Multivariate Multinomial T2 Control Chart using Fuzzy Approach

Quality of a product is often measured through various quality characteristics generally correlated. Multivariate control charts are a response to the need for quality control in such situations. If quality characteristics are qualitative, it sometimes happens that the product quality is defined by linguistic variables – where quality levels are represented by some specific words- and product units are classified into several linguistic forms categories, depending on the degree of fulfillment of expectations, creating a situation of fuzzy classifications. This paper first reviews the concepts found in the literature on the development of fuzzy multivariate control charts. We propose a method to control these fuzzy quality evaluations, with correlated multiple attributes quality characteristics, through the use of a Hotelling T^2 control chart

Keywords: Multinomial Processes, Fuzzy Theory, Hotelling T^2

1. Introduction

Product Quality is often measured through various quality characteristics generally correlated. In such cases, multivariate control procedures, considering the correlation structure between them, is appropriate. Lot of work has been done to monitor multivariate quality characteristics whose value is defined in terms of a numerical value measured over a range, usually identified as variables in quality control language, but comparatively few papers deal with controlling multivariate attributes processes.

Lu et al. (1998) proposed a control chart for multivariate attribute processes in which the quality status of a product unit is classified as conforming or nonconforming, this approach deals only with binomial quality features however, in many cases, binary classification may not be appropriate. In fact, for many products quality does not change abruptly from perfect to worthless, so there is a need for intermediate assessments for quality characteristics such as appearance, smoothness and color, which cannot be expressed numerically (Taleb, Limam and Hirota 2006). To complement binary classification, several intermediate levels can be expressed in linguistic terms. For example, a quality characteristic of a product can be classified in terms: "perfect", "good", "average", "poor" or "bad" depending on their ability to meet specifications, then, the significance of these terms is determined by fuzzy sets that associate a membership function to each of them.

A research paper incorporating uncertainty into decision analysis is basically done through the probability theory and/or the fuzzy set theory (Gülbay and Kahraman 2007). The first one represents the stochastic nature of decision analysis while the second captures the subjectivity of human behavior. A rational approach toward decision-making should take human subjectivity into account, rather than employing only objective probability measures. The fuzzy set theory is a perfect means for modeling uncertainty (or imprecision) arising from mental phenomena which is neither random nor stochastic. When human subjectivity plays an important role in measuring the quality characteristics, classical control charts may not be applicable, since they require precise information.

The use of fuzzy control charts becomes inevitable when statistical data considered are vague or affected by uncertainty, or information available is incomplete or includes human subjectivity.

In this paper, we have integrated the concepts employed in the development of Hotelling T^2 multivariate control chart and fuzzy control charts and we propose a method to design a Fuzzy Control Chart for correlated multi-attribute quality characteristics in order to simplify procedures and improving statistical modeling to avoid excessive dependence on simulation processes.

2. Multivariate Fuzzy Control Charts

Control charts for multivariate quality variables such as Hotelling T^2(Hotelling 1947), MCUSUM (Crosier, 1988) and MEWMA (Lowry 1992) have been extensively studied in the literature. Moreover, there is increased effectiveness of techniques to identify the cause of an out-of-control signal (Mason et al, 1997; Runger et al, 1996; Aparisi et al, 2006, etc.).

The application of multivariate control charts for multinomial attributes processes depends on the sampling procedure used (Taleb and Limam, 2004, 2005).

In the case where the items are classified successively with respect to all quality characteristics, i.e., sample´s items are classified by each quality characteristic separately Taleb, Limam and Hirota (2006) have suggested two approaches for the construction of process control charts for multivariate attribute when data are presented in a linguistic form. They developed two monitoring statistics T_f^2 and W^2 based on fuzzy and probability theories. The first one T_f^2 is similar to T^2 Hotteling statistics and is based on representative value of fuzzy sets. The W^2 statistics distribution, being a linear combination of dependent chi-square variables, it is obtained by Satterthwaite approximation. If each item is controlled simultaneously with respect to all the quality characteristics, then Taleb (2009) proposes two methods for the construction of control charts for monitoring multivariate process based on multi-dimensional linguistic data. The first is based on probability theory and the second on the fuzzy theory. These graphs suffer from a lack of formalization and high dependence on simulation/bootstrap processes, which makes such an essential topic in control charts as is the very definition of the control limits is performed by simulation.

When product quality characteristics are evaluated by a panel of experts who rate them according to a ranking, Kumar and Mohapatra (2012) integrate the concepts presented by other researchers (e.g. in 2005 Chen proposed a new approach to the development of fuzzy control charts by association of fuzzy numbers with scores of experts on quality characteristics) and develop a method for the modeling of fuzzy control charts to monitor a product with multiple features (attributes) that are correlated and measured by experts, on a linguistic scale (subjectively). The overall quality of the sample is calculated by adding interactive weighted fuzzy values assigned to each quality characteristic. Control charts are drawn using possibility and necessity measures, following the definition in Prade (1982).

3. Proposed Methodology

For this paper we assume that the items are classified successively with respect to all quality characteristics.

3.1 Modelling and nomenclature

Suppose that p correlated attribute quality characteristics $Q_1, Q_2, ..., Q_p$ are controlled jointly. Each $Q_j, j = 1,2,...p,$ is a linguistic variable that describe the quality of a product unit through a set of linguistic terms q_{jk}, $k = 1,2,..., s_j$, where s_j is the number of categories of the set of terms of the quality characteristic Q_j. Each q_{jk} term is associated with a fuzzy subset F_{jk} described by a membership function $\mu_{jk}(x)$ which associates to each x value of the standardized variable basis *(in fuzzy terms)*, a number in the interval [0, 1]. This number represents the degree to which the x value belongs to the fuzzy subset F_{jk}. There are several methods of preparation and selection of membership functions essentially based on statistical data (Civilnar and Trussell 1986).

It is possible to perform arithmetic operations on linguistic variables represented as fuzzy sets by applying the definitions and fuzzy mathematics techniques that have been developed extensively (Raz and Wang 1990). These require that the membership function is normal (there is at least one value of x whose degree of membership is equal to 1) and convex. For an introduction to fuzzy arithmetic, see Kaufmann and Gupta (1985). Throughout the rest of this document all membership functions associated to the linguistic terms are assumed normal, convex and standardized in the range [0, 1], where 0 represents the best possible quality, and 1 represents the worst quality, so as to ensure compliance with those conditions will be used triangular membership functions and therefore each category q_{jk} of the quality characteristics will be represented by a triangular fuzzy number.

$$F_{jk} = \left(a_{jk}, b_{jk}, c_{jk}\right) \tag{1}$$

Thus its corresponding $\mu_{jk}(x)$ membership functions would be defined as:

$$\mu_{jk}(x) = \begin{cases} 0, & if \quad x \le a_{jk} \\ \dfrac{1}{b_{jk} - a_{jk}}\left(x - a_{jk}\right), & if \quad a_{jk} \le x \le b_{jk} \\ \dfrac{1}{b_{jk} - c_{jk}}\left(x - c_{jk}\right), & if \quad b_{jk} \le x \le c_{jk} \\ 0, & if \quad x \ge c_{jk} \end{cases} \tag{2}$$

A sample A_i of n observations can then be represented by the set

$$A_i = \left\{\left\{\left(F_{11}, n_{i11}\right), \ldots, \left(F_{1s_1}, n_{i1s_1}\right)\right\}; \ldots; \left\{\left(F_{j1}, n_{ij1}\right), \ldots, \left(F_{jk}, n_{ijk}\right), \ldots, \left(F_{js_j}, n_{ijs_j}\right)\right\}; \ldots; \left\{\left(F_{p1}, n_{ip1}\right), \ldots, \left(F_{ps_p}, n_{ips_p}\right)\right\}\right\}$$

where n_{ijk} is the number of observations of the quality characteristic Q_j of the sample A_i classified by the linguistic variable q_{jk} and it holds that

$$\sum_{k=1}^{s_j} n_{ijk} = n, \quad , \quad j = 1, \ldots, p.$$

Each quality characteristic Q_j measured in a sample A_i is a multinomial variable with s_j categories such that $Q_{ij} \sim M\left(p_{ij1},\ldots,p_{ijk},\ldots p_{ijs_j},n\right)$

where

$$p_{ijk} = \frac{n_{ijk}}{n} \tag{3}$$

Using fuzzy arithmetic, it is possible to combine the fuzzy subsets of each attribute. Kaufmann and Gupta (1985) showed that the multiplication of a triangular fuzzy number (TFN) **T** and a real number k is a TFN, and adding two TFNs **T** and **S** is a TFN too, so is therefore a linear combination of TFNs obtains a TFN. For example, if **T** and **S** are respectively represented by triples (t_1,t_2,t_3) and (s_1,s_2,s_3) then a linear combination C=k_1**T**+k_2**S** will be represented by the triplet $(k_1t_1+k_2s_1,k_1t_2+k_2s_2,k_1t_3+k_2s_3)$.

Therefore by assuming that F_{jk} fuzzy variables corresponding to the k categories of the jth characteristic are TFNs obtained by equation (1) then, a linear combination can represent the quality characteristic Q_j of the sample i by a single fuzzy number provided by

$$F_{ij} = \frac{1}{n}\sum_{k=1}^{s_j} n_{ijk}F_{jk} = \frac{1}{n}\left(\sum_{k=1}^{s_j} n_{ijk}a_{jk}, \sum_{k=1}^{s_j} n_{ijk}b_{jk}, \sum_{k=1}^{s_j} n_{ijk}c_{jk}\right) \tag{4}$$

it will also be a TFN and can be written as

$$F_{ij} = \left(a_{ij},b_{ij},c_{ij}\right)$$

The sample can then be expressed as

$A_i = \{F_{i1},\ldots,F_{ip}\}$

In the literature, fuzzy control charts have been developed by converting the fuzzy sets associated with linguistic variables in scalars called representative values (Wang and Raz 1990). This conversion facilitates the layout of the observations in the graph and can be done in several ways, but the most popular methods are four, which are similar to the measures of central tendency used in descriptive statistics: the fuzzy mode, the $\alpha - level$ fuzzy midrange, the fuzzy median and the fuzzy average. However there is no theoretical backup for choosing one transformation method (Taleb et al. 2006). The choice between them should be based primarily on ease of calculation or user preference (Gülbay and Kahraman 2007).

The defuzzification method used to derive the representative value in this document is the fuzzy average f_{avg} based on Zadeh (1965), defined by

$$f_{avg} = Av(x:F) = \frac{\int_0^1 x\mu_F(x)dx}{\int_0^1 \mu_F(x)dx} \tag{5}$$

Considering triangular membership function $\mu_{ij}(x)$ corresponding to F_{ij}, similar to that calculated in equation (2), and replaced in equation (5), the representative value R_{ij} for characteristic Q_j of the sample A_i would be obtained by

$$R_{ij} = \frac{\int_0^1 x\mu_{ij}(x)dx}{\int_0^1 \mu_{ij}(x)dx} = \frac{\int_{a_{ij}}^{b_{ij}} \frac{x(x-a_{ij})}{b_{ij}-a_{ij}}dx + \int_{b_{ij}}^{c_{ij}} \frac{x(x-c_{ij})}{b_{ij}-c_{ij}}dx}{\frac{c_{ij}-a_{ij}}{2}} = \frac{a_{ij}+b_{ij}+c_{ij}}{3} \tag{6}$$

replacing in equation (6) the corresponding values of a_{ij}, b_{ij} y c_{ij} calculated by equation (4) yields

$$R_{ij} = \frac{1}{3\mathrm{n}}\left(\sum_{k=1}^{s_j} n_{ijk} a_{jk} + \sum_{k=1}^{s_j} n_{ijk} b_{jk} + \sum_{k=1}^{s_j} n_{ijk} c_{jk}\right) = \frac{1}{3\mathrm{n}}\sum_{k=1}^{s_j} n_{ijk}\left(a_{jk} + b_{jk} + c_{jk}\right) \qquad (7)$$

If we call vr_{jk} to the value given by

$$vr_{jk} = \frac{\left(a_{jk} + b_{jk} + c_{jk}\right)}{3}$$

As we can see, it would be the representative value corresponding to the membership function of the category *k* of the quality characteristic *j* which is obtained using the method of fuzzy average defuzzification. Replacing it in (7), we have

$$R_{ij} = \frac{1}{n}\sum_{k=1}^{s_j} n_{ijk} vr_{jk} \qquad (8)$$

This final expression, allows moving from a multinomial variable for each quality characteristic Q_j to a numeric value. As we can see the representative value to use is a linear combination of the n_{ijk} corresponding to the categories q_{jk} of Q_j .

Then the sample A_i of n observations would now be represented by the vector

$$\mathbf{R}_i = \left(R_{i1}, R_{i2}, \ldots, R_{ip}\right)' \qquad (9)$$

where R_{ij} approximates a normal distribution (according to the theory of large numbers) whose mean is obtained by

$$\mu_{ij} = E(R_{ij}) = E\left(\frac{1}{n}\sum_{k=1}^{s_j} n_{ijk} vr_{jk}\right) = \frac{1}{n}\sum_{k=1}^{s_j} E(n_{ijk} vr_{jk}) = \frac{1}{n}\sum_{k=1}^{s_j} E(n_{ijk}) vr_{jk} = \frac{1}{n}\sum_{k=1}^{s_j} np_{ijk} vr_{jk}$$

but from equation (3) $p_{ijk} = \frac{n_{ijk}}{n}$. then

$$\mu_{ij} = \frac{1}{n}\sum_{k=1}^{s_j} n_{ijk} vr_{jk} = R_{ij}$$

and its variance is calculated by

$$var(R_{ij}) = var\left(\frac{1}{n}\sum_{k=1}^{s_j} n_{ijk} vr_{jk}\right) = \frac{1}{n^2} var\left(n_{ij1} vr_{j1} + n_{ij2} vr_{j2} + \cdots + n_{ijs_j} vr_{js_j}\right)$$

$$var(R_{ij}) = \frac{1}{n}\sum_{k=1}^{s_j} p_{ijk}(1 - p_{ijk}) vr_{jk}^2 - \frac{2}{n}\sum_{k=1}^{s_j - 1}\left[p_{ijk} vr_{jk}\left(\sum_{l=k+1}^{s_j} p_{ijl} vr_{jk}\right)\right]$$

The process for obtaining the covariance matrix is very complex, so it requires to be estimated. The estimator to use is discussed in the next section. The set of representative values of the p quality characteristics are represented by the $p \times 1$ dimension vector $\boldsymbol{R_i}$ given in equation (9).

Like Taleb, Liman and Hirota (2006), the test statistic to be represented in the control chart for each sample is

$$T_i^2 = (\mathbf{R_i} - \boldsymbol{\mu}_\mathbf{R})' \boldsymbol{\Sigma}_\mathbf{R}^{-1} (\mathbf{R_i} - \boldsymbol{\mu}_\mathbf{R}) \quad (10)$$

where $\boldsymbol{\mu}_\mathbf{R} = (\mu_1, \ldots, \mu_p)$ is the in-control mean vector for each quality characteristic and $\boldsymbol{\Sigma_R}$ is its covariance matrix.

3.2. Parameter Estimation $\boldsymbol{\mu_R}$ and $\boldsymbol{\Sigma_R}$

Since $\mathbf{R}_i$ has a p-dimensional multivariate normal distribution with mean vector $\boldsymbol{\mu}_\mathbf{R}$ and covariance matrix $\boldsymbol{\Sigma}_\mathbf{R}$ whose calculation is very complex, these parameters will be estimated by $\overline{\mathbf{R}}$ and **S** respectively from the data matrix **R** shown in equation (15).

Montgomery (2008) distinguish two phases in the construction and use of T^2 Phase I corresponds to a pre-control step oriented to obtain initial control limits, while phase II is the monitoring of the process.

According to Chou, Mason and Young (2001), in Phase I, if $R_1, R_2, \ldots, R_m$ represent a historical data set (HDS) of m observations with sample mean vector $\overline{\mathbf{R}}$ and sample covariance matrix **S**, T^2 statistic value for R_i are obtained by

$$T_i^2 = \left(\mathbf{R_i} - \overline{\mathbf{R}}\right)' \mathbf{S^{-1}} \left(\mathbf{R_i} - \overline{\mathbf{R}}\right); \quad i = 1,2,\ldots,m \tag{11}$$

these values can be represented in a T^2 control chart with UCL calculated by

$$UCL_1 = \frac{m-1}{m} B_{\left(\frac{p}{2}, \frac{m-p-1}{2}; \alpha\right)} \tag{12}$$

where $B_{(a,b;\alpha)}$ is the $(1-\alpha)^{th}$ quantile of a Beta distribution $B_{(a,b)}$

Given a HDS of size m, and a single future observation Y in Phase II, the statistical value

$$T_i^2 = \left(\mathbf{Y} - \overline{\mathbf{R}}\right)' S^{-1} \left(\mathbf{Y} - \overline{\mathbf{R}}\right) \tag{13}$$

is represented in a T^2 control chart with

$$UCL_2 = \frac{p\left(m^2 - 1\right)}{m(m-p)} F_{(p, m-p; \alpha)} \tag{14}$$

where $F_{(p,m-p;\alpha)}$ is the $(1-\alpha)^{th}$ quantile of a Fisher distribution $F_{(p,m-p)}$

Parameters $\boldsymbol{\mu}_{\mathbf{R}}$ and $\boldsymbol{\Sigma}_{\mathbf{R}}$ must then be estimated by $\overline{\mathbf{R}}$ and **S** from the analysis of the preliminary HDS consisting of samples of size n, taken when the process is in-control.

Let´s suppose that m of such samples are available. The m data samples are stored in the table 1. Representative values for the p quality characteristics of the m samples can be expressed as

$$\mathbf{R} = \begin{pmatrix} R_{11} & R_{12} & \cdots & R_{1P} \\ R_{21} & R_{22} & \cdots & R_{2P} \\ \vdots & \vdots & \ddots & \vdots \\ R_{m1} & R_{m2} & \cdots & R_{mp} \end{pmatrix} \tag{15}$$

whose column $\mathbf{R}_{\mathbf{j}}$ (*j=1,2,...,p*) can be obtained from

Table 1. Data table for preliminary analysis.

	Q_1			…	Q_j							…	Q_p		
	q_{11}	…	q_{1s_1}	…	q_{j1}	q_{j2}	…	q_{jk}	…	q_{js_j}		…	q_{p1}	…	q_{ps_p}
A_1	n_{111}	…	n_{11s_1}	…	n_{1j1}	n_{1j2}	…	n_{1jk}	…	n_{1js_j}		…	n_{1p1}	…	n_{1ps_p}
⋮	⋮		⋮		⋮	⋮		⋮		⋮			⋮		⋮
A_i	n_{i11}	…	n_{i1s_1}	…	n_{ij1}	n_{ij2}	…	n_{ijk}	…	n_{ijs_j}		…	n_{ip1}	…	n_{ips_p}
⋮	⋮		⋮		⋮	⋮		⋮		⋮			⋮		⋮
A_m	n_{m11}	…	n_{m1s_1}	…	n_{mj1}	n_{mj2}	…	n_{mjk}	…	n_{mjs_j}		…	n_{mp1}	..	n_{mps_p}

$$\mathbf{R_j} = \begin{pmatrix} R_{1j} \\ R_{2j} \\ \vdots \\ R_{mj} \end{pmatrix} = \frac{1}{n} \begin{pmatrix} n_{1j1} & n_{1j2} & \cdots & n_{1js_j} \\ n_{2j1} & n_{2j2} & \cdots & n_{2js_j} \\ \vdots & \vdots & \ddots & \vdots \\ n_{mj1} & n_{mj2} & \cdots & n_{mjs_j} \end{pmatrix} \begin{pmatrix} vr_{j1} \\ vr_{j2} \\ \vdots \\ vr_{js_j} \end{pmatrix}$$

where vr_{jk} are the representative values of fuzzy subsets F_{jk} corresponding to the s_j categories of the quality characteristic Q_j .

The mean vector $\boldsymbol{\mu}_{\mathbf{R}}$ will be estimated by $\overline{\mathbf{R}} = \left(\overline{R}_1,\ldots,\overline{R}_{\mathrm{j}},\ldots,\overline{R}_p\right)$

where $\overline{R}_j = \frac{1}{m}\sum_{i=1}^{m} R_{ij}$ and R_{ij} is the representative value of the fuzzy subset associated with the *jth* quality characteristic of the *ith* sample, as mentioned above.

A significant issue in the case of individual observations is estimating the covariance matrix Σ (Montgomery 2008). Sullivan and Woodall (1996) provide an excellent discussion and analysis of this problem, and compare several estimators. They showed a more efficient graph that estimates the covariance matrix of the difference vector between successive observations, as suggested by Holmes and Mergen (1993), under the assumption that successive observations tend to have the same, or nearly the same mean vector, which is a more robust estimation of the in-control covariance matrix for individual observations.

Holmes and Mergen (1993) call this estimator mean square successive difference (MSSD). Under this approach, the estimator of the variance is obtained by

$$S_j^2 = \frac{1}{2(m-1)}\sum_{i=1}^{m}\left(R_{i+1\mathrm{j}} - R_{ij}\right)^2$$

and the estimator of covariance between the representative values of the quality characteristics Q_j and Q_h will be

$$S_{jh} = \frac{1}{2(m-1)} \sum_{i=1}^{m} \left(R_{i+1\mathrm{j}} - R_{ij}\right)\left(R_{i+1\mathrm{h}} - R_{ih}\right), \quad for \quad j \neq h \tag{16}$$

The covariance matrix of the sample S is then expressed by

$$\mathbf{S} = \begin{pmatrix} S_1^2 & S_{12} & \cdots & S_{1P} \\ S_{21} & S_2^2 & \cdots & \vdots \\ \vdots & \vdots & \ddots & \vdots \\ S_{P1} & S_{P2} & \cdots & S_P^2 \end{pmatrix}$$

3.3. Interpreting out-of-control signal

One difficulty encountered with any multivariate control chart is the practical interpretation of out-of-control signals. To do this, we will use the procedure proposed by Montgomery (2009) for the diagnosis of an out-of-control signal that is to break the T^2 statistic into components that reflect the contribution of each individual variable. If T^2 is the current value of the statistic, and $T_{(i)}^2$ is the statistical value of all process variables except the *ith*, Runger, Alt and Montgomery (1996b) showed that

$$d_i = T^2 - T_{(i)}^2 \tag{17}$$

is an indicator of the relative contribution of the *ith* variable to the global statistic. When an out-of-control signal is generated, it is recommended to calculate the values of $d_i (i = 1,2,\ldots, p)$ and focus on the one of the variables for which the d_i values are relatively large.

4. Numerical Examples

4.1. Frozen food example

To show the application of proposed graphic, we use the example provided in the paper by Taleb, Liman and Hirota (2006) of a food processing industry in which three quality characteristics are joint measure: appearance (Q_1), colour (Q_2), and flavor (Q_3), which sets linguistic terms are calculated by:

$$T(Q_1) = \{q_{11}, q_{12}, q_{13}\} = \{good, medium, poor\}$$

$$T(Q_2) = \{q_{21}, q_{22}, q_{23}\} = \{standard, acceptable, rejected\}$$

$$T(Q_3) = \{q_{31}, q_{32}, q_{33}, q_{34}\} = \{perfect, good, medium, poor\}$$

Membership functions associated with these three term sets are shown in Figure1. The membership functions can be represented by their corresponding fuzzy triangular numbers:

$$F_{11} = (0,0,0.25) \; F_{12} = (0,0.25,0.75) \; F_{13} = (0.25,1,1)$$

$$F_{21} = (0,0,0.5) \; F_{22} = (0,0.5,0.75) \; F_{23} = (0.5,1,1)$$

$$F_{31} = (0,0,0.25) \; F_{32} = (0,0.25,0.75) \; F_{33} = (0.25,0.75,1) \; F_{34} = (0.75,1,1)$$

The data from the process of food are shown in columns 2 to 11 of Table 2.

4.1.1 Process analysis Phase I

From these data representative values and corresponding T^2 statistic values are calculated for each sample and are summarized in the columns 12 to 15 of Table 2. The inverse of the covariance matrix **S** for the 20 samples is obtained using equation (16) and is:

$$\mathbf{S}^{-1}=\begin{pmatrix} 29825.62 & -13398.86 & -649.52 \\ -13398.86 & 25060.69 & 4826.80 \\ -649.52 & 4826.80 & 19917.60 \end{pmatrix}$$

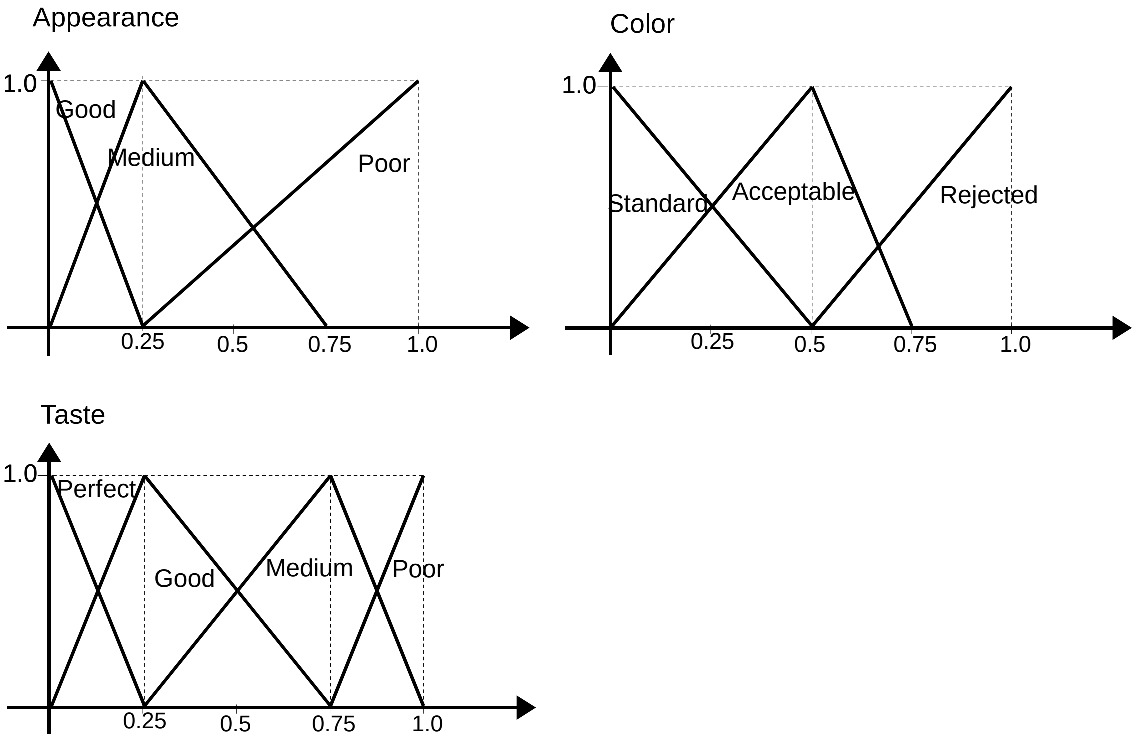

Figure2. Sets of membership functions

The upper control limit for Phase I, for a type I $\alpha = 0.05$ error is obtained by:

$$UCL=\frac{(m-1)^2}{m}B\left(\frac{p}{2},\frac{m-p-1}{2};\alpha\right)\ \alpha = 0.05, p = 3 \quad y \quad m = 20 \rightarrow B = 0.3778$$

$$UCL=\frac{19^2}{20}0.3778 = 6.82$$

Table 2 data from the process of food (taken from Taleb, Limam, and Hirota 2006), representative values of samples i and statistic

T2.

i	q_{11}	q_{12}	q_{13}	q_{21}	q_{22}	q_{23}	q_{31}	q_{32}	q_{33}	q_{34}
1	210	7	3	206	9	5	167	48	3	2
2	211	6	3	207	8	5	176	42	2	0
3	206	9	5	202	12	6	163	55	2	0

4	211	5	4	207	8	5	163	51	5	1
5	203	16	1	194	18	8	175	45	0	0
6	210	6	4	206	9	5	174	44	1	1
7	208	7	5	204	9	7	174	40	5	1
8	207	7	6	204	9	7	169	46	3	2
9	206	7	7	202	9	9	169	48	2	1
10	186	25	9	200	12	8	169	48	3	0
11	196	13	11	196	13	11	163	46	10	1
12	203	12	5	200	13	7	167	44	9	0
13	203	9	8	198	11	11	174	42	3	1
14	202	9	9	198	11	11	174	40	6	0
15	209	6	5	207	9	4	172	42	5	1
16	210	3	7	205	5	1	172	44	4	0
17	205	11	4	201	13	6	172	45	2	1
18	210	6	4	206	8	6	169	48	2	1
19	206	13	4	203	13	4	172	46	0	2
20	206	12	2	202	14	4	169	46	5	0

The control chart obtained is shown in figure 4(b), three possible assignable causes are identified and eliminated, and then, the parameters are calculated again.

4.1.2 Analysis of the process, Phase II

After clearing the samples considered out-of-control due to assignable causes, we obtain the parameters $\overline{\mathbf{R}}$ and $\mathbf{S}^{-1}$ that will be used in Phase II to calculate the statistic T^2 for additional samples. The new inverse of the covariance matrix $\mathbf{S}^{-1}$ and sample mean estimator $\overline{\mathbf{R}}$ are respectively:

$$\mathbf{S}^{-1} = \begin{pmatrix} 238033.45 & -148759.62 & -100416.03 \\ -148759.62 & 124035.97 & 11579.35 \\ -10041.03 & 11579.35 & 24350.10 \end{pmatrix}. \ \overline{\mathbf{R}} = (0.107, 0.198, 0.147)$$

The upper control limit for this phase is obtained by:

$$UCL_2 = \frac{p(m^2-1)}{m(m-p)} F_{(p,m-p;\alpha)}$$

$$\alpha = 0.05, p = 3 \quad y \quad m = 17 \rightarrow F = 3.344$$

$$UCL_2 = \frac{3(17^2-1)}{17(17-3)} 3.344 = 12.139$$

Additional data are shown in the columns 2 to 11 of the table 3. From these additional data, representative values and corresponding T^2 statistic values are calculated for each sample and summarized in columns 12 to 15 of the Table 3. Then we have the corresponding control chart for Phase II shown in Figure 5(b).

4.1.3 Interpretation of the out-of- control signal

The figure shows that the process is out-of-control when the samples 21, 22, 23 and 25 are taken. To determine which variables are responsible for each of the three cases, the corresponding d_i are calculated and shown in Table 4.

From table 6 we see that the more contributor variable to the to the out-of-control signal detected when samples 21 and 22 was taken, is the QC Q_1 (appearance), since in both cases is the variable with the highest value d_i. While in the case of the sample 23 the largest contributor is the QC Q_2 (colour) and to the sample 25 is the Q_3 (flavor)

4.2. Porcelain process example

In this example we have used the data in the paper by Taleb (2009) of a porcelain processing industry in which three quality characteristics are joint measure: appearance (Q_1), translucence (Q_2), and Whiteness (Q_3), which sets linguistic terms are:

$T(Q_1) = \{q_{11}, q_{12}, q_{13}\} = \{standard, \text{sec}ond\ choice, third\ choice\}$

$T(Q_2) = \{q_{21}, q_{22}, q_{23}\} = \{perfect\ translucence, good\ translucence, not\ translucence\}$

$T(Q_3) = \{q_{31}, q_{32}, q_{33}, q_{34}\} = \{high, medium, poor\}$

To apply the proposed methodology to these data we made a count by category independently. Like the above example, the membership functions can be represented by their corresponding fuzzy triangular numbers:

$F_{11} = (0,0,0.4)$ $F_{12} = (0.2,0.4,0.5)$ $F_{13} = (0.6,1,1)$

$F_{21} = (0,0,0.4)$ $F_{22} = (0.3,0.5,0.7)$ $F_{23} = (0.6,1,1)$

$F_{31} = (0,0,0.4)$ $F_{32} = (0.4,0.6,0.8)$ $F_{33} = (0.6,1,1)$

The data from the porcelain process are shown in columns 2 to 10 of Table 5.

4.2.1 Porcelain process analysis Phase I

From these data representative values and corresponding T^2 statistic values are calculated for each sample and are summarized in the columns 11 to 14 of Table 5. The inverse of the covariance matrix **S** for the 23 samples is obtained using equation (16) and is:

$$\mathbf{S}^{-1} = \begin{pmatrix} 4656.33 & 2717.10 & 74.98 \\ 2717.10 & 5884.70 & 1982.79 \\ 74.98 & 1982.79 & 3957.63 \end{pmatrix}$$

Table 3. Additional samples

muestra	q_{11}	q_{12}	q_{13}	q_{21}	q_{22}	q_{23}	q_{31}	q_{32}	q_{33}	q_{34}	R_1	R_2	R_3	T^2
21	202	10	8	204	11	5	169	44	5	2	0.119	0.194	0.154	45.56
22	184	25	11	206	12	2	174	44	1	1	0.145	0.186	0.140	495.96
23	208	7	5	196	13	11	174	44	1	1	0.106	0.215	0.140	37.76
24	206	6	8	196	13	11	174	40	5	1	0.114	0.215	0.146	11.21
25	210	2	8	198	12	10	165	44	1	10	0.110	0.211	0.173	35.48

The upper control limit for Phase I, for a type I $\alpha = 0.05$ error is obtained by:

$$UCL = \frac{(m-1)^2}{m} B\left(\frac{p}{2}, \frac{m-p-1}{2}; \alpha\right) \quad \alpha = 0.05, p = 3 \quad y \quad m = 23 \rightarrow B = 0.33056$$

$$UCL = \frac{22^2}{23} 0.33056 = 6.956$$

Table 4. d_i Values for interpretation of out-of-control signal

muestra	T^2	$T^2_{(1)}$	$T^2_{(2)}$	$T^2_{(3)}$	d_1	d_2	d_3
21	45.56	1.31	9.82	45.56	44.25	36.74	0.01
22	495.96	6.44	84.17	476.15	489.53	411.79	19.81
23	37.76	8.74	1.38	37.74	29.02	36.39	0.02
25	35.48	25.53	17.46	11.14	9.95	18.03	24.35

The control chart obtained is shown in Figure 2, we found that sample 11 was outside the upper control limit, the sample is removed and the parameters are calculated again.

4.2.2 Analysis of the porcelain process, Phase II

After clearing the samples considered out-of-control due to assignable causes, we obtain the parameters $\overline{\mathbf{R}}$ and $\mathbf{S}^{-1}$ that will be used in Phase II to calculate the statistic T^2 for additional samples. The new inverse of the covariance matrix $\mathbf{S}^{-1}$ and sample mean estimator $\overline{\mathbf{R}}$ are respectively:

$$\mathbf{S}^{-1} = \begin{pmatrix} 6844.75 & 3059.0 & -1082.66 \\ 3059.0 & 5658.54 & 1574.57 \\ -1082.66 & 1574.57 & 4151.20 \end{pmatrix} \qquad \overline{\mathbf{R}} = (0.330, 0.378, 0.226)$$

The upper control limit for this phase is obtained by:

$$UCL_2 = \frac{p(m^2-1)}{m(m-p)} F_{(p, m-p; \alpha)} \quad \alpha = 0.05, p = 3 \quad y \quad m = 22 \rightarrow F = 3.127$$

$$UCL_2 = \frac{3(22^2-1)}{22(22-3)} 3.127 = 10.841$$

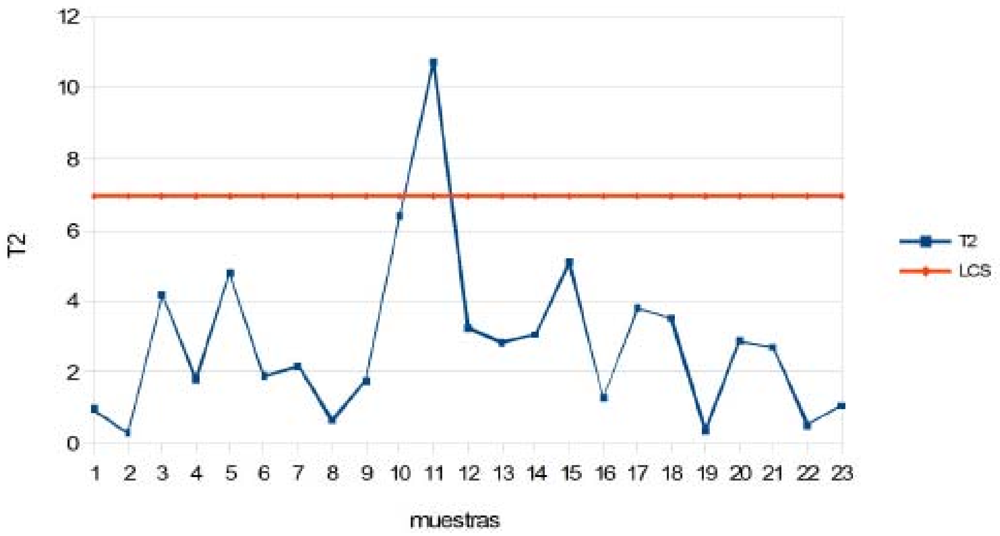

Figure 2. Fuzzy multivariate control chart phase I.

Table 5. Data from the porcelain process (taken from Taleb 2009), representative values of samples i, and statistics T2.

muestra	q_{11}	q_{12}	q_{13}	q_{21}	q_{22}	q_{23}	q_{31}	q_{32}	q_{33}	R_1	R_2	R_3	T^2
1	107	66	27	101	91	8	161	27	12	0.320	0.375	0.227	0.95
2	100	71	29	96	99	5	165	24	11	0.334	0.383	0.217	0.29
3	108	57	35	101	96	3	150	30	20	0.338	0.368	0.261	4.16
4	106	65	29	103	89	8	162	32	6	0.326	0.370	0.214	1.78
5	106	57	37	110	87	3	159	33	8	0.345	0.347	0.223	4.78
6	119	49	32	95	100	5	171	18	11	0.316	0.385	0.206	1.88
7	97	66	37	104	91	5	161	27	12	0.357	0.364	0.227	2.16
8	106	63	31	94	101	5	164	19	17	0.331	0.387	0.230	0.63
9	116	53	31	98	94	8	171	17	12	0.318	0.382	0.208	1.74
10	113	56	31	85	109	6	155	33	12	0.322	0.410	0.238	6.39
11	97	64	39	90	101	9	166	29	5	0.362	0.402	0.205	10.72
12	113	59	28	86	106	8	169	21	10	0.315	0.410	0.208	3.23
13	101	59	40	101	94	5	156	31	13	0.359	0.371	0.238	2.82
14	107	63	30	86	109	5	169	21	10	0.327	0.406	0.208	3.06
15	112	59	29	109	83	8	157	27	16	0.318	0.356	0.241	5.09
16	107	60	33	97	99	4	156	28	16	0.334	0.379	0.243	1.29
17	112	55	33	109	84	7	156	26	18	0.328	0.355	0.247	3.79
18	119	56	25	93	99	8	165	27	8	0.300	0.394	0.212	3.52
19	105	65	30	100	92	8	165	24	11	0.330	0.377	0.217	0.35
20	105	64	31	104	91	5	152	30	18	0.332	0.364	0.254	2.86
21	92	75	33	108	85	7	165	21	14	0.354	0.357	0.223	2.70
22	108	56	36	98	94	8	159	31	10	0.340	0.382	0.227	0.52
23	108	62	30	93	100	7	169	21	10	0.326	0.392	0.208	1.05

Additional data are shown in the columns 2 to 10 of the table 5. From these additional data, representative values and corresponding T^2 statistic values are calculated for each sample and summarized in columns 11 to 14 of the Table 6. Then we have the corresponding control chart for Phase II shown in Figure 3.

Table 6. Additional samples of Porcelain Process (taken from Taleb 2009) and its representative values

muestra	q_{11}	q_{12}	q_{13}	q_{21}	q_{22}	q_{23}	q_{31}	q_{32}	q_{33}	R_1	R_2	R_3	T^2
24	62	76	62	101	91	8	151	37	12	0.462	0.375	0.245	112.24
25	100	71	29	63	97	39	165	24	11	0.334	0.502	0.217	87.17
26	88	77	35	111	86	3	110	40	50	0.364	0.345	0.390	89.33

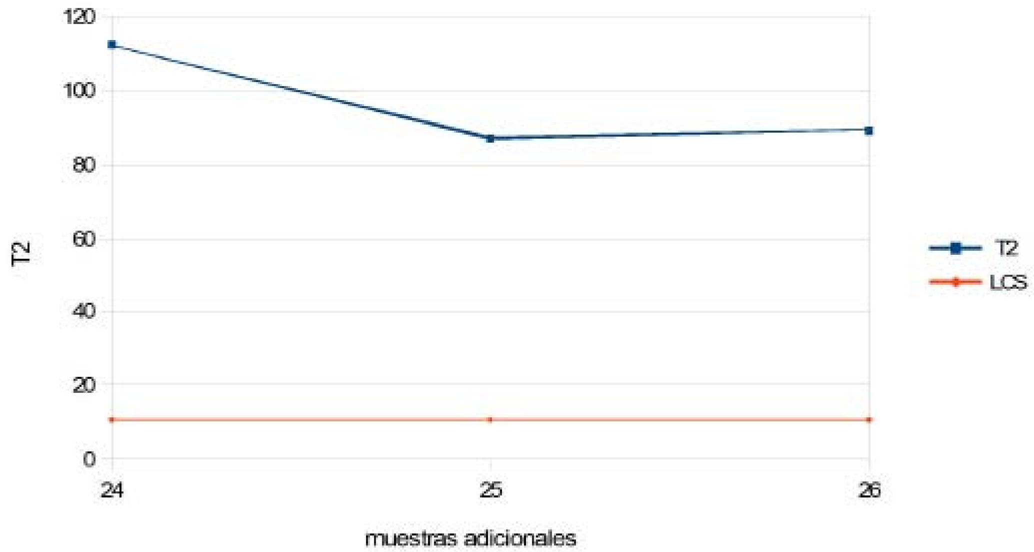

Figure 3. Fuzzy multivariate control chart for additional samples phase II.

4.2.3 Interpretation of the out-of- control signal

The figure shows that the process is out-of-control when all additional samples 24, 25 and 26 are taken. To determine which variables are responsible for each of the three cases, the corresponding d_i are calculated and shown in Table 7.

From table 7 we see that the more contributor variable to the to the out-of-control signal detected when samples 24 was taken, is the QC Q_1 (appearance), since is the variable with the highest value d_i. Likewise for sample 25 is Q_2 (translucence), while in the case of the sample 25 the largest contributor is the QC Q_3 (whiteness).

5. Discussion

As we can see in the figure 4 the results obtained by the proposed methodology are quite similar to those proposed by Taleb et al. (2006).

In phase I there is a significant overlap between the two graphs in the general shape of the curves and in identifying abnormalities, with the advantage that we use a well-known chart and avoid using methodologies based on simulation only.

Table 7. d_i values for interpretation of out-of-control signal Porcelain process

muestras	T^2	$T^2_{(1)}$	$T^2_{(2)}$	$T^2_{(3)}$	d_1	d_2	d_3
24	112.24	1.20	81.76	111.07	111.04	30.47	1.17
25	87.17	61.92	0.50	81.42	25.26	86.67	5.75
26	89.33	89.02	83.95	5.29	0.31	5.39	84.04

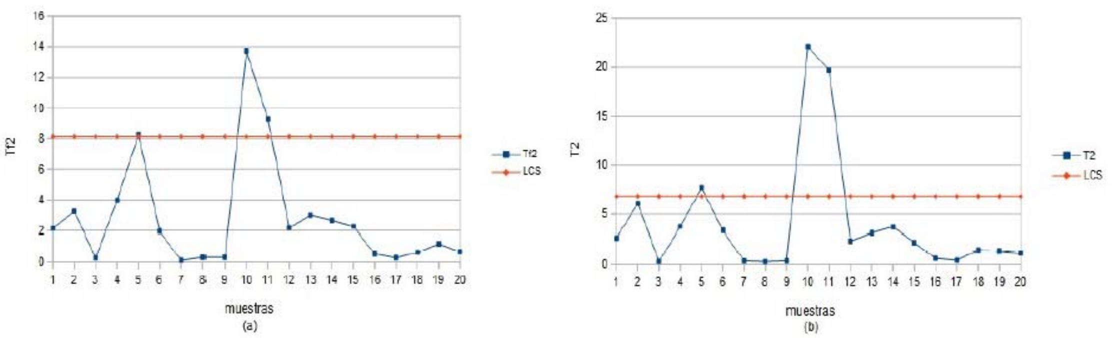

Figure 4. (a) Statistics $T^2{}_f$ (Taleb et al.1996), (b) T^2 fuzzy multivariate control phase I.

Table 8. Comparison between statisticals proposed by Taleb, Limam, and Hirota (2006) and the statistic T^2

muestra	T^2_f (LCS=8.12)	T^2 (LCS=11.80)
21	4.18	**45.56**
22	**55.03**	**495.96**
23	**8.78**	**37.76**
24	4.84	11.21
25	**21.67**	**35.48**

For additional data on the phase II shown in table 8 and figure 5, the graphics match the identification of key points that show anomalies.

Regarding to the second example, we can say that although both methodologies and statistical sampling in Taleb (2009) are different, the final results in terms of control output signal detection and identifying the variables responsible for the control output signal are the same, showing the generality of our approach.

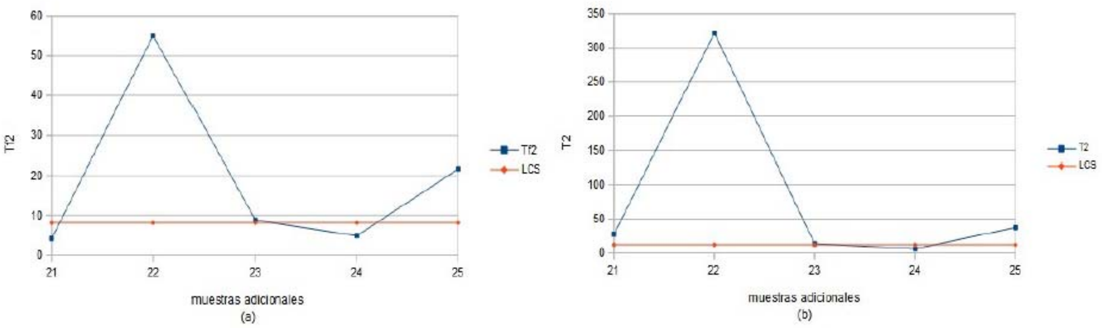

Figure 5. (a) T^2_f control chart for additional data (Taleb, Limam, and Hirota 2006), (b) T^2 fuzzy multinomial control chart for additional data

6. Conclusions

After the literature review relative to fuzzy multivariate quality control charts we decided to work under the assumption that the items are classified by each quality characteristic separately, as it is the sampling methods having higher degree of feasibility in their implementation.

Fuzzy Theory has been used to transform the multivariate variables into their corresponding representative values. These values follow an approximate multivariate normal distribution when the defuzzification method used is the fuzzy average. With this approach, the inherent complexity of multivariate multinomial is avoided. Instead a Hotelling T^2 Control Chart for a single observation can be designed.

Performance of this chart has been compared with previous alternatives (Taleb et al. 2006) showing similar behavior, with the operative important advantage of avoiding dependence on simulation.

The next step is the optimization of the parameters of the chart to preset under control ARL values. Also we believe it is important to complete this work by a sensibility analysis with respect to the selection of the membership functions and the degree of fuzzines.

References

Aparisi, F. , Avendaño, G. y Sanz, J. (2006). Techniques to interpret T2 control chart signals. IIE Transactions (Revista JCR). VOLUMEN: 38 (8): 647-657

Cheng, C. (2005). *Fuzzy process control: construction of control charts with fuzzy numbers*. Fuzzy sets and systems; 154:287-303

Civilnar, M. R. and Trussell, H. J. (1986). *Constructing membership function using statistical data*. Fuzzy Sets and Systems, 18, 1-14.

Crosier, R. B. (1988). *Multivariate generalizations of cumulative sum quality control schemes.* Technometrics, 30, 291-303.

Donald S. Holmes & A. Erhan Mergen (1993): I*mproving The Performance Of The T2 Control Chart*, Quality Engineering, 5:4, 619-625

Gülbay, M., & Kahraman, C. (2007). *An alternative approach to fuzzy control charts: Direct fuzzy approach*. Information Sciences, 177, 1463–1480.

Lowry, C. A., Woodall, W. H., Champ, C. W. and Rigdon, S. E. (1992). *A Multivariate Exponentially Weighted Moving Average Chart*. Technometrics, 34, 46-53.

Lu, X. S., Xie, M., Goh, T. N. and Lai, C. D. (1998). *Control chart for multivariate attribute processes,* International Journal of Production Research, 36, 3477-3489

Kaufmann, A. and Gupta, M. M. (1985). *Introduction of Fuzzy Arithmetic Theory and Applications*. (New York: Van Nostrand Reinhold

Kumar, Mohapatra (2012), *A New Approach to Design of Fuzzy Multi-Attribute Control Charts*, Abstract Number: 025-0620. POMS 23rd Annual Conference Chicago, Illinois, U.S.A. April 20 to April 23, 2012.

Mason, R. L., Tracy, N. D. and Young, J. C. (1997). A practical approach for interpreting multivariate T2 control chart signals. Journal of Quality Technology, 29,396-406.

Montgomery, Douglas, C. (2009). *Statistical Quality Control, a Moder Introductión*. Sixth Edition. Jhon Wiley & Sons (Asia).

Prade, H. (1982). Modal Semantics and fuzzy set theory, in: R.R. Yager (Ed.), *Fuzzy Set and Possibility Theory*: Recent Developments, Pergamon, New York.

Raz, T., & Wang, J. H. (1990). *Probabilistic and membership approaches in the construction of control charts for linguistic data.* Production Planning & Control, 1, 147–157.

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

Taleb, H., and Limam(2004), M. *Fuzzy Multivariate Control Charts for Multinomial Attribute Processes*, JT-CSIT04, , 96-104.

Taleb, H., & Limam, M. (2005). *Fuzzy multinominal control charts. In AI*IA 2005: 9th* Congress of the Italian Association for Artificial Intelligence (Vol. 3673, pp. 553–563). September 21–23, Milan, Italy.

Taleb, H., Limam, M., & Hirota, K. (2006). *Multivariate Fuzzy Multinomial Control Charts*. Quality Technology & Quantitative Management, 3, 437–453.

Taleb (2009). *Control charts applications for multivariate attribute processes.* Computers & Industrial Engineering 56 399–410.

Wang & Raz (1990): *On the construction of control charts using linguistic variables*, International Journal of Production Research, 28:3, 477-487.

Youn-Min Chou , Robert L. Mason & John C. Young (2001): *The Control Chart For Individual Observations From A Multivariate Non-Normal Distribution*, Communications in Statistics - Theory and Methods, 30:8-9, 1937-1949.

Zadeh, L. A. (1965). *Fuzzy sets*. Information and Control, 8, 338-353