12th World Congress on Structural and Multidisciplinary Optimization
5th - 9th, June 2017, Braunschweig, Germany

Characterization of Geometric Uncertainty in Gas Turbine Engine Components using CMM Data

Jennifer Forrester[1], Andy Keane[2]

[1] University of Southampton, England, jennifer.forrester@soton.ac.uk
[2] University of Southampton, England, Andy.Keane@soton.ac.uk

1. Abstract

Measurements of component geometry are routinely made for inspection during manufacturing. Typically this results in 'clouds' of points or pixels depending upon the measuring system. Examples include points form laser-based or touch-probe co-ordinate measuring machines (CMMs). The point density may vary as will the cost and time taken to make measurements. There can also be gaps and occlusions in data, and sometimes it is only practical to collect sparse sets or points in a single dimension.

This data often provides an untapped source of quantitative uncertainty information pertaining to manufacturing methods. It is proposed that state-of-the-art uncertainty propagation and robust design optimization approaches, often demonstrated using assumed normal input distributions in existing parameters, can be improved by incorporating these data. Inclusion of this information requires, however, that the point cloud be converted to an appropriate parametric form.

Although the design intent of a component may be described using simple geometric primitives joined with tangency or at vertices, manufactured geometry may not exhibit the same simple form, and line and surface segment end locations are notoriously difficult to locate where there is tangency or shallow angles. In this paper we present an approach to first characterise point cloud measurements as curves or surfaces using Kriging, allowing for gaps in data by extension to universal Kriging. We then propose a novel method for the reduction of variables to parameterize curves and surfaces again using Kriging models in order to facilitate practical analysis of performance uncertainty. The techniques are demonstrated by application to a gas turbine engine blade to disc joint where the contact surface shape is measured and the notch stresses are critical to component performance.

2. Keywords: point-cloud, variable reduction, uncertainty propagation, probability distribution, Kriging.

3. Introduction

To investigate the effects of irreducible geometric uncertainty on component performance and to make design decisions based on these effects, we turn to uncertainty propagation techniques (such as Monte-Carlo sampling [14], sparse quadrature [15], and Monte Carlo sampling of response surfaces (also known as *emulators* or *metamodels*) [6]), and robust design optimization (RDO) or reliability based design optimization (RBDO) (typically using global search routines and often also employing emulators)*.

Robust design optimization, based on the definition by Park et al. [11], is formulated as the multi-objective problem: find $\mathbf{v} \in \mathbb{R}^n$ to minimise a combination of $\mu_f(\mathbf{v},\mathbf{p})$ and $\sigma_f(\mathbf{v},\mathbf{p})$, subject to the constraints $g_j\left(\mathbf{v}+\mathbf{z}_\mathrm{v},\mathbf{p}+\mathbf{z}_\mathrm{p}\right) \leq 0$ where $j = 1,...,r$, and $\mathbf{v}_\mathrm{L} \leq \mathbf{v} \leq \mathbf{v}_\mathrm{U}$.

$\mathbf{p}$ are 'uncertain' or 'uncontrollable'; they contain irreducible uncertainty (often described as 'noise') which may be distributed normally or otherwise. $\mathbf{v}^\dagger$ are design variables, which can also be uncertain. μ_f and σ_f are the mean and standard deviation of the performance function f, respectively, and are given by the well-documented definitions (for example [10]):

$$\mu_f = \mathrm{E}[f(\mathbf{u})] = \int_{-\infty}^{+\infty} f(\mathbf{u})p(\mathbf{u})\mathrm{d}\mathbf{u}, \tag{1}$$

and

$$\sigma_f{}^2 = \mathrm{E}[(f(\mathbf{u})-\mu_f)^2] = \int_{-\infty}^{+\infty} [f(\mathbf{u})-\mu_f]^2 p(\mathbf{u})\mathrm{d}\mathbf{u}, \tag{2}$$

where p is the *joint probability density function* of the variables $\mathbf{u}$ which is a combined vector of design $\mathbf{v}$ and $\mathbf{p}$.

*Here we refer only to 'black-box' processes, but the difficulty associated with ensuring accuracy of input uncertainties is just as relevant to 'intrusive' approaches.

†It is typical to use $\mathbf{x}$, but to avoid confusion with $\mathbf{x}$ from a measurement set, we refer to the design variable vector as $\mathbf{v}$.

For any study to be accurate and not misleading, the input uncertainties must be correctly modelled, i.e., the correct distributions must be attributed to them [9] and any correlations accounted for. In addition, UP, RDO, and RBDO routines are hampered and often deemed entirely impracticable by the number of variables (the length of the vectors **v** and **p**); the 'curse of dimensionality' [3] when building models of the space of interest. The flowchart in figure 1 illustrates the process, simplified to a single objective, and uncertainty only in **p** (not **v**). The performance analysis code is typically expensive and in order to perform robust design it must be evaluated at numerous points within the design and uncertain spaces. Typically, the methods mentioned above require evaluation of performance at carefully chosen points, e.g., sparse grids [7] or Latin hypercubes [8], in order to build models. The exception to this is Monte Carlo, which requires untenably large samples in order to provide accurate variance predictions.

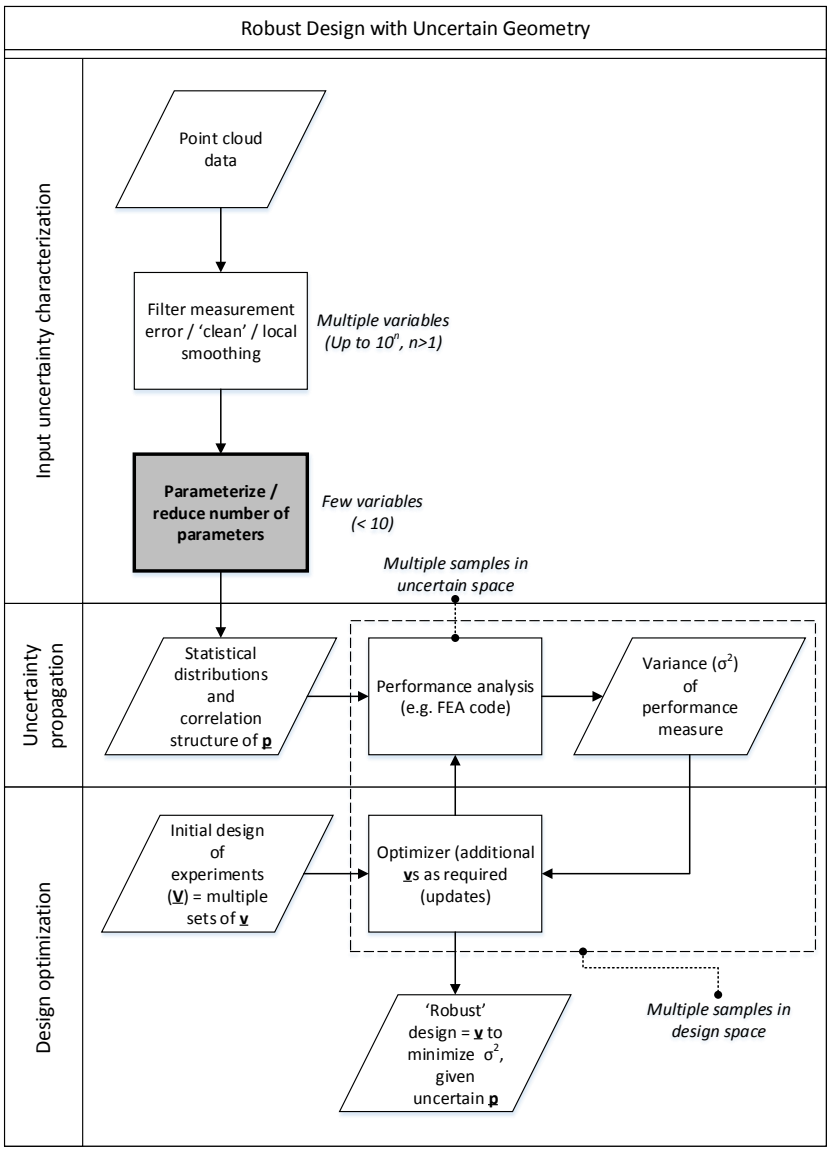

Figure 1: The proposed robust design optimization (RDO) workflow

We present an approach where, given a set of measurement data, Kriging can be used to parameterize a generic curve with a reduced number of parameters (**p**) from which pseudo-curves can be generated at specified sample points for UP and RDO in the face of irreducible geometric uncertainty.
We begin with a description of Kriging and the salient points relevant to geometric fitting in the context of uncertainty. We then present a brief summary of the Iman-Conover method for sampling to maintain correlation. This paper then takes an example of a nominally straight line that represents a contact surface on a turbine blade dovetail or firtree joint and demonstrates that, by fitting (to remove measurement noise) and then refitting a Kriging model

using the approach described, a useful reduction in variables can be achieved. The defining variables of the refitted curve are the correlated y locations of a number of points fixed in x, and hypothesis tests are used to evaluate the supposition of normality of these variables at a chosen confidence level.
It is shown that an initial sample of 36 measurement points can be reduced to 8 evenly distributed points, from which a 'pseudo-sample' of any size can be generated with a given distribution or within a given range.

4. Kriging; an Overview

4.1. Ordinary Interpolating Kriging

Kriging is an approach to fitting models to data points which uses a specific form of *radial basis function* [1, 3, 4]; a sum of basis functions (ψ). We consider the bases to be at the sample points and in ordinary Kriging each point is taken to be a realization of a normally distributed random variable:

$$\mathbf{y} \sim \mathcal{N}(\boldsymbol{\mu}, \sigma^2\boldsymbol{\Psi}). \tag{3}$$

In other words, any points on the curve or surface are random variables with mean $\boldsymbol{\mu}$ and covariance $\sigma^2\boldsymbol{\Psi}$ (correlation matrix $\boldsymbol{\Psi}$). The correlation matrix assumed to take the form of a Gaussian kernel function, and as such is made up of the elements:

$$\psi_{ij} = \exp\left(-\sum_{k=1}^{d} \theta_k |x_k^{(i)} - x_k^{(j)}|^{p_k}\right), \tag{4}$$

where d is the number of dimensions and θ_k and p_k are constants. $\theta_k \geq 0$ is a constant related to how rapidly values change (usually given a value between 10^{-3} and 10^2 [3]) and $0 < p_k \leq 2$ determines the smoothness of the model (often fixed at $p = 2.00$ (or slightly lower at $p = 1.99$ for numerical stability [13])). Here, we simplify things by setting $p = 1.99$, and consider only a single dimension. Therefore, $\boldsymbol{\Psi}$ is a function of the sample point locations $\mathbf{x}$ and θ only:

$$\psi_{ij} = \exp\left(-\theta |x_i - x_j|^{1.99}\right). \tag{5}$$

In general, the *Kriging predictor* (the function estimate $\hat{y}$ at any new point $\mathbf{x}$ is the the maximum likelihood estimate (MLE) and is expressed as:

$$\hat{y}(\mathbf{x}^*) = \hat{\mu} + \boldsymbol{\psi}^T\boldsymbol{\Psi}^{-1}(\mathbf{y} - \mathbf{1}\hat{\mu}), \tag{6}$$

where estimates for μ and σ^2 in terms of $\boldsymbol{\Psi}$ can be found analytically to be:

$$\hat{\mu} = \frac{\mathbf{1}^T\boldsymbol{\Psi}^{-1}\mathbf{y}}{\mathbf{1}^T\boldsymbol{\Psi}^{-1}\mathbf{1}}, \tag{7}$$

and

$$\hat{\sigma}^2 = \frac{(\mathbf{y} - \mathbf{1}\hat{\mu})^T\boldsymbol{\Psi}^{-1}(\mathbf{y} - \mathbf{1}\hat{\mu})}{n}. \tag{8}$$

The remaining constant θ must then be determined by using some form of numerical optimization routine (usually a global search such as a genetic algorithm is employed followed by a terminal gradient descent search) to determine the value for which the *negative ln-likelihood* is minimized. The likelihood can be expressed as the *concentrated ln-likelihood function*:

$$\ln(L) \approx -\frac{n}{2}\ln(\hat{\sigma}^2) - \frac{1}{2}\ln(|\boldsymbol{\Psi}|). \tag{9}$$

The process employed to search for θ is termed *training*, and implies that θ is in fact directly determined by the other parameters σ^2 and $\boldsymbol{\Psi}$, and in turn μ, $\mathbf{x}$, and $\mathbf{y}$.

4.2. Regressing Kriging

Regressing Kriging extends this concept to allow for noise in $\mathbf{y}$ by representing it as an additional constant parameter λ, added to the leading diagonal of $\boldsymbol{\Psi}$ [4]. λ is related, in our case, to the measurement error, which should be known. Picheny et al. [12] suggest adding the noise as the true variance τ^2 to the leading diagonal of the covariance matrix $\sigma^2\boldsymbol{\Psi}$ instead. The λ parameter becomes another constant, like θ, to be found by minimization. This is also true for τ^2, however, there is no-longer an analytical solution for σ^2, which becomes a third constant to find during the procedure. This method is useful as it allows us to specify the level of regression based on a known measurement variance value.

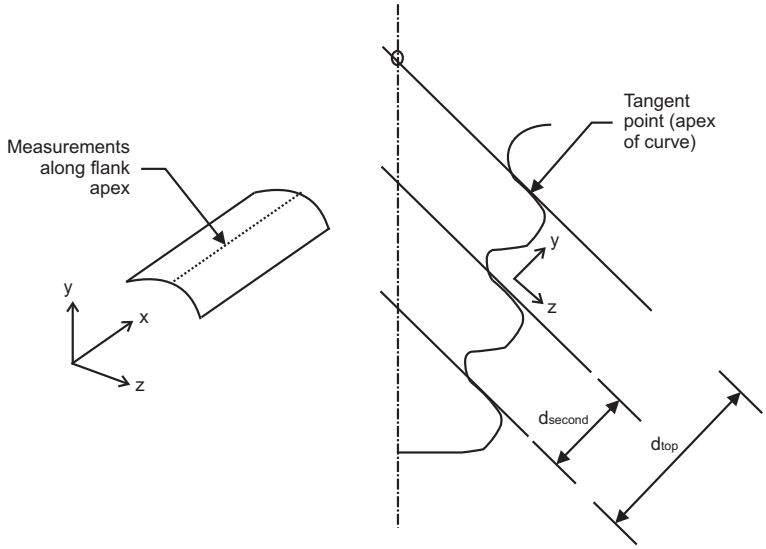

Figure 2: Schematic showing location of measurement sets on turbine blade 'firtree' joints

4.2. Sampling for Kriging in the measurement space
The response surface methodology, of which Kriging is just one approach, is founded on the requirement to model multi-dimensional spaces accurately in regions of interest using as few sample points as is possible. These methods, in optimization, typically make use of a space-filling sampling strategy such as an *optimal Latin hypercube* to create an initial *design of experiments* (DOE). A global model of the space is then constructed using this DOE, and further samples (updates) are taken in regions of interest, such as those where the expectation of finding a better optimum are high (using, for example, a criterion based on *expected improvement* (EI)).
In the case of fitting measurement data, there are a few key differences:

- we wish to fit multiple curves models using the same variables (points),
- we are only concerned with models in one and two dimensions, and
- we are not interested in refining the model in a specific area (or if we are, then we don't yet know which area is of greatest importance), but in getting the best representative overall shape.

5. Sampling with Correlation; the Iman-Conover Approach

The Iman-Conover method for inducing rank correlation among samples [5] is widely used for Monte Carlo (MC) sampling and is independent of distribution type. The procedure takes a target rank correlation $\mathbf{C}^*$ (in this case, the rank correlation of the points in our measurement data set). If $\mathbf{C}$ (the rank correlation of our sampled set) is equal to $\mathbf{C}^*$, if $\mathbf{P}$ is a lower triangular matrix that results in $\mathbf{PP}' = \mathbf{C}$, and if our sample set $\mathbf{R}$ is an n×k matrix with rows $\mathbf{R}_i$ ($i = 1, ...n$), then $\mathbf{R}_i\mathbf{P}$ is a vector with desired correlation matrix $\mathbf{C}$ and $\mathbf{RP}' = \mathbf{R}^*$ (a matrix with the same distribution in its rows as $\mathbf{R}_i\mathbf{P}'$). Therefore, the rank correlation of $\mathbf{R}^*$ should be close to $\mathbf{C}$.
Calculating $\mathbf{P}$ by Cholesky factorization and re-ordering $\mathbf{R}$ as the corresponding column of $\mathbf{R}^*$ results in rank correlation of the input sample set $\mathbf{R}$ that is approximately equal to $\mathbf{C}$.

6. Refitting Pseudo-curves to Measured Data

6.1. Local 'Cleaning' of Data using Regressing Kriging
Measured data sets tend to take the form of 'point clouds' in Cartesian space. The data sets may contain gaps and are not always composed of evenly spaced points. The larger the point cloud, and the greater the precision of the instrument used, the more complex, but potentially more accurate, any fitted curve or surface becomes. The example set used in this study (figure 2 and 3) represents a nominally straight line. There is also a 'gap' in the data where a fixture held the part.
There are various possible approaches that could be employed to 'clean' the raw data. We have used a regressing Kriging fit [12], specifying the measurement noise, and we have extended this to Universal Kriging [2, 4] to main-

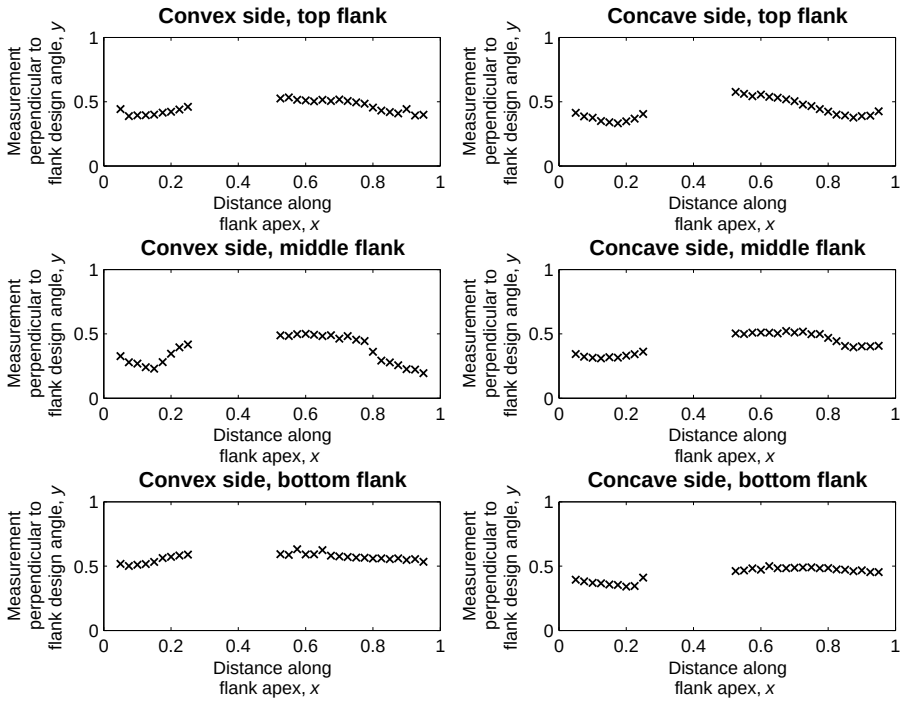

Figure 3: Example data set for a single blade (six flanks) (normalized, nominal flank line at y=0.5)

tain the general trend across the 'gap'[‡].

6.2. Variable Reduction: Equally Spaced Points

Using the approach described above, the resulting curves for each of the six flanks can be plotted on the same axes (figure 4 shows fitted curves, seven blades only for clarity). It is possible to observe that there are distinct trends in the data for each of the flanks. Given that one can observe these trends 'by eye', one can hypothesize that it should be possible to represent these curves and their variation (uncertainty) mathematically.

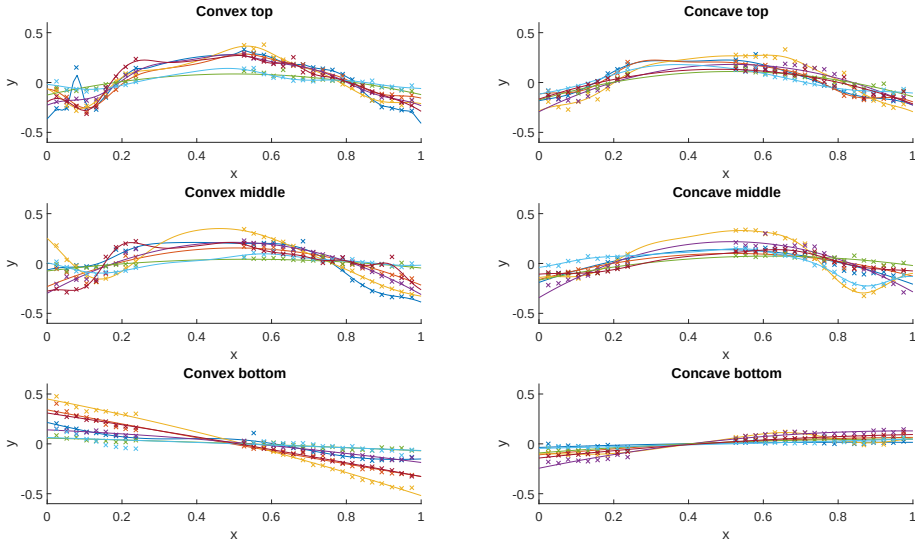

Figure 4: Measured curves (seven blades), fitted using universal regressing Kriging

It is noted at this point that one can identify that the universal Kriging technique used for fitting may be introducing a 'dip' in the region of missing data, which may bot be 'real' (a 'relic' of the 'cleaning' process, which is not the

[‡]This methodology may have its own flaws but is not the focus of this paper, where it is assumed that the 'cleaned' curve is 'correct'.

focus of this paper). For this study we therefore continue with the fitted curves assuming that they are a true representation of the shape of the flank. Given that not all measurement point clouds exhibit large 'gaps', as a generic process for representing point cloud data this approach is acceptable.
The proposed process for parameterizing the curves to enable them to be sampled and 'pseudo-curves' created is described by the flowchart in figure 5. We then show how this process has been applied to the curves from the convex side top flank of the 'firtree'.

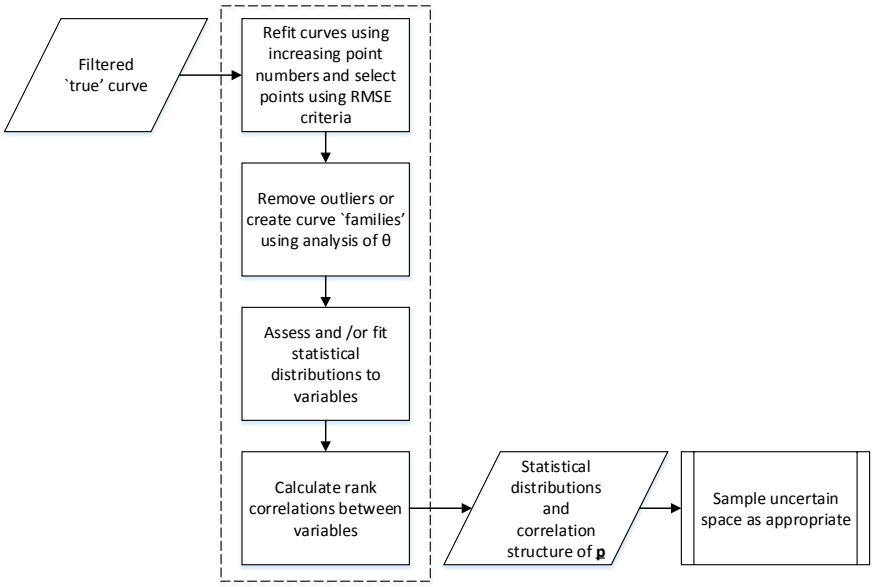

Figure 5: The proposed parameterization for uncertainty workflow

The measured curves are refitted using interpolating Kriging by evaluating the Kriging predictor at increasing numbers of evenly distributed points. The convergence of the curve fit is evaluated over the full set of 46 blades and 6 flanks. An example of the convergence of a single blade is shown in figure 6.

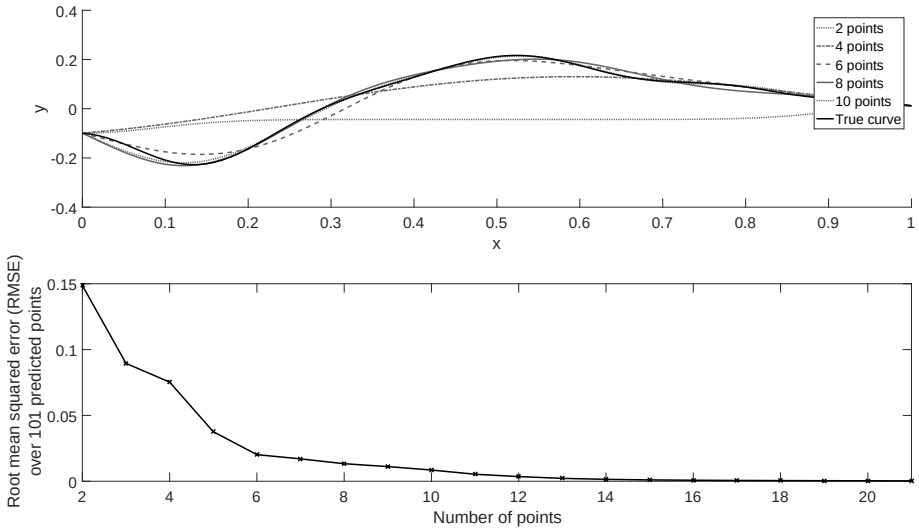

Figure 6: Convergence of curve refit using increasing numbers of evenly distributed points for a single blade firtree, convex side top flank

A convergence criteria can be set as, for example, RMSE$<$0.05. To specify the number of points required, we select the number of points for which 99.87% of curves reach the convergence criteria, i.e., curves requiring greater than three standard deviations above the mean number of points are discounted. This results in a requirement for 8 and 7, 9 and 8, and 4 and 4 points for the top, middle, and bottom flanks respectively. The refits based on this number of points are plotted for the convex side top flank (figure 7).

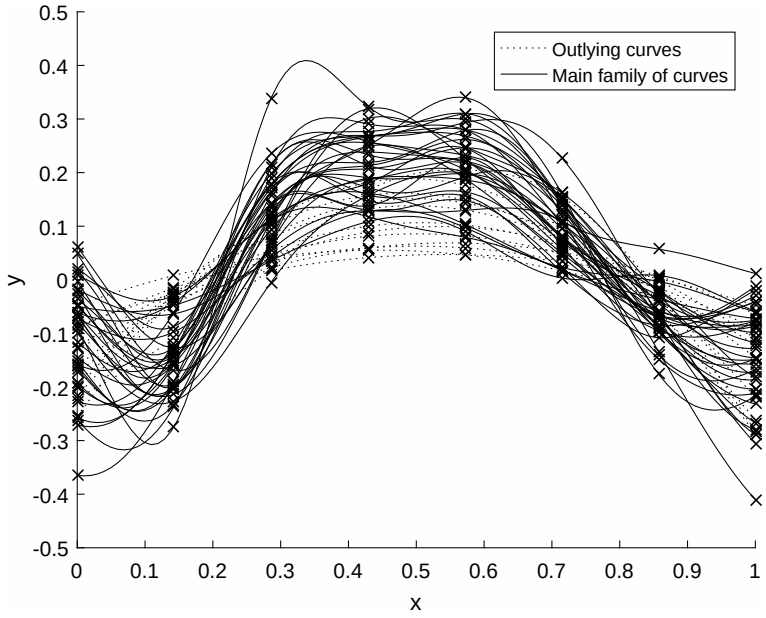

Figure 7: Measured firtree flank (convex top) curves refitted using 8 evenly distributed points (46 blades)

Plotting histograms of the eight sets of y-values at the points (figure 8) reveals an impression of their distributions. A histogram of the fitted θs can also be plotted (figure 9). It is clear that there are a number of 'outlying' curves, where the θ value indicates curves with a different underlying shape (values changing along x at a significantly different rate, or 'smoothness'). These are shown in figure 7 and removed for sampling. However, where there are significant numbers this could suggest more than one distinct 'family' of curves to be sampled. Chi-square significance tests have been conducted to test the hypothesis that the samples of y have been drawn from a normal distribution with mean and standard deviation estimated by the sample values. The hypotheses cannot be rejected at the 5% level for any of the 8 point locations.

6.3. Variable Reduction: Locating Principal Points

One might suggest that the point distribution is most likely to define the cuves with the smallest possible error if, instead of selecting evenly distributed x-values, we choose points that define the curve. In particular, it could be preferable to select the points of peak curvature and inflection. Initial investigations have proven that the RMSE for the refit of individual curves can be reduced, or the number of required points reduced slightly. However, the benefit is then lost when selecting the x-values at which to fix to refit all of the curves and determine distributions of ys (the inflection points won't be in the same x locations so need to be averaged in some way). Further investigation into fitting and sampling from a bi-variate distribution in (x,y) is being considered.

6.4. Creating Pseudo-curves for Uncertainty Propagation (Sampling)

Given a set of x locations, with distributions of y, it is now possible to sample y values as required. It is clear, however that there is not only correlation between the x-values as defined by $\mathbf{\Psi}$, but also in the values of y. The Iman-Conover approach to sampling with correlation is used. To demonstrate the approach, we show a set of 30 'pseudo-curves' sampled randomly from the y distributions (figure 10). The correlation structure is shown using the scatter plot, 11. This set is used for demonstration purposes, but one could also sample using some kind of space-filling algorithm over the range of ys, thereby enabling models of the uncertain space to be made.

To demonstrate that the process works wherever the points are chosen, the process is repeated with a ninth point. The 'pseudo-curves' are plotted in figure 12 and correlation structure shown in figure 13.

7. Review and Future Work

In this study we have demonstrated that the number of variables used to define a Kriging curve, fitted from measurement data, can be reduced to a much smaller set of uncertain variables (the locations of a reduced set of y-values). It was shown that the hypothesis that the variables in the test data were drawn from normal distributions

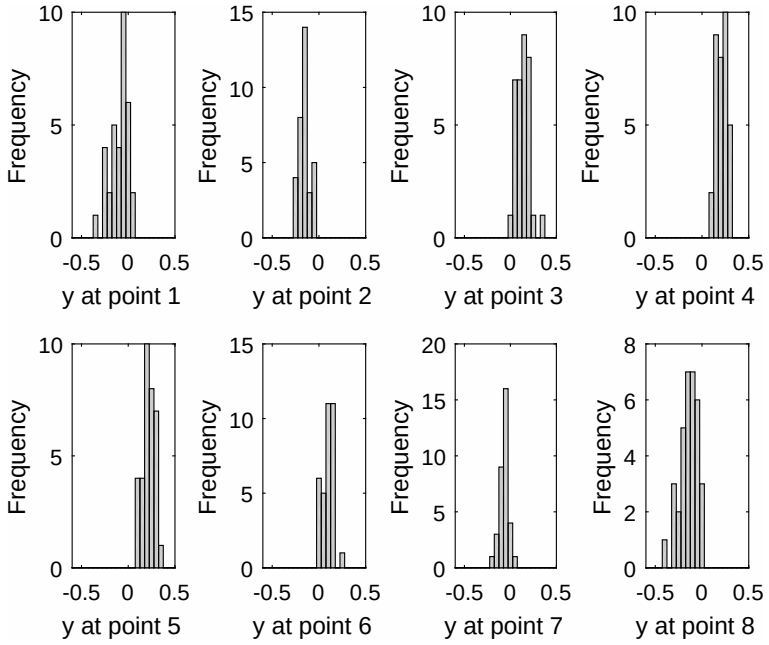

Figure 8: Histograms of y-values drawn from refitted firtree flank (convex top) curves with evenly distributed x-values (34 blades (outlying sets removed)).

could not be rejected with the specified statistical test. Given this reduced variable set with defined uncertainty, and using the Iman-Conover approach to maintain correlation, it has been possible to create 'pseudo-curve' sets. Such an approach could enable expedited UP and RDO in the face of geometric uncertainties where parametric curve and surface definitions are not immediately obvious. However, the methodology could potentially be improved still further by:

- careful selection of the principal curve-defining points,
- including distributions in x when sampling (sampling each point location from a bi-variate probability distribution function), and
- extending the work to include surfaces.

These options are currently being investigated.

8. Acknowledgements
This work is supported by Rolls-Royce plc. and conducted in the Rolls-Royce University Technology Centre for Computational Engineering, in the Computational Engineering and Design research group, Aeronautics, Astronautics, and Computational Engineering, Faculty of Engineering and the Environment, at the University of Southampton, England.

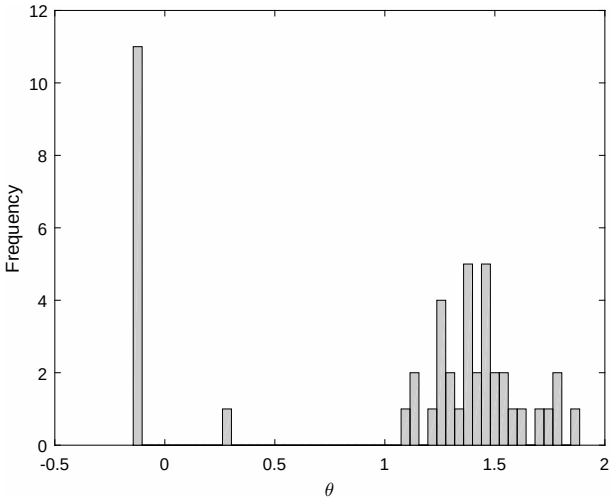

Figure 9: Histograms of θ values drawn from refitted firtree flank (convex top) curves with evenly distributed x-values (46 blades)

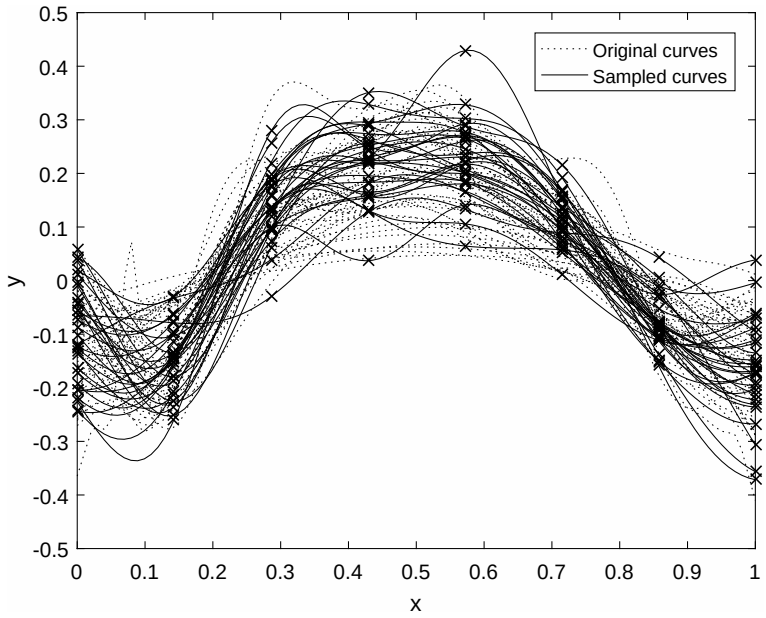

Figure 10: Random sample of 30 'pseudo-curves' from the determined extracted y-distributions using the Iman-Conover correlated sampling method (8 variables)

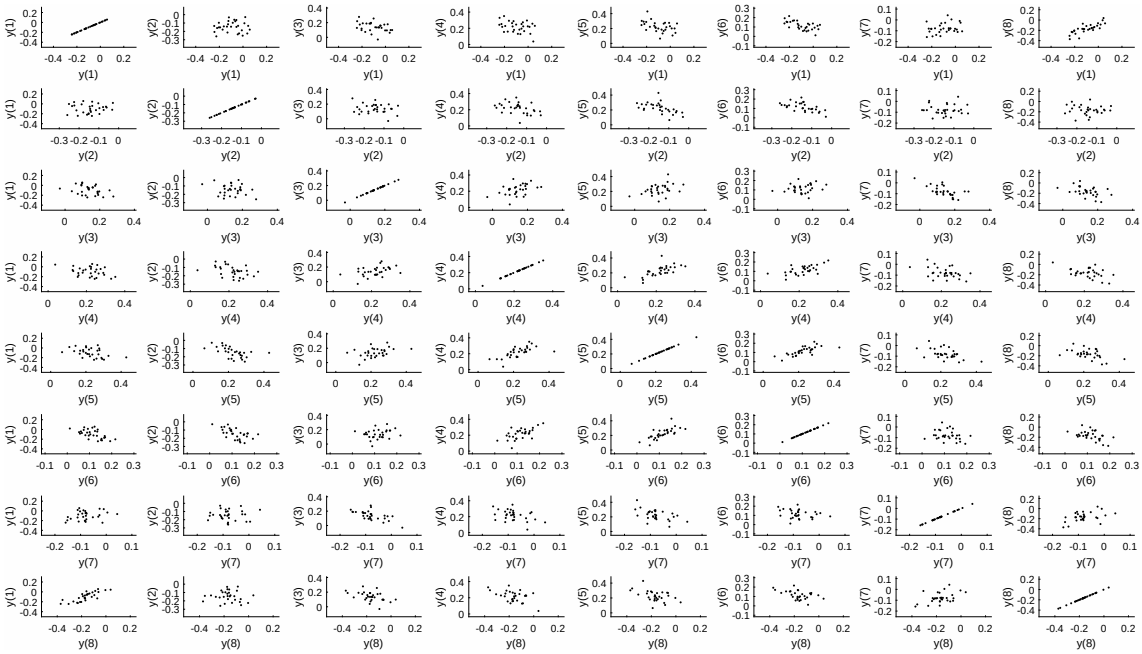

Figure 11: Scatter plot of 30-point Monte Carlo sample of curves using 8 correlated variables and the Iman-Conover correlated sampling method

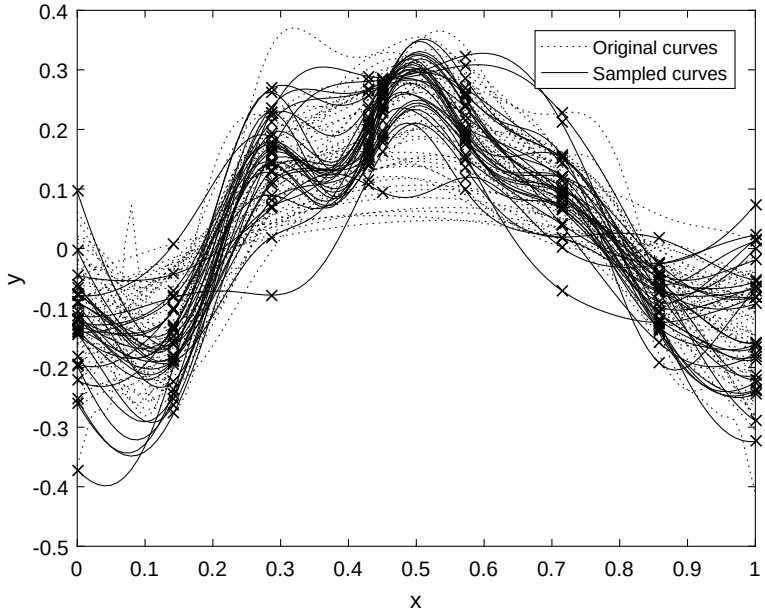

Figure 12: Random sample of 30 'pseudo-curves' from the determined extracted y-distributions using the Iman-Conover correlated sampling method (9 variables)

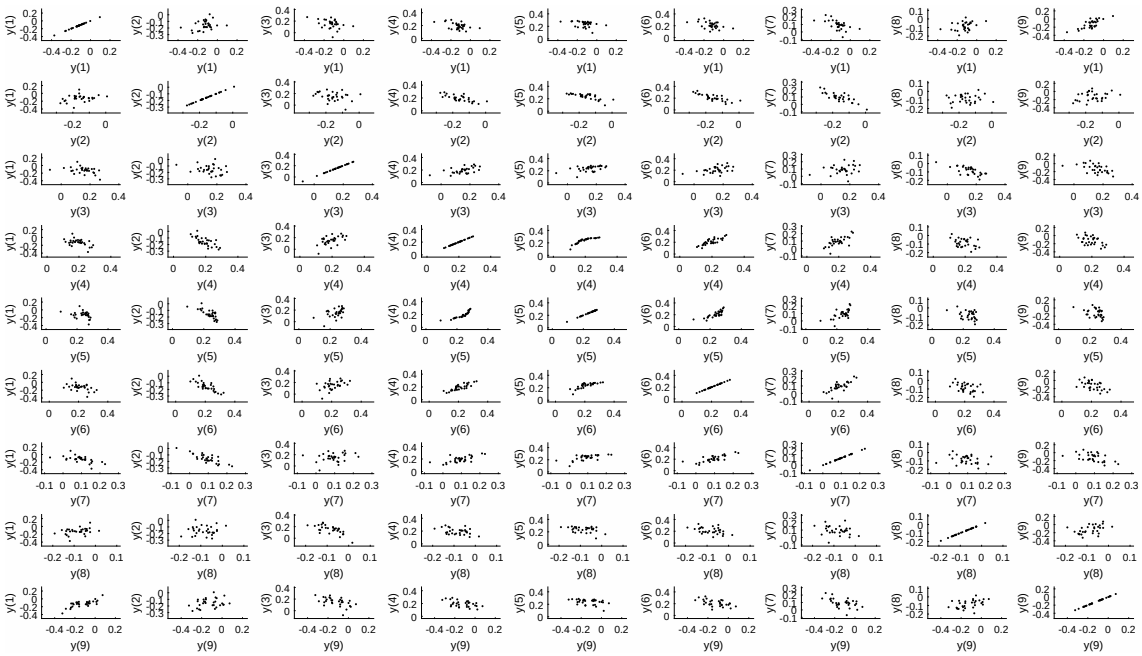

Figure 13: Scatter plot of 30-point Monte Carlo sample of curves using 9 correlated variables and the Iman-Conover correlated sampling method

9. References

[1] D. S. Broomhead and D. Lowe. Multivariate functional interpolation and adaptive networks. *Complex Systems*, 2:321–355, 1988.

[2] N. A. C. Cressie. *Statistics for Spatial Data*. John Wiley & Sons, Inc., revised edition edition, 1993.

[3] A. Forrester, A. Sóbester, and A. Keane. *Engineering Design via Surrogate Modelling - A Practical Guide*. John Wiley & Sons, Inc., 2008.

[4] A. I. J. Forrester and A. J. Keane. Recent advances in surrogate-based optimization. *Progress in Aerospace Sciences*, 45:50–79, January 2009.

[5] R. L. Iman and W. J. Conover. A distribution-free approach to inducing rank correlation among input variables. *Communications in Statistics B*, (11):311–334, 1982.

[6] F. Jurecka. Robust design optimization based on metamodelling techniques, 2007.

[7] A. Klimke. *Sparse Grid Interpolation Toolbox User's Guide*. Universität Stuttgart, Institut für Angewandte Analysis und Numerische Simulation, February 2008. Version 5.1.

[8] M. D. Morris and T. J. Mitchell. Exploratory designs for computational experiments. *Journal of Statistical Planning and Inference*, 43:381–402, 1995.

[9] W. L. Oberkampf, J. C. Helton, C. A. Joslyn, S. F. Wojtkiewicz, and S. Ferson. Challenge problems: uncertainty in system response given uncertain parameters. *Reliability Engineering and System Safety*, 85:11–19, 2004.

[10] A. Papoulis and S. U. Pillai. *Probability, Random Variables and Stochastic Processes*. McGraw-Hill Companies, Inc., 2002.

[11] G.-J. Park, T.H. Lee, K. H. Lee, and H.-H. Hwang. Robust design: An overview. *AIAA Journal*, 44(1): 181–191, January 2006.

[12] V. Picheny, T. Wagner, and D. Ginsbourger. A benchmark of kriging-based infill criteria for noisy optimization. *Journal of Structural and Multidisciplinary Optimization*, 2013.

[13] N-H. Quttineh and K. Holstr om. Implementation of a one-stage efficient global optimization (ego) algorithm. 20th International Symposium on Mathematical Programming (ISMP), Chicago, 2009.

[14] I. M. Sobol'. On the systematic search of a hypercube. *SIAM Journal on Numerical Analysis*, 16(5):790–793, October 1979.

[15] F. Xiong, S. Greene, W. Chen, Y. Xiong, and S. Yang. A new sparse grid based method for uncertainty propagation. *Structural and Multidisciplinary Optimization*, 41:335–349, 2010. doi: 10.1007/s00158-009-0441-x.