A Dynamic Sampling Method for Kriging and Cokriging Surrogate Models

Markus P. Rumpfkeil *
University of Dayton, Ohio, 45469, USA
Wataru Yamazaki †
Nagaoka University of Technology, 1603-1, Kamitomiokamachi Nagaoka, Niigata 940-2188, Japan
Dimitri J. Mavriplis ‡
University of Wyoming, Laramie, 82071, USA

In this paper we describe our gradient and Hessian enhanced Kriging surrogate model with dynamic sample point selection. We demonstrate the quality of the surrogate by comparison with higher-dimensional analytic test functions. We also apply the surrogate model to uncertainty quantification and robust optimization problems using inexpensive Monte-Carlo simulations. All applications benefit from the additional gradient and Hessian information as well as the dynamic sample point selection by requiring fewer function evaluations and overall less computational time.

Nomenclature

C_d	Drag coefficient
C_l	Lift coefficient
D	Design variables
$\mathcal{J}$	Objective function
$\frac{d\mathcal{J}}{dD_j}$	Gradient of objective function
$\frac{d^2\mathcal{J}}{dD_j dD_k}$	Hessian of objective function
M	Number of design variables
N	Number of sample points
μ_D	Mean of design variables
$\mu_{\mathcal{J}}$	Mean of objective function
σ_{D_j}	Standard deviation of design variable j
$Var_{\mathcal{J}}$	Variance of objective function

I. Introduction and Motivation

Computational methods have been playing an increasingly important role in science and engineering analysis and design over the last several decades, due to the rapidly advancing capabilities of computer hardware, as well as increasingly sophisticated and capable numerical algorithms. However, in spite of the rapid advances and acceptance of numerical simulations, serious deficiencies remain in terms of accuracy, uncertainty, and validation for many applications. Many real-world problems involve input data that is noisy or uncertain, due to measurement or modeling errors, approximate modeling parameters,[1] manufacturing tolerances,[2] in-service wear-and-tear, or simply the unavailability of the information at the time of the

*Assistant Professor, Dept. of Mechanical and Aerospace Engineering, Markus.Rumpfkeil@udayton.edu, Member AIAA
†Assistant Professor, Dept. of Mechanical Engineering, wyamazak@uwyo.edu, Member AIAA
‡Professor, Dept. of Mechanical Engineering, mavripl@uwyo.edu, Associate Fellow AIAA

decision.[3] These imprecise or unknown inputs are important in the design process and need to be quantified in some fashion. To this end, uncertainty quantification (UQ) has emerged as an important area in modern computational engineering. Today, it is no longer sufficient to predict specific objectives using a particular physical model with deterministic inputs. Rather, a probability distribution function (PDF) or interval bound of the simulation objectives is required as a function of the uncertainties inherent in the simulation input parameters, in order to establish confidence levels over a range of performance predictions.

Probabilistic assessment of uncertainty in computational models consists of three major phases: (i) data assimilation in which the input parameters are characterized (in terms of PDFs) from observations and physical evidence; (ii) uncertainty propagation in which the input variabilities are propagated through the mathematical model; and (iii) certification in which the output of the numerical predictions are characterized in terms of their statistical properties and confidence bounds are derived.[4] Arguably, the computationally most expensive part of UQ is the second phase. The simplest approach to obtain the output statistics in response to input distributions is the Monte-Carlo (MC) method,[5] in which a large number of independent calculations are computed; however, in many practical cases the number of realizations required is too large and results in prohibitively high computational cost, especially for complex high-fidelity physics-based simulations. Thus, the use of surrogate models for UQ has become popular. The idea of a surrogate model is to replace expensive function evaluations with an approximate but inexpensive functional representation which can be probed exhaustively if required. For example, when a surrogate model for an optimization problem is constructed with given training data, the most promising locations in the model can be explored with cheap computational cost. The accuracy of surrogate models can be increased efficiently by adding the exact function information in the most promising locations. This approach can save a lot of computational cost and enables the exploration of wider design spaces efficiently. For instance, in Peter and Marcelet[6] the performances of major surrogate models, such as least square polynomial, multi-layer perceptron, radial basis function (RBF), and Kriging, have been compared for a two dimensional turbomachinery problem. The Kriging and RBF models showed the best performance. Thus, the Kriging model, originally developed in the field of geological statistics, has gained popularity.[6–16] The Kriging surrogate model predicts the function value by using stochastic processes, and has the flexibility to represent multimodal functions.

An efficient gradient evaluation method based on adjoint formulations has been adopted by the computational community for data-assimilation and design optimization problems over the last several decades.[17,18] Thus, the introduction of gradient information within surrogate models as additional training data has also attracted attention. A gradient enhanced direct as well as indirect Kriging (called direct or indirect co-Kriging) has been developed in the surrogate model community and has shown very beneficial results.[11,14,15] While adjoint methods provide an effective approach for computing first-order sensitivity derivatives, the ability to compute second-order sensitivity derivatives is also highly desirable for many science and engineering simulation problems.[19–23] On the one hand, the availability of Hessian information allows the use of much stronger Newton optimization strategies, which holds the potential for greatly reducing the expense of solving difficult optimization problems. On the other hand, second-order sensitivity information can be used effectively to devise efficient uncertainty propagation methods and inexpensive Monte-Carlo (IMC) techniques[22] for characterizing PDFs of computed simulation results. Since an efficient Hessian evaluation method has been developed by us,[24,25] it is very promising to utilize the Hessian information within surrogate models in addition to the gradient information.[26,27] The reason for this optimism is the observation that, for computational high-fidelity applications targeting a single output objective, the effort for computing the full gradient is, thanks to the adjoint, comparable to the effort of computing the objective function itself. Therefore, as the number of inputs, M, increases, using the output function and its derivative information is appealing, because it provides $M + 1$ pieces of information for roughly the cost of two function evaluations. Similarly, the Hessian provides $M \cdot (M + 1)/2$ pieces of information for roughly the cost of M function evaluations since, in general, the most efficient full Hessian constructions require the solution of M forward linear problems (one corresponding to each input parameter).[19,22] Thus, one can reasonably expect to have to compute the output function overall far fewer times to obtain a good surrogate model when using gradient and Hessian information and this should also scale more reasonably to higher dimensions.

The outline of this paper is as follows. In Section II we describe our gradient and Hessian enhanced surrogate model with dynamic sample point selection and demonstrate the quality of these surrogate models by means of analytic test functions in Section III. Section IV then applies the surrogate model to UQ as well as robustness analysis problems. Section V concludes this paper.

II. Construction of Dynamically Sampled Kriging Model

Recently we have developed a gradient enhanced direct as well as indirect Kriging model.[16,26,27] In the direct co-Kriging approach, the covariances between function values, function values and gradients, as well as gradients have to be considered within the correlation matrix as opposed to the original Kriging formulation where only the covariances between function values have to be considered. Thus, the correlation matrix R_{dcok} becomes asymmetric and its size increases to $N \cdot (M+1)$ (from N), where N is the number of sample points, and M represents the number of input parameters. On the other hand, the formulation of the indirect co-Kriging model is exactly the same as that of the original Kriging model. In this approach, additional sample points are constructed around a real sample point by using a Taylor series extrapolation. Usually, one additional point for each gradient direction is produced for any real sample point. If all real sample points have gradient information, the augmented number of sample points becomes $N \cdot (M+1)$ which is the same as the size of R_{dcok}. The principal advantage of the indirect co-Kriging approach is its ease of implementation. The major disadvantage of indirect co-Kriging is the fact that additional sample points will be close to a real sample point (to reduce the error from the Taylor series extrapolation), which tends to produce ill-conditioned correlation matrices. Direct and indirect co-Kriging models produce identical results in the limit of small real to extrapolated point step sizes, although the direct approach is preferable due to its robustness and lack of tunable parameters. Utilizing Hessian information in addition to gradient information within the Kriging surrogate model is not a trivial task, since for the indirect approach the correlation matrix becomes even more ill-conditioned and for the direct approach up to fourth-order derivatives of the covariance are required (we use automatic differentiation for this in our implementation).

In order to obtain a globally accurate surrogate model, we refine the building of the model by a dynamic sample point selection with a stopping criteria rather than just specifying the sample size in the beginning and picking the sample points randomly. This is similar to the concept of expected improvement[9,28] (EI) when optimizing with a Kriging model where a potential for improvement is used which considers both estimated function values and uncertainties in the surrogate model, thereby keeping the balance between global and local search performance.

We construct a local response surface using a hybrid of extrapolation and interpolation involving a few, already existing, sample points D_i, $i = 1 \ldots, I$ in order to guide the sampling process. The function values and available derivatives at each sample point are used to construct extrapolated function values of order n_e for a test candidate location, D. The extrapolations from the sample points are then weighted with a low-order interpolant of order n_i to find a unique function value $\mathcal{J}(D)$. This approach has been coined Dutch Intrapolation[29] (DI) and it has been shown that the order of accuracy of the intrapolant, n, is equal to $n_i + n_e$, that is, using function, gradient, and Hessian information for the extrapolations and second-order interpolation leads to a fourth-order accurate intrapolant. This situation is shown in Figure 1 where the two-dimensional Rosenbrock function (a fourth-order polynomial) is represented exactly using function, gradient, and Hessian information in six sample points chosen randomly using latin hypercube sampling.[30]

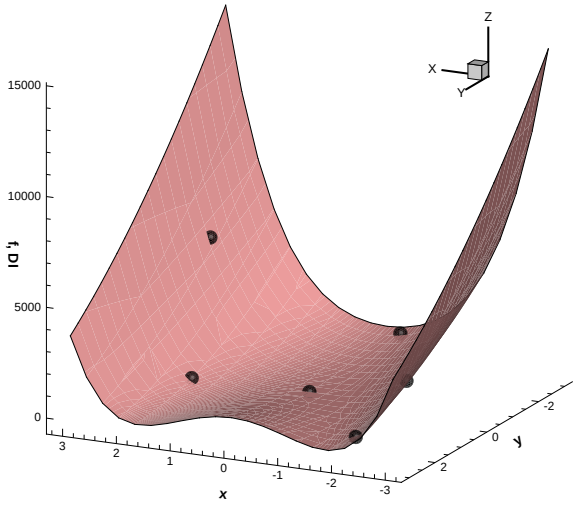

Figure 1. Comparison between two-dimensional Rosenbrock function and Dutch Intrapolation. The six spheres are the sample point locations with function, gradient and Hessian information.

The Dutch extrapolation functions are normal multivariate Taylor expansions of order n_e with a correction term given in multi-index notation by[29]

$$\mathcal{T}^{n_e}(D, D_i) = \sum_{|k|\geq 0}^{|k|\leq n_e} \frac{a_k^{n_e}}{k!}(D - D_i)^k \partial^k \mathcal{J}(D_i) \quad \text{for} \quad i = 1, \ldots, I \quad \text{with} \quad a_k^{n_e} = 1 - k/(n_e + 1), \tag{1}$$

where $\mathcal{J}(D_i)$ is the function value of sample point D_i. It is important to note that although the Dutch Taylor expansions are discussed here for general order n_e, practical applications are usually restricted to low values of n_e. The range of practical applicability is similar to that of "normal" Taylor expansions. High-order Taylor expansions are often used in theoretical formulations, however, in practical applications their use is limited because the convergence with increasing order is typically very slow, and the region of convergence very small. Thus, the Dutch Taylor expansions are to be used in small regions where the function to be approximated is well represented by a low-order polynomial, that is, where the Taylor expansion coefficients decrease quickly for increasing order. In addition, it becomes impractical to calculate higher-order derivatives of the objective function for high-fidelity physics-based simulations.

For the interpolation a true linear or quadratic interpolating polynomial for arbitrary nodes in any dimension is used which requires a total of $I = \binom{M+n_i}{M} = \frac{(M+n_i)!}{M!n_i!}$ nodes. For example, for $M = 2$ and $n_i = 2$ the intrapolated function value for each test candidate $D = (D_x, D_y)$ is given by the values of six Dutch extrapolations $\mathcal{T}^{n_e}(D, D_i)$ around the nodes $D_i = (D_{i,x}, D_{i,y})$, $i = 1, \ldots, 6$:

$$\mathcal{J}_{\text{DI}}(D) = (1 \;\; D_x \;\; D_y \;\; D_x^2 \;\; D_x D_y \;\; D_y^2) \cdot \mathbf{b} \qquad \text{with} \qquad \mathbf{V}\,\mathbf{b} = \mathbf{f}$$

where

$$\mathbf{V} = \begin{pmatrix} 1 & D_{1,x} & D_{1,y} & D_{1,x}^2 & D_{1,x}D_{1,y} & D_{1,y}^2 \\ 1 & D_{2,x} & D_{2,y} & D_{2,x}^2 & D_{2,x}D_{2,y} & D_{2,y}^2 \\ 1 & D_{3,x} & D_{3,y} & D_{3,x}^2 & D_{3,x}D_{3,y} & D_{3,y}^2 \\ 1 & D_{4,x} & D_{4,y} & D_{4,x}^2 & D_{4,x}D_{4,y} & D_{4,y}^2 \\ 1 & D_{5,x} & D_{5,y} & D_{5,x}^2 & D_{5,x}D_{5,y} & D_{5,y}^2 \\ 1 & D_{6,x} & D_{6,y} & D_{6,x}^2 & D_{6,x}D_{6,y} & D_{6,y}^2 \end{pmatrix} \qquad \mathbf{b} = \begin{pmatrix} b_1 \\ b_2 \\ b_3 \\ b_4 \\ b_5 \\ b_6 \end{pmatrix} \qquad \mathbf{f} = \begin{pmatrix} \mathcal{T}^{n_e}(D, D_1) \\ \mathcal{T}^{n_e}(D, D_2) \\ \mathcal{T}^{n_e}(D, D_3) \\ \mathcal{T}^{n_e}(D, D_4) \\ \mathcal{T}^{n_e}(D, D_5) \\ \mathcal{T}^{n_e}(D, D_6) \end{pmatrix}.$$

The matrix $\mathbf{V}$ is a generalization of a one-dimensional Vandermonde matrix.[31]

The proposed dynamic sampling method works as follows: Start by evaluating the function (gradient and Hessian) value at the center of the domain. Then pick an additional amount of sample points via latin hypercube sampling[30] and evaluate their function (gradient and Hessian) such that the initial number of sample points is equal to $\binom{M+n_i}{M}$. Then repeat the following steps until convergence or until a maximum amount of function (gradient and Hessian) evaluations has been reached

1. Specify a set of test candidates via latin hypercube sampling.
2. Construct a local function value for each test candidate using Dutch Intrapolation as described above involving an appropriate number of closest neighbors with function (gradient and Hessian) information.
3. Compare the global Kriging surrogate model function value predictions for the test candidates with the local Dutch Intrapolations.
4. Add a user-specified number of test candidates (we add M) with the worst discrepancy between the two values to the set of sample points, only then evaluating the real function (gradient and Hessian).

We define convergence as having the worst discrepancy below a certain threshold. We also augment the selection process by geometric criteria, for example, we make sure that the distance of a test candidate to the nearest existing sample point is above the average distance of all test candidates to their respective closest sample point. One could also use the discrepancy in step three as a mean to decide whether one wants to evaluate the real gradient or Hessian which may not be necessary if this particular area of the design space is relatively flat.

Another option for the first step in the procedure is to generate a mesh using a high-dimensional Delaunay triangulation in the sample space (see Figure 2 for a two-dimensional example) and to specify a set of test candidates geometrically (we pick the centers of the hyper-triangles and the midpoints of the edges) rather than using latin hypercube sampling. This, however, requires that we start with all the corners of the domain as initial sample points which scales with 2^M rather than $\binom{M+n_i}{M}$. In addition, the Delaunay triangulation becomes the bottleneck of this procedure for more than half a dozen or so inputs.

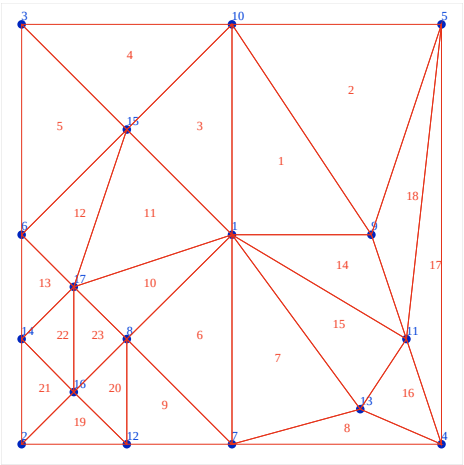

Figure 2. An example of a two-dimensional Delaunay triangulation in the sample space. Points numbered 1 to 5 are the initial points and 12 points have been added at this time.

III. Accuracy of Dynamically Sampled Kriging Model

We will use three different analytic test functions on the hypercube $[-2, 2]^M$ to demonstrate the quality of our surrogate:

1. A multidimensional Cosine function: $f_1(x_1, \ldots, x_M) = \cos(x_1 + \ldots + x_M)$
2. The multidimensional Runge function: $f_2(x_1, \ldots, x_M) = \frac{1}{1+x_1^2+\ldots+x_M^2}$
3. The multidimensional Rosenbrock function: $f_3(x_1, \ldots, x_M) = \sum_{i=1}^{M-1} \left[(1 - x_i)^2 + 100(x_{i+1} - x_i^2)^2\right]$

Plots of all three functions in two dimensions together with the direct co-Kriging surrogate constructed using the five function and gradient values in the center and four corners of the domain are shown in Figure 3.

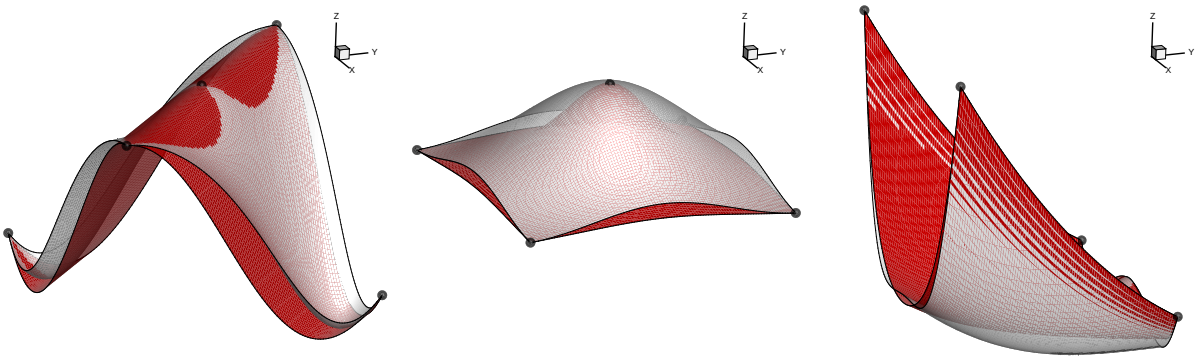

Figure 3. Three analytic test functions in two dimensions (red) together with the surrogate constructed from five function and gradient values (white). Left: Cosine function f_1. Middle: Runge function f_2. Right: Rosenbrock function f_3 (with different scaling in the z-direction than f_1 and f_2).

In Figure 4 we compare the root-mean square error between the two-dimensional test functions and the Kriging surrogate models (calculated by comparing the real and predicted function values on a Cartesian mesh with 101×101 nodes) versus the number of sample points used to construct the surrogate. The sample points always include the center of the domain and the others are either all selected through latin hypercube sampling[30] (dashed lines) or we start with five latin hypercube sampled points and add additional points through dynamic sampling (thick solid lines) as described in the previous section using a second-order interpolant. The thin solid lines show the results of the dynamic sampling with the Delaunay triangulation.

One can clearly see that the gradient (FG) as well as the gradient and Hessian enhanced Kriging models (FGH) perform much better than the model that is only based on function evaluations (F) for all three functions. One can also infer that the dynamic sample point selection performs better than just selecting all

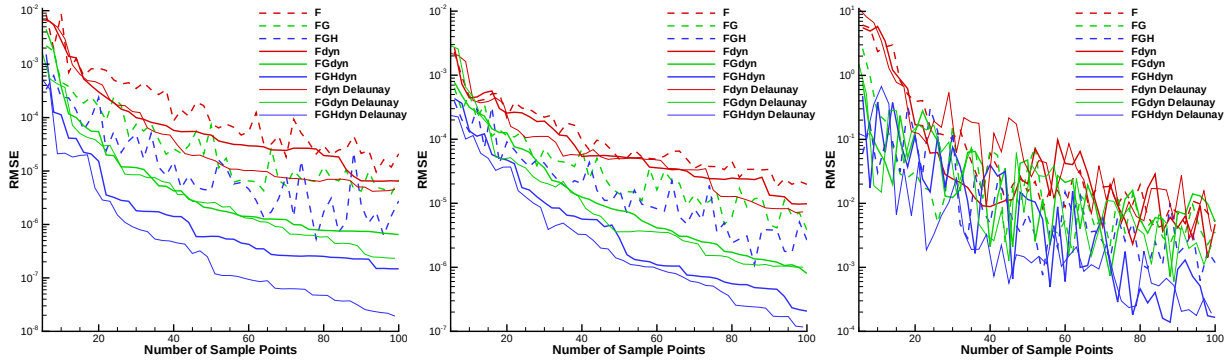

Figure 4. The RMS error (RMSE) between the two-dimensional test functions and the surrogate models vs. number of sample points. Sample points are either selected through latin hypercube sampling (dashed line) or added via dynamic sampling using random test candidates (thick solid line) or Delaunay triangulation (thin solid line). Left: Cosine function f_1. Middle: Runge function f_2. Right: Rosenbrock function f_3.

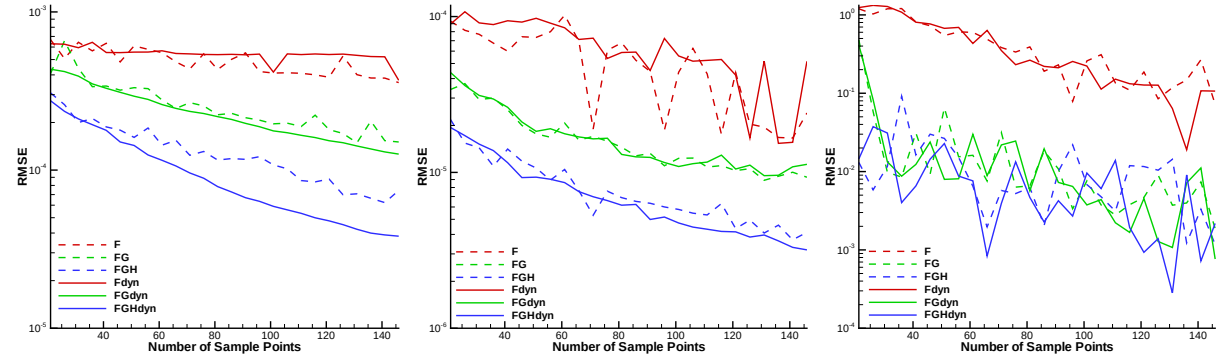

Figure 5. The RMS error (RMSE) between the five-dimensional test functions and the surrogate models vs. number of sample points. Sample points are either selected through latin hypercube sampling (dashed line) or added via dynamic sampling (solid line). Left: Cosine function f_1. Middle: Runge function f_2. Right: Rosenbrock function f_3.

sample points through latin hypercube sampling. It also helps to reduce the effect due to randomness that more sample points do not necessarily lead to a better surrogate model. Figure 5 shows similar results for the five-dimensional test functions (compared on a Cartesian mesh with 18 nodes in each dimension). Here, we omit the results of the Delaunay triangulation dynamic sampling because there is already a significant overhead in five dimensions and the results are similar to the dynamic sampling with random test candidates.

IV. Uncertainty Quantification

It is important in UQ to differentiate between epistemic and aleatory uncertainty. Epistemic uncertainty (or type B, or reducible uncertainty) represents a lack of knowledge about the appropriate value to use for a quantity.[32] In contrast, uncertainty characterized by inherent randomness is called aleatory uncertainty (or type A, or irreducible uncertainty). Epistemic uncertainty may or may not be modeled probabilistically, however, regulatory agencies and design teams are increasingly being asked to specifically characterize and quantify epistemic uncertainty and separate its effect from that of aleatory uncertainty.[33]

For epistemic uncertainties MC methods may be employed, but the results can only be interpreted with regards to the interval produced on the output functional, with no inferred statistical distribution. Other approaches for propagating epistemic uncertainties, such as Dempster-Shafer evidence theory[34–36] also typically require a large number of function evaluations in part because it is a generalization of classical probability theory which is non-intrusive. An even simpler approach for determining epistemic uncertainty output intervals is to pose the problem as a constrained optimization problem of finding the minimum and maximum value of the output functional. One could use traditional Newton or Quasi-Newton optimization techniques, which often scale only weakly with the dimensionality of the problem but may produce only local optima, or global optimization methods which again require a large number of function evaluations. For all these methods, the construction of surrogate models is one of the most effective options since one can then capital-

ize on the cheap function evaluations of the surrogate. Thus, the development of efficient surrogate models potentially enhanced by gradient and Hessian information constitutes an important avenue for reducing the cost of UQ for both aleatory and epistemic uncertainties. To address the "curse of dimensionality" whereby the cost of quantifying uncertainty increases rapidly with the number of inputs, we combine two different strategies: firstly, select only the input parameters that are truly relevant to the simulation outcome through a sensitivity analysis and thus reduce the dimension of the problem at the outset; secondly, exploit the information gain at reduced additional cost with sensitivities as described in the introduction.

If one is only interested in the mean and standard deviation of an objective function, moment methods can be a good choice.[19,37] Moment methods are based on Taylor series expansions of the original non-linear objective function $\mathcal{J}(D)$ about the mean of the input, μ_D, given standard deviations, σ_{D_j}. The resulting mean $\mu_{\mathcal{J}}$ and variance $\mathrm{Var}_{\mathcal{J}}$ of the objective function are given to first order (MM1) by

$$\mu_{\mathcal{J}}^{(1)} = \mathcal{J}(\mu_D) \qquad\qquad \mathrm{Var}_{\mathcal{J}}^{(1)} = \sum_{j=1}^{M} \left(\frac{d\mathcal{J}}{dD_j}\bigg|_{\mu_D} \sigma_{D_j} \right)^2, \tag{2}$$

and to second order (MM2) by

$$\mu_{\mathcal{J}}^{(2)} = \mu_{\mathcal{J}}^{(1)} + \frac{1}{2}\sum_{j=1}^{M} \left(\frac{d^2\mathcal{J}}{dD_j^2}\bigg|_{\mu_D} \sigma_{D_j}^2 \right) \qquad \mathrm{Var}_{\mathcal{J}}^{(2)} = \mathrm{Var}_{\mathcal{J}}^{(1)} + \frac{1}{2}\sum_{j=1}^{M}\sum_{k=1}^{M} \left(\frac{d^2\mathcal{J}}{dD_j dD_k}\bigg|_{\mu_D} \sigma_{D_j}\sigma_{D_k} \right)^2. \tag{3}$$

Note that in the latter case, the non-linear shift between the mean of the output and the output of the mean is accounted for by the Hessian diagonal elements. On the other hand, the method of moments provides no information on the distribution function of the output and when a complete PDF of the objective function is desired, a full non-linear MC simulation represents the most straight-forward approach for propagating uncertainties through the simulation process. Because this approach relies on a large number of repeated simulations, it is most often not practical for use with high-fidelity simulations. However, since we have an accurate surrogate model as demonstrated in the previous section, we can instead probe the surrogate exhaustively for an inexpensive MC (IMC) at relatively low cost. We prescribe the mean value of the MC samples as the center of the Kriging domain and the boundary is three standard deviations away in all dimensions. This means that for a normally distributed input variable more than 99 percent of all samples fall within the domain and the less accurate extrapolation capabilities of the Kriging surrogate model have only to be used for a small fraction.

An even cheaper method for an IMC simulation is to simply use extrapolation[22] around the function value of the mean of the inputs, $\mathcal{J}(\mu_{\mathcal{D}})$. A linear extrapolation (Lin) for the function value of the sample point, D, is given by

$$\mathcal{J}_{\mathrm{Lin}}(D) = \mathcal{J}\Big(\mu_D\Big) + \frac{d\mathcal{J}}{dD}\bigg|_{\mu_D} \cdot (D - \mu_D), \tag{4}$$

and a quadratic extrapolation (Quad) by

$$\mathcal{J}_{\mathrm{Quad}}(D) = \mathcal{J}_{\mathrm{Lin}}(D) + \frac{1}{2}(D-\mu_D)^T \cdot \frac{d^2\mathcal{J}}{dD^2}\bigg|_{\mu_D} \cdot (D - \mu_D). \tag{5}$$

We apply all of these methods to the analytic test functions of the previous section and to a transonic airfoil problem in the next two Subsections.

IV.A. Analytic Test Functions

As a first test we generate $50,000$ normally distributed MC sample points in two dimensions through latin hypercube sampling with a mean of $\mu_D = (0,0)$ and a standard deviation of $\sigma_{D_1} = \sigma_{D_2} = 0.6$ for the three analytic test functions presented in Section III. Results for the mean and variance predictions using the real function values, MM1 and MM2, as well as linear and quadratic extrapolation around the mean value are shown in Table 1. As can be seen MM1 and Lin yield very similar results as expected from the leading error. Also, MM2 and Quad give similar results for the same reason. In Figure 6 we show the error in the mean and variance predictions by using the Kriging surrogate models for an IMC versus the number of sample points used to construct the surrogate. One can see that the gradient (FG) as well as the gradient and Hessian enhanced Kriging models (FGH) perform better than the model that is only based on function evaluations (F) for all three functions. One can also infer that the dynamic sample point selection usually performs better than just selecting all the sample points through latin hypercube sampling. Another important observation is that the mean and variance predictions are already quite good just using twenty or so sample

points with function, gradient (and Hessian) information. Table 2 and Figure 7 show similar results for the five-dimensional test functions using $\mu_D = 0$ and $\sigma_{D_j} = 0.6$.

Table 1. Comparison of Mean and Variance predictions in two dimensions.

	Cosine fct		Runge fct		Rosenbrock fct	
	Mean	Variance	Mean	Variance	Mean	Variance
Real	0.697	0.133	0.658	4.11×10^{-2}	76.19	2.51×10^{4}
MM1	1.0	0.0	1.0	0.0	1.0	1.44
Lin	1.0	0.0	1.0	0.0	1.0	1.44
MM2	0.640	0.259	0.280	5.18×10^{-1}	37.36	2.59×10^{3}
Quad	0.639	0.259	0.280	5.19×10^{-1}	37.36	2.59×10^{3}

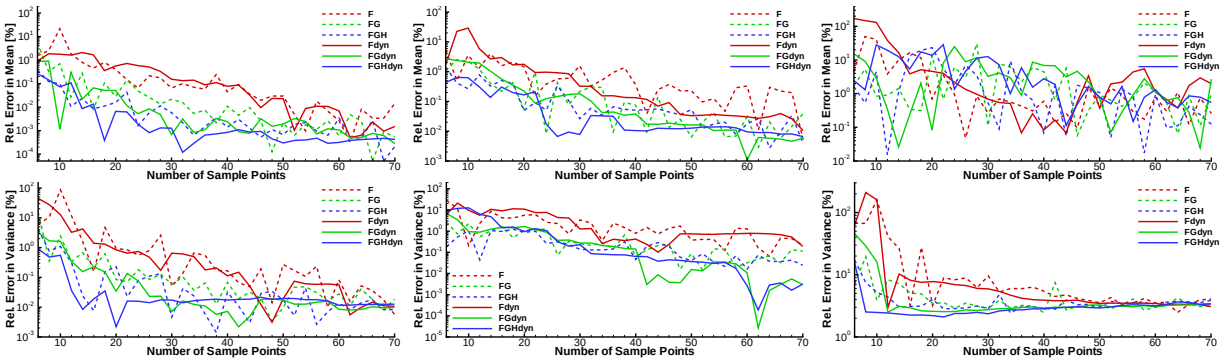

Figure 6. The error in the mean and variance between the two-dimensional test functions and the surrogate models vs. number of sample points. Sample points are either selected through latin hypercube sampling (dashed line) or added via dynamic sampling (solid line). Left: Cosine function f_1. Middle: Runge function f_2. Right: Rosenbrock function f_3.

Table 2. Comparison of Mean and Variance predictions in five dimensions.

	Cosine fct		Runge fct		Rosenbrock fct	
	Mean	Variance	Mean	Variance	Mean	Variance
Real	0.407	0.349	0.413	2.42×10^{-2}	304.6	1.32×10^{5}
MM1	1.0	0.0	1.0	0.0	4.0	5.76
Lin	1.0	0.0	1.0	0.0	4.0	5.77
MM2	0.1	1.62	−0.8	1.30	149.4	1.05×10^{4}
Quad	0.101	1.60	−0.80	1.29	149.4	1.05×10^{4}

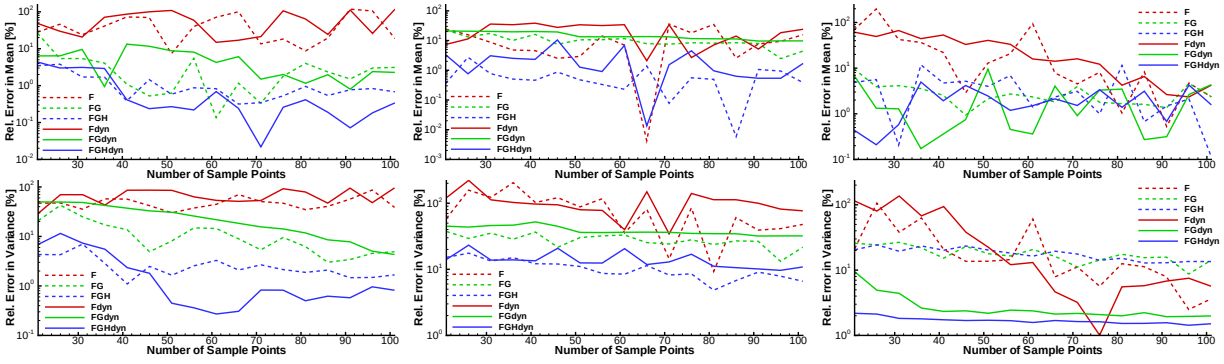

Figure 7. The error in the mean and variance between the five-dimensional test functions and the surrogate models vs. number of sample points. Sample points are either selected through latin hypercube sampling (dashed line) or added via dynamic sampling (solid line). Left: Cosine function f_1. Middle: Runge function f_2. Right: Rosenbrock function f_3.

IV.B. Robustness Analysis of a Transonic Airfoil

We consider the steady inviscid case of a transonic NACA 0012 airfoil as flow example which is described in more detail in Mani and Mavriplis.[38,39] The computational mesh has about 20,000 triangular elements and the free-stream Mach number is 0.755 with an angle of attack of 1.25 degrees. The non-dimensionalized pressure contours for this flow are shown in Figure 8 leading to a lift and drag coefficient of $C_l = 0.268$ and $C_d = 0.00521$, respectively.

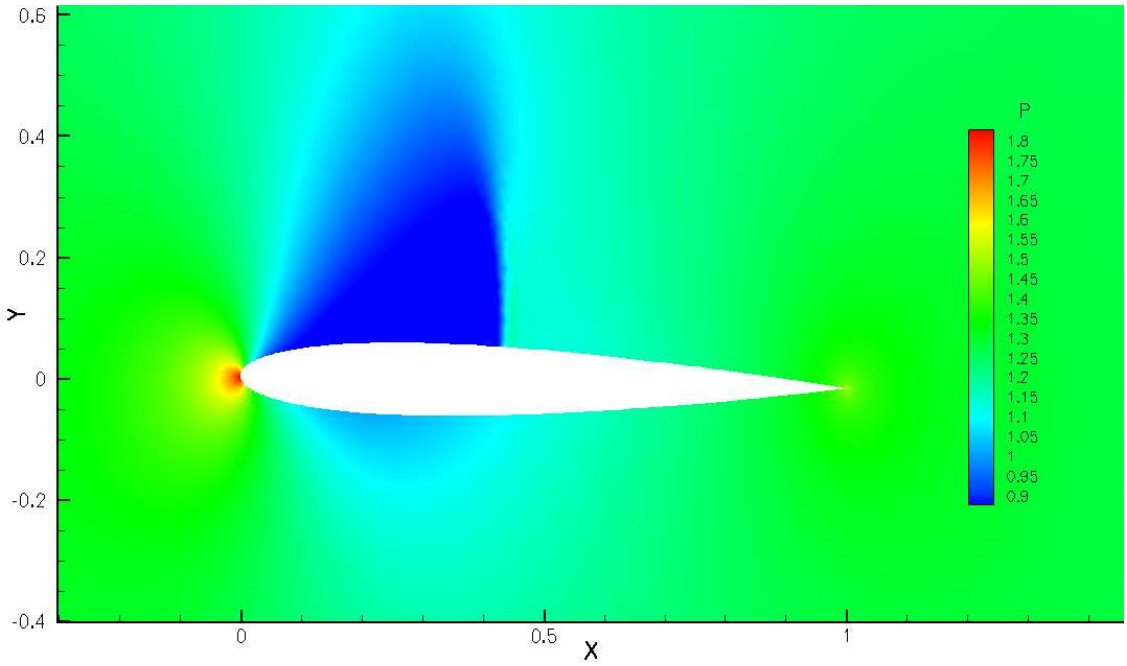

Figure 8. Non-dimensionalized pressure contours for $M_\infty = 0.755$ and $\alpha = 1.25$.

We want to perform a robustness analysis of the lift coefficient C_l with respect to shape variations of the airfoil. Therefore, we allow three shape design variables on the upper surface and three on the lower surface to vary which control the magnitude of Hicks-Henne sine bump functions.[40] The required deformation of the mesh is calculated via a linear tension spring analogy.[38,41] We assume that all six design variables have aleatory uncertainties which we model with the same normal distributions. The mean is set to zero (corresponding to the NACA 0012 airfoil) and different standard deviations between 0.0 and 0.01 are considered. Figure 9 shows the original NACA 0012 airfoil and the airfoils resulting from perturbations of ± 0.005. The required MC samples are generated using latin hypercube sampling with a sample size of 3,000.

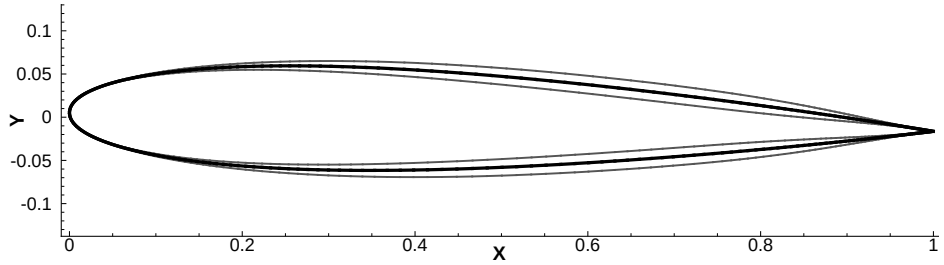

Figure 9. The NACA 0012 airfoil (in black) and airfoils resulting from perturbations of ± 0.005 (in gray).

One flow solve takes about 10 seconds on twelve Intel Xeon processors with 3.33 GHz each and the adjoint solve for the gradient as well as the forward solves for each design variable for the Hessian calculation take about the same time. However, the Hessian was only used for the second-order moment method because it did not improve the Kriging models due to a noisy design space and convergence problems for large perturbations of the design variables.

We show the mean and variance of the lift coefficient versus the standard deviation of the six shape design

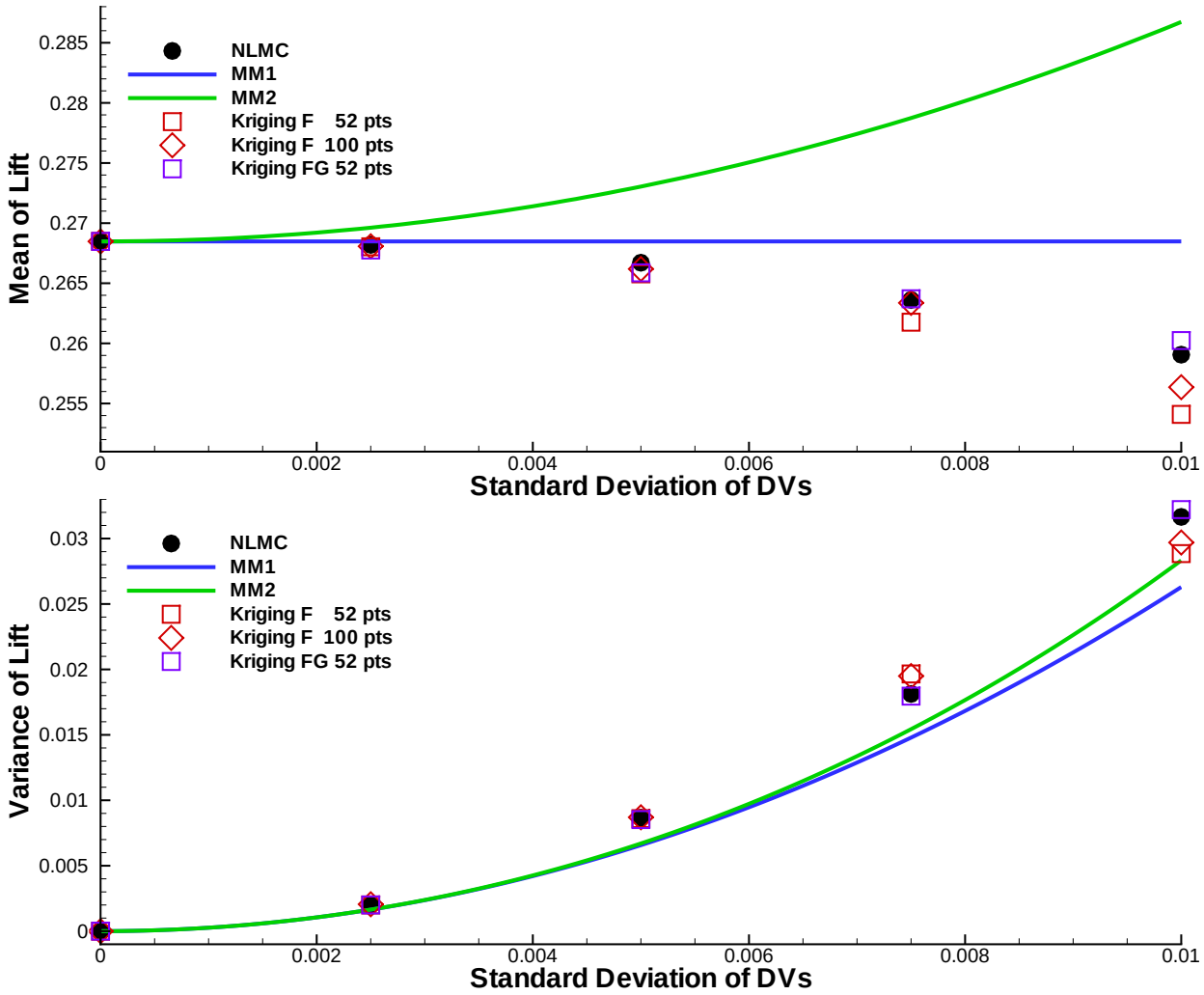

Figure 10. Two-dimensional transonic NACA 0012 shape robustness analysis.

variables in Figure 10. The blue line shows the results of the first-order moment method (MM1) given by equation (2) and the blue line that of the second-order moment method (MM2) given by equation (3). For comparison purposes a full non-linear MC (NLMC) simulation for a few selected standard deviations is also shown in the same figure as black circles. The red and purple squares show the results of the original Kriging (Kriging F) and the gradient enhanced Kriging (Kriging FG) using 52 sample points each and built with our dynamic sampling approach. Lastly, the red diamonds represent the results probing the original Kriging model built using 100 sample points and dynamic sampling which is close in computational cost to the gradient enhanced Kriging model with 52 sample points.

The moment methods give reasonable answers for smaller standard deviations but once these are above 0.005 the moment methods start to deviate. MM2 even gives the wrong tendency and predicts a higher mean lift than that of the original airfoil which is clearly not supported by the NLMC results. On the other hand, the Kriging models yield reasonable answers for a fraction of the cost of a full NLMC simulation.

V. Conclusions

We described our gradient and Hessian enhanced Kriging surrogate model with dynamic sample point selection. We demonstrated the quality of the surrogate by comparison with higher-dimensional analytic test functions. We also applied the surrogate model to uncertainty quantification and robustness analysis problems using inexpensive Monte-Carlo simulations. All applications benefited from the additional gradient and Hessian information by requiring fewer function evaluations and overall less computational time.

Acknowledgments

This work was partially supported by the US Air Force Office of Scientific Research under AFOSR Grant number FA9550-07-1-0164. The funding of the second author by the Japan Society for the Promotion of Science is gratefully acknowledged. We are also very grateful to Karthik Mani for making his flow and adjoint solver available to us.

References

[1]Luckring, J. M., Hemsch, M. J., and Morrison, J. H., "Uncertainty in computational aerodynamics," AIAA Paper, 2003-0409, Jan. 2003.

[2]Gumbert, C. R., Newman, P. A., and Hou, G. J., "Effect of random geometric uncertainty on the computational design of 3-D wing," AIAA Paper, 2002-2806, June 2002.

[3]Ben-Tal, A., Ghaoui, L. E., and Nemirovski, A., "Foreword: special issue on robust optimization," *Mathematical Programming*, Vol. 107(1-2), 2006, pp. 1–3.

[4]Oberkampf, W. and Barone, M., "Measures of agreement between computation and experiment: validation metrics," *Journal of Computational Physics*, Vol. 217, 2006, pp. 5–36.

[5]Metropolis, N. and Ulam, S., "The Monte Carlo method," *Journal of the American Statistical Association*, Vol. 44, 1949, pp. 335–341.

[6]Peter, J. and Marcelet, M., "Comparison of Surrogate Models for Turbomachinery Design," *WSEAS Transactions on Fluid Mechanics*, Vol. Vol. 3(1), 2008.

[7]Cressie, N., "The Origins of Kriging," *Mathematical Geology*, Vol. Vol. 22, No. 3, 1990, pp. 239–252.

[8]Koehler, J. R. and Owen, A. B., "Computer Experiments," *Handbook of Statistics, Ghosh, S., Rao, C.R., (Eds.)*, pp. 261-308, 1996.

[9]Jones, D. R., Schonlau, M., and Welch, W. J., "Efficient Global Optimization of Expensive Black-Box Functions," *Journal of Global Optimization*, Vol. Vol. 13, 1998, pp. 455492.

[10]Simpson, T. W., Korte, J. J., Mauery, T. M., and Mistree, F., "Comparison of Response Surface and Kriging Models for Multidisciplinary Design Optimization," AIAA Paper, 98-4755, 1998.

[11]Chung, H. S. and Alonso, J. J., "Using Gradients to Construct Cokriging Approximation Models for High-Dimensional Design Optimization Problems," AIAA Paper, 20020317, 2002.

[12]Martin, J. D. and Simpson, T. W., "Use of Kriging Models to Approximate Deterministic Computer Models," *AIAA Journal*, Vol. Vol. 43, No.4, 2005, pp. 853–863,.

[13]Jeong, S., Murayama, M., and Yamamoto, K., "Efficient Optimization Design Method Using Kriging Model," *Journal of Aircraft*, Vol. Vol. 42, No. 2, 2005, pp. 413–420.

[14]Laurenceau, J. and Sagaut, P., "Building Efficient Response Surfaces of Aerodynamic Functions with Kriging and Cokriging," *AIAA Journal*, Vol. Vol. 46, No. 2, 2008, pp. 498–507.

[15]Laurenceau, J. and Meaux, M., "Comparison of Gradient and Response Surface Based Optimization Frameworks Using Adjoint Method," AIAA Paper, 2008-1889, 2008.

[16]Yamazaki, W., Mouton, S., and Carrier, G., "Efficient Design Optimization by Physics-Based Direct Manipulation Free-Form Deformation," AIAA Paper, 2008-5953, 2008.

[17]Pironneau, O., "On Optimum Design in Fluid Mechanics," *Journal of Fluid Mechanics*, Vol. 64, No. 1, 1974, pp. 97–110.

[18]Errico, R. M., "What is an adjoint model?" *Bulletin of the American Meteorological Society*, Vol. 8(11), 1997, pp. 25772591.

[19]Sherman, L. L., Taylor III, A. C., Green, L. L., and Newman, P. A., "First- and second-order aerodynamic sensitivity derivatives via automatic differentiation with incremental iterative methods," *Journal of Computational Physics*, Vol. 129, 1996, pp. 307 – 331.

[20]Dixon, L. C., "Automatic Differentiation: Calculation of the Hessian," *Encyclopedia of Optimization*, Kluwer Academic Publishers, Dordrecht, The Netherlands, 2001, pp. 82–86.

[21]Taylor III, A. C., Green, L. L., Newman, P. A., and Putko, M., "Some Advanced Concepts in Discrete Aerodynamic Sensitivity Analysis," *AIAA Journal*, Vol. 41 No. 7, 2003, pp. 1224–1229.

[22]Ghate, D. and Giles, M. B., "Inexpensive Monte Carlo uncertainty analysis," *Recent Trends in Aerospace Design and Optimization, Tata McGraw-Hill, New Delhi*, 2006, pp. 203–210.

[23]Chalot, F., Dinh, Q., Herbin, E., Martin, L., Ravachol, M., and Roge, G., "Estimation of the impact of geometrical uncertainties on aerodynamic coefficients using CFD," AIAA Paper, 2068-2008, Apr. 2008.

[24]Rumpfkeil, M. P. and Mavriplis, D. J., "Efficient Hessian Calculations using Automatic Differentiation and the Adjoint Method," AIAA Paper, 2010-1268, January, 2010.

[25]Rumpfkeil, M. P. and Mavriplis, D. J., "Efficient Hessian Calculations using Automatic Differentiation and the Adjoint Method with Applications," *AIAA Journal*, Vol. 48, No. 10, 2010, pp. 2406–2417.

[26]Rumpfkeil, M. P., Yamazaki, W., and Mavriplis, D. J., "Uncertainty Analysis Utilizing Gradient and Hessian Information," Sixth International Conference on Computational Fluid Dynamics, ICCFD6, St. Petersburg, Russia, July 12-16, 2010.

[27]Yamazaki, W., Rumpfkeil, M. P., and Mavriplis, D. J., "Design Optimization Utilizing Gradient/Hessian Enhanced Surrogate Model," AIAA Paper, 2010-4363, June, 2010.

[28]Huang, D., Allen, T. T., Notz, W. I., and Zeng, N., "Global Optimization of Stochastic Black-Box Systems via Sequential Kriging Meta-Models," *Journal of Global Optimization*, Vol. 34(3), 2006.

[29]Kraaijpoel, D. A., *Seismic ray fields and ray field maps: theory and algorithms*, Ph.D. thesis, Universiteit Utrecht, The Netherlands, 2003.

[30]Iman, R. L., Helton, J. C., and Campbell, J. E., "An approach to sensitivity analysis of computer models, Part 1. Introduction, input variable selection and preliminary variable assessment," *Journal of Quality Technology*, Vol. 13(3), 1981, pp. 174–183.

[31]von Mises, R., *Mathematical Theory of Probability and Statistics*, New York: Academic Press, 1964.

[32]Helton, J. C., Johnson, J. D., Oberkampf, W. L., and Storlie, C. B., "A sampling-based computational strategy for the representation of epistemic uncertainty in model predictions with evidence theory," Tech. Rep. SAND2006-5557, Sandia National Laboratories, 2006.

[33]Diegert, K., Klenke, S., Novotny, G., Paulsen, R., Pilch, M., and Trucano, T., "Toward a More Rigorous Application of Margins and Uncertainties within the Nuclear Weapons Life Cycle A Sandia Perspective," Tech. Rep. SAND2007-6219, Sandia National Laboratories, 2007.

[34]Shafer, G., *A Mathematical Theory of Evidence*, NJ: Princeton University Press, 1976.

[35]Yager, R. R. and Liu, L., *Classic works of the Dempster-Shafer theory of belief functions. Studies in fuzziness and soft computing series, v. 219*, Berlin: Springer, 2008.

[36]Helton, J. C., Oberkampf, J. D. J. W. L., and Sallaberry, C. J., "Representation of Analysis Results Involving Aleatory and Epistemic Uncertainty," Tech. Rep. SAND 2008-4379, Sandia National Laboratories, 2008.

[37]Putko, M. M., Newmann, P. A., Taylor III, A. C., and Green, L. L., "Approach for uncertainty propagation and robust design in CFD using sensitivity derivatives," AIAA Paper, 2001-2528, June 2001.

[38]Mani, K. and Mavriplis, D. J., "Unsteady Discrete Adjoint Formulation for Two-Dimensional Flow Problems with Deforming Meshes," *AIAA Journal*, Vol. 46 No. 6, 2008, pp. 1351–1364.

[39]Mani, K. and Mavriplis, D. J., "Adjoint-Based Sensitivity Formulation for Fully Coupled Unsteady Aeroelasticity Problems," *AIAA Journal*, Vol. 47 No. 8, 2009, pp. 1902–1915.

[40]Hicks, R. and Henne, P., "Wing Design by Numerical Optimization," *Journal of Aircraft*, Vol. 15 No. 7, 1978, pp. 407 – 412.

[41]Batina, J. T., "Unsteady Euler Airfoil Solutions Using Unstructured Dynamic Meshes," *AIAA Journal*, Vol. 28, No. 8, 1990, pp. 1381 – 1388.