International Journal of Pure and Applied Mathematics
Volume 64 No. 3 2010, 367-375

ON-LINE MULTIVARIATE QUALITY CONTROL

Gabriele Stoppa
Department of Computer and Management Sciences
University of Trento
Cia Inama, 5, Trento, 38100, ITALY
e-mail: gabriele.stoppa@economia.unitn.it

Abstract: The on-line quality control must be made on all individual items, namely it is reasonable to consider all produced units. In fact, in case of automated inspection, every manufactured item is accurately monitored and calibrated so that a great part of traditional quality control techniques does not work or is inappropriate. This paper concerns a method for monitoring the multidimensional production process for variables based on the measurement of all items manufactured in real time. We use all the disposable data for determining whether in the *first phase* (retrospective data) there is an *in-control-state* with the advantage of maximizing the amount of information collected. Mainly the following questions arise: 1) the *technical rank* between the variables, that puts the measures in a-symmetric position, is often disregard; 2) individual control charts for the variability and the co-relations do not exist. The **xoc**-method here suggested accounts for the a-symmetry and can handle the dispersion and the co-relations in the same way as the calibrations. We define the item calibrations, the tool calibrations and the calibration sequence and evaluate them measuring the *loss-in-quality* derived from the T^2-distance. Several topics will be discussed, including: calibrations, oscillations, co-relations and detection of the responsible variable(s). Every chart is associated with the chart of the variable contributions, which is a natural extension of the multivariate chart here suggested to visualize involved tools. The use of the **xoc**-method is illustrated with an example of industrial data and also some of the issues related to practical interpretations is discussed.

AMS Subject Classification: 46N30, 47N30
Key Words: individual control chart, variable contribution, N_p-distribution, χ^2-distribution, T^2-distance, hypothesis testing

Received: March 14, 2010

1. Introduction

In presence of measurement technology and automated inspection, namely when all items are simultaneously monitored in real time, it is useful to have a quality control for providing the calibration data by using individual charts. If so, only the on-line individual charts for items (units, parts, subassemblies, finished articles, etc.) make sense and the traditional quality control techniques are inappropriate. There are many situations in which all items are monitored; some examples of these situations are: 1) there are no reasons for sampling the automated data; 2) the production rate is very slow and it is inopportune to group the items for accumulating information before the analysis and a long interval between items may cause problems with the rational grouping; 3) the measurements come from the same produced unit, such as those taken in different locations along a wafer in a semiconductor manufacture. However, individual charts turn out to be very economical and, when the normality assumption is satisfied, it is unnecessary to group the information and the individual charts are more immediate and easily comprehensible. The sequence of variables reflects a logical/time order typical of the specific manufacturing process. This means that the declared sequence reflects the importance of the variables and each variable depends on the relations to the previous ones. The *technical rank* between the variables causes an a-symmetric situation, but unfortunately this rank is an often disregarded extra-information. The suggested **xoc**-method incorporates this extra-information and concerns: calibrations ($\boldsymbol{x}$), oscillations ($\boldsymbol{o}$) and co-relations ($\boldsymbol{c}$). But at present in the literature on individual multivariate cases there are no dispersion and co-relation charts at all. For item dispersions we suggest to analyse the mobile oscillations and we named co-relations the mobile oscillations of the cumulated sums of variables. The paper is organized as follows: Section 2 presents the points of view about the process and the charts; Section 3 outlines the method and Section 4 concerns an application to a switch-drum case.

2. The On-Line Quality Control

The multivariate quality control is usually concentrated on detecting the *out-of-control* items due to anomalies coming from special causes, on the recognition of control lack and on taking action as soon as possible. When the target of calibrations is $\boldsymbol{\mu} = \mathbf{0}$ (the control condition) the calibrations become real deviations from the target; a sufficiently large n (the number of items available) assures that the covariance matrix $\mathbf{\Sigma}$ can be considered as known or conveniently estimated; and if each of the p variables involved is associated with p tools there is a perfect correspondence between the out-of control signal and the actions to do. But by monitoring a production process we must also carry out a permanent analysis of the process evolution, examining above all: item calibrations, tool calibrations, persistency of the calibrations during a sequence, dynamic of the *loss-in-quality* and responsible variable(s), besides changes in oscillation and in the co-relational structure among variables.

2.1. Calibrations

Calibrations are a short run aspect and are related to the instant positioning of the tools. Firstly, $\boldsymbol{x}$, $\overline{\boldsymbol{x}}$ and $\{\boldsymbol{x}\}$ define the item calibrations, the tool calibrations (averages of calibrations) and the calibration sequence respectively. The average of calibrations of the n items may change, for instance when a tool setting gets out of position or breaks down. When the production is perfectly calibrated the process quality is definable in terms of variability. Calibration problems reduce the quality and appear very frequently, but they can be easily handled by the operators. From the point of view of the *persistency* we suggest the measure $f_{\{i,...,i'\}} = (i' - i + 1)/n$, $(i < i' = 2, ..., n)$, that is the fraction of items of the sequence $\{i, ..., i'\}$ with reduced spread. As definition of *loss-in-quality* of each calibration we suggest two statistics directly derivable from T^2-distance (referred also as χ^2-distance): $L = (2p)^{-1/2}T^2$ and $I = k_\alpha^{-1}T^2$, where k_α is the critical value of the χ^2-distribution at level α and α is the probability of false alarm. The utility depends on the fact that L may be interpreted as *quality loss* in terms of standard deviation and I is a *quality loss* index for *in-control* items. Coherently, the statistics on the *loss-in-quality* of tool calibrations are: $L_j = 2^{-1/2}\tilde{\Delta}_j$, and $I_j = k_\alpha^{-1}\tilde{\Delta}_j$, $j = 1, 2, ..., p$, where $\tilde{\Delta}_j$ is the median of the *contributes* of j variable as defined below. For a sequence of items we may take: $L_{\{i,...,i'\}} = \max_{\{i,...,i'\}} L$ and $I_{\{i,...,i'\}} = \max_{\{i,...,i'\}} I$. Using the *technical rank* between the variables it is possible to estimate the co-relation structure in a-symmetric manner and to work on the in-correlated $\boldsymbol{x}^*$ variables, where the $\boldsymbol{x}^*$ are obtained by conditioning the variables with respect to all previous ones involved. Moreover, if the independent identical $\boldsymbol{x}$ come from $N_p(\boldsymbol{\mu_x}, \boldsymbol{\Sigma_x})$, under the null hypothesis of perfect calibrations, $\mathbf{x}^*$ is distributed as a $N_p(\boldsymbol{\mu_{x^*}}, \boldsymbol{\Sigma_{x^*}})$.

2.2. Oscillations

To measure the item dispersion we define the mobile oscillations $\boldsymbol{o}$, where $o_i = x_i - x_{i-1}$ $(i = 2, ..., n)$, here named *vertical* oscillations also, are the differences between successive determinations along each variable. My opinion is that the mobile oscillations capture the variability of individual items in the sense that as the dispersion increases the oscillation enlarges. There are no multivariate individual charts for the dispersion at present. The $\boldsymbol{o}$, nearest similar to the traditional mobile range, may be used to monitor the item variability. Typically large oscillations are a medium run problem and could happen, for instance, when a fixed tool becomes loose or wears out or the machine needs ordinary maintenance and/or usual cleaning care. We evaluate: item oscillation $\boldsymbol{o}$, tool oscillation $\overline{\boldsymbol{o}}$ and sequence oscillation $\{\boldsymbol{o}\}$ for handling the oscillations $\boldsymbol{o}$ in the same manner as the calibrations. In fact this is important: if the independent identical $\boldsymbol{x}$ comes from $N_p(\boldsymbol{\mu_x}, \boldsymbol{\Sigma_x})$, under the null hypothesis of unchanged dispersion, $\mathbf{o}^*$ is distributed as $N_p(\boldsymbol{\mu_{o^*}}, \boldsymbol{\Sigma_{o^*}})$.

2.3. Co-Relations

For monitoring the co-relation structure we use the mobile oscillations of the cumulated sums of the variables $\boldsymbol{c}$ and we name it co-relations directly. We suggest to consider the cumulated sums (*horizontal* sums) because their oscillations get the inter-dependences. Co-relation structure problems could happen, for instance, when a machine wears out or needs extra-ordinary maintenance or unusual care. Obsolete machine and their wear out represent a long run situation. In fact, the oscillations of the sums $\boldsymbol{c}$ capture the co-relation structure in the sense that when the co-relations change, the cumulated sums tend to enlarge. This means that $\boldsymbol{c}$ may also be handled in the same manner as calibrations and oscillations. We study the item co-relation $\boldsymbol{c}$, the tool co-relation $\overline{\boldsymbol{c}}$ and the sequence co-relations $\{\boldsymbol{c}\}$. In literature there are no multivariate individual charts for co-relations now. Note that, if the variables have the same measure unit and are comparable in magnitudes, if the independent identical $\boldsymbol{c}$ make sense and $\boldsymbol{x}$ come from $N_p(\boldsymbol{\mu_x}, \boldsymbol{\Sigma_x})$, under the null hypothesis of stable co-relation, $\mathbf{c}^*$ is distributed as $N_p(\boldsymbol{\mu_{c^*}}, \boldsymbol{\Sigma_{c^*}})$.

2.4. Variable Contributions

In the multivariate process control it is difficult to determine which of the variables is responsible for the out-of-control signal and for this reason we develop the contribution chart. Generally, variable contributions to a signal have two principal disadvantages: the direction of the correction is not considered and the co-relation structure between the variables is not properly involved. First we use the star-plot chart of $\boldsymbol{x}^*$ to detect the direction of the deviations and then we look at the contribution of each variable appearing in an associated control chart, where the contributions can be measured observing the changes of the T^2-distance obtained by dropping each variable in turns. The value of $T^2_{(j)}$ is the corresponding value by dropping the j-th variable. So the contribution of x_j variable, $\Delta_j = T^2 - T^2_{(j)}$, is a diagnostic metric that can be properly computed by any software package that calculates a χ^2-distance. To compare the magnitude of each Δ_j, because under *null* hypothesis the x^*_j does not contribute to the signal, Δ_j has the χ^2-distribution with one degree of freedom and the corresponding $1 - \alpha$ percentile is the critical value that gives a guideline to the importance of each Δ_j.

3. The Method

Throughout this paper it is assumed: i) the homogeneity of the production (this happens, for instance, when articles are produced by the same machine; in case of several machines we must adopt control charts for each machine separately); ii) the variables have the same measure unit and comparable magnitudes; iii) the sequence of the variables reflects their *technical rank*; iv) the $\boldsymbol{x}^*$ is a multinormal random variable, that is $\mathbf{x}$ is distributed as $N_p(\mu_x, \Sigma_x)$. We express in detail the calibration problems because the other two aspects, oscillations and co-relations, are reducible to the first one. Here we remark only that under ii) hypothesis the vertical differences, the horizontal sums and their use make sense. The knowledge of local shifts (short run problem) are usually of primary interest, i.e. where a set of items in sequence differs from the others, can also provide useful information: for example, after a control action the method permits to see how closely we come to the target. For evaluating item shifts the calibration analysis usually needs the 99%-*control limit* and the 95%-*attention level*. We incorporate the extra-information contained in the *technical rank* using $\boldsymbol{x}^*$. Note that $T^2_x = T^2_{x^*}$ (where $\boldsymbol{x}$ and $\boldsymbol{x}^*$ origins are remarked) because of the invariance property, but the stars and the contributions look differently. Then the most important necessary steps are: 1) producing the $\mathbf{x}^*$ variables by conditioning each x_j on the previous ones involved and evaluating $f_{\{i,\ldots,i'\}}$, L and I; 2) drawing control chart of $\mathbf{x}^*$; 3) studying the persistency of the sequences and evaluating the achievable best quality; 4) calculating the contributions of $\mathbf{x}^*$ and drawing the associated variable contribution chart; 5) operating the same steps as in 2-6 for $\boldsymbol{o}$ and $\boldsymbol{c}$. The method consists in observing for all pieces the *calibration*, the *oscillation* and the *co-relation*. In fact the **xoc**-method shows all pieces and controls if $\boldsymbol{x^*}$, $\boldsymbol{o^*}$ and $\boldsymbol{c^*}$ are within the control limits. In this case the production is retained *under control*, otherwise it is *out-of-control* and the process must be stopped for carrying out tool revisions and calibrations. The superior limit, i.e for $\boldsymbol{x^*}$, is derived from the probability of false alarm $P[T^2_{x^*} \geq k_\alpha|\mu_{x^*}]$ where k_α is the critical value of the χ^2-distribution at α level. Here we have only superior limits, because the lower levels of χ^2 are good accepted and not problematic. However, the unique superior limit is justified by the fact that a distance $T^2 \leq k_{1-\alpha}$ is less likelihood from technological point of view. This common level k_α is the same for $\boldsymbol{x^*}$, $\boldsymbol{o^*}$ and $\boldsymbol{c^*}$ but, in general, it may be different. Note that it is possible to make inference on item calibrations not only on distance, but also on the *loss-in-quality* measure L and on the index I. In fact we achieve the following results: L and I are distributed as $\chi^2_p[(2p)^{-1/2}, 1]$ and $\chi^2_p[pk_\alpha^{-1}, 2pk_\alpha^{-2}]$, respectively (where the values in parenthesis are the average and the variance). The power function of the method is: $\gamma(\mu_{\mathbf{x}^*}, \mu_{\mathbf{o}^*}, \mu_{\mathbf{c}^*})=P[(T^2_{x^*} \geq k_\alpha) \cup (T^2_{o^*} \geq k_\alpha) \cup (T^2_{c^*} \geq k_\alpha)]$ and under the conjecture that $T^2_{x^*}$, $T^2_{o^*}$ and $T^2_{c^*}$ are independent distances the probability of false alarm for the **xoc**-method becomes $1-(1-\alpha)^3$.

4. A Practical Application

It follows here an example about a switch-drum (Kreuter factory, data in Flury [2], p. 151). We begin with a on-line star chart (see Statgraphics [3]) on $p = 5$ variables, using at first stage (past data) $n = 35$. In this case the target (in mm) is $\mu = (18, 10, 13.5, 11, 7.5)$ and the tool calibrations are $\overline{\mathbf{x}}=(-0.171; +0.171; +0.129; -0.171; +0.014)$. We can estimate in an a-symmetric manner a reasonable inter-dependence between the variables of the following type: $(x_1, x_2) \to x_3$; $x_2 \to x_4$; $(x_2, x_3, x_4) \to x_5$, so that the conditioned variables are: $x_1^* = x_1$; $x_2^* = x_2$; $x_3^* = x_3|(x_1, x_2)$; $x_4^* = x_4|x_2$; $x_5^* = x_5|(x_2, x_3, x_4)$; the conditioned target is: $(18; 10; -0.0401; 0.3544; -0.0866)$ and the tool calibrations are $\overline{\mathbf{x}}^*=(-0.171; +0.171; +0.04; -0.354; 0.087)$. Working on $\boldsymbol{x^*}$ we see

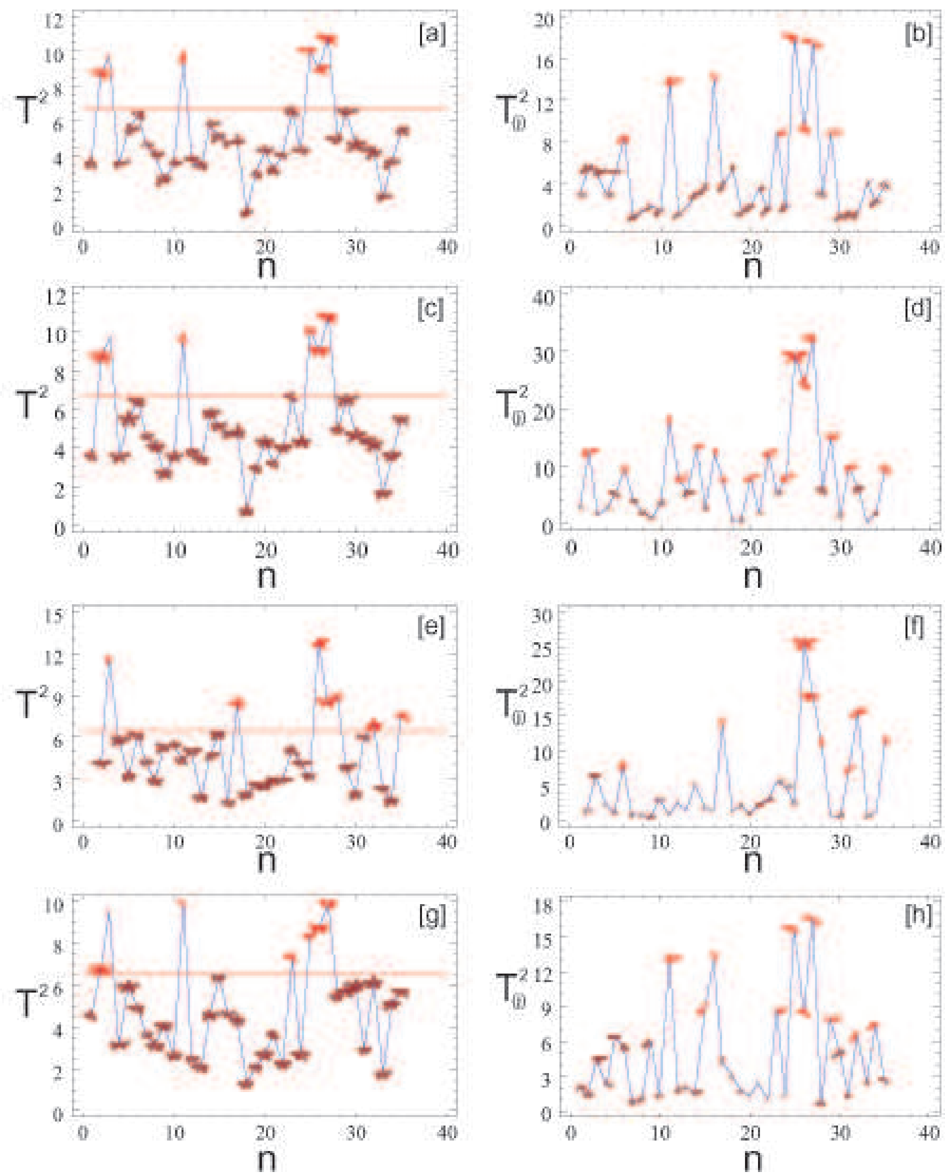

Figure 1: T^2-distance and attention level (red line) for $\boldsymbol{x}^*$, $\boldsymbol{x}$, $\boldsymbol{o}^*$, $\boldsymbol{c}^*$ (on the left) and related contribute charts (on the right)

individual item calibrations in the on-line chart (see Figure 1 a) and, item by item, the directions of the variables: for example, item 3 has all calibrations under the target and the problem is on tool number two (see contribution chart Figure 1 b). The control limit line 11.07 is associated with the 99% confidence level and 6.46 is the attention level corresponding to 95% threshold. Note that over the attention threshold the stars are big or small, but they are of medium size under this line. The variables on the stars are placed in clockwise order beginning from 12 o'clock. A big ray in a star means that the calibration is over the target, while a small ray in a star shows that the dimension is under the target. There is a first relative stable calibration sequence during $\{7, ..., 13\}$ interrupted only by item 11 because of x_2^* which is in the negative direction (the value is smaller than the target); there is a second relative good calibration period during $\{18, ..., 24\}$, apart from item 23, yet because of x_2^* which is also in the negative direction, and a third relative good calibration sequence during $\{30, ..., 35\}$ with a better item 33, because the variables x_2^* appear adjusted in positive direction (the values is greater than the target). In the contribution chart the mainly persistence of variable x_2^* is relevant and the instrument number 2 must be constantly monitored. If we are able to realize the production conditions such as in $\{18, ..., 22\}$-sequence we may achieve the maximum quality gain. The evolution of the quality-loss for item calibrations is: $L_{18} = 0.22 \leq L \leq L_{27} = 3.406$ and $I_{18} = 0.029 \leq I \leq I_{27} = 0.973$. Good items must usually have $L_1 \leq 1$ and $I \leq 30\%$. Tool calibrations are $(-0.228; 0.104; 0.182; 0.186; 0.18)$, so the reduction in quality due to tool calibrations is about 2/10 (different from that in $\overline{\boldsymbol{x}}$). Through the best sequence we have $f_{\{18,...,22\}} = 0.143$, $I_{\{18,...,22\}} = 0.431$ and this means: we can gain about 57% in quality. The conditioning for $\boldsymbol{o}$ is the same as for $\boldsymbol{x}$, while for $\boldsymbol{c}$ it is: $c_1^* = c_1$; $c_2^* = c_2|c_1$; $c_3^* = c_3|(c_1, c_2)$; $c_4^* = c_4|(c_1, c_2, c_3)$; $c_5^* = c_5|(c_2, c_3, c_4)$. For oscillations $\boldsymbol{o^*}$ see Figures 1 e-f; co-relations $\boldsymbol{c*}$ are in Figures 1 g-h. The charts show no serious oscillations before the 25-th items and that the co-relations are relatevely stable. If we realize production conditions as during time units-$\{18, ..., 25\}$ the maximum quality gain in oscillation reduction is estimated above 65.6%.

5. Conclusive Remarks

The individual multivariate charts for calibrations, oscillations and co-relations are typical for on-line situations. So it is necessary and useful to convince traditional multivariate chart users of the value of individual multivariate charts when extra-information related to *technical rank* is introduced. We recommend to use properly the metric after conditioning each variable in turns. Similar traditional methods do not utilize co-relations in a proper way. It is very convenient that the procedure about the oscillation and co-relation aspects works in the same manner as that about the calibrations. Standard charts, usually not individual, are: profile-plot, poly-plot, star-plot and multivariate box-plot, as in Atienza [1]. Unfortunately in the literature the standard variable contribution procedures and the multivariate charts are wrong or inappropriate when process variables are related. We suggest six new multivariate charts: a calibration chart, an oscillation chart, a co-relation chart and the associated contribution charts. It is possible to show that, when μ_{x^*} and Σ_{x^*} are unknown, the Hotelling distribution is more useful than the χ^2-distribution (Wierda [5]). The method also works even by no correspondence between variables and tools: when a single variable is detected, there is often a direct relationship between that variable and a corrective action (for example, the material dilatation could not involve some equipment but the effect of the expansion may depends on pressures, temperature, flows, etc.). If the variables are incomparable in terms of magnitudes it is possible to use the *relative principal components* technique (Stoppa [4]). In a future work we could study the charts for the second phase.

References

[1] O.O. Atienza, L.C. Tang, B.W. Ang, Simultaneous monitoring of univariate and multivariate spc information using boxplots, *International Journal of Quality Science*, **3**, No. 2 (1998), 194-204.

[2] B. Flury, H. Riedwyl, *Multivariate Statistics: A Practical Approach*, Chapman and Hall, London (1988).

[3] Statgraphics 4.1, *User's Manual*, Manugistic, Maryland, USA (1998).

[4] G. Stoppa, Relative principal components, In: *Workshop on Laser Spectroscopy for Laser Gas Detection*, 18-20/02, Sardagna-Trento, Italy (2004).

[5] S. Wierda, Multivariate statistical process control – recent results and directions for future research, *Statistica Neerlandica*, **48**, No. 2 (1994), 147-68.