Management and Production Engineering Review
Volume 3 • Number 4 • December 2012 • pp. 80–86
DOI: 10.2478/v10270-012-0036-7

SOME NOTES ON MULTIVARIATE STATISTICAL PROCESS CONTROL

Michał Rogalewicz

Poznan University of Technology, Chair of Management and Production Engineering, Poland

Corresponding author:
Michał Rogalewicz
Poznan University of Technology
Chair of Management and Production Engineering
ul. Piotrowo 3, 61-131 Poznań, Poland
phone: +48 61 665 27 98
e-mail: michal.rogalewicz@put.poznan.pl

Received: 29 October 2012
Accepted: 23 November 2012

Abstract
In the article some features of multivariate statistical process control were introduced. Advantages and disadvantages of multivariate control charts (Hotelling charts) – counterpart of traditional Shewhart control charts in multivariate space – were also shown. The author pointed out the advantage of Hotelling charts over traditional ones in some situations. The main part of the paper is a multivariate surveillance procedure based on the division on two stages: Phase I (phase of stabilization) and Phase II (phase of monitoring). After describing some elements of multivariate statistical process control implementation (considered from the methodological point of view) author introduced two most popular control charts used in industrial practice: control charts for individual observations and for means. The article is summarized by some author's conclusions concerning popularity of multivariate surveillance and future directions of his work.

Keywords
Hotelling control charts, MSPC (multivariate statistical process control), multivariate surveillance.

Introduction

There is no such a manufacturing process being able to produce the same products taking into account repeatability of their final parameters and properties. There is always some variability existing in process which causes a difference between pieces of product manufactured even directly one after another. The essential cause of differences between values of product characteristics coming from the same process is existing of some disturbances which can result from natural (random) or special causes. Natural causes of variation (called also non-assignable causes) are practically unavoidable in specified conditions. There is many of them but none points out dominant influence on process and is subject to meaningful change in time. Natural variability generated by this kind of causes is very difficult to limit and requires very serious changes e.g. in technological process, machine selection or supplier choice. The second group of causes of process variation is called special causes (or assignable causes). They tend to affect a process in systematic (process characteristics change gradually in time) or sporadic way (suddenly change affecting process). Variability generated by this kind of causes is comparatively easy to identify and limit or eliminate [1].

Because the process characteristics are dependent on various sources of variation affecting the manufacturing process there is a need to control it to manufacture high quality products. The set of statistical methods and tools which help in specifying if the process is affected only by natural causes or maybe also by special causes of variation is called statistical process control (SPC). The most well-known SPC tool is control chart designed in 1924 by Walter Shewhart.

Traditional control charts let one monitor stability and capability of the process or a result of this process – product. They are very simple graph charts based on probability distribution (the most often normal distribution). To supervise the process one draws against time some statistics from the process and compares its values against earlier specified control limits. The position of control limits results from the distribution of drawn statistic. If the points on the control chart do not indicate special patterns (e.g. some consecutive points above or below the center line, some consecutive points rising or falling or points outside the control limits) the process is supposed to be in-control (there is only natural - random variation present), otherwise one should believe it is out-of-control (there are special causes of variation present). The most popular and wise is using two kinds of control charts simultaneously – the first monitors a measure of location, the other a measure of variation (the most well-known and prevalent in industry control chart is Shewhart's $\overline{x} - R$).

Because of its simplicity and transparency Shewhart control charts became one of the most popular tools of controlling the process. Unfortunately they have some serious limitations.

Traditional Statistical Process Control charting (called also Univariate SPC) consists in plotting only one characteristic on the control chart at the same time. In today's industry there are sometimes many variables deciding on process/product quality. In this case process engineer should control all of them to keep the process stable and assure its high quality. That is why the most popular approach in industrial practice is designing a univariate control chart for each quality characteristic. In this case process operator, who is responsible for supervising the process (sometimes some processes simultaneously) can have big problems with this wide range of duties. He must collect measurements and register them, compute and plot all the points on each control chart and look for symptoms of process maladjustment (Fig. 1).

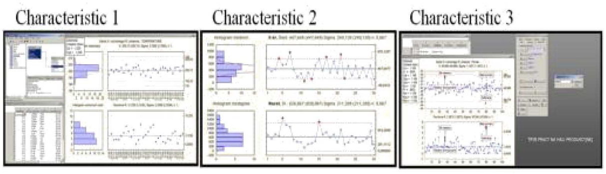

Fig. 1. Use of single control chart for each quality characteristic (source: StatSoft).

Usage of many control charts not only overwhelms the operator but can also lead to wrong conclusions about the process state. Simultaneous monitoring of many characteristics is conducted assuming its independence. Very often this statement does not hold true and can lead to serious errors in process of inference. Besides many control charts for one process monitoring means serious distortion for I type error (I type terror (α) – specifies a probability that the statistic plotted on the control chart will lie outside the control limits, assuming that process is stable. One should remember that such a symptom leads to unnecessary stopping the process and adjusting it that is why α should not be too high) of the whole procedure. Any control chart has its own I type error and specifying what is a value of an accumulated error is (assuming dependence between characteristics) practically impossible. An answer for these limitations of traditional Shewhart control charts are their counterparts in the multivariate space – multivariate control charts.

Multivariate control charts – advantages and disadvantages

Multivariate control charts were introduced in 1947 by Harold Hotelling [2] and that is the reason why they are often called Hotelling charts. They enable to aggregate information concerning a few process/product variables on one control chart using so called T^2 statistic. This statistic is a measure of the distance between these variables' values (or its sample statistics) and vector of its means (known or estimated from the base sample). Besides it takes into account a structure of correlations between variables in the form of covariance matrix. Such way of computing the T^2 statistic lets one find the process malfunction in multivariate space not only in a few univariate spaces of individual variables. The difference between these two ways of controlling the process is shown in Fig. 2.

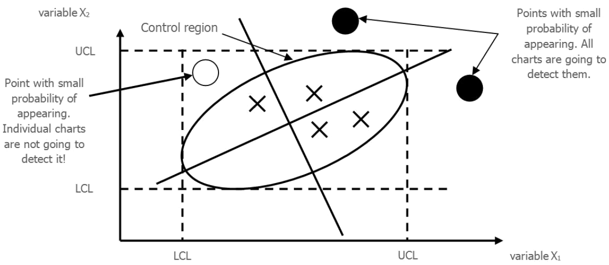

Fig. 2. Two-dimensional control chart in comparison with two individual control charts [own work on the basis of 3].

Multivariate control chart is clearly more effective in detecting some symptoms of the process being out of control than traditional one. In the figure above a difference between two-dimensional control chart (shown in the form of control region; control region is an analogy to control limits on the traditional Shewhart control charts. Field of this region (ellipse) is smaller than rectangle made by two separate control charts because of taking into account a correlation between variables) and two individual univariate control charts were demonstrated. It can be seen that traditional Shewhart control charts do not make possible to detect distinct symptom of maladjustment (marked white), which is outside the control region on two-dimensional control chart. It takes place because traditional control charts do not take into account that the white point is clearly against the positive correlation on which the control region was based. When the variable X_1 takes the small value, so should the variable X_2. When it is not in this case, it means that the process went out of control.

Apart from taking into consideration the correlation between variables the serious advantage of multivariate control chart is a constant and possible to determine value of I type error. It is an error which is fixed during the designing of the control chart.

Another advantage of Hotelling charts is mentioned above aggregating information from many variables on the one chart. It is particularly important for the operator because his task is to monitor the process using only one, not many, control charts. It is definitely easier for him to assure the quality of the process.

Unfortunately multivariate control charts have also some disadvantages. The most important one is connected with the interpretation of signals appearing on the control chart. In multivariate surveillance it can be a serious problem because out of control point can stand for at least some situations: process is out-of control because of one variable, because of acting together two or more variables or because of the change in the covariance matrix. The procedure of establishing what really happened with the process can be very complicated.

Another drawback of Hotelling charts is its comparatively complicated apparatus and using matrix algebra. It is surely not easy to understand even basic principles of this procedure without strong mathematical background. From the other side even a statistic drawn on the Hotelling chart is not friendly for the operator. Because of the aggregation of some variables T^2 is not in units of any particular variable what can be a little bit confusing.

In spite of these disadvantages multivariate control charts can be very useful in industrial practice, especially taking into account that many processes are multivariate in nature and characteristics deciding on quality of the process/product are very often correlated. It seems that these days when computational methods and computers are still rapidly developing their implementation in production environment should not meet such barriers as even some years ago. Taking into account the fact that in many processes huge amount of data is gathered often in automatic way and that very frequently this data does not serve any purpose because of its complexity and weak knowledge about methods of its analysis one can assume that there is a need to develop Multivariate SPC tools and adjust them to industrial practice.

Multivariate control procedure

There are some very important steps in putting the multivariate statistical process control into practice. In this chapter some assumptions are specified taking into account rather methodological not organizational steps.

There are two distinct phases of establishing control of the process:

- I phase – phase of stabilization – it serves to establish a statistical control of the process. The base sample is collected and the process is checked for stability and capability. Some common assumptions are also checked.
- II phase – phase of monitoring – it serves to keep process stable.

Assuming there is a process chosen for implementing multivariate statistical process control using Hotelling control charts and some organizational mechanisms which lead to obtain a vector of controlled variables were implemented one must take into account some very important assumptions connected particularly with the Phase I. There are some main assumptions to check and assure when introducing Hotelling charts [3]:

- selecting a base sample of independent (random) observations,
- the data should follow a multivariate normal distribution (MVN). Because there is very difficult to find out if the data follows MVN, Mason and Young suggest checking the distribution of T^2 statistic. It is proved that if data distribution is multivariate normal then T^2 statistic follows the univariate Beta distribution so it is possible to check it the other way. If T^2 follows Beta distribution then with high probability the data distribution is described by MVN,
- collecting a sufficient sample size – one has to remember that in multivariate control procedure there is often a need to estimate a great number of parameters, e.g. for two variables and control charts using samples there are 2 means, 2 variances and 1 covariance but for three variables there are already 9 parameters to estimate. The

more variable the definitely more parameters to estimate, much more than in univariate case. That is why the size of sample for multivariate procedure should be sufficiently big.

After checking these assumptions at the beginning of the Phase I one should examine the stability of the process i.e. stability of variables' parameters. For this purpose one must collect so called base sample from the process and estimate these parameters – mean vector and covariance matrix (assuming both are unknown) – it is particularly important to use a consistent covariance matrix estimator. Then one should design the control chart for the Phase I and check collected sample for outliers. It is particularly important because this base sample will be used to examine a capability of the process and compute accurate control charts for the Phase II. If any point plots above the control limit there is a need to identify and remove the source of variation, discard the sample from base sample and recalculate control limits. One should repeat this actions until the moment when base sample is clear of outliers.

The next step in Phase I is specifying the capability of the process. Once again this computation is not as easy as in univariate case (e.g. Cp and Cpk). Taking into account earlier assumptions about multivariate normality of the distribution, the data create an ellipsoid in multivariate space and tolerance is specified as a range for each individual variable. There are at least some approaches to computing multivariate capability indices [4]:

- indices based on vector representation,
- indices based on comparison of tolerance area to area of process variability,
- indices defined using Principal Components Analysis (PCA),
- indices based on the fraction of nonconforming units,
- other approaches.

It seems that the first group of indices is the most informative one because they usually consist of three components. Some author's analysis lead to conclusion that probably one of the best indices was introduced by Shahriari and Abdollahzadeh [4]. New vector of multivariate capability contains three components: NMCpm, PV and LI which represent comparison of tolerance and natural variability area, location of the centers of the areas and mutual location of areas respectively. The author recommends using this capability index.

If the values of the capability indices are sufficient one can use the parameters computed from the base sample to calculate control limits for the Phase II – phase of monitoring. Also in this situation there is a distinct difference between multivariate and individual control charts. In univariate case limits computed in Phase I were used directly as control limits in Phase II. In multivariate case very often there is a need to compute new control limits to monitor the process. When one computed the control limits for the Phase II it is possible to look for patterns appearing on the chart which indicate an existence of special sources of variation. When such signal appears an operator should stop the process and adjust it.

As it was indicated before the signal produced by the control chart is not easy in interpretation and can be caused by one variable, by acting together two or more variables or because of the change in the covariance matrix. There are some methods of interpreting the signals on the multivariate control charts [5–7] but the most popular is so called MYT decomposition [8]. The T^2 statistic is divided into two parts: conditional terms and unconditional terms. The decomposition terms contain information about the residuals which are generated by all possible linear regressions of one variable on the any other subset of variables. It is a great aid in locating the source of special variation in terms of individual variables or subset of variables.

All procedures of multivariate SPC are followed by the assumption that there is no change in covariance matrix. That is why it is so important to collect a base sample of sufficient size and use a consistent estimator of covariance matrix. But what about the situation when during the process the change in covariance matrix appears? The most popular approach to monitor changes in covariance matrix is so called generalized variance chart [Fig. 3]. The statistic drawn on the chart is a determinant of the covariance matrix. If its change is significant it is very possible that the covariance matrix also changed.

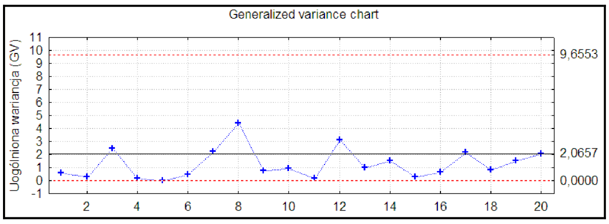

Fig. 3. Generalized variance chart (source: StatSoft).

Multivariate control charts for individual observations and means

Depending on the fact if the sample collected from the process contains single observation or a few observations it is possible to distinguish between two most popular Hotelling control charts: charts for individual observations and chart for means. Below some formulas concerning these two charts were demonstrated [9].

Multivariate control charts for individual observations

Multivariate control charts for individual observations are used whenever there is no possibility or there is no economic explanation of collecting the bigger sample than $n = 1$ e.g. when production rate is too slow, measurements are particularly expensive or destructive. The common and most important assumption in this case is connected with the distribution of data – it should be checked for multivariate normality.

Below there are some formulas for computing control limits for multivariate control charts for individual observations as well as the formula for point plotted on the chart. The formulas are divided into two parts connected with the phase of establishing control of the process.

Phase I:

Lower Control Limit:

$$\mathrm{LCL} = \frac{(m-1)^2}{m} B_{1-\frac{\alpha}{2},\frac{p}{2},\frac{m-p-1}{2}}. \tag{1}$$

Upper Control Limit:

$$\mathrm{UCL} = \frac{(m-1)^2}{m} B_{\frac{\alpha}{2},\frac{p}{2},\frac{m-p-1}{2}}. \tag{2}$$

Plot point:

$$Q_i = (X_i - \overline{X}_m)' S_m^{-1} (X_i - \overline{X}_m), \tag{3}$$

where $B_{\frac{\alpha}{2},\frac{p}{2},\frac{m-p-1}{2}}$ is a 1-α-th percentile of Beta distribution with parameters $\frac{p}{2}$ and $\frac{m-p-1}{2}$, m – size of base sample, p – number of variables, X_i – vector of individual observations, $\overline{X}_m$ – vector of means (computed from the base sample), S_m^{-1} – inverse of covariance matrix, $()'$ – denotes transparent matrix, α – significance level.

Phase II:

Lower Control Limit:

$$\mathrm{LCL} = \frac{p(m+1)(m-1)}{m(m-p)} F_{1-\frac{\alpha}{2},p,m-p}. \tag{4}$$

Upper Control Limit:

$$\mathrm{UCL} = \frac{p(m+1)(m-1)}{m(m-p)} F_{\frac{\alpha}{2},p,m-p}. \tag{5}$$

Plot point – the same as in Phase I, where $F_{\alpha,p,m-p}$ is 1-α-th percentile of F distribution with p and $m-p$ degrees of freedom, the rest symbols as above.

Multivariate control charts for means

Multivariate control charts for means are used when there is a possibility to collect some homogenous samples with size $= n$ which describe the process at the moment of collecting.

Below there are some formulas for computing control limits for multivariate control charts for means as well as the formulas for point plotted on the chart. The formulas are divided into two parts connected with the phase of establishing control of the process.

Phase I:

Upper Control Limit:

$$\mathrm{UCL} = \frac{p(m-1)(n-1)}{mn-m-p+1} F_{\alpha,p,mn-m-p+1}. \tag{6}$$

Plot point:

$$T^2 = n(\overline{X} - \overline{\overline{X}})' S_{pooled}^{-1} (\overline{X} - \overline{\overline{X}}), \tag{7}$$

where $F_{\alpha,p,mn-m-p+1}$ is 1-α-th percentile of F distribution with p and *mn-m-p-1* degrees of freedom, m – number of samples, p – number of variables, n – size of the sample, $\overline{X}$ – vector of sample means, $\overline{\overline{X}}$ – vector of grand averages, S_{pooled}^{-1} – inverse of covariance matrix pooled over all samples.

Phase II:

Upper Control Limit:

$$\mathrm{UCL} = \frac{p(m+1)(n-1)}{mn-m-p+1} F_{p,mn-m-p+1} \tag{8}$$

Plot point – the same as in Phase I, where $F_{\alpha,p,mn-m-p+1}$ is 1-α-th percentile of F distribution with p and *mn-m-p-1* degrees of freedom, the rest symbols as above.

The multivariate control charts for individual observations and for means are very similar so there was only one control chart demonstrated as an example (Fig. 4).

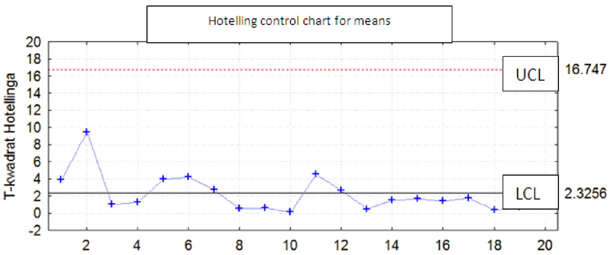

Fig. 4. Hotelling control chart for means (source: StatSoft).

Conclusions and further research

During the last years multivariate statistical process control becomes more and more popular. It is connected with the development of computational techniques and usage of computers in industrial practice. Besides in the area of research some modifications of a traditional multivariate control chart were made for example to treat small shifts in the process mean (e.g. MEWMA [10], MCUSUM [11]). Many articles have been published concerning recognition of signals on these charts or interpretation which variable is responsible for process out-of-control conditions. There are also more and more implementations of above mentioned method and tools in industrial e.g. in semiconductors production, cars assembling, chemical and petrochemical industry, mining industry, biochemistry, steel production and paper production (e.g. [12–15]). It is supposed that using very advanced and sophisticated tools of MSPC is the case in the companies characterized by highly developed quality awareness or complex manufacturing processes taking into account many quality variables.

The author made some initial research concerning using of MSPC tools in Wielkopolska region. The results of this research show that unfortunately in polish companies still using univariate SPC definitely dominates over using MSPC. It is supposed that many complicated tools and sophisticated mathematical apparatus of Multivariate Statistical Process Control discourage quality practitioners from getting acquainted with them. Definitely easier is using traditional control charts and capability indices because usage of them and interpretation does not make any problem and is clear.

It seems obvious that the next reason for using mostly univariate SPC is a fact that the knowledge about MSPC is not too popular among practitioners. A great number of books or even manuals concerning SPC does not mention about multivariate case. It is very easy to check out that the bible of automotive industry – SPC handbook published by Ford, General Motors and Chrysler [16] – devotes only few pages to that subject.

The third important reason of the lack of popularity of MSPC in Poland is lack of requirements from the side of big companies concerning making use of more sophisticated control procedures. Any requirements for suppliers concern only traditionally used methods and tools so it is nothing weird that companies do not practice MSPC. As it was mentioned above it is supposed that only enterprises with very complex processes and aware of the meaning of high quality of processes/products are interested in making their control schemes better and better.

Taking into account these conclusions it seems very important to promote an idea of using multivariate SPC tools. In the era of modern, sophisticated and very often automated manufacturing processes it should not be a problem to implement even complicated methods to industrial practice. An operator doesn't have to fully understand the background of MSPC. It should be enough for him to catch the main idea of multivariate surveillance.

In author's opinion there is however a strong need for clear methodology of practical implementing multivariate control charts to process control. There is a lack of guidelines giving such a step-by-step instructions. Except of [3] and [17] there is very few books containing detailed characteristic of all important MSPC tools and methods together with instructions concerning its use and accompanying assumptions. Working out such a methodology is a purpose of author's doctoral thesis. The assumptions concerning this methodology were described in [18].

Multivariate Statistical Process Control is a promising set of tools and methods which can be very helpful in controlling the process. However step-by-step instruction of implementing MSPC is the strongest need form the practitioners point of view can definitely help in building a bridge between scientists and practitioners. In author opinion in few next years it will be well-seen a systematic growth of industrial implementations of Multivariate Statistical Process Control contributing to transforming data into knowledge about process.

References

[1] Hamrol A., *Quality control with examples*, PWN, Warszawa, 2008.

[2] Hotelling H., *Multivariate Quality Control Illustrated by Air Testing of Sample Bombsights*, C. Eisenhart et al., 111–184, 1947.

[3] Mason R.L., Young J.C., *Multivariate Statistical Process Control with Industrial Applications*, ASA-SIAM, Philadelphia, 2002.

[4] Shahriari H., Abdollahzadeh M., *A new multivariate process capability vector*, Quality Engineering, 21, 290–299, 2009.

[5] Hawkins D.M., *Multivariate quality control based on regression-adjusted variables*, Technometrics, 33(1), 61–75, 1991.

[6] Hawkins D.M., *Regression adjustment for variables in multivariate quality control*, Journal of Quality Technology, 25(3), 170–182, 1993.

[7] Mason R.L., Tracy N.D., Young C.H., *Decomposition of T^2 for multivariate control chart interpretation*, Journal of Quality Technology, 27(2), 99–108, 1995.

[8] Mason R.L., Tracy N.D., Young J.C., *A practical approach for interpreting multivariate T^2 control chart signals*, Journal of Quality Technology, 29(4), 396–406, 1997.

[9] Ryan T., *Statistical methods for quality improvement*, Wiley, 2011.

[10] Lowry C.A., Woodall W.H., Champ C.W., Rigdon S.E., *A multivariate exponentially weighted moving average control chart*, Technometrics, 34, 46–53, 1992.

[11] Pigniatello R., Runger G., *Comparisons of multivariate CUSUM charts*, Journal of Quality Technology, 22(3), 173–186, 1990.

[12] Fuchs C., Kenett R., *Multivariate quality control. Theory and applications*, Quality and Reliability 64, M. Dekker Inc., New York, 1998.

[13] Li P., Wang X., Wang O., *Approach to using MSPC for power plant process monitoring and fault diagnosis*, Huagong Xuebao/Journal of Chemical Industry and Engineering, 59(7), 1790–1796, 2008.

[14] Kourti T., Lee J., MacGregor J., *Experiences with industrial applications of projection methods for multivariate statistical process control*, Computers and chemical engineering, 20, 747–752, 1996.

[15] Lahman-Schalem S., Haimovitch N., Shauly E., Daniel R., *MBPCA for fault detection in NMOS fabrication*, IEEE Transactions on Semiconductor Manufacturing, 15(1), 60–69, 2002.

[16] Chrysler, Ford, General Motors, *Statistical Process Control (SPC) reference manual*, Second Edition, 2005.

[17] Rogalewicz M., Kujawińska A., *Traditional or multivariate SPC – what to choose?*, Conf. Management of Technology Step to Sustainable Production, Rovinj (Croatia), 2–4.06.10, Conference Proceedings – [CD-ROM].

[18] Yang K., Trewn J., *Multivariate statistical methods in quality management*, Mc Graw-Hill, 2003.