OXFORD CDE

Journal of Computational Design and Engineering, 2020, 7(1), 44–55

doi: 10.1093/jcde/qwaa005
Journal homepage: www.jcde.org
Advance Access Publication Date: 19 March 2020

RESEARCH ARTICLE

Robust design optimization using surrogate models

Andy J. Keane* and Ivan I. Voutchkov

University of Southampton, Faculty of Engineering and Physical Sciences, Southampton SO17 1BJ, UK

*Corresponding author. E-mail: ajk@soton.ac.uk

Abstract

The use of surrogate models (response surface models, curve fits) of various types (radial basis functions, Gaussian process models, neural networks, support vector machines, etc.) is now an accepted way for speeding up design search and optimization in many fields of engineering that require the use of expensive computer simulations, including problems with multiple goals and multiple domains. Surrogates are also widely used in dealing with uncertainty quantification of expensive black-box codes where there are strict limits on the number of function evaluations that can be afforded in estimating the statistical properties of derived performance quantities. Here, we tackle the problem of robust design optimization from the direction of Gaussian process models (Kriging). We contrast two previously studied models, co-Kriging and combined Kriging (sometimes called level 1 Kriging), and propose a new combined approach called *combined co-Kriging* that attempts to make best use of the key ideas present in these methods.

Keywords: Surrogate; Robust; Design; Optimization

1. Introduction

The use of optimization methods to design products is now completely routine in many companies and research institutes; see Keane and Nair (2005) for an introduction to this approach, written nearly 10 years ago. For example, in aerodynamic design, typically a computer-aided design (or other geometry design) system, meshing tool, computational fluid dynamics (CFD) solver, and post-processor are used, coupled to a range of different optimization tools to give the required design search and optimization (DSO) capability. Recently, the significant run times that can be encountered in using high-fidelity codes for design optimization (such as reynolds averaged navier-stokes (RANS) CFD or contact mechanics finite element analysis (FEA)) are often addressed by the use of surrogate schemes to speed up search.

Increasingly, designers must also contend with the fact that their designs and analyses are subject to a range of uncertainties and that designs optimized purely for nominal performance may suffer from significantly degraded performance when subject to such variations. These uncertainties stem from a range of sources such as the aleatoric uncertainties of

- limits on accuracy during manufacture,
- variations in operating conditions,
- wear and degradation in service,

and the epistemic uncertainties of

- limited accuracy in the physics of the computational models invoked during design,
- limited convergence in any iterative numerical scheme used in computations,
- round-off and discretization errors.

It is therefore of increasing importance to study designs from a stochastic perspective and to formally quantify the impact of such uncertainties on predicted design performance, rather than simply relying on ad hoc factors of safety and tolerance settings. One way of doing this is to invoke the formalism of robust design optimization (RDO) (see e.g. Kumar, Keane, Nair, & Shahpar, 2006; Ng & Willcox, 2014; Welch, Yu, Kang, & Sacks, 1990), which leads naturally to a multi-objective (Pareto front) problem where designers seek to improve the mean performance of a

Received: 3 December 2018; **Revised:** 20 February 2019; **Accepted:** 4 September 2019

design while guaranteeing that fall-off away from mean conditions is strictly controlled.

RDO can, however, be carried out in a range of ways and can additionally make very effective use of surrogate approaches to model building (see e.g. Jones, Schonlau, & Welch, 1998; Keane, 2003). In this paper, we apply some of these ideas to a problem of aerodynamic design in the face of geometric (aleatoric) uncertainty using standard and well-accepted analysis methods. We recommend the adoption of an advanced multifidelity surrogate approach, but this is, however, not the only way multifidelity schemes can be used in RDO. The most obvious alternative is based on the control variate logic as set out by Ng and Willcox (2014) and further elaborated in Peherstorfer, Willcox, and Gunzburger (2016). In both approaches, the aim is to exploit correlations between alternative approximations of the goals being sought, so as to speed up calculations. Here, however, our differing fidelity models will simply be constructed by using different numbers of samples in the uncertainty quantification (UQ) process, rather than by using different models of the underlying problem, and this makes the use of control variate approaches less effective. (It is possible to use a subset of the available UQ samples to construct a surrogate to support control variate approximations, but they turn out to be no better than using all the available samples in a larger surrogate of the same type and then carrying UQ on the larger surrogate directly.)

There are also a number of alternative Kriging-based methods directly related to the current work, such as that proposed by Han, Zimmermann, and Görtz (2012), the simplified method of Zimmermann and Han (2010), and the hierarchical Kriging method of Han and Görtz (2012). See also Courrier, Boucard, and Soulier (2015), who compared various multifidelity methods.

This paper builds on two previous studies (Keane, 2009, 2012) where co-Kriging response surface methods were combined with pseudo-Monte Carlo (low-discrepancy sequence) approaches for assessing the mean and standard deviation in performance of a 2D gas turbine compressor blade section subject to damage in service and uncertainty in manufacture. It then adds in ideas taken from Dellino, Kleijnen, and Meloni (2012) where combined, or level 1, Krigs are proposed and demonstrates a new approach making use of ideas from both these methods – a method we term combined co-Kriging.

The paper is laid out as follows: in Section 2 we introduce the more advanced techniques that represent the central ideas in this work. Section 3 then describes a real 2D CFD design case and shows how this design can be modified by design or uncertainty and then analysed with a RANS-based CFD approach. Section 4 applies the various advanced schemes proposed here to the RDO of this design case, while Section 5 draws conclusions from the work presented.

2. Robust Design Optimization with Advanced Surrogates

To improve the behaviour of search schemes based on the simple application of surrogates to either DSO or UQ separately, we introduce three advanced ways of dealing with the combined problem of uncertainty propagation and Pareto front search.

2.1. Co-Kriging

Our first advanced method makes use of the formalism of co-Kriging where results with multiple levels of fidelity can be combined during the search (see Kennedy & O'Hagan, 2000). To do this, we combine results from limited sample size UQ (the low-fidelity model) with those for more expensive UQ with many more samples (the high-fidelity model). To make this approach worthwhile, we can use the high-fidelity calculation only very sparingly – e.g. we start a response surface-based search with a design of experiments (DoE) of design vectors where we calculate low-fidelity results at all points but high-fidelity ones at less than 5% of these. We then build a pair of multifidelity co-Krigs (one for mean and one for standard deviation) with all the results and use these to estimate the response statistics of the functions being searched (see Fig. 1).

In co-Krigs, the inputs to the low-fidelity (cheap) and high-fidelity (expensive) calculations, $\mathbf{x}_c$ and $\mathbf{x}_e$, are taken to be related to the outputs (responses) y_c and y_e by computational functions, $y_c = f_c(\mathbf{x}_c)$ and $y_e = f_e(\mathbf{x}_e)$. The responses resulting from the DoE over these codes are used to construct an approximation

$$\hat{y}_e = \rho \hat{f}_c(\mathbf{x}) + \hat{f}_d(\mathbf{x}), \tag{1}$$

which is the sum of two Gaussian process models, each of which depends on the distances between the sample data used to construct them. Here, the hat symbols indicate the models are approximations, the subscript 'd' indicates a model of the differences between the low- and high-fidelity functions (all the high-fidelity evaluations are carried out at locations where low-fidelity calculations have already been run), and ρ is a scaling parameter.

The distance measure used here is

$$d(\mathbf{x}^{(i)}, \mathbf{x}^{(j)}) = \sum_{h=1}^{k} \theta_h \left(x_h^{(i)} - x_h^{(j)} \right)^{p_h}, \tag{2}$$

where θ_h and p_h are hyperparameters tuned to the data in hand and k is the number of dimensions in the problem. The correlation between points $\mathbf{x}^{(i)}$ and $\mathbf{x}^{(j)}$ is then given by

$$\mathbf{R}(\mathbf{x}^{(i)}, \mathbf{x}^{(j)}) = \exp[-d(\mathbf{x}^{(i)}, \mathbf{x}^{(j)})] + \Lambda \delta_{ij}, \tag{3}$$

where Λ is a regularization constant that governs the degree of regression in the model (when set to zero, the Krig strictly interpolates the data supplied) and δ_{ij} is the Dirac delta function. When the response at a new point $\mathbf{x}^{new}$ is required, a vector of correlations between the new point and those used in the DoE is formed:

$$\mathbf{c}(\mathbf{x}^{new}) = \begin{pmatrix} \rho \sigma_c^2 \mathbf{R}_c(\mathbf{x}_c, \mathbf{x}^{new}) \\ \rho^2 \sigma_c^2 \mathbf{R}_c(\mathbf{x}_e, \mathbf{x}^{new}) + \sigma_d^2 \mathbf{R}_d(\mathbf{x}_e, \mathbf{x}^{new}) \end{pmatrix}, \tag{4}$$

where σ^2 are the variances in the cheap and difference Gaussian models. The prediction is then given by

$$\hat{y}_e(\mathbf{x}^{new}) = \mu + \mathbf{c}^T \mathbf{C}^{-1}(\mathbf{y} - 1\mu), \tag{5}$$

where $\mu = \mathbf{1}^T \mathbf{C}^{-1} \mathbf{y} / \mathbf{1}^T \mathbf{C}^{-1} \mathbf{1}$ and $\mathbf{C} = \begin{pmatrix} \sigma_c^2 \mathbf{R}_c(\mathbf{x}_c, \mathbf{x}_c) & \rho \sigma_c^2 \mathbf{R}_c(\mathbf{x}_c, \mathbf{x}_e) \\ \rho \sigma_c^2 \mathbf{R}_c(\mathbf{x}_e, \mathbf{x}_c) & \rho^2 \sigma_c^2 \mathbf{R}_c(\mathbf{x}_e, \mathbf{x}_c) + \sigma_d^2 \mathbf{R}_d(\mathbf{x}_e, \mathbf{x}_e) \end{pmatrix}$. When building any form of Krig, it is necessary to carefully tune the sets of hyperparameters to match the data inuse – for co-Krigs this tuning is applied to the low-fidelity data, data representing the differences between the low- and high-fidelity series and the ratio ρ parameter to link the various data sets. Fortunately, for the small numbers of results typically available in such work, this is not overly expensive for co-Kriging – further details of the co-Kriging approach are provided in Forrester, Sobester, and Keane (2008).

The resulting co-Krigs are then searched with the NSGA-II algorithm (a multi-objective evolutionary search; Deb, Samir, Amrit, & Meyarivan, 2000) and updated using new design vec-

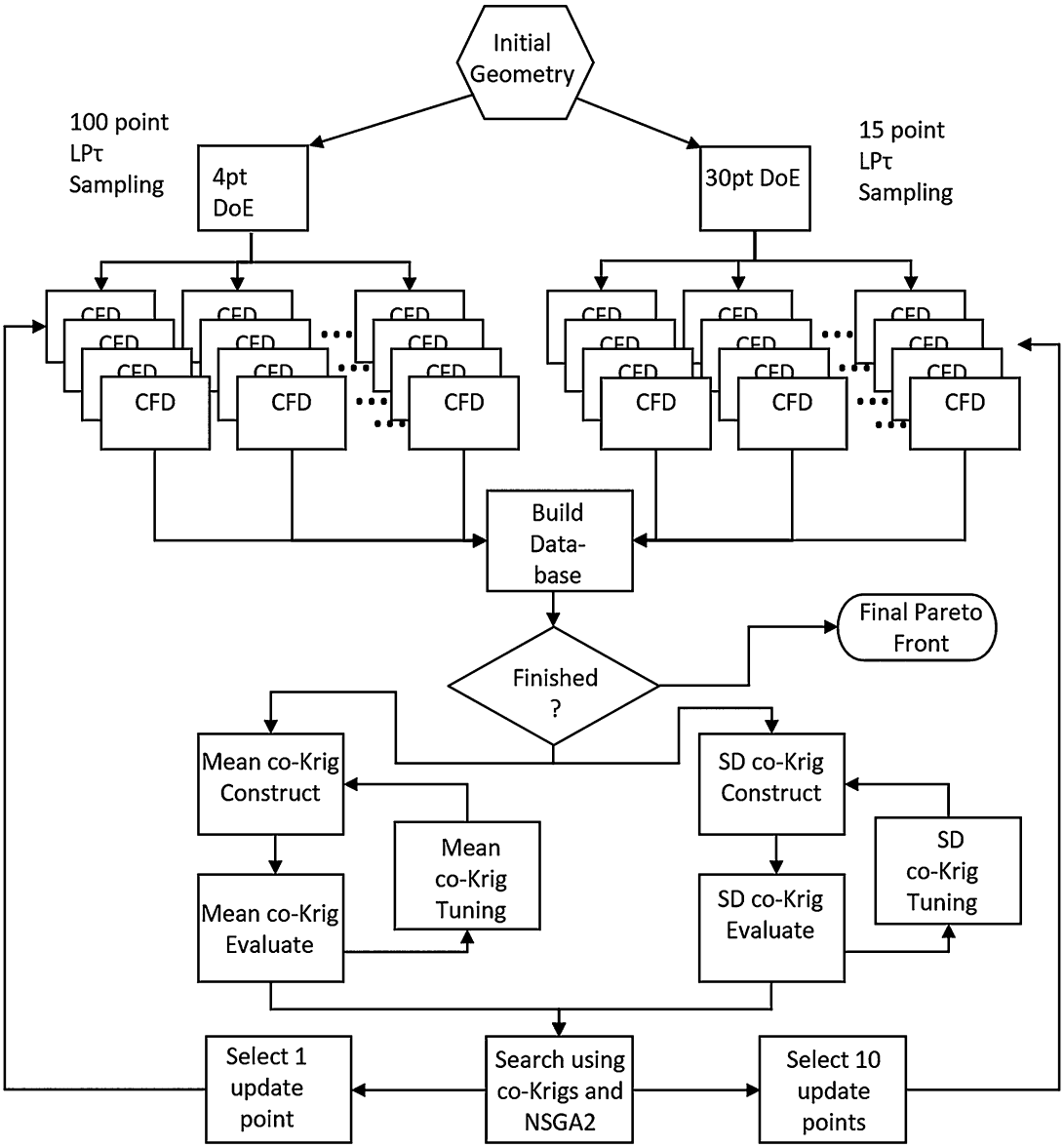

Figure 1: Flow chart illustrating the update sequences used in co-Krig approaches.

tors taken from the approximated Pareto front at each stage, but of these only a small subset is analysed using both sample sizes. Thus, at the end of the search we have evaluated the low-fidelity model many times more than the high-fidelity model. This should be compared to a direct Kriging search where only low- or high-fidelity calculations are used. The aim is to achieve results comparable to using only high-fidelity calls with costs more comparable to working solely at the lower fidelity – this approach is intended to mitigate the problem identified by Keane (2009) that, when working solely with limited sized (Monte Carlo) uncertainty samples on real-world problems, searches typically returned designs that failed to fulfil their promise when evaluated with larger sample sizes.

A particular benefit of this approach is that at the end of the process one is left with the collection of expensive evaluations that can then be used directly without further effort when selecting the best design to proceed with – this is not generally the case when using limited sample sizes for UQ.

It is also possible to build co-Krigs from the results of direct UQ Krig models where the lower level Krigs are built with different numbers of sample points. By occasionally using larger numbers of sample points to build the UQ Krigs, the aim is to try and provide better results that can be fed back to the search algorithm. This approach is generally not much more successful, however: for simpler noise problems, a co-Krig built on good design of experiments sampling is usually sufficiently accurate because of the addition of the few high-fidelity points and this makes the construction of low-level Krigs essentially a waste of effort; conversely, on more difficult problems the low-level Krigs add only a modest advantage given the extra cost in their construction. We do not pursue this idea further here.

2.2. Combined Kriging

The second advanced approach we demonstrate requires a slightly more intrusive change to the problem handling: we build

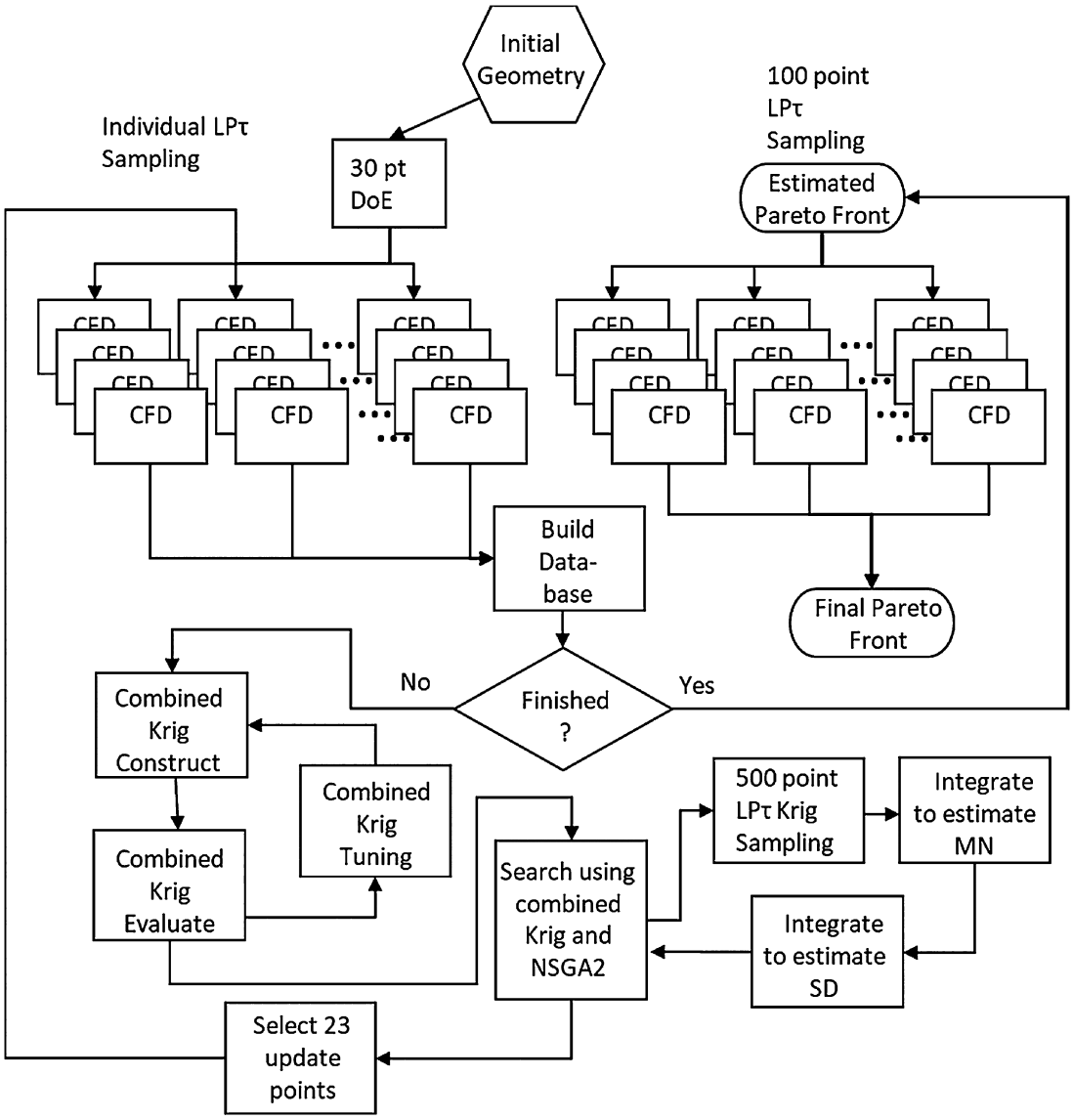

Figure 2: Flow chart illustrating the update sequences used in combined (level 1) Krig approaches.

a single combined Krig that is used to both support search *and* carry out UQ (see e.g. Dellino et al., 2012), an approach we term 'combined Kriging' in this paper (Dellino et al. refer to it as 'level 1 Kriging'). Clearly, if the low-level Krig can accurately model the underlying functions, it will allow very accurate statistics to be computed, albeit at the cost of sampling the Krig multiple times. (Note that to be competitive such Krigs have to be built on many more samples than for methods where UQ is used to integrate out the noise quantities before building the RDO surrogates.)

To begin with, an LPτ DoE (Statnikov & Matusov, 1995) is carried out where both the design variables and noise variables are *all* varied simultaneously. Next, a multidimensional Krig is constructed through these data. Then, low-discrepancy sequence sampling over the noise variables is carried out on the combined Krig, at any desired set of design variables, so as to return the desired predictions of mean and standard deviation of performance (see Fig. 2).

Note that when evaluating *new* update points, the search engine can no longer simply specify sets of design variables for each evaluation; it also has to manage the noise sampling at the same time – as each set of design variables is added, information also has to be supplied on what values the noise variables should take for these update samples – here we use a space filling approach where we carry out a short search to best place each new noise sample in the existing data set, maximizing its Euclidean distance from the existing samples. Also, once the search is completed, it will be necessary to confirm the statistics of the final Pareto front with direct UQ as the combined Krig will be unlikely to be able to supply completely accurate statistics directly.

This approach increases the number of variable combinations being sampled as compared to the number of noise variables. Since it is realistic to assume that design variables play a greater part in influencing final performance as compared to noise variables, this approach has obvious appeal. However, it does, again, depend on how nonlinear any noise effects are.

Moreover, the sizes of the Krigs being built become significantly larger as now they contain sample information on both design and noise variable changes. Since the cost of tuning and sampling Krigs rapidly rises with both the number of variables and the number of samples, this can become the limiting factor in using the combined approach. Tuning is generally an order of magnitude slower than that for the co-Krig approach.

Having completed the search process, it will typically be necessary to re-evaluate any final design being selected by additional UQ; typically around 10% of the front can be checked using full sampling when using the same budget as the co-Krig approach.

As expected, the accuracy of the UQ depends on the total number of samples taken and the difficulty of the problem. In general, the resulting Pareto fronts are often better than those achieved using the co-Krig approach, but this very much depends on the dimensions of the problem being studied as well as its inherent nonlinearities.

2.3. Combined co-Kriging

Because combined Kriging does not carry out a full UQ process on the problem functions directly at any point, the user is forced to carry out a series of UQ checks on the results of the search after it is completed, to allow for any inaccuracies in the Krigs being built. This can be a significant additional cost that can, to some extent, be mitigated by borrowing ideas from co-Kriging. To do this, instead of placing new points across a range of design and noise variable values at every update cycle, we occasionally carry out a UQ process focused on a single promising design *during the search*. So, for example, at every fifth cycle, we take the suggested design variables for a single design point midway along the Pareto front, as provided by the NSGA-II search of the combined Krig, and use all the available function evaluations for the update to simply explore the noise variables at this one point using a low-discrepancy sequence. We then use fewer calls for final Pareto front UQ at the end of the search.

Since they are chosen from the centre of the predicted Pareto front, towards the end of the search process these designs are very likely to be those the user is most interested in taking forward for final checking. Two benefits arise: first, the errors in the combined Krig prediction are much reduced at these points, and second, full UQ is cheaper at these locations, because one already has many directly useful noise samples to begin with. When dealing with expensive real-world problems with many degrees of freedom, both aspects become significant as we shall see. Once the search is finished, we carry out additional UQ samples on these promising designs combined with a smaller sample of points taken from the final Pareto front for full UQ as before.

If we follow this approach, and assume the same budget of total function evaluations, because we no longer need so many calls for final design UQ, we can carry out more cycles of the update process and then carry out additional UQ on the mid-Pareto points for which extended UQ has already been carried out, together with full UQ on a few further designs, spaced out along the Pareto front.

3. Design Example

We next consider a practical aerodynamic design example: gas turbine blade section design. It is, of course, commonplace to use design optimization methods to change the shape of aerodynamic sections when designing new gas turbines. Here, we assess the advanced approaches just described on an unconstrained gas turbine aerodynamic section model introduced previously (Keane, 2009, 2012) and focus on the effects on gas turbine compressor blade performance of foreign object damage, erosion, and uncertainty in manufacture.

In the previous papers cited, we compared and contrasted five approaches. First, we simply optimized a baseline geometry for nominal performance using direct (Goldberg, 1989) and surrogate-based search (Jones et al., 1998). Then, we employed a noisy phenotype (Tsutsui & Ghosh, 1997) approach to allow for possible sensitivity in the mean performance but did not explicitly seek a trade-off between mean performance and robustness. Third, we used NSGA-II to deliberately construct a trade surface (Pareto front) adopting both direct and surrogate-assisted approaches. Fourth, we used a surrogate-based multi-objective expected improvement formulation (Keane, 2006) to construct the trade-off using single-objective search methods. Finally, we used co-Kriging-based surrogates, which were shown to give the most efficient approach to robust design improvement among the methods tested.

We next demonstrate how the combined Kriging and combined co-Kriging approaches, which both use a single surrogate to seek design improvements *and* carry out UQ at the same time, can improve RDO on this industrial aerodynamics problem, as compared to the co-Krig-based method shown to be the best approach in the earlier studies. To begin with, we outline for convenience the problem being studied; complete details can be found in the references.

3.1. Geometry modification

To carry out any form of automated aerodynamic section optimization, some form of geometry modification is required. Here, we adopt the parametric design and rapid meshing (PADRAM) code (Shahpar & Lapworth, 2003) that allows for a range of uncertainty studies that have been described elsewhere (Kumar et al., 2006). In this process, a baseline airfoil section shape is altered by the addition of a series of Hicks–Henne functions in various ways (Hicks & Henne, 1978). Here, this scheme is used to make four sets of changes to the baseline airfoil that allow for overall design improvement, modelling of manufacturing uncertainty, modelling of foreign object damage, and modelling of flank erosion. Figure 3a–f illustrates the baseline airfoil section and each of these modification processes in turn. Note that all the changes are controlled such that the trailing edge form and angle are kept fixed as would be needed to ensure unchanged inflow angles for the next stage in the compressor. Ten Hicks–Henne functions are used for the design shape changes, distributed around the section, and this permits very significant alterations in geometry.

In all the tables that follow, the normalized amplitude changes of these design variables are provided. These are defined in such a way that a value of 0.5 generates the base geometry, with values less than this removing material from the section and those greater adding to it – full details may be found in the thesis of Kumar (2008). The other three sets of changes work in the same way but make much smaller shifts:

1. the manufacturing variability changes affect the entire section (again using 10 Hicks–Henne functions distributed around the section, but now much smaller in scale) paying particular attention to the leading edge;
2. the foreign object damage changes are localized to the leading edge using a single Hicks–Henne function that can only remove material but can vary in position and shape as well as depth (a peak height setting of zero means no damage);

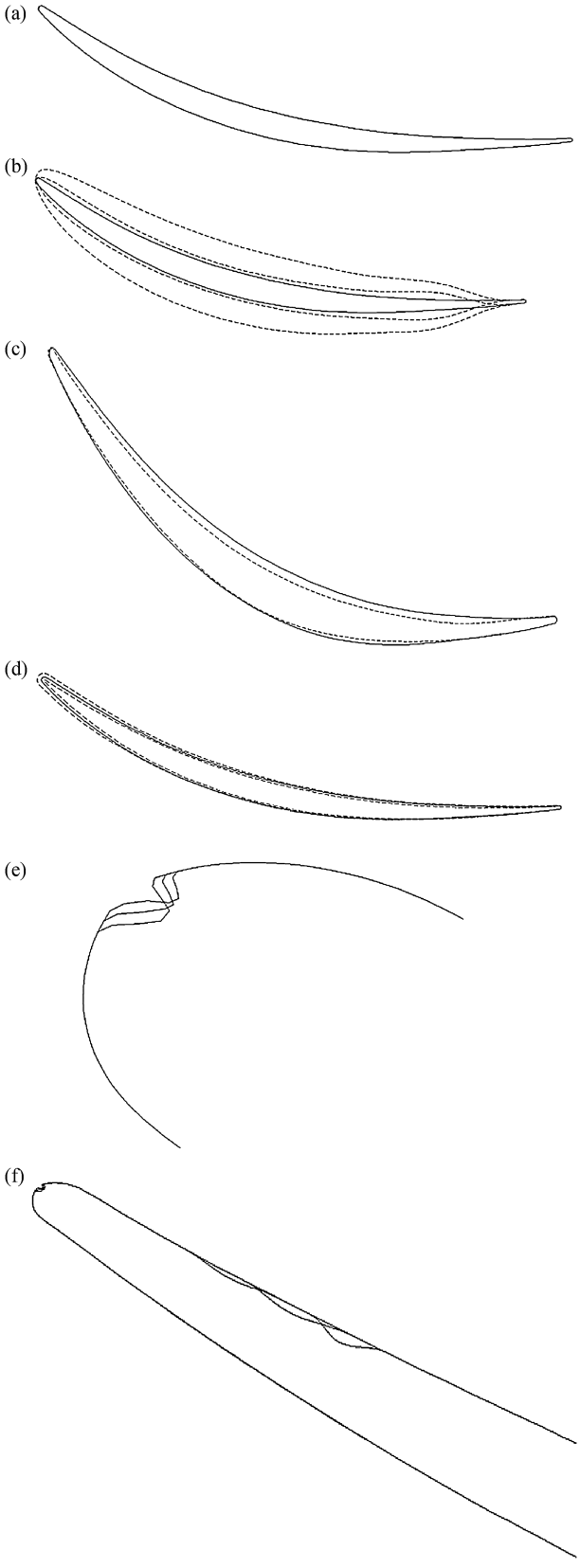

Figure 3: (a) Baseline airfoil section. (b) Overall design variations' maximum possible changes. (c) Typical design change (shown distorted for clarity). (d) Maximum range of changes aiming to model uncertainty in manufacturing processes. (e) Leading edge damage models. (f) Flank erosion models.

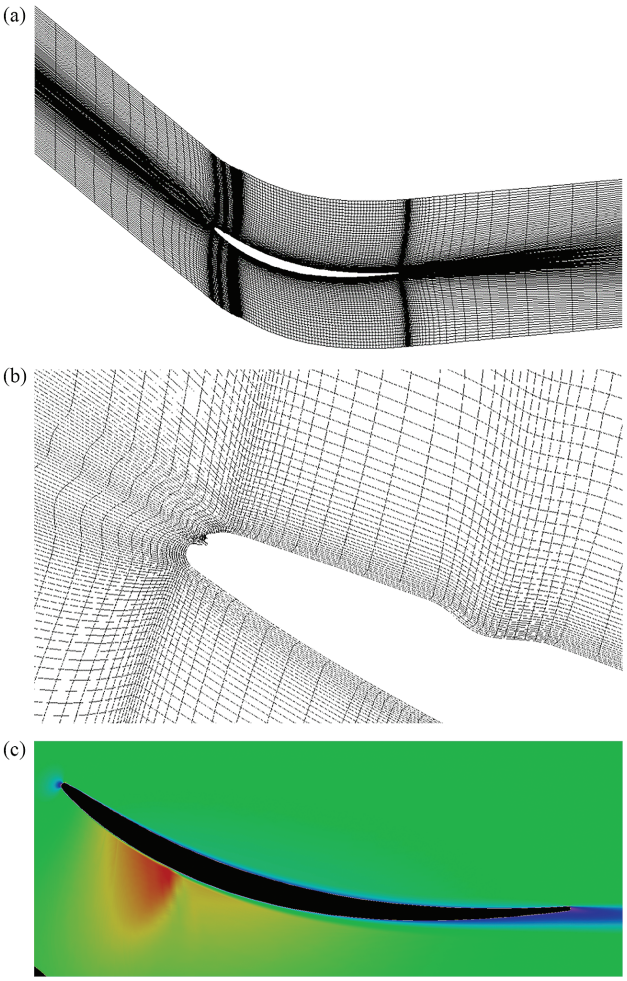

Figure 4: PADRAM meshes and resulting flow field: (a) overall mesh; (b) in way of leading edge and flank erosion; and (c) section profile and streamwise velocity field.

3. the erosion changes are localized to the flank, again using a single Hicks–Henne function that can again vary in position and shape as well as depth and can only remove material (a peak height setting of zero means no damage) – it is applied where typical erosion events occur.

The modelling of the damage and erosion events is clearly far from accurate since the Hicks–Henne functions give smooth and continuous pockets, whereas in reality these are typically irregular and sharp edged. However, as used here, the pockets still act to trip and disturb the flow in a way that is not unrealistic. Note that the flank erosion pockets are allowed to extend over a greater area of the blade than the damage pockets at the tip.

3.2. Analysis

The PADRAM geometry modification process also permits the automated adaptation of high-quality 'OCH' meshes around the airfoil section in a setting appropriate for application to gas turbine blade design (here 'OCH' refers to the shapes of the various mesh blocks, e.g. ring-like, cup-shaped, or with legs). Figure 4 shows a typical mesh and the way that this mesh is distorted around the localized erosion geometries introduced

Table 1: Comparisons of various blade designs: nominal and six fully evaluated co-Kriging designs.

Quantity	Baseline design	NSGA-II co-Krig optimized design (last six designs)					
Shape fn. 1	0.5	0.5220	0.4582	0.4565	0.4984	0.4375	0.5125
Shape fn. 2	0.5	0.4666	0.4576	0.4000	0.5165	0.5125	0.4375
Shape fn. 3	0.5	0.4000	0.4164	0.4007	0.5000	0.4375	0.5125
Shape fn. 4	0.5	0.4760	0.4800	0.4611	0.5500	0.5125	0.4375
Shape fn. 5	0.5	0.4000	0.4904	0.4160	0.4167	0.4375	0.5125
Shape fn. 6	0.5	0.4643	0.4695	0.4810	0.4647	0.5125	0.4375
Shape fn. 7	0.5	0.4657	0.4695	0.5144	0.4714	0.4375	0.5125
Shape fn. 8	0.5	0.4652	0.5500	0.4395	0.4107	0.5125	0.4375
Shape fn. 9	0.5	0.4648	0.4754	0.4954	0.4802	0.5125	0.4375
Shape fn. 10	0.5	0.4047	0.4759	0.4756	0.4305	0.4375	0.5125
Mean $loss_{100}$	1.1571	1.3439	1.1772	1.1720	1.3076	1.1872	1.3345
Std. $dev._{100}$	0.1234	0.1051	0.1218	0.0997	0.0941	0.1033	0.1228
Average mean $loss_{100}$	1.1571	1.2537					
Average std. $dev._{100}$	0.1234	0.1078					
Total fn. calls/search	1/100	$(30 + 10 \times 10) \times 15 + (4 + 10 \times 1) \times 85 = 3140$					

when studying leading edge damage or flank erosion. Here, the 'O' mesh is 4 cells deep and slightly over 27 000 cells are used in total, along with the resulting flow field. The topology of this mesh is unchanged throughout the study presented here so that any effects caused by remeshing can be avoided. If geometry changes result in overly large distortions to the mesh, the relevant run is aborted and appropriate action taken during any subsequent optimization or statistical processes.

Having set up the mesh, a full steady-state RANS solve can be carried out using the parallel version of the Hydra CFD code in around 3 min on a 12-core PC (Moinier & Giles, 1999). The section is run in isolation with the following boundary conditions: inlet temperature, 290 K; inlet total pressure, 63 400 Pa; whirl angle, −37.3°; and outlet static pressure, 52 000 Pa. Typical root-mean-square (RMS) flow residuals are 4.7e−9 and RMS turbulence residuals are 2.4e−15. Here, we use pressure losses as the main design goal. These are all normalized by that of this chosen baseline blade, to preserve confidentiality – it therefore has a normalized loss coefficient of 1.0; see column 2 of Table 1. Full validation studies of this meshing and solution process are presented by Kumar (2008), where it is shown that the results achieved are reasonable for the kind of statistical work being presented here. (Note that using this mesh it is accepted that accurate modelling of the flow in the damage pockets cannot be achieved with this number of cells – in this study, the intention is that these regions should act as trips to the flow as rough edge pockets would in practice. Clearly, detailed flow analysis around an eroded section that attempted to capture all of the flow physics would require a significantly increased mesh count with consequent increases in run time.)

If this baseline blade is then made subject to uncertainties in leading edge damage, flank erosion, and manufacturing, it is possible to estimate the mean pressure loss and standard deviation in this loss using the previously described Monte Carlo approaches. To do this, 100 independent perturbations are made to each of the three sets of Hicks–Henne functions defining these three sets of uncertainties, using the standard LP_τ space filling design of experiment approach, and averages taken. The perturbations made are significantly beyond the extremes in tolerance band typically found in manufacture or in service, but they provide a good measure of the resilience of a section design – such variations would be unlikely in practice. This calculation, of course, requires 100 times as much processing effort as finding the nominal performance but provides a datum for comparison

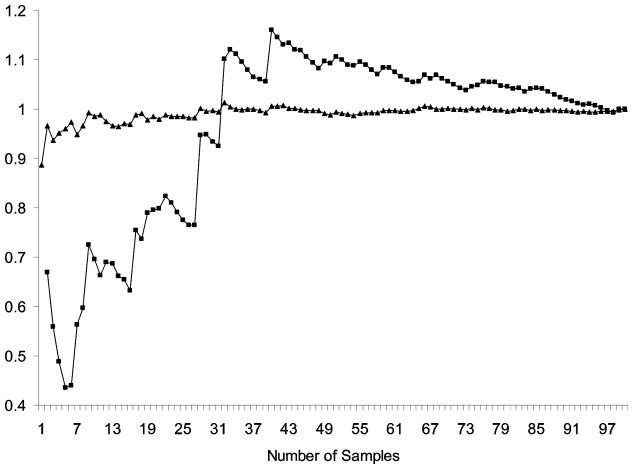

Figure 5: Convergence of averaging process used in assessing uncertainties in shape (triangles show normalized mean and squares show normalized standard deviation).

– here the normalized mean loss coefficient is 1.157 with a standard deviation of 0.123 (i.e. the mean loss is 16% worse than the deterministic calculation on the nominal baseline shape); see again column 2 of Table 1. Figure 5 shows the normalized convergence history of this calculation – note that since larger values of the perturbations almost inevitably lead to greater losses, the mean and standard deviation both rise until the 16D hypercube being sampled is fully explored by the pseudo-random number sequence in use – here after around 50 calculations – thereafter the convergence is essentially unbiased. Unfortunately, the use of such a large sample size in design searches cannot generally be justified, particularly if 3D analyses are being considered.

The central idea of this paper is to show how calculations using fewer samples can be linked with those using 100 samples via the formalisms of co-Kriging, combined Kriging, and combined co-Kriging. In this work, the results gained when using 100 samples are treated as being 'correct' and thus form an 'expensive' evaluation, while those using just 15 samples are considered to be approximations to these correct results and thus form an alternative 'cheap' evaluation. In combined Kriging and combined co-Kriging, the available computational budget is spread over both design and uncertainty sampling in the hope that the

resulting single, large Krig can model both aspects of the design sufficiently well – the combined co-Kriging approach focusing more attention on establishing good predictive power for the statistical quantities being modelled, as discussed earlier. Results from these last two approaches are then confirmed using the 'exact' calculation of LPτ sampling over 100 samples once the searches are complete, as before.

Of course, in this real-world problem we do not know the true Pareto front of mean versus standard deviation, even for 100-point sampling. This makes comparing any proposed method with a datum somewhat difficult. However, as we have accumulated tens of thousands of CFD runs in our work on this problem, we can construct an approximation to the true performance using a very large combined Krig model based on some 47 000 CFD runs. It is not feasible to tune the Krig over this full data set since the repeated matrix solutions needed during tuning become too expensive, so instead the Krig is first tuned using a subset of 6000 points to establish the relevant hyperparameters. Then, these hyperparameters are used during prediction over the full data set. Since we are studying mean and standard deviation, we need to integrate the results coming from this large combined Krig using LPτ sampling – here we use 500-point sampling to do this. The results are then fed into an NSGA-II search for the 'true' Pareto front, which we can then use to characterize our proposed search methods. This front is shown in subsequent figures and is also used when normalizing the hypervolume metrics used to illustrate search progression. It thus provides the baseline against which to compare results, but it must be borne in mind that the resulting front is still only an estimate of the true one. Even so, this process takes several weeks to perform on a dedicated 12-core system over and above the cost of gathering the thousands of CFD runs used.

4. Robust Design Optimization

Having set out the performance of the datum blade and how we generate a 'true' Pareto front, we begin our RDO comparisons by first adopting the formalisms of co-Kriging to combine results with 15 LPτ damage samples and those where 100 samples are taken – this is the search process detailed before in Keane (2012). We then look at the combined Krig methods.

4.1. Co-Kriging

To apply the co-Kriging process, an initial DoE of 30 design geometries is used and each geometry analysed with 15-point LPτ pseudo-Monte Carlo sampling of the damage models. The first four of these geometries are then further analysed with 100-point LPτ pseudo-Monte Carlo sampling. Krigs are built and carefully tuned on the 15-point data and the difference data and then combined as a co-Krig following the procedure set out earlier. The resulting co-Krig is then searched using the NSGA-II formalism with a population of 100 designs over 200 generations. Updates are taken from the final predicted Pareto front – in this case, 10 designs are chosen evenly spread along the estimated front and each analysed with 15-point LPτ pseudo-Monte Carlo sampling. Of these 10 designs, one is additionally analysed with 100-point sampling, taken from the middle of the estimated front. These results are then added to the existing DoE data and the Krigs retuned. This update procedure is repeated 10 times. At the end of this co-Krig process, $30 + 10 \times 10$ 15-point samples and $4 + 10 \times 1100$-point samples (i.e. 85 extra samples at these points) have been taken, giving a total of 3140 CFD runs. Moreover, at the end this means full 100-point statistical analyses have been carried out on 14 designs, a number of which being created towards the end of the search will hopefully be directly usable without further computation; see columns 3–8 of Table 1.

Figure 6 illustrates these searches, with six (upper) showing the expensive and cheap update points produced during a typical search (Note that the objective function values for the cheap evaluations are obtained here by taking the design variables used during the search and using the very large offline trained Krig-based approach to correct them to their 'true' positions – the expensive points being based on 100-point samples do not need this correction.) alongside the large combined Krig-based estimate of the 'true' Pareto front, and six (lower) showing the normalized hypervolume metric convergence for reference point (5, 0.5), also based on the combined Krig estimate of the 'true' Pareto front, here for nine independent runs. Figure 6 (upper) also shows where the co-Krig believes the front to lie at the end of the search.

It is clear that this is a hard problem for the Krig model to deal with. The co-Krig estimated front lies some way from the 'true' one, while the solutions are very far from converged over the nine runs plotted. Note also that only one of the expensive designs has improved on the mean loss, while a number of them have improved standard deviations, probably because more attention was paid to performance than variability in designing the initial section. This kind of results' spread is typical of real-world problems where the analyst has to settle for what is achievable within a tight computational budget. The best expensive point here is probably that at (1.1772, 0.1218), column 5 of Table 1, and since this has been evaluated with 100-point LPτ sampling it could be carried forward directly for further design consideration without additional UQ. The average mean and standard deviation taken over the tabulated designs are 1.2537 and 0.1078, respectively.

4.2. Combined Kriging

Next, a combined Kriging approach with the same budget of CFD runs is used. So to begin with, an LPτ DoE is carried out where both the design variables and noise variables are *all* varied simultaneously.

For this example problem, a 26D Krig must be constructed through these data – tuning this Krig is now a rather expensive process, which can be mitigated to some extent by the use of GPU-based calculations (see e.g. Toal, 2016). Then, 500-point low-discrepancy sequence sampling over the 16 noise variables is carried out on the combined Krig, again using GPU computations, at any desired set of 10 design variables, so as to return the desired predictions of mean and standard deviation of performance. Note again that once the search is completed, it will be necessary to confirm the statistics of the final Pareto front with direct 100-point UQ as the combined Krig will be unlikely to be able to supply sufficiently accurate statistics directly, given the large number of noise variables in use and the modest number of CFD samples available in total within a single search run.

As mentioned earlier, the cost of tuning and sampling Krigs can become the limiting factor in using the combined approach. Here, the Krigs are restricted to just 950 data points at any one time, 30 points in an initial DoE sample plus 40 updates each of 23 more points (of which 18 are derived from searching the evolving combined Krig and 5 are used to infill the domain based on nearest-neighbour methods). This is less than half the total number of function evaluations used when using co-Krigs to represent the design variables and UQ separately, because it leads to an extremely time-consuming Krig tuning process and

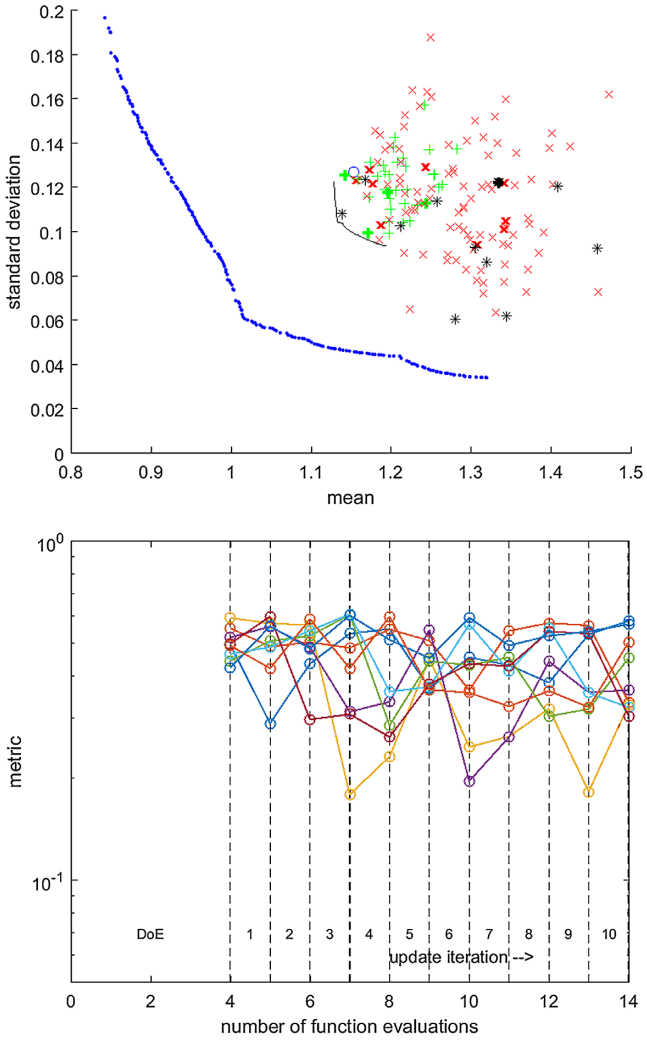

Figure 6: Co-Krig-assisted NSGA-II results – upper: results for 15- and 100-point (bold) LPτ pseudo-Monte Carlo sampling, (□) final generation, (+) initial generation, and (×) intermediate generations, along with the estimated Pareto front (solid line), 'true' Pareto front (dotted line), and (○) initial baseline design; lower: Pareto front normalized hypervolume metric convergence for reference point (5, 0.5) based on the estimate of the 'true' Pareto front, for nine independent runs.

also because of the need to spare some resources so as to be able to validate final designs. So, having completed the search process, 22 final designs taken from the NSGA-II search over the final combined Krig are confirmed by additional UQ using 100-point sampling so that the whole process uses the same budget as the co-Krig search ($950 + 22 \times 100 = 1350$). Of these 22 designs, the best 6 are shown in Table 2, while Fig. 7 illustrates this approach in the same format as Fig. 6.

Now the estimated Pareto front lies much closer to the 'true' one and the convergence plots show a noticeably better performance. However, when one goes on to examine the best suggested designs with full 100-point LPτ sampling, things are not quite so good. Nonetheless, as Fig. 8 shows, the results after full UQ on the final designs are clearly much improved over the co-Krig approach. The average mean and standard deviation taken over the tabulated designs are now both improved at 1.0854 and 0.0763, respectively.

4.3. Combined co-Kriging

Finally, we adopt the newly proposed combined co-Kriging approach, where during the Krig update process we place greater focus on evaluating response statistics, still keeping the same budget of CFD runs. Again, an LPτ DoE is carried out where both the design variables and noise variables are *all* varied simultaneously, followed by 500-point low-discrepancy sequence sampling over the 16 noise variables on the combined Krig at any desired set of 10 design variables. This time, however, at every fifth update cycle, a single design point is chosen for more rigorous UQ sampling so that at the end of the search this collection of points is more directly usable, and also, hopefully, the Pareto front has been more accurately positioned in terms of the design variables. Nonetheless, after the search is completed, it will still be necessary to confirm the statistics of the final Pareto front with an amount of direct 100-point UQ as the combined co-Krig is still unlikely to be able to supply sufficiently accurate statistics directly.

As mentioned earlier, the cost of tuning and sampling Krigs can become the limiting factor in using the combined approaches. Here, the Krigs are restricted to 1170 data points at any one time, 30 points in an initial DoE sample plus 30 updates each of 38 more points (of which 25 are derived from searching the evolving combined Krig and 15 are used to infill the domain based on nearest-neighbour methods for four

Table 2: Comparisons of various blade designs: nominal and six designs from the combined Kriging final Pareto front.

Quantity	Baseline design	NSGA-II co-Krig optimized design (last six designs)					
Shape fn. 1	0.5	0.4000	0.4000	0.4000	0.4000	0.4000	0.4000
Shape fn. 2	0.5	0.4001	0.4000	0.4009	0.4000	0.4567	0.4567
Shape fn. 3	0.5	0.4039	0.4000	0.4411	0.4813	0.4000	0.4024
Shape fn. 4	0.5	0.5500	0.5500	0.5500	0.5500	0.5500	0.5500
Shape fn. 5	0.5	0.5053	0.5038	0.4668	0.5100	0.4000	0.4000
Shape fn. 6	0.5	0.5500	0.5500	0.5500	0.5500	0.5500	0.5500
Shape fn. 7	0.5	0.4989	0.4638	0.4987	0.4703	0.4541	0.4543
Shape fn. 8	0.5	0.4749	0.4999	0.4810	0.4944	0.5500	0.5158
Shape fn. 9	0.5	0.5227	0.5176	0.5189	0.5211	0.5130	0.5152
Shape fn. 10	0.5	0.5500	0.5500	0.5500	0.5500	0.4949	0.5500
Mean loss_{100}	1.1571	1.0885	1.0716	1.0944	1.0372	1.1146	1.1060
Std. dev._{100}	0.1234	0.0730	0.0811	0.0783	0.0804	0.0727	0.0727
Average mean loss_{100}	1.1571	1.0854					
Average std. dev._{100}	0.1234	0.0763					
Total fn. calls/search	1/100	$950 + 22 \times 100 = 3150$					

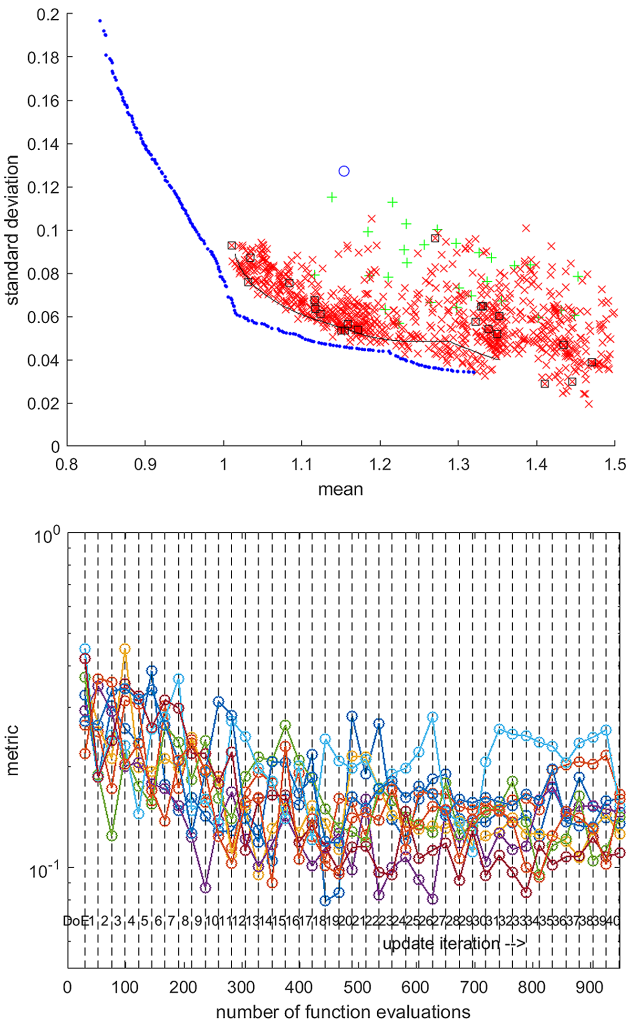

Figure 7: Combined Krig-assisted NSGA-II results – upper: estimated results for 500-point LPτ pseudo-Monte Carlo sampling on the combined Krig, (□) final generation, (+) initial generation, and (×) intermediate generations, along with the estimated Pareto front (solid line), 'true' Pareto front (dotted line), and (○) initial baseline design; lower: Pareto front normalized hypervolume metric convergence for reference point (5, 0.5) based on the estimate of the 'true' Pareto front, for nine independent runs.

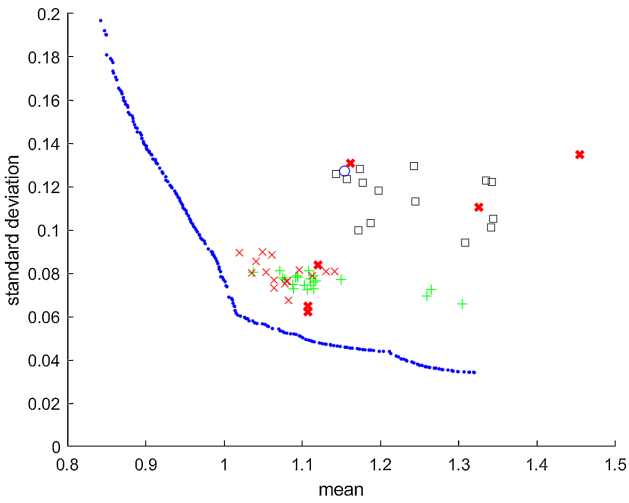

Figure 8: Final Pareto fronts after 100-point UQ for the three approaches discussed in the text: (□) co-Kriging, (+) combined Kriging, and (×) combined co-Kriging (update points are shown in bold), along with the estimated 'true' Pareto front and (○) initial baseline design.

of every five update cycles, while on the fifth the whole update process is focused on UQ of a single design variant). This budget is still less than half the total number of function evaluations used when using co-Krigs to represent the design variables and UQ separately, because again it leads to an extremely time-consuming Krig tuning process. However, now since UQ sampling is playing a more prominent role, fewer final UQ calculations are needed to validate the final Pareto front. So, having completed the search process, although 22 final designs are again confirmed by additional UQ using 100-point sampling the whole process uses the same budget as the previous two searches ($1170 + 6 \times 60 + 16 \times 100$). (That is, 6 of these being those already tested with 40-point UQ sampling during the update process only require further 60 runs to fully validate; the remaining 16 are taken from the NSGA-II search over the final combined co-Krig.) Figure 9 and Table 3 illustrate this approach in the same format as Figs 6 and 7 and Tables 1 and 2. In this case, the estimated Pareto front is not as extensive as for the simple combined Krig, but it still lies quite close to the 'true' one and the convergence plots are broadly similar to those for the com-

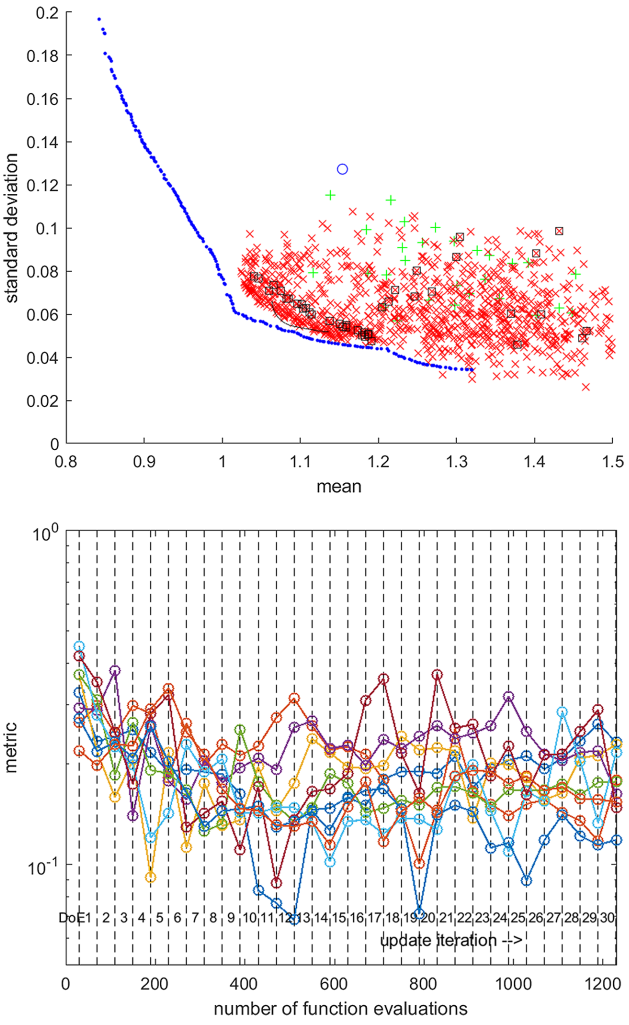

Figure 9: Combined co-Krig-assisted NSGA-II results – upper: estimated results for 500-point LPτ pseudo-Monte Carlo sampling on the combined co-Krig, (□) final generation, (+) initial generation, and (×) intermediate generations, along with the estimated Pareto front (solid line), 'true' Pareto front (dotted line), and (○) initial baseline design; lower: Pareto front normalized hypervolume metric convergence for reference point (5, 0.5) based on the estimate of the 'true' Pareto front, for nine independent runs.

bined Krig runs. Moreover, when one goes on to examine the best suggested designs with full 100-point LPτ sampling, things are rather better: the average performance of the designs shown in Table 3 is now a mean of 1.0639 and standard deviation of 0.0749 and the final designs validated in this way offer a clearly improved Pareto front (see again Fig. 8). Note that several of the points subjected to 40-point UQ during the search are broadly similar in performance to those coming from the direct co-Krig, but at the end of the search, the best designs seen to date have been delivered by the combined approach: the points tested with full UQ dominate all those coming from either the co-Krig or simple combined Krig methods, demonstrating the benefits of combined co-Kriging uncertainties in calculated results.

Even so, the results coming from all the search schemes still differ from the 'true' results to some extent. This is simply the result of the large number of design and noise variables in use, which leads to a huge search space that can only be sampled to a very limited extent given realistic computational budgets. Each update cycle helps and a range of possible updates can be considered, such as those based on expected improvement or hypervolume indicators – we use a mixture of such methods. In the end, however, in work such as this, final convergence must be given up in favour of worthwhile improvements.

Note that in all the searches reported here some form of random number or pseudo-random number sequence has been used. These are required because searches such as NSGA-II rely on random operations to work. It is therefore the case that all of the results reported vary with different sequences of random numbers and so to draw completely definite conclusions some more thorough investigation of averaging than just nine independent runs would ideally be needed, even though tens of thousands of individual design calculations have been carried out – unfortunately when working with resource-intensive RANS codes, as here, this is not possible even given the extensive computing power available in the cluster being used. It does, however, demonstrate the clear benefit of including variance measures in design optimization and various forms of Kriging as a method for compensating for limited number of Monte Carlo evaluations. Finally, it should be noted that it is current practice when designing blades for aero-engines to work mainly with 3D CFD models and over a range of operating points – the use of 2D approaches at a single condition has been adopted here simply to limit the computational cost of the study.

Table 3: Comparisons of various blade designs: nominal and combined co-Kriging designs.

Quantity	Baseline design	NSGA-II co-Krig optimized design (last six designs)					
Shape fn. 1	0.5	0.4000	0.4000	0.4000	0.4001	0.4000	0.4000
Shape fn. 2	0.5	0.4611	0.4871	0.5500	0.5500	0.4541	0.5102
Shape fn. 3	0.5	0.4000	0.4472	0.5271	0.5214	0.4012	0.4769
Shape fn. 4	0.5	0.5500	0.5500	0.5500	0.5500	0.5500	0.5498
Shape fn. 5	0.5	0.4000	0.5500	0.5500	0.5500	0.5494	0.5500
Shape fn. 6	0.5	0.5500	0.5500	0.5500	0.5500	0.5500	0.5500
Shape fn. 7	0.5	0.5159	0.4446	0.4356	0.4167	0.4723	0.4384
Shape fn. 8	0.5	0.4001	0.5394	0.5500	0.5500	0.4586	0.5500
Shape fn. 9	0.5	0.5500	0.5500	0.5500	0.5500	0.5500	0.5500
Shape fn. 10	0.5	0.5076	0.5500	0.5500	0.5500	0.5500	0.5500
Mean loss_{100}	1.1571	1.1075	1.0638	1.0540	1.0827	1.0200	1.0557
Std. dev._{100}	0.1234	0.0647	0.0732	0.0806	0.0675	0.0895	0.0737
Average mean loss_{100}	1.1571	1.0639					
Average std. dev._{100}	0.1234	0.0749					
Total fn. calls/search	1/100	$1170 + 6 \times 60 + 16 \times 100 = 3130$					

5. Conclusions

In this paper, we have shown how advanced surrogate models can be used to support robust design geometry optimization problems, both speeding up the search for good designs and supporting the process of design UQ. We have introduced a new scheme that takes ideas from co-Kriging and the use of combined surrogates that model both design and noise variable effects. This new scheme outperforms the other methods investigated here when applied to a real-world aerodynamic design problem carried out with industrial strength analysis codes.

Acknowledgements

The use of Rolls-Royce plc codes and the financial support of the ATI under the GemINIDS grant are gratefully acknowledged.

Conflict of interest statement

Declarations of interest: none.

References

Courrier, N., Boucard, P. A., & Soulier, B. (2015). Variable-fidelity modeling of structural analysis of assemblies. *Journal of Global Optimization*, 64, 577–613.

Deb, K., Samir, A., Amrit, P., & Meyarivan, T. (2000). A fast elitist non-dominated sorting genetic algorithm for multi-objective optimization: NSGA-II. *Lecture Notes in Computer Science*, 1917, 848–849.

Dellino, G., Kleijnen, J. P. C., & Meloni, C. (2012). Robust optimization in simulation: Taguchi and Krige combined. *INFORMS Journal on Computing*, 24, 471–484.

Forrester, A. I. J., Sobester, A., & Keane, A. J. (2008). *Engineering design via surrogate modelling: A practical guide*. Chichester, UK: Wiley.

Goldberg, D. E. (1989). *Genetic algorithms in search, optimization and machine learning*. Addison-Wesley Longman Publishing Co.

Han, Z.-H., & Görtz, S. (2012). Hierarchical Kriging model for variable-fidelity surrogate modeling. *AIAA Journal*, 50, 1885–1896.

Han, Z.-H., Zimmermann, R., & Görtz, S. (2012). Alternative cokriging model for variable-fidelity surrogate modeling. *AIAA Journal*, 50, 1205–1210.

Hicks, R. M., & Henne, P. (1978). Wing design by numerical optimization. *Journal of Aircraft*, 15, 407–412.

Jones, D. R., Schonlau, M., & Welch, W. J. (1998). Efficient global optimization of expensive black-box functions. *Journal of Global Optimization*, 13, 455–492.

Keane, A. J. (2003). Wing optimization using design of experiment, response surface, and data fusion methods. *Journal of Aircraft*, 40, 741–750.

Keane, A. J. (2006). Statistical improvement criteria for use in multiobjective design optimization. *AIAA Journal*, 44, 879–891.

Keane, A. J. (2009). Comparison of several optimisation strategies for robust turbine blade design. *Journal of Propulsion and Power*, 25, 1092–1099.

Keane, A. J. (2012). Use of co-kriging for robust design optimization. *AIAA Journal*, 50, 2351–2364.

Keane, A. J., & Nair, P. B. (2005). *Computational approaches for aerospace design*. Chichester, UK: Wiley.

Kennedy, M. C., & O'Hagan, A. (2000). Predicting the output from a complex computer code when fast approximations are available. *Biometrika*, 87, 1–13.

Kumar, A. (2008). *Robust design methodologies: Application to compressor blades* (PhD thesis). University of Southampton, Southampton, UK.

Kumar, A., Keane, A. J., Nair, P. B., & Shahpar, S. (2006). Robust design method of compressor fan blades against erosion. *ASME Journal of Mechanical Design*, 128, 864–873.

Moinier, P., & Giles, M. B. (1999). Preconditioned Euler and Navier–Stokes calculations on unstructured grids. In *Proceedings of the 6th ICFD Conference on Numerical Methods for Fluid Dynamics*.

Ng, L. W. T., & Willcox, K. E. (2014). Multifidelity approaches for optimization under uncertainty. *International Journal for Numerical Methods in Engineering*, 100, 746–772.

Peherstorfer, B., Willcox, K., & Gunzburger, M. (2016). Optimal model management for multifidelity Monte Carlo estimation. *SIAM Journal on Scientific Computing*, 38, 3163–3194.

Shahpar, S., & Lapworth, L. (2003). Parametric design and rapid meshing systems for turbomachinery optimization. In *ASME Turbo Expo* (pp. 579–590).

Statnikov, R. B., & Matusov, J. B. (1995). *Multicriteria optimization and engineering*. New York: Chapman & Hall.

Toal, D. J. J. (2016). A study into the potential of GPUs for the efficient construction & evaluation of Kriging models. *Engineering with Computers*, 32, 377–404.

Tsutsui, S., & Ghosh, A. (1997). Genetic algorithms with a robust solution searching scheme. *IEEE Transactions on Evolutionary Computation*, 1, 201–208.

Welch, W., Yu, T.-K., Kang, S. M., & Sacks, J. (1990). Computer experiments for quality control. *Journal of Quality Technology*, 22, 15–22.

Zimmermann, R., & Han, Z.-H. (2010). Simplified cross-correlation estimation for multi-fidelity surrogate cokriging models. *Advances and Applications in Mathematical Sciences*, 7, 181–201.