chosen between 0.05 and 0.1 for these three distributions, and the predicted result with misclassifications is tabulated below.

Table 4: Diagnosis using UCL based on different distributions and  $\alpha$ -values

Original Class \ UCL value	approx $\chi^2(0.05)$	approx $F(0.05)$	exact $\beta(0.05)$	approx $\chi^2(0.1)$	approx $F(0.1)$	exact $\beta(0.1)$
	11.07	16.755	11.586	9.2364	14.098	10.225
0	0	0	0	0	0	0
0	0	0	0	0	0	0
0	0	0	0	0	0	0
0	0	0	0	1	0	1
0	0	0	0	0	0	0
0	0	0	0	0	0	0
0	0	0	0	0	0	0
0	0	0	0	0	0	0
0	0	0	0	0	0	0
0	0	0	0	0	0	0





It has been observed from the above table that  $\chi^2$  and  $\beta$  distributions are giving better results in terms of threshold set and the result of misclassifications. In fact, the prediction results obtained using UCL based on  $\chi^2$  distribution and  $\beta$  distribution with  $\alpha = 0.1$ , are exactly the same as that depicted when using QLF approach.

This has been possible since  $p$  is relatively small to fit  $\chi^2$  distribution with small data (Hawkins 1981). However, if we use  $F$ -distribution, which is generally used for monitoring purpose of the process, the type-II error i.e. misclassifying an 'abnormal' situation into a 'normal' group would be increased.

This type of situation should always be taken care, especially with respect to the definition of 'abnormal' event as explained by Taguchi. It is also observed from Tracy et al.1992 that UCLs computed from approximated distribution like  $\chi^2$  and  $F$  are always larger than the UCL computed from exact distribution like  $\beta$ , resulting in increase of type-II error spontaneously with the increase in number of variables.

## 6. CONCLUSION

This paper addresses the shortcomings of Taguchi's QLF approach in determining the threshold value for detecting abnormals, where domain knowledge is a prerequisite.

To overcome this situation and to apply the MTS approach in general, the concept of using control limits based on established statistical distributions is used. The objective of this paper is to hybridize the statistical theories into MTS approach.

The system of MTS should not be divided in start-up phase and monitoring phase, as in case of traditional multivariate control charts, because, unlike the control chart concept here no "cleaning" process is required to establish the statistical control in the process. Also the results of misclassification after using threshold based on UCL values are found quite satisfactory. Rather, the need for detecting 'abnormal' event with the conservative estimate of threshold is much more important.

Therefore, one can opt for statistical distribution based UCL using either  $\chi^2$  or  $\beta$  distribution as threshold value, if the 'normal' group observations of the data follow approximately multivariate normal distribution, instead of determining the threshold using QLF approach based on expert's opinion to fix the value of functional limit as illustrated in MTS.

This will hold good on the basis of associated statistical distribution with MD. In essence, the subjective method of threshold determination using QLF approach can definitely be augmented by multivariate control limits to justify the philosophy of MTS, which is said to be data analytic.

## REFERENCES:

- [1] Mahalanobis, P.C. (1936), On the generalized distance in statistics, *Proceedings, National Institute of Science of India*, 2, 49-55.
- [2] Taguchi, G. and Jugulum, R. (2000), "New Trends in Multivariate Diagnosis", *Sankhya: The Indian Journal of Statistics*, 62, Series B, 2, pp. 233-248.
- [3] Woodall, W. H., Koulelik, R., Tsui, K. L., Kim, S. B., Stoumbos, Z. G., and Carvounis, C. P., (2003), "A Review and Analysis of the MTS", *Technometrics*, Feb., 45, 1, pp. 1-15.
- [4] Tracy, N. D., Young, J. C. and Mason, R. L., (1992), "Multivariate Control Charts for Individual Observations", *Journal of Quality Technology*, April, 24, 1, pp. 88-95.
- [5] Taguchi, G. (1986), *Introduction to Quality Engineering*, Asian Productivity Organization, UNI-PUB, White Plains, N.Y.
- [6] Taguchi, G., Chowdhury, S. and Wu, Y. (2001) *The Mahalanobis Taguchi System*, McGraw-Hill, New York.
- [7] Taguchi, G. and Jugulum, R. (2002), "*The Mahalanobis-Taguchi Strategy: A Pattern Technology System*", John Wiley & Sons, Inc., pp.107-116.
- [8] Seber, G.A.F., (1984), "*Multivariate Observations*", John Wiley & Sons, New York.
- [9] Gnanadesikan, R. and Kettenring, J.R., (1972), "Robust estimates, residuals and outlier detection with multiresponse data", *Biometrics*, 28, pp. 81-124.
- [10] Hawkins, D.M., (1981), "A new test for multivariate normality and homoscedasticity", *Technometrics*, 23, pp.37-44.

Received: 15.12.2009

Accepted: 05.05.2010

Open for discussion: 1 Year