Adjoint Methods for Uncertainty Quantification

Dimitri Mavriplis *
Department of Mechanical Engineering
University of Wyoming
Laramie WY USA

September 14-17, 2015

Contents

*email:mavripl@uwyo.edu

1 Introduction

As mentioned in the first chapter, the ability to compute not only simulation outputs as a function of inputs, but also the derivatives or sensitivities of these outputs with respect to the inputs is becoming increasingly recognized as an important capability not only for enabling optimization and error estimation, but also for uncertainty quantification. As simulation capabilities advance, practitioners are demanding more than simple deterministic results from their simulations. Rather, these are being replaced with the prediction of mean values, accompanied with information on the probability distribution of these outputs about the mean as a function of variations in the simulation inputs or model parameters. For example, in gas turbine engines, where blade life is correlated with exposure to high temperatures, the prediction of a single deterministic blade surface temperature provides little information for assessing overall reliability compared to the ability to predict a probabilistic range of possible temperatures given the various uncertainties in the blade characteristics, the operating conditions, and the simulation model itself. The additional information provided by uncertainty quantification is especially important for situations in which real world data is difficult or impossible to obtain. A prime example of this situation is hypersonic flow. The simulation of hypersonic flow is characterized by high velocities, leading to strong shocks, high temperatures and pressures, the excitation of internal energy modes and often chemical reactions. To simulate these physical phenomena, a number of constitutive relations and empirical models are required, each of which contain a large number of often experimentally derived inputs. For engineering systems featuring hypersonic flow, such as atmospheric re-entry of spacecraft, acquiring experimental data is costly and, in some cases, impossible, leading to a heavy reliance on simulation for the design of these systems and for ensuring the proper performance. Because of this heavy reliance, uncertainty quantification is a vital tool for the simulation of hypersonic flow. Using uncertainty quantification, the quality of simulation results can be assessed and the reliability of the system can be characterized in a probabilistic sense.

The goal of uncertainty quantification is to determine the range and distribution of possible values of important simulation outputs or quantities of interest, given a probabilistic distribution of simulation input parameters. The uncertain input parameters can include variations in geometry due to manufacturing tolerances and/or wear and tear, variations in flow conditions (for example due to atmospheric changes for flight systems), and imprecise knowledge of model parameters, such as empirically measured chemically reacting real gas reaction rates for the case of hypersonic flow problems. While other types of errors are present in most simulations, such as model form error due to physical models (such as turbulence models) that fail to correctly represent all underlying physics, these require specialized techniques of their own, and in general we will only be concerned with uncertainties that can be expressed as variations in parametric inputs.

The variability of simulation inputs can come in two forms: aleatory or epistemic. The quantification of each form, as well as the case of mixed form, requires a unique set of methods. Aleatory uncertainties arise due to the inherent randomness of a variable and are characterized by a probability distribution [42]. For aleatory inputs, the goal of uncertainty quantification is to determine the distribution of an output quantity due to these input distributions. Depending on the application, this characterization may consist of constructing the full empirical distribution function or may be limited to calculating

statistics of the distribution that can be used within design, such as the average, variance or a specified quantile. Epistemic uncertainty arises from a lack of knowledge regarding the true value of a parameter. Because of this lack of knowledge, the parameter has no associated probability distribution function and is typically only specified using an interval. The goal of uncertainty quantification for epistemic uncertainties is to determine the output interval of a quantity due to specified input intervals.

In general, uncertainty quantification problems involve many input parameters that interact with each other in non-linear manners and one cannot study the effect of one parameter in isolation from variations in the others. Therefore, the most consistent approach for determining output statistics is to perform many simulations with different values of the input parameters for each simulation in order to build up a statistical representation of the simulation output through sampling of the uncertain input parameters. This is known as Monte Carlo simulation and various approaches for sampling or determining the values of the input parameters can be used such as Latin Hypercube sampling, stratified sampling, and Markov Chain Monte Carlo methods. Although we will use Monte Carlo methods extensively, we will not concern ourselves with the particular characteristics of different sampling strategies. Monte Carlo methods are not dependent on the number of input parameters although these methods converge relatively slowly with the number of samples ($O(\sqrt{N})$). The principal disadvantage with Monte Carlo methods is the large number of simulations required, often exceeding thousands or tens of thousands of simulations required to obtain good output statistics. Thus, for high fidelity simulations which may take hours or even days for a single simulation, Monte Carlo methods quickly become intractable.

A typical approach for reducing the expense of Monte Carlo sampling is the use of an inexpensive surrogate model. This surrogate approximates the relationship between the true functional value and the input parameters and is built based on a limited number of functional evaluations, each of which corresponds to a full simulation. Because the surrogate is inexpensive to evaluate, exhaustive sampling of this model can be performed to build the required statistics of the output. Surrogate models range in complexity from simple extrapolations [9, 10] to more sophisticated models, such as least-squares polynomials [2, 3, 41, 16], support vector regression [12], radial basis functions [4], and Kriging. In computational fluid dynamics (CFD), Kriging methods in particular have gained popularity [7, 21, 19, 45, 6, 31, 18, 37, 24, 23, 49]. For the quantification of uncertainty in hypersonic flows, surrogates based on polynomial chaos have been employed with success [2, 3]. Despite this success, these polynomial approaches have been limited to a small number of variables. One drawback of surrogate-based methods is the "curse of dimensionality", whereby the number of samples required for an accurate surrogate increases exponentially as the number of input parameters grows.

Adjoint methods can be used to reduce the cost of uncertainty quantification in various ways. For example, the sensitivities of the output functional with respect to all input parameters can be used to determine which parameters are least likely to influence the output values (lowest gradients or sensitivities) and these parameters can be discarded prior to construction of the surrogate model, thus reducing the dimensionality of the problem. This corresponds to a local sensitivity analysis, for which adjoint methods are ideally suited, since the sensitivity of the output with respect to all inputs can be computed for the cost of a single adjoint solution. However, by considering only local

information and by neglecting non-linear effects, local sensitivity analysis may lead to erroneous conclusions about which parameters can be discarded.

Adjoint methods can also be used to reduce the cost of surrogate model construction. In addition to requiring the surrogate model to match computed functional values at specific locations in parameter space, we may also require the model to match the gradients of the functional with respect to the input parameters at those same locations. In doing so, we may leverage the power of the adjoint formulation in obtaining the entire gradient vector with respect to all input parameters for the same cost as the functional evaluation. Thus for an N-dimensional problem (i.e. N input parameters) we can supply N+1 pieces of information to the surrogate model for a cost of one functional and one adjoint evaluation. In addition to first-order derivatives, we may also seek to match the second-order derivatives of the surrogate and simulation at these locations. For this, the adjoint can be combined with N forward sensitivity analysis problems (one per parameter) to obtain the full Hessian of the functional with respect to the inputs. Thus we can supply an additional $\frac{N(N+1)}{2}$ pieces of information (individual elements of the symmetric Hessian matrix) at the cost of N forward sensitivity solutions. Of course, the success of higher order functional interpolation is highly dependent on the smoothness of the functional itself, and the effectiveness in improving the accuracy of the surrogate model must be weighed against the cost of evaluating these additional derivatives.

Epistemic uncertainty may be quantified via sampling based approaches or via optimization. Typically, Latin hypercube sampling [17] is used for epistemic uncertainties, although other methods such as approaches based on random sampling and Dempster-Shafer evidence theory can be used [15, 48, 22]. For Latin hypercube sampling in particular, the required number of samples grows quickly as the dimension of the problem increases, making the quantification of epistemic uncertainties for large-dimension problems difficult [42]. As was the case with aleatory uncertainty, one possible solution is to replace sampling with a surrogate model; however, this approach will again eventually encounter the curse of dimensionality as the input dimension increases. The other main approach for epistemic uncertainty quantification is to pose the problem as a bound-constrained optimization problem, defined as: given input parameters within specified ranges, determine the maximum and minimum values of an output functional. Although this approach entails solving a complicated global optimization problem with the possibility of multiple extrema [38], the number of functional evaluations to solve the optimization problem scales more readily to high-dimensional problems if a gradient-based optimizer is employed, particularly when an adjoint method is available to compute the gradients at low cost. For situations in which traditional gradient-based approaches, such as those built from Newton's method, are inadequate, efficient global optimization techniques based on Kriging surrogate models have been successfully demonstrated within the field of CFD [18]. For these methods, a Kriging surrogate is used to represent the design space, and traditional global optimization techniques, such as genetic algorithms, are applied to this surrogate. As is the case for aleatory uncertainty, these Kriging methods can be enhanced with derivative and Hessian values to improve the performance of the model in higher dimensions [24, 49].

In the following sections we illustrate the application of adjoint methods to sensitivity analysis and uncertainty quantification problems. We first provide an overview of the forward sensitivity and adjoint formulation for parameter sensitivities and extend this

to the calculation of the full Hessian for second derivatives using a combination of the adjoint and forward sensitivity solutions. We then illustrate the application of adjoint derived first and second-order derivatives for uncertainty quantification problems using the method of moments, simple extrapolation, and the construction of polynomial regression surrogate models as well as Kriging models. In a second part we demonstrate the systematic application of these techniques to a hypersonic reacting gas simulation with numerous uncertain model parameters.

2 Calculation of First and Second Derivative Information

Although the derivation of the forward and adjoint sensitivity equations has previously been described, the derivation is repeated here for completeness. This is followed by the derivation of the formulation for the Hessian matrix of second derivatives.

2.1 Forward and Adjoint Methods

We consider a simulation which takes a set of N input parameters $D_j, j = 1, 2, ...N$ and produces a set of outputs or quantities of interest L based on a computed state $\mathbf{U}$, where $\mathbf{U}$ is a field vector, for example a flow field solution for a computational fluid dynamics (CFD) problem. Thus, the functional dependence can be written as

$$L = L(\mathbf{U}(D_j), D_j) \tag{1}$$

where $\mathbf{U}$ depends implicitly on the inputs D, since it is obtained as the solution of the residual equation

$$\mathbf{R}(\mathbf{U}(D_j), D_j) = 0 \tag{2}$$

In order to compute the derivate of L with respect to D_j, we invoke the chain rule as

$$\frac{dL}{dD_j} = \frac{\partial L}{\partial D_j} + \frac{\partial L}{\partial \mathbf{U}} \frac{\partial \mathbf{U}}{\partial D_j} \tag{3}$$

$\frac{\partial \mathbf{U}}{\partial D_j}$ represents the flow field sensitivities which can be obtained by differentiating the residual equation or constraint $\mathbf{R}(\mathbf{U}(D_j), D_j) = 0$ as

$$\left[\frac{\partial \mathbf{R}}{\partial \mathbf{U}}\right] \frac{\partial \mathbf{U}}{\partial D_j} = -\frac{\partial \mathbf{R}}{\partial D_j} \tag{4}$$

Substituting this into equation (3) we obtain:

$$\frac{dL}{dD_j} = \frac{\partial L}{\partial D_j} - \frac{\partial L}{\partial \mathbf{U}} \left[\frac{\partial \mathbf{R}}{\partial \mathbf{U}}\right]^{-1} \frac{\partial \mathbf{R}}{\partial D_j} \tag{5}$$

The forward linearization approach corresponds to evaluating the last two terms on the right hand side first, while the adjoint approach consists of evaluating the first two terms

on the right-hand side first. Therefore, the forward linearization can be written as:

$$\left[\frac{\partial \mathbf{R}}{\partial \mathbf{U}}\right] \frac{\partial \mathbf{U}}{\partial D_j} = -\frac{\partial \mathbf{R}}{\partial D_j} \tag{6}$$

$$\frac{dL}{dD_j} = \frac{\partial L}{\partial D_j} + \frac{\partial L}{\partial \mathbf{U}} \frac{\partial \mathbf{U}}{\partial D_j} \tag{7}$$

whereas the adjoint procedure corresponds to:

$$\left[\frac{\partial \mathbf{R}}{\partial \mathbf{U}}\right]^T \mathbf{\Lambda} = -\frac{\partial L}{\partial \mathbf{U}}^T \tag{8}$$

$$\frac{dL}{dD_j} = \frac{\partial L}{\partial D_j} + \mathbf{\Lambda}^T \frac{\partial \mathbf{R}}{\partial D_j} \tag{9}$$

In both of these formulations, most of the computational expense occurs in the solution of the first equation, which requires the inversion of a matrix of the size of the exact flow Jacobian. In the first case, a new solution of the flow sensitivity problem must be repeated for each input parameter D_j, whereas in the second case, a new adjoint solution is required for each objective L. Thus the forward sensitivity approach is best suited for cases with a single input and multiple objectives, while the adjoint approach is best suited for cases with a single objective and multiple input parameters.

2.2 Hessian Computation

Consider now the computation of the Hessian matrix of second derivatives of the form $\frac{d^2L}{dD_j dD_k}$, where $j = 1, 2, ..N$ and $k = 1, 2, ...N$. The Hessian is a $N \times N$ square symmetric matrix. For a given objective we will show that it is possible to compute all entries in the matrix using a single adjoint solution and N forward sensitivity solutions [10, 43]. Differentiating equation (3) again we obtain an expression for the second derivative of the objective as:

$$\frac{d^2L}{dD_j dD_k} = \mathfrak{D}_{ik}L + \frac{\partial L}{\partial \mathbf{U}} \frac{\partial^2 \mathbf{U}}{\partial D_j \partial D_k} \tag{10}$$

where $\mathfrak{D}_{ik}L$ is given by

$$\mathfrak{D}_{ik}L = \frac{\partial^2 L}{\partial D_j \partial D_k} + \frac{\partial^2 L}{\partial \mathbf{U} \partial D_k} \frac{\partial \mathbf{U}}{\partial D_j} + \frac{\partial^2 L}{\partial \mathbf{U} \partial D_j} \frac{\partial \mathbf{U}}{\partial D_k} + \frac{\partial^2 L}{\partial \mathbf{U}^2} \frac{\partial \mathbf{U}}{\partial D_j} \frac{\partial \mathbf{U}}{\partial D_j} \tag{11}$$

The expression for $\frac{\partial^2 L}{\partial D_j \partial D_k}$ in equation (10) is broken down in this manner because the $\mathfrak{D}_{ik}$ is easily computable provided the N flow sensitivity vectors $\frac{\partial \mathbf{U}}{\partial D_j}$ $(j = 1, 2, ..., N)$ are available, for example by solving equation (6) N times. However, the second derivative term $\frac{\partial^2 \mathbf{U}}{\partial D_j \partial D_k}$ in equation (10) must be computed by solving the equation obtained by differentiating the residual equation twice, i.e. by differentiating equation (4) or (6) a second time to obtain:

$$\left[\frac{\partial \mathbf{R}}{\partial \mathbf{U}}\right] \frac{\partial^2 \mathbf{U}}{\partial D_j \partial D_k} = -\mathfrak{D}_{ik}\mathbf{R} \tag{12}$$

where $\mathfrak{D}_{ik}\mathbf{R}$ is given by analogy with equation (11) as:

$$\mathfrak{D}_{ik}\mathbf{R} = \frac{\partial^2 \mathbf{R}}{\partial D_j \partial D_k} + \frac{\partial^2 \mathbf{R}}{\partial \mathbf{U} \partial D_k}\frac{\partial \mathbf{U}}{\partial D_j} + \frac{\partial^2 \mathbf{R}}{\partial \mathbf{U} \partial D_j}\frac{\partial \mathbf{U}}{\partial D_k} + \frac{\partial^2 \mathbf{R}}{\partial \mathbf{U}^2}\frac{\partial \mathbf{U}}{\partial D_j}\frac{\partial \mathbf{U}}{\partial D_j} \tag{13}$$

Therefore, solving for all $\frac{\partial^2 \mathbf{U}}{\partial D_j \partial D_k}$ requires $N(N+1)/2$ solutions of equation (12), for all possible values of j and k. However, if we substitute equation (12) into equation (10) and make use of our definition of the adjoint variable (e.g. equation (8), we obtain:

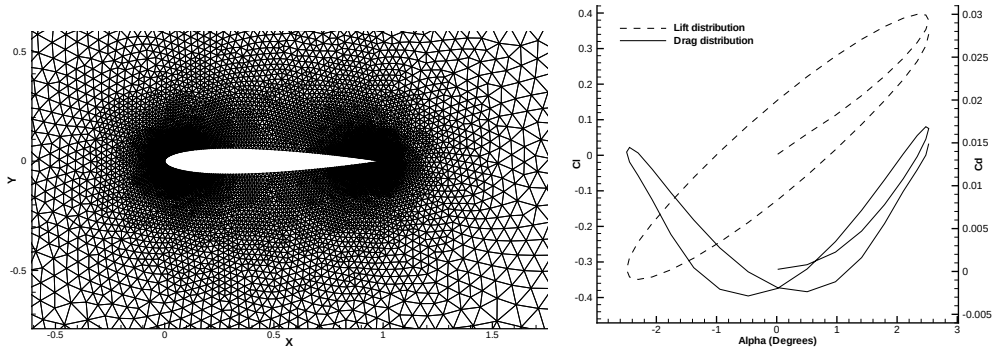
$$\frac{d^2 L}{dD_j dD_k} = \mathfrak{D}_{ik} L + \mathbf{\Lambda}^T \mathfrak{D}_{ik}\mathbf{R} \tag{14}$$

Since $\mathfrak{D}_{ik}L$ and $\mathfrak{D}_{ik}\mathbf{R}$ can be computed provided all flow sensitivity vectors $\frac{\partial \mathbf{U}}{\partial D_j}$ are known, all entries in the Hessian matrix can therefore be assembled with equation (14) at the cost of one adjoint solution and N forward sensitivity solutions.

2.3 Application to Uncertainty Quantification

In the following section, we illustrate the use of gradient and Hessian information for a simple uncertainty quantification problem. The problem consists of quantifying the uncertainty in the aerodynamic performance of a time-dependent pitching airfoil given uncertainties in the shape of the airfoil. Using both the first and second derivatives of the output objective with respect to the airfoil shape parameters, we construct an extrapolation model of the airfoil objective about the mean that can be used to approximate the output for relatively small changes in the input values. We also enable the propagation of uncertainty statistics from the inputs to the output using the method of moments. These techniques are compared against a brute force Monte Carlo approach that performs a large number of flow simulations with different input values in order to build up statistics for the output objective.

The test case consists of a sinusoidally pitching airfoil about its quarter-chord location solved using the Euler equations in arbitrary Lagrange Eulerian form as described in [30, 43]. The computational mesh contains about 20,000 triangular elements and is shown in Figure 1(left). The required deformation and movement of the mesh is performed via a linear tension spring analogy. The freestream Mach number is M = 0.755 with a mean angle of attack of 0.016 degrees with a periodic pitch amplitude of 2.51 degrees at a reduced frequency of 0.0814. A second-order backwards difference (BDF2) time-stepping scheme is used with 32 time steps per period. The baseline simulation consists of N = 40 time steps after a steady-state solution at the mean angle of attack. The resulting time-dependent lift and drag profiles are displayed in Figure 1(right). We are interested in the variations displayed in the time-averaged lift coefficient (the objective) with respect to uncertainties in the shape of the airfoil. Geometric uncertainty is modeled by introducing two airfoil shape design parameter, one Hicks-Henne bump function on the top surface of the airfoil and one on the lower surface. The two design variables are treated as random variables with normal distribution. The mean is set to zero (corresponding to the NACA 0012 airfoil) and the standard deviations are taken to be $\sigma_{D_1} = \sigma_{D_2} = 0.01$. Figure 2 shows the NACA 0012 airfoil and the airfoils resulting from perturbations of $\pm\sigma_{D_j}$.



1.1: 1.2:

Figure 1: (Left) Computational mesh for pitching airfoil problem; (right) Time history of lift and drag coefficients

The time-dependent adjoint and Hessian for the two-dimensional inviscid flow solver have been implemented and are used to obtain first and second derivatives of the objective with respect to the variations in the shape design parameters. Therefore, for given variations in the design parameters, the objective may be linearly extrapolated as:

$$L_{\text{Lin}}\Big(D, \mathbf{x}(D), \mathbf{U}(D)\Big) = L\Big(D_0, \mathbf{x}(D_0), \mathbf{U}(D_0)\Big) + \sum_{j=1}^{N} \left.\frac{dL}{dD_j}\right|_{D_0} \cdot \Delta D_j \tag{15}$$

using only first-order derivative information suppled by the adjoint solution, or quadratically as

$$L_{\text{Quad}}\Big(D, \mathbf{x}(D), \mathbf{U}(D)\Big) = L_{\text{Lin}}\Big(D, \mathbf{x}(D), \mathbf{U}(D)\Big) + \sum_{j=1}^{N}\sum_{k=1}^{N} \frac{1}{2} \left.\frac{\partial^2 L}{\partial D_j \partial D_k}\right|_{D_0} \cdot \Delta D_j \Delta D_k \tag{16}$$

using both first and second-order derivatives obtained from the adjoint and Hessian solutions, where D_0 represents the mean value of the input paramters D, and ΔD_j represents the deviation of the j^{th} parameter from its nominal value.

The simplest approach for obtaining statistics for the objective as a function of the input distributions is to perform a full Monte Carlo simulation where the inputs are sampled and used to generate a new objective value through a full CFD simulation for each set of samples. Alternately, we may use the same sampling of the inputs, but generate approximate objective values for each set of inputs using either linear of quadratic extrapolation, based on the above equations. This approach requires a single CFD simulation (about the mean values) and a single adjoint calculation in the linear case, and two additional forward sensitivity calculations (one for each input) in the quadratic case. In essence, the extrapolation models are used as simple surrogate models upon which an inexpensive Monte Carlo (IMC) simulation is performed.

Moment methods represent an alternate approach for obtaining output objective statistics [44, 39]. Moment methods are based on Taylor series expansions of the original nonlinear objective function $L(D)$ about the mean of the input D_0 given standard deviations σ_{D_j}. The resulting mean μ_L and standard deviation σ_L of the objective function are given

to first order (MM1) by

$$\mu_L^{(1)} = L(D_0) \tag{17}$$

$$\sigma_L^{(1)} = \sqrt{\sum_{j=1}^{M} \left(\frac{dL}{dD_j}\bigg|_{D_0} \sigma_{D_j} \right)^2}, \tag{18}$$

and to second order (MM2) by

$$\mu_L^{(2)} = \mu_L^{(1)} + \frac{1}{2} \sum_{j=1}^{M} \left(\frac{d^2L}{dD_j^2}\bigg|_{D_0} \sigma_{D_j}^2 \right) \tag{19}$$

$$\sigma_L^{(2)} = \sqrt{\sum_{j=1}^{M} \left(\frac{dL}{dD_j}\bigg|_{D_0} \sigma_{D_j} \right)^2 + \frac{1}{2} \sum_{j=1}^{M} \sum_{k=1}^{M} \left(\frac{d^2L}{dD_j dD_k}\bigg|_{D_0} \sigma_{D_j} \sigma_{D_k} \right)^2}. \tag{20}$$

In the first-order moment method the mean output is given as the output evaluated at the mean input values, whereas the shift in the output mean due to non-linear effects is captured in the second-order method. On the other hand, moment methods assume Gaussian distributions and do not give and information about the probability density function (PDF), and higher order moment methods require the computation of higher derivatives which becomes impractical in most cases.

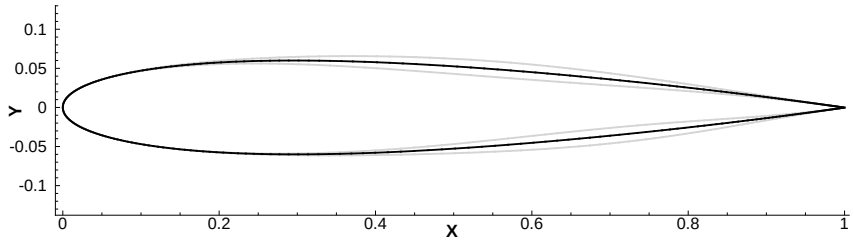


Figure 2: The NACA 0012 airfoil (in black) and airfoils resulting from perturbations of $\pm\sigma_{D_j}$ (in gray).

For the full non-linear Monte Carlo and inexpensive Monte Carlo methods (based on linear or quadratic extrapolation) we use stratified sampling with a sample size of $10,000$. One flow solve takes about 15 minutes on four 2GHz cores 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. Comparisons of the mean and standard deviation predictions of the objective function (time-averaged lift) using the various methods as well as approximate running times are displayed in Table 1.

The 99 per cent confidence interval for the mean calculated with the full nonlinear Monte Carlo simulation is $[5.52 \times 10^{-2}, 5.58 \times 10^{-2}]$. As can be seen first-order moment method (MM1) and linear extrapolation (Lin) yield very similar results as expected from the leading error. Also, second-order moment method (MM2) and quadratic extrapolation (Quad) give similar results for the same reason. Finally, as can be seen in Figure 3 the IMC methods capture the actual histograms and consequently PDFs of the time-averaged lift distribution quite well.

Table 1: Comparison of Mean and Standard deviation predictions.

	Mean	Standard deviation	Run time (minutes)
Nonlinear	5.55×10^{-2}	1.03×10^{-1}	150,000
MM1	5.81×10^{-2}	1.02×10^{-1}	30
MM2	5.39×10^{-2}	1.02×10^{-1}	60
Lin	5.82×10^{-2}	1.02×10^{-1}	30
Quad	5.39×10^{-2}	1.03×10^{-1}	60

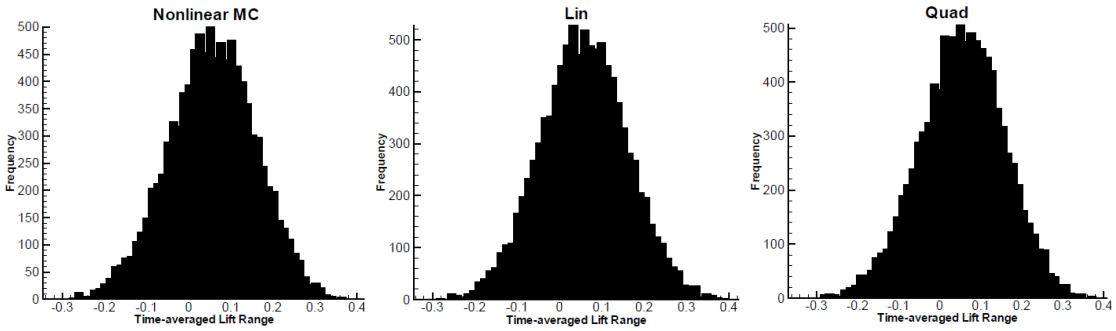


Figure 3: Histograms for time-averaged lift perturbations using various methods.

3 Gradient and Hessian Enhanced Surrogate Models

A surrogate model is a function that can be used to inexpensively approximate one or more outputs of a simulation over a range of parameter values based on a relatively small number of simulation results performed with specific input parameter values. The linear and quadratic extrapolation methods demonstrated in the previous section are simple examples of surrogate models. However, extrapolation methods only make use of local information and quickly become unreliable for non-linear functions away from the sampled region. In order to be more uniformly accurate and reliable, a set of samples that covers a broad range of the parameter space is required. The main drawback with surrogate models is that the number of samples and thus the cost associated with constructing an accurate surrogate typically grow exponentially as the number of design variables increases. This cost can be mitigated to some degree by the incorporation of gradient information into the surrogate model through the use of adjoint methods. Although each functional value used to build the surrogate model comes at the cost of one full CFD simulation, for the equivalent cost of one adjoint solution, we can obtain the entire vector of gradients at that location for use in the model. Thus while the functional value supplies one piece of information for training the model, the adjoint supplied N pieces of information for the same cost, where N represents the number of parameters or inputs. Similarly, a Hessian calculation can supply $\frac{N(N+1)}{2}$ pieces of information for the additional cost of N forward sensitivity solutions. On the other hand, the lower cost of this local information must be weighed against the effectiveness of obtaining more global information by sampling additional functionals at different locations in parameter space. At the same time, gradient and especially Hessian information will be less useful for globally approximating non-smooth functions, illustrating the problem dependent nature of surrogate model construction.

In the following section, we describe the construction of a polynomial regression model that includes gradient information, as well as the formulation of Kriging models that may include gradient and Hessian information and illustrate the performance of gradient and Hessian enhanced Kriging models on simple model problems.

3.1 Gradient Enhanced Polynomial Regression Methods

A polynomial regression model is based on the assumption that the output is approximated by a linear combination of polynomials given by equation:

$$y(D) = \sum_s \beta_s \Psi_s(D) \tag{21}$$

where $\Psi(D)$ represent a series of polynomials in D with degree less than p and β are a set of undetermined coefficients. The coefficients are determined by finding the set that best describe the simulation results at various values of D. Based on the results of N simulations, a system of equations can be solved to determine β.

$$\begin{bmatrix} \Psi_1(D_1) & \Psi_2(D_1) & \cdots & \Psi_s(D_1) \\ \Psi_1(D_2) & \Psi_2(D_2) & \cdots & \Psi_s(D_2) \\ \vdots & \ddots & \ddots & \vdots \\ \Psi_1(D_{N-1}) & \Psi_2(D_{N-1}) & \cdots & \Psi_s(D_{N-1}) \\ \Psi_1(D_N) & \Psi_2(D_N) & \cdots & \Psi_s(D_N) \end{bmatrix} \begin{bmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_s \end{bmatrix} = \begin{bmatrix} y(D_1) \\ y(D_2) \\ \vdots \\ y(D_{N-1}) \\ y(D_N) \end{bmatrix} \tag{22}$$

Here, the matrix on the left-hand side of the equation is defined as the collocation matrix H. Because N is typically greater than S (normally by a factor of two), this collocation matrix is inverted in a least-squared sense. This process is demonstrated below. Let the vector Y represent the results of N simulations using the input parameters D_i for $i = 1, 2, ..., N$. Additionally, define the collocation matrix H as the rectangular matrix (dimension $N \times S$) whose elements represent the basis functions evaluated at the input parameters D_i. The coefficients β are determined by solving a set of linear equations, represented below.

$$H\beta = Y \tag{23}$$

Because H is a rectangular matrix, the coefficients are determined in a least-squares sense. To solve the problem in this manner, both sides of equation (23) are multiplied by the transpose of the collocation matrix. The product of the collocation matrix with its transpose is denoted as the regression matrix, A.

$$H^T H\beta = A\beta = H^T Y \tag{24}$$

Because the regression matrix A is square, it can be inverted to determine the coefficients that best describe the simulation outputs Y. These coefficients are given as:

$$\beta = A^{-1} H^T Y \tag{25}$$

Gradient information can be incorporated into the regression model by differentiating equation (21).

$$\frac{\partial y(D)}{\partial D_k} = \sum_s \beta_s \frac{\partial \Psi_s(D)}{\partial D_k} \tag{26}$$

These derivative observations can be incorporated into the collocation matrix, providing additional equations without increasing the number of samples required for the regression.

$$\begin{bmatrix} \Psi_1(D_1) & \Psi_2(D_1) & \cdots & \Psi_s(D_1) \\ \frac{\partial \Psi_1(D_1)}{\partial D_1} & \frac{\partial \Psi_2(D_1)}{\partial D_1} & \cdots & \frac{\partial \Psi_s(D_1)}{\partial D_1} \\ \vdots & \ddots & \ddots & \vdots \\ \frac{\partial \Psi_1(D_1)}{\partial D_d} & \frac{\partial \Psi_2(D_1)}{\partial D_d} & \cdots & \frac{\partial \Psi_s(D_1)}{\partial D_d} \\ \vdots & \ddots & \ddots & \vdots \\ \Psi_1(D_N) & \Psi_2(D_N) & \cdots & \Psi_s(D_N) \\ \frac{\partial \Psi_1(D_N)}{\partial D_1} & \frac{\partial \Psi_2(D_N)}{\partial D_1} & \cdots & \frac{\partial \Psi_s(D_N)}{\partial D_1} \\ \vdots & \ddots & \ddots & \vdots \\ \frac{\partial \Psi_1(D_N)}{\partial D_d} & \frac{\partial \Psi_2(D_N)}{\partial D_d} & \cdots & \frac{\partial \Psi_s(D_N)}{\partial D_d} \end{bmatrix} \begin{bmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_s \end{bmatrix} = \begin{bmatrix} y(D_1) \\ \frac{\partial y(D_1)}{\partial D_1} \\ \vdots \\ \frac{\partial y(D_1)}{\partial D_d} \\ \vdots \\ y(D_N) \\ \frac{\partial y(D_N)}{\partial D_1} \\ \vdots \\ \frac{\partial y(D_N)}{\partial D_d} \end{bmatrix} \tag{27}$$

With the collocation defined as in equation (27), the regression coefficients are determined using the least-squares procedure given in equations (23) through (25).

The choice of basis functions Ψ_D used for this regression is general. In previous work, Hermite polynomials have been used within gradient-enhanced regression models [41] and multidimensional basis have been constructed by means of a tensor product of one dimensional polynomials [16].

$$\Psi_k(D) = \prod_{i=1}^{d} H_{m_i^j}(D_i) \tag{28}$$

Here, m_i^j is a multi-index dictating the order of the polynomial H. The basis of a given order p will contain products of polynomials less than or equal to degree p. In this way, the basis is hierarchical in that the basis of degree p contains all the terms of basis of degree $p-1$. An example basis of degree 2 in 3 dimensions is given below.

$$\Psi(D) = \begin{bmatrix} 1, \\ H_1(D_1), H_1(D_2), H_1(D_3) \\ H_1(D_1) \times H_1(D_2), H_1(D_1) \times H_1(D_3), H_1(D_2) \times H_1(D_3), \\ H_2(D_1), H_2(D_2), H_2(D_3) \end{bmatrix} \tag{29}$$

Here, H_k is the hermite polynomial of degree k. As the above set demonstrates, the number of terms in the regression grows rapidly as order and/or dimension is increased.

The cost of constructing a polynomial regression model is dictated by the number of terms required in the regression, as the number of simulation results required for the surrogate must be greater than or equal to the number of terms in the regression. The number of terms in the regression is given by the following relation:

$$S = \frac{(d+p)!}{d!p!} \tag{30}$$

where d is the dimension of the space and p is the highest polynomial order. When gradient information is incorporated into the training of the surrogate, each simulation

result (analysis followed by adjoint) provides $d+1$ pieces of information. Hence, the number of simulation results required for a surrogate using gradient information is given by:

$$N \geq \left\lceil \frac{(d+p)!}{d!p!(d+1)} \right\rceil \tag{31}$$

For example, if we impose a requirement of worst-case linear growth in the number of required simulations, regression orders of $p = 1$ and $p = 2$ may be considered. With this constraint, the required number of simulation results is given as:

$$N \geq \begin{cases} 1 & \text{for } p = 1, \\ \left\lceil \frac{(d+2)}{2} \right\rceil & \text{for } p = 2. \end{cases} \tag{32}$$

As these equations show, when gradient information is used, there is no constraint on the number of simulations required to construct a linear model. For a quadratic model, the expense associated with training the model increases only linearly. Equation (32) places a lower bound on the number of simulations required to construct the regression model. In practice, the regression model is typically over-determined and the coefficients are determined in a least-squares sense.

3.2 Gradient and Hessian Enhanced Kriging Methods

A Kriging model is a form of regression that is able to account for correlation between the data points, enabling more accurate functional representations. An overview of Kriging models is given here based on the descriptions in Reference [40]. The Kriging model is premised on the assumption that the output data y obey a Gaussian process, specified as:

$$y = N(m(x), K(x, x'; \theta)) \tag{33}$$

where $m(x)$ is the mean function, $K(x, x'; \theta)$ represents the covariance between data points, and θ represent parameters used to govern the covariance function, known as hyperparameters. The choice of mean function can vary widely based on the application of the Kriging model. Within machine learning applications where training data is abundant, a zero mean function is often used, known as simple Kriging. For the purposes of uncertainty quantification, a non-zero mean function may be required. This mean function can be explicitly defined or incorporated as part of the Kriging construction. For an explicitly defined mean function, the output is represented as the sum of the mean function with a zero mean Gaussian process [40].

$$y = m(x) + N(0, K(x, x'; \theta)) \tag{34}$$

For an explicit mean function, the zero mean Kriging model is built based on the residual between the measured output y and the explicitly defined mean function $m(x)$.

To construct the Kriging model, it is first "trained" using a number of simulation results, represented as Y, evaluated at a set of input parameters, $\vec{X}$. Using these results, the parameters in the Gaussian process are fitted and predictions away from these training points can be made. The prediction from the Kriging model is itself a Gaussian process with an associated mean value and variance. Using the covariance between points in the

domain, model predictions throughout the domain are determined by sampling from the conditional distribution $y_*|\vec{X}, Y$ where $\vec{X}, Y$ are the input and output training data. The posterior mean predictions for an explicit mean are given by the formula [40]:

$$y(\vec{x}_*)|\vec{X}, Y, m(x) = m(x) + k_*^T K^{-1}(Y - m(x)) \tag{35}$$

where k_*^T represents the covariance between the test point, $\vec{x}_*$, and the training points $\vec{X}$ (a row vector of length N). The term K is the covariance between the training data, represented by a matrix of dimension $N \times N$.

In addition to an explicitly defined mean function, a non-zero mean function can be incorporated into the construction of the Gaussian process. This non-zero mean function usually takes the form of a polynomial regression and the coefficients within the regression are informed by the correlation within the Kriging model. When a polynomial mean function is used, the Kriging model is referred to as Universal Kriging. The special case of a zeroth order regression, meaning a constant mean function, is referred to as ordinary Kriging. For a universal Kriging model, the functional form becomes [40]:

$$y(\vec{x}) = N(h(\vec{x})\beta, K(\vec{x}, \vec{x}; \theta)) \tag{36}$$

where $h(\vec{x})$ is a column vector containing the basis functions of the regression evaluated at the point $\vec{x}$ and β are the regression parameters. Using a regression-based mean function, predictions can be made based on this model using the formula:

$$y(\vec{x}_*)|\vec{X}, Y = h(\vec{x}_*)\beta + k_*^T K^{-1}(Y - H\beta) \tag{37}$$

where H is the collocation matrix of the regression, Y is the vector of training function values, K is the covariance matrix between the training points and k_* is the vector of covariances between the training points and the test point (x_*). Since the regression is built from a limited number of training points, it is prudent to assume that the regression parameters belong to a distribution of parameters. In the limit of zero knowledge of this distribution (vague prior assumption), the optimal regression parameters are given by [40]:

$$\hat{\beta} = (H^T K^{-1} H)^{-1} H^T K^{-1} Y = A^{-1} H^T K^{-1} Y \tag{38}$$

where A is the regression matrix defined as $H^T K^{-1} H$. Unlike the regression matrix used in polynomial regression, the regression matrix for the Kriging model includes the correlation between the data points, K. The case of polynomial regression is found by assuming no correlation between the data points, represented as K reducing to the identity matrix. Using this definition of regression parameters, mean predictions can be made using the following single formula:

$$y(\vec{x}_*)|\vec{X}, Y = k_*^T K^{-1} Y + (h(\vec{x}_*) - k_*^T K^{-1} H)\hat{\beta} = k_*^T K^{-1} Y + R(\vec{x}_*)\hat{\beta} \tag{39}$$

The elements of the covariance matrix represent the covariance between the function values. For Kriging, the covariance between function values is assumed to be a function of distance between the two data points. This functional form is known as a stationary covariance function [40].

$$K_{i,j} = cov(y_i, y_j) = k(|\vec{x}_i - \vec{x}_j|) + \sigma_n \delta_{i,j} \tag{40}$$

Here, $k(|\vec{x}_i - \vec{x}_j|)$ is the covariance between training points i and j and σ_n represents the noise in the training data. For machine learning applications, this noise is treated as a hyperparameter and fitted through the likelihood equation, which is described below. For the application of uncertainty quantification, this noise is a specified value and is set to ensure proper conditioning of the covariance matrix. The multidimensional covariance function is formed as the product of one dimensional covariance functions as this form produces better conditioned covariance matrices [13]. The covariance between two points (denoted as $\vec{x}$ and $\vec{x}'$ to avoid confusion) is given below.

$$k(\vec{x}, \vec{x}'; \theta) = \sigma^2 \prod_{i=1}^{d} k_i(x_i - x_i'; \theta_i) = \sigma^2 \prod_{i=1}^{d} k_i(r_i; \theta_i) \tag{41}$$

Here, σ represents the covariance magnitude and θ_i is the length scale for each dimension. A common one dimensional covariance function is the Matern function with the parameter $\nu = 3/2$, given below [40].

$$k_i(x_i - x_i') = \left(1 + \sqrt{3}\left|\frac{x_i - x_i'}{\theta_i}\right|\right) e^{-\sqrt{3}\left|\frac{x_i - x_i'}{\theta_i}\right|} \tag{42}$$

The parameters in the covariance function are determined via the likelihood equation. The likelihood equation gives the probability that the data with a given set of parameters satisfy the assumed Gaussian process. By maximizing this probability, the optimal parameters can be determined. The log-likelihood equation for a universal Kriging model is given by [40]:

$$\log(p(y|X;\theta)) = -\frac{1}{2}(Y - H^T\beta)^T K^{-1}(Y - H^T\beta) - \frac{1}{2}\log|K| - \frac{N}{2}\log 2\pi \tag{43}$$

where N is the number of training points. Based on this likelihood equation, the optimal covariance magnitude can be found analytically. By differentiating the likelihood equation with respect to the covariance magnitude, the optimal magnitude can be determined. Denoting the covariance matrix with a magnitude of 1 as $\hat{K}$, the optimal magnitude is given as [13]:

$$\sigma(\theta)^2 = \frac{(Y - H^T\beta(\theta))\hat{K}(\theta)^{-1}(Y - H^T\beta(\theta))}{N} \tag{44}$$

Using this explicit equation for the magnitude, the likelihood equation can be rewritten as:

$$\log(p(y|X;\theta)) = -n\log\sigma(\theta)^2 - \log(|\hat{K}(\theta)|) - \frac{N}{2}\log 2\pi \tag{45}$$

The only parameters without analytic expressions are the length scales θ. These length scales are determined via numerical optimization. For this work, a pattern search [33] is used to determine the optimal parameters as a global deterministic optimization technique is preferable to non-deterministic methods, such as genetic algorithms, as deterministic methods allow for repeatability of results. This optimization represents the most expensive

part of constructing the Kriging model, as the covariance matrix must be constructed and inverted for each function evaluation in the optimization. Additionally, the dimension of the optimization problem corresponds to the dimension of the space approximated by the model.

Like the polynomial regression presented in the previous section, the Kriging model is also susceptible to the "curse" of dimensionality. Although the Kriging model does not have an explicit relationship for the number of training points as the dimension of the space expands, the amount of data required for an accurate model typically increases exponentially fast as the number of parameters increases. This fact is because an accurate surrogate requires training points that fill the design space. To reduce the cost of training the model as the dimension expands, gradient information can be incorporated into the Kriging model. This can be done in an indirect approach, or direct approach. In the indirect approach, new functional values are obtained by extrapolating the current functional value using the available gradient, and these are supplied as additional training points to the unmodified Kriging formulation. Although this allows the inclusion of gradient information without modifcations to the Kriging formulation, the indirect approach can be sensitive to the magnitude of the extrapolation distance used, with large distances resulting in loss of accuracy, and small distances resulting in poor matrix conditioning. In the direct approach for including gradient information in the Kriging model, the covariance matrix must first be extended to include the covariance between derivative values. When these correlations are included, the covariance becomes a block matrix with the following components.

$$\underline{K} = \begin{bmatrix} cov(Y,Y) & cov(Y,\nabla Y) \\ cov(\nabla Y,Y) & cov(\nabla Y,\nabla Y) \end{bmatrix} \tag{46}$$

The covariance between function and gradient components is found by differentiating the covariance function [46]:

$$cov(\frac{\partial y}{\partial x_k}, y') = \frac{\partial}{\partial x_k} k(\vec{x},\vec{x}') \tag{47}$$

Differentiating once more (now w.r.t to the second argument of the covariance function) gives the covariance between gradient components:

$$cov(\frac{\partial y}{\partial x_k}, \frac{\partial y'}{\partial x_l'}) = \frac{\partial^2}{\partial x_k \partial x_l'} k(\vec{x},\vec{x}') \tag{48}$$

The gradient vector and Hessian matrix resulting from equations (47) and (48) can then be arranged into the matrices $cov(\nabla Y, Y)$ and $cov(\nabla Y, \nabla Y)$ respectively.

In addition to extending the covariance matrix, the training data must be redefined as:

$$Y = \begin{bmatrix} Y \\ \nabla Y \end{bmatrix} \tag{49}$$

and the collocation matrix must be extended to include the derivative of the basis function evaluated at the training points. This extension is performed by differentiation of

the basis functions, as was demonstrated in the previous section. Similarly, by extending the covariance matrix to include the covariance between second derivatives, Hessian information may be included in the Kriging model, as shown in Reference [49].

3.2.1 Kriging Model Examples

The performance of surrogate models can be evaluated by using them to fit analytic functions. An impressive example of the potential of gradient and Hessian enhanced Kriging models can be demonstrated through the Rastrigin function. The Rastrigin function is a non-convex multimodal function that is often used as a performance test for optimization problems and is defined as:

$$Rastrigin_{(x)} = 10m + \sum_{i=1}^{m} \left[x_i^2 - 10cos(2\pi x_i)\right] \qquad (-5.12 \leq x_i \leq 5.12) \tag{50}$$

where m denotes the dimension (i.e. $m = 2$ in this case). The shape of the 2D Rastrigin function is illustrated in Figure 4, showing the multimodal form of this function. We construct a Kriging model of this function using 80 sample points generated by latin hypercube sampling (LHS) which may include gradient and Hessian components evaluated analytically at these sample points. In Figure 5, three different surrogate models con-

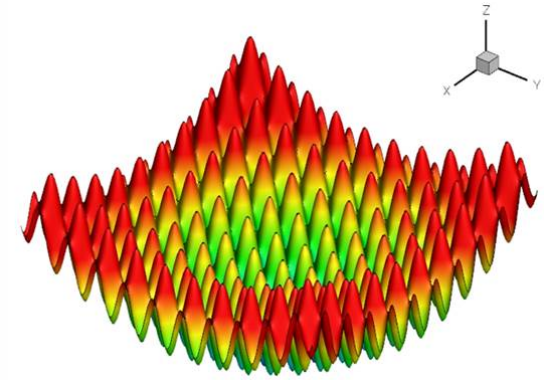


Figure 4: Visualization of 2D Rastrigin function

structed by function, function/gradient and function/gradient/Hessian information are compared. The gradient and gradient/Hessian-enhanced Kriging models are constructed by the direct Kriging approach. Although the function values show agreement in the vicinity of the 80 sample locations, the function-based conventional Kriging model does not capture the global trend of the Rastrigin function. Although the gradient-enhanced Kriging model displays better accuracy, only the gradient/Hessian-enhanced Kriging model reproduces the global trends accurately, which is rather remarkable given that only 80 sample locations are used in this case.

For a more quantitative investigation with higher dimensionality, we use the same Kriging model to fit the 5D Rosenbrock function by using various sets of sample points (from 10 to 600) generated by LHS. The Rosenbrock function is defined as:

$$Rosenbrock_{(x)} = \sum_{i=1}^{m-1} \left[(1 - x_i)^2 + 100(x_{i+1} - x_i^2)^2\right] \qquad (-2 \leq x_i \leq 3) \tag{51}$$

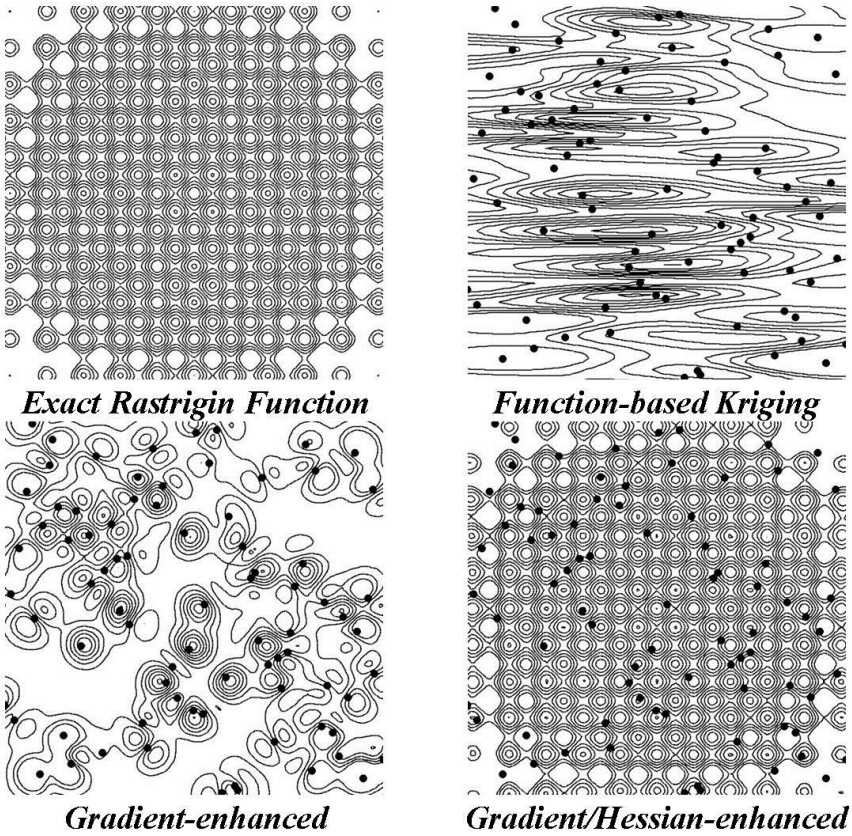


Figure 5: Kriging models used to fit Rastrigin function with 80 sample points

where m denotes the dimension of the function (i.e. $m = 5$ in this case). Once the surrogate model is constructed, the accuracy of the model is evaluated by using the root mean squared error (RMSE) between the exact function and approximate function values given by the surrogate model. The RMSE is given as follows:

$$RMSE = \frac{1}{M}\sqrt{\sum_{j=1}^{M}(\hat{y}_{(x_j)} - y_{(x_j)}^{exact})^2} \tag{52}$$

where the coordinates $\mathbf{x_j}$, $j = 1, ...M$ define an equally spaced Cartesian mesh which covers the entire design space. In Figure 6 the RMSE values are compared between the direct/indirect Kriging approaches. For the gradient/Hessian-enhanced indirect Kriging approach, three runs are executed with different sets of the extrapolation distance Δx and number of extrapolation points m_{add}. For the gradient/Hessian-enhanced direct Kriging approach, two runs are executed: one where the full Hessian matrix is used at each point, and one where only the diagonal elements of the Hessian are used. The direct approaches do not have any parameters to be specified, and in general the performance of the direct approaches is much better than the indirect approaches. This superiority is due to the exact enforcement of derivative information and the better conditioning of the correlation matrix. In Figure 6(b), the x-axis is modified to represent the number of pieces of information and the approximated computational time factor for a detailed

comparison. The number of pieces of information is the sum of the number of function values, gradient as well as Hessian net components among all sample points, which is generally equal to the size of the correlation matrix. The approximated computational time factor is defined as the sum of the local sample points, with a weight of 1 for function only, 2 for function/gradient, and 3 for function/gradient/Hessian sample points. This definition assumes that the computational times of gradient/Hessian evaluations are

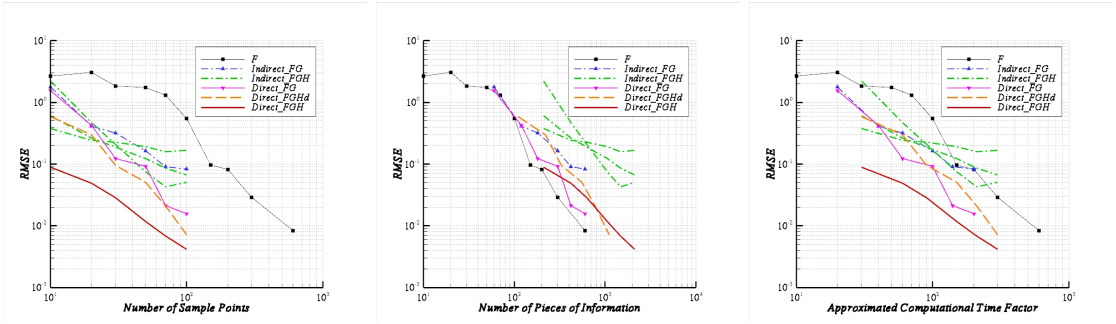


Figure 6: Comparison of RMS error of various Kriging models for the 5D Rosenbrock function as versus (a) number of sample points (b) pieces of information supplied to surrogate models, (c) approximate computational time

respectively comparable to one function evaluation. Although this is the case when using an adjoint approach for obtaining the gradient information, the Hessian construction scales with the number of design variables or input parameters. However, the current implementation parallelizes this aspect of the Hessian resulting in an overall wall clock timing close to that assigned above. Of course, this represents a best-case scenario for the Hessian enhanced model. The original function-based Kriging shows the best performance when compared by the number of pieces of information. This result implies that the scattering of only function information for the whole design space is better than the concentration of function/gradient/Hessian information at a limited number of sample points to obtain a more accurate fitting of the whole design space. When compared by the approximated computational time factor, the gradient/Hessian-enhanced direct Kriging approaches show better performance than the original Kriging approach. These results demonstrate that, although including additional local information is not as effective as corresponding increases in the amount of supplied global information, the principal benefit comes from the reduced cost afforded by the adjoint for evaluating the local gradient (and to a lesser degree Hessian) information.

The developed Kriging approaches are next used to fit an airfoil two-dimensional aerodynamic database. Moving now to a realistic aerodynamic problem, we demonstrate the use of a gradient-enhanced Kriging model for fitting a two-dimensional airfoil aerodynamic database. We consider the steady inviscid flow around a NACA0012 airfoil. The governing Euler equations of the flow problem are discretized by a finite-volume approach and are solved with second-order spatial accuracy. The computational mesh is the same as shown previously in Figure 1. In this study, two parameters are considered; Mach number and angle of attack. Their ranges are specified as $0.5 \leq Mach \leq 1.5$ and $0^o \leq \alpha \leq 5^o$. The exact surface or validation data is obtained through a series of 21x21=441 flow computations covering the entire parameter space at equally spaced increments. The exact

hypersurfaces of lift and drag coefficients are shown in Figures 7(a) and 8(a) showing rapid variations in the transonic regions. The Kriging models are constructed by utilizing the function and adjoint gradient values at ten sample points, and the resulting hypersurfaces for the function alone and function-gradient Kriging models are depicted in Figures 7 and 8. The figures show how the addition of gradient information vastly improves the approximation of the hypersurfaces. However, in this case, the Hessian information could not be used to improve the Kriging model. Upon closer examination, the exact hypersurfaces were found to contain small noise levels (attributed to shock alignment with cell faces) which resulted in highly varying second derivatives and thus precluded effective use of the Hessian information.

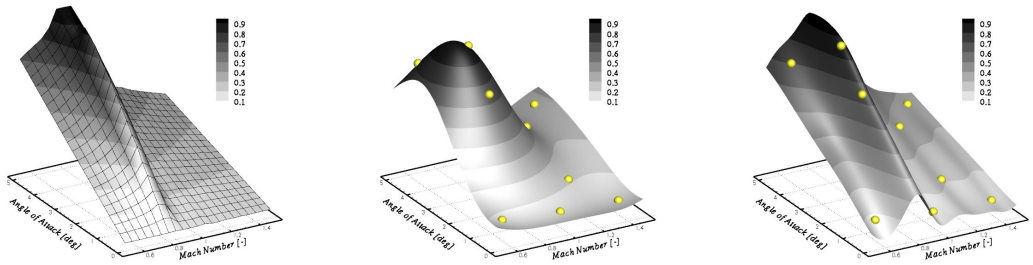


Figure 7: Hypersurface of lift coefficient; (a) Exact, (b) Kriging using function values, (c) Kriging using function and gradient values

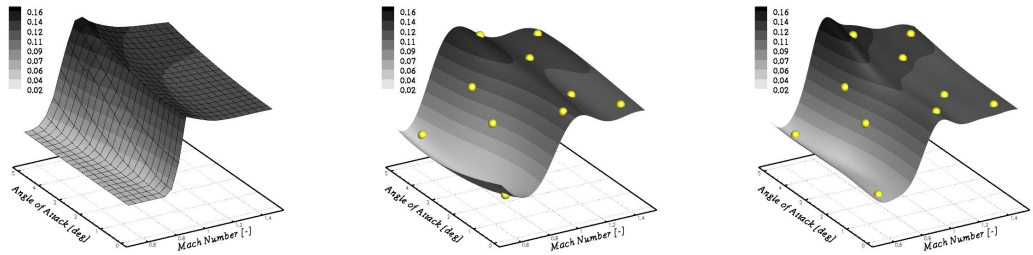


Figure 8: Hypersurface of drag coefficient; (a) Exact, (b) Kriging using function values, (c) Kriging using function and gradient values

3.2.2 Uncertainty Quantification using Kriging Model

We now compare the results of an uncertainty quantification analysis about the $Mach = 0.8$, $\alpha = 2.5^o$ point on the hypersurface using a full non linear Monte Carlo simulation and an inexpenive Monte Carlo (IMC) simulation on the Kriging surface. In this analysis, uncertainties are given to the two parameters ($Mach$ and α) based on a normal distribution. The mean of the normal distribution is fixed to the nominal values and the output statistics are examined as a function of the standard deviation of the input parameters (σ_D) which is varied from 0 to 0.1 The location of the mean values in parameter space is chosen to coincide with a region of rapid non-linear variation as noted in Figure 9. The

number of function calls for a Monte-Carlo simulation is fixed to 1000 in this study, and these are used to calculate the mean and variance of the aerodynamic functions. The function calls are directly solved by non-linear CFD calculations in the non-linear Monte Carlo (NLMC) simulation. This is carried out for seven different values of σ_D, which means 7000 CFD function calls are required for this analysis. In the IMC simulation, on the other hand, the function calls are solved on the Kriging surrogate model, which dramatically reduces the computational cost for the uncertainty analysis. In Figure 9, the locations of all 1000 sample points are visualized on the exact hypersurfaces for σ_D of 0.1. In Figure 10, the variations in the mean of lift/drag coefficients with respect to the variation of σ_D are compared for Kriging models constructed using different numbers of function evaluations versus the full NLMC case. As can be seen from the plots, using upwards of 60 function evaluations for the Kriging model results in close agreement for the mean values with those obtained from the NLMC simulation at a fraction of the cost, and the agreement increases as more function evaluations are used.

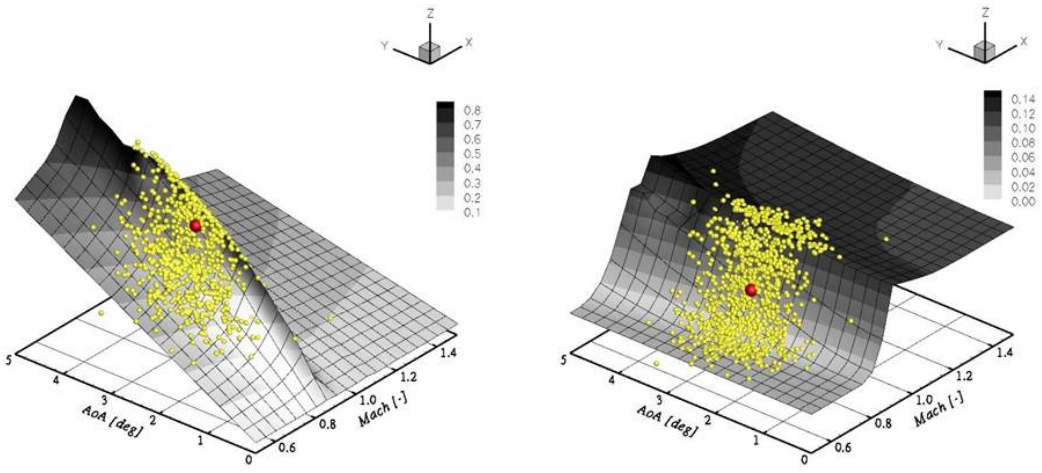


Figure 9: Location of nominal values (red) and 1,000 Monte Carlo sample points (yellow) for $\sigma_D = 0.1$ on lift (left) and drag (right) hypersurfaces

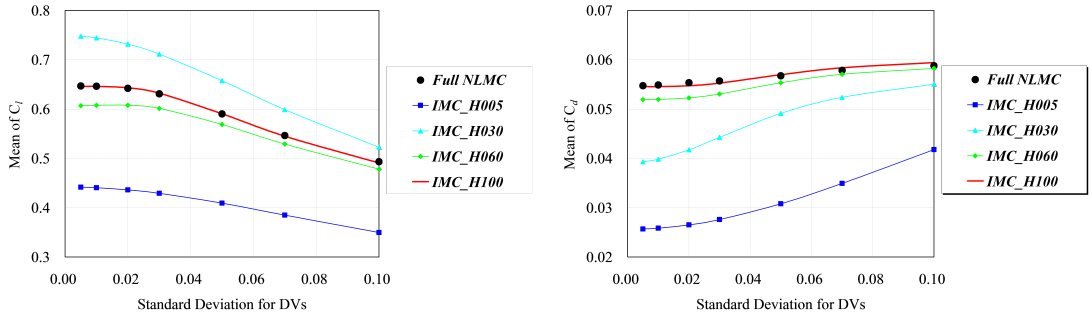


Figure 10: Mean value of lift and drag computed by full non-linear Monte Carlo versus inexpensive Monte Carlo on Kriging model constructed with increasing numbers of samples as a function of standard deviation of design variables

4 Gradient-based Epistemic Uncertainty Quantification

For epistemic uncertainty, the goal is to determine the interval of the output given intervals for the input parameters. As was the case with aleatory uncertainty, this output interval can be approximated by assuming a functional form of the output over the design space. Using a linear functional representation, the output interval width can be approximated using a moment-method type formula, given as:

$$y_o = f(x_o) \tag{53}$$

$$\Delta_y = \sum_{i=1}^{d} \left| \frac{\partial f}{\partial x_i} \Delta_{x_j} \right| \tag{54}$$

where x_o represents the design variables at the center of the input interval and y_o represents the center of the output interval. Because epistemic uncertainties are often specified using a plus/minus, x_o and y_o can be thought of as the unperturbed values. Although these values are analogous to mean values for the aleatory case, the use of mean implies an associated probability distribution function so the term unperturbed will be used throughout this section.

For cases in which the linear approximation is insufficient, the determination of the output interval can be recast as a bound-constrained optimization problem. The optimization problem is specified as: given a set of intervals for the input parameters, determine the minimum and maximum possible output values. This problem statement gives two optimization problems, given in as:

$$y_{min} = \min_{x \in I} f(x) \tag{55}$$

$$y_{max} = \max_{x \in I} f(x) \tag{56}$$

Here, the input intervals are given by the space I and the minimum and maximum values are determined based on inputs in this space. The process by which these minimum and maximum values are determined can vary widely in expense and complexity. The most straight-forward method for determining the extrema is by exhaustive sampling of the function. Typically, uniform Latin hypercube sampling is performed using a fixed number of samples, usually 3, for each variable. For this method, the required number of samples grows exponentially fast as the number of input parameters is increased, with the formula given as $N = 3^d$. For complex simulations, this exhaustive sampling is prohibitively expensive. As was the case with aleatory uncertainty, the simulation output can once again be replaced with a gradient-enhanced surrogate model and the Latin hypercube sampling can be performed on this surrogate. As was discussed previously, the expense associated with the construction of these surrogate models increases dramatically with dimension; hence, for large numbers of epistemic variables, the expense of a surrogate based approach may become prohibitively large.

In addition to exhaustive sampling approaches, more sophisticated optimization methods can be applied directly to the problems posed in equations (55) and (56). In particular, we will focus on gradient-based methods for optimization as these method can scale to large input dimension without an exponential increase in cost, by leveraging the adjoint capability. The optimization method used for this work is the limited-memory Broyden-Fletcher-Goldfarb-Shanno (L-BFGS) algorithm [50]. The L-BFGS algorithm is a quasi-Newton solver, meaning that only the function and gradient at each iteration is required and an approximate Hessian is constructed based on the convergence history of the solver. Using this approximate Hessian, Newton's method is used to drive the gradient to zero. For a BFGS method, the design space is approximated at iteration k as [33]:

$$m_k(p) = y_k + \nabla y_k^T p + \frac{1}{2} p^T B_k p \tag{57}$$

where the subscript k indicates current iteration values, B_k is the approximate Hessian and p represents the search direction. With this functional form, the direction of the minimum location is given by:

$$p_k = -B_k \nabla y_k \tag{58}$$

With the search direction determined, a line search is performed in this direction to determine a step size that gives a sufficient decrease in the functional. The approximate Hessian is updated at each iteration based on the requirement that the new approximate Hessian B_{k+1} must accurately predict the difference between the current gradient value ∇y_{k+1} and the previous gradient ∇y_k when used in a Taylor series approximation. The exact details of the construction of this approximate Hessian can be found in Reference [33]. For the limited-memory BFGS algorithm, this approximate Hessian is never explicitly constructed and the effect of the approximate Hessian is built up using only a limited number of previous gradient values. The limited storage requirements of this method make it particularly suited for high dimensional problems.

The main drawback of gradient-based optimization is that the methods are inherently localized in nature. For epistemic uncertainty propagation, the global maximum and minimum are required. For problems where gradient-based optimization is insufficient, a global optimization method must be employed. Popular global optimization methods include genetic algorithms, pattern searches, Monte-Carlo sampling and swarm-based algorithms. Global optimization methods typically require a large number of function evaluations. In light of this drawback, efficient global optimization is an area of active research. One such efficient global optimization technique often used in CFD is based on the previously outlined Kriging model [18, 24, 49]. Although no results will be shown using this method, the method will be outlined due to its close relation to the other techniques outlined in this chapter.

The process of Kriging-based global optimization is given by the following steps. This process is outlined only for minimization as the case of maximization can be found by reversing the sign on the function. First, a handful of function evaluations are performed and a Kriging surface is constructed based on these results. Second, based on the Kriging surface, promising candidate locations for the minimum value are determined. Finally, additional function evaluations are performed at these candidate locations and a new Kriging surface is constructed with these new function values added to the surrogate

training data. This process is repeated until a termination criteria is met, typically a relative convergence, gradient norm or maximum function evaluation criteria.

An effective way of determining promising candidate locations is based on the underlying Gaussian process representation of the Kriging surface. Because the Kriging surrogate is stochastic, function predictions have an associated variance. Hence, the surrogate predictions have an associated uncertainty representing the accuracy of the surrogate at a particular location. For optimization applications, this variance, along with the mean predictions, can be used to determine the location in the design space with the highest probability of containing a new minimum value. This criteria is known as the expected improvement (EI) method and is given by the following formula [18]:

$$EI(x) = \begin{cases} (y_{min} - y^*(x))\Phi\left(\frac{y_{min}-y^*(x)}{s(x)}\right) + s_x\phi\left(\frac{y_{min}-y^*(x)}{s(x)}\right) & \text{if } s(x) > 0, \\ 0 & \text{if } s(x) = 0 \end{cases} \tag{59}$$

where y_{min} is the minimum value from the previous optimization iteration, $y^*(x)$ is the mean prediction from the Kriging model, $s(x)$ is the Kriging standard deviation prediction, Φ is the normal cumulative distribution function and ϕ is the normal probability density function. As the formula shows, the EI criteria contains two terms. The first term is proportional to the mean Kriging prediction and grows when the mean prediction is less than the current minimum value. The second term is proportional to the variance of the Kriging prediction and grows for x values away from existing sample points. Hence, the EI criteria maintains a balance between a local search, where points around the minimum of the surrogate model are added, and a global search, where points are added in unexplored areas of the design space. The candidate locations are the points in the design space where the expected improvement function is maximized. Because this maximization is performed on the Kriging model, any global optimization technique can be used. In previous works, genetic algorithms have been used to determine the location with highest expected improvement [18]. Obviously, because this optimization technique is dependent on the construction of a Kriging model, gradient and Hessian information can easily be incorporated into the training of the surrogate, improving the accuracy of the Kriging surface [49].

5 Uncertainty Quantification of Hypersonic Flow Problem

In this section, we demonstrate the gradient enhanced uncertainty quantification techniques developed in the preceding sections for a realistic problem consisting of hypersonic flow over a circular cylinder. The physical model involves the solution of the two-dimensional laminar Navier-Stokes equations with a non-equilibrium real gas model. The real gas model is a five-species, two-temperature model for non-ionizing air [14]. Both the Dunn-Kang chemical kinetics model and the Park model [36] have been implemented. Despite the superior accuracy of the Park model, the Dunn-Kang model is used for uncertainty quantification because of the ease with which uncertain parameters within the model may be specified. The complete real gas model contains approximately 250 parameters, embedded within the constitutive models for the reaction rates, transport coefficients,

relaxation times, and caloric equations of state. However, a number of parameters are not included in the sensitivity analysis either because previous sensitivity and uncertainty analyses have indicated the parameters to be unimportant or reliable uncertainty specifications are unavailable [47]. Thus, a total of 66 model parameters are examined for this sensitivity analysis. These parameters cover freestream conditions, transport properties and the chemical kinetics model.

5.1 Problem Definition

In vector form, the Navier-Stokes equations for chemically reacting real gas mixtures are given by

$$\frac{\partial \mathbf{U}}{\partial t} + \nabla \cdot \vec{F}(\mathbf{U}) = \nabla \cdot \vec{F}_v(\mathbf{U}) + \mathbf{S}(\mathbf{U}), \tag{60}$$

where $\mathbf{U}$ are the conserved flow variables, $\vec{F}$ is the inviscid flux, $\vec{F}_v$ is the viscous flux, and S contains any source terms required for the physical model, such as reaction or energy coupling terms. For the real gas model, specific heats are calculated via fourth order polynomial curve fits covering various temperature ranges. The total enthalpy is calculated simply by integrating these curve fits and incorporating the proper heat of formation information [11]. The transport model is a collision integral model. For this model, viscosity, thermal conductivity, and diffusion coefficients are calculated based on linear interpolation of collision integrals between 2000 K and 4000 K[11, 34].

The governing equations described above are first discretized in space, using a cell-centered second-order accurate finite-volume approach on two-dimensional unstructured meshes and the solution is advanced in time to steady state using a fully implicit approach. In semi-discrete form, the equations have the following form:

$$\frac{\partial \mathbf{U}}{\partial t} + \mathbf{R}(\mathbf{U}) = 0 \tag{61}$$

The residual within each cell is given by the sum of the normal inviscid and viscous fluxes over all faces plus a cell-centered contribution due to source terms. The inviscid flux is calculated by using gradient reconstruction of primitive variables, and gradients are calculated using Green-Gauss contour integration over the cell. The limiter used in this code is a combination of a pressure switch and the smooth Van Albada limiter[14, 34, 8]. The AUSM+UP flux function is used because of the ease with which it can be extended to additional equations and its applicability across a wide range of Mach numbers. In order to extend this flux function to the real gas model, a frozen speed of sound is used [8, 25]. The equations are solved using a preconditioned Newton-Krylov method that is converged to machine precision for each steady-state analysis case. The exact discrete adjoint of this model is also implemented and solved in an analogous fashion [27, 26].

This solver is validated using the standard test case of 5 km/s flow over a circular cylinder with a super-catalytic, fixed-temperature wall. The conditions for this test case can be found in Table 2. The results of this test case are compared with those of the well-validated code LAURA[5, 32] and are depicted in Figure 11. For these comparisons, the Park chemical kinetics model was used. Although this model shows better agreement with the validation codes, the Dunn-Kang model is ultimately used for all of the demonstration sensitivity and uncertainty quantification results due to the ease with which uncertain

Table 2: Benchmark flow conditions

$V_\infty =$	5 km/s
$\rho_\infty =$	0.001 kg/m^3
$T_\infty =$	200 K
$T_{wall} =$	500 K
$M_\infty =$	17.605
$Re_\infty =$	376,930
$Pr_\infty =$	0.72

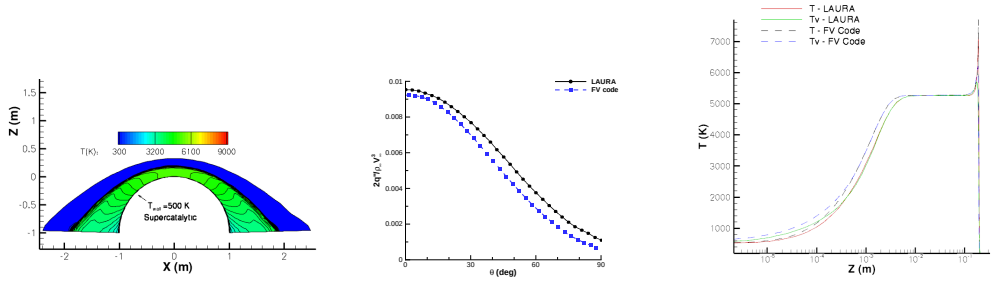


Figure 11: Validation of solver for 5 km/s flow over circular cylinder. Left: Computed flow field temperature contours. Middle: Surface heating distribution. Right: Comparison of temperatures along centerline with LAURA [5] results running on equivalent mesh.

parameters within the model may be specified. As the figure shows, the solver is able to match the LAURA validation results closely.

For the subsequent sensitivity analyses and uncertainty quantifications, the objective of interest is the integrated surface heating over the surface of the cylinder, given as:

$$L = -\frac{\int_{\partial\Omega} \left(k\nabla T \cdot \vec{n} + k_v \nabla T_v \cdot \vec{n}\right) dA}{\frac{1}{2}\rho_\infty V_\infty^3} \tag{62}$$

In this equation, T is the translational-rotational temperature, k is the translational-rotational thermal conductivity, T_v is the vibrational temperature, and k_v is the vibrational thermal conductivity.

In the following sections, we first perform a local sensitivity analysis in order to assess which parameters contribute the most uncertainty to the surface heating objective. This is followed by a global sensitivity analysis, using a polynomial regression model, and the results of local and global sensitivity analyses are compared. We then focus on the construction and use of gradient enhanced Kriging models for aleatory uncertainty quantification, as well as gradient-based optimization for epistemic uncertainty quantification. We conclude with an example of mixed aleatory-epistemic uncertainty quantification for the hypersonic cylinder case.

5.2 Local Sensitivity Analysis

The first variables examined relate to the specification of reaction rates. For the Dunn-Kang chemical kinetics model used within this work, the reaction rates take the following form.

$$K_f = C_f T_a^{\eta_f} e^{-\frac{E_{a,f}}{k_B T_a}} \tag{63}$$

$$K_b = C_b T_a^{\eta_b} e^{-\frac{E_{a,b}}{k_B T_a}} \tag{64}$$

Here $E_{a,f}$ and $E_{a,b}$ represent the activation energy for the forward and backward reactions respectively, k_B is Boltzmann's constant, and T_a is a characteristic temperature for the reaction. The parameters examined in this case were C_f and C_b for each reaction, giving a total of 34 parameters for the 17 reactions in the model. Figure 12 depicts the computed sensitivity of surface heating with respect to the forward and backward reaction rates using the adjoint procedure for the 5 km/s benchmark case. Note that because of the large discrepancy between the design variable and the objective, the sensitivity is expressed as fractional change in objective per fractional change in design variable (i.e. $\frac{\frac{dL}{L}}{\frac{dD}{D}}$) [28]. As the results demonstrate, the reactions governing the production and breakdown of NO, as well as the oxygen recombination reactions, have the greatest influence on integrated surface heating.

In addition to reaction rates, the sensitivity with respect to parameters within the transport model was calculated. For the collision integral model, measured collision integrals between the five species at 2000 K and 4000 K are used, and linear interpolation is used to determine the collision integral at the appropriate temperature [11].

$$log_{10}(\Omega_{s,r}^{k,k}) = log_{10}(\Omega_{s,r}^{k,k})_{2000} + \left[log_{10}(\Omega_{s,r}^{k,k})_{4000} - log_{10}(\Omega_{s,r}^{k,k})_{2000}\right] \frac{ln(T) - ln(2000)}{ln(4000) - ln(2000)} \tag{65}$$

In the above equation, $\Omega_{s,r}^{k,k}$ represents the collision integral between species s and species r. For the five-species model, 15 independent collision interactions are possible. This model gives a total of 60 parameters, since two separate collision integrals ($\Omega_{s,r}^{1,1}$ and $\Omega_{s,r}^{2,2}$) are used at each temperature. The collision integrals are treated as the parameters of interest and the variables examined for this analysis take the form of a multiplicative constant, $A_{s,r}^{k}$, on the input collision integrals $\hat{\Omega}_{s,r}^{k,k}$ [35]. Because the inputs at 2000 K and 4000 K are correlated, a single parameter is prescribed that accounts for shifts in the cross-section across the entire temperature range. This parameter modifies the collision integrals as follows:

$$\Omega_{s,r}^{k,k}(T) = A_{s,r}^{k} \hat{\Omega}_{s,r}^{k,k}(T) \tag{66}$$

where $\hat{\Omega}_{s,r}^{k,k}(T)$ is the unperturbed collision integral and $A_{s,r}^{k}$ is a multiplicative constant for each unique collision integral (30 total). The results of this calculation are also presented in Figure 12. As the results show, the collisions involving N_2 have the greatest effect on integrated surface heating. This result is unsurprising because N_2 is the predominant species at the super-catalytic wall boundary condition.

In order to determine the relative importance of parameters drawn from the various parts of the real gas model, the derivative magnitude over all the model parameters are

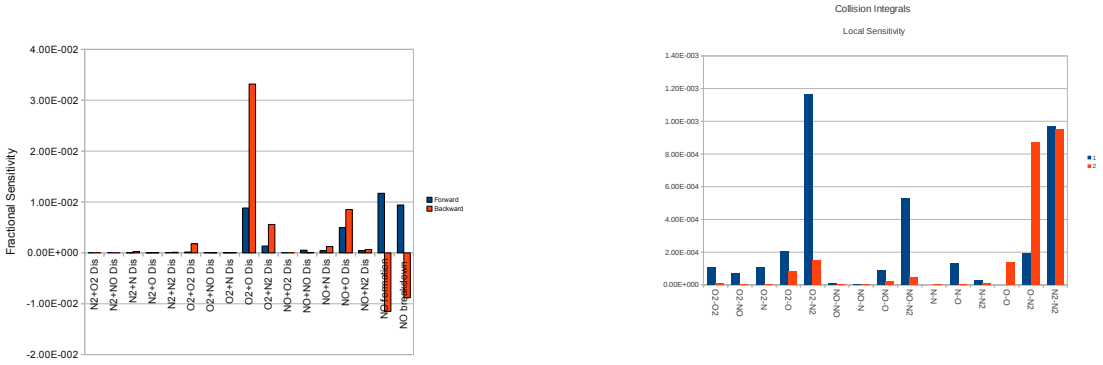


Figure 12: Left: Sensitivity of surface heating with respect to forward and backward reaction rates. Right: Sensitivity of surface heating with respect to collision integral parameters

directly compared. In addition to the parameters for the transport model and chemical kinetics model, the freestream density and velocity are also included in the sensitivity analysis in order to measure the relative importance of freestream conditions. The total number of uncertain parameters considered amounts to 66, which includes the 30 collision integral parameters described above, the 34 reaction rate parameters, and the 2 additional freestream flowfield parameters (density and velocity). The top ten parameters based on the derivative magnitude are given in Table 3. As these results demonstrate, the freestream density has the greatest effect on surface heating by several orders of magnitude. The rest of the top parameters are composed of the previously identified important collision integrals and reaction rates with the top collision integrals ranking above the majority of the top reaction rates.

5.3 Parameter Reduction through Global Sensitivity Analysis

The previously presented derivative sensitivity analysis did not consider the relative input uncertainties for each parameter. Also, the analysis is inherently localized in nature and may give inaccurate results for large perturbations of the input parameters or due to interference effects between variables [47]. To overcome these limitations, a global sensitivity analysis can be performed and the importance of each variable can be estimated based on its contribution to the simulation output. This global sensitivity analysis is performed using Monte Carlo sampling. For this method, design variable values are drawn from the associated input distribution and a CFD simulation is run to acquire an output sample. Statistics are then calculated based on these output samples. Although the work associated with this sampling is independent of the number of design variables, the expense of this sampling is often prohibitively high due to the slow convergence of output statistics. The sensitivity itself can be quantified using a linear regression analysis. This linear analysis calculates the correlation coefficient for each variable based on the Monte Carlo results. Although this method cannot fully separate the effect of each variable on the output, it can provide valuable information on the overall effect each variable has on the output measured over the entire design space. It can also account for interference effects between the variables and give a measure of the contribution each variable's uncertainty

Table 3: Ranking of input parameters based on derivative values

Number	Variable	Derivative Value
1	ρ_∞	4.887
2	O2-N2 (k=1)	1.1630×10^{-3}
3	N2-N2 (k=1)	9.6708×10^{-4}
4	N2-N2 (k=2)	9.4977×10^{-4}
5	O-N2 (k=2)	8.6974×10^{-4}
6	$O_2 + O \leftrightarrows 2O + O$ (f)	6.2120×10^{-4}
7	NO-N2 (k=2)	5.2660×10^{-4}
8	$NO + O \leftrightarrows N + O + O$ (b)	4.1447×10^{-4}
9	$N_2 + O \leftrightarrows NO + N$ (f)	3.9861×10^{-4}
10	$N_2 + O \leftrightarrows NO + N$ (b)	3.8755×10^{-4}

makes to the overall output uncertainty. The correlation coefficient for variable D_i is given by the following [47]:

$$r_i = \frac{cov(D_i, y)}{\sigma_{D_i}\sigma_y} \tag{67}$$

Here, y represents the output of interest from the simulation, σ_{D_i} represents the standard deviation of the input design variable and σ_y represents the standard deviation of the output. The standard deviation of the input design variable is a quantity that must be taken from the relevant literature or estimated based on some expert judgment or experience [42, 47]. The quantity σ_y is measured empirically from the Monte Carlo data set. Because the output variance is required, global sensitivity analysis is intricately tied to uncertainty quantification.

In practice, the correlation coefficient must be estimated based on the Monte Carlo data using estimators for each of the statistical quantities in equation (67). The correlation coefficient can therefore be estimated as [47, 1]:

$$r_i = \frac{\sum_k (D_{i,k} - \bar{D}_i)(y_k - \bar{y})}{\sqrt{\sum_k (D_{i,k} - \bar{D}_i)^2 \sum_k (y_k - \bar{y})^2}} \tag{68}$$

where $D_{i,k}$ represents the k^{th} sample of the i^{th} parameter input , $\bar{D}_i$ represents the mean value of the i^{th} parameter input, y_k denotes the output value at the k^{th} sample, $\bar{y}$ corresponds to the average output computed over all samples, and where the summation over k indicates a sum over the Monte Carlo sample points.

The expense of performing a full Monte Carlo analysis can be mitigated by constructing a surrogate model and performing an inexpensive Monte Carlo analysis directly on the surrogate model. The gradient-enhanced polynomial regression model detailed in section 3.1 can provide a basis for rapid global sensitivity analysis even in the limit of large input dimension when the degree of the regression is limited to 2. To demonstrate this capability, a global sensitivity analysis is performed on the real gas solver using the previously enumerated 66 model parameters based on a polynomial regression of degree 2. The results of this regression-based analysis are then compared with a Monte Carlo global sensitivity analysis performed directly using the CFD solver.

For 66 dimensions, the size of the regression basis is 2278 terms for $p = 2$. When gradient values are included, these facts indicate only 34 simulation results are required for a $p = 2$ regression. These numbers represent the minimum number of simulations and in practice additional simulations can be incorporated into the regression by additional over-determination of the system. In practice, a total of 68 function/gradient evaluations are used to train the model, computed from input samples distributed uniformly throughout the design space via Latin Hypercube sampling. For this analysis, these parameters are assumed to follow a Gaussian distribution. The results of the regression model global sensitivity analysis are compared against high-fidelity Monte Carlo sampling using 6,331 CFD simulation results based on the square of the correlation coefficient. The 66 variables used for this sensitivity analysis along with their mean and standard deviations taken from [47] are listed in Table 4. The error in the square of the correlation coefficient is estimated at 0.02 based on the results of a similar global sensitivity analysis with a similar number of samples [47]. Hence, when comparing coefficients, only differences greater than 0.04 should be considered when ranking parameters.

The results of this regression-based global sensitivity analysis are summarized in Table 5 and compared to the corresponding result from the high-fidelity Monte Carlo analysis. In this table, the ten most influential parameters on the surface heating are identified based on the regression-based global sensitivity analysis and compared with the ranking from the Monte Carlo analysis. Additionally, the square of the correlation coefficient predicted by each method is compared. Based on the results in Table 5, two conclusions can be made. First, the regression-based analysis produces parameter rankings and uncertainty contributions in relatively good agreement with the Monte Carlo results at a fraction of the cost (68 function/gradient results vs. 6,331 function results). Second, both the Monte Carlo and regression-based analysis indicate that the majority of the output uncertainty, measured by the square of the correlation coefficient, is the result of a small number of parameters, with these top ten accounting for over 90% of the output variance. Because the output variance is the result of a handful of variables, this sensitivity analysis can provide the justification for the dimension reduction used within the uncertainty quantification presented in the next section.

We also compare the results of the global sensitivity analysis with those obtained from the simpler local sensitivity analysis. For this comparison, the top parameters from the local sensitivity analysis are identified and compared with their ranking from the global sensitivity analysis in Table 6, using the high-fidelity Monte Carlo simulation results in

Table 4: Input Model Parameters with Uncertainty

Number	Variable	Mean	Standard Deviations
1	$\rho_\infty\ (kg/m^3)$	1×10^{-3}	5%
2	$V_\infty(m/s)$	5000	15.42
3-17	A^1_{s-r}	1	5%
18-32	A^2_{s-r}	1	5%
33-49	ξ_f	0	0.25
50-66	ξ_b	0	0.25

Table 5: Top 10 parameters from P=2 regression sensitivity analysis compared with global

Rank	Variable	Global Rank	Regression Contribution	Global Contribution
1	ρ_∞	1	0.56879	0.60055
2	$O_2 + O \leftrightarrows 2O + O$ (f)	2	1.0002×10^{-1}	1.0610×10^{-1}
3	$O_2 + O_2 \leftrightarrows 2O + O_2$ (b)	6	5.7669×10^{-2}	2.1621×10^{-2}
4	$NO + O \leftrightarrows N + O + O$ (b)	3	4.0057×10^{-1}	5.1914×10^{-2}
5	N2-N2 (k=1)	5	3.7461×10^{-2}	3.1617×10^{-2}
6	O2-N2 (k=1)	4	3.3299×10^{-2}	4.2121×10^{-2}
7	N2-N2 (k=2)	8	2.1163×10^{-2}	1.9019×10^{-2}
8	O-N2 (k=2)	9	1.7395×10^{-2}	1.3874×10^{-2}
9	V_∞	14	1.3497×10^{-2}	4.8401×10^{-3}
10	$O_2 + O \leftrightarrows 2O + O$ (b)	13	1.1734×10^{-2}	7.4280×10^{-3}

the case of the global sensitivity analysis. In addition to comparing the ranking of each parameter, the uncertainty contribution is also compared. For the local sensitivity results, this is computed using the method of moments. The inverse of Table 6 is given in Table 7 where the top parameters from the global sensitivity are given with their associated local ranks.

As this comparison shows, the local sensitivity analysis does a reasonable job identifying the most sensitive parameters with the global and local analysis. The two analyses identify the same top 3 parameters and identify 7 of the same parameters within their respective top 10 lists. Unfortunately, outside of the top three, the two types of analysis begin giving dramatically different rankings and the local sensitivity analysis seems to miss a number of the important collision integrals. Based on the uncertainty in the global sensitivity results, the disagreement in rankings is only important for variables with correlation coefficients significantly greater than 0.02, namely $O2 - N2(k = 1)$ and $N2 - N2(k = 1)$. The fact that the local analysis misidentifies the ranking of these parameters demonstrates a significant disagreement between the two forms of sensitivity analysis and illustrates the importance large perturbations and non-linear interference effects can have on overall sensitivities.

5.4 Aleatory Uncertainty Quantification using Kriging Model

We now demonstrate the use of a Kriging model for uncertainty quantification of the surface heating objective for the hypersonic cylinder case based on the 66 uncertain parameters as described previously in Table 4.

From the global sensitivity analysis presented in the previous section, it was shown that the majority of the uncertainty is the result of a relatively small number of variables. Because of this fact, the dependence of the output on the lower importance variables can be neglected, allowing for the creation of a lower-cost Kriging model with respect to the variables contributing the majority of the uncertainty. In order to demonstrate the performance of this strategy, Kriging models based on the top 15 variables identified by

Table 6: Top 10 parameters from local sensitivity analysis compared with global

Rank	Variable	Global Rank	Local Contribution	Global Contribution
1	ρ_∞	1	0.43230	0.60055
2	$O_2 + O \leftrightarrows 2O + O$ (f)	2	1.7490×10^{-1}	1.0610×10^{-1}
3	$NO + O \leftrightarrows N + O + O$ (b)	3	7.7860×10^{-2}	5.1914×10^{-2}
4	$N_2 + O \leftrightarrows NO + N$ (f)	7	7.2017×10^{-2}	2.0647×10^{-2}
5	$N_2 + O \leftrightarrows NO + N$ (b)	10	6.8076×10^{-2}	1.2155×10^{-2}
6	$NO + O \leftrightarrows N + O + O$ (f)	11	2.5979×10^{-2}	8.3596×10^{-3}
7	O2-N2 (k=1)	4	2.4524×10^{-2}	4.2121×10^{-2}
8	$NO + O \leftrightarrows O_2 + N$ (f)	16	2.2155×10^{-2}	1.4609×10^{-3}
9	$NO + O \leftrightarrows O_2 + N$ (b)	63	1.7353×10^{-2}	3.7488×10^{-6}
10	N2-N2 (k=1)	6	1.6956×10^{-2}	3.1617×10^{-2}

Table 7: Top 10 parameters from global sensitivity analysis compared with local

Rank	Variable	Local Rank	Global Contribution	Local Contribution
1	ρ_∞	1	0.60055	0.43230
2	$O_2 + O \leftrightarrows 2O + O$ (f)	2	1.0610×10^{-1}	1.7490×10^{-1}
3	$NO + O \leftrightarrows N + O + O$ (b)	3	5.1914×10^{-2}	7.7560×10^{-2}
4	O2-N2 (k=1)	7	4.2121×10^{-2}	2.4524×10^{-2}
5	N2-N2 (k=1)	10	3.1617×10^{-2}	1.6956×10^{-2}
6	$O_2 + O_2 \leftrightarrows 2O + O_2$ (b)	13	2.1621×10^{-2}	1.3120×10^{-2}
7	$N_2 + O \leftrightarrows NO + N$ (f)	4	2.0647×10^{-2}	7.2017×10^{-2}
8	N2-N2 (k=2)	11	1.9019×10^{-2}	1.6354×10^{-2}
9	O-N2 (k=2)	12	1.3874×10^{-2}	1.3714×10^{-2}
10	$N_2 + O \leftrightarrows NO + N$ (b)	5	1.2155×10^{-2}	6.8076×10^{-2}

the Monte Carlo based global sensitivity analysis as accounting for 95% of the variance, given in Table 8, were constructed. Both function and function/gradient Kriging models were constructed, and the accuracy of these models was compared with a full Monte Carlo simulation, using 6,331 samples, as a function of the number of training points used to build the Kriging models. The error in the surrogate is calculated by sampling the Kriging models at the same points used for the high-fidelity Monte Carlo sampling and measuring the difference between the Kriging predictions and the Monte Carlo results. Because the full Monte Carlo results include the effect of all uncertain parameters, the errors incurred by the Kriging models include both differences due to the omission of the low ranking uncertain parameters, and the accuracy of the Kriging model itself due to a finite number of training points.

The error as the number of training points increases is plotted in Figure 13 for both the function-only models and the function/gradient-enhanced models. As the figure shows, the gradient-enhanced model consistently produces a more accurate surrogate than the function-only model. Additionally, the error for the gradient-enhanced model quickly levels off while the function-only model shows large variation over all but the largest numbers of training points. This variation may indicate that the sampling density for the function-only model may not be sufficient until a large number of points is reached.

Figure 14 shows the average and variance predictions as the number of training points increases for both the function only model and the function/gradient-enhanced model. As the figure demonstrates, the average predictions for the gradient-enhanced model show less variation and have better agreement with the Monte Carlo results when compared to the function-only results. For the variance, both models show discrepancies with the Monte Carlo results but the predictions for the gradient-enhanced model show less variation as the number of training points increases, indicating some level of convergence for the model training.

Finally, to examine the prediction of statistics beside the mean and variance based on a Kriging model, the cumulative distribution function (CDF) of the output based on the gradient-enhanced Kriging model was constructed using 68 training points. This CDF is plotted in Figure 15 along with the CDF from the high-fidelity Monte Carlo results. The CDF curve produced by the Kriging model matches the CDF from the Monte Carlo results well in spite of the fact that a majority of the input variables have been ignored within the construction of the Kriging model and only 68 simulations are required, compared to 6,331 simulations for the full Monte Carlo method.

5.5 Epistemic Uncertainty Quantification

We now consider gradient-based approaches for epistemic uncertainty quantification. Because of the expense associated with traditional methods of episetmic uncertainty, a subset of the input variables used in the previous sections is chosen and treated as epistemic. Using this variable subset, the output interval for the hypersonic cylinder surface heating objective is predicted using a linear surrogate and an optimization approach. The performance of these methods is measured by comparing against the bounds produced by exhaustive sampling.

The traditional method for quantifying epistemic uncertainty is Latin Hypercube sampling. This sampling is performed using a fixed number of samples (~ 3) in each direction

Table 8: Variables for the Kriging model based on Monte Carlo global sensitivity analysis

Rank	Variable	Uncertainty Contribution	Total Contribution
1	ρ_∞	0.60055	0.60055
2	$O_2 + O \leftrightarrows 2O + O$ (f)	1.0610×10^{-1}	0.70665
3	$NO + O \leftrightarrows N + O + O$ (b)	5.1914×10^{-2}	0.75857
4	O2-N2 (k=1)	4.2121×10^{-2}	0.80069
5	N2-N2 (k=1)	3.1617×10^{-2}	0.83231
6	$O_2 + O_2 \leftrightarrows 2O + O_2$ (b)	2.1621×10^{-2}	0.85393
7	$N_2 + O \leftrightarrows NO + N$ (f)	2.0647×10^{-2}	0.87457
8	N2-N2 (k=2)	1.9019×10^{-2}	0.89359
9	O-N2 (k=2)	1.3874×10^{-2}	0.90747
10	$N_2 + O \leftrightarrows NO + N$ (b)	1.2155×10^{-2}	0.91962
11	$NO + O \leftrightarrows N + O + O$ (f)	8.3596×10^{-3}	0.92798
12	NO-N2 (k=1)	7.6676×10^{-3}	0.93565
13	$O_2 + O \leftrightarrows 2O + N$ (b)	7.4280×10^{-3}	0.94308
14	V_∞	4.8401×10^{-3}	0.94792
15	$O_2 + N_2 \leftrightarrows 2O + N_2$ (f)	3.5848×10^{-3}	0.95150

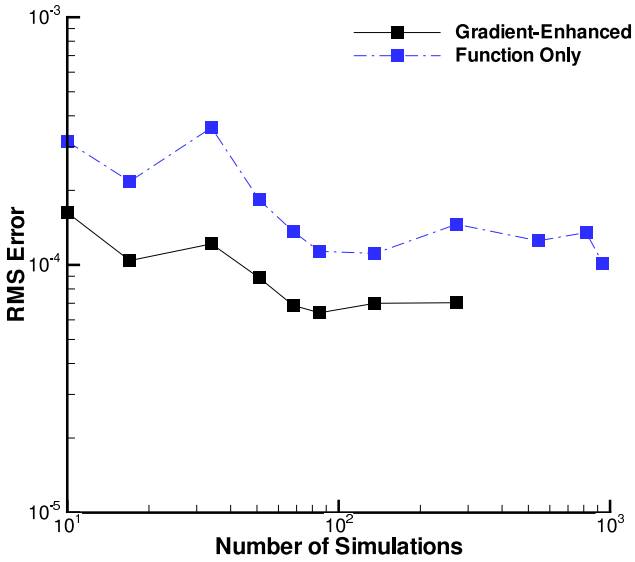


Figure 13: Variation of surrogate error as number of training points increases for gradient-enhanced and function-only Kriging models

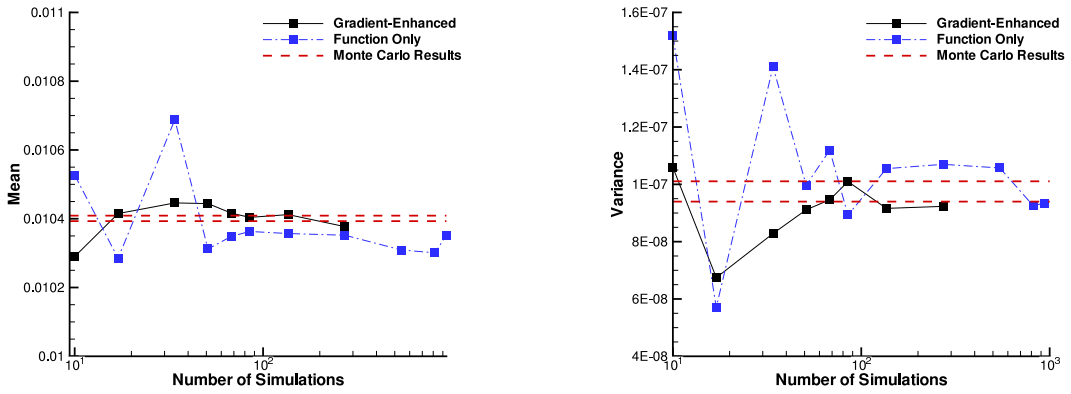


Figure 14: Variation of mean (Left) and variance (Right) of surface heating objective as number of training points increases for gradient-enhanced and function-only Kriging models

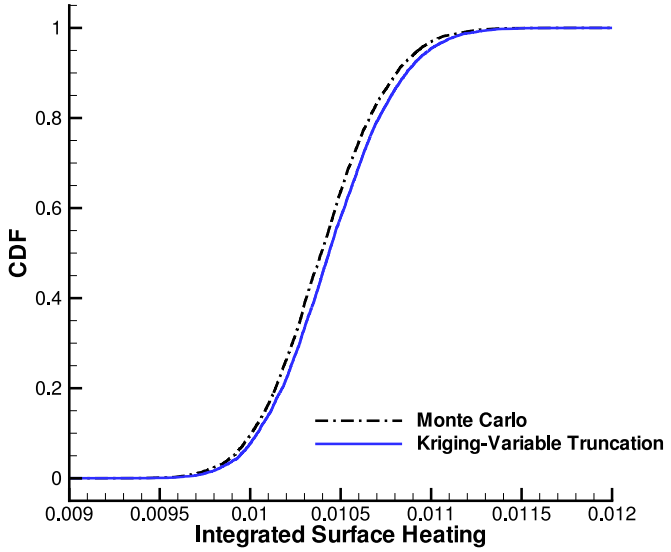


Figure 15: CDF prediction of surface heating objective for Kriging model based on 68 training points in 17 dimensions compared with Monte Carlo CDF.

(with a total number of samples increasing as 3^d with d representing the number of epistemic variables). Due to the expense of sampling approaches, the dimension for this problem is reduced relative to the aleatory uncertainty problem presented in the previous section. For the purposes of validation, the set of 66 variables used for the aleatory uncertainty problem is reduced to a set of 8 variables for the epistemic problem. Limiting the validation to 8 variables allowed the output interval to be computed in a cost equivalent to that of an aleatory problem. Using 3 sample points in each direction, as indicated in Reference [42], the choice of 8 epistemic variables requires $6,561$ simulation results to provide the appropriate interval, a cost similar to the 6,331 results used to validate the aleatory results. The choice of which 8 parameters to include in the epistemic study was motivated by both theoretical and practical concerns. Of the 66 parameters used in the sensitivity analysis and aleatory uncertainty quantification, the parameters corresponding to the collision integrals and chemical kinetics model are most accurately described by epistemic uncertainty due to their experimental nature and the lack of information in the literature regarding the distribution of these values.

Limiting ourselves to the 30 transport parameters in the model, the final set of 8 parameters was chosen based on the sensitivity analysis presented in section 5.3. From this analysis, the collisions involving diatomic nitrogen have the greatest effect on surface heating. The final set of 8 variables are chosen as the collision integrals of diatomic nitrogen with itself and the other predominant species near the wall, namely O_2, O, and N, for k equal to 1 and 2. These variables are given in Table 9 with their associated intervals. The assumption of $\pm 20\%$ uncertainty for the collision integrals is taken from Reference [35], although that work assumed a Gaussian for the uncertainty while an interval is used in this work. It should be noted that the parameters chosen for the epistemic problem do not correspond exactly with the 8 most important collision integrals from the sensitivity analysis, although that criteria would indicate 6 of the same parameters. Although the sensitivity analysis played a role in selecting the parameters used for the epistemic problem, the variable set was modified slightly to more closely match the experiences of Reference [20]. Had similar studies not been available, the sensitivity analysis could have been used exclusively.

The output interval due to variations in these 8 parameters was determined based on Latin Hypercube sampling using 3 samples in each dimension ($6,561$ total samples). For this sampling, variations were only allowed for these 8 variables and the other variables were frozen at their unperturbed or mean values. With the samples determined, the minimum and maximum objective values from the data set are determined. Because no distribution can be associated with epistemic uncertainty, this output interval is the only

Table 9: Epistemic Model Parameters

Variable	Unperturbed Value	Lower Bound	Upper Bound
A^1_{N2-N2}, A^2_{N2-N2}	1	0.8	1.2
A^1_{N2-N}, A^2_{N2-N}	1	0.8	1.2
A^1_{N2-O}, A^2_{N2-O}	1	0.8	1.2
A^1_{N2-O2}, A^2_{N2-O2}	1	0.8	1.2

useful result of the sampling, as the epistemic nature of this problem prevents inferring any other properties. The choice of using 3 samples in each dimension was made based on Reference [42]. As will be shown, this choice may lead to underestimation of the output interval. Hence, when comparing the results of other methods to the interval predicted by sampling, an interval prediction larger than the sampling result is preferred, as the sampling result is limited by the number of samples used and is used to guard against the possibility of the optimization falling into a local optimum. Using these validation results, the performance of the gradient-based methods can be measured.

In addition to this 8 dimensional problem, the gradient-based methods are used to predict the output interval due to uncertainty in all 30 transport parameters. Although interval predictions are made, no validation of the interval based on sampling is possible due to the prohibitively high cost associated with sampling. Using 3 samples in each direction for 30 dimensions would require over 100 trillion samples. Instead of comparing the results of the gradient-based method to a sampling based approach, the methods can only be compared with one another. The interval for all 30 transport parameters was $[0.8, 1.2]$ and again corresponds to $\pm 20\%$ uncertainty [35].

5.5.1 Linear Results

When gradient values are available, the width of the output interval can be predicted based on a single function/gradient evaluation at the unperturbed parameter values. The half width of the interval is predicted using the gradient values (according to equations (53) and (54)) and the lower and upper bounds of the output interval are predicted by subtracting and adding this half width to the function prediction of the simulation. The results of this linear approximation for the 8 dimensional case are given in Table 10 and compared to the interval produced by Latin Hypercube sampling.

Table 10: Interval prediction from moment method compared with LHS result

	Linear Method	LHS interval
Center	1.0370E-002	1.0449E-002
Interval Half Width	8.6634E-004	7.1266E-004
Upper	1.1237E-002	1.1161E-002
Lower	9.5040E-003	9.7361E-003
Percentage	$\pm 8.35\%$	$\pm 6.82\%$

As this table demonstrates, the linear method gives a conservative estimate for both the lower bound and the upper bound of the interval. Also included in Table 10 is the center of the interval predicted by the LHS sampling and linear method. As the table shows, the center of the interval from sampling deviates significantly from the function value evaluated using parameters at the center of the interval. This fact gives an indication of the non-linearity of the simulation result. Additionally, it demonstrates the difficulty in determining a representative result for the simulation value. While the center of the interval prediction may be an intuitive representative result, any value inside the interval bounds is equally valid. Despite this difficulty, the center of the interval is used as a

representative value so the uncertainty can be represented as a plus/minus value about this value. When this percentage error about the center is computed, it is clear that the linear method produces a significantly wider output uncertainty when compared to the sampling results.

Because a single function/gradient evaluation is used to predict the output interval, this method can be applied to the 30 dimensional problem without increasing the cost of the prediction. Although no validation results are presented for this case, the 30 dimensional interval prediction is given in Table 11. As expected, the size of the output interval, relative to the 8 dimension case, is larger as a result of adding more uncertain parameters. Although no validation of this result is possible, it will be compared to the corresponding optimization result in the next section.

Table 11: Interval prediction using moment method for 30 dimensional case.

	Linear Method
Center	1.0370E-002
Interval Half Width	1.1787E-003
Upper	1.1549E-002
Lower	9.1916E-003
Percentage	11.37%

5.5.2 Optimization results

In order to provide an inexpensive interval prediction even as the problem dimension increases, gradient-based optimization (leveraging the adjoint capability) is used to determine the minimum and maximum values possible given specified intervals on the inputs. Specifically, the L-BFGS algorithm [50] is used for the optimization. The optimization approach is first tested for the 8 dimensional test case. The convergence for the minimum and maximum optimization results is shown in Figure 16 in terms of function/gradient evaluations. Plotted with the convergence are the minimum and maximum values from the LHS results. As the figure demonstrates, the interval produced by optimization is larger than the sampling result, producing conservative estimates for the minimum and maximum values.

Because the optimization result produces interval bounds beyond the values produced by sampling, the optimization results should be viewed as the correct result. Because the results of the optimizations represent actual function values achieved with inputs contained in the specified intervals, the bounds produced by optimization are the correct solution to equations (55) provided the minimum and maximum values are the global extrema. Hence, the sampling results are used to ensure the optimization has not fallen into any local extrema. Although the sampling does not ensure that the optimization gives the global minimum and maximum values, the sampling results provide no evidence to undermine the optimization results. Even though L-BFGS is an inherently local optimization, in this case it appears to produce the global minimum and maximum values. In addition to producing the more accurate interval estimate, the optimization produces it

at significantly reduced cost, requiring only 43 function/gradient evaluations. The rapid convergence of these optimization problems is likely due to the relatively smooth behavior of the design space and the fact that the optimal values typically occur at the bounds for most variables. Although a smooth design space is typical for most outputs of engineering calculations [29, 41], it is likely more complex simulations will require more work to solve the required optimization problems.

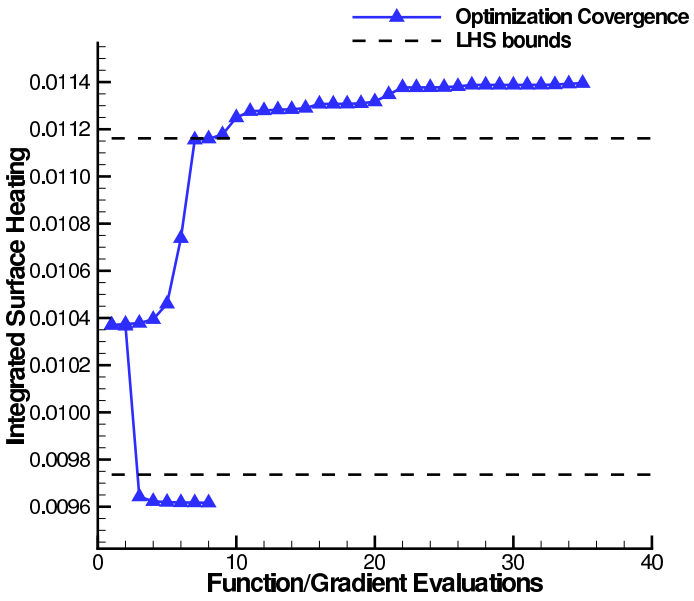


Figure 16: Convergence of optimization over epistemic variables (8 dimensional case) for fixed aleatory variables compared with bounds from Latin hypercube sampling.

In order to further compare the optimization to the other results, its bounds are compared with sampling and the linear method in Table 12. In addition to comparing the bounds, the center of the interval as well as a symmetric uncertainty percentage are given. Based on these results, the optimization and linear method produce similar measures of uncertainty as a percentage of the center of the interval. Despite this fact, the bounds produced by the linear method show disagreement with the optimization result, with the linear method underestimating the maximum and producing an overly conservative minimum bound.

Because the gradient-based optimization can scale to larger dimension, the bounds for the 30 dimensional problem can be determined. These bounds are compared with the results from the linear method. Although the agreement of these two methods does not imply that the bound is necessarily correct, the fact that the predictions are based on differing methods should give an indication of the quality of the result. Because the optimization result is based on actual function values, instead of extrapolation, the optimization result is likely the more correct result. The convergence of the 30 dimensional optimization problems is given in Figure 17. Using 83 function/gradient evaluations, the optimization method is able to produce the output interval for the 30 dimensional problem. Despite the large increase in the dimension of the problem, the cost of the optimization only increased by approximately a factor of two. This fact is likely because

Table 12: Comparison of Interval prediction for Optimization, Sampling and Linear Method (8 dimensional case)

	Linear Method	LHS interval	Optimization
Center	1.0370E-002	1.0449E-002	1.0506E-002
Interval Half Width	8.6634E-004	7.1266E-004	8.8912E-004
Upper	1.1237E-002	1.1161E-002	1.1395E-002
Lower	9.5040E-003	9.7361E-003	9.6168E-003
Percentage	8.35%	6.82%	8.46%

the optimal values are located at the bounds of the domain for most variables, allowing the optimizer to easily find the optimum. Additionally, because the sensitivity of many collision integrals is small, the addition of these variables to the optimization does not drastically effect the overall design space. The interval produced by optimization on the 30 dimensional problem is compared to the linear method in Table 13. As the table shows, the linear method again underestimates the maximum value but is overly conservative for the minimum value. Although the width of the interval and center of the interval do not agree with the optimization, the uncertainty expressed as a percentage of the center value agree remarkably well for the two methods, as was the case for the 8 dimensional case. Given the difference in bound predictions, this fact is likely coincidental.

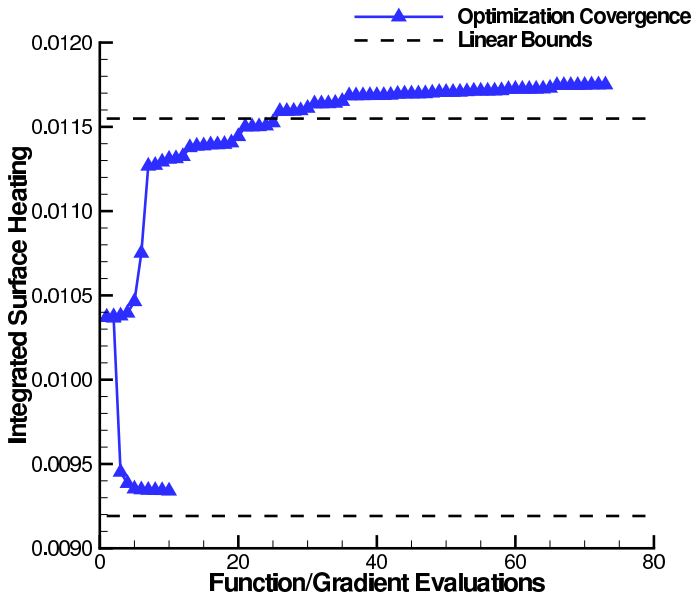


Figure 17: Convergence of optimization over epistemic variables (30 dimensional case) for fixed aleatory variables.

Table 13: Interval prediction for optimization method for 30 dimensional problem.

	Linear Method	Optimization
Center	1.0370E-002	1.0543E-002
Interval Half Width	1.1787E-003	1.2031E-003
Upper	1.1549E-002	1.1746E-002
Lower	9.1916E-003	9.3400E-003
Percentage	11.37%	11.41%

5.6 Combined Aleatory-Epistemic UQ

To quantify uncertainty for the mixed case, a combined optimization/Kriging approach is applied in a statistics-of-intervals manner, meaning that multiple optimizations are performed over different values of the aleatory variables and a surrogate is created for these optimization results. To test this strategy, ten variables are treated as uncertain, given in Table 14. The epistemic variables for the mixed case are the same eight variables as those used for the pure epistemic case. Two aleatory variables are used for this study, corresponding to the freestream density and velocity. These parameters and associated uncertainty were chosen based on the results of similar uncertainty quantification studies[35, 2]. The choice of which parameters to treat as aleatory was taken from an additional study[3].

The goal of a mixed aleatory/epistemic uncertainty quantification is to produce an interval containing the solution with a specified probability, known as a P-box [42]. The performance of the statistics-of-interval approach is measured based on its predictions of a 99% P-box as well as its ability to predict the CDF curves associated with the minimum and maximum values of the optimization, enabling any P-box to be predicting in principle. These predictions can be validated in one of two ways. First, a nested sampling approach can be used, in which an aleatory uncertainty quantification is performed for each epistemic sample [42]. Using three points in each dimension for the sampling of epistemic variables and 5000 samples for the aleatory uncertainty quantification, this nested approach would require approximately 30 million simulation results for the variables used in this study. Aside from nested sampling, the statistics associated with the

Table 14: Uncertain model parameters

Variable	Type	Uncertainty
ρ_∞ (kg/m^3)	Aleatory	$\pm 10\%$ ($\sigma = 5\%$)
$V_\infty (m/s)$	Aleatory	± 30.84 ($\sigma = 15.42$)
$\Omega^{1,1}_{N2-N2}, \Omega^{2,2}_{N2-N2}$	Epistemic	$\pm 20\%$
$\Omega^{1,1}_{N2-N}, \Omega^{2,2}_{N2-N}$	Epistemic	$\pm 20\%$
$\Omega^{1,1}_{N2-O}, \Omega^{2,2}_{N2-O}$	Epistemic	$\pm 20\%$
$\Omega^{1,1}_{N2-O2}, \Omega^{2,2}_{N2-O2}$	Epistemic	$\pm 20\%$

interval bounds can be validated by exhaustive sampling of the optimization results over the aleatory variables. Hence, thousands of pairs of optimization results would be required to validate the statistics associated with the interval bounds. If 5000 samples are performed to characterize the distribution of the interval bound due to the aleatory variables, approximately 300,000 simulation results with gradients would be required for this test. For both approaches, this validation is essentially intractable. To validate the Statistics-of-Intervals/Kriging method applied to a real gas simulation, each element of the method was validated separately against exhaustive sampling. With each element validated, the mixed aleatory/epistemic uncertainty was calculated by using successively more accurate surrogate models to demonstrate convergence of the statistic predictions.

The optimization portion of the Statistics-of-Intervals method was validated in the previous section. With the optimization confirmed, the ability of the Kriging model to capture the aleatory variation of the integrated surface heating was tested. Although the previously discussed method for aleatory uncertainty quantification was based on a gradient-enhanced Kriging model, the Statistics-of-Intervals method uses a function-only Kriging model. Because the Statistics-of-Intervals method requires the construction of a Kriging model for the optimization results, the gradient of the optimal results would be required for a gradient-enhanced method, a quantity that is difficult to calculate. Luckily, within hypersonic flows, the number of aleatory variables is relatively small, reducing the need for a gradient-enhanced surrogate model.

In order to determine the number of training points required to capture the design space associated with the two aleatory variables identified in Table 14, Kriging models were constructed using an increasing number of simulation results and statistics were predicted based on each model. For this test, the epistemic variables were frozen at their non-perturbed values (1 in the terms of the parameters defined in Table 14), and sampling was performed over the aleatory variables. In order to provide validation data, Monte Carlo sampling was performed over the two aleatory variables, and the distribution was characterized both by constructing a CDF curve and by calculating specific statistics. In order to acquire accurate statistics, 4,564 samples were used, and a separate simulation was performed for each. With the validation data acquired, ordinary (constant mean function) Kriging models with increasing numbers of training points were constructed. Because the epistemic variables for this test were fixed, each training point required only a single CFD simulation.

As a first test, the convergence of the mean, variance, and $99th$ percentile are shown for Kriging models with increasing numbers of training points. The convergence of this metric as a function of training point number is given in Table 15. As the results show, predictions of the Kriging model rapidly converge toward the Monte Carlo results. In addition to predicting distribution statistics, a CDF of the output is constructed based on samples extracted from the Kriging model and compared with that of Monte Carlo sampling. Figure 18 shows the predicted CDF curve for a Kriging model with 8 training points and the CDF from Monte Carlo sampling. Using only 8 samples, the Kriging model produces a CDF curve nearly identical to the curve produced through Monte Carlo sampling, at a fraction of the cost. Based on this result, it can be expected that a similar number of optimization results should be required to accurately predict the mixed result. Since this problem only considers the uncertainty due to two aleatory variables, this cost is most likely overly optimistic for typical problems.

Table 15: Convergence of Kriging statistic predictions for aleatory uncertainty with fixed epistemic variables with increasing number of training points

Training Points	Average	Variance	$99th$ Percentile
8	1.036110×10^{-2}	6.061055×10^{-8}	1.098518×10^{-2}
16	1.036622×10^{-2}	6.075630×10^{-8}	1.097558×10^{-2}
31	1.034997×10^{-2}	6.145065×10^{-8}	1.098506×10^{-2}
59	1.037171×10^{-2}	6.185576×10^{-8}	1.099184×10^{-2}
121	1.036669×10^{-2}	6.120957×10^{-8}	1.097695×10^{-2}

Monte Carlo Results

Samples	MC Average	MC Variance	MC $99th$ Percentile
4564	1.036082×10^{-2}	6.103365×10^{-8}	1.098385×10^{-2}

With each element of the Statistics-of-Intervals/Kriging approach validated independently, the complete mixed aleatory/epistemic uncertainty is predicted by using optimization for the epistemic dependence and an ordinary Kriging model for the aleatory dependence. In order to demonstrate the validity of the full results, the convergence of the minimum and maximum $99th$ percentile predictions are shown as the number of training points for the Kriging model is increased in Table 16. For the mixed results, a training point now represents a pair of optimizations and has a cost of approximately 60 function/gradient evaluations on average. As the table demonstrates, the statistic predictions quickly converge to asymptotic values. Included in Table 16 is the total cost in terms of function/gradient evaluations. While the nested sampling and exhaustive sampling of the optimization were prohibitively expensive for the CFD model, the Statistics-of-Intervals/Kriging model was able to capture converged statistics with a number of function/gradient evaluations within the computational budget (although still most likely prohibitively high for complex simulations). Nevertheless, by using the Kriging model combined with optimization, the Statistics-of-Intervals/Kriging method was able to quantify the mixed aleatory/epistemic uncertainty problem where other methods could not be used.

Figure 19 shows the convergence of the average and variance prediction based on

Table 16: $99th$ percentile predictions for SOI method using ordinary Kriging model for real gas CFD simulation

Training Data Size	Number of F/G Evaluations	99^{th} percentile of Min	99^{th} percentile of Max
8	500	1.017556×10^{-2}	1.206949×10^{-2}
15	900	1.016681×10^{-2}	1.207132×10^{-2}
23	1400	1.018928×10^{-2}	1.207939×10^{-2}
52	3000	1.020232×10^{-2}	1.210513×10^{-2}
104	6176	1.020243×10^{-2}	1.210416×10^{-2}

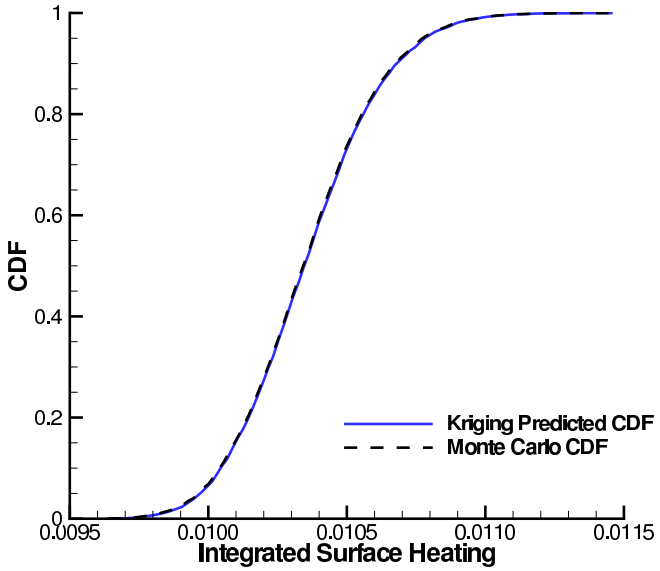


Figure 18: CDF based on Kriging model using 8 sample points compared with CDF of Monte Carlo results with fixed epistemic variables.

Kriging models with increasing numbers of training points. As this figure shows, the statistics produced from each model show little variation as the number of training points increases and the variation is small compared with the overall interval produced due to the epistemic uncertainty. In addition to calculating specific statistics of the output interval, the CDF of the minimum and maximum values can be predicted by sampling from the Kriging surface. The bounding CDF curves are plotted in Figure 20 for a Kriging model based on 8 and 104 pairs of optimizations. As the figure demonstrates, the CDF curves are nearly identical, suggesting that the Kriging model has reached some level of convergence.

6 Conclusion and Further Work

In these notes, we have shown how adjoint techniques can be used to accelerate and enhance applications in sensitivity analysis and uncertainty quantification (UQ). The principal advantage adjoint techniques bring to UQ is the ability to compute the sensitivity of a simulation objective or quantity of interest with respect to any number of input parameters at a cost similar to a single analysis simulation. Additionally, adjoint methods enable the computation of the Hessian matrix of second derivatives at an additional cost equivalent and N forward sensitivity solutions for N input parameters. The resulting first and second-order sensitivities can be used directly in sensitivity analysis to identify the most important parameters, or in uncertainty quantification of the objective, either by propagating statistical metrics of uncertainties through the simulation using the method of moments, or by providing additional information at reduced computational expense for the construction of surrogate models. Additionally, for epistemic uncertainties, the UQ problem can be posed as the solution of a bound optimization problem for which adjoint

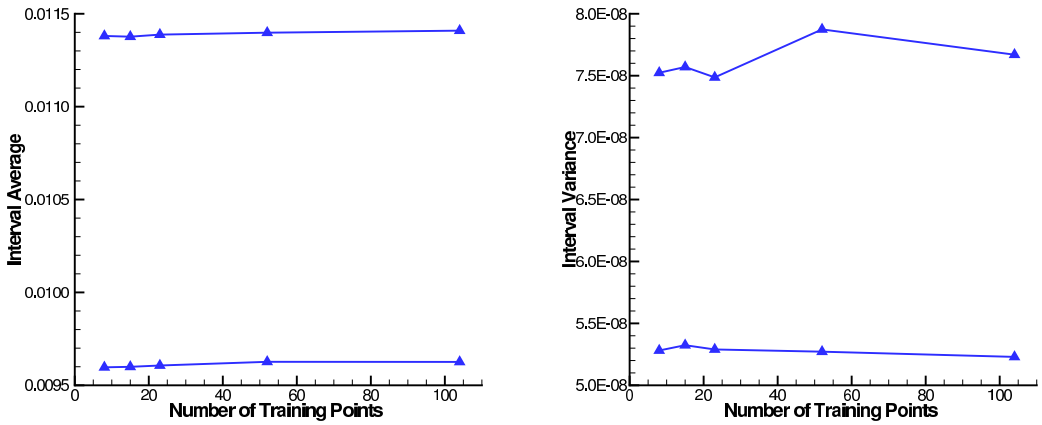


Figure 19: Convergence of average (Left) and variance (Right) prediction for minimum and maximum distribution using Kriging models built from increasing numbers of optimization results for real gas CFD simulation.

methods can provide efficient gradient computations.

While the use of first-order sensitivities in uncertainty quantification problems is relatively well established, the potential for Hessian information to accelerate uncertainty quantification techniques remains relatively unexplored. In the examples described in these notes, the Hessian was shown to be advantageous for model problems based on analytic functions, but was not used on practical uncertainty quantification problems due to the cost and associated lack of smoothness of the output functional in some cases. Further investigations into the use of Hessian or Hessian-vector product information for uncertainty quantification should be an active area of research in the future.

The sensitivity analysis and uncertainty quantification examples described in these notes were mostly developed for steady-state single disciplinary problems. Based on the development of adjoint methods for time-dependent and multi-disciplinary problems in the preceding chapters, these techniques should be extensible to more complex problems of these types, although the overall expense of high-fidelity UQ will remain an impediment to widespread use. Finally, the potential of adjoint methods to be used in conjunction with more elaborate UQ techniques, involving combinations of the approaches described herein, but also involving Bayesian methods such as Bayesian belief networks for multidisciplinary problems remains relatively unexplored and can be considered to be an area of research with potentially high payoff.

7 Acknowledgements

Significant contributions to these notes were provided by Dr. Brian Lockwood, whose PhD thesis and prior publications provided most of the material for hypersonic flow uncertainty quantification sections. Other material is drawn from joint work with Dr. Wataru Yamazaki, who was a visiting scientist at the University of Wyoming, and Dr. Markus Rumpfkeil, who was a postdoctoral researcher at the University of Wyoming.

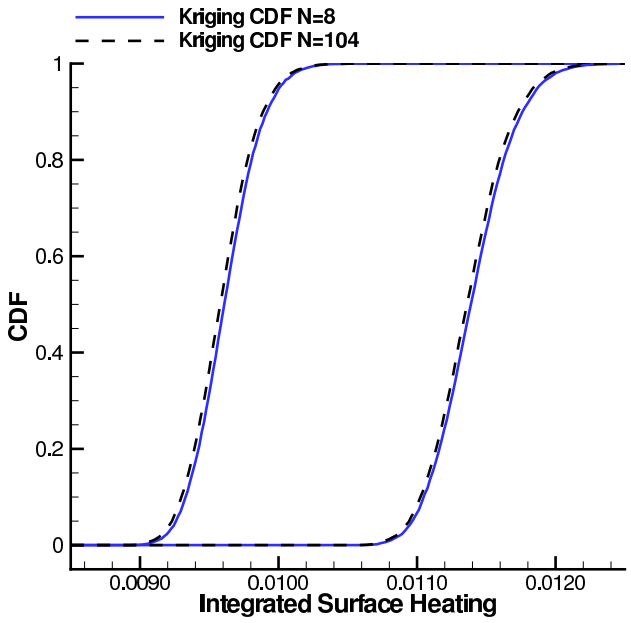


Figure 20: Kriging-predicted CDF curves for maximum and minimum values using 8 and 104 optimization pairs.

References

[1] Covariance, March 2012. http://en.wikipedia.org/wiki/Covariance, Last modified 30 March 2012.

[2] A. Alexeenko, A.B. Weaver, R.B. Greendyke, and J.A. Camberos. Flowfield uncertainty analysis for hypersonic CFD simulations. In *48th* AIAA *Aerospace Sciences Meeting and Exhibit*, Orlando, FL, January 2010. AIAA Paper, 2010-1180.

[3] B. R. Bettis and S. Hosder. Uncertainty quantification in hypersonic reentry flows due to aleatory and epistemic uncertainties. In *49th* AIAA *Aerospace Sciences Meeting and Exhibit*, Orlando, FL, January 2011. AIAA Paper, 2011-252.

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

[5] F. Cheatwood and P. A. Gnoffo. User's manual for the Langley aerothermodynamic upwind relaxation algorithm (LAURA). Technical report, NASA, April 1996.

[6] H. S. Chung and J. J. Alonso. Using gradients to construct cokriging approximation models for high-dimensional design optimization problems. In *40th* AIAA *Aerospace Sciences Meeting and Exhibit*, Reno, NV, January 2002. AIAA Paper, 2002-0317.

[7] N. Cressie. The Origins of Kriging. *Mathematical Geology*, 22(3):239–252, 1990.

[8] J. R. Edwards. A low-diffusion flux-splitting scheme for Navier-Stokes calculations. *Computers & Fluids*, 26(6):635–659, 1997.

[9] D. Ghate and M.B. Giles. Inexpensive Monte Carlo uncertainty analysis. In B. Uthup, S. Koruthu, R.K. Sharma, and P. Priyadarshi, editors, *Recent Trends in Aerospace Design and Optimization*, pages 203–210. Tata McGraw-Hill, New Delhi, 2006.

[10] D. P. Ghate and M. B. Giles. Efficient Hessian calculations using automatic differentiation. AIAA-Paper 2007-4059, 18th AIAA Computational Fluid Dynamics Conference, Miami FL, June 2007.

[11] P. A. Gnoffo, R. N. Gupta, and J. L. Shinn. Conservation equations and physical models for hypersonic air flows in thermal and chemical nonequilibrium. Technical report, NASA, February 1989.

[12] Steve R. Gunn. Support vector machines for classification and regression. Technical report, University of Southampton, 1998.

[13] Zhong-Hua Han, Stefan Görtz, and Ralf Zimmermann. On improving efficiency and accuracy of variable-fidelity surrogate modeling in aero-data for loads context. In *CEAS 2009 European Air and Space Conference*, Manchester, UK, October 2009.

[14] B. Hassan. *Thermo-Chemical Nonequilibrium Effects on the Aerothermodynamics of Hypersonic Vehicles*. PhD thesis, North Carolina State University, Albuquerque, NM, December 1993.

[15] J. C. Helton, J. D. Johnson, W. L. Oberkampf, and C. J. Sallaberry. Representation of Analysis Results Involving Aleatory and Epistemic Uncertainty. Technical Report SAND 2008-4379, Sandia National Laboratories, 2008.

[16] S. Hosder and R. W. Walters. Non-intrusive polynomial chaos methods for uncertainty quantification in fluid dynamics. In *48th* AIAA *Aerospace Sciences Meeting*, Orlando, IL, January 2010. AIAA Paper, 2010-129.

[17] Ronald L. Iman. *Latin Hypercube Sampling*. John Wiley & Sons, Ltd, 2008.

[18] S. Jeong and M. Murayama. Efficient Optimization Design Method Using Kriging Model. *Journal of Aircraft*, 42(2):413–420, 2005.

[19] D. R. Jones, M. Schonlau, and W. J. Welch. Efficient Global Optimization of Expensive Black-Box Functions. *Journal of Global Optimization*, 13:455–492, 1998.

[20] W. L. Kleb, R. A. Thompson, and C. O. Johnston. Blurring the inputs: A natural language approach to sensitivity analysis. In *18th* AIAA *Computational Fluid Dynamics Conference*, Miami, FL, June 2007. AIAA Paper, 2007-4206.

[21] J. R. Koehler and A. B. Owen. Computer Experiments. In *Handbook of Statistics*. pp. 261-308, 1996.

[22] V. Kreinovich and S.A. Ferson. A new Cauchy-based black-box technique for uncertainty in risk analysis. In *Reliability Engineering and Systems Safety*, pages 267–279, 2002.

[23] J. Laurenceau and M. Meaux. Comparison of gradient and response surface based optimization frameworks using adjoint method. In *49th* AIAA*/ASME/ASCE/AHS/ASC Structures, Structural Dynamics, and Materials Conference*, Schaumburg, IL, April 2008. AIAA Paper, 2008-1889.

[24] J. Laurenceau and P. Sagaut. Building Efficient Response Surfaces of Aerodynamic Functions with Kriging and Cokriging. AIAA *Journal*, 46(2):498–507, 2008.

[25] M. S. Liou. A sequel to AUSM, Part II: AUSM+-up for all speeds. *Journal of Computational Physics*, 214:137–170, 2006.

[26] B. Lockwood. Gradient-based approaches for sensitivity analysis and uncertainty quantification within hypersonic flows. PhD Thesis, Department of Mechanical Engineering, University of Wyoming, May 2012.

[27] B. Lockwood and D. Mavriplis. Gradient-based methods for uncertainty quantification in hypersonic flows. *Computers & Fluids*, 85:27–38, 2013.

[28] B. A. Lockwood and D. J. Mavriplis. Parameter sensitivity analysis for hypersonic viscous flow using a discrete adjoint approach. In *48th* AIAA *Aerospace Sciences Meeting and Exhibit*, Orlando, FL, January 2010. AIAA Paper, 2010-447.

[29] Brian A. Lockwood and Mihai Anitescu. Gradient-enhanced universal Kriging for uncertainty propagation in nuclear engineering. *Nuclear Science and Engineering*, 170(2):168–195, 2012.

[30] K. Mani and D. J. Mavriplis. An unsteady discrete adjoint formulation for two-dimensional flow problems with deforming meshes. *AIAA Journal*, 46(6):1351–1364, 2008.

[31] J. D. Martin and T. W. Simpson. Use of Kriging Models to Approximate Deterministic Computer Models. AIAA *Journal*, 43(4):853–863,, 2005.

[32] NASA. *FUN3D: Fully Unstructured Navier-Stokes Manual*, May 2009. http://fun3d.larc.nasa.gov/index.html.

[33] J. Nocedal and S. J. Wright. *Numerical Optimization*. Springer-Verlag, New York, 1999.

[34] David R. Olynick. *A New LU-SGS Flow Solver for Calculating Reentry Flows*. PhD thesis, North Carolina State University, 1992.

[35] G. E. Palmer. Uncertainty analysis of CEV LEO and lunar return entries. In *39th* AIAA *Thermophysics Conference*, Miami, FL, June 2007. AIAA Paper, 2007-4253.

[36] C. Park. *Non-equilibrium hypersonic aerothermodynamics*. John Wiley and Sons, New York, NY, 1990.

[37] J. Peter and M. Marcelet. Comparison of Surrogate Models for Turbomachinery Design. *WSEAS Transactions on Fluid Mechanics*, 3(1):10–17, 2008.

[38] M. Pilch, T. G. Trucano, and J. C. Helton. Ideas Underlying Quantification of Margins and Uncertainties (QMU): A white paper. Technical Report SAND2006-5001, Sandia National Laboratories, 2006.

[39] M. M. Putko, P. A. Newmann, A. C. Taylor III, and L. L. Green. Approach for uncertainty propagation and robust design in cfd using sensitivity derivatives. In *15th* AIAA *Computational Fluid Dynamics Conference*, Anaheim,CA, June 2001. AIAA Paper, 2001-2528.

[40] C.E. Rasmussen and C.K.I. Williams. *Gaussian Processes for Machine Learning*. The MIT Press, 2006.

[41] O. Roderick, M. Anitescu, and P. Fischer. Polynomial Regression Approaches Using Derivative Information for Uncertainty Quantification. *Nuclear Science and Engineering*, 164(2):122–139, 2010.

[42] C. J. Roy and W. L. Oberkampf. A complete framework for verification, validation and uncertainty quantification in scientific computing. In *48th* AIAA *Aerospace Sciences Meeting and Exhibit*, Orlando, FL, January 2010. AIAA Paper, 2010-124.

[43] M. Rumpfkeil and D. Mavriplis. Efficient Hessian calculations using automatic differentiation and the adjoint method. *AIAA Journal*, 48(10):2406–2417, 2010. doi:10.2514/1.51239.

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

[45] T. W. Simpson, J. J. Korte, T. M. Mauery, and F. Mistree. Comparison of response surface and Kriging models for multidisciplinary design optimization. In *7th* AIAA*/USAF/NASA/ISSMO Symposium on Multidisciplinary Analysis and Optimization*, 1998. AIAA Paper, 98-4758.

[46] E. Solak, R. Murray-Smith, W. Leithead, C. Rasmussen, and D. Leith. Derivative observations in Gaussian process models of dynamic systems, 2003. http://books.nips.cc/papers/files/nips15/AA70.pdf.

[47] M. J. Wright, D. Bose, and Y.-K. Chen. Probabilistic modeling of aerothermal and thermal protection material response uncertainties. AIAA *Journal*, 45(2):399–425, February 2007.

[48] R. R. Yager and L. Liu. *Classic Works of the Dempster-Shafer Theory of Belief Functions. Studies in Fuzziness and Soft Computing Series*, volume 219. Berlin: Springer, 2008.

[49] W. Yamazaki, M. P. Rumpfkeil, and D. J. Mavriplis. Design optimization utilizing gradient/hessian enhanced surrogate model. In *40th Fluid Dynamics Conference and Exhibit*, Chicago, IL, June 2010. AIAA Paper, 2010-4363.

[50] C. Zhu, R. H. Byrd, P. Lu, and J. Nocedal. L-BFGS-B: A Limited Memory FORTRAN Code for Solving Bound Constrained Optimization Problems. Technical Report NAM-11, Department of Electrical Engineering and Computer Science, Northwestern University, Evanston, Illinois, USA, 1994.