From profile to surface monitoring: SPC for cylindrical surfaces via Gaussian Processes

BIANCA M. COLOSIMO, PAOLO CICORELLA

Politecnico di Milano, Dipartimento di Meccanica, via La Masa 1, 20156 Milano, Italy

MASSIMO PACELLA and MARZIA BLACO

Università del Salento, Dipartimento di Ingegneria dell'Innovazione,

Piazza Tancredi 7, 73100 Lecce, Italy

Quality of machined products is often related to the shapes of surfaces that are constrained by geometric tolerances. In this case, statistical quality monitoring should be used to quickly detect unwanted deviations from the nominal pattern. The majority of the literature has focused on statistical profile monitoring, while there is little research on surface monitoring. This paper faces the challenging task of moving from profile to surface monitoring. To this aim, different parametric approaches and control charting procedures are presented and compared with reference to a real case study dealing with cylindrical surfaces obtained by lathe-turning. In particular, a novel method presented in this paper consists of modeling the manufactured surface via Gaussian processes models and monitor the deviations of the actual surface from the target pattern estimated in Phase I. Regardless of the specific case study in this paper, the proposed approach is quite general and can be easily extended to deal with different kinds of surface or profiles.

KEYWORDS: GP model, kriging, profile monitoring, spatial statistics, geometric specifications, cylindrical surfaces.

Introduction

In the last years, profile monitoring has become an important new area of statistical process control (SPC). Starting from the simplest case of linear profiles, (Kang and Albin (2000), Kim *et al.* (2003), Mahmoud and Woodall (2004)) literature on profile monitoring has now covered different parametric and nonparametric approaches and different application domains. Woodall *et al*. (2004) and Woodall (2007) presented reviews of the literature on this topic while the recent book (Noorossana *et al*. (2012)) on profile monitoring presents a brief overview of different approaches and applications of profile monitoring.

Among the different approaches for profile monitoring, one stream of research focused on monitoring 2D or 3D profiles which are related to physical shapes of the workpiece constrained by geometric tolerances (Colosimo and Pacella (2007, 2010), Colosimo *et al*. (2008), Colosimo and Senin (2011), Colosimo and Pacella (2011), Del Castillo *et al*. (2013)).

Moving from profiles to surfaces, Colosimo *et al*. (2010) showed how surface monitoring can be considered as a generalization of profile monitoring. Specifically, they showed that cylindrical surfaces of lathe-turned items can be modeled using linear regression with spatially correlated error terms. By combining this model with control charting, the authors showed how different types of out of control of the manufactured surfaces can be detected.

The present paper starts from the approach proposed in Colosimo *et al*. (2010) for surface monitoring and presents a novel method for surface quality monitoring. In this approach, Gaussian processes (GPs) (Matheron (1963), Cressie (1993)) are used as main tool for surface modeling. The use of GPs instead of regression allows one to simplify the modeling step, since selection of the appropriate regressor variables is not required anymore. As a drawback, the standard solution of monitoring model's coefficients to detect change of the profile/surface pattern is not working when GP replies regression as modeling tool. In this paper, we will show that this approach can be ineffective and different solutions

have to be found. In particular, we will propose a method based on monitoring the GP-predicted deviations of the surface from the in-control pattern estimated in Phase I. This solution is similar to the approach recently proposed by Grimshaw *et al*. (2013) to monitor the mean of a process via Gaussian processes or to the approach recently proposed by Zhang *et al*. (2013) in the context of linear profile monitoring. Similarly, some connections can be also found with approaches recently proposed for nonparametric profile monitoring (Qiu *et al.* (2010), Chipman *et al.* (2010)). We will discuss and evaluate performance of these other methods in the last section of the performance comparison study.

It is worth noting that all these approaches share the same idea of our proposed procedure, which is using a statistic that summarizes discrepancies between profile/surface data predicted via GP (or other nonparametric approaches) and their corresponding "in-control" values. However, none of these methods discuss on how to select the number and locations of these design points. Our paper provides some guidelines on possible strategies to locate these points.

The remainder of the paper is organized as follows. In the next section, the real case study and the parametric model for cylindrical surface modeling will be introduced. In the same section, the in-control and out-of-control simulation models will be presented. Then, the GP-based approaches for monitoring cylindrical surfaces will be described, while the following section will report results of a simulation study used to compare performances of all the existing approaches for surface monitoring. Conclusions and final remarks will be eventually provided in Section 0.

The case study and the simulation model

The case study was originally presented in Colosimo *et al*. (2008) to show how to perform profile monitoring on roundness profiles. The data used in that paper refer to a circular profile obtained by an ideal cross section of cylindrical components obtained by lathe-turning. In this paper, we will use all the data measured on the cylindrical surfaces to move from profile to surface monitoring.

Data refer to 100 rolled C20 carbon steel bars which were machined by lathe-turning (cutting speed 163 m/min and feed rate 0.2 mm/rev). The initial diameter of the bar was 30 mm and it was reduced to a final diameter of 26 mm through two cutting steps, 1 mm depth each.

Each machined surface was measured by a Coordinate Measuring Machine (CMM) on a given cylindrical grid of $N_C \times N_G$ equally spaced positions, N_C measurements along the axial and N_G along the angular directions, respectively. N_C and N_G correspond to the number of circumferences and generatrices sampled on the cylindrical surfaces. In practice, the CMM touch trigger approached the nominal position of each grid point, and then stored the actual coordinates of the probe when it touched the surface. Surface data in this case study were measured on a 61×68-point cylindrical grid, i.e., $N_C = 61$ and $N_G = 68$.

For each item, final surface data were computed as deviations of the measured radiuses from the substitute geometry, which is the ideal pattern of the form error under study (i.e., a cylinder). This substitute geometry was computed using a least-square approach, i.e., by minimizing the sum-of-squared distances between the observed radiuses and the ideal feature. By subtracting the radius of the substitute geometry, the final set of measurements consists of a set of radial deviations from a perfect cylinder, measured at each position of the cylindrical grid.

Cylindricity requirement measures the difference between the actual feature and a perfect cylinder, regardless of the position of the actual object or its orientation in space (ISO (2012)). Therefore, preliminary steps of registration were implemented for each cylindrical surface. Registration by translation is performed by centering each cylindrical surface on the origin of the coordinate system. This step is accomplished by subtracting the mean value to the surface data expressed in Cartesian coordinates. After this step, all of the objects share the same axis. Hence, only registration by rotation around the vertical axis is still needed. This final registration step is based on maximizing the cross-

correlation between roundness profiles obtained by cross-sectioning the cylinders, as detailed in Colosimo and Pacella (2007).

Notation and parametric model for cylindrical surfaces

In order to measure the actual surface, a common sampling strategy consists of considering a fixed and regular grid on the ideal cylinder (substitute geometry) and let the CMM trigger approach radially each point of the grid to register a signal as the trigger touches the real surface. Let $\mathbf{s}_i = (\rho, \theta_i, z_i)$, denote the location at which the actual radius is measured, where ρ is the nominal radius, which is equal for all the points on the cylindrical grid, while θ_i and z_i are the angular and vertical coordinates associated to the i-th point (Figure 1). Let $r_h(\mathbf{s}_i) = r_h(\rho, \theta_i, z_i)$ represent the difference between the actual radius and the nominal radius observed at location $\mathbf{s}_i$, $(i = 1, \ldots, N)$ on the h-th surface. Note that $r_h(\mathbf{s}_i) = r_h(\rho, \theta_i, z_i)$ represents the response variable of interest as a function of other location variables, similarly to what is usually done for representing 2.5D surfaces, where the vertical deviation of the real surface from its ideal pattern is commonly modeled as a function of the other two coordinates (i.e., $z_h(x_i, y_i)$). Thanks to this clear connection, approaches presented in this paper for cylindrical surfaces monitoring can be easily extended to 2.5D surface monitoring.

Usually the same sampling strategy is assumed for all the surfaces, defining an ideal grid of $N = N_C \times N_G$ equally spaced locations, where N_C and N_G are the number of vertical and angular levels at which the actual radii are measured. In this case, angular and vertical locations at which data are observed are selected among fixed levels, i.e., $\theta_i \in [\theta_1, \theta_2, \ldots, \theta_G]$ and $z_i \in [z_1, z_2, \ldots, z_C]$.

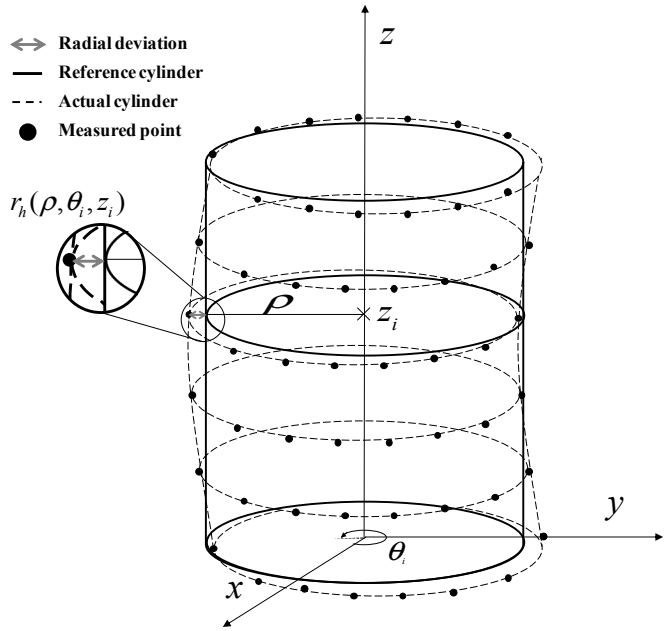


Figure 1: An example of deviation of a real cylindrical surface from the nominal one.

Following Henke *et al.* (1999) and Colosimo *et al*. (2010, 2011), we will assume that typical deviations of a machined cylinder from its ideal feature can be described as a combination of a *large*- and a *small-scale* components of a SARX model (Cressie (1993)), described in the first and second rows of the following equation, respectively:

$$\begin{aligned} \mathbf{r}_h &= \mathbf{X}\boldsymbol{\beta}_h + \mathbf{u}_h \\ \mathbf{u}_h &= \left(\alpha_{1h}\mathbf{W}^{(1)} + \alpha_{2h}\mathbf{W}^{(2)}\right)\mathbf{u}_h + \boldsymbol{\varepsilon}_h. \end{aligned} \tag{1}$$

In the first row $\mathbf{r}_h = \left[r_h(\mathbf{s}_1) \quad \ldots \quad r_h(\mathbf{s}_i) \quad \ldots \quad r_h(\mathbf{s}_N)\right]'$ represents the $N \times 1$ vector of the radial deviation from the nominal radius observed on the h-th item, $\mathbf{X}$ is the regressor matrix, whose i-th row $\mathbf{x}_i' = \left[x_{1i} \quad \ldots \quad x_{ki} \quad \ldots \quad x_{Ki}\right]$ is a $K \times 1$ vector, $\boldsymbol{\beta}_h' = \left[\beta_{1h} \quad \ldots \quad \beta_{kh} \quad \ldots \quad \beta_{10h}\right]$ is the $K \times 1$ vector of parameters. The second row in (1) describes the small-scale component which is given by the $N \times 1$ vector of error terms $\mathbf{u}_h$. Error terms are assumed to be spatially correlated and are represented as a Spatial Autoregressive (SAR) process of order 2. $\boldsymbol{\alpha}_h' = \left[\alpha_{1h} \quad \alpha_{2h}\right]$ is the vector of coefficients of the

SAR model related to the h-th cylindrical surface. $\mathbf{W}^{(1)}$ and $\mathbf{W}^{(2)}$ are the first and the second order adjacency matrices (Cressie (1993)). As shown in Colosimo *et al*. (2010), the rook-based contiguity is suitable to represent real data and will be assumed in this paper, too. Finally, $\boldsymbol{\varepsilon}_h$ is a $N \times 1$ vector of independently and normally distributed residuals, $\boldsymbol{\varepsilon}_h \sim N\left(0, \sigma_\varepsilon^2 \mathbf{I}_N\right)$, where $\mathbf{I}_N$ is the N-dimension identity matrix.

Table 1 shows the expression of the $K = 10$ regressor functions which were shown to be significant to describe the 100 cylindrical surfaces of the case study (Colosimo *et al*. (2011)). These regressor functions combine Chebyshev polynomials to represent axial errors and sinusoidal functions to model circular errors.

In particular, sinusoidal functions represent bilobe and trilobe patterns, which can be observed along the angular direction, while Chebyshev polynomials of order 0 to 2 are used to represent linear and quadratic patterns observed along the vertical direction (Figure 2). Following Henke's model, Chebyshev polynomials $T_0\left(\xi_i\right) = 1;\ T_1\left(\xi_i\right) = 2\xi_i;\ T_2\left(\xi_i\right) = 4\xi_i^2 - 1\ T_l\left(\xi_i\right)$ are expressed as a function of a scaled z-coordinate, given by $\xi_i = 2\dfrac{z_i - z_{\min}}{z_{\max} - z_{\min}} - 1$ which ranges in [-1 , +1].

Coefficients of model in (1) can be stored in a 12-dimension parameter vector $\hat{\mathbf{c}}_h' = \left[\hat{\beta}_{1h} \quad \hat{\beta}_{2h} \quad \ldots \quad \hat{\beta}_{10h} \quad \hat{\alpha}_{1h} \quad \hat{\alpha}_{2h}\right]$ that can be used to simulate in-control profiles.

With reference to the actual data of 100 lathe-turned surfaces, measured on a regular grid of $N_C = 61$ axial locations and $N_G = 68$ angular locations, the unknown coefficients $\mathbf{c}_h$ are approximately distributed as a multinormal distribution with mean $\boldsymbol{\mu}_c$ and covariance matrix $\boldsymbol{\Sigma}_c$ $\left(\mathbf{c}_h \sim MN\left(\boldsymbol{\mu}_c, \boldsymbol{\Sigma}_c\right)\right)$.

Table 1: The regressor functions x_{ki} as function of the index k, Colosimo *et al.* (2010).

K	Order of Chebyshev polynomial	Order of the periodic (Fourier) component	x_{ki}
1	0	2	$T_0(\xi_i)\cos(2\theta_i)$
2	0	-2	$T_0(\xi_i)\sin(2\theta_i)$
3	0	3	$T_0(\xi_i)\cos(3\theta_i)$
4	0	-3	$T_0(\xi_i)\sin(3\theta_i)$
5	1	0	$T_1(\xi_i)$
6	1	2	$T_1(\xi_i)\cos(2\theta_i)$
7	1	-2	$T_1(\xi_i)\sin(2\theta_i)$
8	1	3	$T_1(\xi_i)\cos(3\theta_i)$
9	1	-3	$T_1(\xi_i)\sin(3\theta_i)$
10	2	0	$T_2(\xi_i)$

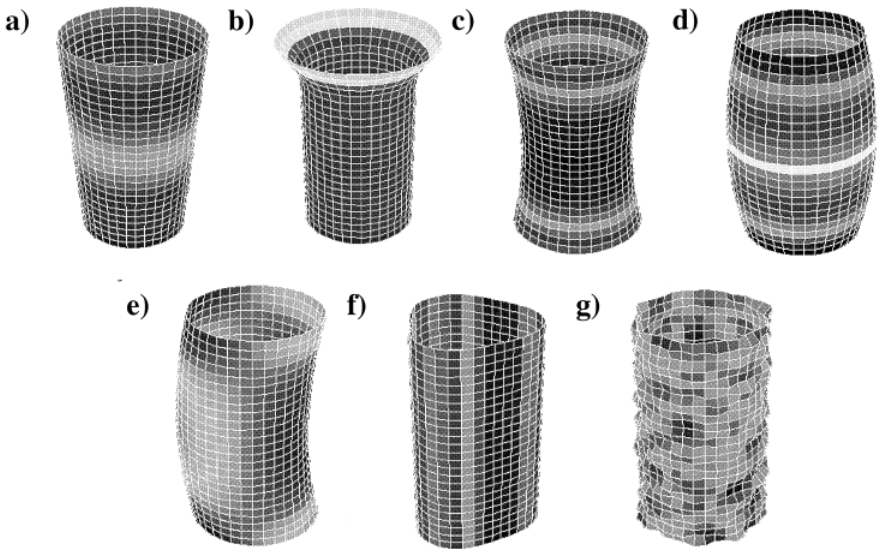


Figure 2: Typical form errors for cylindrical features: a) taper; b) bellmouth; c) hourglass; d) barrel; e) banana; f) lobing; g) random (Henke *et al.* (1999))

In-control and out-of-control surfaces simulation

In-control surfaces can be simulated by using model (1) after performing Monte Carlo simulation of the parameter vector $\mathbf{c}_h$. Starting from this in-control baseline model, out-of-control cylindrical surfaces can be simulated by adding different terms, corresponding to possible occurrences of assignable causes that can affect lathe-turning processes (Cho and Tu (2001), Henke *et al*. (1999) and Zhang *et al*. (2005)). Each condition is characterized by an amplitude parameter (δ) related to the severity of the out-of-control.

Let $r_h(\mathbf{s}_i)$ and $\tilde{r}_h(\mathbf{s}_i)$ represent the surface data under the in- and out-of-control models, respectively, five out-of-control conditions can be represented as follows:

- Quadrilobe, which can be due to a spindle-motion error that was not present in the baseline model,

$$\tilde{r}_h(\mathbf{s}_i) = r_h(\mathbf{s}_i) + \delta_{quadrilobe}\left[\cos(4\theta_i) + T_1(\xi_i)\cos(4\theta_i)\right], \qquad \delta_{quadrilobe} = 0.00185, 0.002, 0.0025. \tag{2}$$

- Trilobe, which can be due to an increase in the trilobe motion error already existing in the baseline model of the spindle or to an excessive force imposed by the clamping fixture,

$$\tilde{r}_h(\mathbf{s}_i) = r_h(\mathbf{s}_i) - \delta_{trilobe}T_1(\xi_i)\cos(3\theta_i), \qquad \delta_{trilobe} = 0.0025, 0.003, 0.0035. \tag{3}$$

- Half-frequency, which can be due to wear on one ball bearing spindle or to whirling in a hydrodynamic bearing,

$$\tilde{r}_h(\mathbf{s}_i) = r_h(\mathbf{s}_i) + \delta_{halffreq}\left[\cos(0.5\theta_i) + T_1(\xi_i)\cos(0.5\theta_i)\right], \qquad \delta_{halffreq} = 0.00075, 0.001, 0.002. \tag{4}$$

- Tapering, which can be due to an increased inflection of the workpiece axis while machining,

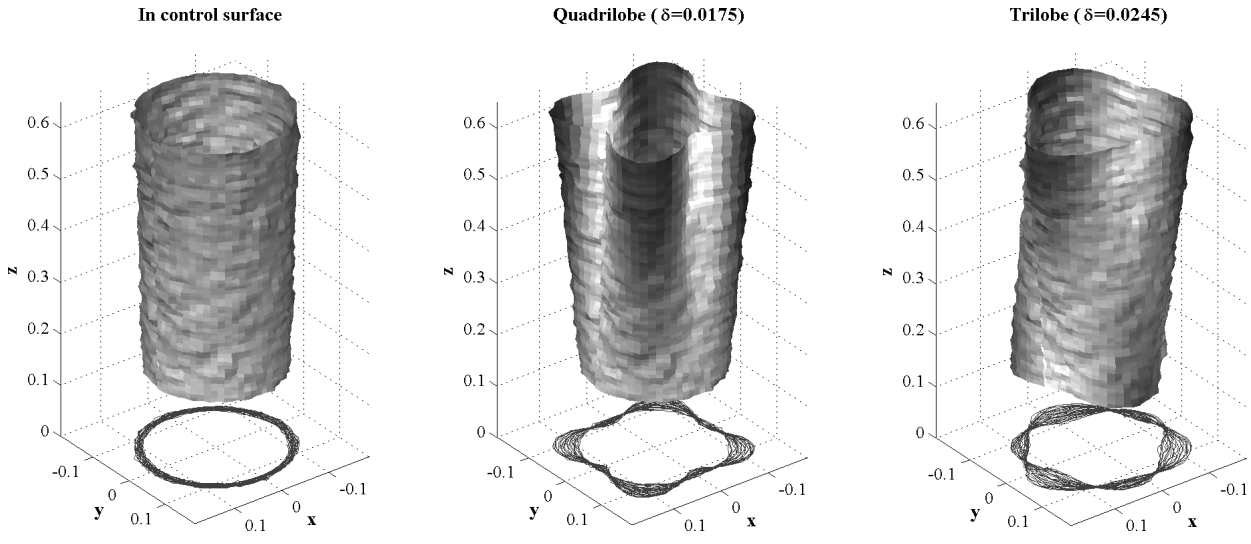
$$\tilde{r}_h(\mathbf{s}_i) = r_h(\mathbf{s}_i) + \delta_{taper}\beta_{5h}T_1(\xi_i), \qquad \delta_{taper} = 0.1, 0.15, 0.2. \tag{5}$$

- Residuals variance. This kind of out-of-control condition can be due to possible changes in the material property or machining conditions. Values $\delta_{incvar} = 5\%,\ 10\%,\ 15\%$ are considered.

In Figure 3, the in-control and out-of-control surfaces are graphically shown, using values of the factor δ to better represent the specific shape under study.

Monitoring cylindrical surfaces via Gaussian Processes

The parametric model just presented requires one to identify functional forms that can characterize the final shape (e.g., regressors shown in Table 1). Clearly, this task can be easy when we deal with simple patterns or standard shapes that have been studied in the literature for long time, as the cylindricity obtained by lathe-tuning. For new shapes and processes, selection of the proper set of regressor functions can be a cumbersome activity and represents the main barrier to application of surface monitoring in industrial practice. This is why a second solution based on GPs is presented in this paper.



a) b) c)

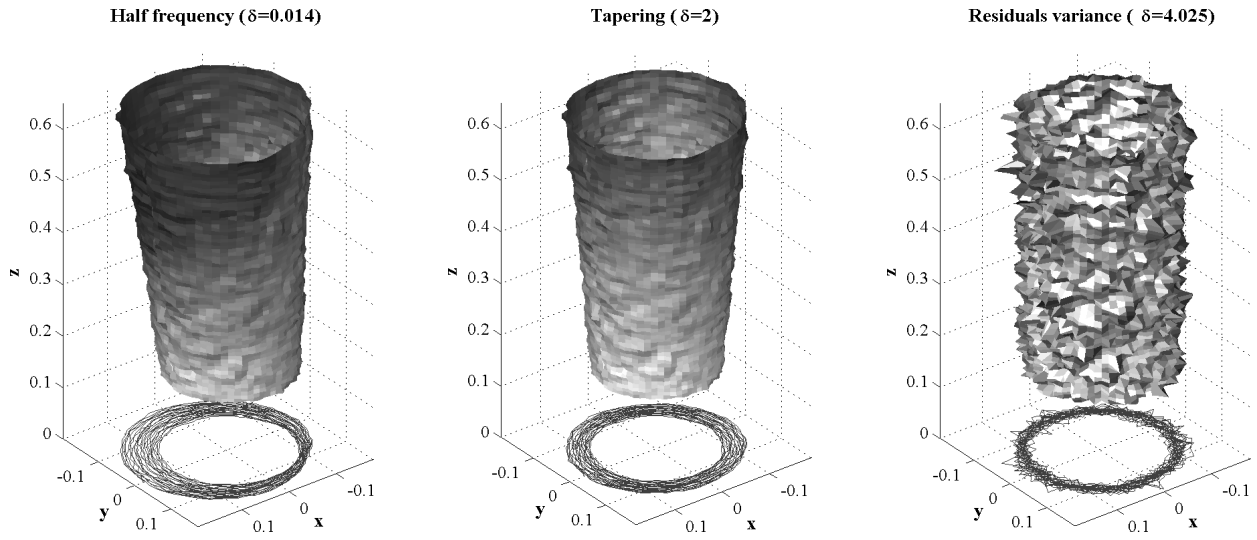


Figure 3: In control and out control simulated surfaces (grids of 61×68 points). a) In-control surface, b) Quadrilobe error, c) Trilobe error, d) Half-frequency error, e) Tapering form error, f) Residuals variance. The radial deviations related to each surface are represented on a nominal cylinder of diameter and height equal to dimensions of the real case cylinders, respectively 26mm and 50mm. The effects of out-of-control conditions are highlighted by representing the nominal cylinder on 1:100 scale. Gray scale is related to radial deviations. Dark colors represent deviations greater than zero, while bright colors represent deviation smaller than zero.

1.1 **Modeling cylindrical surfaces via GP**

Xia *et al*. (2008) and del Castillo *et al.* (2013) showed how GPs can be used for modeling surfaces. GP is a particular type of random process where all the joint probability distributions of any finite subset of values $r_h(\mathbf{s}_1) \ \ldots \ r_h(\mathbf{s}_i) \ \ldots \ r_h(\mathbf{s}_N)$ are normally distributed. In particular, Xia *et al*. (2008) showed how the deviation from the perfect shape can be modeled as a function of the spatial location via GP.

In the case of a cylindrical surface measured via CMM, the model by Xia *et al*. (2008) can be applied after an appropriate transformation from cylindrical to Cartesian coordinates is considered. In fact, if location $\mathbf{s}_i$, at which $r_h(\mathbf{s}_i)$ is observed, is expressed in cylindrical coordinates, contiguity in space is

not correctly represented. Consider for instance radial deviations observed on a cylinder at θ_i equal to 0° and 360°. They are close in space but seem far apart if the angle is taken as driver of contiguity.

In order to overcome this lack of the GP formulation for closed surfaces, we decided to pre-process data by firstly transforming cylindrical coordinates into three-dimensional Cartesian coordinates, namely $\mathbf{s}_i = (\rho, \theta_i, z_i) \rightarrow \mathbf{t}_i = (x_i, y_i, z_i)$ and then represent the radial deviation as a function of this last indicator, i.e., $r_h(\mathbf{t}_i)$. By transforming the cylindrical coordinates into the Cartesian ones we are still dealing with a three-dimensional spatial domain. Herein, we are not considering any transformation into a low-dimensional domain, as could be a map projection of the cylindrical surface to a two-dimensional plane.

Then, the following GP model was assumed:

$$r_h(\mathbf{t}_i) = \mathrm{f}_h(\mathbf{t}_i) + \varepsilon_h, \tag{6}$$

where $\varepsilon_h \sim N(0, \sigma_\varepsilon^2)$ and $\mathrm{f}_h(\mathbf{t}_i)$ is a GP model characterized by a mean $m_h(\mathbf{t}_i)$ and covariance function $k_h(\mathbf{t}_i, \mathbf{t}_i + \mathbf{d})$. Following a usual practice, we set the mean equal to a constant value:

$$m_h(\mathbf{t}_i) = a_h, \tag{7}$$

given that most of the prediction ability of the GP model is usually due to the correlation function.

The covariance function represents the relationship between data observed at location $\mathbf{t}_i$ and $\mathbf{t}_i + \mathbf{d}$, where $\mathbf{d}$ is a displacement vector. It is usually modeled using the correlation function $R(\mathbf{d}, \ell_h)$, namely $k_h(\mathbf{t}_i, \mathbf{t}_i + \mathbf{d}) = \sigma_{\eta h}^2 R(\mathbf{d}, \ell_h)$. Different kinds of correlation functions are proposed in Rasmussen and Williams (2006). A very popular choice is the isotropic squared exponential correlation function, which is very simple and requires one to estimate a small number of parameters. The squared exponential function (a.k.a. isotropic Gaussian correlation) has the form:

$$R(\mathbf{d}, \ell_h) = \exp\left(-\frac{|\mathbf{d}|^2}{2\ell_h{}^2} \right). \tag{8}$$

In (8), $|\cdot|$ denotes the Euclidean norm and ℓ_h is the scale parameter controlling how rapidly the correlation decays as the distance between two points, i.e. the Euclidean norm of the displacement vector $\mathbf{d}$, increases.

In general, a correlation function for a GP model depends on the metric used to compute the distance between points. The Euclidian distance of two points on the cylindrical surface, on which equation (8) is based, actually corresponds to the three-dimensional "chordal" metric. More accurate metrics could be used instead, such as a geodetic distance, which is defined as the length of the shortest path(s) linking these points and remaining on the spatial domain defined by the cylindrical surface (Banerjee, 2005). By using the chordal distance, a slight underestimation of the geodetic distance is expected, because the chord "penetrates" the domain, producing a straight-line approximation to the geodetic arc.

In most practical settings, as in the reference case study where the spatial domain has negligible curvature, the approximation of the chordal metric can be considered appropriate (del Castillo *et al*., 2013). Furthermore, this approximation of the chordal metric has an important theoretical implication for the modeler. A troublesome aspect of geodetic distances is that they are not necessarily valid arguments for correlation functions, such as the exponential, squared exponential and Matérn (Curriero, 2006), which yield positive definite correlation matrices only in Euclidean spaces. Valid classes of correlation functions for Euclidean spaces are generated by Bochner's theorem, (ch. 2 of Stein, 1999), highlighting the theoretical importance of Euclidean metrics.

However, the approximation of the chordal metric (which is Euclidean) ensures valid correlation such as the squared exponential yield positive definite correlation matrices and enables proper convergence of the statistical estimation algorithms.

The Gaussian correlation function shown in (8) is one of the possible choices. In order to evaluate alternative selection, we also tried a different isotropic correlation function. In particular, we consider the class of Matérn correlation functions, where each function is indexed by a smoothness parameter

$\nu > 0$. The larger ν is, the smoother is the correlation function at zero, and the smoother is the spatial process (details can be found in Rasmussen and Williams (2006)). The Matérn class gives a broad range of shapes of correlation functions allowing any degree of smoothness at the origin (from continuous but non-differentiable, $\nu < 1$, to infinitely differentiable as $\nu \to \infty$), which includes, in the limit as $\nu \to \infty$, the Gaussian correlation in (8). The integer part of ν determines the number of times the underlying spatial process is differentiable.

In this paper, we considered the Matérn correlation function of smoothness parameter $\nu = 5/2$, which is 2-times differentiable and is given by

$$R_{Matern,\nu=5/2}\left(\mathbf{d}, \ell_h\right) = \left(1 + \frac{\sqrt{5}|\mathbf{d}|}{\ell_h} + \frac{5|\mathbf{d}|^2}{3\ell_h^{\,2}}\right)\exp\left(-\frac{\sqrt{5}|\mathbf{d}|}{\ell_h}\right). \tag{9}$$

In order to decide which of the correlation functions is better for the application at hand, we compared the log marginal likelihood values (Rasmussen and Williams (2006)) of the two corresponding GP models. Starting from the original set of $61 \times 68 = 4148$ surface points simulated by using the model in (1), we sub-sampled $N_1 = 31 \times 34 = 1054$ data points and used it as learning data set to fit the GP model, i.e., estimate the GP parameters and compute the log marginal likelihood, under both the squared exponential and the Matérn correlations functions. Through a pairwise t-test we compared the log marginal likelihood values obtained repeating the procedure on $h = 1, \ldots, 100$ simulated surfaces. The test procedure consisted of computing the difference of the pair of log marginal likelihood values on each of the h-th surface and then testing the hypothesis that the mean of the difference is zero. The obtained test statistic is $t_0 = -1.80$ and the 95% confidence interval is [-23.1582, 1.1448]. Choosing $\alpha = 0.05$ results in $t_{0.025,99} = -1.98$ and we concluded that there was no strong evidence to indicate that the two GP models differ in their mean log marginal likelihood (the p-value is P=0.08). For this reason we proceeded by comparing the fitting ability of the two corresponding GP models. After the estimate of the GP parameters, we then used the two GP models to predict surface data $\hat{r}_h\left(\mathbf{t}_i\right)$ at the remaining set

of $\tilde{N} = 4148 - 1054 = 3094$ locations and compared predicted data with the original values $r_h(\mathbf{t}_i)$ observed at the same locations. The Mean Square Prediction Error (MSPE) was eventually used as synthetic indicator of the prediction ability:

$$MSPE_h = \frac{1}{\tilde{N}} \sum_{i}^{\tilde{N}} \left(r_h(\mathbf{t}_i) - \hat{r}_h(\mathbf{t}_i) \right)^2 \quad h = 1, \ldots, 100, \tag{10}$$

Also in this case, the same procedure was repeated for $h = 1, \ldots, 100$ simulated surfaces and we collected 100 values of MSPE estimated using both the two correlation functions under study.

Figure 4 shows the plots of a surface simulated using the parametric model in (1) (Figure 4a), the corresponding surfaces predicted by the GP models under a squared exponential (Figure 4b) and a Matérn (Figure 4c) correlation functions. Visual comparison of the two predicted surfaces shown in Figure 4b and Figure 4c does not show significant differences in terms of reconstruction of the surfaces starting from a subset of data points. This qualitative result is confirmed by Figure 5, where the interval plots of the 100 MSPE values of the two competitive GP models are shown. As it can be observed, the difference between interval plots is not significant.

Similar results (not shown here) were also obtained using Matérn correlation functions characterized by different smoothness parameters ν. Considering these results, we decided to keep one correlation function only as reference for the following of the study, namely the squared exponential correlation function shown in (8).

Fitting a GP model requires one to estimate all the unknown parameters for each given surface. Let $\mathbf{p}_h = \left[a_h, \ell_h, \sigma^2_{\eta h} \right]$ represent the vector of unknown parameters for the h-th surface. The Gaussian Process for Machine Learning (GPML) Toolbox in the MATLAB environment was used in this paper for GP fitting and prediction (Rasmussen and Williams (2006), Rasmussen and Nickisch (2010, 2011)). Once unknown parameters are estimated, the GP model can be used to predict an expected observation at any spatial location (both observed and unobserved).

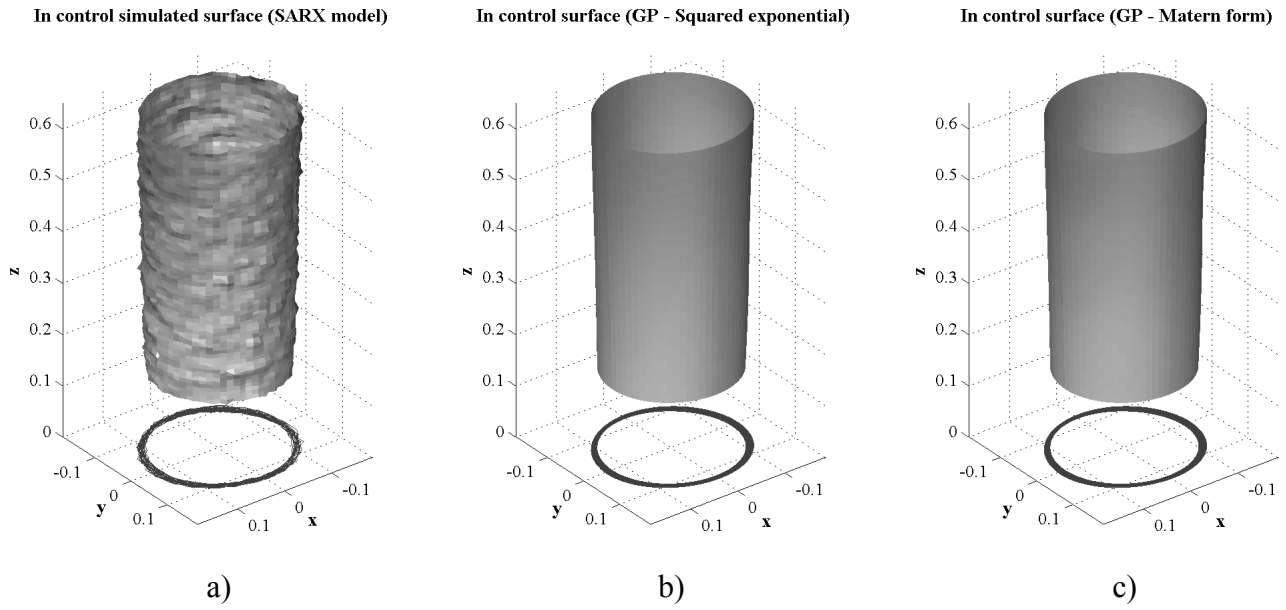


Figure 4: Simulated surface and GP models predicted surfaces (61×68 points). a) SARX model simulated surface. b) Surface predicted by a GP model with constant mean function and squared exponential covariance function. c) Surface predicted by a GP model with constant mean function and Matérn covariance function with $\nu = 5/2$. Gray scale is related to radial deviations. Dark colors represent deviations greater than zero, while bright colors represent deviation smaller than zero.

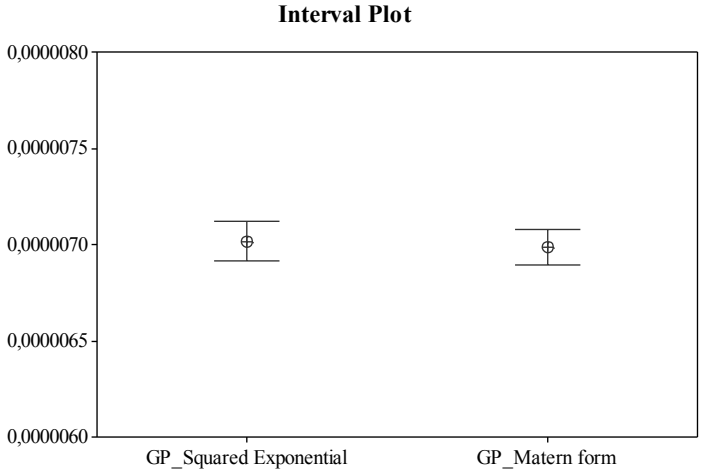


Figure 5: Interval plots on 100 values of MSPE obtained by using a GP model with constant mean function and squared exponential covariance function and a GP model with constant mean function and Matérn covariance function with $\nu = 5/2$.

Control charts for GP-based approach

Different approaches can be used to combine a GP model of the surface data with control charts. These different methods are described in the following.

T^2 control chart on GP parameters

The first approach mimics traditional methods for profile monitoring and consists of fitting a new GP model to each surface data, and then use a T^2 statistic for monitoring the vector of estimated parameters, namely

$$T_h^2 = \left(\hat{\mathbf{p}}_h - \hat{\boldsymbol{\mu}}_{par}\right)' \hat{\boldsymbol{\Sigma}}_{par}^{-1} \left(\hat{\mathbf{p}}_h - \hat{\boldsymbol{\mu}}_{par}\right), \tag{11}$$

where $\hat{\mathbf{p}}_h$ denote the 3×1 vector of GP parameters estimated on the h-th surface data, i.e. $\hat{\mathbf{p}}_h = \left[\hat{a}_h, \hat{\ell}_h, \hat{\sigma}_{\eta h}^2\right]$ while $\hat{\boldsymbol{\mu}}_{par}$ and $\hat{\boldsymbol{\Sigma}}_{par}$ are the mean vector and the covariance matrix of the parameters estimated in Phase I using a set of in-control surfaces.

Results achieved with this method will be referred to as 'GP_par' from now on. As we will see in a while, this approach has poor performance because a similar set of GP parameters can result in very different shapes of the surface to be monitored. For this reason, we decided to try a second approach described in the following subsection.

T^2 control chart on GP-predicted values

A second approach for GP surface monitoring consists of building a model of the in-control surface pattern by estimating the GP model on each of the Phase-I surfaces. Then, attention focuses on a set of opportunely selected locations, referred to as *checkpoints*, from now on. A multivariate T^2 control chart is then computed on each new profile by comparing deviations observed from the average in-control pattern estimated in Phase I. Abdel-Salam *et al.* (2010) showed that this method has similar performance when compared with approaches based on monitoring coefficients of a parametric or semi-parametric model.

Let $\hat{\boldsymbol{\mu}}_{par}$ be the mean value of GP parameters estimated in Phase I, $\hat{r}_k(\mathbf{t}_j)|\mathbf{r}_k, \hat{\boldsymbol{\mu}}_{par}$ denotes the value predicted at the *j*-th checkpoint $\mathbf{t}_j$, given all the data $\mathbf{r}_k$ observed on the *k*-th surface and the GP model with known parameter $\hat{\boldsymbol{\mu}}_{par}$. Repeating this procedure for all the Phase I in-control surfaces, we can eventually estimate the mean and covariance of predicted data at checkpoints, $\hat{\boldsymbol{\mu}}_{\hat{\mathbf{r}}}$ and $\hat{\boldsymbol{\Sigma}}_{\hat{\mathbf{r}}}$

For each new *h*-th surface the T^2 statistic is then given by:

$$T_h^2 = \left(\hat{\mathbf{r}}_h - \hat{\boldsymbol{\mu}}_{\hat{\mathbf{r}}}\right)' \hat{\boldsymbol{\Sigma}}_{\hat{\mathbf{r}}}^{-1} \left(\hat{\mathbf{r}}_h - \hat{\boldsymbol{\mu}}_{\hat{\mathbf{r}}}\right), \tag{12}$$

where $\hat{\mathbf{r}}_h$ represents the $n \times 1$ vector of surface data predicted at checkpoints of the *h*-th surface, where for simplicity of notation we skipped to show '$|\mathbf{r}_h, \hat{\boldsymbol{\mu}}_{par}$', i.e., prediction is done given the other data observed on the same surface and the vector of GP parameters estimated in Phase I.

Number and locations of the checkpoints

We start designing the T^2 control chart in (12) considering as checkpoints the whole number of points used to estimate the GP model. Results referring to this approach are labeled 'GP_all' henceforth. Usually, when a dense data grid is used to compute the T^2 statistic, numerical problems can arise as discussed in the Appendix. To overcome these drawbacks we found the nearest positive semidefinite matrix to the given sample matrix (Higham (2002).

Secondly, we design the T^2 control chart in (12) considering a reduced number of *n* checkpoints, which is not necessarily a subset of the original set of *N* measured points. In particular, for each new surface *N* points are measured as in the first case. However all this point cloud is used to predict via GP the surface pattern in a reduced number of locations. In other words, we are planning to work assuming a constant measurement effort (acquiring always a point cloud of fixed sample size *N*).

We explored two different strategies to locate checkpoints on the surface. The first assumes an equally-spaced cylindrical grid of checkpoints (obtained by combining 5 uniformly distributed values of the

heights with 5 uniformly distributed values of the angular position). Results referring to this approach are labeled 'GP_sub_unif' henceforth. The second method locates the same number of n=25 checkpoints using a latin hypercube design considered in the literature on computer experiments (McKay *et al.* (1979), Joseph and Hung (2008)). In particular, the maximin distance criterion (Morris and Mitchell (1995)) and the related Matlab routine shown in Forrester *et al.* (2008) is considered. The label 'GP_sub_lh' is used to show results of this approach.

In all the approaches (GP_par, GP_all, GP_sub_unif and GP_sub_lh), the T^2 control chart is coupled with a univariate control chart on the estimated residual variance:

$$\hat{\sigma}_h^2 = \frac{\hat{\mathbf{v}}_h' \hat{\mathbf{v}}_h}{N-1}, \tag{13}$$

where $\hat{\mathbf{v}}_h = \mathbf{r}_h - \hat{\mathbf{r}}_h$ is the vector of difference between the vector of actual surface data $\mathbf{r}_h$ and the vector $\hat{\mathbf{r}}_h$ of values predicted via GP at the same locations of the h-th surface.

Performance comparison

In order to speed up performance comparison and stay closer to production practice, down-sampling was performed on the original set of 61×68-point cylindrical surfaces. In particular, 1 point out of 2 was selected in both the axial and the angular directions to obtain a grid of $N_1 = 31\times34 = 1054$ in the first simulated scenario. Similarly, by selecting 1 point out of 3 in the original grid, a set of $N_2 = 20\times23 = 460$ data points defined the second productive scenario.

In this paper, we will concentrate on Phase II analysis, using the Average Run Length (ARLs) as performance indicator. For all the approaches, Phase I or the design stage was based on a set of 100 in-control surfaces simulated according to the parametric model shown in equation (1). A second set of 20000 in-control cylindrical surfaces, was then used to set the control limits of the multivariate and the

univariate control charts in order to achieve an overall in-control ARL=100, using $\alpha = 1 - \sqrt{1 - 1/ARL}$ as false alarm probability of each single control chart.

Once completed the design stage, Phase II performances are computed by generating new surfaces for all the out-of-control scenarios described in Section 2.1.1 and computing the ARL as the mean number of surfaces required to detect the specific out-of-control under study.

Comparison of approaches based on GP model of the surface

This first comparison study is aimed at comparing performance achieved by directly monitoring the GP parameters (GP_par) or using the GP model to predict surface data and then controlling deviations of the predicted values from the in-control pattern observed at that same set of locations in Phase I (GP_all, GP_sub_unif or GP_sub_lh). Note that the first approach -GP_par - is time-consuming, since it requires fitting a GP model for each new surface. On the contrary, the second set of approaches (GP_all, GP_sub_unif or GP_sub_lh) are faster since predicting new data with a known GP model is quick.

Table 2 summarizes the simulation results. In particular, the ARLs were estimated by computing a set of 250 run lengths, given new surfaces simulated according to each specific out-of-control model.

Table 2 clearly shows that the first method, which consists of monitoring the GP parameters (GP_par), is inefficient in signaling all the types of out of controls. In fact, in all the examined cases GP_par presents the lowest power of detection. This behavior can be mainly ascribed to the fact that variation of the cylindrical patterns does not translate in change of the GP parameters, which are in turn related to the type of correlation or similarity between adjacent points of the surface. We outline that a further main weakness of the first approach is the computational time, since it requires fitting the GP model for each cylindrical surface. Consequently, monitoring the GP parameters is not recommended in actual applications.

Table 2: Performance comparisons of methods GP_par, GP_all (computed on the whole set of N_2=460 data), GP_sub_unif and GP_sub_lh (computed on a subset of n=25 checkpoints): ARLs and Corresponding Standard Deviations within brackets (250 trials).

	Shift size	GP_par	GP_all	GP_sub_unif	GP_sub_lh
In control		102.83 (7.14)	102.29 (7.18)	101.28 (6.97)	101.60 (6.62)
Quadrilobe	0.002	57.18 (3.40)	1.61 (0.07)	2.46 (0.11)	1.99 (0.08)
Trilobe	0.003	68.08 (4.29)	15.10 (0.91)	19.71 (1.14)	1.35 (0.04)
Half frequency	0.001	78.96 (5.12)	43.59 (2.87)	9.15 (0.54)	3.80 (0.20)
Tapering	0.15	102.53 (6.10)	83.60 (5.68)	64.87 (3.79)	63.55 (4.06)
Residuals variance	1.1	58.17 (3.28)	7.29 (0.39)	7.52 (0.44)	7.40 (0.39)

As shown in Table 2, the best performance are associated to GP_sub_lh, where $n = 25$ checkpoints are placed according to the maximin distance criterion (Morris and Mitchell (1995)).

When compared to GP_sub_unif, GP_all presents improved performance only for out-of-control conditions characterized by high frequency content, as for the trilobe and the quadrilobe pattern. Poor performances of GP_sub_unif method in these cases can be mainly ascribed to an inappropriate locations of the checkpoints. In fact, by simply varying the location of the $n = 25$ checkpoints (i.e., moving from GP_sub_unif to GP_sub_lh), we observe a clear performance improvement.

Similar results (not reported in this paper) were obtained by using n=460 checkpoints with locations established through a maximin distance criterion. Also in this case, the power of detection of out-of-control conditions at low harmonic frequency was poor when compared to GP_sub_unif and GP_sub_lh methods. Furthermore it has to be highlighted that a high density of checkpoints makes the maximin

distance criterion computational time requesting and less effective due to the reduced degrees of freedom on the design space.

From obtained simulation results we can conclude that high numbers of checkpoints may lead to poor performance, especially for out-of-control conditions characterized by low harmonic frequencies. These results are consistent with other studies showing performances of multivariate control charts as a function of the number of process variables to be monitored (Prabhu and Runger (1997), Champ *et al.* (2005), Chenouri *et al*. (2009)). As a matter of fact, as the number of variables increase the average run-length performance to detect a specified shift in the mean of the variables for multivariate control charts also increases, because the shift in the mean of the variables is diluted in the multidimensional space of the process variables (Montgomery (2009)).

On the contrary, both the methods based on monitoring a reduced number of checkpoints, i.d. $n = 25$, (GP_sub_unif and GP_sub_lh) are well performing. In particular, the second approach, which locates the checkpoints using a space-filling criterion instead of placing them uniformly, seems the more promising.

A comparison between different approaches for surface monitoring

In this second comparison study, performances of the GP methods (GP_sub_unif and GP_sub_lh) are compared to two benchmarks. The first benchmark consists of using a standard univariate control chart for monitoring a synthetic indicator associated to each surface. This indicator is the cylindricity form error, which will be referred to as *out-of-cylindricity* (OOC), from now on. This indicator is the minimum radial distance between two coaxial cylinders that contain among them the actual surface (ISO (2012)). Typically, OOC values are computed for quality inspection of machined surfaces in order to decide whether an item can be considered conforming to the requirements. In fact, when the OOC is greater than the tolerance value specified in the technical drawing, the item has to be scrapped or reworked. This monitoring approach is selected as representative of the industrial practice.

The second benchmark was proposed for cylindrical surface monitoring (Colosimo *et al.* (2010, 2011)) and consists of using a T^2 control chart for monitoring the regression-based parameters of model shown in (1), coupled with a univariate control chart for monitoring the estimated residuals variance.

As a summary, for each scenario ($N_1 = 1054$ and $N_2 = 460$), four competing approaches are considered: i) the individual control chart on the form errors (OOC); ii) the SARX-based approach proposed by Colosimo *et al.* (2010, 2011); iii) the GP-based approach with the T^2 control chart designed on a uniform grid of checkpoints (GP_sub_unif); iv) the GP-based approach with the T^2 control chart designed on a Latin Hypercube distribution of checkpoints (GP_sub_lh).

Also OOC and SARX-based approaches were first tuned in order to achieve the same in-control ARL value of about 100, the control limits were obtained by simulation on 20000 instances of in-control cylindrical surfaces. The simulation results of our simulation study for scenario 1 ($N_1 = 1054$) and scenario 2 ($N_2 = 460$) are shown in Figure 6-10 and reported in Table 3.

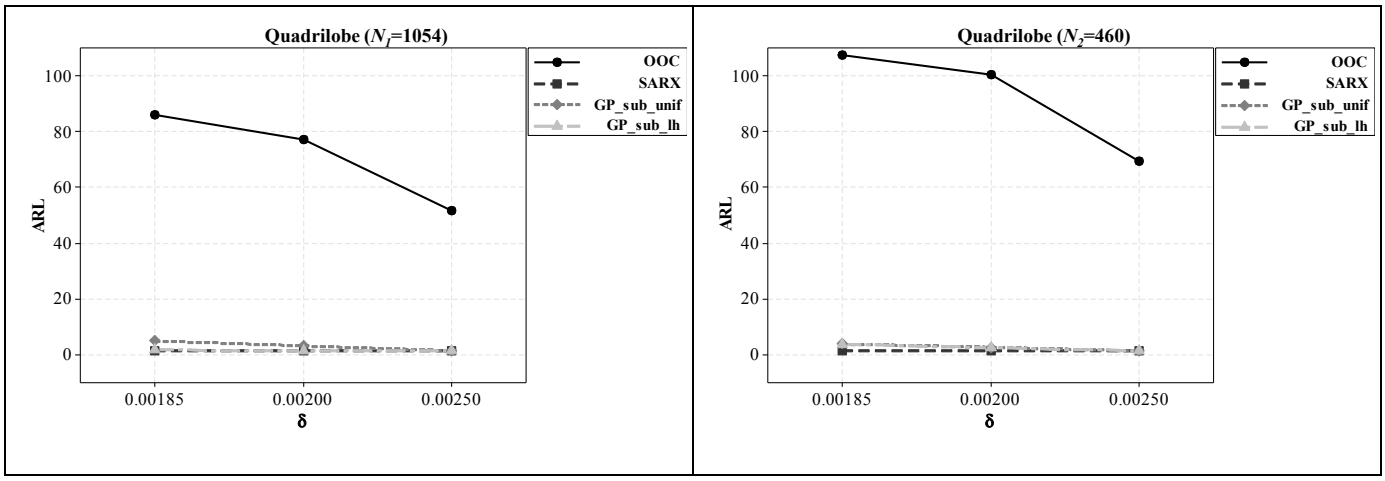


Figure 6: Performance comparison of OOC, SARX, GP_sub_unif and GP_sub_lh for quadrilobe out-of-control condition: ARL values (1000 trials) in scenario 1 (N_1=1054) and in scenario 2 (N_2=460).

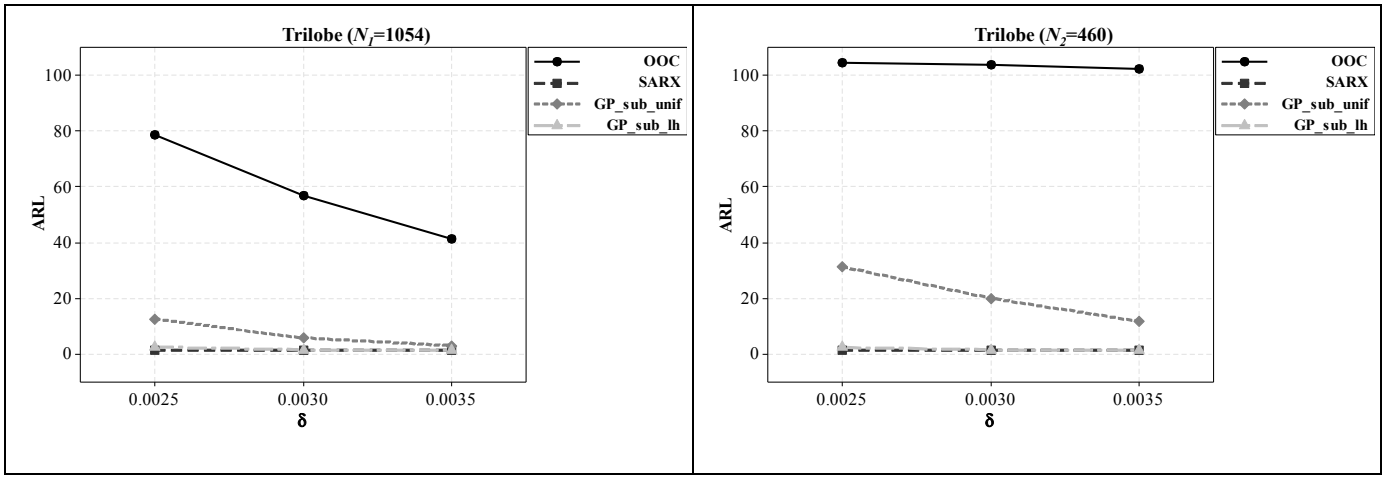


Figure 7: Performance comparison of OOC, SARX, GP_sub_unif and GP_sub_lh for trilobe out-of-control condition: ARL values (1000 trials) in scenario 1 (N_1=1054) and in scenario 2 (N_2=460).

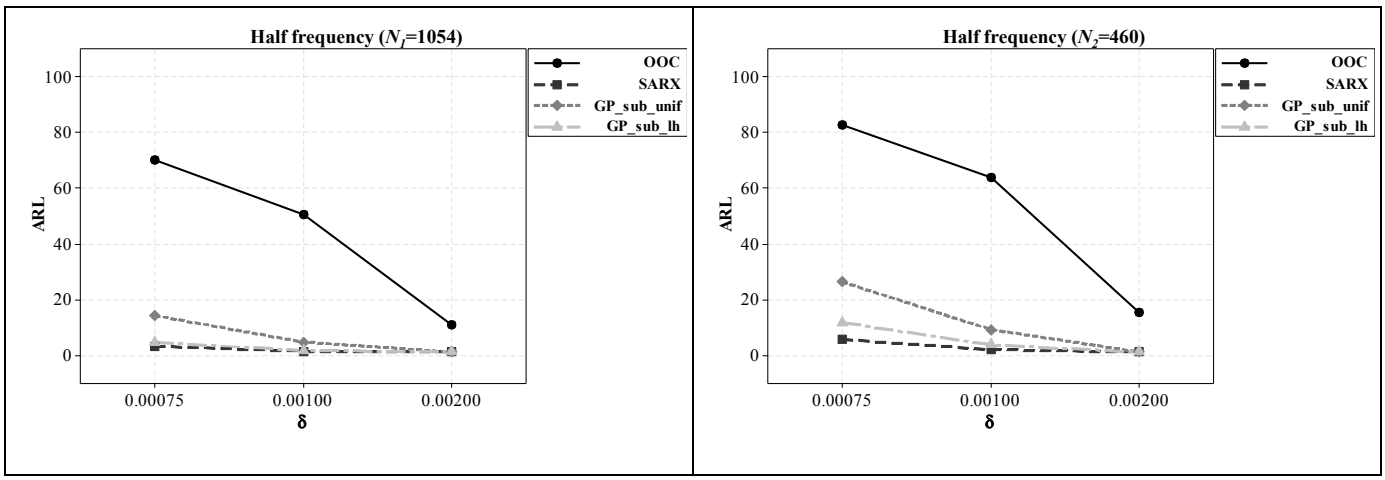


Figure 8: Performance comparison of OOC, SARX, GP_sub_unif and GP_sub_lh for half frequency out-of-control condition: ARL values (1000 trials) in scenario 1 (N_1=1054) and in scenario 2 (N_2=460).

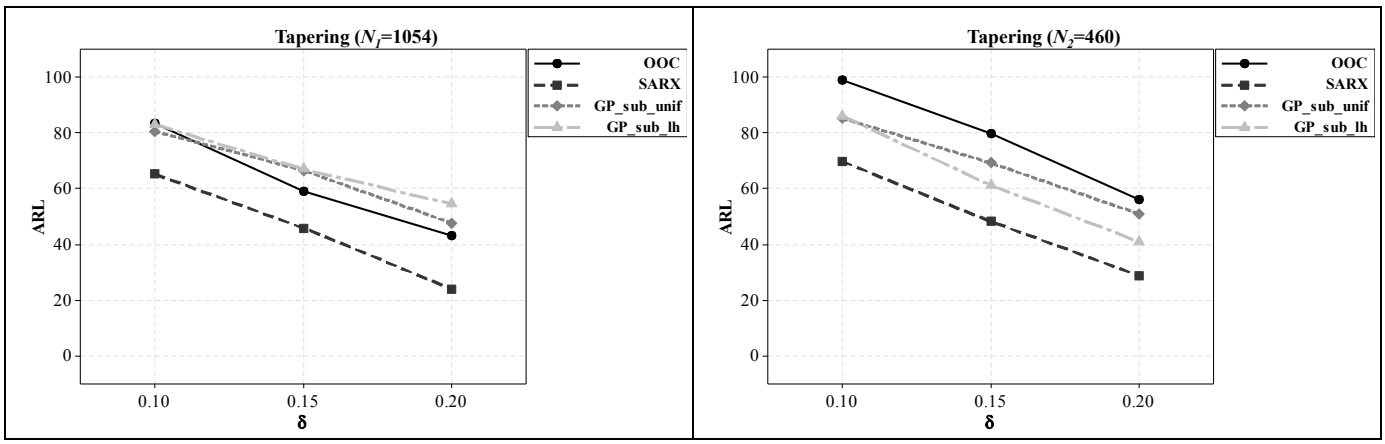


Figure 9: Performance comparison of OOC, SARX, GP_sub_unif and GP_sub_lh for Tapering out-of-control condition: ARL values (1000 trials) in scenario 1 (N_1=1054) and in scenario 2 (N_2=460).

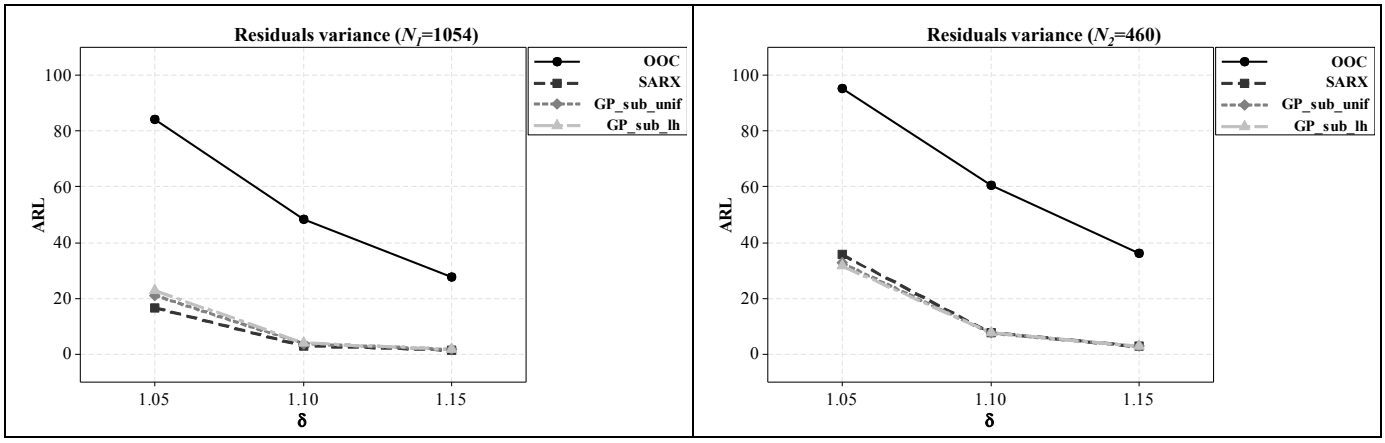


Figure 10: Performance comparison of OOC, SARX, GP_sub_unif and GP_sub_lh for increase of residuals variance out-of-control condition: ARL values (1000 trials) in scenario 1 (N_1=1054) and in scenario 2 (N_2=460).

The ARLs values are the averages of 1000 run lengths in each out-of-control condition for different values of the parameter δ (shown in the abscissa of Figure 6-10), which represents the level of surface change.

As it can be observed, the GP-based monitoring approaches (GP_sub_unif and GP_sub_lh) and the regression based method (SARX) outperform the simple OOC control chart in almost all the simulation scenarios. In fact, in most cases the individual control chart of the OOC values presents the lowest power of detection. The only exception is the tapering out-of-control (Figure 9), where the OOC control chart seems to perform as well as the GP-based approach when $N_1 = 1054$.

Comparing the two GP-based approaches (GP_sub_unif and GP_sub_lh), GP_sub_lh presents improved performance when compared to GP_sub_unif, as we observed in the previous performance comparison study. In other words, using GP with an appropriate distribution of checkpoints allows us to obtain better performances with respect to the case where a uniform grid of checkpoints is used.

Eventually, the regression-based method (SARX) performs as better as the GP_sub_lh in all the scenarios but the tapering one. This result is particularly worth, since we expected the SARX method to outperform the GP-based methods by construction. As a matter of fact, all the surfaces were simulated

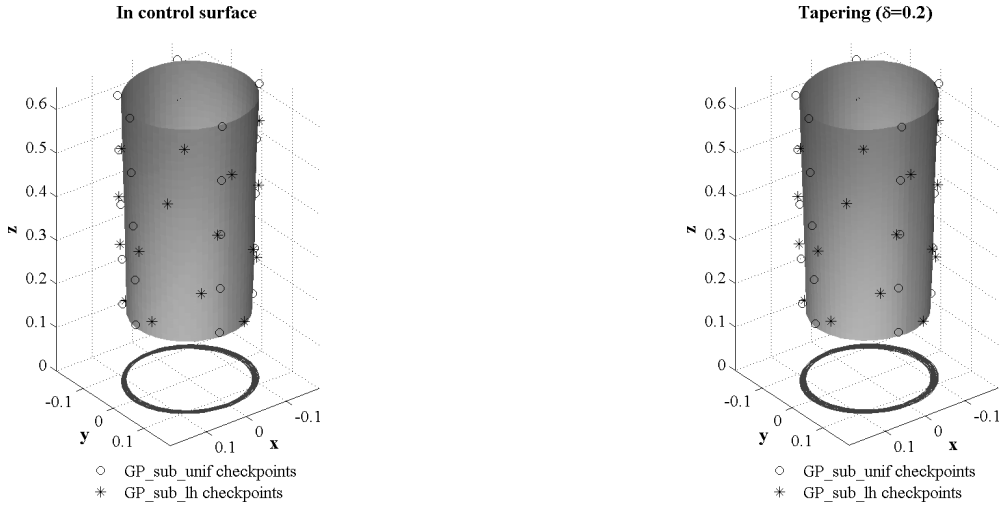
using the SARX model and this is why we were expecting that a control charting procedure based on this model should outperform competing methods.

As we already noted, the only case where the GP_sub_lh approach does not produce satisfactory results is that characterized by the tapering error, a case in which all the approaches show unsatisfactory results. In this case, the difference between the approaches shows that GP_sub_lh is still close to the optimal performance, which is achieved by the SARX-based method.

From our simulation study, we observed that this last behaviour might be mainly ascribed by a combination of two factors: i) the locations of checkpoints, which cause only a small subset of them to be influenced by the effect of tapering (e.g. the checkpoints at higher azimuthal locations) and ii) low amplitude parameters ($\delta_{taper} = 0.1, 0.15, 0.2$), which make the tapering to have a limited effect on the out-of-control condition.

As shown in Figure 11, the effect of tapering error is irrelevant also for δ_{taper}=0.2 when compared to the corresponding in-control surface. Values of δ_{taper} greater than 0.2 were not considered in this work, since they could represent unrealistic out-of-control conditions of a lathe-turning process.

From the results shown in Figure 6-10, by comparing graphs on the left with the corresponding ones on the right, we can observe that reducing the number of measured points (from $N_1 = 1054$ to $N_2 = 460$), increases the ARLs without affecting the ranking of methods.



a) b)

Figure 11: Checkpoints for GP_sub_unif and GP_sub_lh approaches. a) In control surface predicted by GP_model. b) Tapering ($\delta = 0.2$) predicted by GP_model. Circle points and asterisk points represent the 25 checkpoints of GP_sub_unif and GP_sub_lh respectively. Gray scale is related to radial deviations. Dark colors represent deviations greater than zero, while ~~cold~~ bright colors represent deviation smaller than zero.

In general, we can observe that deciding on the location of checkpoints is crucial, as the performance of the control charts combined to GP may be greatly influenced by how the checkpoints are distributed on the cylindrical surface. Adopting a simple strategy in which checkpoints are uniformly distributed over a grid of equally spaced locations (along the vertical and angular coordinates) is not recommended. This is particularly true when it is important to signal timely out-of-control conditions characterized by periodic shape errors along the angular direction (e.g., Quadrilobe, Half Frequency and Trilobe errors). In fact, checkpoints uniformly distributed over a grid of locations at fixed vertical and angular coordinates, may cause several of these points result situated on positions, which share the same angular coordinate. In these cases, the T^2 control chart may have poor power in detecting periodic out-of-control conditions.

Table 3: Performance comparisons of OOC, SARX, GP_sub_unif and GP_sub_lh: ARLs and Corresponding Standard Deviations within brackets (1000 trials) in scenario 1 (N_1=1054) and in scenario 2 (N_2=460).

	Shift size	OOC	SARX	GP_sub_unif	GP_sub_lh
			Scenario 1 ($N_1 = 1054$)		
In control		101.53 (2.98)	103.40 (3.26)	99.69 (3.08)	100.77 (3.19)
Quadrilobe	0.00185	86.12 (2.80)	1.03 (0.00)	4.70 (0.13)	1.39 (0.02)
	0.002	77.29 (2.35)	1.01 (0.00)	2.90 (0.07)	1.16 (0.01)
	0.0025	51.70 (1.68)	1.00 (0.00)	1.14 (0.01)	1.00 (0.00)
Trilobe	0.0025	78.70 (2.47)	1.09 (0.01)	12.40 (0.38)	2.27 (0.05)
	0.003	56.75 (1.79)	1.00 (0.00)	5.57 (0.15)	1.29 (0.01)
	0.0035	41.42 (1.35)	1.00 (0.00)	2.78 (0.07)	1.04 (0.01)
Half frequency	0.00075	70.27 (2.17)	3.06 (0.08)	14.11 (0.43)	4.51 (0.13)
	0.001	50.51 (1.48)	1.27 (0.02)	4.61 (0.13)	1.54 (0.03)
	0.002	10.99 (0.34)	1.00 (0.00)	1.00 (0.00)	1.00 (0.00)
Tapering	0.1	83.68 (2.59)	65.34 (2.29)	80.81 (2.46)	83.18 (2.64)
	0.15	59.12 (1.80)	45.88 (1.44)	66.66 (2.01)	67.11 (2.10)
	0.2	43.01 (1.36)	23.98 (0.72)	47.63 (1.48)	54.53 (1.77)
Increase of variance	1.05	84.23 (2.56)	16.51 (0.49)	20.86 (0.68)	22.77 (0.71)
	1.1	48.31 (1.44)	2.69 (0.06)	3.55 (0.10)	3.70 (0.10)
	1.15	27.36 (0.79)	1.21 (0.02)	1.38 (0.02)	1.37 (0.02)
			Scenario 2 ($N_2 = 460$)		
In control		102.38 (3.24)	101.33 (3.43)	98.76 (3.01)	102.39 (3.35)
Quadrilobe	0.00185	107.79 (3.30)	1.28 (0.02)	3.58 (0.09)	3.48 (0.09)
	0.002	100.73 (3.12)	1.09 (0.00)	2.41 (0.06)	2.23 (0.05)
	0.0025	69.36 (2.16)	1.00 (0.00)	1.07 (0.01)	1.08 (0.01)
Trilobe	0.0025	104.65 (3.31)	1.25 (0.01)	31.25 (0.94)	2.16 (0.05)
	0.003	104.04 (3.30)	1.02 (0.00)	19.84 (0.58)	1.28 (0.02)
	0.0035	102.75 (3.30)	1.00 (0.00)	11.60 (0.35)	1.05 (0.01)
Half frequency	0.00075	83.00 (2.63)	5.46 (0.15)	26.34 (0.83)	11.63 (0.35)
	0.001	63.80 (2.03)	1.88 (0.04)	9.10 (0.27)	3.59 (0.10)
	0.002	15.33 (0.44)	1.00 (0.00)	1.01 (0.00)	1.00 (0.00)
Tapering	0.1	99.05 (3.06)	69.98 (2.34)	85.44 (2.78)	86.33 (2.71)
	0.15	80.00 (2.58)	48.21 (1.48)	69.35 (2.17)	61.33 (1.93)
	0.2	56.24 (1.77)	28.66 (0.92)	50.92 (1.65)	40.92 (1.24)
Increase of variance	1.05	95.68 (3.06)	35.59 (1.10)	32.55 (1.01)	31.50 (0.98)
	1.1	60.49 (1.86)	7.39 (0.21)	7.32 (0.21)	7.55 (0.21)
	1.15	35.89 (1.13)	2.48 (0.06)	2.45 (0.06)	2.51 (0.07)

On the other hand, in a space-filling distribution, points are scattered throughout the surface by optimizing a specific criterion. Our simulation results showed that, one of the most widely used measures to evaluate the 'space-fillingness' of points over a region, i.e. the 'maximin' metric introduced by Johnson *et al.* (1990), can be usefully adopted in order to decide on the locations of checkpoints. By adopting this method, the performance of the GP-based approach is improved when compared to those obtained by uniformly distributing the points on a regular grid.

With reference to the out-of-control condition characterized by increase of residual variance the performances of two compared GP methods are similar. In this case, the shape error is 'randomly' distributed over the cylindrical surface. Performance of GP control charts is not influenced by the distribution of checkpoints.

Comparison with other approaches for profile monitoring

Different authors presented approaches based on the idea of predicting profile data on a set of points (that we called *checkpoints* in our paper) by using appropriate smoothing techniques and then monitoring the deviations between the actual and the in-control data at these set of locations. As a matter of fact, a similar approach was proposed by Qiu *et al*. (2010) and Chipman *et al*. (2010) for nonparameteric profile monitoring. Zhang *et al*. (2013) and Grimshaw *et al*. (2013) presented similar approaches for GP-based monitoring.

In this section, our aim is to compare performance of different ways of summarizing discrepancies between the observed and the in-control data points in the statistic to be monitored.

In order to keep comparison fair, we will re-conduct all the different methods, sometimes involving exponential weighting schemes, to the Hotelling control chart, as the one we used in our paper. The main idea behind this choice is to keep a monitoring scheme that stays alert for both single outlying profiles and sustained profile shifts. Secondly, we will focus the attention only on the different statistics to be

plotted on the control chart against the control limit, while keeping the GP as reference smoothing technique.

Qiu *et al*. (2010) presented an approach for nonparametric profile monitoring, which can be adapted to our framework by eliminating the EWMA weighting scheme and considering GP-based prediction in the checkpoints, thus resulting in the following statistic (written in matrix form):

$$Q_{QZW\_h} = \frac{1}{n}(\hat{\mathbf{r}}_h - \hat{\boldsymbol{\mu}}_{\hat{r}})' \hat{\boldsymbol{\Sigma}}_1^{-1} (\hat{\mathbf{r}}_h - \hat{\boldsymbol{\mu}}_{\hat{r}}) \tag{14}$$

where $\hat{\boldsymbol{\Sigma}}_1 = diag\left(\hat{\boldsymbol{\Sigma}}_{\hat{\mathbf{r}}}\right)$, $\hat{\boldsymbol{\mu}}_{\hat{\mathbf{r}}}$ and $\hat{\boldsymbol{\Sigma}}_{\hat{\mathbf{r}}}$ are the Phase-1 estimate of the mean and covariance of predicted data at checkpoints and $\hat{\mathbf{r}}_h$ represents the vector of surface data predicted at checkpoints of the *h*-th surface.

Equation (15) is clearly an oversimplified version of the Qiu *et al*.'s approach, since the main contribution of their method is what we neglected, i..e. the nonparametric exponentially smoothing scheme used to predict surface data. However, by separating the smoothing approach from the structure of the statistic to be monitored, we can draw some fair conclusions on the two separate effects. In the discussion to the paper by Qiu *et al.* (2010), Chipman *et al*. (2010) suggested a different statistic for nonparametric profile monitoring, which is basically equivalent to our proposal when the exponentially weighting scheme is neglected. This is why we will not include the Chipman *et al*.'s statistic as competitor method in this study.

Zhang *et al*. (2013) presented a different approach to model the within-profile correlation estimated via GP. In this paper, the linear trend term (if any) is monitored with a first statistic while the remaining within-profile correlation is modeled via GP and monitored using the successive difference of profile data at selected locations:

$$T^2_{MMR\_h} = \frac{1}{2}\left(\hat{\mathbf{r}}_h - \hat{\mathbf{r}}_{h-1}\right)' \hat{\boldsymbol{\Sigma}}_{\hat{\mathbf{r}}}^{-1} \left(\hat{\mathbf{r}}_h - \hat{\mathbf{r}}_{h-1}\right) \tag{15}$$

A third possible competitor is the GP-based method proposed by Grimshaw *et al*. (2013). However, we will not include this further approach as competitor because the statistic they proposes is very similar to ours.

It is worth noting that none of these methods discussed about locations of the checkpoints. Some of the methods (Grimshaw *et al*. (2013), Zhang *et al.* (2013) simply assumed to use a fixed set of locations for Phase 1 and Phase 2. Qiu *et al*. (2010) discussed about a different design matrix when monitoring has to be performed instead of smoothing, but they did not deepen discussion on how to select the design points.

Table 4 shows the ARLs performance and the corresponding standard deviations (within brackets, 1000 replicates) of our GP_sub_lh method, the MMR (equation 15) and QZW (equation 14) approaches. All the approaches now will share the same N_2=460 points were used to estimate the in-control model in Phase 1 and to predict data at the checkpoints of each new surface. Once again, we use all the N_2=460 points measured on each surface to build the statistic which is based only on the set of 25 checkpoints.

Once again, we found that our proposed approach has the best performance on almost all the simulated scenarios. The only exception is the case of tapering out of control, where the approach by Qiu *et al.* (2010) seems to outperforms all the other two.

Table 4: Performance comparisons of GP_sub_lh, MMR and QZW: ARLs and corresponding standard deviations within brackets (1000 trials) in scenario 2 (N_2=460).

	Shift size	**GP_sub_lh**	**MMR**	**QZW**
In control		102.39 (3.35)	100.36 (3.39)	99.82 (3.09)
Quadrilobe	0.002	2.23 (0.05)	101.87 (3.45)	89.50 (2.84)
Trilobe	0.003	1.28 (0.02)	102.03 (3.63)	73.15 (2.42)
Half frequency	0.001	3.59 (0.10)	96.66 (3.09)	11.96 (0.36)

Tapering	0.15	61.33 (1.93)	92.86 (3.33)	24.17 (0.78)
Residuals variance	1.1	7.55 (0.21)	31.83 (1.12)	80.50 (2.55)

Conclusion

In this paper, a GP-based approach for surface monitoring is presented and analyzed through an extensive simulation study. A real case study concerning lathe-turned items subject to cylindricity tolerance was used for the study. The main conclusions of the study are summarized below.

1) The GP model can be effectively used to represent surface patterns. Compared with traditional regression-based approaches to surface modeling, the main advantage is its easiness of use, since no cumbersome activity of regressor selection is required. Further work should be devoted to deepen advantages of using geodesic distances instead of Euclidean distances in the correlation function (del Castillo *et al.* (2013)).
2) Control charts based on GP model cannot monitor the estimated GP parameters, as commonly done for profile monitoring. In fact, similar values of the GP parameters can result in very different shapes of the manufactured surfaces.
3) Our proposed method consists of monitoring the vector of deviations between surface data predicted via the in-control GP model and the real surface data observed on the surface. This approach is able to detect unwanted change of the surface, provided that appropriate number and sampling strategy of surface points to be monitored is considered. In our study, we showed that a minimax space filling Latin Hypercube design can be very effective. Further work evaluating the effect of sampling strategy is required.

As a summary, research on surface monitoring via GP appeared to be a promising direction for further research. Considering that the approach proposed is quite general, a revised version of the proposed method to profile monitoring could be also tested and compared with existing procedures. In particular, extensions of the proposed method can be effectively considered to monitor image data (Megahed, *et al.*, 2011), which can be interpreted as a special case of 2.5D surface.

Acknowledgements

The authors wish to thank an anonymous referee and the editor for their comments and suggestions which have resulted in an improved presentation. This work was partially funded by M.I.U.R. Grant: PRIN 2008 – "Large-scale coordinate metrology: study and realization of an innovative system based on a network of distributed and cooperative wireless sensors". This research was partially supported by the European Union's Seventh Framework Programme (FP7/2007-2013) under grant agreement number 285075 - MuProD

Appendix: Numerical problems in computing T^2 statistics using GP models

Two main problems can arise when using surface data predicted via GP to compute the T^2 statistic shown in equation (12). A first problem can arise because of the *round-off error* i.e., the error that occurs through replacing real numbers with floating-point decimal numbers with only a fixed number of digits. This error can firstly cause the propagation of absolute or relative errors in all the computations that follow. Secondly, as a consequence of rounding and truncation a second type of error called *cancellation* can happen. In fact, it is possible to show that subtractive cancellation causes relative errors or uncertainties already present in the matrix to be magnified; in other words, subtractive cancellation brings earlier errors into prominence (see Higham (1996)). The second issue is *ill-posedness* of the problem as surface data become more dense. In particular, the main drawback is related to the inversion

of a matrix (in our case the covariance matrix), which can become rank deficient or close to rank deficient due to multicollinearity brought by the proximity of points in the design matrix. The spectral condition number of a symmetric matrix, A, defined as the ratio between the absolute values of the largest and smallest magnitude eigenvalues, measures the sensitivity of the solution of a problem to perturbations in the data, either in initial data or introduced by a solution process, to the errors in the final result (Wilkinson (1961)). For this reason, large condition numbers are undesirable in practice.

Ababou *et al*. (1994) investigate the behavior of the condition number of the covariance matrix with respect to two factors in order to find which characteristic affects results in terms of condition number: choice of the covariance function and number of data points in the domain. At the end of their paper, they founded out that the condition number diverges as sampling density increases while domain size is fixed, especially for covariance models which are "flat" at the origin resulting in a smooth function of the data. Posa (1989) showed that the introduction of noise in the data reduces smoothness, and therefore results in a less ill-conditioned covariance matrix, partially solving the influence of the choice of the correlation function. The relationship between the size of the matrix and previous numerical issues is well known and, as already stated, one would expect them to increase as the size increased (Higham (1996)).

To overcome these difficulties we decided to follow two different methods: during the estimate of optimal parameters inside the optimization algorithm, where numerical errors are only due to round-off, we simply discard solutions that don't lead to a positive semidefinite correlation and covariance matrices; instead, when we estimate the prediction covariance matrix which may not be positive semidefinite due to the cancellation effect, we decide to find the nearest positive semidefinite matrix to the given sample matrix by the approach proposed in Higham (2002), who formalized the notion of nearness using a weighted Frobenius norm and provided a method for computing the nearest covariance or correlation matrix.

References

Ababou, R., Amvrossios C., Bagtzoglou C., Wood, E.F. (1994), ”On the Condition Number of Covariance Matrices in Kriging Estimation, and Simulation of random Fields”, *Mathematical Geology*, 26, pp. 99-133.

Abdel-Salam, A.G., Birch, J.B., and Jensen, W.A. (2010). “Nonparametric and semiparametirc mixed models methods for Phase I profile monitoring”. Technical Report, Department of Statistics, Virginia Polytechnic Institute & State University http://www.web-e.stat.vt.edu/dept/web-e/tech_reports/TechReport10-2.pdf.

Banerjee, S. (2005) “On geodetic distance computations in spatial modeling”. *Biometrics*, 61, pp. 617-625.

Champ, C. W., Jones-Farmer, A.L. and Rigdon S. E. (2005). “Properties of the T^2 Control Chart When Parameters Are Estimated”. *Technometrics*, 47, 437-445.

Chenouri, S., Steiner, S. and Variyath, A.M. (2009). “A Multivariate Robust Control Chart for Individual Observations”. *Journal of Quality Technology,* 41, pp.259-271

Chipman H.A., MacKay R.J., Steiner SH. (2010) “Comment: nonparametric profile monitoring by mixed effects modeling”. *Technometrics*, 52(3), pp. 280–283.

Cho, N.W., and Tu, J.F. (2001). “Roundness modeling of machined parts for tolerance analysis”. *Precision Engineering*, 25, pp. 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, pp. 707-725.

Colosimo, B.M., and Pacella, M. (2010). “A comparison study of control charts for statistical monitoring of functional data”, *International Journal of Production Research*, 48, pp. 1575-1601.

Colosimo, B.M., and Pacella, M. (2011). "Analyzing the effect of process parameters on the shape of 3D Profiles". *Journal of Quality Technology*, 43, pp.169-195.

Colosimo, B.M., and Senin, N., eds. (2011). *Geometric Tolerances. Impact on Product Design, Quality Inspection and Statistical Process Monitoring*. London: Springer.

Colosimo, B.M., Mammarella, F., and Petrò, S. (2010). "Quality control of manufactured surfaces". In *Frontiers of Statistical Quality Control*, edited by HJ Lenz and PT Wilrich, Springer.

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, pp.1-18.

Colosimo, B.M., Pacella, M., and Semeraro, Q. (2011). "Statistical Process Control for Geometric Specifications". In: Noorossana, R., Saghaei, A., and Amiri, A., eds. *Statistical Analysis of Profile Monitoring*. pp. 217-252, John Wiley & Sons, Hoboken.

Cressie, N.A.C. (1993). *Statistics for spatial data*. Revised edn. John Wiley & Sons, NewYork.

Curriero, F.C., (2006). "On the use of non-Euclidean distance measures in geostatistics". *Mathematical Geology*, 38, pp. 907-926.

Del Castillo, E., Colosimo, B.M., and Tajbakhsh, S. (2013). "Geodesic Gaussian Processes for the Reconstruction of a Free-Form Surface", submitted paper – available as technical report http://www2.ie.psu.edu/Castillo/research/EngineeringStatistics/publications.htm.

Forrester, A., Sobester, A., and Keane, A. (2008). *Engineering design via surrogate modelling: a practical guide*. John Wiley & Sons.

Grimshaw, S.D., Blades, N.J., and Miles, M.P. (2013). "Spatial Control Charts for the Mean". *Journal of Quality Technology*, 45- 2, pp. 130-148.

Henke, R.P., Summerhays, K.D., Baldwin, J.M., Cassou, R.M., and Brown, C.W. (1999). "Methods for evaluation of systematic geometric deviations in machined parts and their relationships to process variables". *Precision Engineering*, 23, pp. 273-292.

Higham, N.J. (1996). *Accuracy and stability of Numerical Algorithms*, SIAM.

Higham, N. (2002). "Computing the nearest correlation matrix-A problem from finance", *JIMA*, 22, pp. 329–343.

ISO (2012). Geometrical product specifications (Gps)-Geometrical tolerancing-Tolerances of form, Orientation, Location and Run-Out. ISO 1101:2012, International Organization for Standardization.

Johnson, M.E., Moore L.M., and Ylvisaker, D. (1990). "Minimax and maximin distance designs". *Journal of Statistical Planning and Inference*, 26, pp.131-148.

Joseph, V.R., and Hung, Y. (2008). "Orthogonal-Maximum Latin Hypercube Designs". *Statistica Sinica* 18, pp.171-186.

Kang, L., and Albin, S.L. (2000). "On-line monitoring when the process yields a linear profile". *Journal of Quality Technology*, 32, pp.418-426.

Kim, K., Mahmoud, M.A., and Woodall, W.H. (2003). "On the monitoring of linear profiles". *Journal of Quality Technology*, 35, pp.317-328.

Mahmoud, M.A., and Woodall, W.H. (2004). "Phase I analysis of linear profiles with calibration applications". *Technometrics*, 46, pp.377-391.

Matheron, G. (1963). "Principles of geostatistics". *Economic Geology*, 58, pp.1246-1266.

McKay, M.D., Beckman, R.J., and Conover, W. J. (1979). "Comparison of three methods for selecting values of input variables in the analysis of output from a computer code". *Technometrics*, 21, pp. 239-245.

Megahed, F.M., Woodall, H.W., and Camelio, J.A. (2011). "A Review and Perspective on Control Charting with Image Data". *Journal of Quality Technology*, 43, pp.83-98.

Montgomery, D.C. (2009). *Introduction to statistical quality control*. 6th ed. New York, NY: John Wiley & Sons, Inc.

Morris, M.D., and Mitchell, T.J. (1995). "Exploratory designs for computational experiments". *Journal of Statistical Planning and Inference*, 43, pp.381-402.

Noorossana R.R., Saghei A., and Amiri A., eds. (2012). *Statistical Analysis of Profile Monitoring*, Wiley.

Posa, D. (1989). "Conditioning of the stationary kriging matrices for some well-known covariance models", *Mathematical Geology*, 21, pp. 755-765.

Prabhu, S.S., and Runger G.C. (1997). "Designing a Multivariate EWMA Control Chart". *Journal of Quality Technology*, 29, pp. 8-15.

Qiu P., Zou C., and Wang Z. (2010). "Nonparametric Profile Monitoring by Mixed Effects Modeling" – with discussion. *Technometrics*, 52, 3, pp. 265-293.

Rasmussen, C.E., and Nickisch, H. (2010). "Gaussian Process for Machine Learning Toolbox". *Journal of Machine Learning Research*, 11, pp. 3011–3015.

Rasmussen, C.E., and Nickisch, H. (2011). The GPML Toolbox - Version 3.1. url: www.GaussianProcess.org/gpml [accessed May 2013].

Rasmussen, C.E., and Williams, C.K.I. (2006). *Gaussian processes for machine learning*. Cambridge-MA: the MIT Press.

Stein, M.L. (1999). *Statistical Interpolation of Spatial Data: Some Theory for Kriging*. Springer, New York.

Wilkinson, J. H. (1961), "Error Analysis of Direct Methods of Matrix Inversion", *Journal of the Association for Computing Machinery*, 8, pp. 281-330.

Woodall, W.H. (2007). "Current research on profile monitoring". *Produção*, 17, pp.420-425.

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, pp.309-320.

Xia, H., Ding, Y., and Wang, J. (2008). "Gaussian process method for form error assessment using coordinate measurements". *IIE Transactions*, 40, pp.931-946.

Zhang, X.D., Zhang, C., Wang, B., and Feng, S.C. (2005). "Unified functional tolerancing approach for precision cylindrical components". *International Journal of Production research*, 43, pp.25-47.

Zhang, Y., He, Z., Zhang, C., & Woodall, W. H. (2013). Control Charts for Monitoring Linear Profiles with Within-Profile Correlation Using Gaussian Process Models. *Quality and Reliability Engineering International*.