JOURNAL OF AIRCRAFT
Vol. 44, No. 2, March–April 2007

Ranking Based Uncertainty Quantification for a Multifidelity Design Approach

J. Umakant*
Defence Research and Development Laboratory, Hyderabad 500 058, India
K. Sudhakar† and P. M. Mujumdar‡
Indian Institute of Technology, Powai, Mumbai 400 076, India
and
C. Raghavendra Rao§
University of Hyderabad, Hyderabad 500 046, India

DOI: 10.2514/1.22424

Computer simulation based design processes are being extensively used in complex systems like scramjet powered hypersonic vehicles. The computational demands associated with the high-fidelity analysis tools for predicting the system performance restrict the number of simulations that are possible within the design cycle time. Hence, analysis tools of lower fidelity are generally used for design studies. To enable the designer to make better design decisions in such situations, the lower fidelity analysis tool is complemented with an uncertainty model. An approach based on ranks is proposed in this study to aggregate high-fidelity information in a cost effective manner. Based on this information, a cumulative distribution function for the difference between high-fidelity response and low-fidelity response is constructed. The approach is explained initially for uncertainty quantification in a synthetic example. Subsequently an uncertainty model for estimating the mass flow capture of air, a typical disciplinary performance metric in hypersonic vehicle design, is presented.

Nomenclature

H_{cr}	=	cruise altitude
M_∞	=	Mach number
m_a	=	mass flow capture of air
θ_1, θ_2, θ_3	=	forebody compression angles

I. Introduction

THE design of any system involves tradeoff among various options and selecting one that best meets the requirements. For complex systems, like scramjet powered hypersonic vehicles, it is desirable to assess the various options using a high-fidelity analysis (HFA) that accurately evaluates the performance metrics of the system. This requirement is critical because the thrust minus drag margins for such vehicles are typically small. An accurate estimate of the various performance metrics, like mass flow capture of air and drag of the vehicle, is therefore essential for successful realization of the design. A computer based design framework may demand a large number of simulations employing HFA tools like computational fluid dynamics (CFD) and finite element methods (FEM) for this purpose. HFA tools are computationally very intensive. They are typically used to analyze a particular configuration in great detail rather than to evaluate a large number of configurations in the design phase. Thus there is a restriction on the number of HFA simulations that are possible to carry out within the design cycle time. Design processes therefore employ computationally efficient methods known as low-fidelity analysis (LFA) or medium-fidelity analysis that use simplified physics or coarse discretization and are therefore faster but less accurate. The designer is thus confronted with the challenge of making decisions in an environment wherein uncertainty associated with less accurate analysis tools is ever present, and hence there is a need for an uncertainty model that can mitigate the effect of the lack of fidelity in the analysis.

This type of uncertainty that arises due to simplifying the level of fidelity in the analysis is referred to as epistemic uncertainty in the literature. Epistemic uncertainty is defined [1] as “a potential deficiency in any phase or activity of the modeling process that is due to lack of knowledge.” In this paper we use the term “fidelity uncertainty” to mean epistemic uncertainty. This type of uncertainty may also arise when new classes of systems like scramjet powered hypersonic vehicles are developed for the first time. In such situations, there is a scarcity of high-fidelity information. Probabilistic approaches to handle uncertainty associated with low-/medium-fidelity analysis and their application in design scenarios have been demonstrated recently. Quantification of uncertainty using a Bayesian approach to update the uncertainty model was proposed by Mantis [2] in the context of an aerospace vehicle design. DeLaurentis [3] discretized probability density function (PDF) for various confidence levels and created a response surface model to achieve aircraft design that is robust in performance with respect to stability and control disciplines. Charania et al. [4] used engineering methods for various participating disciplines in reusable launch vehicle design, together with a multiplier coefficient that is characterized by an assumed probability distribution. Alternative approaches that are not based on probability theory are also being investigated to characterize epistemic uncertainty. Agarwal et al. [5] presented an approach based on evidence theory to quantify uncertainty in multidisciplinary design optimization for the aircraft sizing problem. However, in this paper we continue to treat epistemic uncertainty within the framework of probability theory. The authors of this paper demonstrated, in [6], a probabilistic design approach for a hypersonic vehicle. Fidelity uncertainty in a disciplinary performance metric, mass flow capture of air, was characterized through a Weibull distribution, using four arbitrarily selected high-fidelity observations, and its effect was propagated onto a system metric, namely, thrust deliverable. A design that

Presented as Paper 2149 at the 8th AIAA Non-Deterministic Approaches Conference, Hyatt Regency Newport, Newport, Rhode Island, 1–4 May 2006; received 13 January 2006; revision received 29 April 2006; accepted for publication 16 June 2006.

*Scientist, Aerodynamics Division, also Research Student at Department of Aerospace Engineering, IIT, Powai.
†Professor and Head, Department of Aerospace Engineering, CASDE.
‡Professor, Department of Aerospace Engineering and Head, CASDE.
§Reader, Department of Mathematics and Statistics.

maximized the system metric was sought through formal optimization. However, there are still several issues that merit further research.

Development of a probabilistic design process involves three major steps, namely:

1) Construct an input uncertainty model, based on the information available.

2) Propagate the effect of disciplinary uncertainty onto system performance metrics.

3) Assess the system performance and make a design decision under uncertainty.

In the studies discussed above, the focus has been on the last two steps. An uncertainty model was assumed either on the basis of a disciplinary expert's recommendation regarding the prediction accuracy of the lower fidelity analysis or on the basis of an evaluation of the analysis tool with respect to similar applications. For example, if the disciplinary expert declared that the lower fidelity analysis is accurate within $\pm 10\%$, then a normal distribution $N(1, 3.3)$, that is, with normalized mean 1 and standard deviation 3.3, is assumed as the uncertainty. However, as discussed above, a challenging situation arises when there is a scarcity of high-fidelity information, especially in the context of the design of a new class of vehicles. The designer is thus faced with a dilemma about the choice of points in the design space where high-fidelity simulations are to be conducted so as to enable characterization of the epistemic uncertainty in the lower fidelity analysis tool in a computationally efficient manner.

The present study seeks to address this dilemma by proposing a method, based on ranks, to aggregate high-fidelity information in a systematic manner. Using this information, an empirical cumulative distribution function (CDF) is constructed to complement the use of the LFA tool for application in a multidisciplinary design environment. It is envisaged that the approach will enable robust design decisions.

II. Rank Transformation Approach

This section describes the rank transformation approach to model uncertainty in the estimation of a function. It is assumed that the function is computationally expensive, restricting its frequent use while searching the design space during optimization. Hence a LFA tool that is computationally efficient is used, albeit with less accuracy. Uncertainty therefore arises in the estimation of the function. The concept of ranks is introduced and a method that exploits this concept for uncertainty modeling is proposed.

A. Concept of Ranks

Ranking refers to the process of ordering a sample (say of size N) with respect to a system performance metric. For minimization problems, the observation with the least value receives the highest rank (rank N) while the observation with the maximum value receives the lowest rank (rank 1). Rank based approaches have been mainly used in agricultural research and property valuation research. Ranking procedures are one approach in multiple decision theory [7] where a simple loss function (zero–one) is used and the risk is an incorrect decision. Dell and Clutter use "judgment ordering" in [8] and enable use of the sample mean as an estimator for the population mean, when it is difficult or expensive to obtain the characteristic of interest for the whole population. Cronan et al. in [9] report that for small sample sizes, the rank regression technique produces a model with better estimates of residential property value as compared to the model based on the multiple regression analysis technique. However to the best of the author's knowledge there has been no study in the area of modeling uncertainty based on the rank transformation approach.

As discussed in the previous section, creating a sample of high-fidelity information constitutes the initial step in the process of modeling uncertainty. Instead of modeling the uncertainty over the entire design space, a method is proposed to sequentially aggregate a high-fidelity sample from the regions where the value of the expensive function is potentially attractive. At a design point, the value of the function is considered attractive if its value is relatively lower as compared to other design points. A formal definition of the concept of an attractive zone is given in the next section. Thus, design points with lower response values receive higher ranks while those with higher response values receive lower ranks. A stopping criterion is used to limit the sample size to a reasonable number. The reasonable number can also be specified as a computational budget. This number is intended to be less than that demanded in a design of experiments to create a response surface model. Based on this information, a probabilistic model for the difference between the high-fidelity response value and the corresponding low-fidelity response value, defined as residual, is constructed. Uncertainty model for the estimation of the expensive function is now defined as the LFA tool complemented with a probabilistic model of the residual. It may be noted that in the context of optimization (minimization), the inaccuracy of the model in the regions where the function is relatively higher is not of much interest. Rank transformation of the response enables one to introduce the preferential characteristics in the uncertainty model.

B. Notations

Let $Z = F(\mathbf{X})$ denote a HFA tool that describes the system behavior. The LFA tool is represented by $z = f(\mathbf{X})$. $\mathbf{X} \in D \subseteq \mathbf{R}^n$ denotes the input vector for both the functions and D is the design space in n dimensional real space $\mathbf{R}^n$. $F(\cdot)$ is the high-fidelity or expensive function and Z is the high-fidelity response. $f(\cdot)$ is the low-fidelity function and z is the low-fidelity response. $\mathbf{X}$ typically describes the parameterization of the system while Z or z describes a performance metric of the system. Then for the rank transformation approach the following notations are adopted:

$D = [\mathbf{a} \quad \mathbf{b}]$: for design space. This defines the space in $\mathbf{R}^n$ over which the function is defined.

$\mathbf{X}_1, \ldots, \mathbf{X}_K, \ldots$,: design points in D.

$Z_1, \ldots, Z_K, \ldots$,: high-fidelity responses corresponding to the design points.

$z_1, \ldots, z_K, \ldots$,: low-fidelity responses corresponding to the design points.

$S_X = \{\mathbf{X}_1, \ldots, \mathbf{X}_N, \mathbf{X}_{N+1}, \ldots, \mathbf{X}_M\}$: set of design points.

$S_R = \{Z_1, \ldots, Z_N, z_{N+1}, \ldots, z_M\}$: set of high-fidelity and low-fidelity responses corresponding to the points in S_X (i.e., there are N high-fidelity responses and M–N low-fidelity responses).

r_i: rank of the ith response in S_R.

P_i: empirical probability density at $\mathbf{X} = \mathbf{X}_i$ in S_X.

C_i: empirical cumulative probability density at $\mathbf{X} = \mathbf{X}_i$ in S_X.

e_i: $Z_i - z_i$ at the point $\mathbf{X} = \mathbf{X}_i$. This denotes the residual at the point X_i.

$X_0 = \mathbf{a} < \mathbf{X}_1 < \cdots < \mathbf{X}_{p-1} < \mathbf{X}_p = \mathbf{b}$: partition of D with size p. This divides the design space D into p levels.

$f_L = z_0 < z_1 < \cdots < z_{p-1} < z_p = f_U$: partition of $f(D)$ with size p. This divides the range of LFA response into p levels.

$f(D) = [f_L \quad f_U]$: low-fidelity response space for D.

$S_L = \{(\mathbf{X}, z) \mid \mathbf{X} \in D, z = f(\mathbf{X})\}$: low-fidelity system state descriptor. This defines a set of design points where the system behavior is described using its low-fidelity response, as modeled by $f(\mathbf{X})$.

$S_H = \{(\mathbf{X}, Z) \mid \mathbf{X} \in D, Z = F(\mathbf{X})\}$: high-fidelity system state descriptor. This defines a set of design points where the system behavior is described using its high-fidelity response, as modeled by $F(\mathbf{X})$.

$S_A = \{(\mathbf{X}, Z) \mid (\mathbf{X}, Z) \in S_H \text{ and } Z \leq z_1\}$: system state descriptor in the attractive zone. The attractive zone is the set of high-fidelity responses belonging to S_H that are less than a specified threshold response. In this study, the specified threshold response is the low-fidelity response z_1 corresponding to the first partition level of the low-fidelity response space.

n_A: cardinality of S_A indicating the number of observations of the sample in the attractive zone.

The expensive function is estimated using the model, $Z \sim z + U(Z)$, where $U(Z)$ is a probability distribution of the residual and represents the epistemic or fidelity uncertainty in the estimation of Z when a LFA tool is used instead of the HFA tool.

Table 1 Univariate bimodal function

	Function	Design space	Attractive zone
High fidelity	$Z = (X - 0.5)(X - 2)(X - 4)(X - 3.25) + 10$	$D = [0\ 5]$	$Z \leq 12.8$
Low fidelity	$z = 1.8785X^2 - 7.8888X + 15.7313$	$D = [0\ 5]$	

C. Rank Transformation Approach

The various steps involved in the rank transformation approach are now discussed. For ease of explanation, a synthetic example is used to illustrate the various steps. The HFA is a univariate bimodal function and is assumed to be computationally expensive. The LFA is a quadratic unimodal function. This is assumed to be computationally efficient and may be used a large number of times in simulation. The functions are defined in Table 1.

1. Initial Sample

Assume that to begin with there are K ($K \geq 2$) high-fidelity responses or observations available. A minimum of two responses is required because we need to order the responses. However, in this study, we have chosen $K = 3$. This enables a smoother transition of the responses between the attractive and nonattractive zones. Typical data consisting of three initial points are shown in Table 2.

2. Augmentation of the Sample with Low-Fidelity Responses

LFAs are now performed at uniformly gridded points in the design space to augment the sample. Augmenting is done to encourage global representation of the range of the function and the design space, during ranking. In case of multimodal functions this also helps to avoid aggregating high-fidelity information that is restricted to a local valley. However, caution must be exercised to ensure that the number of low-fidelity responses in the sample is not significantly higher as compared to the number of high-fidelity responses. Otherwise the trend of the function behavior as predicted from the sample will be dominated by the LFA tool. Table 3 shows the data after the addition of six low-fidelity observations. It may be argued that low-fidelity responses may be added sequentially. However, adding one low-fidelity response at a time may not encourage global representation in the sample, especially in the initial stages when the available number of responses is small.

3. Rank Transformation of Responses

The high-fidelity and low-fidelity responses are combined to form the dataset S_R and the elements of the set are sorted in a descending order. The responses are now transformed by assigning ranks to them in a serially increasing manner, starting from 1 to the value of the maximum rank. The maximum rank has a value equal to M, the total number of high-fidelity and low-fidelity responses. Thus the response with the minimum value receives the maximum rank while the response with maximum value receives a rank of 1. If a response value occurs more than once, then the same rank is assigned to all its occurrences. A typical rank transformation of the responses is shown in the fourth column of Table 4.

Table 2 Typical sample of high-fidelity responses

S. no.	X	Z
1	0.35	12.61
2	2.86	10.90
3	4.64	19.72

Table 3 Typical augmented data

S. no.	S_X	S_R
1	0.35	12.61
2	2.86	10.90
3	4.64	19.72
4	0.00	15.73
5	1.00	9.72
6	2.00	7.46
7	3.00	8.97
8	4.00	14.23
9	5.00	23.25

4. Mapping of the Ranks onto Design Space

The ranks for the responses are now mapped to their corresponding design points in S_X. Thus the design points are now given an additional attribute, namely, the rank that defines its preference for selection.

5. Computation of the Empirical PDF and CDF of the Design Space

The design points, together with their associated ranks, are now sorted in an ascending order with respect to their respective values. The empirical probability density of the ith point is then computed as the ratio of its rank to the summation of the ranks of all the points, and is given as

$$P_i = \frac{r_i}{\sum_{i=1}^{M} r_i} \tag{1}$$

Because the points are sorted in ascending order, the empirical cumulative density of the ith point in the table is the ratio of the sum of ranks of all the points above it to the total summation of ranks and is denoted as

$$C_i = \frac{\sum_{j=1}^{i} r_j - 0.5}{\sum_{i=1}^{M} r_i} \tag{2}$$

The factor 0.5 in Eq. (2) is used for a continuity correction to connect the first point and last point in the sample, respectively, with the lower and upper bounds of the design space. Piecewise linear interpolation gives the cumulative density at any other point in the design space. Empirical PDF and CDF for the design space are thus constructed. The fifth and sixth columns of Table 4 depict the probability density and cumulative density values of the design points. The lower and upper bounds of the design space will now, respectively, take 0 and 1 as their empirical CDF values.

6. Selection of a New Point for HFA Evaluation

The CDF of the design space is now employed to sample a new design point where the expensive function is evaluated. It is desired to select the new design point such that the function value is lower than any of the high-fidelity responses available currently in the sample.

Table 4 Typical data showing rank transformation of responses and cumulative probabilities of design points

S. no.	S_X	S_R	Rank	Probability density	Cumulative probability
4	0.00	15.73	3	0.06	0.05
1	0.35	12.62	5	0.11	0.16
5	1.00	9.72	7	0.15	0.32
6	2.00	7.46	9	0.20	0.52
7	2.86	10.90	6	0.13	0.65
2	3.00	8.97	8	0.17	0.83
8	4.00	14.23	4	0.08	0.92
3	4.64	19.72	2	0.04	0.96
9	5.00	23.25	1	0.02	0.98

Table 5 Typical high-fidelity sample after selecting one point from the CDF of design space

S. no.	X	Z
1	0.35	12.61
2	2.86	10.90
3	4.64	19.72
4	0.87	6.86

To sample from the distribution, first a random number, u is generated from uniform distribution $U[0, 1]$. Note that the CDF value also varies in a monotonic manner from 0 to 1. Performing piecewise linear interpolation on the data, given in Table 4, for the cumulative distribution value u yields the new design point. This procedure of generating random variates is referred to as the inverse-transform method [10]. Because the design points with a lower function value have higher ranks and therefore a higher probability value, the CDF of the design space favors the selection of a new point from the regions where the function value is relatively lower. However, since the distribution value u is chosen randomly, there is also a chance, with low probability, that the new design point has a higher function value. Thus there are now $K + 1$ high-fidelity observations. An updated table is shown at Table 5.

Steps 2, 3, and 4, that is, augmenting the sample with low-fidelity observations, ranking the collection of the responses, and mapping the responses onto the design space are repeated to update the CDF of the design space. The updated distribution is used for selection of the next design point for high-fidelity evaluation. This process is repeated till the desired number of high-fidelity responses is aggregated. Alternatively a heuristic stopping criteria described in the next section can be used.

7. *Stopping Criterion*

A heuristic criterion is used to decide when to stop the process of collecting high-fidelity information. The CDF for the residual is examined after every update and if the variation between the successive trials is negligible, the process is terminated.

Let $F_i(e)$ and $F_{i+1}(e)$ be the CDFs of residual e for the ith and $(i + 1)$th trials or updates, respectively. The variation in successive CDFs is expressed as follows:

$$d = \max_X |F_{i+1}(e) - F_i(e)| \quad (3)$$

The variation is considered negligible when $d \leq \varepsilon$, where ε is a small number compared to 1. Thus, there are $N + K$ high-fidelity observations when the process is terminated after N trials are performed.

D. Limitations of the Rank Transformation and Means to Circumvent Them

The rank transformation procedure described in the previous section has the following limitations:

1) The procedure may be sensitive to the initial sample and consequently influence the selection of subsequent high-fidelity responses. For example, if the initial design points are such that their response values are not significantly different from each other, the new point selected would also have greater probability of having a response value in the same range. However, it is desired that initially the entire range of the LFA response is represented.

2) Because the inverse transformation for selecting a new design point is random, there is a possibility of a clustering of the design points. This depends on the nature of the CDF of the design space. If the successive values of u are close, the corresponding random variates obtained from inverse transformation may yield design points that are within a close neighborhood. It is desired to avoid points that are spatially close to each other.

The first limitation can be largely circumvented by starting with an initial sample that spans not only the entire design space but also the range of the response. A strategy based on stratification is adopted to implement this. The second limitation is addressed by defining a minimum distance criterion.

1. *Stratification Algorithm*

Stratification refers to partitioning or coding the range of the design space D and the range of the response space $f(D)$ into K number of levels or strata with the following properties:

$X_s = i$, iff $X_{i-1} < X \leq X_i$, and

$z_s = i$, iff $z_{i-1} < z \leq z_i$; $i = 1, 2, \ldots, K$.

A typical system state descriptor $S_L = (X, z)$ is coded as (X_s, z_s).

Each level can be interpreted as an isocontour. Our purpose is to choose the initial points such that all the strata in X and z are represented, at least once. Thus, if K points are chosen, then, each point should uniquely represent a strata in X and z. This requirement can be formulated as an assignment problem (or an integer programming problem), in Boolean space. However, in the present study this is implemented as described below. For the synthetic example, the process can be initiated with two initial observations.

The entire design space is uniformly gridded into a large number of points and a LFA tool is used for evaluating the response at each of the grid points. Typical representation for 15 grid points is shown in Table 6a). The grid points and the corresponding response values are then coded or stratified according to the definition above. A stratification table of coded values is thus set up. Table 6b) illustrates this process for three stratification levels. A consequence of the stratification is that the design points falling in the same strata lose their distinct identity, resulting in duplicate points that do not possess any additional information. For further analysis, such points are not retained in the stratification table. Table 6c) highlights the data after deletion of such points. A frequency table indicating the number of occurrences of X_s and z_s is defined and a typical result is shown in Table 6d). Strata level corresponding to the maximum number of occurrences either in X_s or z_s is identified and a point from the set of coded values corresponding to this strata level is selected. Any tie that occurs is resolved randomly. Other points having the same stratification identity either in support space or in the response space are removed. For example, Table 6d) shows that the maximum number of occurrences is three for both z_s and X_s. We choose to select the point at serial no. 6 in Table 6c). Subsequently all the other entries in Table 6c) having strata level three either in X_s or z_s are deleted from the table. An updated stratification table is illustrated in Table 6e). The process of defining the frequency table and selecting a point corresponding to the maximum frequency is repeated for the specified number of strata levels. This results in K distinct coded design points whose corresponding response codes are also distinct. An illustration of the updated table after selecting the second point is shown in Table 6f). Table 6g) summarizes the selected strata levels for the design space points and the corresponding coded response levels. The selected stratified levels are then mapped back to the physical domain. For each strata level, there exist multiple design points that satisfy the mapping. The points may be chosen randomly. At these initial points the expensive function evaluations are carried out. For example, the entry at serial no. 3 in Table 6g) corresponds to a design point in strata level 2 with its response in strata level 1. In the physical domain, the entries at serial nos. 6–10 of Table 6a) correspond to these strata levels and any one of these points is randomly selected.

Thus an initial sample of size K is aggregated through stratification.

2. *Minimum Distance Criterion*

Inverse transformation of the cumulative distribution function to select the new design points can sometimes result in the new point to lie within the neighborhood of existing points. To avoid this, the L_1 norm of the new point with respect to the existing points is calculated. HFA is performed only if the norm is greater than a specified tolerance bound; otherwise another point is selected from the distribution. L_1 is defined as the absolute difference between the coordinates of two design points in S_X. The tolerance bound is given as a fraction of the range of the design variables. In an engineering

Table 6 Typical data for stratification

S. no.	z	X	z_s	X_s
a) Physical domain data				
1	15.73	0.00		
2	13.15	0.35		
3	11.05	0.71		
4	9.43	1.07		
5	8.29	1.42		
6	7.63	1.78		
7	7.45	2.14		
8	7.75	2.50		
9	8.52	2.85		
10	9.78	3.21		
11	11.51	3.57		
12	13.73	3.92		
13	16.42	4.28		
14	19.59	4.64		
15	23.25	5.00		
b) Coded data				
1			2	1
2			2	1
3			1	1
4			1	1
5			1	1
6			1	2
7			1	2
8			1	2
9			1	2
10			1	2
11			1	3
12			2	3
13			2	3
14			3	3
15			3	3
c) Coded data after deleting duplicate points				
1			1	1
2			1	2
3			1	3
4			2	1
5			2	3
6			**3**	**3**
d) Frequency of coded data				
Level 1			3	2
Level 2			2	1
Level 3			1	3
e) Coded data after selecting first point				
1			1	1
2			1	2
3			**2**	**1**
f) Coded data after selecting second point				
1			**1**	**2**
g) Summary of coded data selected for three levels				
1			**3**	**3**
2			**2**	**1**
3			**1**	**2**

design environment, the tolerance bound may be set based on a designer's intuition on the sensitivity of the function value with respect to the design variables. An approximation of this measure may be obtained using the LFA tools.

E. Uncertainty Modeling

The first step of aggregating a sample of high-fidelity responses is now complete. Residual, denoted by e, is estimated by taking the difference between the high-fidelity response Z and the corresponding low-fidelity response, z, that is, $e = Z - z$. Based on this information, a CDF for the residual is constructed. This distribution characterizes the epistemic uncertainty or fidelity uncertainty.

Diagnosis of the trajectory of the residues with respect to the low-fidelity response enables one to infer whether or not the residues are correlated with low-fidelity response. In case the trajectory exhibits a random path, it is inferred that the differences are random. On the other hand, a systematic variation in the trajectory exemplifies a correlation between the low-fidelity and the high-fidelity responses.

1. CDF for the Residual

The residuals for the $N + K$ high-fidelity responses are sorted in ascending order with respect to their values. Homogenous distribution, for the residual, is assumed for computing the probability distribution $U(Z)$. The cumulative density for the ith residual is defined as

$$C_i(e) = \frac{(i - 0.5)}{N + K}, \qquad i = 1, 2, \ldots, N + K \tag{4}$$

The lower bound of the residues is defined as the minimum value of the observed residues decremented by one unit. Similarly the upper bound of residue is defined as the maximum value of the observed residue incremented by one unit. Alternatively, the slope based on the first two residual values can be extrapolated to yield the lower bound. Similarly, the slope based on the last two residuals is extrapolated to yield the upper bound. Cumulative densities of 0 and 1 are assigned, respectively, to the lower and upper bounds. A fit for this data will give a smooth representation of empirical CDF for $U(Z)$. This type of distribution is referred to as a nonparametric distribution.

The CDF is the characterization of fidelity uncertainty and the HFA value for any design point is predicted as $\hat{Z} = z + U(Z)$, where $\hat{Z}$ is the predicted value for Z. Residuals may be sampled from the distribution $U(Z)$ and propagated through another analysis to characterize uncertainty in system level metrics. Also residual, $e_{0.95}$, for $u = 0.95$ helps to predict with 95% confidence that $Z < z + e_{0.95}$.

2. Restricted CDF of Residue for the Attractive Zone

It may be recalled that using the LFA as a guiding tool, the range of the response was stratified into three contour levels. The zone defined by the first contour level is designated as the attractive zone, because the response value corresponding to the first level is smaller as compared to that for the second and third levels. In the context of minimization, we are interested only in those high-fidelity observations that are contained in the first contour level. Hence, the $N + K$ responses are screened to retain only those responses that are within the attractive zone. The cumulative distribution function of the residual for the attractive zone is constructed in a similar manner as described in the previous section. It may be noted that the uncertainty bounds of the residual will be typically lower than that obtained in the previous section and hence the estimate of the expensive function is not unnecessarily conservative. This is consistent with the philosophy to create an uncertainty model that is more appropriate in the regions of interest, rather than trying to characterize the uncertainty for the entire design space.

F. Verification of Algorithm

In the above procedure, the new design point for high-fidelity estimation has been chosen randomly using inverse transformation of the cumulative distribution function of the design space. Hence the repeatability or robustness of the algorithm needs to be verified by conducting a large number of simulations and examining the results. Verification is based on the following criteria:

1) The probability of the number of observations in the attractive zone should be nontrivial with at least 95% confidence.

2) The empirical cumulative distribution of the residual is examined to check if the same type of distribution is obtained in most of the simulations. A criterion similar to the one defined in Eq. (3) above may be used to check the similarity of distribution. The metric d is computed for all pairs of CDFs obtained through the Monte Carlo simulation and is given as

$$d_{ij} = |F_i(e) - F_j(e)|; \qquad i = 1, 2, \ldots, ns; \qquad j = 1, 2, \ldots, ns \tag{5}$$

where ns is the number of simulations and

$$d = \max(d_{ij}) \tag{6}$$

If $d \leq \varepsilon$, where ε is a small number compared to 1, then the CDFs may be considered to be similar.

3) The coefficient of variation [11] cv, defined as the percentage ratio of standard deviation to mean, is computed for the lower and upper bounds of the residual (when the mean is not 0) from the CDFs generated in Monte Carlo simulation. The algorithm can be considered to be stable if the coefficient of variation is less than 33%. The value 33% is a heuristic criteria suggested by a statistical expert.

III. Uncertainty Quantification for Synthetic Example

Results illustrating the performance of the above algorithm are presented initially for a univariate bimodal function defined in the previous section. Figure 1 shows the high-fidelity and low-fidelity function contours. It may be noted that the function has a global minima at $X = 1.05$ and another local minima at $X = 3.75$. It is desired to construct an uncertainty model that can complement the LFA tool. The uncertainty model is expected to enable robust prediction bounds on the function value at the design points where the HFA values are lower. Robust prediction here means that, if an actual HFA is conducted at the point, the probability of function value Z being lower than the predicted upper bound is high.

A LFA tool is used to obtain a low-fidelity system state descriptor S_L at 15 uniformly distributed design points in the design space D. S_L is stratified into three levels and three design points, one from each level, are selected based on the stratification algorithm discussed previously. The range of the variation of LFA response is about 16 with a minimum value of about 7.5. The first stratification level is then considered as the minimum value plus $1/3$ of the range of the LFA response. An initial sample of three high-fidelity observations is thus obtained. The threshold response z_1 for defining the attractive zone corresponds to the first stratification level and its value is defined in Table 1. Subsequently, 10 high-fidelity simulations were carried out at the design points suggested by the rank transformation approach. During aggregation of the sample, the number of LFA responses is twice the number of HFA responses available in the sample at that stage. The symbols depicted in Fig. 1 represent the selected design points. The initial points selected by stratification are distinguished by the symbol *. It can be observed that though the LFA tool is unimodal, the algorithm has selected design points near both the modes of the high-fidelity function and three points out of a sample size of 13 are in the attractive zone.

The CDF for the design space is shown in Fig. 2. It can be noticed that when X is between 0 and 1, the probability of selecting a point from this distribution is 0.2. However, when X is between 1 and 2, the probability of selecting a point is about 0.4. Recall that the function has a global minima at $X = 1.05$. It can therefore be inferred from the result that CDF of the design space favors selection of the design points from the regions where the function is attractive. Similarly, the probability of selecting a point from the region X greater than 4 is less than 0.02. Referring to Fig. 1, it can be noticed that for X greater than 4 the value of the HFA function Z is higher as compared to the region when X is between 1 and 4. This demonstrates that the algorithm has enabled the construction of the distribution function such that design points with high function value have a low probability of getting selected.

The CDF of the residual based on the total sample of high-fidelity information is depicted in Fig. 3. The lower and upper bounds of the residual from the CDF are used to predict the function F, as shown in Fig. 4. It can be seen that the uncertainty has been characterized over most of the design space. Using the predicted upper bound values in an optimization exercise, where the function F is to be minimized, will always ensure that the HFA value is lower than the predicted value. It may be noted that the motivation of the exercise is to make a robust prediction of F, using few HFA simulations, and not actually try to find the minimum of F. Figure 5 shows the restricted CDF of the residual based on restricting the sample to the HFA observations that are within the attractive zone. The lower and upper bounds from the restricted CDF are now used to predict the function F. It can be seen from Fig. 6 the uncertainty in the estimation of the function is less conservative, as compared to the result shown in Fig. 5, near the two valleys of the function.

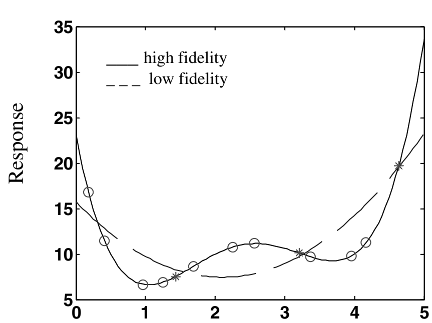


Fig. 1 High-fidelity and low-fidelity contours with selected design points.

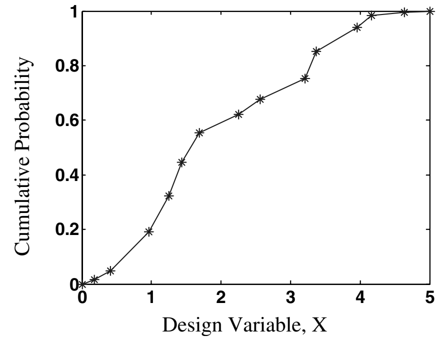


Fig. 2 Cumulative distribution function of design space.

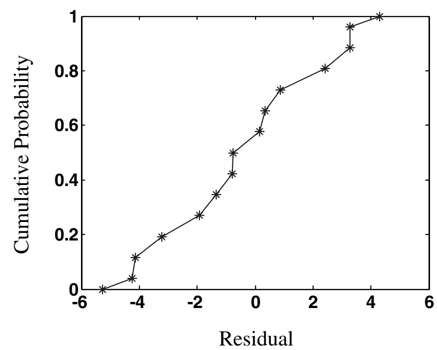


Fig. 3 Cumulative distribution function of residual.

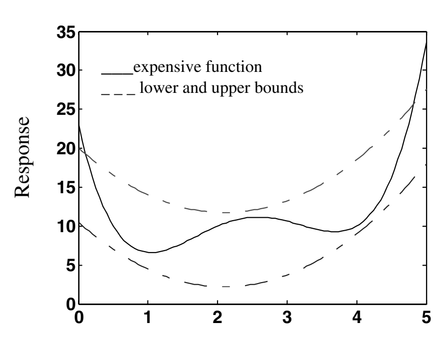


Fig. 4 Uncertainty in estimation of expensive function.

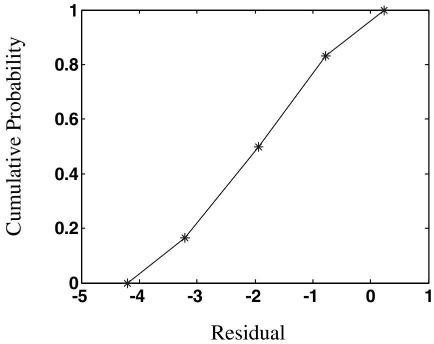


Fig. 5 Restricted cumulative distribution function of residual.

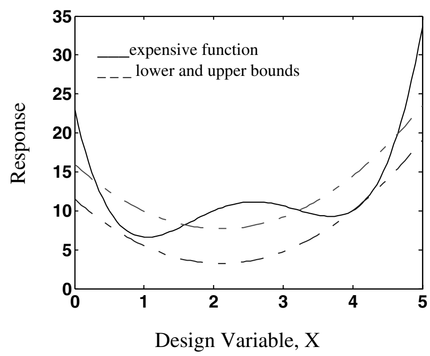


Fig. 6 Uncertainty in estimation of expensive function using restricted CDF.

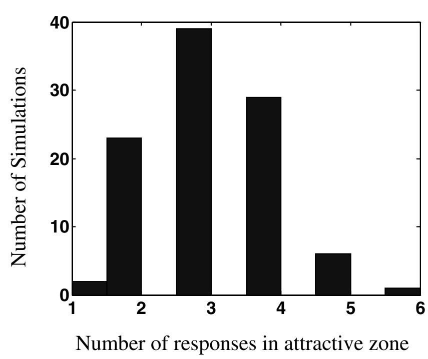


Fig. 7 Histogram of number of hits in the attractive zone.

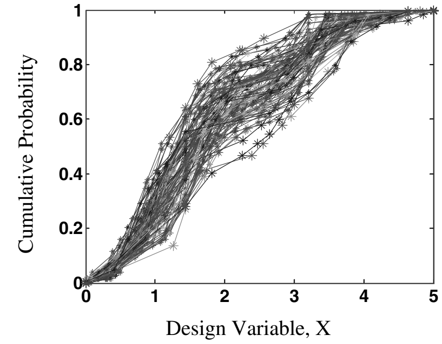


Fig. 8 Cumulative distribution function of design space for various simulations.

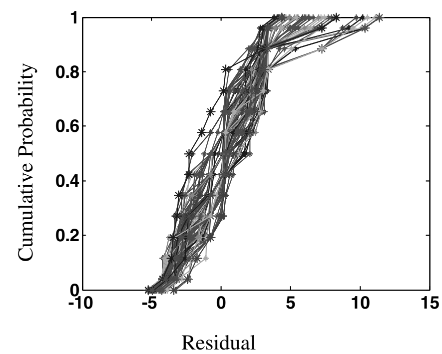


Fig. 9 Cumulative distribution function of residual for various simulations.

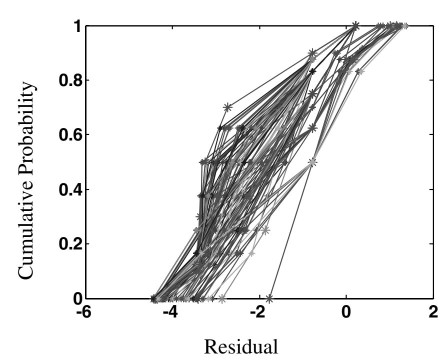


Fig. 10 Restricted cumulative distribution function of residual for various simulations.

Results of Monte Carlo simulations are shown in Figs. 7–10. One hundred simulations were performed and in each simulation the rank transformation approach was applied to yield a sample of 13 HFA observations. From Fig. 7, we can observe that only three simulations show that there are no points in the attractive zone. In other words, 97 simulations out of the total 100 simulations performed record at least one sample in the attractive region. Hence, the nontrivial probability of the number of observations in the attractive zone is $1/13$ with 97% confidence. Figures 8 and 9 show that the CDF for the design space and residual exhibit similar characteristics in most of the simulations. The coefficient of variation for the bounds is tabulated in Table 7. It can be noticed that the lower bounds of the residual have a coefficient of variation less than 33%. However, the upper bound of the restricted CDF, in Fig. 10, is significantly greater than 33%. Increasing the value of the threshold enables more numbers of the aggregated responses to fall within the attractive zone. This helps to reduce the coefficient of variation in such situations.

The effect of sample size on the CDF for a residual is shown in Fig. 11. Trial 1 refers to the sample with the initial three points selected based on the stratification algorithm. From the results it can be observed that the CDF of the residual exhibits negligible variation for trial numbers more than 10. For example, in trial no. 5, the value

Table 7 Metrics from Monte Carlo simulations of univariate bimodal function

	Residual		Restricted residual	
	Lower bound	Upper bound	Lower bound	Upper bound
Mean	−3.70	5.70	−3.11	−0.59
Standard deviation	0.49	2.35	0.50	0.36
Coefficient of variation, %	13.20	41.20	16.00	60.00

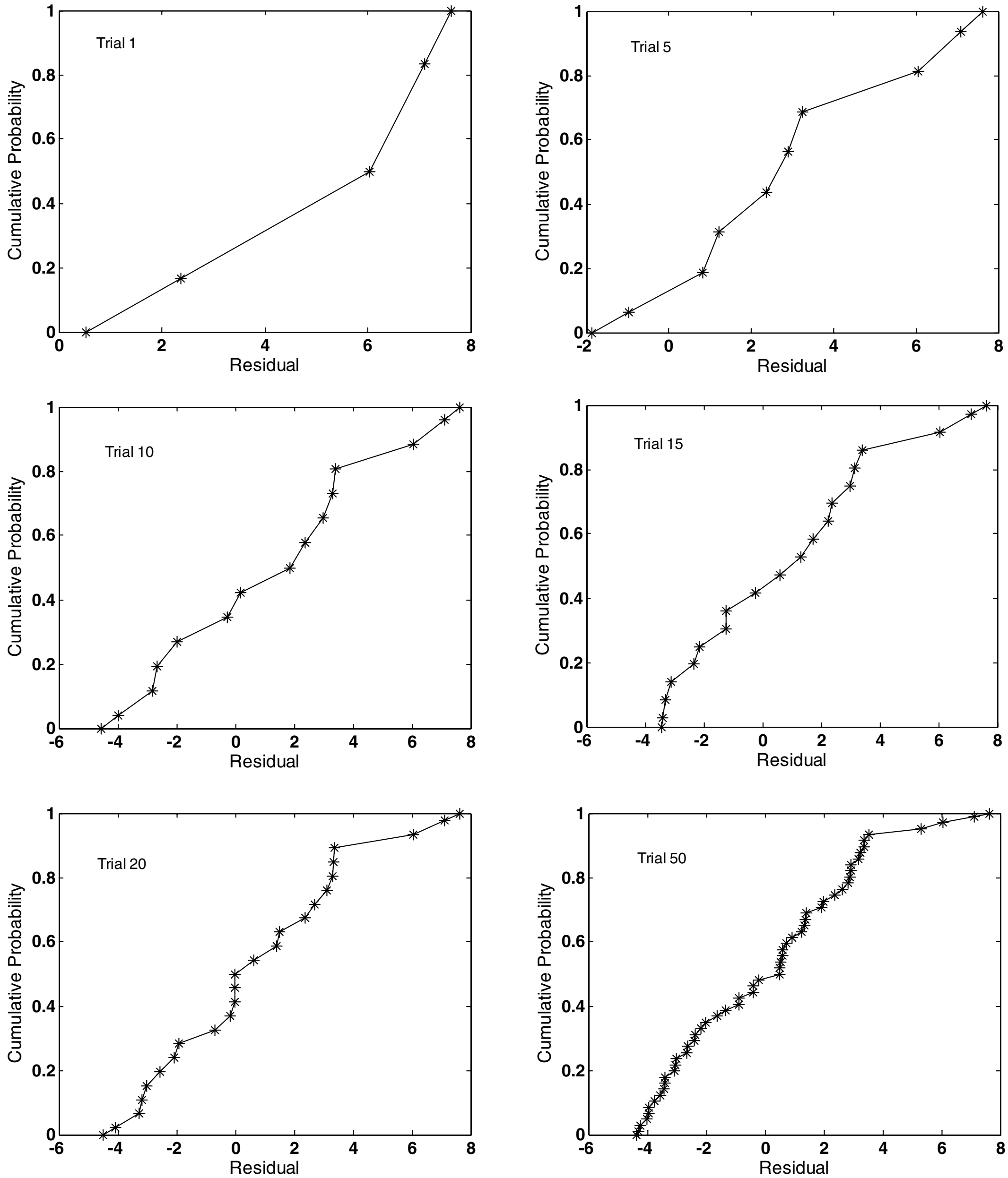


Fig. 11 History of cumulative distribution function of residual.

of CDF is 0.81 when the residual is 6. In trial no. 10, for the same value of residual, the corresponding value of CDF is 0.89, while in trial no. 15, it is 0.91. Thus, when the residual is 6, the variation in the value of CDF is about 0.08 between trial nos. 5 and 10, and the variation reduces to about 0.02 between trial nos. 10 and 15. Hence we may limit the size of the sample to 10.

In the next section the method is applied to quantify uncertainty for a typical disciplinary metric in hypersonic vehicle design.

IV. Uncertainty Quantification of a Typical Disciplinary Metric in Hypersonic Vehicle Design

A scramjet powered hypersonic vehicle typically exhibits a highly integrated airframe engine. A generic representation of such a vehicle is shown in Fig. 12. Mass flow capture of air is a critical disciplinary performance metric needed as input in the design of the intake. The design of the forebody largely dictates the mass flow capture of air of the configuration and the intake entry conditions. At the same time the forebody has a strong influence on the body aerodynamics and also affects the sizing of the vehicle. Because the flow past the forebody is dominated by the presence of strong shocks and viscous phenomena, a CFD code is ideally needed for calculating the disciplinary metrics. The computational demands of such codes allow only a few numbers of high-fidelity simulations in the design environment. Hence an LFA tool is used. However the fidelity uncertainty in the estimation of the performance metric affects the entire downstream design of the propulsion flowpath. Thus, it is important to quantify the uncertainty in the estimation of mass flow capture of air to enable a robust prediction. In the context of this

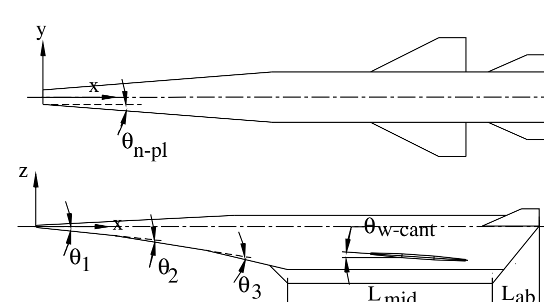


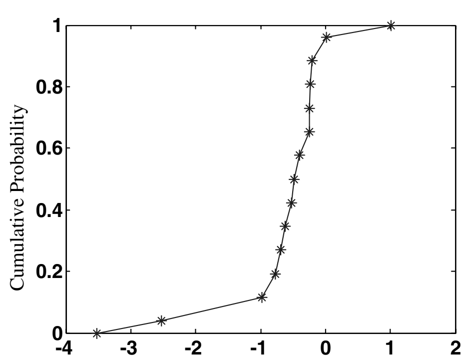
Fig. 12 Generic hypersonic vehicle configuration.

example, robust prediction would mean that the HFA value should always be higher than the predicted lower bound.

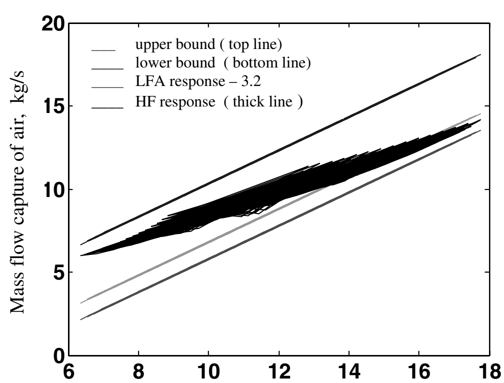
In the present study, the forebody is assumed to have three compression ramps with a body width of 0.8 m. The design variables are the three compression angles and each of these angles is allowed to vary between 0 and 6 deg. The rank transformation approach is thus now applied in a three-dimensional space. The intake entry dimensions are 0.240 m × 0.500 m. The freestream conditions are as follows: $M_\infty = 6.5$, $H_{\text{cruise}} = 32.5$ km, and $\alpha = 4$ deg. The rank transformation approach described in the previous section is used to construct the uncertainty model. Because the thrust deliverable is directly proportional to the mass flow capture of air, it is desired to select design points that maximize this metric. The high-fidelity analysis tool is an inviscid CFD based model whereas the low-fidelity analysis tool is based on oblique shock theory.

Stratification is carried out for three levels and an initial sample of three design points is selected. The LFA value of function corresponding to the first level is 10.8, and this value is set as the threshold for defining the attractive zone. In this application, it was observed that the HFA values are always lower as compared to the LFA values. If the bias is large, then there may not be any points in the attractive zone. To avoid this, a constant is included in the LFA tool. This constant is defined as $\bar{Z} - \bar{z}$, where $\bar{Z}$ denotes the mean of the initial HFA sample obtained by stratification and $\bar{z}$ denotes the mean of the corresponding LFA values.

Because visualizing the CDF in three-dimensional space is difficult, results pertaining to design space are not presented. Figure 13a shows the CDF of the residual. Using the lower and upper bounds of the CDF defined in Fig. 13a, the disciplinary metric is computed for the complete design space. For a given value of mass flow capture of air based on the LFA tool, the corresponding tolerance bounds are shown in Fig. 13b. It can be observed from the simulation results that the HFA values are always higher than the predicted lower bounds thereby enabling robust predictions. For system studies, the uncertainty model can be used to propagate the disciplinary uncertainty onto a system level metric and enable robust design decisions.

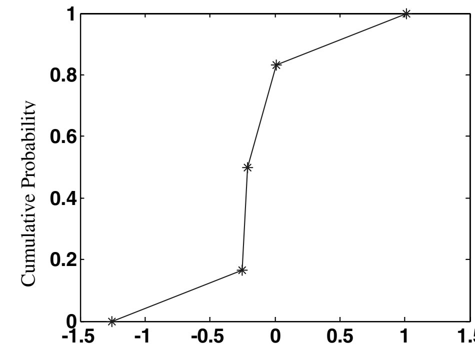


a) Residual, kg/s

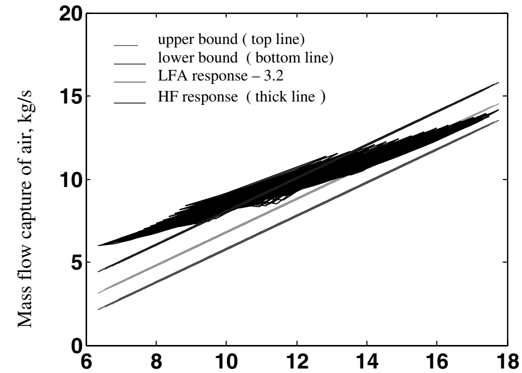


b) LFA value of Mass flow capture of air, kg/s

Fig. 13 Mass flow capture of air a); CDF of residual b) uncertainty in estimation.



a) Residual, kg/s



b) LFA value of Mass flow capture of air, kg/s

Fig. 14 Mass flow capture of air a); restricted CDF of residual b) uncertainty in the estimation.

The refinement in the CDF based on the attractive zone is shown in Fig. 14a. The corresponding uncertainty bounds in the estimation of m_a are shown in Fig. 14b. It can be observed that the width of the tolerance bounds is now less conservative in the regions where the response values of m_a are higher.

It may be noted that for a design problem in three dimensions, 10 high-fidelity simulations at the design points suggested by the rank transformation approach together with an initial sample of three observations having been used to develop the uncertainty model. Developing a surrogate model for the same performance metric required 32 high-fidelity simulations as described in [6]. Thus, for the purpose of making robust predictions in the design phase, the suggested approach requires fewer HFA observations. However, it may be noted if higher design cycle time and higher computational budget are permissible, the surrogate model described in [6] offers better prediction capabilities.

V. Conclusions

A rank based approach has been developed to enable quantification of uncertainty in disciplinary performance metrics when a low-fidelity analysis tool is used instead of a computationally expensive analysis tool. Rather than relying on subjective opinions regarding the accuracy of the low-fidelity analysis tools, the proposed method seeks to aggregate a limited number of high-fidelity information in a sequential manner. Based on this information, an empirical cumulative distribution function for the residual is constructed. Results for a synthetic example demonstrate the validity of the method. The approach has been subsequently used to develop a cumulative distribution function to represent the uncertainty in the estimation of a typical disciplinary metric in the design of a hypersonic vehicle.

The method can be extended to develop similar uncertainty models for other performance metrics. The models can be incorporated in the system synthesis design tool and used in a multidisciplinary design optimization environment for a hypersonic vehicle to enable robust design decisions.

Acknowledgments

J. Umakant would like to thank S. Panneerselvam, Technology Director, Hypersonics and Advanced Technology, Defence Research and Development Laboratory (DRDL), and M. Nagarathinam, Technology Director, Aerodynamics, DRDL, for their continuous encouragement during the course of this work.

References

[1] Oberkampf, W. L., *Bibliography for Verification and Validation in Computational Simulation*, SAND 98-2041, 1998.

[2] Mantis, G. C., "Quantification and Propagation of Disciplinary Uncertainty via Bayesian Statistics," Ph.D. Thesis, Georgia Institute of Technology, Atlanta, GA, 2002.

[3] DeLaurentis, D. A., "A Probabilistic Approach to Aircraft Design Emphasizing Stability and Control Uncertainties," Ph.D. Thesis, Georgia Institute of Technology, Atlanta, GA, 1998.

[4] Charania, A. C., Bradford, J. E., Olds, J. R., and Graham, M., "System Level Uncertainty Assessment for Collaborative RLV Design," *2002 JANNNAF 38th Combustion Subcommittee/26th Airbreathing Propulsion Subcommittee/20th Propulsion Systems Hazards Subcommittee/2nd Modeling and Simulation Subcommittee Joint Meeting*, JANNAF-2002-2B-4-MSS, 2002.

[5] Agarwal, H., Renaud, J. E., Preston, E. L., and Padmanabhan, D., "Uncertainty Quantification Using Evidence Theory in Multidisciplinary Design Optimization," *Reliability Engineering and System Safety* Vol. 85, Nos. 1–3, July–Sep. 2004, pp. 281–294.

[6] Umakant, J., Sudhakar, K., Mujumdar, P.M ., and Panneerselvam, S., "Configuration Design of Air-Breathing Hypersonic Vehicle Considering Fidelity Uncertainty," AIAA Paper 2004-4543, 2004.

[7] Gupta, S. S., and Huang, D. Y., *Multiple Statistical Decision Theory*, Springer–Verlag, New York, 1981.

[8] Dell, T. R., and Clutter, J. L., *Biometrics*, Vol. 28, No. 2, June 1972, pp. 545–555.

[9] Cronan, T. P., Epley, D. R., and Perry, L. G., "The Use of Rank Transformation and Multiple Regression Analysis in Estimating Residential Property Values with a Small Sample," *Journal of Real Estate Research*, Vol. 1, No. 1, 1986, pp. 19–31.

[10] Law, A. M., and Kelton, D. W., *Simulation Modeling and Analysis*, McGraw–Hill International Series, McGraw–Hill, New York, 2000.

[11] Agarwal, B. L., *Basic Statistics*, 3rd ed., New Age International (P) Limited Publishers, New Delhi, India, 1996.