euspen's 19th International Conference & Exhibition, Bilbao, ES, June 2019
www.euspen.eu

A new Gaussian process-based approach for uncertainty propagation in surface metrology profile parameters estimation

Olusayo Obajemu[1], Mahdi Mahfouf[1], Thomas E McLeay[2], Xiang Jiang[3], Paul J Scott[3] and Visakan Kadirkamanathan[1]

[1]*Department of Automatic Control and Systems Engineering, The University of Sheffield, U.K.*
[2]*Advanced Manufacturing Research Centre with Boeing, The University of Sheffield, Rotherham, U.K.*
[3]*Centre for Precision Technologies, University of Huddersfield, U.K.*

o.obajemu@sheffield.ac.uk

Abstract

Surface texture parameters are important indicators for understanding and controlling manufacturing processes. Deriving these parameters is however beset by ambiguities and uncertainties. The inclusion of confidence bands should provide valuable information on the reliability of the derived surface parameters. Existing methodologies of uncertainty modelling assume non-random interpolation functions, which do not adequately allow for the inclusion of interpolation uncertainties. This paper presents a new approach based on Gaussian processes, including a mechanism for the derivation and inclusion of such interpolation-based uncertainties when calculating surface texture parameters. The interpolation-based uncertainties are assumed to be independent of measurement uncertainties and as such, they can be independently modelled and propagated onto the derived parameters. When tested using real machining surface data, validation results show that the newly proposed technique has the advantage over the ISO-based approach of systematically characterising interpolation-based uncertainties in the form of confidence bands in the estimated profile parameters.

Keywords: Surface metrology, profile parameters, uncertainty, kriging

1. Introduction

Surface texture parameters are important functionality indicators of manufactured components. Their measurements and utilisation provide a number of useful benefits such as helping to understand the manufacturing process, controlling the manufacturing process and providing an assessment of the quality of the manufactured components [1]. Indeed, surface metrology can be viewed as the fingerprint of the whole manufacturing process whereby if there is a change in the manufacturing process settings, such as personnel and/or tool changes, then this will be reflected somewhere on the work-piece. It is for this reason that surface metrology has been an integral part of components manufacturing. There are two groups of surface texture parameters: the profile parameters, which have been standardised according to the ISO 4287 [2] and the areal parameters, which is standardised in the ISO 25178 [3]. These ISO documents define a comprehensive list of important surface metrology parameters as well as provide procedures and formulas for calculating these parameters. The process of calculating the parameters in practice involves making assumptions which makes this process inherently uncertain. For example, definitions of many of the parameters (such as the arithmetic mean deviation - Ra) involve the use of mathematical integration which implicitly assumes that the surface from which parameters are to be derived is continuous. Usually however, the surface data is recorded at discrete intervals and it is typical to fit a continuous surface to the surfaces after which some form of analytical or numerical integration method is used to calculate the parameters [4]. For example, the widely used Ra parameter is given by the following formula:

$$Ra = \frac{1}{l}\int_0^l |z(x)| \qquad (1)$$

where l is the sampling length and $z(x)$ is the height at location x.

There exists a plethora of approaches that can be used for calculating these parameters. A common approach proposed in [4] involves using cubic splines for the interpolation problem. The advantage of using the cubic spline interpolant is that it facilitates analytically calculating the integral such as the one defined in Equation (1). One major drawback of this approach is that such a spline interpolation approach does not adequately account for the fact that the values between interpolated points are not observed directly. In other words, existing methodologies do not allow for characterising the uncertainties because the data is only observed at discrete points. The method of constant uncertainty interval over the interpolating formula was proposed in [5]. In this approach, data points on the surface are assumed to be bounded by a constant uncertainty interval. A formula based on the Guideline for uncertainty propagation (GUM) [6] is then utilised to calculate and produce and uncertainty distribution in the surface profile parameter values. The use of uncertainty intervals in the interpolation function can be useful for understanding the distribution of the parameter uncertainties and can provide an explanation as to why there are sometimes ambiguities in the performance of the manufactured components even though the same profile parameter values were observed [6]. However, using constant bounds as uncertainty interval does not adequately and systematically reflect the data distribution. Indeed, one would expect that the locations where data are observed should include lower uncertainties than the locations where data are not observed. Statistical interpolation techniques provide a means of systematically including such types of uncertainty intervals. This paper proposes to use an interpolation approach based on the Gaussian process regression approach. Gaussian processes have been extensively utilised for characterising uncertainties in machine learning [7] and geostatistics [8]. Gaussian process regression, otherwise known as kriging, provides a powerful framework for interpolation and for determining interpolation uncertainties. This approach also allows for including uncertainties due to other errors (such as measurement error). These uncertainties can then be propagated in calculating the surface profile parameter values using the GUM approach as will be discussed. To the best of the authors' knowledge, this is the first time such an approach has been applied to calculate surface texture parameters. The rest of the paper is organised as follows: section 2 introduces the relevant surface metrology terms as well as the theoretical foundations of Gaussian process regression which will henceforth be known by the geostatistics name – Kriging. The section also briefly discusses the GUM uncertainty propagation framework. Section 3 discusses the surface metrology datasets used in the paper which include synthetic and real data sets. Section 4 presents the results and section 5 concludes the paper and provides suggestions for future work.

2. Surface Texture Parameters and Kriging

2.1. Surface Parameters

This work will be concerned with calculating the Ra surface profile parameters. The Ra parameter is the mostly widely used surface profile parameter in academia and industry [1] where it is typically utilised to predict work-piece function. It should be noted however that the proposed methodology readily extends to other surface texture parameters. As already mentioned, calculating the parameters typically involve a two-step procedure. The first step is to fit a continuous function to the measured surface data and the second step involves using some form of analytical or numerical scheme to solve the integral equations (for example Equation (1)) in order to derive the relevant surface profile parameters [4]. For example, the study performed in [4] used the cubic spline interpolant to fit the surface profile data which facilitates analytical computation of the integrals. The approach also utilised a constant uncertainty bound on the interpolation function where the degree of the uncertainty bound is determined by an expert or from instrument limitations. However, as already mentioned, approaches which use non-random interpolants do not allow for systematically characterising the interpolation uncertainties in the distribution of the parameters as they fail to address the fact that the uncertainties in places with sparse data distribution are expected to be lower than in places with a denser data distribution. A framework that allows for a systematic characterisation of uncertainties involved in interpolation is Kriging. Kriging can be interpreted as a Bayesian framework and is discussed briefly in the next section.

2.2. Kriging

Kriging is a statistical interpolation technique. Given a set of points $X = [x_1, x_2, \dots, x_N]$, $x_i \in R^D$ and a set of observations $Z = [z_1, z_2, \dots, z_N]$, $z_i \in R$, kriging interpolation allows to find a function $f: R^D \rightarrow R$ such that $f(x_i) = z_i$ for $i = 1, 2, \dots N$. Additionally, the kriging formula also provides a second function $\sigma(x)$ which expresses the uncertainty involved in the interpolation process. For an arbitrary point x^*, the kriging formula gives the best linear unbiased predictor ($f(x^*)$) as well as the degree of uncertainty $\sigma(x^*)$. The uncertainty distribution can then be propagated to any quantity derived from the interpolation functions. In the case of surface metrology, x_i refers to a sampled location and z_i refers to the height at the ith point. The covariance function is central to the kriging methodology and provides a degree of correlation between data points across the domain. It is worth noting that a zero-mean function as well as the squared exponential covariance function was utilised in this paper. This function defines the prior as follows:

$$\begin{gathered} m(x) = 0 \\ \operatorname{cov}\left(f(x), f(x')\right) = k(x, x') = \exp\left(-\frac{1}{2}|x - x'|^2\right) \end{gathered} \tag{2}$$

where $m(x)$ is the mean function, which is assumed to be zero and $\operatorname{cov}\left(f(x), f(x')\right)$ is the covariance function.

It can be shown that the prediction $f(x^*)$ for an arbitray point x^* is a Gaussian distribution defined by the following equation:

$$\begin{gathered} p(f(x_*)|x_*, D) = \mathcal{N}(\mu_*, \Sigma_*) \\ \mu_* = k(x_*, x)k(x, x)^{-1}z \\ \Sigma_* = k(x_*, x_*) - k(x_*, x)k(x, x)^{-1}k(x, x_*) \end{gathered} \tag{3}$$

It is worth noting that the covariance function contains hyperparameters which is to be estimated. The estimation was carried-out using one-fold cross validation – a procedure discussed in [9]. Additionally, as the paper is only concerned with interpolation uncertainties, the measurement uncertainties were not included in the kriging formulas (Equation (3)). However, extending the approach to include measurement uncertainties is straightforward as shown in [7].

2.3. Propagating the Uncertainties

When a kriging model has been found, numerical integration was used to perform the calculation of the parameters. In particular, the data sets are sampled (M number of points) and then integrated using these points. Typically M>>N for an accurate estimation. For a 2 mm sampling length, the value of M = 2 000 was used in this paper. For each of these sampled points, Equation (3) provides the formula for calculating uncertainty interval. This uncertainty is then propagated when calculating the parameter as defined by GUM. The GUM framework is shown in Fig. 1.

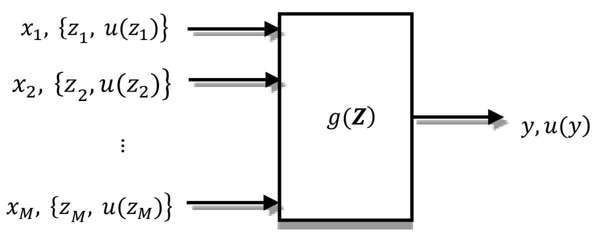

Figure 1. GUM propagation framework allows for propagating the uncertainty involved in an input quantity to an output quantity.

It should be emphasized that vector $\boldsymbol{Z} = [z_1, z_2, \dots, z_M]$ represents the observed data at locations $\boldsymbol{X} = [x_1, x_2, \dots, x_M]$, $u(y)$ defines the uncertainty (typically a probability distribution or any other measure of uncertainty) of the derived quantity. $g(\boldsymbol{z})$ is the formula for calculating the surface parameter of interest such as the Ra as defined in Equation (1) . It can be shown (according to the GUM document [6]) that the following set of equations can be used to calculate the propagated uncertainty distribution $u(y)$:

$$u(y) = \int u(\boldsymbol{Z})\delta\big(y - g(\boldsymbol{Z})\big)d\boldsymbol{Z}$$
$$\bar{y} = \int y u(y)dy = \int u(\boldsymbol{Z})g(\boldsymbol{Z})d\boldsymbol{Z} \quad (4)$$
$$varr(y) = \int (y - \bar{y})^2 u(\boldsymbol{Z})d\boldsymbol{Z} = \int u(\boldsymbol{Z})(g(\boldsymbol{Z}) - \bar{y})^2 d\boldsymbol{Z}$$

where $\bar{y}$ is the mean of the uncertainty distribution and $varr(y)$ its variance. Equation (4) above defines the procedure for deriving the uncertainties involved in calculating the surface profile parameters.

3. Datasets

Two data sets were investigated in this paper. The first data set relates to a synthetic data defined in [5]. The data points are generated from a sinusoidal function ($f = \sin(x)$). Two sets of data samples (15 and 19 samples) were taken from the synthetic data set to investigate how the number of data points affect the uncertainty distribution. The sampled data points are used to calculate the surface profile parameters after using the kriging interpolant. The function from which the synthetic data was generated facilitates analytical integration so that the real values of the parameters to be calculated are known in advance. This will provide a means to compare the calculated parameters (from the data points) with the real values in order to validate the proposed approach. Note that such a comparison cannot be carried-out in real data sets, as it would be impossible to know what the real values of the parameters are. The synthetic function as well as data points are shown in Fig 2.

The second data set is derived from a real surface from a machining experiment. A surface texture of about 2 mm x 2 mm was measured. As this paper is concerned with only profile measurements, a profile was extracted from this areal measurement. The profile was filtered with a Gaussian filter to remove long scale components. The extracted profile is shown in Fig. 3.

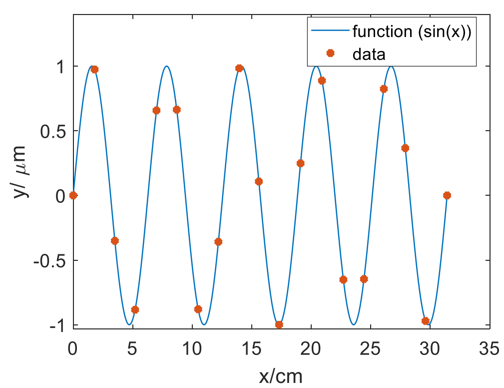

Figure 2. Synthetic Data.

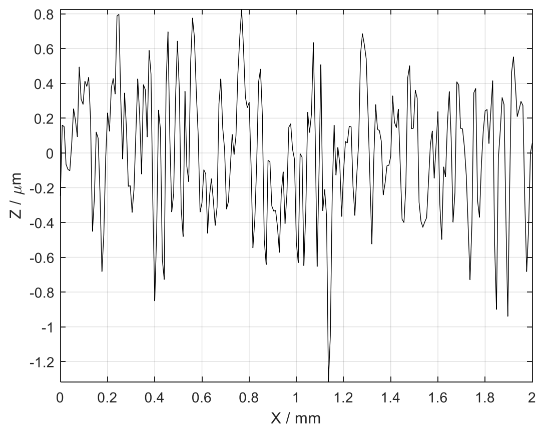

Figure 3. Real Milling Data set (Gaussian filtered).

4. Results

Table 1 shows the results of applying the proposed framework on the two data sets. As can be seen, results from the proposed approach were compared with the spline technique from [5] for both sets of data. It can be seen from Table 1 that the proposed approach is able to accurately determine the correct value of the profile parameter.

Table 1. Comparison of results for both real and synthetic data sets. The numbers represent the Ra values (µm) for the data sets.

		Proposed	Spline [5]	Real Value
Synthetic Data set	15 samples	0.6366	0.6366	0.6366
	19 samples	0.6366	0.6366	0.6366
Real Data set		0.5643	0.5870	NA

In particular, in the case of the synthetic data set, both approaches are able to give good accuracy especially for two different number of sampled data points. The proposed Gaussian process approach is however able to accurately capture the uncertainties embedded in the interpolation function as can be seen in Fig. 4. The uncertainty intervals increase when there are less data as shown in Fig. 5. This result should be compared with the constant uncertainty interval approach shown in Fig. 6. In the proposed approach, it can be seen that the uncertainty is high when there is a sparsity of data around a region and vice versa. This uncertainty across the data domain is then propagated through in the calculation of the surface parameters using the GUM approach as described in Section 2.3 which consequently provides a characterisation of the interpolation uncertainties when deriving the particular parameter.

In the case of the real data set, the interpolation function as well as the data points are shown in Fig. 7. The uncertainty bounds are much lower than those obtained from the synthetic data set because of the high data density (256 data points for a 2 mm evaluation length). The propagated uncertainty is shown in Fig. 8. This distribution provides the interpolation uncertainties involved in the calculation of the Ra value.

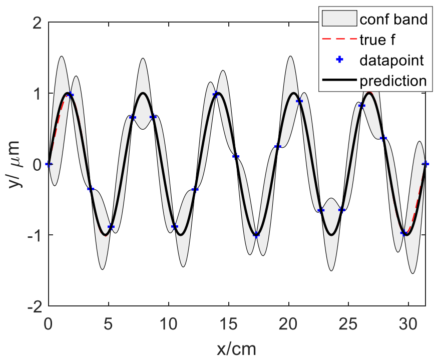

Figure 4. Interpolating function and uncertainty interval across the data domain for the synthetic function. Nineteen (19) data points were sampled from the true function.

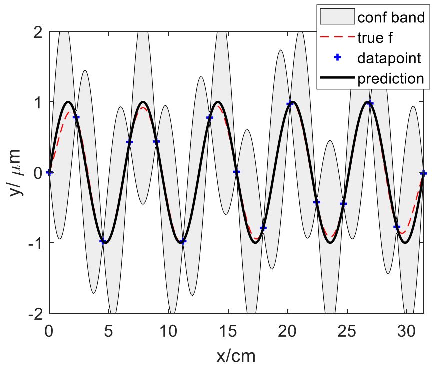

Figure 5. Interpolating function and uncertainty interval across the data domain for the synthetic function. Fifteen (15) data points were sampled from the true function.

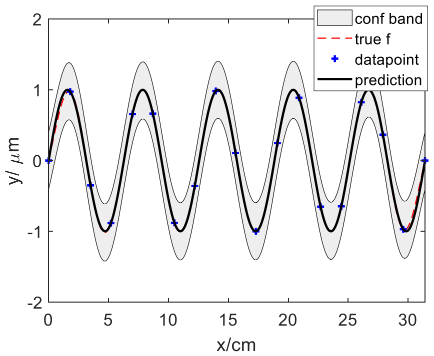

Figure 6. Constant uncertainty interval proposed in an existing study [5]. Note that this approach does not systematically provide a characterisation of the interpolation uncertainties

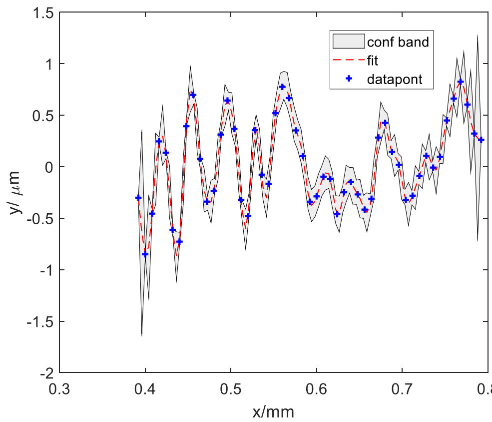

Figure 7. Interpolation and uncertainty band on the real milling data set (zoomed).

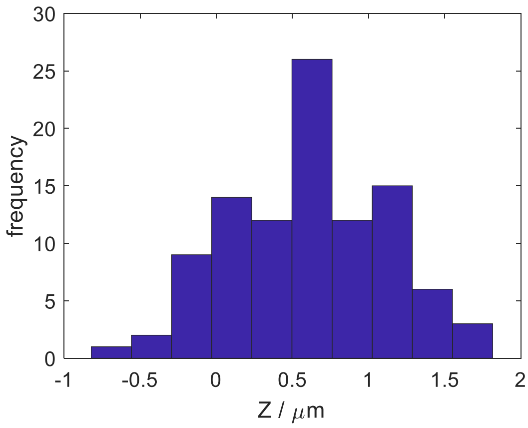

Figure 8. Propagated uncertainty in the calculation of the Ra parameter for the synthetic data set with 19 sampled point. Note Z represents the Ra parameter.

5. Conclusion

The paper has presented a new approach for characterising the uncertainty embedded in the calculation of surface profile parameters. The approach, which involves using a statistical interpolation technique based on the Gaussian process, provides a more systematic approach for modelling the interpolation-based uncertainties. The proposed technique is better than existing methodologies which typically use non-random interpolants and/or constant uncertainty bands. The proposed approach was tested on a synthetic as well as real machining data sets. Results proved that not only are the profile parameter values able to be correctly extracted, the proposed approach can correctly provide a more credible measure of the uncertainties.

References

[1] Whitehouse, D J, 2010. *Handbook of surface and nanometrology*. CRC press.

[2] International Organization for Standardization. Technical Committee ISO/TC 57, 1997. ISO 4287. Geometrical Product Specifications (GPS). Surface Texture: Profile Method. Terms, Definitions and Surface Texture Parameters

[3] ISO 25178-2, 2012. Geometrical product specifications (GPS)– surface texture: areal–part 2: terms, definitions and surface texture parameters.

[4] Brennan J K, Crampton A, Jiang X, Leach R and Harris P, 2005 vol 13 No 1, Approximation of surface texture profiles. *In Journal of Physics*: Conference Series, IOP Publishing, p 264.

[5] Brennan, J K, Crampton A, Jiang X, Leach R K and Harris P M, 2006. Propagation of uncertainty in discretely sampled surface roughness profiles. *In Advanced mathematical and computational tools in metrology* VII, p 271-75.

[6] ISO I and OIML B, 1995. Guide to the Expression of Uncertainty in Measurement. Geneva, Switzerland.

[7] Rasmussen C E, 2004. Gaussian processes in machine learning. *In Advanced lectures on machine learning*. Springer, Berlin, Heidelberg, p 63-71.

[8] Thimmisetty C A, Ghanem R G, White J A and Chen X, 2018. High-dimensional intrinsic interpolation using Gaussian process regression and diffusion maps. *Math.l Geosciences*, 50(1), p 77-96.

[9] Obajemu O, Mahfouf M and Catto J W, 2018. A New Fuzzy Modeling Framework for Integrated Risk Prognosis and Therapy of Bladder Cancer Patients. *IEEE Trans. on Fuzzy Systems*, 26(3), p 1565-77.