A Comparison Study of Control Charts for Statistical Monitoring of Functional Data

Bianca M Colosimo, Massimo Pacella

HAL Id: hal-00548947

https://hal.archives-ouvertes.fr/hal-00548947

Submitted on 21 Dec 2010

A Comparison Study of Control Charts for Statistical Monitoring of Functional Data

RESEARCH ARTICLE

A Comparison Study of Control Charts for Statistical Monitoring of Functional Data

Bianca M. Colosimo[a] and Massimo Pacella[b*]

[a]*Dipartimento di Meccanica, Politecnico di Milano, Milano, ITALY*;
[b]*Dipartimento di Ingegneria dell'Innovazione, Università del Salento, Lecce, ITALY*

(*October 2008*)

Quality of products and processes is more and more often related to functional data, which refer to information summarized in form of profiles. Recent literature pointed out that traditional control charting methods cannot be directly applied in these cases and new approaches for profile monitoring are required. While many different profile monitoring approaches have been proposed in the scientific literature, few comparison studies are available up to now. This paper aims at filling this lack by comparing three representative profile monitoring approaches in different productive scenarios. The performance comparison will allow one to select a specific approach in a given situation. The competitor approaches are chosen for representing different levels of complexity, as well as different types of modelling approach. In particular, at a lower level of complexity, the "location control chart" (where the upper and lower control limits are $\pm K$ standard deviations from the sample mean at each profile location) is considered as representative of the industrial practice. At a higher complexity level, approaches based on combining a parametric model of functional data to multivariate and univariate control charting are considered. Within this second class, we analyse two different approaches. The first is based on regression and the second focuses on using principal component analysis for modelling functional data. A reference case study in manufacturing is used throughout the paper, namely, profiles measured on machined items subjected to geometrical specification (roundness).

Keywords: Quality, control chart, profile monitoring, functional data, PCA, regression, spatial statistic.

1. Introduction

With the development of computerized data-acquisition systems and of modern measuring equipments, quality of products or processes is more and more often related to functional data. Functional data (Ramsay and Silverman 2005) refer to information summarized in form of *profiles* where each data point is the observed response at a given location (spatial or temporal). Since functional data are usually observed only on a finite set of points, profiles can be stored as multivariate vectors. Hence, the most direct way to deal with these data consists in treating each profile as a realization of a multivariate process. However, use of standard multivariate charts is ill-advised because when the number of monitored points exceeds the number of collected samples, the covariance matrix is singular and the common multivariate statistics cannot be calculated. This condition can often be encountered in actual applications, especially in the case of machined profiles

*Corresponding author. Email: massimo.pacella@unile.it

subject to geometrical specification (e.g., roundness, straightness, and free-form tolerance) where, in order to have an accurate estimate of the form error, the number of observations sampled is usually in the order of hundreds. Hence, when the quality of a process or product is characterized by functional data, new approaches are required. Recently, there has been much research activity in this new area of statistical monitoring. In particular, Woodall *et al.* (2004) discussed general issues to be addressed when monitoring quality profiles, and presented a complete review of the literature on the topic of *profile monitoring*. The approaches for profile monitoring proposed in the literature share a common structure which consists of: *i*) identifying a parametric model of the functional data; *ii*) estimate the model parameters; *iii*) design a multivariate control chart on the estimated parameters and a univariate control chart on the residual variance. The proposed approaches can be then classified with reference to the type of application faced (i.e., calibration study, process signal or geometric specification monitoring) or to the modelling approach considered (mainly linear regression or approaches for multivariate data reduction as the principal/independent component analysis). In particular, new approaches for profile monitoring have been extensively used in calibration (Kang and Albin 2000, Kim *et al.* 2003, Mahmoud and Woodall 2004, Gupta *et al.* 2006) where the profiles that have to be monitored are straight lines. The aims of these researches were to establish performance of the measurement method and to verify that it remained unchanged over time. Profile monitoring was also exploited when the analysed response is represented by a generic physical signal (e.g., an acoustic signal or an electrical power signal). For instance, Jin and Shi (1999, 2001) referred to profiles as waveform signals and cited examples of force and torque signals collected from online sensors on a press in a stamping process.

A further example of functional data discussed in the literature refers to the set of points measured on a machined profile which is subject to geometrical specifications. In this case, each point collected on the machined profile is related to a specific spatial location. Colosimo *et al.* (2008) explored a method in this area of research. Their study focused on the spatial correlation which often characterizes adjacent points of the machined profile. In fact, adjacent points measured on a machined profile are obtained in similar conditions of the manufacturing process and related to similar local properties of the machined material, and hence profile data are spatially correlated. Colosimo *et al.* (2008) showed that an appropriate parametric model of machined profiles is a Spatial AutoRegressive regression (SARX) model (Cressie 1993) to represent the "signature" left by the machining process. Since profile is assumed to be observed on an equally spaced set of points which can be supplemented with neighborhood information, spatial model for lattices was used in their work. Furthermore, in order to signal any deviation from the in-control behavior, coefficients used to describe process' signature were used to design a T^2 control chart for profile monitoring, while the estimate of the residual variance was monitored with a univariate control chart.

Colosimo and Pacella (2007) explored the advantages of using a data-reduction approach for geometric product specification modelling. In particular, they showed that principal component analysis (PCA) can be an excellent exploratory method for interpreting profiles obtained from manufacturing processes. PCA has been shown to be useful in exploratory analysis, i.e., for detecting and modelling the systematic ways in which profiles are varying around the mean profile. The choice of using a data reduction method instead of regression can be motivated by its inner suitability in dealing with complex profiles, without requiring a specific model type selection. In fact, the selection of a proper type of regressors or the use of non-linear regression can become cumbersome activities, which can represent an

obstacle to the introduction of profile monitoring in actual industrial applications. Ding *et al.* (2006) compared PCA to a different type of data reduction method, namely, independent component analysis (ICA), for Phase I monitoring of profiles. According to their study, ICA has to be preferred to PCA when the assumption of multivariate normality of the original data set (the profile data) is not respected when significant components have to be identified (in Phase I). In other words, ICA has to be preferred to PCA when the monitoring procedure has to be designed starting from contaminated profile data. In the following, we will not deal with this type of design issue and this is why we will use from now on PCA as representative of multivariate data reduction techniques.

Despite of the specific modelling issues behind approaches for profile monitoring, practitioners may be tempted to skip their adoption at all, attracted by the unbeatable simplicity of the approach referred to as "location control chart". The location control chart can be considered the simplest approach for functional data monitoring as it consists in applying a traditional control chart to data observed at each given location (Boeing 1998, Woodall *et al.* 2004). This approach results in a control region where the upper and lower control limits are $\pm K$ standard deviations from the sample mean at each location. According to this method, an alarm is issued when at least one point, in the whole set of data observed in a profile, exceeds the control limits.

The objective of the present research is to provide a comparison of different approaches for monitoring functional data, namely, *i*) the location control chart that, thanks to its inner simplicity, can be considered as representative of industrial practice; *ii*) a regression-based approach, *iii*) a PCA-based approach. The comparison is based on different simulation scenarios which are obtained by perturbing an actual industrial case study of machined profiles subject to geometrical specification (roundness). The perturbed scenarios are designed for representing different although realistic productive situations. Performance are measured both as the ability to obtain a predefined false alarm rate in the design phase of the control chart (the so-called Phase I) as well as to detect unusual patterns in the functional data during the operating phase of the control chart (known as Phase II).

For Phase I analysis, only groups of no contaminated process samples are considered in our study. These in-control profiles are obtained by simulation, while performance of the competing methods are compared in terms of the probability of obtaining at least one statistic outside the control limits when performing control charting using the set of simulated samples. Control chart parameters estimated from these Phase-I samples are not used in the subsequent Phase-II comparison study, where performance comparison is based on the ideal assumption that the in-control parameters are known. Indeed, computer simulation is used in our work to obtain a large dataset of in-control profiles in order to estimate as closely as possible control charts parameters.

This paper is basically aimed at investigating situations where each approach should be preferred to a different one, thus providing some guidelines for implementing profile monitoring in industrial manufacturing. The remainder of this paper is organized as follows. In section 2, the competing approaches for profile monitoring are presented. In section 3 the reference case study is presented and the simulation model is detailed along with the different scenarios considered for the comparison. In section 4, the approaches for profile monitoring are compared with reference to Phase I, while in section 5 performance are compared with reference to Phase II. Eventually, section 6 reports the conclusions and some final remarks.

2.   Different approaches for profile monitoring

2.1   *The location control chart*

A method for profile monitoring in discrete part manufacturing was presented by Boeing (1998, pp. 89-92) with reference to applications in which measurements of the same variable (e.g., a dimension such as thickness) are made at several locations on each manufactured part. The method, also known as the location control chart, consists in applying a traditional Shewhart control chart separately to each data point observed at a given location of the part.

Assume to collect a sample of n profiles, where each profile is represented by a vector of p measurements observed at a fixed set of locations. Assuming this set of n profiles is observed from an in-control process, the control limits can be separately computed for each location by using the Shewhart approach, i.e., by considering the sample mean and the sample standard deviation of the n data observed at that location, and by computing the control limits as the traditional $\pm K$ standard deviations from the sample mean.

Given a specific profile, an alarm is issued when at least one point, in the set of p observations, exceeds either the upper or the lower control limit at the specific location. It is worth noting that the control limits used at each location depend only on the responses at that specific position, thus, the main disadvantage with this method is that the multivariate structure of data is ignored at all.

2.1.1   *Control limits of the location control chart*

Let $y_j(k)$ denote the data measured at a specific location of index k on the j-th profile, where $k = 1, \ldots p$ and $j = 1, \ldots n$. The control limits can be separately computed for each location k as follows.

$$
\begin{aligned}
UCL(k) &= \bar{y}(k) + z_{\alpha/2}s(k) \\
CL(k) &= \bar{y}(k) \\
LCL(k) &= \bar{y}(k) - z_{\alpha/2}s(k),
\end{aligned}
\tag{1}
$$

where $\bar{y}(k) = \frac{1}{n}\sum_{j=1}^{n} y_j(k)$ and $s(k) = \sqrt{\frac{1}{n-1}\sum_{j=1}^{n}\left(y_j(k) - \bar{y}(k)\right)^2}$ are respectively the sample mean and the sample standard deviation of data observed at location k, while $z_{\alpha/2}$ represents the $100(1 - \alpha/2)$ percentile of the standard normal distribution.

Given that p dependent control rules are simultaneously applied, Bonferroni's rule for dependent events should be used to attain an actual false alarm rate not greater than a predefined value. Therefore, let α' denote the upper bound of the first type probability error (false alarm probability), the value $\alpha = \alpha'/p$ is used for designing the p control limits of equation (1).

2.2   *The regression based approach*

A different approach, proposed in the recent literature on profile monitoring (Kang and Albin 2000, Kim *et al.* 2003, Woodall *et al.* 2004), consists in combining a parametric model for functional data to multivariate and univariate control charting. The parametric model is intended to represent functional data as characterized by a predictable behavior and a natural variability.

With reference to the vector of points measured on a machined profile which is

subject to geometrical specification, the predictable behavior is the so-called process' *signature*, defined as the systematic pattern that characterizes all the features machined with that process. The knowledge of this signature can be used to design proper tools for profile monitoring. Colosimo *et al.* (2008) presented an approach for modelling the manufacturing signature based on fitting a Spatial Autoregressive Regression (SARX) model (Cressie 1993). The regression model presented by Colosimo *et al.* (2008) is summarized in the following.

Assume to collect a sample of n profiles observed on a fixed set of p equally spaced locations (note that this assumption is not required for the location control chart presented in previous section). Let $y_j(k)$ denote the observation of location with index k on the j-th profile, where $k = 1, \ldots p$ and $j = 1, \ldots n$. Assuming to organize the p data observed on the j-th profile into a (column) vector $\mathbf{y}'_j = [\, y_j(1) \ldots y_j(k) \ldots y_j(p) \,]$, the general SARX model can be written in matrix notation as follows.

$$\mathbf{y}_j = \mathbf{X}\mathbf{b}_j + \boldsymbol{\nu}_j$$
$$(\mathbf{I} - \mathbf{R}_j)\boldsymbol{\nu}_j = \boldsymbol{\varepsilon}_j \tag{2}$$
$$\mathbf{R}_j = \sum_{s=1}^{q} a_{sj}\mathbf{W}^{(s)}.$$

The first expression in (2) describes the $p \times 1$ vector of responses for the j-th profile $\mathbf{y}_j$ as formed by a *large scale* and a *small scale* component (Cressie 1993).

The large scale component is given by $\mathbf{X}\mathbf{b}_j$, where $\mathbf{X}$ is a $p \times r$ matrix of r regressor variables that are assumed to be known and constant and $\mathbf{b}'_j = [\, b_{1j} \ldots b_{lj} \ldots b_{rj} \,]$ is the $r \times 1$ vector of regression parameters which are normally distributed with mean $\boldsymbol{\beta}' = [\, \beta_1 \ldots \beta_l \ldots \beta_r \,]$ and covariance matrix $\mathbf{B}$ ($\mathbf{b}_j \sim MN(\boldsymbol{\beta}, \mathbf{B})$).

The small-scale component is the $p \times 1$ vector of error terms $\boldsymbol{\nu}_j$ of equation (2). Error terms are assumed to be spatially correlated and are represented as a generic Spatial AutoRegressive process (SAR) of order q. The SAR(q) model expression is given in the last two expressions of (2), where $\mathbf{I}$ is the $p \times p$ identity matrix, $\boldsymbol{\varepsilon}_j$ is a $p \times 1$ vector of independently and normally distributed errors $(\boldsymbol{\varepsilon}_j \sim MN(\mathbf{0}, \sigma_\varepsilon^2 \mathbf{I}))$, $\mathbf{a}'_j = [\, a_{1j} \ldots a_{sj} \ldots a_{qj} \,]$ is the vector of coefficients of the SAR(q) model for the j-th profile which is assumed to be normally distributed with mean $\boldsymbol{\alpha}' = [\, \alpha_1 \ldots \alpha_s \ldots \alpha_q \,]$ and covariance matrix $\mathbf{A}$ ($\mathbf{a}_j \sim MN(\boldsymbol{\alpha}, \mathbf{A})$).

Eventually, the $p \times p$ matrix $\mathbf{W}^{(s)}$ of elements $w^{(s)}(k, t)$ ($k, t = 1, \ldots p$) represents the core of the small scale model, since it is the s-th order adjacency matrix (or *spatial weights matrix*). For instance, the generic element of the first-order adjacency matrix $w^{(1)}(k, t)$ is set equal to 1 if the t-th point is the neighbor of the k-th point and 0 otherwise. Analogously, the element of a second-order adjacency matrix, $w^{(2)}(k, t)$, is set equal to 1 if the t-th point is a neighbor of the original first-generation neighbors of the k-th point, and so on. By definition, all the adjacency matrices are binary and symmetric matrices whose diagonal elements are zero (Cressie 1993).

For each profile, two vectors of coefficients $\mathbf{b}_j$ and $\mathbf{a}_j$ are related to the large scale and the small scale behaviors, respectively. In order to let the model have the most general form, we further assume that these two vectors could be also correlated, i.e., $cov(\mathbf{b}_j, \mathbf{a}_j) = \mathbf{D}$. In other words, with reference to the parametric model structure given in (2), we merge the two vectors characterizing the observed

6   *Taylor & Francis and I.T. Consultant*

pattern in a single coefficient vector related to the j-th profile:

$$\mathbf{c}'_j = \begin{bmatrix} \mathbf{b}'_j & \mathbf{a}'_j \end{bmatrix} = \begin{bmatrix} b_{1j} \ldots b_{lj} \ldots b_{rj} & a_{1j} \ldots a_{sj} \ldots a_{qj} \end{bmatrix}; \tag{3}$$

$$\mathbf{c}_j \sim MN(\boldsymbol{\mu}, \boldsymbol{\Sigma}) \quad \text{where} \quad \boldsymbol{\mu}' = \begin{bmatrix} \boldsymbol{\beta}' & \boldsymbol{\alpha}' \end{bmatrix} \quad \text{and} \quad \boldsymbol{\Sigma} = \begin{bmatrix} \mathbf{B} & \mathbf{D} \\ \mathbf{D}' & \mathbf{A} \end{bmatrix}.$$

2.2.1 *Control limits of regression based control charts*

The signature model for the j-th profile shown in equation (2) is completely defined by a SARX model which requires one to estimate the $d = r + q$ parameters which are components of the vector $\mathbf{c}_j$, besides the residual variance $s_j^2 = \widehat{\sigma}_\varepsilon^2$. Let $\widehat{\mathbf{c}}'_j = \begin{bmatrix} \widehat{\mathbf{b}}'_j & \widehat{\mathbf{a}}'_j \end{bmatrix}$ represent the vector of d parameters' estimates for the j-th profile, a T^2 control chart can be designed with reference to the statistic:

$$T_j^2 = (\widehat{\mathbf{c}}_j - \widehat{\boldsymbol{\mu}})' \widehat{\boldsymbol{\Sigma}}^{-1} (\widehat{\mathbf{c}}_j - \widehat{\boldsymbol{\mu}}), \tag{4}$$

where $\widehat{\boldsymbol{\mu}}$ and $\widehat{\boldsymbol{\Sigma}}$ are respectively the sample mean vector and covariance matrix of the d-dimensional vectors of coefficients estimated during the design phase (Phase I) of control charting.

Given a set of n profiles collected in Phase I, for each profile one can estimate the vector $\widehat{\mathbf{c}}_j$ ($j = 1, \ldots n$) and compute $\widehat{\boldsymbol{\mu}}$ as the sample mean of this set of vectors. Starting from the same set of vectors, a possible estimator of $\boldsymbol{\Sigma}$ is given by (Sullivan and Woodal 1996, Vargas 2003)

$$\widehat{\boldsymbol{\Sigma}} = \frac{1}{2} \frac{\mathbf{V}'\mathbf{V}}{p - 1}, \tag{5}$$

where the $(p - 1) \times d$ matrix $\mathbf{V}$ consists of row vectors of the differences $\mathbf{v}'_j$ given by:

$$\mathbf{v}_j = \widehat{\mathbf{c}}_{j+1} - \widehat{\mathbf{c}}_j \quad j = 1, \ldots n - 1.$$

Williams *et al.* (2006) studied the performance of different control limits to be used in Phase I for the T_j^2 statistic when the estimated covariance matrix is based on equation (5). As a result of their study, while the distribution of the T_j^2 statistic does not have a simple closed form, its asymptotic distribution is a Chi-squared. In particular, when the number of samples n is at least twice the number of parameters estimated $(d + [d(d + 1)/2])$, the upper control limit based on the asymptotic distribution of T^2 can be used

$$UCL = \chi^2_{\alpha,d},$$

where $\chi^2_{\alpha,d}$ is the $100(1 - \alpha)$ percentile of the Chi-squared distribution with d degrees of freedom.

The $p \times 1$ vector of estimated residuals associated to the j-th profile can be described as

$$\mathbf{e}_j = \left(\mathbf{I} - \widehat{\mathbf{R}}_j\right)\left(\mathbf{y}_j - \mathbf{X}\widehat{\mathbf{b}}_j\right),$$

where $\widehat{\mathbf{R}}_j = \sum_{s=1}^{q} \widehat{a}_{sj}\mathbf{W}^{(s)}$. Hence, the estimated variance of residuals s_j^2 is given

by

$$s_j^2 = \frac{\mathbf{e}_j' \mathbf{e}_j}{p-1}.$$

In order to monitor the residual variance, traditional Shewhart-type control chart can be used. In this work, the control limits used to monitor the residual variance are based on the χ^2 distribution with $p-1$ degrees of freedom where p is the number of points monitored. In particular, denoting by $\bar{s}^2$ the sample residuals' variance of the n profiles, the control limits are computed as follows.

$$UCL = \frac{\bar{s}^2}{p-1} \chi_{\alpha/2, p-1}^2$$

$$CL = \bar{s}^2$$

$$LCL = \frac{\bar{s}^2}{p-1} \chi_{1-(\alpha/2), p-1}^2,$$

where $\chi_{\alpha/2, p-1}^2$ and $\chi_{1-(\alpha/2), p-1}^2$ are the upper and lower $\alpha/2$ percentage points of the Chi-squared distribution with $p-1$ degrees of freedom associated with the residuals (Montgomery 2004).

Given two control charts (the T^2 and the Shewhart control chart) are contemporarily used, the Bonferroni's rule for independent events was used to design the control limits of the two charts. Therefore, given a nominal false alarm probability α', the control limits of each of the two control charts are set by assuming a false alarm probability $\alpha = 1 - \sqrt{1 - \alpha'}$.

2.3   *The PCA-based approach*

Ramsay and Silverman (2005) presented an extension of PCA to functional data, i.e., an approach which allows one to find a set of orthonormal functions (also called functional principal components - PCs) so that the original data can be approximated in terms of a linear combination of these basis functions.

In particular, Ramsay and Silverman (2005) showed that, in the case of equally spaced observations, the easiest way to compute the PCs consists in modelling the curve data sampled at regular intervals as a multivariate vector, and performing a traditional PCA on the set of samples collected over different curves. When the PCA outlines a set of significant PCs to be retained, the coefficients (or loadings) defining these significant PCs can be interpreted as eigenfunctions (also called empirical orthogonal functions). These eigenfunctions do not have a parametric expression and are empirical, since they are obtained by the data at hand. A rough sketch of how PCA works is reported in the following.

Assume to organize a sample of n vectors of $p \times 1$ data $\mathbf{y}_j$ $(j = 1, \ldots n)$ into a $n \times p$ data matrix $\mathbf{Y}$ whose j-th row is equal to the transpose of the j-th data vector $\mathbf{y}_j$. PCA consists in performing a spectral decomposition of the covariance matrix of $\mathbf{Y}$. The covariance matrix describes the variability of data observed at each location with respect to the mean value observed at the same location in all the profiles. Therefore, a first step in PCA consists in centring the data by subtracting the average profile.

Say $\mathbf{S}$ the covariance matrix, i.e., $\mathbf{S} = \frac{1}{n-1} \sum_{j=1}^{n} (\mathbf{y}_j - \bar{\mathbf{y}}) (\mathbf{y}_j - \bar{\mathbf{y}})'$, where $\bar{\mathbf{y}} =$

8   *Taylor & Francis and I.T. Consultant*

$\frac{1}{n}\sum_{j=1}^{n}\mathbf{y}_j$ is the sample mean profile, the spectral decomposition consists in finding the $p \times p$ matrices $\mathbf{U}$ and $\mathbf{L}$ which satisfy the following relationship

$$\mathbf{U}'\mathbf{S}\mathbf{U} = \mathbf{L}, \tag{6}$$

where $\mathbf{L}$ is a diagonal matrix which contains the eigenvalues of $\mathbf{S}$ (say l_k), while $\mathbf{U}$ is an orthonormal matrix whose k-th column, $\mathbf{u}_k$ is the k-th eigenvector of $\mathbf{S}$, i.e., the so-called loadings.

With reference to the j-th profile $\mathbf{y}_j$, denote with $\mathbf{z}_j$ the vector

$$\mathbf{z}_j = \mathbf{U}'(\mathbf{y}_j - \bar{\mathbf{y}}) = \begin{bmatrix} z_{j1} & \cdots & z_{jk} & \cdots & z_{jp} \end{bmatrix}',$$

where z_{jk} are the the so-called scores. Each profile can be then expressed as a linear combination of loadings $\mathbf{u}_k$, where the weights of the linear combination are the scores z_{jk}.

$$\mathbf{y}_j = \bar{\mathbf{y}} + z_{j1}\mathbf{u}_1 + z_{j2}\mathbf{u}_2 + \ldots + z_{jp}\mathbf{u}_p.$$

Since the PCs are statistically independent and each PC has variance equal to the corresponding eigenvalue (l_k), we can rank the PCs, according to the associated eigenvalue and decide to retain just the most important ones (i.e., the ones which are associated with larger variances). Different approaches can be used to select the proper set of PCs (Jackson 2003). For instance, cross-validation can be effectively used to chose the number m of significant PCs (Colosimo and Pacella 2007). When a subset of m out of the whole number of p PCs is retained ($m < p$), the original profile can be estimated as:

$$\widehat{\mathbf{y}}_{j(m)} = \bar{\mathbf{y}} + z_{j1}\mathbf{u}_1 + z_{j2}\mathbf{u}_2 + \ldots + z_{jm}\mathbf{u}_m. \tag{7}$$

2.3.1 Control limits of PCA-based control charts

Similar to the regression based approach, also in the case of PCA a T^2 control chart can be used for monitoring the vector of the first m retained PCs. In this case, the Hotelling statistic is given by (Jackson 2003)

$$T_j^2 = \frac{z_{j1}^2}{l_1} + \frac{z_{j2}^2}{l_2} + \ldots + \frac{z_{jm}^2}{l_m}. \tag{8}$$

If an unexpected event leaded the process to change in a direction orthogonal to that of the first m PCs, the control chart will not be able to issue an alarm. For this reason, another control chart based on the Q statistic (sometimes referred to as Squared Prediction Error or SPE control chart) has to be used as well (Jackson 2003). Given the estimate in (7), the Q statistic can be computed as the sum of the squared errors obtained by reconstructing each observation by the first m PCs:

$$Q_j = (\mathbf{y}_j - \widehat{\mathbf{y}}_{j(m)})'(\mathbf{y}_j - \widehat{\mathbf{y}}_{j(m)}). \tag{9}$$

The upper control limit of the T^2 statistics in equation (8) can be computed as (Williams *et al.* 2006)

$$UCL = \chi_{\alpha,m}^2,$$

where $\chi^2_{\alpha,m}$ is the $100(1-\alpha)$ percentile of the Chi-squared distribution with m degrees of freedom. With reference to the Q statistic, according to Nomikos and MacGregor (1995) the upper control limit can be computed as

$$UCL = g\chi^2_{\alpha,h}, \tag{10}$$

where g and h can be estimated as $\hat{g} = \hat{\sigma}^2_Q/(2\bar{Q})$, $\hat{h} = 2\bar{Q}^2/\hat{\sigma}^2_Q$, while $\bar{Q}$ and $\hat{\sigma}^2_Q$ are the sample mean and the sample variance obtained by computing the Q statistics via equation (9) for the set of n profiles.

Since the two control charts are related to two disjoint set of PCs, the Bonferroni's rule for independent events is used to design the control limits. Thus, given a nominal false alarm probability α', the control limits of each of the two control charts are set by assuming a false alarm probability $\alpha = 1 - \sqrt{1-\alpha'}$.

3. Roundness profiles

3.1 *The reference case study*

The case study used as reference in this paper is described in details by Colosimo *et al.* (2008). It consists of 99 items obtained by turning (cutting speed: 163 m/min; feed rate: 0.2 mm/rev; two steps of cutting depth: 1 mm) C20 carbon steel cylinders (supplied as ⌀30 mm rolled bars), where each item was characterized by a roundness profile of 748 evenly distributed measurements of its radius. The original measurements sampled by using a Coordinate Measuring Machine were scaled by subtracting the least squares estimate of the radius and centred at the least squares estimate of the centre. A further step was eventually applied to register all the sampled profiles by minimizing the phase delay caused by the random contact angle (Colosimo and Pacella 2007).

Starting from actual measurements, the general SARX model of equation (2) was fitted to data. In this case, $\mathbf{y}_j$ represents the column vector of radial deviation from the nominal radius measured at the angular position $\theta_k = k(2\pi/p)$, where $k = 1, \ldots p$ ($p = 748$) is the index of the equally spaced observations on each profile.

Consider the l-th column vector of the regressor matrix $\mathbf{X}$ in equation (2), denote such a $p \times 1$ vector by $\mathbf{x}_l$, where $l = 1, \ldots r$. In general, each element of vector $\mathbf{x}_l$ is described as function of the index location k. In the specific case of roundness profiles, $x_l(k)$ can be expressed either as $x_l(k) = \cos((k-1)f_h)$ or $x_l(k) = \sin((k-1)f_h)$, i.e., as a sinusoidal function of frequency equal to $f_h = h(2\pi/p)$ rad/sample. h is the frequency ($h \in \{1, 2, \ldots p/2\}$) measured in undulation per revolution (upr), which is fixed for all the elements of $\mathbf{x}_l$.

Two harmonics were selected for modelling the radial deviations in the actual test case, namely, the second and the third one. Indeed, process signature was mainly affected by ovality and triangularity. The oval contour was possibly due to a bi-lobe error motion affecting the spindle's lathe or to eccentricity caused by an improper setup, while the three-lobe pattern was due to a similar error motion of the spindle.

Therefore, the regressor matrix $\mathbf{X}$ in equation (2) has 4 columns ($r = 4$) since two sinusoidal functions are needed to model the amplitude and phase of each specific harmonic. The k-th row of matrix $\mathbf{X}$ is equal to $[\cos((k-1)f_2)\ \sin((k-1)f_2)\ \cos((k-1)f_3)\ \sin((k-1)f_3)]$, where $f_h = h(2\pi/p)$ represents the frequency in rad/sample of the h-th harmonic ($h = 2, 3$).

Eventually, the vector of random error $\boldsymbol{\nu}_j$ in equation (2) was fitted as a SAR

Table 1.   Model parameters estimated on 99 actual round-ness profiles (Colosimo *et al.* 2008). Mean vector $\widehat{\boldsymbol{\mu}}' = \left[\widehat{\boldsymbol{\beta}}'\widehat{\boldsymbol{\alpha}}'\right]$ and covariance matrix $\widehat{\boldsymbol{\Sigma}} = \begin{bmatrix} \widehat{\mathbf{B}} & \widehat{\mathbf{D}} \\ \widehat{\mathbf{D}}' & \widehat{\mathbf{A}} \end{bmatrix}$ of the co-efficient vector $\widehat{\mathbf{c}}_j$.

$\widehat{\boldsymbol{\beta}}'$	-0.0341	0.0313	0.0080	-0.0322
$\widehat{\boldsymbol{\alpha}}'$	0.3021	0.2819		
$\widehat{\boldsymbol{B}}$	4.07E-04	-2.02E-04	6.54E-05	2.65E-05
	-2.02E-04	3.90E-04	1.49E-04	6.10E-06
	6.54E-05	1.49E-04	2.24E-04	-1.07E-05
	2.65E-05	6.10E-06	-1.07E-05	3.12E-04
$\widehat{\boldsymbol{D}}$	-8.84E-05	-2.41E-04		
	-1.21E-04	1.96E-04		
	-1.18E-04	5.96E-05		
	-1.50E-04	-3.72E-04		
$\widehat{\boldsymbol{A}}$	3.80E-03	1.59E-03		
	1.59E-03	4.32E-03		

model of order 2 ($q = 2$) (Colosimo *et al.* 2008), using the algorithm implemented in the Spatial econometrics toolbox (LeSage 1999).

With reference to the model in equations (2, 3) each specific profile of index j is associated to the vector of $d = p+q = 4+2$ estimated parameters $\widehat{\mathbf{c}}_j \sim MN\left(\widehat{\boldsymbol{\mu}}, \widehat{\boldsymbol{\Sigma}}\right)$, where $\widehat{\boldsymbol{\mu}}' = \left[\widehat{\boldsymbol{\alpha}}' \ \widehat{\boldsymbol{\beta}}'\right]$ and $\widehat{\boldsymbol{\Sigma}} = \begin{bmatrix} \widehat{\mathbf{B}} & \widehat{\mathbf{D}} \\ \widehat{\mathbf{D}}' & \widehat{\mathbf{A}} \end{bmatrix}$. Values of vectors and matrices estimated on the actual roundness data are shown in Table 1.

3.2   *Productive scenarios under study*

Starting from the real case under study, different scenarios are exploited to anal-yse various, although realistic, characteristics of the machined profiles. The main objective is to let the different scenarios outline possible situations in which one approach for profile monitoring should be preferred to the others. Starting from the characteristics of our real case, we will perturb this reference scenario with ref-erence to different factors, characterizing either the large-scale or the small-scale components of the profile model, or both. With reference to the large-scale compo-nent, fixed-effect models are usually assumed in traditional approaches for profile monitoring. This situation can be seen as a special case of our general model, where $\mathbf{B} = \mathbf{0}$ is the variance matrix of the large-scale model coefficients $\mathbf{b}_j \sim MN(\boldsymbol{\beta}, \mathbf{B})$. Alternately, when $\mathbf{B} \neq \mathbf{0}$ is considered while simulating profiles, random effects are included.

Note that when PCA is performed in the first case (fixed-effect model), no sig-nificant PCs are identified (Colosimo and Pacella 2007). In fact, PCA is usually performed after data centring and this first step consists in subtracting the mean pattern (described by the fixed-effect model) from each profile data. In this case, PCA is thus performed just on the error terms and hence no significant PC is correctly reported. In terms of PCA-based control charting, the T^2 statistic in (8) is not defined since $m = 0$, and only the univariate control chart based on the Q statistic in (9) can be used to detect out-of-control profiles. Note also that in this case the Q statistic is given by the sum of the squared difference between data observed at each location and the average profile at that location.

On the other hand, in the case of random effects in the large-scale model, the number m of retained PCs is greater than zero. In the reference case study of this

paper, Colosimo and Pacella (2007) showed that the number of PCs to be retained is equal to $m = 4$. In this last case, the PCA approach can take full advantage of both the T^2 and the Q control charts.

With reference to the small-scale component, a first source of perturbation focuses on the spatial correlation structure. In fact, we will consider the case in which no spatial correlation occurs, assuming only independently and identically distributed (iid) errors included in the model. This situation can be modelled as a particular case of the general model by assuming $q = 0$ and hence $\mathbf{R}_j = \mathbf{0}$ in (2). The situation of iid errors can arise in industrial practice when the sampling frequency of the machined profile is reduced. In fact, measurements on a machined profile are spatially correlated when adjacent points are obtained in similar conditions of the machining process and related to similar (local) properties of the machined material. This situation can be observed in particular when the sampling frequency on the machined profile is high, i.e., when adjacent measurements are separated by a very short spatial distance. Obviously, as the distance between consecutive points increases, the spatial correlation between adjacent points starts decreasing. Therefore, iid errors represent situations in which the sampling frequency on each profile is low, while the cases of spatially correlated errors, represent an higher sampling frequency.

Similarly to what is done for the large-scale component, variability of the small-scale model coefficients is the second source of perturbation concerning the error terms. In fact, also for the spatial structure coefficients $\mathbf{a}_j \sim MN(\boldsymbol{\alpha}, \mathbf{A})$ we can assume $\mathbf{A} = \mathbf{0}$ to represent a low degree of variability from profile to profile. In fact, assuming that the input material is very stable and homogeneous, profile-to-profile variability of the spatial structure can be probably neglected. On the other hand, if $\mathbf{A} \neq \mathbf{0}$ is considered while simulating profiles, random effects are included in the small-scale coefficients. However, when the properties of the input materials are stabler and/or more controlled, this type of profile-to-profile variability tends to vanish.

It is worth noting that, if random and fixed effects are both present, then a linear mixed-effect model should be preferred for the regression-based approach, especially when observations within profiles are correlated. In fact, Jensen *et al.* (2008) pointed out that a linear mixed model can be used on the whole dataset of profiles instead of simply using standard least squares separately for each profile and then estimating the covariance matrix of the estimated coefficients. With reference to Phase I analysis of no contaminated data, the study of Jensen *et al.* (2008) showed that standard least squares and mixed model give similar results when data are balanced, as in the case we are dealing with. Nonetheless, in order to obtain in our study an accurate estimation of coefficients when observations within profiles are correlated, a SARX model (Cressie 1993) is implemented in the regression-based approach instead of using standard least squares. The algorithm implemented in the spatial econometrics toolbox (LeSage 1999) is implemented for this purpose.

Eventually, all the characteristics of the simulated scenarios are summarized in Table 2. Note that the actual real case used as starting reference is labelled "scenario 6", while all the other perturbed scenarios are labelled from 1 to 5.

4.   Phase I: Performance Comparison

In Phase I, a historical set of n process samples is analysed in order to both evaluate the stability of the process and estimate the in-control state's parameters. During Phase I, performance of the competing control chart methods are compared in

Table 2.   Simulated scenarios.

		Small-scale component of the model Spatial correlation	
	no	yes	
		variability	
		no	yes
Large-scale variability		($\mathbf{A} = \mathbf{0}$)	($\mathbf{A}$ in Table 1)
no ($\mathbf{B} = \mathbf{0}$)	scenario 1	scenario 3	scenario 5
yes ($\mathbf{B}$ in Table 1)	scenario 2	scenario 4	scenario 6

terms of the probability of deciding whether or not the process is stable. This is the probability of obtaining at least one statistic outside the control limits when performing control charting using the set of n process samples.

Given a desired false alarm or Type I error rate α', assumed equal to $\alpha' = 1\%$ in the following, the three competing approaches can be compared in terms of the probability of actually achieving this nominal value.

To compare the performance of the alternative approaches we considered $10,000$ replicates of Phase I control charting for different values of the number of profiles used in this phase, specifically $n = 50, 100, 150$ and 200. We recorded the false alarm rate in each replicate (computed as the number of out-of-control signals divided by n) and then we stored the average false alarm rate obtained in the whole set of $10,000$ replicates. Table 3 summarizes the actual Type I error rates for each of the competing methods in the six simulation scenarios considered in our work. Given the large number of simulations considered in our study, a negligible variability characterizes the estimated alarm rates. In fact, the number of false alarm in $n \times 10,000$ trials has a binomial distribution and hence the standard deviation of the Type I error rate α is given by $\sqrt{\alpha(1-\alpha)/(10,000 \cdot n)}$.

For all the scenarios it can be observed that the false-alarm rates produced by the location chart are always smaller than the nominal value (1%), although the gap between actual and nominal rate decreases as the number n of profiles used in Phase I increases. Indeed, this is an expected result as the Bonferroni's inequality allows one to set only an upper bound on the actual false alarm rate. Note that a reduced false alarm rate could seem an advantage at a first sight. However, it means that control limits are too far from the centre line, thus resulting ineffective to detect out-of-control profiles when they arise.

Dissimilarly from the location control chart, the actual false alarm rates obtained by the regression-based approaches are much closer to the nominal value. In fact, as shown in Table 3, despite of both the specific scenario considered and the set dimension n, there is no practical differences between the actual false alarm rate produced by the regression-based approach and the nominal value.

The PCA-based approach is a little more sensitive to the specific scenario under study, since it gives false alarm rate closer to the nominal value in all the scenarios but the last two. Note that in the second and the fourth scenarios (when large scale variability is assumed) at least around 150 profiles are required in Phase I to have good performance (actual false alarm rate closer to the nominal value). Eventually, in the last two scenarios (where the small scale variability is assumed too), an excessive rate of false alarm is achieved. This is mainly due to the Q chart. Such a result shows that the upper control limit for the Q statistic proposed by Nomikos and MacGregor (1995) is substantially adequate when a iid error term is considered in the model, as well as in the case of correlated errors with fixed effects in the small-scale component of the model. However, the limit in (10) is not suitable in the case of correlated errors with random effects in the small-scale component of the model.

Table 3.   Phase I simulation results. Actual Type I error rate for different scenarios and different approaches.

		Scenario 1	Scenario 2	Scenario 3	Scenario 4	Scenario 5	Scenario 6
LOC CC	$n = 50$	0,076%	0,070%	0,048%	0,057%	0,072%	0,070%
	$n = 100$	0,352%	0,307%	0,208%	0,240%	0,298%	0,283%
	$n = 150$	0,513%	0,431%	0,306%	0,343%	0,426%	0,391%
	$n = 200$	0,614%	0,504%	0,353%	0,404%	0,493%	0,454%
REG CC	$n = 50$	0,967%	0,973%	1,049%	1,053%	1,042%	1,048%
	$n = 100$	0,995%	0,981%	1,013%	1,029%	1,008%	1,037%
	$n = 150$	0,999%	0,981%	0,985%	1,037%	1,017%	1,044%
	$n = 200$	0,988%	1,001%	0,988%	1,026%	1,004%	1,031%
PCA CC	$n = 50$	0,825%	0,290%	0,935%	0,392%	1,798%	1,210%
	$n = 100$	0,924%	0,670%	1,022%	0,790%	1,861%	1,495%
	$n = 150$	0,945%	0,809%	1,048%	0,908%	1,901%	1,562%
	$n = 200$	0,958%	0,867%	1,065%	0,939%	1,906%	1,607%

Nominal Type I error rate 1.000%.

Finally, with reference to Phase I simulation results in Table 3, it is fairly to conclude that the regression-based control charts can be easily designed, since the actual false alarm rate can be achieved in practice by using an analytic computation of the control limits. Similar conclusions can be also drawn with reference to the PCA-based approach but only in specific productive scenarios and for large dimension of the data set. On the contrary, the design of a location control chart can be a difficult task, since the actual false alarm rate can be very different from the expected one.

4.1   *A modified Bonferroni procedure for the location control chart*

The location control chart is one of the most simple methods for monitoring profiles. It consists in a "control region" where the upper and lower control limits are $\pm K$ standard deviations from the sample mean at each location of the profile (where K is a constant). A profile can be plotted against this control region with the advantage to allow a simple identification of the locations where problems arise. In the previous section, K has been computed as a function (percentile of the standardized normal distribution) of the desired Type I error corrected by the standard Bonferroni's method. However, different procedures can be also used.

For instance, the Simes' modified Bonferroni procedure (Simes 1986) can be implemented. This procedure makes use of the ordered p-values for each location of a given profile. For a desired false alarm rate α', the ordered p-value of index k, where $k = 1, \ldots p$, is compared to the critical value $\alpha' \cdot k/p$. The profile is considered out-of-control when there exists at least one ordered p-value which is lesser than, or equal to, the critical value.

With reference to this approach, Table 4 summarizes the actual Type I errors for the location control chart in the six scenarios considered in our work (10,000 Phase-I replicates for groups of $n = 50, 100, 150$ and 200 in-control profiles in each scenario). By comparing the results of Table 4 to those of Table 3, it can be observed that, despite of the productive scenario considered, there is only a slight increase of the actual false alarm rates produced by the location chart. Thus, with reference to Phase-I analysis of no contaminated profiles, when compared to the standard Bonferroni's method, the Simes' procedure does not produce significant effects on the actual Type I errors obtained by the location control chart. Furthermore, since the Simes' procedure does not allow the graphical representation of the control region as Bonferroni procedure does, we decide to keep the last one for its inner simplicity and easiness of use as representative of the industrial practice.

Table 4.   Phase I simulation results. Actual Type I error rate for different scenarios when the Simes' modified Bonferroni procedure is applied for the location control chart.

		Scenario 1	Scenario 2	Scenario 3	Scenario 4	Scenario 5	Scenario 6
LOC CC	$n = 50$	0,076%	0,087%	0,053%	0,078%	0,077%	0,091%
	$n = 100$	0,348%	0,324%	0,218%	0,292%	0,304%	0,340%
	$n = 150$	0,519%	0,465%	0,301%	0,411%	0,432%	0,459%
	$n = 200$	0,631%	0,555%	0,370%	0,461%	0,506%	0,525%

Nominal Type I error rate 1.000%.

5.   Phase II: performance comparison

The objective in Phase II of control charting is to quickly detect any change in the process from its in-control state. In this phase of control charting, the competing methods are compared in terms of the average run length (ARL), where the run length is defined as the number of samples taken until an out-of-control signal is issued.

In order to evaluate performance of control charts in this operating phase, occurrence of assignable causes were simulated by a total of six out-of-control conditions, similarly to what was done in (Colosimo *et al.* 2008). Each condition is then characterized by a parameter directly proportional to the "severity" of the out-of-control introduced in the baseline model.

Five out of these six out-of-controls are simulated by spindle-motion errors (Cho and Tu 2001). These kinds of error, which can affect the motion of the spindle, are modelled by introducing one spurious harmonic in the baseline model of roundness profile data, as summarized in Table 5. Note that the spurious harmonic can influence just one frequency (as for the bi-lobe, tri-lobe and four-lobe out-of-control), or more frequencies at once (as in the case of the half frequency, which is known to influence several harmonics at once (Cho and Tu 2001)). Furthermore, frequencies interested by the change can be already included in the original baseline model, as for the case of bi-lobe and tri-lobe spindle-motion error, where the change concerns the sole amplitude of the second or third harmonic, respectively. On the contrary, the out-of-control can influence harmonics which are not included in the original profile data model (as in the case of the four-lobe out-of-control). In this latter case, two sub-cases are considered, namely, a spurious harmonic with either fixed or random phase (uniformly distributed in $[0, 2\pi]$).

Dissimilarly from the previous cases, the sixth out-of-control behavior (not shown in Table 5) does not produce a shift in the amplitude or phase of one or more specific harmonics, but an increase of the error term variance. To simulate this out-of-control the standard deviation of the error term was multiplied by a factor greater than 1. Different values are used to simulate incremental severity of this behavior (the values of 1.01, 1.02, 1.03 and 1.04 were implemented in our simulation study).

Data obtained under these Phase II models were always scaled (by subtracting the least squares estimation of the radius) and centred (on the least squares estimation of the centre) before applying the control charting method (note that the introduction of the half-frequency harmonic affects both the null and the first harmonics, which are related to the average radius and to the position of the centre, respectively). In fact, we are assuming that centring and scaling are standard steps applied to data when the focus is on the out-of-roundness value only (Cho and Tu 2001).

Table 5.   Five out-of-control behaviors modelled as spindle-motion errors

Spindle-motion error	Model	Parameter values
Half frequency	$\sqrt{\frac{2}{p}}\delta_1\sin((k-1)\pi/p)$	$\delta_1 = 0.1, 0.15, 0.2, 0.25$
Bi-lobe	$\sqrt{\frac{2}{p}}\delta_2[c_{1j}\cos((k-1)f_2)+c_{2j}\sin((k-1)f_2)]$	$\delta_2 = 0.1, 0.2, 0.3, 0.4$
Tri-lobe	$\sqrt{\frac{2}{p}}\delta_3[c_{3j}\cos((k-1)f_3)+c_{4j}\sin((k-1)f_3)]$	$\delta_3 = 0.1, 0.2, 0.3, 0.4$
Four-lobe (fixed phase)	$\sqrt{\frac{2}{p}}\delta_4\cos((k-1)f_4)$	$\delta_4 = 0.005, 0.01, 0.015, 0.02$
Four-lobe (random phase)	$\sqrt{\frac{2}{p}}\delta_4\cos((k-1)f_4-\phi)$	$\delta_4 = 0.005, 0.01, 0.015, 0.02$

$k = 1, \ldots p$. Random phase of fourth harmonic distributed as $\phi \sim U[0, 2\pi]$

5.1   *Simulation results*

We used simulation to find the ARLs for the three control charts considered in our comparison study. For each productive scenario under study, all simulations were conducted by first tuning the competing approaches in order to achieve the same in-control ARL value of about 100. A large number of profiles was used to design the control charts for each productive scenario. In fact, control limits for each method were obtained by simulation given $n = 30,000$ instances of in-control profiles.

Hence, the performance of competing approaches in Phase II were related to the ability of detecting out-of-control profiles, given that all the approaches are designed to achieve the same false alarm probability approximatively equal to 1%.

Table 6 through Table 9 summarize simulation results for each of the six productive scenarios under study. Each table reports the ARL performance estimated by computing a set of $1,000$ run lengths, given that the new profiles are simulated according to a specific out-of-control model. Standard deviations estimated for such mean values are also included in the tables within parenthesis.

In order to select the best approach in each different case study (where each case study is characterized by a specific scenario and a specific out-of-control), we need to determine the method corresponding to the lowest value of ARL. Since ties are possible, we further used a very simple rule which identify two approaches as equally performing when the two (approximated) confidence intervals for ARL are overlapping. In particular the approximated confidence interval is simply computed as mean ± 3 times the corresponding standard deviation. Eventually, the ARL values corresponding to the best approaches are highlighted using a bold-type font in Table 6 through Table 9.

5.2   *Discussion*

Two main factors characterize the simulated case studies: the type of out-of-control and type of scenario. With respect to the type of out-of-control, three different classes of changes can be outlined. The first class includes all the out-of-control models where the change is affecting a component which was originally present in the baseline (Phase I) model. Specifically, the half-frequency, the bi-lobe and the tri-lobe models belong to this first class. The second class includes out-of-controls where the change is due to an additional component which was not originally present in the model but it is supposed to suddenly appear affecting the profile signature. The four-lobe (both with fixed and random phase) model belongs to this second class. Eventually, the last class includes out-of-control whose change is affecting the random error variance.

With reference to the type of scenario under study, a further distinction will aid the following discussion. The first type of scenarios will refer to situations in which

Table 6. Phase II simulation results for productive scenarios 1, 3 and 5 (fixed effects in the large-scale component of the model). Out-of-controls influencing one or more harmonics already modelled in the baseline model. Actual ARLs and corresponding standard deviations within parenthesis.

	severity	LOC CC		REG CC		PCA CC	
		Scenario 1					
Half Frequency	0.1	54.37	(1.73)	**1.00**	(0.00)	6.21	(0.18)
	0.15	25.45	(0.85)	**1.00**	(0.00)	1.38	(0.02)
	0.2	10.58	(0.34)	**1.00**	(0.00)	1.01	(0.00)
	0.25	4.74	(0.14)	**1.00**	(0.00)	**1.00**	(0.00)
Bi-lobe	0.1	88.29	(2.80)	**1.99**	(0.05)	45.37	(1.40)
	0.2	61.08	(1.94)	**1.00**	(0.00)	7.35	(0.22)
	0.3	35.68	(1.09)	**1.00**	(0.00)	1.55	(0.03)
	0.4	21.51	(0.65)	**1.00**	(0.00)	1.01	(0.00)
Tri-lobe	0.1	91.93	(2.90)	**5.13**	(0.14)	66.85	(2.07)
	0.2	74.91	(2.24)	**1.07**	(0.01)	22.60	(0.71)
	0.3	55.39	(1.65)	**1.00**	(0.00)	5.68	(0.16)
	0.4	39.47	(1.25)	**1.00**	(0.00)	1.87	(0.04)
		Scenario 3					
Half Frequency	0.1	33.31	(1.00)	**1.27**	(0.02)	6.00	(0.17)
	0.15	12.26	(0.38)	**1.00**	(0.00)	1.51	(0.03)
	0.2	4.81	(0.13)	**1.00**	(0.00)	1.03	(0.01)
	0.25	2.37	(0.06)	**1.00**	(0.00)	**1.00**	(0.00)
Bi-lobe	0.1	72.11	(2.13)	**6.75**	(0.19)	45.89	(1.40)
	0.2	37.70	(1.21)	**1.20**	(0.02)	7.02	(0.20)
	0.3	21.47	(0.63)	**1.00**	(0.00)	1.71	(0.04)
	0.4	10.63	(0.32)	**1.00**	(0.00)	1.05	(0.01)
Tri-lobe	0.1	86.85	(2.72)	**16.70**	(0.50)	64.30	(1.94)
	0.2	60.26	(1.92)	**2.44**	(0.06)	20.18	(0.57)
	0.3	33.59	(1.04)	**1.11**	(0.01)	5.20	(0.15)
	0.4	21.56	(0.66)	**1.01**	(0.00)	1.87	(0.04)
		Scenario 5					
Half Frequency	0.1	33.77	(1.04)	**1.27**	(0.02)	45.44	(1.41)
	0.15	13.66	(0.41)	**1.00**	(0.00)	17.82	(0.53)
	0.2	5.97	(0.18)	**1.00**	(0.00)	5.52	(0.16)
	0.25	2.56	(0.06)	**1.00**	(0.00)	1.90	(0.04)
Bi-lobe	0.1	80.06	(2.53)	**8.19**	(0.22)	87.96	(2.88)
	0.2	40.56	(1.21)	**1.22**	(0.02)	47.54	(1.56)
	0.3	19.06	(0.55)	**1.00**	(0.00)	20.53	(0.61)
	0.4	10.31	(0.31)	**1.00**	(0.00)	6.61	(0.19)
Tri-lobe	0.1	90.62	(2.94)	**21.71**	(0.67)	96.49	(3.02)
	0.2	59.74	(1.86)	**2.73**	(0.07)	75.45	(2.50)
	0.3	35.02	(1.11)	**1.15**	(0.01)	41.79	(1.30)
	0.4	22.05	(0.70)	**1.00**	(0.00)	22.42	(0.67)

ARL about equal to 100 in the in-control case for each method.
Bold fonts highlight outperforming ARL values.

the large-scale model is characterized by fixed effects (i.e., productive scenarios 1, 3 and 5). The second type will refer to the remaining set of scenarios 2, 4 and 6 (where the large-scale component has random effects, too).

Let's start discussing results obtained for scenarios with fixed effects in the large-scale model (i.e., productive scenarios 1, 3 and 5). Table 6 shows results obtained in these scenarios considering the first type of out-of-control (i.e., characterized by a change of components which were already present in the the baseline model). In this table, it can observed that the regression-based approach outperforms both the PCA-based approach and the location control chart, even if in a few cases, especially when an high severity of the out-of-control condition is considered, the

regression-based and the PCA-based approaches can have comparable performance. It should be noted that for these out-of-control conditions (half-frequency, bi-lobe and tri-lobe), the majority of alarms released by the regression-based approach are produced by the T^2 control chart on the vector of fitted parameters, while the PCA-based approach consists of the Q control chart only.

With reference to the same productive scenarios (1, 3 and 5), when the out-of-control influences harmonics not in the baseline model (four-lobe with both fixed and random phase), it can be observed from results in Table 7 that the regression-based approach presents a reduced power of detection. As a matter of fact, the T^2 control chart does not signal many changes in the regression parameters, while the majority of alarms are produced by the residual control chart. However, this latter control chart appears to be not suitable in signalling changes in harmonic's amplitudes or phases. As a consequence, in these cases the regression-based approach present inferior performance when compared to the PCA-based approach (with the exception of productive scenarios 1 when a large shift is considered). From results reported in Table 7, it is also worth noting that the location control chart also did well in signalling four-lobe out-of-controls, especially for the productive scenario 5, where in many cases it presents comparable performance to those produced by the PCA-based approach. Finally, in the case of increased variance of the error term (i.e., the last type of out-of-control case), Table 7 shows that PCA-based approach outperforms the competing methods in almost all the productive scenarios, with the exception of scenario 5. As a matter of fact, while the performance of both the location and the regression-based control charts in signaling increased variance of the error term are not influenced by the nature of such a term (iid for scenario 1, or correlated with either fixed or random coefficients for scenarios 3 and 5 respectively), the performance of Q chart of the PCA-based approach are reducing in the case of spatially correlated errors, especially for random effects in the small-scale component of the model (productive scenario 5).

With reference to the remaining set of scenarios (characterized by a large-scale model with random effects, namely production scenarios 2, 4 and 6), the PCA-based approach can take full advantage of both the T^2 control chart based on the first $m = 4$ PCs and of the Q control chart on the residuals. Considering the first type of out-of-control models (where the change is affecting a regressor that was originally present in the baseline model, i.e., half-frequency, bi-lobe and tri-lobe), Table 8 shows that the PCA-based approach presents in many cases outperforming performance when compared to the regression-based approach, particularly for productive scenarios with correlated error term (scenarios 4 and 6) while in the case of iid errors (scenario 2) performance of the two parametric approaches are almost equivalent. As a matter of fact, the regression-based approach is not doing well in signaling these changes in the amplitude of the baseline-harmonics because of the the extra variability (random effects) characterizing the estimated parameters. In fact, this additional variability causes a reduced detection power of T^2 control chart in the regression-based approach.

Furthermore, from Table 8 it can be also observed that the location control chart presents comparable performance to the ones achieved with PCA-based approach. In a few cases, especially for productive scenarios with correlated errors (scenarios 4 and 6), surprisingly the location control chart outperform both the regression-based and the PCA-based approaches. This result further confirms that the simple location control chart can be considered a valuable alternative to parametric methods for profile monitoring.

Similarly to scenarios where fixed effects in the large-scale model are assumed, Table 9 shows that even if random effects are considered the PCA-based approach

Table 7. Phase II simulation results for productive scenarios 1, 3 and 5 (fixed effects in the large-scale component of the model). Out-of-controls influencing one harmonic not in the baseline model and error term. Actual ARLs and corresponding standard deviations within parenthesis.

	severity	LOC CC		REG CC		PCA CC	
		Scenario 1					
Four-lobe	0.005	89.78	(2.91)	76.50	(2.39)	**38.99**	(1.19)
(fixed phase)	0.01	54.55	(1.72)	10.68	(0.33)	**5.41**	(0.16)
	0.015	30.54	(0.94)	1.65	(0.03)	**1.28**	(0.02)
	0.02	17.35	(0.55)	1.01	(0.00)	**1.00**	(0.00)
Four-lobe	0.005	81.53	(2.58)	76.67	(2.42)	**40.41**	(1.24)
(random phase)	0.01	56.15	(1.71)	10.78	(0.32)	**5.14**	(0.14)
	0.015	33.00	(1.03)	1.57	(0.03)	**1.30**	(0.02)
	0.02	16.13	(0.51)	**1.01**	(0.00)	**1.00**	(0.00)
Error term	1.01	79.75	(2.70)	68.51	(2.09)	**38.28**	(1.20)
	1.02	70.42	(2.27)	35.12	(1.10)	**16.49**	(0.50)
	1.03	59.14	(1.82)	17.65	(0.57)	**8.44**	(0.24)
	1.04	49.06	(1.55)	9.26	(0.27)	**4.85**	(0.13)
		Scenario 3					
Four-lobe	0.005	70.33	(2.25)	84.02	(2.71)	**38.03**	(1.09)
(fixed phase)	0.01	35.22	(1.07)	35.52	(1.17)	**4.90**	(0.14)
	0.015	16.04	(0.48)	6.12	(0.18)	**1.41**	(0.02)
	0.02	8.66	(0.26)	1.56	(0.03)	**1.01**	(0.00)
Four-lobe	0.005	69.44	(2.05)	84.68	(2.78)	**35.66**	(1.18)
(random phase)	0.01	34.23	(1.09)	38.80	(1.19)	**5.09**	(0.14)
	0.015	16.99	(0.52)	5.94	(0.17)	**1.39**	(0.02)
	0.02	8.02	(0.24)	1.59	(0.03)	**1.02**	(0.00)
Error term	1.01	75.33	(2.40)	66.42	(2.04)	**49.60**	(1.55)
	1.02	63.42	(1.77)	35.59	(1.07)	**26.68**	(0.81)
	1.03	50.47	(1.48)	**16.64**	(0.53)	14.84	(0.46)
	1.04	39.63	(1.22)	**8.85**	(0.26)	8.70	(0.26)
		Scenario 5					
Four-lobe	0.005	**74.33**	(2.38)	99.25	(3.13)	**80.03**	(2.56)
(fixed phase)	0.01	**36.30**	(1.16)	83.13	(2.56)	**43.31**	(1.30)
	0.015	**16.98**	(0.51)	52.06	(1.58)	**14.83**	(0.44)
	0.02	9.00	(0.27)	32.08	(1.02)	**5.02**	(0.14)
Four-lobe	0.005	**75.40**	(2.53)	97.02	(2.97)	**83.22**	(2.59)
(random phase)	0.01	**36.53**	(1.12)	80.87	(2.66)	**42.83**	(1.32)
	0.015	**17.11**	(0.52)	54.15	(1.71)	**16.10**	(0.46)
	0.02	8.25	(0.24)	34.22	(1.04)	**4.88**	(0.14)
Error term	1.01	**85.19**	(2.73)	**74.52**	(2.37)	**84.64**	(2.65)
	1.02	65.27	(2.00)	**38.21**	(1.15)	63.55	(2.04)
	1.03	51.64	(1.62)	**18.35**	(0.54)	50.82	(1.60)
	1.04	41.72	(1.33)	**9.55**	(0.29)	40.60	(1.28)

ARL about equal to 100 in the in-control case for each method.
Bold fonts highlight outperforming ARL values.

outperforms the competing methods in signaling that an additional component (in this case a four-lobe harmonic) which was not present in the baseline model is introduced. However, in a few cases, especially in the productive scenarios 2 and 6, the competing methods present also comparable performance. Finally, with reference to the third type of out-of-control, i.e., the increase of the error-term variance, Table 9 shows that the PCA-based approach appears the outperforming method only in the case of iid error term (productive scenario 2) while in the other two scenarios (4 and 6), i.e., when the error terms present a spatial correlation, the regression-based approach presents the best performance.

Table 8.   Phase II simulation results for productive scenarios 2, 4 and 6 (random effects in the large-scale component of the model). Out-of-controls influencing one harmonics already modelled in the baseline model. Actual ARLs and corresponding standard deviations within parenthesis.

	severity	LOC CC		REG CC		PCA CC	
		Scenario 2					
Half Frequency	0.1	**75.84**	(2.54)	**79.36**	(2.52)	**67.18**	(2.09)
	0.15	**53.74**	(1.62)	**55.83**	(1.70)	**48.66**	(1.44)
	0.2	37.98	(1.13)	36.69	(1.16)	**29.46**	(0.92)
	0.25	25.91	(0.79)	21.48	(0.69)	**17.70**	(0.55)
Bi-lobe	0.1	**62.98**	(1.93)	**64.62**	(1.99)	**63.33**	(2.03)
	0.2	**38.32**	(1.23)	**36.78**	(1.17)	**37.20**	(1.19)
	0.3	**24.02**	(0.72)	20.89	(0.61)	**21.12**	(0.64)
	0.4	15.34	(0.49)	12.75	(0.41)	**12.49**	(0.38)
Tri-lobe	0.1	**70.95**	(2.24)	**65.62**	(2.03)	**59.79**	(1.92)
	0.2	46.43	(1.42)	**38.16**	(1.19)	**35.44**	(1.13)
	0.3	32.09	(1.00)	**20.36**	(0.61)	**19.77**	(0.63)
	0.4	21.64	(0.68)	**12.98**	(0.39)	**12.56**	(0.38)
		Scenario 4					
Half Frequency	0.1	**78.93**	(2.43)	94.10	(2.92)	**77.36**	(2.37)
	0.15	**52.08**	(1.67)	77.19	(2.46)	**49.00**	(1.61)
	0.2	**35.70**	(1.15)	57.37	(1.83)	**30.25**	(1.01)
	0.25	23.19	(0.73)	36.60	(1.14)	**17.68**	(0.56)
Bi-lobe	0.1	**53.53**	(1.68)	76.27	(2.36)	68.35	(2.16)
	0.2	**32.95**	(1.04)	48.44	(1.50)	39.88	(1.22)
	0.3	**22.32**	(0.68)	29.95	(0.92)	22.39	(0.68)
	0.4	**13.92**	(0.42)	16.42	(0.50)	14.52	(0.45)
Tri-lobe	0.1	**68.28**	(2.09)	**76.96**	(2.44)	**69.07**	(2.13)
	0.2	41.21	(1.31)	48.96	(1.53)	**39.19**	(1.22)
	0.3	29.19	(0.93)	28.15	(0.90)	**22.32**	(0.68)
	0.4	17.99	(0.56)	17.22	(0.53)	**13.23**	(0.40)
		Scenario 6					
Half Frequency	0.1	**73.26**	(2.18)	93.18	(2.83)	**80.04**	(2.56)
	0.15	**50.53**	(1.57)	78.68	(2.63)	61.07	(1.90)
	0.2	**35.61**	(1.16)	68.49	(2.16)	44.43	(1.37)
	0.25	**22.49**	(0.72)	48.50	(1.56)	29.05	(0.89)
Bi-lobe	0.1	**64.08**	(1.92)	**76.41**	(2.42)	**64.29**	(1.89)
	0.2	**36.07**	(1.16)	47.48	(1.60)	**38.03**	(1.22)
	0.3	**24.01**	(0.77)	29.76	(0.92)	**21.21**	(0.67)
	0.4	**14.98**	(0.45)	16.93	(0.53)	**12.61**	(0.38)
Tri-lobe	0.1	**70.44**	(2.16)	**72.51**	(2.27)	**70.08**	(2.19)
	0.2	**43.66**	(1.44)	47.72	(1.50)	**37.55**	(1.17)
	0.3	30.33	(0.94)	27.57	(0.85)	**21.70**	(0.67)
	0.4	19.43	(0.60)	17.39	(0.54)	**12.60**	(0.40)

ARL about equal to 100 in the in-control case for each method.
Bold fonts highlight outperforming ARL values.

5.3   *Overall performance measure*

Given that no specific approach has shown to be preferable to the competing ones in all the simulated case studied (Table 6-Table 9), we decided to compute an overall performance index by averaging all the performance achieved in each productive scenario, despite of the specific out-of-control under study. As a matter of fact, the two factors which are defining the simulated case studies (the type of scenario and the type of out-of-control model) presents a main difference: the type of scenario should in principle be known by the the analyst while the type of out-of-control that is going to affect profiles in the future should not be known in advance. In fact, scenarios depend on the profile model characteristics (i.e., random or fixed effects,

Table 9. Phase II simulation results for productive scenarios 2, 4 and 6 (random effects in the large-scale component of the model). Out-of-controls influencing one harmonic not in the baseline model and error term. Actual ARLs and corresponding standard deviations within parenthesis.

	severity	LOC CC		REG CC		PCA CC	
Scenario 2							
Four-lobe (fixed phase)	0.005	94.17	(2.95)	75.65	(2.29)	**47.98**	(1.54)
	0.01	72.37	(2.26)	10.31	(0.31)	**7.27**	(0.21)
	0.015	50.96	(1.66)	1.57	(0.03)	**1.41**	(0.02)
	0.02	36.08	(1.19)	**1.01**	(0.00)	**1.01**	(0.00)
Four-lobe (random phase)	0.005	89.76	(2.71)	71.29	(2.21)	**50.01**	(1.63)
	0.01	75.49	(2.29)	9.73	(0.30)	**7.07**	(0.21)
	0.015	53.03	(1.69)	**1.58**	(0.03)	**1.42**	(0.02)
	0.02	37.03	(1.17)	**1.01**	(0.00)	**1.01**	(0.00)
Error term	1.01	85.57	(2.87)	69.28	(2.08)	**49.93**	(1.60)
	1.02	85.47	(2.49)	36.33	(1.12)	**21.26**	(0.63)
	1.03	75.11	(2.54)	16.09	(0.49)	**10.90**	(0.31)
	1.04	71.96	(2.28)	8.41	(0.24)	**5.80**	(0.16)
Scenario 4							
Four-lobe (fixed phase)	0.005	82.84	(2.67)	97.27	(3.07)	**50.62**	(1.57)
	0.01	62.98	(1.88)	52.01	(1.54)	**6.67**	(0.19)
	0.015	46.73	(1.52)	9.67	(0.26)	**1.53**	(0.03)
	0.02	29.93	(0.95)	2.01	(0.04)	**1.02**	(0.00)
Four-lobe (random phase)	0.005	85.04	(2.68)	96.51	(3.02)	**50.52**	(1.55)
	0.01	63.38	(2.08)	53.68	(1.61)	**6.89**	(0.20)
	0.015	43.92	(1.39)	9.75	(0.29)	**1.56**	(0.03)
	0.02	29.17	(0.88)	2.03	(0.05)	**1.02**	(0.00)
Error term	1.01	90.45	(2.80)	**78.04**	(2.42)	**65.50**	(2.01)
	1.02	80.35	(2.49)	**37.38**	(1.18)	**38.16**	(1.16)
	1.03	71.94	(2.30)	**17.63**	(0.54)	21.90	(0.70)
	1.04	68.64	(2.07)	**9.36**	(0.26)	13.11	(0.41)
Scenario 6							
Four-lobe (fixed phase)	0.005	**94.32**	(3.06)	**98.13**	(3.13)	**93.35**	(2.96)
	0.01	74.52	(2.31)	95.74	(3.07)	**53.47**	(1.73)
	0.015	50.01	(1.57)	59.25	(1.82)	**20.42**	(0.60)
	0.02	33.61	(1.07)	39.05	(1.18)	**6.84**	(0.20)
Four-lobe (random phase)	0.005	**93.75**	(3.11)	**98.73**	(3.25)	82.94	(2.61)
	0.01	68.78	(2.20)	93.34	(2.72)	**51.94**	(1.57)
	0.015	46.98	(1.49)	58.61	(1.83)	**22.00**	(0.66)
	0.02	33.65	(1.04)	35.44	(1.09)	**6.73**	(0.20)
Error term	1.01	99.60	(3.29)	**70.81**	(2.25)	**82.80**	(2.56)
	1.02	92.75	(2.86)	**37.58**	(1.13)	75.28	(2.39)
	1.03	82.34	(2.56)	**17.88**	(0.52)	60.27	(1.77)
	1.04	76.31	(2.42)	**9.17**	(0.28)	55.34	(1.80)

ARL about equal to 100 in the in-control case for each method.
Bold fonts highlight outperforming ARL values.

iid or correlated noise terms, etc.) which should be investigated in Phase I control charting. On the contrary, the type of out-of-control is in general unknown, and the overall results that we are going to analyse represent the actual performance indicator when all the different types of out-of-controls are equally likely to affect future profiles.

Figure 1 through Figure 3 graphically depict the 99% Bonferroni's confidence intervals of the overall ARLs presented by the three competing approaches in each production scenarios considered in our study. It can be observed that the PCA-based approach should be preferred in almost all the production scenarios with the only exception of scenario 5 in which the regression-based approach outperforms both the location and PCA-based control chart. It is also worth noting that in
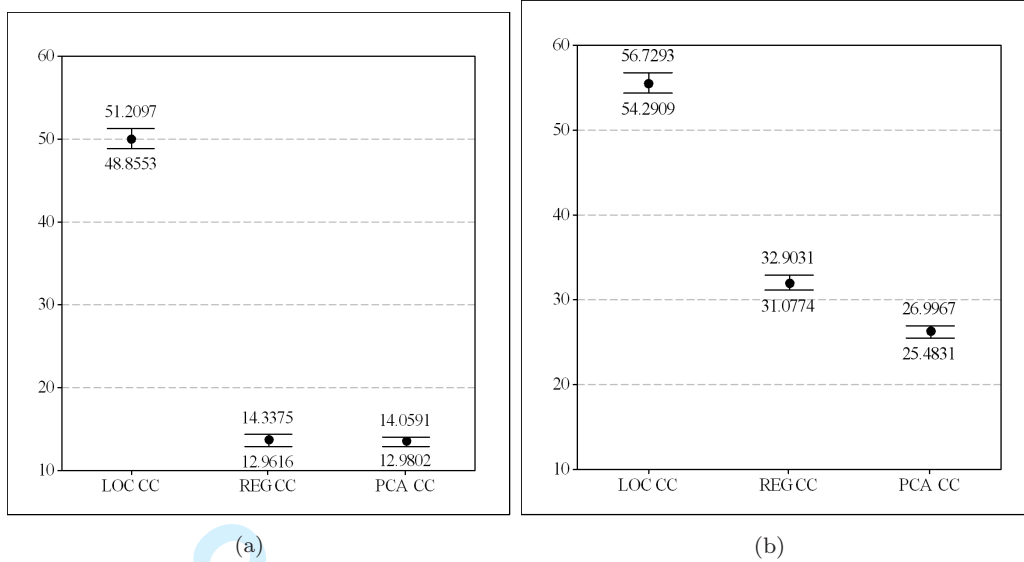
Figure 1.   99% Bonferroni's confidence intervals of the overall ARL for location (LOC CC), regression-based (REG CC) and PCA-based (PCA-CC) control charts. (a) Productive scenario 1. (b) Productive scenario 2.
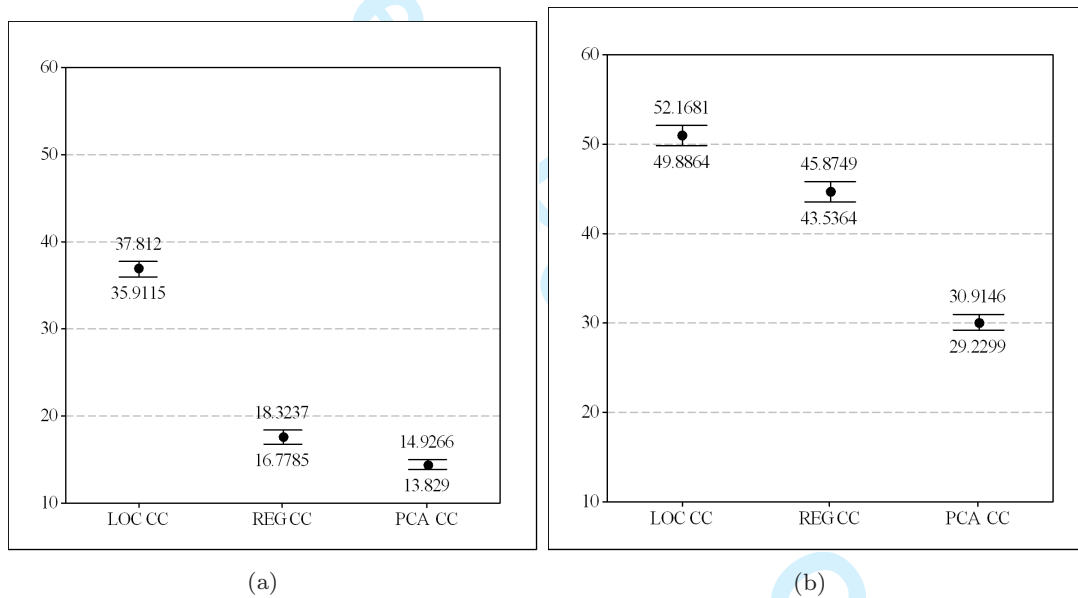

Figure 2.   99% Bonferroni's confidence intervals of the overall ARL for location (LOC CC), regression-based (REG CC) and PCA-based (PCA-CC) control charts. (a) Productive scenario 3. (b) Productive scenario 4.

the scenario 1, the regression-based and PCA-based approaches have performance substantially similar. As expected the location control chart does not present an overall performance better than those which characterize the regression-based and the PCA-based approach, even if in the production scenario 5 the location control chart outperform the PCA-based approach, while in the scenario 6 the performance of the simple location control chart is substantially similar to those of the regression-based approach. Finally, since scenario 6 is the unperturbed one, i.e., the one representing the actual case study, we can conclude that the location control control can be considered a valuable option for profile monitoring of the roundness profiles.
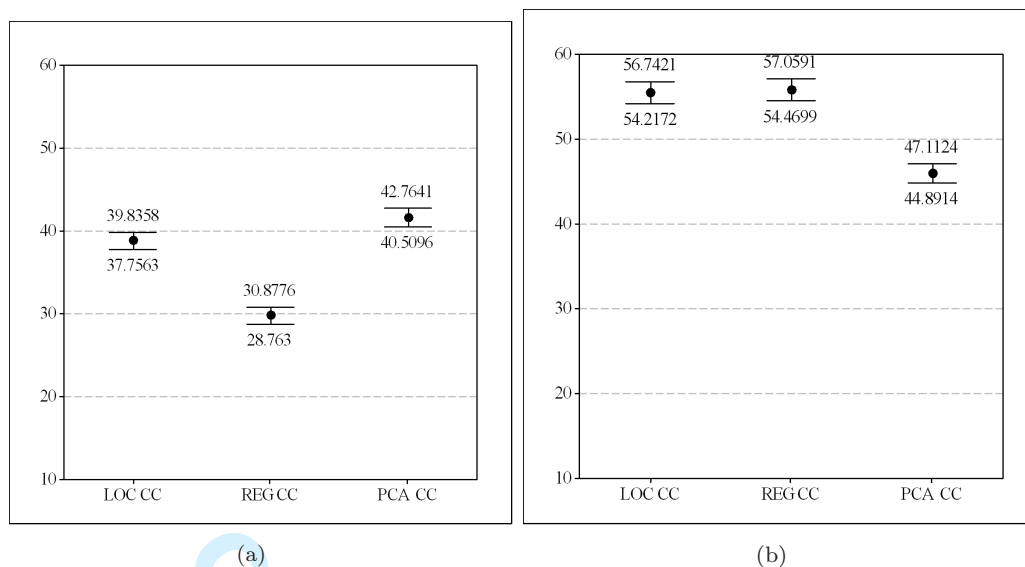
Figure 3. 99% Bonferroni's confidence intervals of the overall ARL for location (LOC CC), regression-based (REG CC) and PCA-based (PCA-CC) control charts. (a) Productive scenario 5. (b) Productive scenario 6.

6. Conclusions

Profile monitoring is a relatively new technique in quality control used when the product or process quality is best represented by a profile at each time period. Much of the work in the scientific literature has focused on presenting new approaches for profile monitoring. This paper presented a comparison study between different approaches (location control chart, regression-based and PCA-based) aimed at defining possible guidelines for selecting the proper approach, depending on the type of problem at hand.

By comparing Phase I performance, which refer to the ability to obtain a pre-defined false alarm rate when no contaminated process samples are considered, we can conclude that the extra-effort required by a regression-based approach is worth. In fact, the regression-based control charts can be easily designed, since the actual false alarm rate can be achieved in practice by using an analytic computation of the control limits despite of the specific productive scenario considered and of the number of historical data available. Similar conclusions can be also drawn with reference to the PCA-based approach but only in specific productive scenarios and for large dimensions of the data set. On the contrary, the design of a location control chart can be a difficult task, since the actual false alarm rate can be very different from the one expected.

With reference to the ability in detecting out-of-control states in Phase II, the simulation results showed that no specific approach should being preferred among the three competing ones. However, the PCA-based technique is more effective in signaling a generic change in the functional data in almost all the production scenarios, while the regression-based approach shows superior performance just in a specific productive scenario.

Furthermore, though the location control chart is less effective in Phase II than the parametric approaches, performance observed by using this method are comparable to those produced by the competing ones in some cases. Given the inner simplicity of location control chart, its use in practice can be justified in these productive situations, even if attention should be paid in properly designing (Phase I) this tool.

Finally, two directions of future research can be identified. The Phase-II comparison study of this paper was based on the ideal assumption that the in-control parameters are known for each of the competing methods. However, since parameters may be unknown in actual applications, inaccuracy in estimation from moderate dataset of Phase-I data can have important effect on the Phase-II performance of the methods. Therefore, the "estimated parameters" problem, which has been recently reviewed in the literature by Jensen *et al.* (2006), has to be investigated in future research.

Moreover, only Shewhart-type approaches were compared in this paper. Indeed, the location control chart consists in applying the standard Shewhart approach at each location. In order to keep a fair comparison, the T^2 control chart was also considered for monitoring the parameters in the competing approaches. Similarly, the estimate of the residual variance was monitored with a univariate control chart. Large shifts and oscillations in the profile data are quickly detected by these Shewhart-type control charts. However, they are not sensitive in detecting small sustained shifts in the profile data. Therefore, our study can be extended to focus on performance for small shifts by combining all the approaches for profile monitoring (location control chart, PCA-based and regression-based) with different control charting procedures, like exponentially weighted moving average (EWMA) or a cumulative sum (CUSUM), if small sustained shifts are of interest (Kang and Albin 2000, Kim *et al.* 2003).

Acknowledgements

References

Boeing Commercial Airplane Group, Quality Assurance Department, 1998. Advanced Quality System Tools, AQS D1-9000-1. Available from: www.boeing.com/companyoffices/doingbiz/supplier/d1-9000-1.pdf. [Accessed June 2008].

Cho, N. W. and Tu, J. F., 2001. Roundness Modeling of Machined Parts for Tolerance Analysis. *Precision Engineering*, 25, 35–47.

Colosimo B. M. and Pacella M., 2007. On the Use of Principal Component Analysis to Identify Systematic Patterns in Roundness Profiles. *Quality and Reliability Engineering International*, 23(6), 707–725.

Colosimo B. M., Pacella M. and Semeraro Q., 2008. Statistical Process Control for Geometric Specifications: On the Monitoring of Roundness Profiles. *Journal of Quality Technology*, 40(1), 1–18.

Cressie, N. A. C., 1993. *Statistics for Spatial Data, Revised Edition*. John Wiley & Sons, New York, NY.

Ding, Y., Zeng L. and Zhou S., 2006. Phase I Analysis for Monitoring Nonlinear Profiles in Manufacturing Processes. *Journal of Quality Technology*, 38(3), 199–216.

Gupta S., Montgomery, D. C. and Woodall W. H., 2006. Performance Evaluation of Two Methods for Online Monitoring of Linear Calibration Profiles. *International Journal of Production Research 44(10)*, 1927–1942.

ISO/TC 213 - Geometrical Product Specification (GPS) Roundness (ISO/TS

12181:2003) and Cylindricity (ISO/TS 12180:2003). Avaiable from: http://isotc213.ds.dk/. [Accessed June 2008].

Jackson J. E., 2003. *A user's guide to principal components*, New York: Wiley.

Jensen, W. A., Jones, L. A., Champ, C. W. and Woodall, W. H., 2006. Effects of parameter estimation on control chart properties: a literature review. *Journal of Quality Technology*, 38(4), 349–364.

Jensen W. A., Birch, J. B. and Woodall, W. H., 2008. Monitoring Correlation Within Linear Profiles Using Mixed Models. *Journal of Quality Technology*, 40(2), 167–183.

Jin, J. and Shi, J., 2001. Automatic Feature Extraction of Waveform Signals for In-Process Diagnostic Performance Improvement, *Journal of Intelligent Manufacturing* 12, 257–268.

Jin, J. and Shi, J., 1999. Feature-Preserving Data Compression of Stamping Tonnage Information Using Wavelets. *Technometrics* 41, 327–339.

Jolliffe I. T., 2002. *Principal Component Analysis*, 2nd edition, Springer Series in Statistics, New-York: Springer-Verlag.

Kang, L. and Albin, S. L., 2000. On-Line Monitoring When the Process Yields a Linear Profile. *Journal of Quality Technology* 32, 418–426.

Kim, K., Mahmoud, M.A. and Woodall, W.H., 2003. On the Monitoring of Linear Profiles. *Journal of Quality Technology* 35, 317–328.

LeSage, J.P., 1999. "The Theory and Practice of Spatial Econometrics" Department of Economics University of Toledo. Available from: www.spatial-econometrics.com. [Accessed June 2008].

Mahmoud, M. A., Woodall, W. H., 2004. Phase I Analysis of Linear Profiles with Calibration Applications. *Technometrics*, 46, 377–391.

Montgomery, D. C., 2004. *Introduction to Statistical Quality Control*, New York: Wiley.

Nomikos P., MacGregor J. F., 1995. Multivariate SPC Charts for Monitoring Batch Processes. *Technometrics*, 37, 41–59.

Ramsay, J. O. and Silverman, B. W., 2005. *Functional data Analysis*, 2nd edition, Springer Series in Statistics, New -York: Springer-Verlag.

Simes, R. J., 1986. An Improved Bonferroni Procedure for Multiple Tests of Significance. *Biometrika*, 73(3), 751–754.

Sullivan, J. H. and Woodall, W. H., 1996. A Comparison of Multivariate Quality Control Charts for Individual Observations. *Journal of Quality Technology*, 28, 398–408.

Vargas N. J. A., 2003. Robust Estimation in Multivariate Control Charts for Individual Observations. *Journal of Quality Technology*, 35, 367–376.

Williams, J. D., Woodall, W.H., Birch, J. B., Sullivan, J. H., 2006. Distribution of Hotellings T^2 Statistic Based on the Successive Differences Estimator. *Journal of Quality Technology*, 38, 217–229.

Woodall, W. H., Spitzner, D. J., Montgomery, D. C. and Gupta, S., 2004. Using Control Charts to Monitor Process and Product Quality Profiles. *Journal of Quality Technology*, 36(3), 309–320.