Title:

A Comparative Study of Different Methodologies for Fault Diagnosis in Multivariate Quality Control.

S. Vidal-Puig and A. Ferrer

Department of Applied Statistics, Operations Research and Quality.

Technical University of Valencia.

Cno. De Vera s/n, Edificio 7A, 46022, Valencia, Spain.

Short Title:

Fault Diagnosis Performance in MQC

Abstract:

Different methodologies for fault diagnosis in multivariate quality control have been proposed in recent years. These methods work in the space of the original measured variables and have performed reasonably well when there is a reduced number of mildly correlated quality and/or process variables with a well-conditioned covariance matrix. These approaches have been introduced by emphasizing their positive or negative virtues, generally on an individual basis, so it is not clear for the practitioner the best method to be used. This paper provides a comprehensive study of the performance of diverse methodological approaches when tested on a large number of distinct simulated scenarios. Our primary aim is to highlight key weaknesses and strengths in these methods as well as clarifying their relationships and the requirements for their implementation in practice.

Keywords: Hotelling´s T^2, Fault Diagnosis, Multivariate Quality Control

Introduction

Industrial quality control usually involves a vector of measurements of either several critical to quality or critical to process parameters rather than a single characteristic. Typically, when these measurements are mutually correlated, a more efficient statistical process monitoring scheme is obtained by using multivariate control charts rather than separate univariate control charts.

Let $\mathbf{x}_i$ represent a K-dimensional vector of measurements made on a process at sampling time i. Assuming that when the process is in control, the $\mathbf{x}_i$ are independent and follow a multivariate normal distribution with a Kx1 mean vector $\boldsymbol{\mu}_{ref}$ and a KxK covariance matrix Σ, *i.e.* $\mathbf{x} \sim N_K(\boldsymbol{\mu}_{ref}, \Sigma)$.

Among the most popular multivariate control charts is the one based on Hotelling´s T^2 statistic[1,2], which is defined as the estimated Mahalanobis squared distance from the K-dimensional sample observation $\mathbf{x}_i$ to its sample mean vector $\bar{\mathbf{x}}$

$$T_i^2 = (\mathbf{x}_i - \bar{\mathbf{x}})^T \mathbf{S}^{-1} (\mathbf{x}_i - \bar{\mathbf{x}})$$

where $\bar{\mathbf{x}}$ and $\mathbf{S}$ are respectively, the usual sample mean vector and covariance matrix calculated from a reference (in-control), with historical data set having N multivariate observations. When the sample observation $\mathbf{x}_i$ is independent of the estimates $\bar{\mathbf{x}}$ and $\mathbf{S}$, the distribution of Hotelling´s T^2 is given by

$$\frac{N(N-K)}{K(N+1)(N-1)} T^2 \sim F(K, N-K)$$

A major advantage of the above statistic is that it is the optimal single-test statistic for a general multivariate shift in the mean vector[3]. However, it has several practical drawbacks: a) it is not optimal for more structured mean shifts (i.e. mean shifts in only selected variables); b) it is not specific to a shift in mean as it is also affected by changes in the covariance matrix; c) it is not immediately interpretable, (i.e. if following a signal, it does not provide information on which specific variable or set of variables is out of control).

In an attempt to improve the interpretability of T^2-based fault diagnostics several approaches have been proposed in multivariate quality control literature. The step-down method of Roy[4] assumes that there is *a priori* ordering among the means of the variables and that tests subsets sequentially using this ordering to determine the sequence. Murphy[5] suggests a method based on a discriminant distance using Hotelling´s T^2 statistic. Mason, Tracy and Young[6,7] introduce a procedure for decomposing Hotelling´s

T^2 statistic into orthogonal components. Hawkins[3,8] uses regression adjustments for each individual variable. Runger and Montgomery[9] define a distance to measure the contribution of a variable to the value of Hotelling´s T^2 statistic. Doganaksoy, Faltin and Tucker[10] propose to rank the variables most likely to have changed according to their relative contribution to Hotelling´s T^2 statistic using a univariate t statistic as a criterion. Hayter and Tsui[11], using a different procedure than Hotelling´s T^2 statistic as a trigger mechanism for out-of-control detection, propose to build exact simultaneous confidence intervals for each of the variable means. Li, Jin and Shi[12] suggest a modification of Mason, Tracy and Young´s method based on the use of bayesian networks for reducing computational cost and improving the diagnosability. The problem with this method is that it can only be applied when *a priori* relationships among process variables and the interrelationships between process variables and quality variables are known.

Mason, Tracy and Young[6] show that some of these methods: the standardized t–based ranking technique of Doganaksoy, Faltin and Tucker[10], the regression-adjusted variables of Hawkins[8], the step-down procedure of Roy[4] and the T^2 discriminant distance procedure of Murphy[5], are imbedded in the partitioning of Hotelling´s T^2. With so many approaches, a practitioner might wonder which one should be used in a particular context.

The objective of this paper is to provide a comprehensive comparison study of the performance of these approaches under a large number of different simulated scenarios where these methods might be implemented.

Fault Diagnosis Methodologies

In this paper the diagnosis performance of the different methods is compared in Phase II (model exploitation). After a previously established statistical monitoring chart detects a new signal, this new observation $\mathbf{x}_{new}$ is used to diagnose the cause of the fault. It must be noted that most of the compared methods use the Hotelling T^2 statistic for the detection of out-of-control observations whilst some like Hawkins´ method, Hayter and Tsui´s method and the Step-down method use their own detection trigger mechanism. In the following it is assumed that $\mu_{k,ref}$ is the in-control mean value for the k^{th} variable and $\mu_{k,new}$ is the mean value of the k^{th} variable after the change (fault).

Doganaksoy, Faltin and Tucker´s Method (DFT)

The diagnostic method proposed by Doganaksoy Faltin and Tucker[10] (DFT) is triggered by an out of control signal from Hotelling´s T^2 chart. The measured variables are ranked according to the univariate t statistic for the difference of two means:

$$t_k = \frac{x_{k,new} - x_{k,ref}}{\left[s_k^2 \left(1 + \frac{1}{N} \right) \right]^{1/2}} \qquad \text{For } k = 1 \text{ to } K$$

Where $x_{k,new}$ is the value of the k^{th} variable in the new observation; $x_{k,ref}$ is the estimated mean of the k^{th} variable in the in-control reference data set; s_k^2 is the estimated variance of the k^{th} variable in the in-control reference data set and N is the size of the reference data set.

This ranking is a valuable guide to diagnose the source of the change. Bonferroni's type of simultaneous confidence intervals for $\mu_{k,\text{ref}} - \mu_{k,\text{new}}$ (k = 1,..,K) are used to provide signals on individual variables. Variables for which the Bonferroni intervals do not enclose zero are highly suspect.

The implementation of this approach is as follows: An observation is considered out of control when Hotelling´s T^2 statistic for the new observation exceeds the upper control limit at the nominal confidence level $CL_{nom.}$. Then, for each variable the smallest confidence level CL_{ind} that would yield an individual confidence interval for $\mu_{k,\text{ref}} - \mu_{k,\text{new}}$ (k = 1,..,.K) that contains the zero is calculated as $CL_{ind} = \left| 2\text{T}(t_{computed}; N-1) - 1 \right|$, where $t_{computed}$ is the calculated value of the univariate t statistic for a variable and T(t;d) is the cumulative distribution function of the t distribution with d degrees of freedom. Variables with larger CL_{ind} values are the ones with relatively larger univariate t statistics which require closer investigation. For each interval the confidence level according to Bonferroni's proposal $CL_{Bonf} = 1 - \alpha_{Bonf}$ is computed, where $\alpha_{\mathrm{Bonf}} = 1 - (1 - \alpha_{\mathrm{sim}})^{1/K} \approx \frac{\alpha_{\mathrm{sim}}}{K}$, being $CL_{sim} = 1 - \alpha_{sim}$ the desired nominal confidence level and α_{sim} the desired overall Type I risk in multiple testing. Then, the variables with $CL_{ind} > CL_{Bonf}$ are classified as being those which are most likely to have changed.

Consequently, the proposed method is correspondent to work out the *p*-value of each individual two sample comparisons, and signalling those variables which *p*-value is lower than α_{Bonf} .

Modifications to the Doganoksoy, Faltin and Tucker´s Method

The Bonferroni test[13] is the simplest multiplicity adjustment procedure to ensure an overall Type I risk in multiple testing (*K*-dimensional measured variables). This method assumes independence throughout the different tests. Therefore, this proposal is too conservative when there are many tests and/or the tests are highly correlated. Being too conservative in the Type I risk derives in less sensitive tests (i.e. lack of power). In the present work we are going to consider some variations of the DFT methodology focused in reducing the risk of being too conservative when applying multiple hypothesis tests. Bonferroni's test will be replaced by different stepwise procedures proposed by Holm[14], Hochberg[15] and Hommel[16]. These approaches are based on the realization that of the *K* null hypotheses tested, the only ones to protect against rejection (at a given step) are those not yet rejected. An example would be Holm´s procedure[14:] a step down approach which conducts the testing in a decreasing order of significance of the ordered hypotheses. The order in which the hypotheses are established is according to *p*-values. In each test the $\alpha_{\mathrm{Holm}} = 1 - (1 - \alpha_{\mathrm{sim}})^{1/K'}$, where *K*'= *K* for the 1th test, *K*'= *K*-1 for the 2th test and *K*'=1 for the *K*th test. Significance testing continues until a null hypothesis is accepted. Then, all remaining (untested) null hypotheses are accepted without further testing. In a similar way Hochberg[15] and Hommel[16] derived step-up procedures. In a step-up procedure the testing is conducted in an increasing order of significance of the (ordered) hypotheses. Significance testing continues until a null hypothesis is rejected. Then all remaining (untested) null hypotheses are rejected without further testing. All these methods proved to be less conservative than the Bonferroni´s approach.

In addition to these methods, Bonferroni's test will also be replaced by two *ad hoc* procedures to take advantage of the correlation information amongst the measured variables. The first procedure, proposed by Tukey, Ciminera and Heyse[17], suggests the adjustments: $p_{ak} = 1-(1-p_k)^{\sqrt{K}}$ and $\alpha_k = 1-(1-\alpha_{\text{sim}})^{1/\sqrt{K}}$, where p_k and p_{ak} are, respectively, the observed and adjusted p-values for the k^{th} variable, and α_k is the adjusted critical α-level for the k^{th} hypothesis for $k = 1,\ldots,K$. In the second procedure, proposed by Dubey[18], Armitage and Parmar[19], and Sankoh, Huque and Dubey[13], the following adjustments were suggested: $p_{ak} = 1-(1-p_k)^{m_k}$ and $\alpha_k = 1-(1-\alpha_{\text{sim}})^{1/m_k}$, where $m_k = K^{1-r_{.k}}$ and $r_{.k} = (K-1)^{-1}\sum_{j\neq k}^{K} r_{jk}$, r_{jk} being the correlation coefficient between the j^{th} and the k^{th} variable.

These variants signal the variables where adjusted p-values, p_{ak}, are lower than α_{sim} or, where equivalent, those variables whose non-adjusted p-values are lower than α_k

Hayter and Tsui´s Method

This procedure operates by calculating a set of simultaneous confidence intervals for each one of the K variables mean (μ_k) with an overall coverage probability of 1-α, assuming a known correlation structure. This method is similar to the bar plot of normalized errors of the variables that can be seen in Kourti and MacGregor[20] or the multivariate profile charts proposed by Fuchs and Benjamini[21]. In Hayter and Tsui´s method the process is deemed to be out of control whenever any of these confidence intervals do not contain its in-control value, $\mu_{k,ref}$, and the identification of the errant variable or variables is immediate.

For a known covariance structure Σ or correlation matrix **R** and a chosen Type I risk α, the experimenter first evaluates the critical point $C_{\mathbf{R},\alpha}$ by simulation. This critical point is defined by:

$$P\left(\frac{\left|x_k - \mu_{k,ref}\right|}{\sigma_k} \leq C_{\mathbf{R},\alpha};\ for\ \ 1 \leq k \leq K\right) = 1-\alpha$$

Then, following any new observation $\mathbf{x}_{new} = \{x_{1,new},....x_{k,new},.....x_{K,new}\}$, simultaneous confidence intervals for the mean of each of the *K* measured variables (μ_k) are obtained:

$$[x_{k,new} - \sigma_k C_{\mathbf{R},\alpha}; x_{k,new} + \sigma_k C_{\mathbf{R},\alpha}]$$

These confidence intervals assume a known variance and they are calculated for a fixed $(1-\alpha)$ confidence level. The process is considered out of control if at least one interval does not contain the corresponding reference value $\mu_{k,ref}$. This is equivalent to consider that a new observation $\mathbf{x}_{new}$ is out of control when:

$$M = \underset{1\leq k\leq K}{Max} \frac{\left|x_{k,new} - \mu_{k,ref}\right|}{\sigma_k} > C_{\mathbf{R},\alpha}$$

The variables x_k whose confidence intervals do not contain $\mu_{k,ref}$ are identified as those responsible for the signal.

Murphy´s Method

Murphy´s method is an approach based on a discriminant distance. This considers a reference population when the process is in control Π_0 where the observations follow a $N_K(\boldsymbol{\mu}_{ref};\Sigma)$ distribution and a new population Π after a change in the process, where the observations follow a $N_K(\boldsymbol{\mu};\Sigma)$ distribution.

Once an out-of-control observation is detected by Hotelling´s T^2 statistic, the method searches for the subset of variables which better discriminates between these two populations. Given a partition of the *K* variables in two subsets: k_1 variables $\mathbf{x}^{(1}$ and k_2 variables $\mathbf{x}^{(2}$, where $K = k_1 + k_2$, in discriminant analysis, the true distance between the populations Π and Π_0 is defined as $\Delta_K^2 = (\boldsymbol{\mu} - \boldsymbol{\mu}_{ref})^T \boldsymbol{\Sigma}^{-1} (\boldsymbol{\mu} - \boldsymbol{\mu}_{ref})$, and the reduced distance as $\Delta_{k_1}^2 = (\boldsymbol{\mu}^{(1} - \boldsymbol{\mu}_{ref}^{(1})^T \boldsymbol{\Sigma}_1^{-1} (\boldsymbol{\mu}^{(1} - \boldsymbol{\mu}_{ref}^{(1})$. Then to test H_0: $\Delta_K^2 - \Delta_{k_1}^2 = 0$ is equivalent to testing that the k_1 subset of variables discriminates just as well as the full set of K variables. Under the assumption that the null hypothesis H_0 is true, the D statistic, $D = T_K^2 - T_{k_1}^2$, follows a $\chi_{k_2}^2$, where T_K is the overall Hotelling´s T^2 statistic (full squared distance): $T_K = (\boldsymbol{\mu}_{ref} - \mathbf{x}_{new})^T \boldsymbol{\Sigma}^{-1} \boldsymbol{\mu}_{ref} - \mathbf{x}_{new})^T$ and $T_{k_1}^2$ is the Hotelling´s T^2 statistic based on the subset of k_1 variables $\mathbf{x}^{(1}$ (reduced squared distance): $T_{k_1}^2 = (\boldsymbol{\mu}_{ref}^{(1} - \mathbf{x}_{new}^{(1})^T \boldsymbol{\Sigma}_1^{-1} (\boldsymbol{\mu}_{ref}^{(1} - \mathbf{x}_{new}^{(1})^T$ where $\boldsymbol{\mu}^{(1}$ and $\boldsymbol{\mu}_{ref}^{(1}$ refer to the mean vector of $\mathbf{x}^{(1}$ and $\boldsymbol{\Sigma}_1$ is the covariance matrix of $\mathbf{x}^{(1}$. If D is large, the hypothesis that the k_1 subset caused the signal is rejected, if it remains small then it is accepted. No *a priori* ordering is assumed in this method and all the possible subsets can be tested. The subset of variables which best discriminates between these two groups is considered the responsible for the observed out-of-control signal and corresponds to the smallest value of the D statistic. A drawback of this methodology is the excessive number of terms to compute. In this paper, the out-of-control variable selection algorithm proposed by Murphy[5] is implemented in order to reduce the intensive computational work.

Hawkins´ Method

The detection and diagnosis in Hawkins' methodology is based on the residual vector $\mathbf{z}_{new}$, whose k^{th} component is the standardized residual resulting when the k^{th} variable is regressed onto all the other variables of **x.**

$$z_{k,new} = \frac{\left(x_{k,new} - \mu_{k,ref}\right) - \sum_{k \neq j} \beta_{kj} \left(x_{j,new} - \mu_{j,ref}\right)}{\sigma_{k|1.2.3..k-1.k+1....K}}$$

Where $\sigma_{k|1.2.3..k-1.k+1....K}$ is the residual standard deviation of the conditional distribution of x_k given all other variables of **x**. Note that if $\boldsymbol{\mu}_{new}$ differs from $\boldsymbol{\mu}_{ref}$ only in its k^{th} component, then the optimal test for a shift is one based on $z_{k,new}$, the k^{th} component of vector $\mathbf{z}_{new}$. These $z_{k,new}$ residuals follow a $N(0,1)$ when the process is in control. Hawkins[8] proposes an easy way of working out the vector of scaled residuals. Let $\mathbf{y} = \boldsymbol{\Sigma}^{-1}(\mathbf{x} - \boldsymbol{\mu}_{ref})$ thus the k^{th} component of **y** is the regression residual when variable x_k is regressed on all other variables, scaled by factor $\sigma^2_{k|1.2.3..k-1.k+1....K}$. When the process is in control

$\mathbf{y} \sim N(\mathbf{0}, \boldsymbol{\Sigma}^{-1})$ and then **z** is just a rescaling of **y**:

$$\mathbf{z} = [\mathbf{diag}(\boldsymbol{\Sigma}^{-1})^{-1/2}]\mathbf{y} = \mathbf{A}(\mathbf{x} - \boldsymbol{\mu}_{\mathbf{ref}})$$

Where the transformation matrix $\mathbf{A} = \mathbf{diag}(\boldsymbol{\Sigma}^{-1})^{-1/2}\boldsymbol{\Sigma}^{-1}$ and when the process is in control $\mathbf{z} \sim \mathrm{N}(\mathbf{0};\mathbf{B})$ where $\mathbf{B} = \left(\mathbf{diag}(\boldsymbol{\Sigma}^{-1})^{-1/2}\right)\boldsymbol{\Sigma}^{-1}\left(\mathbf{diag}(\boldsymbol{\Sigma}^{-1})^{-1/2}\right)$ is the covariance matrix for the vector of scaled residuals $\mathbf{z}_{new}$.

The original proposal consists of monitoring the process using separate control charts for all of the $z_{k,new}$. If the control chart for one of the $z_{k,new}$ signals while charts of others do not, then that indicates that it is $z_{k,new}$ has shifted[8]. Note that the original proposal does not make any correction either for multiple testing or correlation among the scaled residuals. So it is necessary to adjust the Type I risk with an appropriate selection of the number of standard deviations (d) when calculating the upper control limits of the monitoring charts.

Montgomery and Runger´s method.

This methodology tries to establish the contribution of a variable to the value of Hotelling´s T^2 statistic used for monitoring the process when a control chart signals. The contribution c_k for variable $x_{k,new}$ is the required change in the single variable x_k which gives a minimum value of the chi-squared (or F-Snedecor) statistic. So the method looks for the c_k that minimizes the expression:

$$\left[\frac{\mathbf{x}_{new}-c_k\mathbf{e}_k}{\left(\mathbf{e}_k^T\boldsymbol{\Sigma}^{-1}\mathbf{e}_k\right)^{1/2}}\right]^{\mathrm{T}}\boldsymbol{\Sigma}^{-1}\left[\frac{\mathbf{x}_{new}-c_k\mathbf{e}_k}{\left(\mathbf{e}_k^T\boldsymbol{\Sigma}^{-1}\mathbf{e}_k\right)^{1/2}}\right]$$

where $\mathbf{e}_k$ is the unit vector in the direction of the k^{th} coordinate axis and $\left(\mathbf{e}_k^T\boldsymbol{\Sigma}^{-1}\mathbf{e}_k\right)^{1/2}$ is a scale factor so that c_k can be interpreted as a measure of a Euclidean distance.

Variables that require large changes (c_k) aim to the responsible variables. Authors proved that the squared contribution c_k^2 is equivalent to the Murphy´s D difference for the subset of the k_1 variables, $\mathbf{x}^{(1} = (x_1\, x_2 \ldots x_{k-1}\, x_{k+1} \ldots x_{\mathrm{k}})$

$$c_k^2 = D = T_K^2 - T_{k_1}^2$$

where $T_{k_1}^2$ is Hotelling´s T^2 statistic based on a subset $\mathbf{x}^{(1}$ made up of K-1 variables after excluding the k^{th} variable. A large contribution c_k^2 corresponds to a large D statistic, therefore, we reject that the k_1 subset of variables causes the signal, highlighting the variable x_{k} as responsible for the shift.

Mason, Tracy and Young´s Method (MTY)

This method decomposes the overall Hotelling´s T^2 statistic into independent components, each reflecting the contribution of the different variables to the statistic. The Hotelling´s T^2 statistic for a new observation may be iteratively decomposed according to Rencher[22] in two classes of components: a) the *unconditional components* $T_k^2 = \left(\frac{x_{k,new}-x_{k,ref}}{s_k}\right)^2$ that measure the "marginal" contribution of the variable x_k to the statistic T^2 and, therefore, records changes in variable magnitudes but does not account for correlation structure; b) the *conditional components* that, assuming a particular ordering in the variables, can be expressed as $T_{k|1,2\ldots k-1}^2 = \left(\frac{x_{k,new}-x_{k|1,2..k-1}}{s_{k|1,2\ldots k-1}}\right)^2$ for k=2,...,K. These components measure the contribution of the variable $x_{k,new}$ to the value of the T^2 statistic after being adjusted by a regression onto a subset of the other variables and, therefore, records events that break the correlation structure.

The unconditional components are distributed as $T_k^2 \sim \frac{N+1}{N} F_{1,N-1}$, while the conditional components are distributed as $T_{k|1,2,\ldots M}^2 \sim \frac{(N+1)(N-1)}{N(N-M-1)} F_{1,N-M-1}$, where N is the number of observations of the reference data set and M the number of variables conditioning the component distribution.

Each decomposition leads to one unconditional component and K-1 conditional components, as given by the expression:

$$T^2 = T_1^2 + T_{2|1}^2 + T_{3|1,2}^2 + \ldots\ldots + T_{K|1,2,\ldots K-1}^2 = T_1^2 + \sum_{k=1}^{K-1} T_{k+1|1,2\ldots.k}^2$$

As this method does not assume any special order in the variables, there are *K!* different decompositions of the T^2, each one with K independent components. Then, all the components are compared against its corresponding component distribution threshold, and the variables with significant components are identified as responsible for the detected fault. In this paper, the computational scheme proposed by Mason, Tracy and Young[7] (MTY) is implemented.

Step-Down Method

The step-down methodology[23] assumes a certain *a priori* ordering among subsets that can be formed with the K measured variables. According to the ordering, the step-down procedure uses partitioning of the mean vector of the new observation μ_{new} and the mean vector of the reference data μ_{ref} into Q subvectors $\mu_{new,1}$, $\mu_{new,2}$, …$\mu_{new,q}$ …. $\mu_{new,Q}$ and $\mu_{ref,1}$, $\mu_{ref,2}$, …$\mu_{ref,q}$….. $\mu_{ref,Q}$, respectively. Then, it sequentially tests $H_0^{(1}: \mu_{new,1} = \mu_{ref,1}$ versus $H_1^{(1}: \mu_{new,1} \neq \mu_{ref,1}$; then $H_0^{(2}: \mu_{new,2} = \mu_{ref,2}$ versus $H_1^{(2}: \mu_{new,2} \neq \mu_{ref,2}$ given $\mu_{new,1} = \mu_{ref,1}$; then $H_0^{(3}: \mu_{new,3} = \mu_{ref,3}$ versus $H_1^{(3}: \mu_{new,3} \neq \mu_{ref,3}$ given $\mu_{new,1} = \mu_{ref,1}$ and

$\boldsymbol{\mu}_{new,2}$ = $\boldsymbol{\mu}_{ref,2}$; and so on. The test statistics associated with testing these sub hypotheses, G_q^2, are independently distributed under H_0,

where $G_q^2 = \dfrac{T_q^2 - T_{q-1}^2}{1+(T_{q-1}^2/(N-1))}$, q = 1.....Q, T^2_q is the MTY unconditional T^2 term for the first L_q variables with $L_q = \sum_{i=1}^{q} k_i$, k_i is the number of elements of subset i, and $T_0^2 = 0$.

Under H_0 assumption it follows that $G_q^2 \sim \dfrac{(N-1)k_q}{N-L_q} F(k_q, N-L_q)$ and it is possible to use separated control charts for monitoring them with a critical value (i.e. upper control limit, UCL) for the sub-hypothesis q given by:

$$UCL_q = \frac{(N-1)k_q}{N-L_q} F_{\alpha_q}(k_q, N-L_q) \quad \text{for } q=1,2,...,Q$$

where $F_{\alpha_q}(k_q, N-L_q)$ is the $(1-\alpha_q)\times 100$ percentile of the $F(k_q, N-L_q)$ distribution.

The process is considered out of control if at least one G_q^2 exceeds the corresponding threshold UCL_q.

Key drawbacks of this methodology are: a) it assumes the existence of an *a priori* order among the different types of faults; and b) it is impossible to implement this methodology when there are faults that share common measured variables.

Simulation procedure

In order to compare the methods, several faults consisting of small, medium or large shifts in the mean of one (or more) variables under different scenarios of correlation matrices will be simulated.

In the simulation, the different methodologies are applied to a case of four measured variables under eleven different correlation structures shown in Table 1 where the covariance matrix condition numbers tend to increase its value from C1 to C11. The standard deviations of the four variables were uniformly distributed between 0.3 and 0.4. Scenarios leading to unfeasible covariance matrices were discarded.

Reference data sets of 50.000 observations for each of the 11 covariance structures were obtained using the algorithm proposed by Arteaga and Ferrer[24]. These reference data sets were used to adjust the Type I risk when the methodologies under comparison used a different detection trigger mechanism in the detection of the out-of-control observations other than Hotelling´s T^2 statistic (i.e. Hawkins´ method, Hayter and Tsui´s method and Step-down method). For every correlation structure 102 different types of faults were considered. The faults consisted in mean shifts in one, two or three variables. The size of the shifts were small (0.5 units or 1.25 to 1.66 standard deviations), medium (1 units or 2.5 to 3.33 standard deviations) or large (2 units or 5 to 6.6 standard deviations). The shifts involving several means happened in both the same or opposite directions. For each type of fault, 500 observations using the algorithm proposed by Arteaga and Ferrer[24] were simulated. In this study we have only considered faults affecting the mean of the process and excluded faults affecting the covariance structure. The rationale for this decision is: i) this approach is most commonly used to address the performance of different diagnostic methods; ii) this allows the appropriate comparison of the methodologies especially as some of them are not suited for the detection of changes in the covariance matrix of the process.

TABLE 1 [HERE]

These faulty data sets were processed under the different proposed fault diagnosis methodologies and their performance were measured and compared according to several performance indices that were computed for every correlation structure and a particular type of fault. The considered performance indices were the following:

- PTC_0 : Proportion of observations correctly diagnosed

$$PTC_0 = \sum_{i=1}^{500} P_{0,i} \Big/ 500$$

  where $P_0 = 1$ if *all the variables* in the observation are correctly diagnosed, and $P_0 = 0$ on the contrary.

- $PTCv$: Proportion of faulty variables correctly diagnosed (i.e., true positives in variables)

$$PTC_v = \sum_{i=1}^{500} P_{\mathrm{v},i} \Big/ 500$$

where $P_{v,i} = N_{\mathrm{df},i} / N_{\mathrm{f},i}$ with $N_{\mathrm{df},i}$ equals to the number of correctly diagnosed faulty variables in the i^{th} observation, and $N_{\mathrm{f},i}$ equals to the number of faulty variables in the i^{th} observation.

- PWC_0 Proportion of observations with any non faulty variable wrongly diagnosed (i.e. false positives in observations)

$$PWC_0 = \sum_{i=1}^{500} W_{0,i} \Big/ 500$$

where W_0 =1 if there is any non-faulty variable in the observation wrongly classified, and W_0 = 0 on the contrary.

- $PWCv$: Proportion of non faulty variables wrongly diagnosed (i.e. false positives in variables)

$$PWC_v = \sum_{i=1}^{500} W_{v,i} \Big/ 500$$

where $W_{v,i} = NW_{df,i} / N_{nf,i}$ with $\mathrm{NW}_{\mathrm{df},i}$ equals to the number of wrongly diagnosed non-faulty variables in the i^{th} observation, and $N_{\mathrm{nf},i}$ equals to the number of non-faulty variables in the i^{th} observation.

- PND: Proportion of faulty observations which are not detected as faults. This is related to the lack of detection power

$$PND = \sum_{i=1}^{500} P_{\mathrm{d},i} \Big/ 500$$

where $P_{\mathrm{d},i}$ =1 if the i^{th} observation is not detected as a faulty observation, and $P_{\mathrm{d},i}$ = 0 on the contrary.

- PNF : Proportion of detected faulty observations in which no variable is found as responsible. This is related with the lack of isolation power

$$PNF = \sum_{i=1}^{500} P_{f,i} \Big/ 500$$

where $P_{f,i}$ =1 if the i^{th} observation is detected as a faulty observation but no variable is found as responsible, and $P_{f,i}$ = 0 on the contrary.

Type I risk considerations

In order to check the accuracy and precision of the adjusted Type I risk for the 11 covariance matrices under different detection trigger mechanisms, 10 reference data sets under each correlation matrix were simulated and the real Type I risk for each data set were computed.

In the methodologies based on Hotelling´s T^2 the real Type I risk is centered in the desired value as it expected since the Type I risk level is adjusted from a theoretical distribution that takes into account the correlation between variables.

Hawkins´ methodology assumes that the marginal distribution of the monitored residuals follows a standardized normal distribution. The overall Type I risk depends on the number of hypotheses tests and the Type I risk α of each of the hypotheses tests. In the case of four independent variables, the overall Type I risk is $1-(1-\alpha)^4$. For a desired overall rate of $\alpha_{overall} = 0.05$, $\alpha = 1-(1-\alpha_{overall})^{1/4} = 0.01274$ so the number of standard deviations to consider for a two-tail hypothesis test is 2.49σ. Figure 1 a) shows the Type I risk of Hawkins´ methodology after Bonferroni correction for the 11 correlation structures simulated. The underestimation of $\alpha_{overall}$ in most scenarios is due to the lack of independence between the monitored residuals. The **B** matrix of the Hawkins´ methodology shows that the monitored standardized normal residuals are correlated and, consequently, it is necessary to adjust for the Type I risk in every case.

FIGURE 1 [HERE]

Table 2 shows the selection of the number of standard deviation (*d*) to use in the construction of the upper control limit (UCL) in Hawkins´ methodology in order to get an overall Type I risk, $\alpha_{overall}$=0.05 in the 11 correlation matrices simulated.

TABLE 2 [HERE]

. Figure 1 b) shows that after the adjustment the objective of overall Type I risk of 5% is accomplished.

In the case of Hayter and Tsui´s and the Step-down´s methodologies the monitored statistic follows known theoretical distributions what makes easier to adjust them for the overall Type I risk ($\alpha_{overall}$=0.05).

Statistical comparison of methodologies

The results for the different performance indices obtained from the simulation study were analyzed with a multifactor analysis of variance (ANOVA) considering the factors: number of faulty variables, N_f (3 levels: 1, 2 and 3 faulty variables); diagnostic method, *M* (14 levels, see Table 3); and correlation structure, *C* (11 levels, see Table 1).

The ANOVA results show that all the factors and most of their interactions are statistically significant (*p*-value< 0.05) for all the performance indices.

In the Step-down method two *a priori* ordering among the different types of faults were considered: profile 1-1-1-1 (fault in x_1, fault in x_2, fault in x_3, fault in x_4) and profile 1-1-2 (fault in x_1, fault in x_2,

TABLA 3 [HERE]

fault in x_3 and x_4). A variant of Hawkins´ methodology to detect faults affecting one single variable (Hawkins' one single variable method) was also considered in this paper. In this variant, the algorithm identifies as responsible the variable with the largest significant residual $z_{k,new}$.

The mean and 95% least significance difference (LSD) intervals plots displayed in Figure 2 show that the MTY (M10), the *ad hoc* and Bonferroni variants of the DFT method (M7, M8 and M3) give the best results in PTC_0 (correct diagnosis). The MTY (M10) also displays the best results in PTC_v (true positives) and intermediate results in PWC_0 and PWC_v (false positives). The Hawkins' (M1), Murphy's (M9) and Montgomery and Runger's (M11) methods have serious problems of false positives in diagnosis as it can be concluded from their large values in PWC_0 and PWC_v, yielding a low performance in terms of correct diagnosis (PTC_0).

FIGURE 2 [HERE]

The interaction plots displayed in Figure 3 shows that one of the main reasons for the statistically significant interaction between correlation structure and the diagnosis method is the performance in PTC_0, PWC_0 and PWC_v of the methods M1, M9, M11 is much more sensitive to changes in the correlation structure than the others**.**

FIGURE 3 [HERE]

The interaction plots between the number of faults and the fault diagnosis method displayed in Figure 4 show that although M12 and M14 are the best methods in PTC_0 for one single variable faults they perform badly faulty variables 2 and 3. This explains their bad performance in PTC_0 as shown in Figure 5 a). Methods M1, M9, M11 and M13 perform badly in PTC_0 no matter the number of faulty variables. Regarding PTC_v Figure 4 b) shows that the M1 has the best performance for one single variable faults while the M10 gives the best diagnosis performance for 2 and 3 faulty variables. Methods M12, M13 and M14 perform badly for 2 and 3 faulty variables. This explains their bad performance in PTC_v shown in Figure 2 b).

FIGURE 4 [HERE]

In the case of one single variable faults, Figure 5 a) shows that M14 (Hawkins´ one single fault method) and M12 (Step-down method with 1-1-1-1 subsets) perform better in PTC_0 than the rest of the methods. Figures 5 b) and c) shows that M12 and M14 present small values for PWC_0 and PWC_v. The good result of these methods in single variable faults can be explained as they are specially designed for this situation. On the contrary these methods give bad results when the actual fault involves more than one variable as already shown in Figure 4. Another drawback in the Step-down method is the difficulty in implementing the monitoring plots when two different types of faults share a common out-of-control variable (i.e. if one type of fault supposes that the variables 1 and 2 become out of control and a second type of fault supposes that variable 1 and 3 becomes out of control).

FIGURE 5 [HERE]

If the size of fault (3 levels: small, medium and large) is introduced as a new factor in the ANOVA we observe an interesting result in Figure 6 whereby ANOVA interaction plots between the diagnosis method and the size of fault show that large and medium faults are particularly responsible for the excessive false positive rates in methods M1, M9 and M11.

FIGURE 6 [HERE]

As it can be seen in Figure 7 a) and Figure 2 e) *PND* is equal on methods M3 to M11 since the detection on these methods is based on the same Hotelling´s T^2 statistic. Methods M1, M12 and M13 have slightly larger *PND* in all the correlations structures. It can be appreciated that *PND* higher values are obtained in the weakest correlation structures C1, C2 and C3 for all the methodologies with the exception of the M2. The M2 presents a singular behavior since the *PND* results become close similar in all the correlation structures. This method has the worst results in *PND*.

FIGURE 7 [HERE]

FIGURE 8 [HERE]

Figure 7 b) and Figure 2 e) show that the DFT methods (from M3 to M8) give high values on *PNF*, thus indicating a lack of diagnostic power. Figure 8 shows that the bad *PNF* results are mainly associated to the small size faults, being particularly problematic in M3 to M6 methods.

Summary and Conclusions

The simulation showed that the MTY method has a better diagnosis performance than the rest of the methods because it combines better results in PTC_v with similar results in PTC_0 of other methods. Additionally, the MTY provides an easy interpretability of the terms and relationships between variables classifying the out-of-control cases in situations that may or not break the correlation structure between the variables. In the simulation it could be seen that Hawkins´, Murphy´s and Montgomery´s methods increase the number of false positives in the case of strong correlations and, consequently, yielded a bad performance in PTC_0.

In the simulation, the DFT method and its variants manifested problems in "lack of power in fault isolation" (*PNF*). The *ad hoc* methods D/AP and TCH showed a better power in fault isolation and PTC_0 values in the case of faults involving three variables or small faults, than the Bonferroni´s variant. The Holm´s, Hochberg´s and Hommel´s variants had the worst results in all the scenarios simulated.

In the simulation, the step-down method with profile 1-1-1-1 and the Hawkins´method for faults in one single variable yielded the best results in the case of one single variable faults. The problem with these methods is that they cannot be used to diagnose fault situations where more than one variable is responsible

The imperative result of this study is that it has clearly shown that most of the compared methodologies have problems with false positives that have often not been reported in literature. Future research is needed to introduce variants in these methods or improve the algorithms to reduce the impact of the *PWC* indices in the diagnosis performance of these methodologies and, consequently, improve their classification results.

References

1. Jackson JE. (1991). *A User´s Guide to Principal Components*. Wiley & Sons.

2. Mason RL, Tracy ND, Young JC. (1995) Multivariate control charts for individual observations. *Journal of Quality Technology* 24, 2: 88-95.

3. Hawkins DM. (1991) Multivariate Quality Control Based on Regression- Adjusted Variables. *Technometrics* 33: 61-75.

4. Roy J. (1958). Step-down procedures in multivariate analysis. The *Annals of Mathematical Statistics*; 29: 1177-1187.

5. Murphy BJ. Selecting Out of Control Variables with the T^2 Multivariate Quality Procedure. (1987) *The Statistician* 36: 571-583.

6. Mason RL, Tracy ND, Young JC. (1995) Decomposition of T^2 for Multivariate Control Chart Interpretation. *Journal of Quality Technology* 27, 2: 99-108.

7. Mason RL, Tracy ND, Young, J.C. (1997) A Practical Approach For Interpreting Multivariate T^2 Control Chart Signs. *Journal of Quality Technology* 29,4: 396-406.

8. Hawkins DM. (1993). Regression Adjustment for Variables in Multivariate Quality Control. *Journal of Quality Technology* 25, 3: 170-182.

9. Runger GC, Montgomery DC. (1996). Contributors to a Multivariate Statistical Process Control Chart Signal. *Communications in Statistics – Theory and Methods* 25,10: 2203-2213.

10. Doganaksoy N, Faltin FW, Tucker WT. (1991). Identification of Out of Control Quality Characteristics in a Multivariate Manufacturing Environment. *Communications in Statistics-Theory and Methods* 20, 9: 2775-2790.

11. Hayter AJ, Tsui KL. (1994). Identification and Quantification in Multivariate Quality Control Problems. *Journal of Quality Technology* 26, 3: 197-207.

12. Li J, Jin J, Shi J. (2008). Causation-based T2 Decomposition for Multivariate Process Monitoring and Diagnosis, *Journal of Quality Technology* 40, 1: 46-58.

13. Sankoh AJ, Huque MF, Dubey SD. (1997). Some Comments on Frequently Used Multiple Endpoint Adjustment Methods in Clinical Trials. *Statistics in Medicine* 16: 2529-2542.

14. Holm S. (1979). A Simple Sequentially Rejective Multiple Test Procedure. *Scandinavian Journal of Statistics* 6: 65-70.

15. Hochberg Y. (1988). A Sharper Bonferroni Procedure for Multiple Test of Significance. *Biometrika* 75: 800-802.

16. Hommel. (1988). A Comparison of Two Modified Bonferroni Procedures. *Biometrika* 75: 383-386.

17. Tukey JW, Ciminera,JL, Heyse JF. (1985). Testing the Statistical Certainty of a Response to Increasing Doses of a Drug. *Biometrics* 45: 295-301.

18. Dubey SD. (1985). Adjustment of *p*-values for Multiplicities of Intercorrelating Symptoms. *Proceedings of the VIth International Society for Clinical Biostatisticians*. Germany.

19. Armitage P, Parmar M. (1986). Some Approaches to the Problem of Multiplicity in Clinical Trials. *Proceedings of the XII th International Biometrics Conference.* Seattle.

20. Kourti T, MacGregor J. (1996). Multivariate SPC Methods for Process and Product Monitoring. *Journal of Quality Technology* 28, 4: 409-428.

21. Fuchs C, Benjamini Y. (1994). Multivariate Profile Charts for Statistical Process Control. *Technometrics* 36: 182-195.

22. Rencher AC. (1993). The contribution of individual variables to Hotelling´s T^2, Wilks´ $\Lambda$and R^2. *Biometrics* 49 : 479-489.

23. Wierda, S.J. (1993). *Papers 557*, Groningen State, Institute of Economic Research.

24. Arteaga F, Ferrer A. (2010). How to simulate normal data sets with the desired correlation structure. *Chemometrics and Intelligent Laboratory Systems* 101, 1: 38-42.

TABLES

TABLE 1. Correlation structures.

Correlation Structure	Correlation Values	Extreme Correlations (0.9)	Condition number CN= $\frac{\lambda_{max}}{\lambda_{min}}$
C1: Weak correlations	Weak correlation coefficients uniform distributed, U[-0.1 , +0.1]	No	1.57
C2: Moderate positive correlations	Moderate positive correlation coefficients uniformly distributed, U[+0.1 , +0.4]	No	3.24
C3: Moderate mixed correlations	Moderate mixed positive-negative correlations. Absolute correlation coefficients uniformly distributed, U[+0.1 , +0.4]	No	4.91
C4: Moderate negative correlations	Moderate negative correlation coefficients uniformly distributed, U[-0.1 , -0.4]	No	22.32
C5: Weak correlations with one extreme correlation	Weak correlation coefficients uniformly distributed, U[-0.1 , +0.1] with one coefficient +0.9	Yes	20.38
C6: Moderate positive correlations with one extreme correlation	Moderate positive correlation coefficients uniformly distributed, U[+0.1 , +0.4] with one coefficient +0.9	Yes	21.49
C7: Moderate mixed Correlations with one extreme correlation	Moderate mixed positive-negative correlations. Absolute correlation coefficients uniformly distributed, U[+0.1 , +0.4] with one coefficient +0.9	Yes	29.92
C8: Strong positive correlations	Strong positive correlation coefficients uniformly distributed, U[+0.5 , +0.8]	No	17.37
C9: Strong positive correlations with one extreme correlation	Strong positive correlation coefficients uniformly distributed, U[+0.5 , +0.8] with one coefficient +0.9	Yes	38.07
C10: Strong mixed correlations	Moderate strong positive-negative correlations. Absolute correlation coefficients uniformly distributed, U[+0.5 , +0.8]	No	17.91
C11: Strong mixed correlations with one extreme correlation	Moderate strong positive-negative correlations. Absolute correlation coefficients uniformly distributed, U[+0.5 , +0.8] with one coefficient +0.9	Yes	39.35

TABLE 2. Selected number of standard deviations (d) to use in the construction of the UCL in Hawkins' methodology for an overall Type I risk, $\alpha_{overall}$=0.05, in the 11 correlation matrix scenarios

C	1	2	3	4	5	6	7	8	9	10	11
d	2.49	2.44	2.48	2.44	2.31	2.47	2.41	2.46	2.44	2.46	2.45

TABLE 3. List of diagnostic methods

Label	Method
M1	Hawkins
M2	Hayter and Tsui
M3	Doganaksoy, Faltin and Tucker (Bonferroni)
M4	Doganaksoy, Faltin and Tucker (Holm)
M5	Doganaksoy, Faltin and Tucker (Hochberg)
M6	Doganaksoy, Faltin and Tucker (Hommel)
M7	Doganaksoy, Faltin and Tucker (TCH)
M8	Doganaksoy, Faltin and Tucker (D/AP)
M9	Murphy
M10	Mason, Tracy and Young (MTY)
M11	Montgomery and Runger
M12	Step-down with profile (1-1-1-1)
M13	Step-down with profile (1-1-2)
M14	Hawkins' one single variable

FIGURES

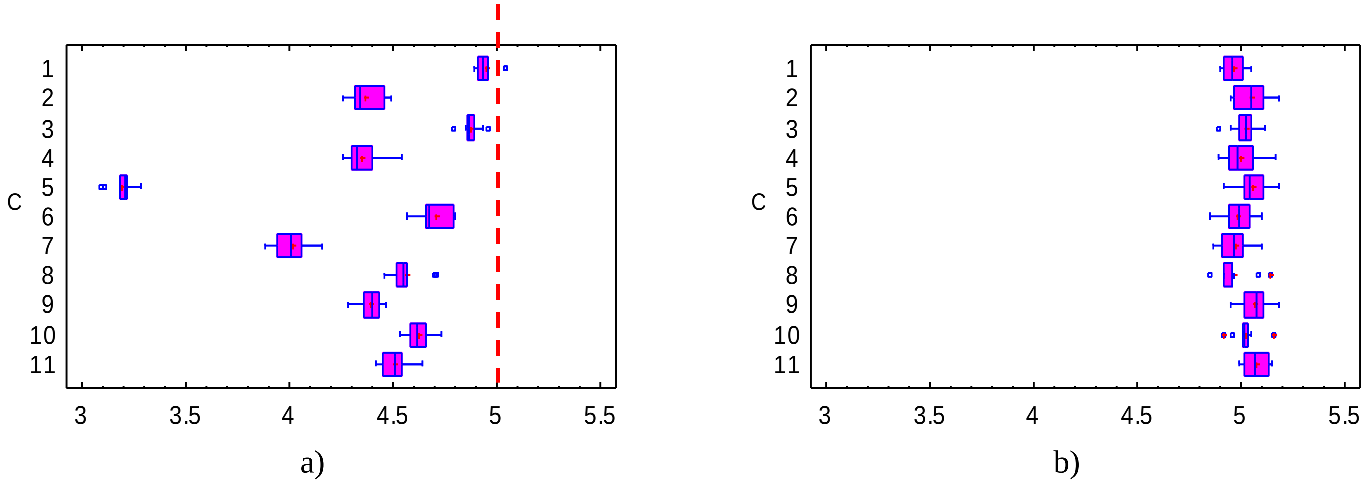

FIGURE 1: Type I risk ($\alpha_{overall} \times 100$) for the 11 correlation structures:
a) Original Hawkins´ method; b) Adjusted Hawkins´ method

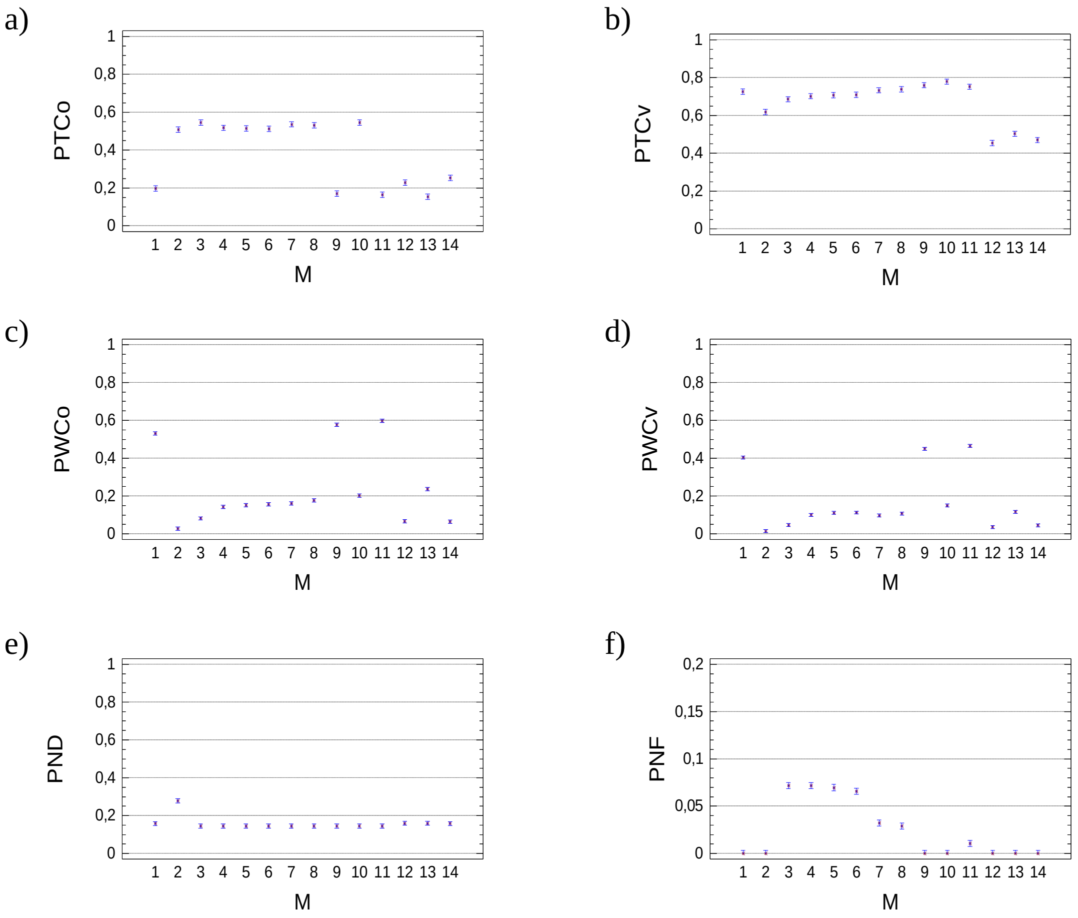

FIGURE 2: Means and 95% LSD intervals plot: a) PTC_0 , b) PTC_v , c) PWC_0 , d) PWC_v , e) PND , f) PNF.

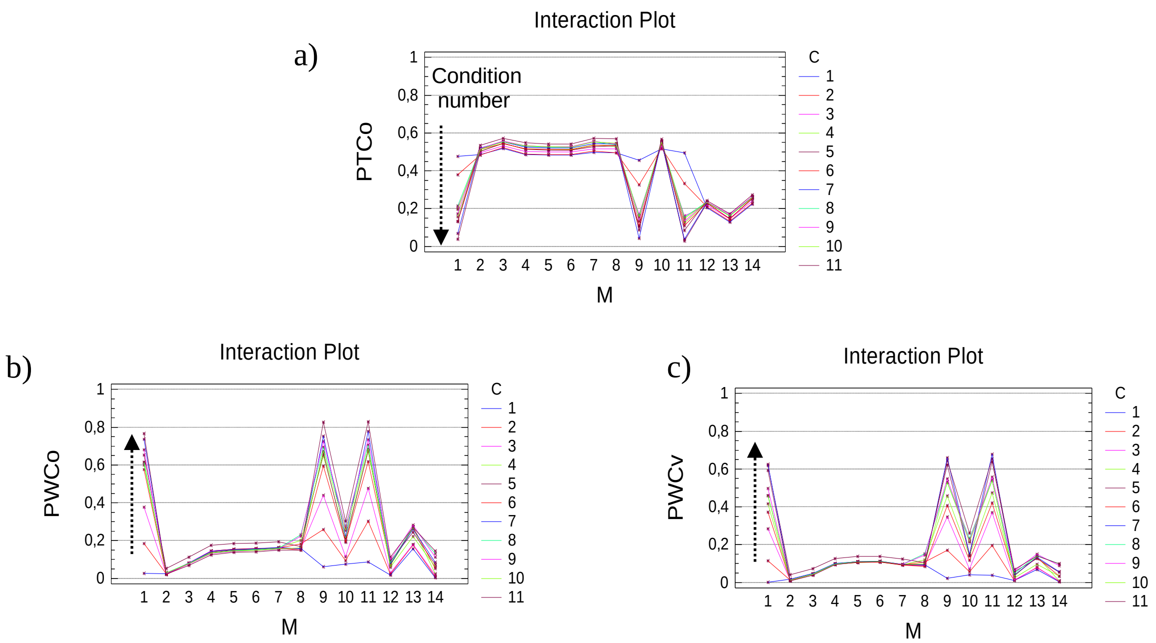

FIGURE 3: Interaction plots for diagnosis method × covariance structure: a) PTC_0 , b) PWC_0 c) PWC_v Arrows in the plots indicate the direction of increment of condition number of the correlation structures.

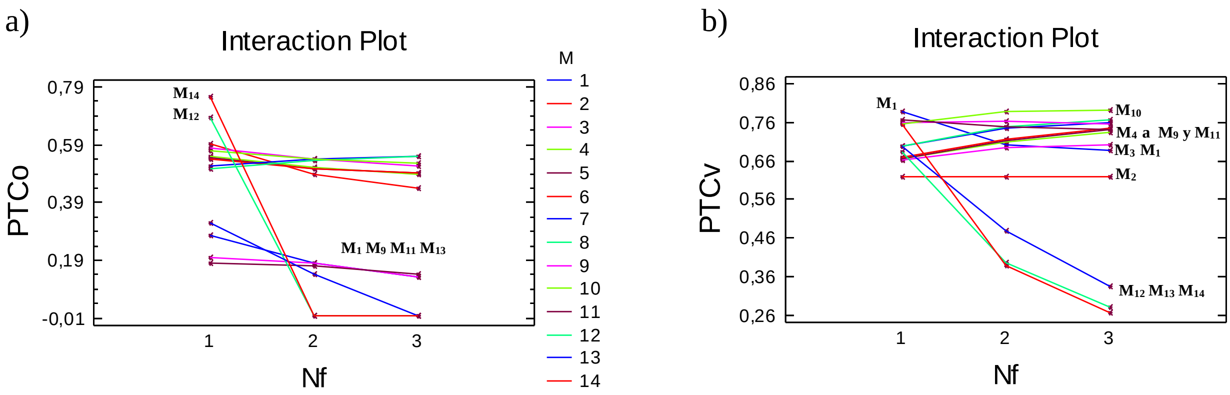

FIGURE 4: Interaction plots for diagnosis method × number of faults: a) PTC_0 , b) PTC_v.

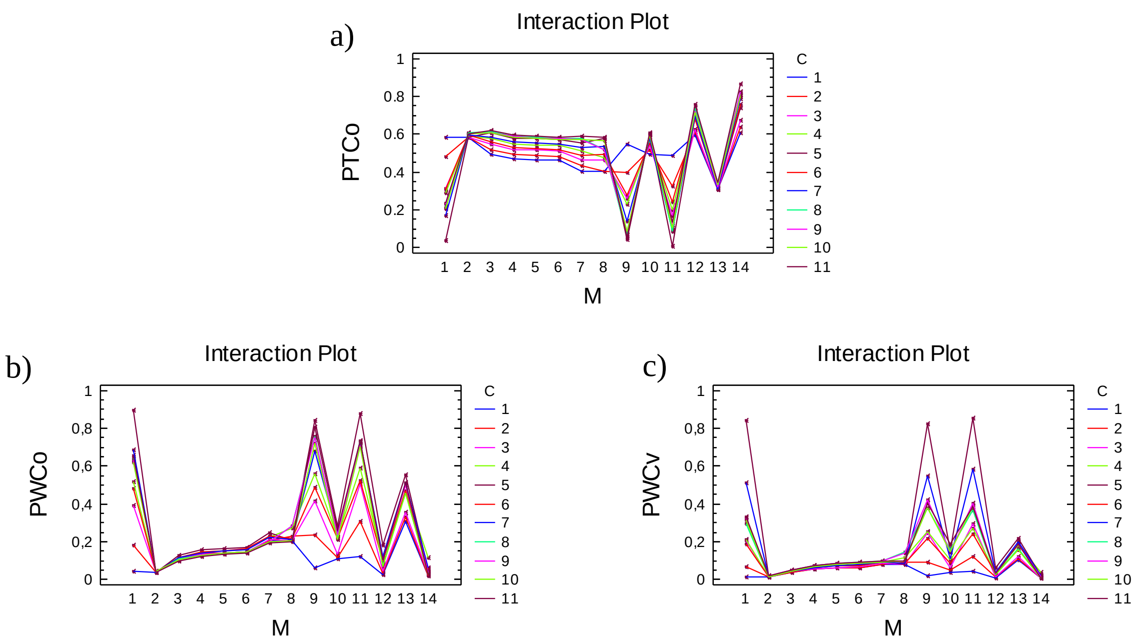

FIGURE 5: One single variable fault interaction plots: diagnosis method × covariance structure:
a) PTC_0 , b) PWC_0, c) PWC_v.

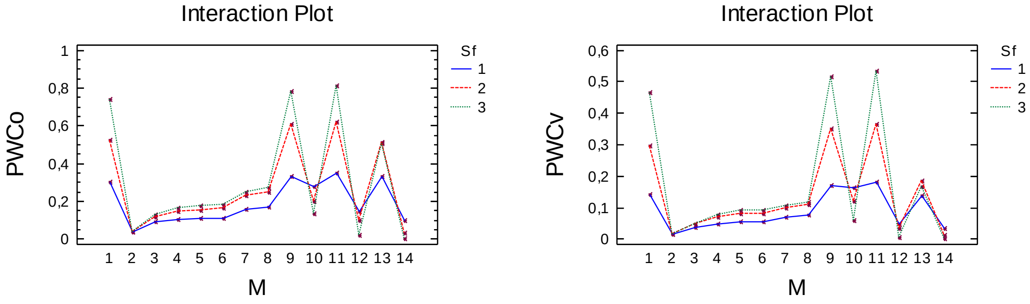

FIGURE 6: One single variable fault interaction plots: diagnosis method × fault size *a)*PWC_0*,b)* PWC_v

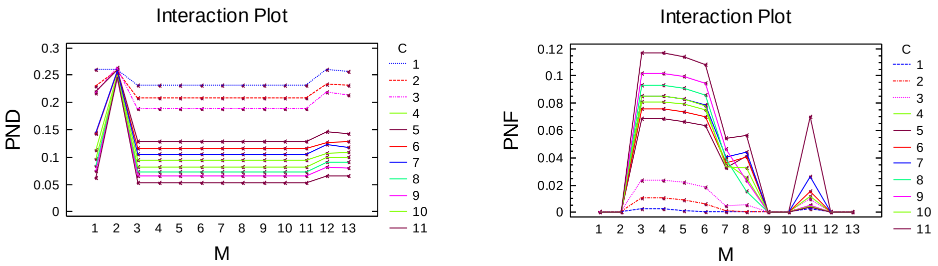

FIGURE 7: Interaction plots: diagnosis method × correlation structure: a) *PND* b) *PNF*

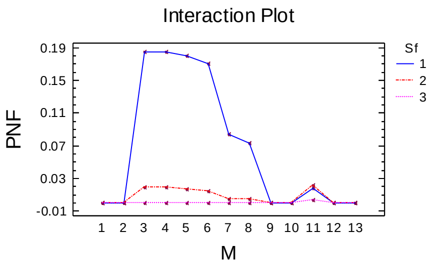

FIGURE 8: *PNF* Interaction plot: diagnosis method × fault size