Structures, Structural Dynamics, and Materials and Co-located Conferences
April 8-11, 2013, Boston, Massachusetts
54th AIAA/ASME/ASCE/AHS/ASC Structures, Structural Dynamics, and Materials Conference

AIAA 2013-1813

Inclusion of Data Uncertainty and Model Error in Multi-disciplinary Analysis and Optimization

Chen Liang[1] and Sankaran Mahadevan[2]
Vanderbilt University, Nashville, TN,37211

This paper combines efficient uncertainty propagation, surrogate modeling, and sampling techniques for uncertainty quantification and design under uncertainty of aerospace structures. A new method is developed for including data uncertainty and model error quantification in multi-disciplinary system analysis (MDA) and optimization that usually requires iterative analyses with a large number of coupling variables. The methodology for MDA estimates the probability distributions of the coupling variables based on computing the probability of satisfying the inter-disciplinary compatibility equations. This idea is used to develop a likelihood-based approach for multidisciplinary analysis (LAMDA). Using the distributions of the feedback variables, the bi-directional coupling can be reduced to unidirectional coupling, while still preserving the mathematical relationship between the coupling variables. In realistic structures, the number of coupling variables is potentially large. Therefore, principal component analysis (PCA) is adopted to decrease the number of coupling variables. For computational efficiency, a Bayesian Belief Network, along with copula-based sampling, is employed to implement probabilistic multi-disciplinary analysis. The multi-disciplinary optimization (MDO) methodology includes three sources of uncertainty: physical variability; data uncertainty (epistemic) due to sparse or imprecise data; and model uncertainty (epistemic) due to modeling errors/approximations. A Gaussian Process (GP) surrogate model is adopted for the sake of computational efficiency in optimization. The proposed methods are illustrated using 3-D aeroelasticity analysis and optimum design of an aircraft wing.

I. Introduction

Uncertainty quantification and design under uncertainty of realistic structures requires multiple functional evaluations in the presence of natural variability, data uncertainty and model uncertainty. When the structural system requires high-dimensional coupled multi-disciplinary analysis between several disciplines, the computational effort multiplies tremendously, creating difficulties in implementing reliability analysis and reliability-based optimization techniques. This paper combines efficient uncertainty propagation, surrogate modeling, and sampling techniques to overcome such challenges.

In this paper, we consider three sources of uncertainty in engineering analysis and design optimization: (1) Variability (aleatory) due to the inherent physical variability in all processes. (2) Data uncertainty (epistemic) due to lack of sufficient information and (3) Model uncertainty (epistemic) caused due to model form assumptions as well as numerical approximations at several stages of the analysis. It is necessary to account for these various sources of uncertainty in both multidisciplinary analysis (MDA) and multidisciplinary optimization (MDO) problems.

To address the issue of uncertainty propagation analysis in multi-level, multi-disciplinary systems, a likelihood-based methodology (LAMDA) [1] is adopted. In this method, the probability of satisfying the inter-disciplinary compatibility is calculated, which is then used to estimate the probability density function (PDF) of the coupling variables. This approach requires no coupled system analysis, i.e. iteratively performing individual disciplinary analyses until convergence, thereby improving the computational cost. For multi-level systems, the difficulty in propagating the uncertainty in the feedback variables to the system output is overcome by replacing the feedback coupling with unidirectional coupling. The direction of coupling can be chosen either way, without loss of generality.

In realistic multidisciplinary analyses, the amount of data exchanged could be extremely large and coupling variables in the same direction could be correlated. (For example, in aero-elasticity analysis shown in Fig.1 and Fig. 2, the aerodynamic response, i.e. pressure distributed on the airfoil, and structural responses, i.e. displacement of the nodes of the wing, are not independent and the sizes are tremendous). As the dimension increases, the computational expense of MDA grows dramatically; meanwhile, in the context of aero-elasticity problem, the coupling variables in each direction are inherently coupled. Thus, a model reduction technique together with fast sampling technique is required to overcome the computational challenge. A principal component analysis (PCA) [2] is adopted to decrease the number of coupling variables into a much smaller size of independent variables. Within the principal component space, a Bayesian Network along with copula-based sampling [3,4,5,6,7] is exerted to construct the probability function of the variables independently.

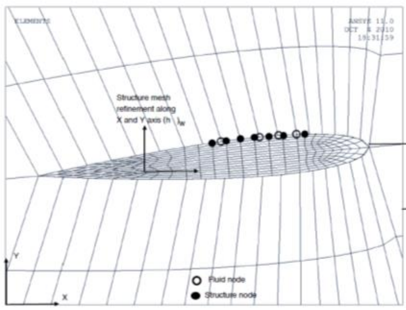


Figure 1: 3-D Wing with NACA 0012 Airfoil

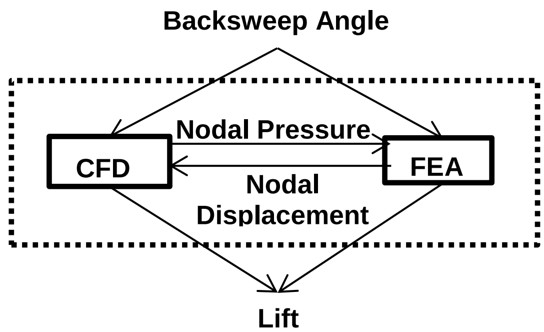


Figure 2: Functional Relationships between Two Analyses

Optimization methods under uncertainty within engineering design have generally been pursued in two directions: (1) reliability-based design optimization (RBDO) [8], where the focus has been on achieving desired reliability levels for constraints, and (2) robust design optimization (RDO), where the primary focus has been on minimizing objective function variations, typically quantified by their respective variances. Design optimization under aleatory uncertainty is a well-researched topic in the literature. However, when epistemic uncertainty is considered, the existing methods address only parts of the entire problem scope. In this paper, we consider the aleatory uncertainty, data uncertainty and model errors within an optimization framework. A likelihood-based approach [9] is used to probabilistically quantify these uncertainties. The constraint and objective functions are assumed to be available only through computationally expensive simulation models. For each function, the model form error and numerical solution error are quantified, as a function of the design variables. The computationally expensive objective/constraint calculation model is replaced by a Gaussian process surrogate model for the sake of computational efficiency [10]. The resulting surrogate model error is also quantified as a function of the design variable inputs. The contributions of this paper can be summarized as follows:

1) Extension of LAMDA to problems with a large number of correlated coupling variables;
2) Development of a probabilistic framework to include natural variability, data uncertainty and model uncertainty within the design optimization problem;
3) Quantification of various types of model errors as functions of design variable values including model form error and discretization error;
4) Uncertainty quantification in model output, through probability distributions, due to variability, data uncertainty, and model errors;
5) Use of a Bayesian Network together with copula-based sampling method to improve the computational efficiency of probabilistic MDA; and
6) Illustration of the proposed methodology with an aircraft wing analysis and design problem.

II. Unidirectional Decoupling for Feedback Multidisciplinary Analysis

Consider the multi-disciplinary system shown earlier in Fig. 3. Given the probability distributions of the inputs $\boldsymbol{x}$, the target problem is to estimate the distribution of each single disciplinary outputs g_1, g_2, and final system output f. An intermediate step is to calculate the PDFs of the coupling variables u_{12} and u_{21} for uncertainty propagation. Consider the deterministic problem of estimating the converged u_{12} and u_{21} values corresponding to given values of $\boldsymbol{x}$. In the conventional fixed point iteration approach, an arbitrary value of u_{12} is taken as input to “Analysis 2” for start, and the resultant value of u_{21} serves as input to “Analysis 1”. This proceeds recursively until the inter-disciplinary compatibility is satisfied, which means the output from “Analysis 1” is the same as the initial u_{12}. Then the analysis is said to have reached convergence and is terminated.

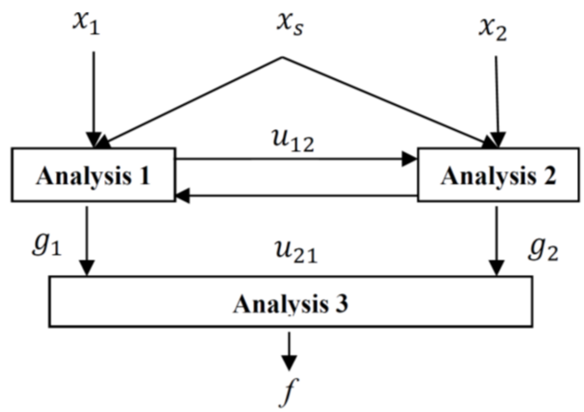


Figure 3: Multidisciplinary System

A unidirectional coupling is formulated by severing the coupling from analysis 1 to analysis 2. A new function G is defined such that: the input of G is the coupling variable u_{12}, in addition to $\boldsymbol{x}$, and the output is denoted by U_{12}, which is obtained by propagating the input through Analysis 2 followed by Analysis 1, as shown in Fig. 4.

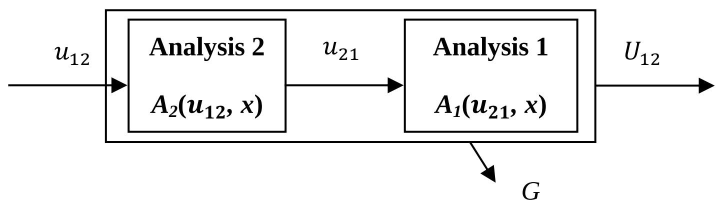


Figure 4: Definition of G

If the multi-disciplinary compatibility is satisfied, then $u_{12} = U_{12}$. In this paper, it is assumed that the inter-disciplinary compatibility for each realization of the input $\boldsymbol{x}$ could be satisfied, i.e. for each value of input u_{12}, function G has a feasible output U_{12}. A likelihood-based approach to calculate this probability si formulated in [1] and briefly summarized below.

Let X be a random variable, whose probability distribution P is dependent to parameter θ. The likelihood function of θ given the outcome of x is probability of observing the given data conditioned on the parameter:

$$\mathcal{L}(\theta) = P(X = x|\theta) \tag{1}$$

In the context of the MDA problem, the probability of the interdisciplinary compatibility condition is satisfied conditioned on u_{12} can be denoted as $P(U_{12} = u_{12} \mid u_{12})$. It is the likelihood that a particular value of u_{12} satisfies the multi-disciplinary constraint. It is only meaningful up to a proportionality constant as following:

$$L(u_{12}) \propto P(U_{12} = u_{12}|u_{12}) \tag{2}$$

Consider the operation $U_{12} \equiv G\,(u_{12}, \boldsymbol{x})$, for a given value of u_{12}, because of the randomness of $\boldsymbol{x}$, the output U_{12} can be calculated by different uncertainty propagation method. The PDF of U_{12} is denoted by $f_{U_{12}}(U_{12}|u_{12})$. The target is to find the PDF value of u_{12} at which the compatibility condition is satisfied, i.e. $U_{12} = u_{12}$. The likelihood is expressed by the integral of the PDF within a small window around the converged u_{12}:

$$L(u_{12}) \propto \int_{u_{12}-\frac{\varepsilon}{2}}^{u_{12}+\frac{\varepsilon}{2}} f_{U_{12}}(U_{12}|u_{12})dU_{12} \propto f_{U_{12}}(U_{12} = u_{12}|u_{12}) \tag{3}$$

Once the likelihood function of u_{12}, i.e the probability of satisfying the multidisciplinary compatibility for a given value of u_{12} is calculated, the PDF of the converged value of the coupling variable u_{12} can be calculated as:

$$f(u_{12}) = \frac{L(u_{12})}{\int L(u_{12})du_{12}} \tag{4}$$

Without loss of generality, the same approach can be used to calculate the PDF of $u_{21}$and cut the coupling from "Analysis 2" to "Analysis 1." The evaluation the likelihood function $L(u_{12})$ for a given value of u_{12} requires the calculation of the PDF $f_{U_{12}}(U_{12} = u_{12}|u_{12})$. A first-order reliability method (FORM) has been employed in [1] to accomplish the local PDF estimation at $U_{12} = u_{12}$. Details of FORM are explained in Haldar and Mahadevan [11].

In the case where the coupling variables are scalars, each likelihood evaluation requires two FORM analyses to obtain the PDF. As the number of coupling variables increases, the integration procedure causes the computational cost to increase exponentially. In the context of aeroelasticity, the coupling analysis between FEA and CFD is associated with the exchange of nodal pressures and nodal displacements, both of which are field quantities. Because of the continuity of structure and fluid, these quantities are inherently coupled. The large number of nodes in FEA and CFD meshes makes it computationally expensive to implement LAMDA using FORM. Moreover, the fluid-structure analysis for an aircraft wing takes large effort to converge even for one input realization; this makes Monte Carlo sampling prohibitive. Therefore, a more efficient sampling and PDF evaluation technique is required.

III. Model Reduction and Sampling Techniques

For every given value of $\boldsymbol{x}$, there is only one converged value of u_{12}, which suggests a functional dependence between the input and coupling variable. However, due to the uncertainty from the input and measurement noise from the output, a statistical dependence describes the relationship in between better. Based on this idea, a Bayesian network is introduced to represent this relationship.

Consider a Bayesian network over a set of nodes: $U = \{X_1 \ldots X_n\}$. The joint probability of $\{X_1 \ldots X_n\}$ is given as:

$$P(U) = P\{X_1 \ldots X_n\} = \prod_{i=1}^{n} P(X_i|\pi_i) \tag{5}$$

where $\pi_i$is the set of parents of node X_i. The marginal probability function of X_i is denoted as:

$$P(X_i) = \sum_{X_{\sim i}} P(U) \tag{6}$$

A common usage of Bayesian network is to infer the beliefs of events given the observation of other events, which is called evidence. The joint probability of events U given evidence e could be evaluated based on Bayes theorem:

$$P(U|e) = \frac{P(U,e)}{P(e)} = \frac{P(U,e)}{\sum_U P(U,e)} \tag{7}$$

Let ε denote the difference between the input u_{12} and output U_{12} of G: $= U_{12} - u_{12}$. The evidence e, which is the interdisciplinary compatibility is satisfied, could be represented as $\varepsilon = 0$. Therefore, the PDF of U_{12} given the compatibility condition is:

$$f_{U_{12}}(U_{12}|U_{12} = u_{12}) = f_{U_{12}}(U_{12}|\varepsilon = 0) \tag{8}$$

To obtain the PDF of the coupling variable given the compatibility condition, a copula sampling technique is introduced. Consider n random variables: X_i $(i = 1 \ldots n)$, all of which have continuous cumulative distribution functions (CDFs): $F_i(x_i)$. A n-dimensional copula function relating to X_i is defined as:

$$C = P[\cup_{i=1}^{n} F_i(x_i) \le u_i] \tag{9}$$

where u_i denotes the CDF value of X_i. A copula is naturally a joint CDF of the CDF of the variables, it doesn't only contains the information about the marginal distribution, but also the dependence relationship between the two variables. The n-variate copula can also be written as the following form for the sake of generating samples:

$$C = P[\bigcup_{i=1}^{n} x_i \leq F_i^{-1}(u_i)] \tag{10}$$

Samples of u_i are drawn from the copula, the inverse CDF of the samples is taken to retrieve the corresponding values of x_i.

There are a variety of choices for copulas. They can be mainly categorized in to (a) families of copulas, which are parametric, and (b) empirical copulas, which are nonparametric. This paper focuses on one of the most wildly used parametric copulas, namely the multivariate Gaussian copula (MGC) [6,7]. Let Σ be a symmetric, semi-positive definite matrix with $\mathrm{diag}(\Sigma) = (1,1 \dots 1)^{\mathrm{T}}$, and Φ_Σ the standardized multivariate normal distribution with correlation matrix Σ, the MGC is defined as:

$$C_\Sigma^{Gauss}(u) = \frac{1}{\sqrt{det(\Sigma)}} \exp\left(-\frac{1}{2}\begin{pmatrix}\Phi^{-1}(u_1)\\ \vdots \\ \Phi^{-1}(u_d)\end{pmatrix} \cdot (\Sigma^{-1} - \mathrm{I}) \cdot \begin{pmatrix}\Phi^{-1}(u_1)\\ \vdots \\ \Phi^{-1}(u_d)\end{pmatrix}\right) \tag{11}$$

where Φ^{-1} is the inverse of standard uni-variate normal distribution function Φ.

Computational efficiency is the major advantage of copula-based sampling compare with MCMC-based sampling for Bayesian updating. One issue raised from this process is that: the correlation matrix calculated from the original sample may not be linear. In fact, the linear correlation represents only the linear dependence between random variables, and when nonlinear transformation is applied on the variables, the linear correlation is no longer preserved. The rank correlation is therefore employed to describe the dependence relations in a more appropriate manner. A rank correlation is the linear correlation between the ranks of observations between two variables. The rank correlation coefficient, such as Kendall's τ or Spearman's ρ, is an invariant under any monotonic transformation on the marginal distribution, and therefore is more useful in describing the dependence between random variables.

To perform the LAMDA approach in a Bayesian manner, samples of coupling variables as input and as output need to be acquired beforehand. These samples could come from any two consecutive iterations of the MDA except for the initial samples and the result after first iteration, since the initial value of the coupling variable is assumed and is not an available piece of information to construct Bayesian network. Then the difference of the corresponding input/output is denoted as ε. Consequently, the samples of U_{12}, u_{12} and ε are used to build the network, and the PDF of the converged coupling variable u_{12} can be estimated by calculate the conditional PDF given $\varepsilon = 0$.

The Bayesian network together with copula sampling technique could largely simplify the MDA process for a high dimensional coupled system. However, when the coupling variables in on direction are correlated with each other, the effort to build the Bayesian network could be huge, and problems may rise particularly when two variables are highly correlated, the correlation matrix might be semi-positive definite. Therefore, principal component analysis is employed to transform the correlated variables into an orthogonal space. Each principal component is a linear combination of the original variables; all the principal components are orthogonal to each other, hence no redundant information is stored. Usually, the variances captured by the first 5 to 10 principal components could exceed 90% of the total variance of the original data. Therefore, PCA can help us to proceed with the construction of Bayesian network using a smaller number of independent variables (first several principal components). To perform the component, the input samples u_{12} and output samples U_{12} should first be assembled as: $\widetilde{U} = \begin{bmatrix} u_{12} \\ U_{12} \end{bmatrix}$. The principal component for this matrix is denoted as: $PC = \begin{bmatrix} PC_{u_{12}} \\ PC_{U_{12}} \end{bmatrix}$. Only the first few columns of PC is required for LAMDA. A simple Bayesian network could be created for one principal component at one time as shown in Fig. 5. $PC_{u_{12}}^i$ and $PC_{U_{12}}^i$ denote the i^{th} column in the upper and lower part of PC. The collection of updated samples from the first several principal components is then used to transform into the original correlated samples.

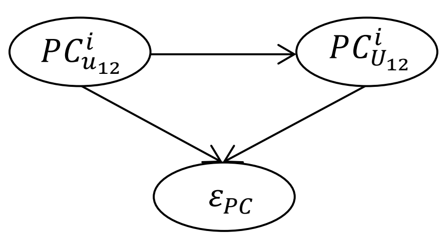


Figure 5: Bayesian Network for Multidisciplinary Analysis

The above method is illustrated in this paper for an areoelasticity problem that requires the multidisciplinary analysis.

IV. RBDO formulation

An RBDO formulation of the above problem can be given as:

$$\min_{\mu_X, d}[\mu_f(X, d, P, p_d)]$$

$$s.t. \quad Prob(g_i(X, d, P, p_d) \leq 0) \geq p_t^i \; i = \{1, \ldots, n_q\} \tag{12}$$
$$Prob(X \geq lb_X) \geq p_{lb}^t$$
$$Prob(X \geq ub_X) \geq p_{ub}^t$$
$$lb_d \leq d \leq ub_d$$

where X is the vector of random design variables with bounds lb_X and ub_X, respectively; μ_f and μ_X are the mean of f and X, respectively; d is the vector of deterministic design variables with bounds lb_d and ub_d; P is the vector of random parameters; P_d is the vector of deterministic parameters; and p_t^i is the target reliability required for the i-th inequality constraint; p_{lb}^t and p_{ub}^t are the target reliabilities for the design variable bounds.

V. Sources of errors and uncertainties

A. Data uncertainty (epistemic): design and non-design variables

A likelihood-based approach is used to represent the data uncertainty due to sparse point data and interval data. The methodology can handle mixed data, e.g., both point data and interval data for the same variable, and fit a non-parametric PDF thereby avoiding the assumption of a distribution type. This PDF can be easily used within uncertainty propagation either using sampling methods such as Monte Carlo simulations or analytical methods such as the First Order Reliability Method. Note that this approach can also handle correlated variables by using a non-parametric joint PDF and the corresponding joint likelihood, given sparse or interval data for multiple variables.

B. Model Errors

Two types of model errors are considered: numerical [12] solution error and model form error. Let G(x) represent the true function value that we are interested in estimating, where x represents a set of point-valued inputs. The following equation can be written as following:

$$G = g_{raw} + \epsilon_{num} + \epsilon_{mf} \tag{13}$$

where g_{raw} is the function value as evaluated by a computational code, ϵ_{num} is the numerical solution error, and ϵ_{mf} is the model form error. Two types of numerical solution errors are considered in this paper: discretization error and error due to surrogate model use.

1) Discretization error

Discretization error is a critical component of numerical solution errors; this arises due to the solution of a continuum problem through finite element or finite difference solution methods that discretize the continuous domain. In this paper, we assume multidimensional mesh refinement, where the mesh can be refined in more than one direction, and possible cases where meshes do not match topologically[13]. A GP model is used for discretization error estimation, as an enhancement of the traditional Richardson extrapolation [14] to alleviate its drawbacks such as requirement of monotonic convergence, and being in the asymptotic convergence region of the mesh size.

2) Surrogate model error

Use of the GP surrogate model introduces surrogate model error. In order to estimate this error as a function of the design variables within the optimization, we construct a GP model for the surrogate model error, denoted as ϵ_{se} , which is also a Gaussian variable for each x. The constraint function can then be written as following:

$$G(x) = g_h(x) + \epsilon_{se}(x) + \epsilon_{mf}(x) \tag{14}$$

Note that at each x, the numerical solution errors of the constraint function, is a Gaussian variable, thereby resulting in a distribution of the model prediction, instead of a point value.

3) Model form error

We assume that the model form error also varies as the inputs to the model change. Assume that a set of observation data from physical experiments is available at a set of q design inputs, $O_{set} = \{X^T_{obs1}, \ldots X^T_{obsq}\}$. The following equation can be written as following:

$$G(x) = g_{obs}(x) + \epsilon_{exp} = g_h(x) + \epsilon_{se}(x) \tag{15}$$

where g_{obs} is the observed experimental value, and ϵ_{exp} is the experimental error. We assume that the experimental error is the same at all design inputs, although this assumption can be changed easily if needed. The model form error can be evaluated as following:

$$\epsilon_{mf}(x) = g_{obs}(x) - [g_h(x) + \epsilon_{se}(x)] + \epsilon_{exp} \tag{16}$$

It is evaluated at each of the values O_{set}, where at each x, ϵ_{mf} (x) is a Gaussian variable, assuming ϵexp is Gaussian. A GP model, $g_{mf}(x)$ is fit with the training data $\{O_{set}; \epsilon_{mf}(O_{set})\}$ to yield the model form error as a function of the input. The expression for G(x) can finally be written as following:

$$g_{true}(x) = g_h(x) + g_{se}(x) + g_{mf}(x) \tag{17}$$

Fig. 6 summarizes the discussion on quantifying various model errors.

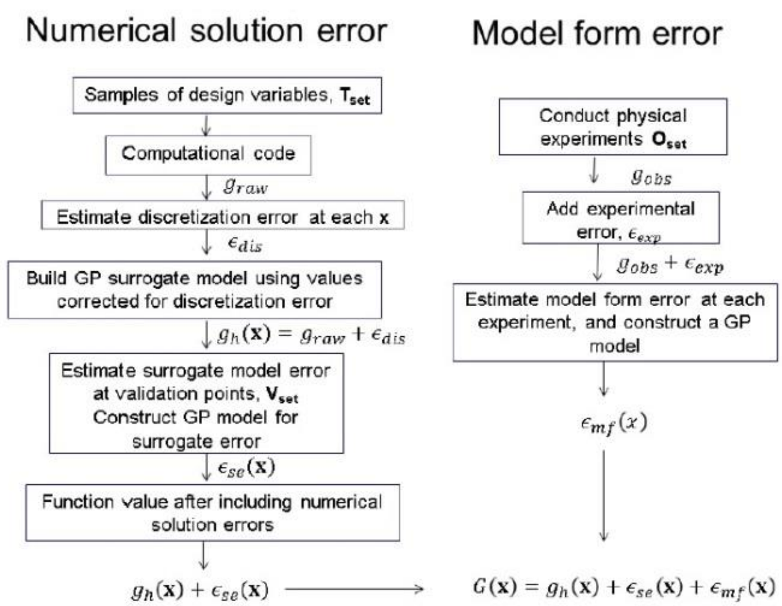


Figure 6: Summary of the model error quantification method

C. Stochastic model prediction in optimization

The model errors discussed in this section are stochastic quantities, thereby resulting in a stochastic model prediction for constraints and objectives at a given design input. These stochastic model predictions are now used in the optimization formulation, and must be appropriately used to yield constraint failure probabilities and the objective function. As shown in Fig. 7, in the traditional method, each random sample is propagated through the objective function model, resulting in a single point value of f; these propagated output samples are then used to compute the mean and standard deviation of f. In other words, the only source of uncertainty in such a case arises because of the input. In the proposed model error treatment on the other hand, each random sample of the input results in a Gaussian distribution of the corresponding prediction. If the optimization requires minimization of mean objective value, μ_f = mean(μ_{fi}), i = {1, ..., n}, or the worst case scenario, μ_f= max(μ_{fi}), i = {1, ..., n} can be used.

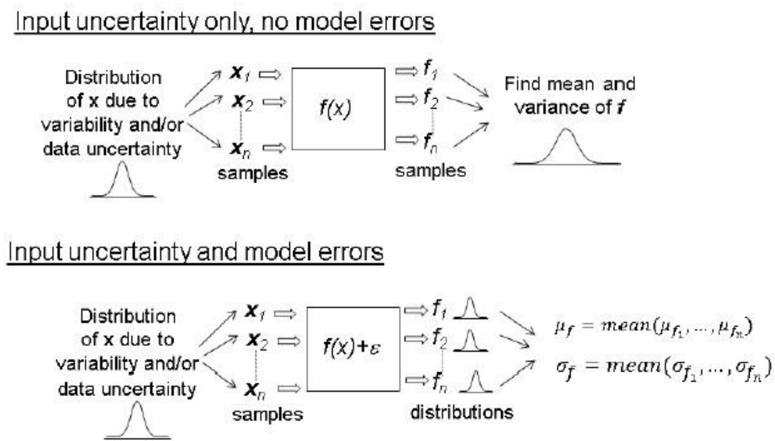


Figure7: Objective function evaluation with stochastic model predictions

For constraint failure evaluation, we exploit the fact that the corrected model value after correction for model form and discretization errors is a Gaussian variable. At each random sample of the input X_i, we find the corresponding probability of failure given that the true model prediction, $g_{true}(x)$ is a Gaussian distribution, as shown in Fig. 8, as P_{f_i} = $Prob(G(X_i) \leq 0)$, which is a simple evaluation of a Gaussian cumulative density function value. The use of the proposed auxiliary variable method can transform the output variability into an additional input uncertainty. Assume the original input variable vector is X, and $\delta$is a standard uniform distribution representing the CDF of the output variability. With the auxiliary variable δ, the input variable vector becomes $X' = [X, \delta]$and can be included in MDA and MDO.

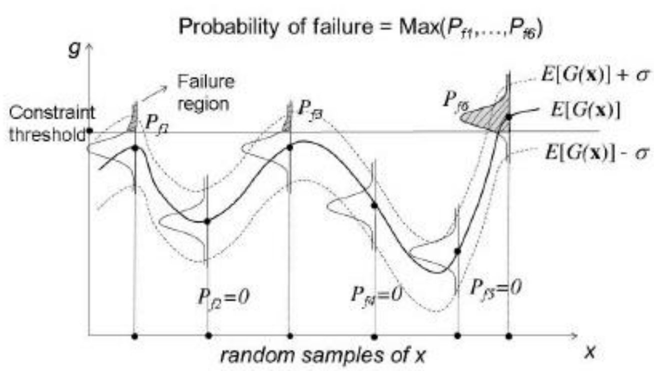


Figure 8: Constraint failure probability evaluation with considering model errors

VI. Design optimization formulation under uncertainty

The optimization problem is presented and the steps involved are outlined as follows:

1) Choose the training data, T_{set}, verification data, V_{set}, and observation data, O_{set} as using a suitable design of experiments.
2) Construct the GP models for the function value after correction for discretization error, surrogate model error, and model form error as functions of the design inputs.
3) Perform uncertainty quantification for all epistemic variables using the likelihood approach. The likelihood based PDF is used to determine the variance of the epistemic variable, which is then used in the optimization problem, and the mean of the epistemic variable is designed. The aleatory variables are described through probability distributions.
4) Solve the following optimization problem:

$$\min_{\mu_{\mathbf{X}},\mathbf{d}} \; E[F(\mathbf{X},\mathbf{d},\mathbf{P},\mathbf{p}_d)]$$

$$\text{s.t}$$

$$Prob(G(\mathbf{X},\mathbf{d},\mathbf{P},\mathbf{p}_d) \leq 0) \geq p_t^i \;\; i = \{1, ..., n_q\}$$

$$Prob(\mathbf{X} \geq \mathbf{lb}_X) \geq p_{lb}^t$$

$$Prob(\mathbf{X} \leq \mathbf{ub}_X) \geq p_{ub}^t$$

$$\mathbf{lb}_d \leq \mathbf{d} \leq \mathbf{ub}_d$$

where F and G are the models corrected for numerical solution error and model form error.

5) The bounds on the means of the epistemic variables can also be posed as constraints

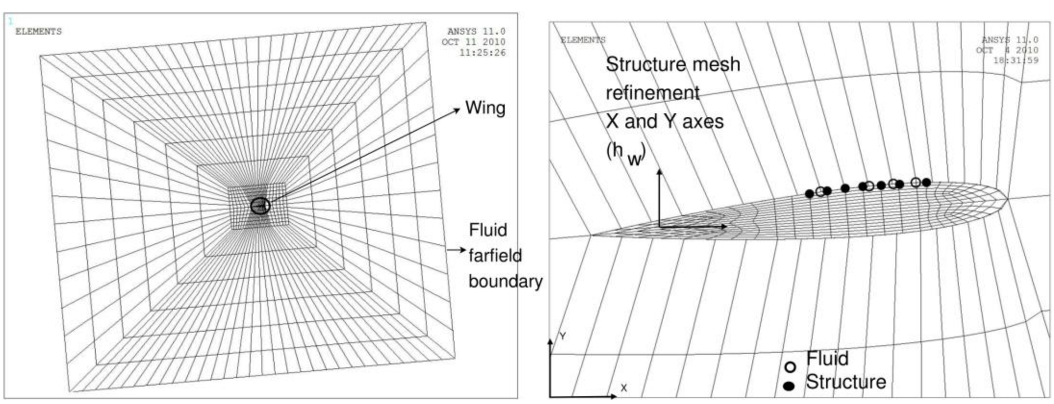
$$\mathbf{lb}_X + \mathbf{k}\sigma_{\mathbf{X}} \leq \mu_{\mathbf{X}} \leq \mathbf{ub}_X - \mathbf{k}\sigma_{\mathbf{X}}$$

where σ_X for the epistemic variable can be estimated from the likelihood-based PDF constructed.

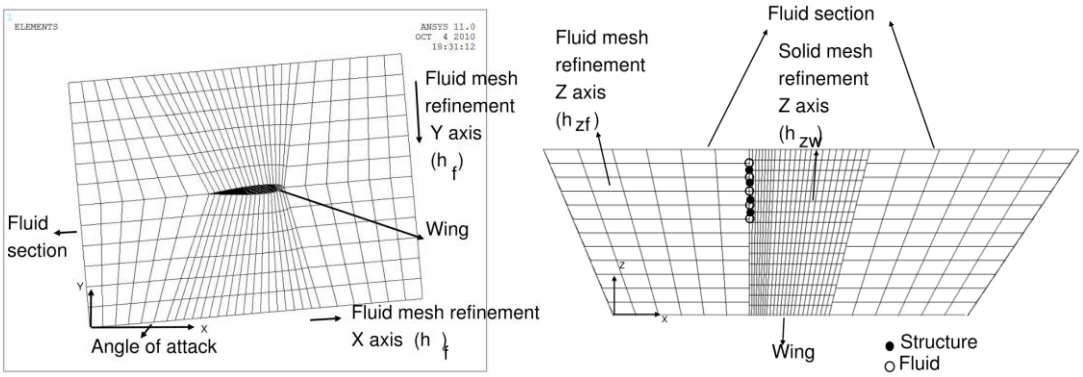
The above probabilistic formulation integrates both aleatory and epistemic uncertainties within the optimization framework.

VII. Numerical Example

3-D aeroelastic analysis and design optimization of an aircraft wing are used to illustrate the proposed methods. A cantilevered wing with a NACA 0012 airfoil is adopted from reference [15]; SI units are used in this paper. We use ANSYS to perform the fluid-structure interaction analysis of the wing. The fluid structure meshes are shown in fig. 9.



a) **Overall view** **(b) Fluid and structure meshes : View I**



c) Fluid and structure meshes: View II (d) Fluid and structure meshes: View III

Figure 9: Non-matching fluid and structure meshes and refinement parameters

A. Multidisciplinary Analysis

The focus of the MDA problem is the propagation of input variability through the fluid-structure system and its impact on the nodal pressure, which is the exchanging variable between the FEA-CFD anaylsis. The backsweep angle bw is chosen to be the input variable with natural variability. It is assumed to be normally distributed as $N(0.4, 0.04)$. 60 samples of backsweep angle are drawn to perform the analysis. 258 nodes are created after mesh. The analysis is terminated after 3 iterations, and the nodal pressure after iteration 2 and iteration 3 are recorded. The 3-iteration analysis is performed on desktop. For the purpose of comparison, converged analysis is performed by Vanderbilt University's ACCRE computer cluster. Five cases are discussed here: (1) LAMDA method using first 10 principal components. (2) LAMDA using 15 principal components. (3) LAMDA using 20 principal components. (4) Results after 2 iterations. (5) Results after 3 iterations.

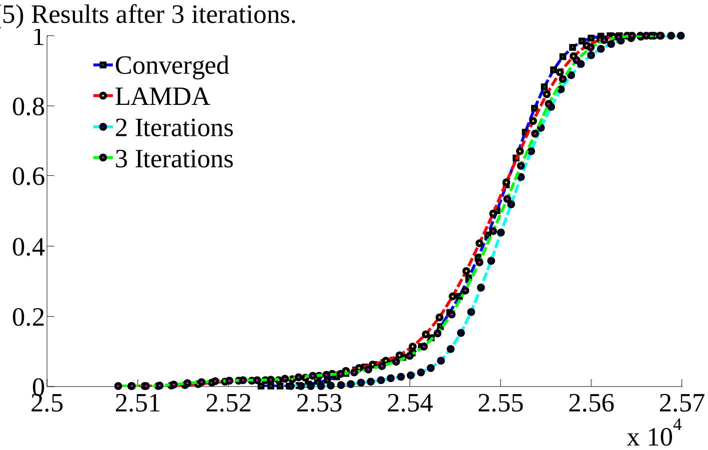


Figure 10: CDF of Pressure from Three Cases at Node I

Fig. 10 shows the CDF of Cases III, IV and V at node I. Both case III and V are closer to the converged distribution compare with case IV at Node I. For further comparison, the Kullback-Leiber Divergence (also called K-L distance) is adopted to quantify the differences between the converged PDF and the approximated result by case I - V at each node. The K-L distance of distribution Q from P, denoted by $D_{KL}(P||Q)$, is a measurement of information lost when P is approximated by Q. Details about the methodology can be found in [16]. For P and Q whose distributions are both continuous functions, the K-L distance is defined to be:

$$D_{KL}(P||Q) = \int_{-\infty}^{+\infty} ln\left(\frac{p(x)}{q(x)}\right)\, dP \tag{18}$$

where p and q are the density functions of P and Q. When continuous density functions are not available, the KLIC can be numerically approximated by discrete probability formulation:

$$D_{KL}(P||Q) = \sum_i ln\left(\frac{f_P(x_i)}{f_Q(x_i)}\right) f_P(x_i) \quad (19)$$

where f_P and f_Q are the PDF value evaluated at the same input. Given the sample of the pressure at a nodes, a kernel density function is estimated. 100 samples of *x* are drawn and their PDF values are estimated by interpolation from the kernel density. For *n*-variate distribution, the K-L distance for marginal distributions are computed, and a total average is estimated in formula (20). In the MDA problem, D_{KL}^i denotes the K-L distance at the ith node.

$$\text{Total Average}= \frac{\sqrt{\sum_{i=1}^{n}(D_{KL}^{i}(P||Q))^2}}{n} \quad (20)$$

Table 1: Problem 1: Kullback-Leiber Divergence for Different Scenario

Node	**LAMDA**			**Case IV**	**Case V**
	Case I	**Case II**	**Case III**		
1	4.0633	3.3345	2.5188	1.8128	2.7615
2	-1.5435	-2.1360	-2.0427	-2.8338	-1.6907
3	1.7146	0.7354	0.9738	-2.9582	-0.0919
4	-1.8010	-2.5998	-1.9280	-4.6760	-3.0249
5	-2.7942	-3.3097	-2.8948	-4.3130	-3.1337
6	-3.9948	-3.8008	-3.7689	6.0775	-0.3395
7	-0.6244	-0.8670	-1.0177	3.7862	0.6780
8	-1.1030	-0.3619	-0.5512	10.9608	3.0114
9	-2.8140	-2.4582	-2.4669	12.0204	2.0541
10	-0.6373	-0.0513	1.2029	-5.2353	-1.1330
Total Average	0.1939	0.1415	0.1386	0.4437	0.1918

Tables 1 summarizes the results of the MDA. The K-L distance at the first 10 nodes are listed. and the following observations are made.

(1) Case I, Case II and Case III: as the utilized principal components increases, the K-L distance tend to be smaller. This is due to the reason that the more principal components are taken, the more variances of the original samples will be caught. Therefore, the results become more accurate. Principal components that higher than 20 has small impact on the final result.

(2) Case IV and Case V: there is a Case V approaches the converged results much better than Case IV. As iteration increases, the stability of system will be enhanced, and the differences between the results from contiguous iterations will become smaller.

(3) Case II, Case III and Case V: The proposed LAMDA method together with the model reduction technique for data uncertainty propagation could overall well quantify the converged distribution of coupling variable just using the result from the first 3 iterations of the analysis instead of performing the converging analysis.

(4) The computational time for 3-iteration analysis on average is 3 min/input on desktop, and it takes 120min/input on average for the analysis to converge. Therefore, the proposed approach as a great superiority in speed.

(5) The K-L distance of LAMDA method could be reduced by adding more input samples to construct the Bayesian network.

Therefore, the proposed LAMDA method together with model reduction technique is evidently computationally economical. It is an efficient tool for dealing with MDA system with large amount of exchanging variables and long converging time. The proposed method can be easily extended to the system with data uncertainty and model error c Bayesian network.

B. Reliability Based Design Optimization

The design optimization problem is to maximize the lift generated by the wing subject to a stress constraint, and the design variable is the backsweep angle. The variability of the manufacturing process is characterized as epistemic described by sparse and interval data. The constraint and objective function values are evaluated by ANSYS, and surrogate models are built to be used in the optimization, thereby introducing discretization error and surrogate model error. In addition, model form error is also considered. All of the above errors are considered as varying functions of the backsweep angle. Five sub-problems are considered within this example to illustrate the proposed developments: (1) Case 1: deterministic optimization problem, (2) Case 2: Optimization under uncertainty with variability alone, and no data or model error considerations, (3) Case 3: Optimization under uncertainty with variability and data uncertainty alone, but no model error considerations, and (4) Case 4: optimization with variability, data uncertainty and mean values of model errors, and (5) Case 5: optimization with variability, data uncertainty and model errors, using the entire distributions of the model errors.

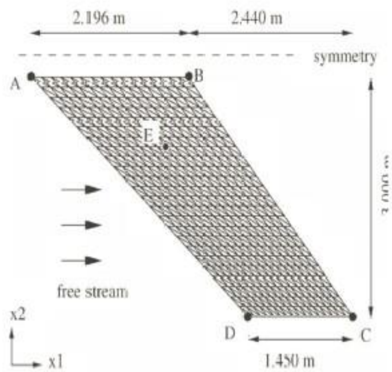


Figure 11: Two-dimensional view of the wing configuration

Table 2: Problem 2: Optimization results summary

Problem	μ_{bw}^*	μ_L^*
Deterministic (Case 1)	0.2794	1927.56
Aleatory (Case 2)	0.4097	1796.52
Variability and data uncertainty (Case 3)	0.3643	1814.05
All uncertainties – mean model errors (Case 4)	0.3326	1919.87
All uncertainties – distribution of model errors (Case 5)	0.3603	1929.16

Table 3: Problem 2: Optimization results summary

	Probability of failure		**Mean objective**	
Design Case	**Ignoring model errors**	**Including distribution of model errors**	**Ignoring model errors**	**Including distribution of model errors**
Case 1	0	0.0422	1889.8	1908.8
Case 2	0	0.0487	1796.51	1907.98
Case 3	0	4.67 × 10−5	1912.15	1929.16
Case 4	-	5.65 × 10−4	-	1908.83
Case 5	-	7.8 × 10−4	-	1929.16

Tables 2 and 3 summarize the results of the above problems, and the following observations are made.

(1) Case 1 versus Cases 2 and 3: When data uncertainty is considered in the problem, the design objective worsened (Table 1), which is generally expected when uncertainty in the design variables is considered in the problem.

(2) Cases 1, 2 and 3 versus Case 4: When model errors are corrected, the worsening of the objective due to data uncertainty is seemingly canceled (in this problem) by the model error corrections applied.

(3) Case 4 versus Case 5: The probability of failure when only mean model errors (Case 4) was observed to be zero, while that when the entire distribution of model error is used (Case 5) was observed to be 7.835×10^{-4}. This is to be expected, as shown in Figure 3.

(4) In Table 2, we conduct a scenario analysis to understand the impact of model errors. In row 1, we use the optimum solution of the deterministic problem, and assume data uncertainty at this solution. Say the system was designed at the deterministic solution of backsweep angle of 0.2794 rad, and there was variability in the actual wing manufactured. How does this impact the probability of failure and the objective function? Furthermore, how does the model error consideration impact this analysis? For row 1, it is seen that the even when mean model errors are considered, the probability of failure is zero. However, if the entire distribution of the model errors is considered, there is non-zero probability of failure. This means that a complete consideration of model errors and the uncertainties in their estimation can increase the probability of failure.

(5) For the objective function, we do a similar study – consider data uncertainty about the deterministic optimum solution, and study how the objective function changes. If no model errors are considered, the objective function performance deteriorates when compared to the one predicted by the deterministic optimization. In this problem, the consideration of distribution of model errors does not cause a profound difference compared to using mean values, since the objective function involves the mean value as against probability of failure.

VIII. Conclusion

In this paper, we proposed approaches to efficiently propogate the data uncertainty through multidisciplinary system with field quantity as coupling variable. The likelihood-based approach for MDA is improved by using Bayesian network together with copula sampling and principal component analysis, such that the high-dimensional coupled problem could be decoupled and the dimension could be decreased. The principal component analysis, which transforms the coupling variables in one direction into independent principal components, provides a convenient situation to construct the Bayesian network. The use of the proposed method is explained by an aeroelasticity problem of a 3-D aircraft wing, which requires fluid-structure analysis.

The proposed method has obvious advantage in computational efficiency, while still providing satisfactory results. We also proposed methods to systematically include aleatory and epistemic uncertainties in the design variables, non-design variables, and model errors into the optimization problem. We employ a likelihood-based approach to handle variables described by sparse and interval data. Numerical solution errors and model form errors are determined prior to the optimization as a function of design variables, and are used in the optimization formulation. It is observed that model errors and the associated uncertainties can have a significant impact on the optimum solution and the constraint reliabilities obtained. In this work, we have used GP models to represent the behavior of model errors with changing design inputs., thereby resulting in Gaussian distributions for the error estimates. Alternate representations of the error models that do not result in a Gaussian distribution might be of interest in other problems.

Acknowledgments

This study was supported by funds from NASA Langley Research Center under Cooperative Agreement No. NNX08AF56A1 (Technical Monitor: Lawrence Green). The support is gratefully acknowledged. Valuable discussion from Dr. Roger Cooke and Dr. Dan Ababei are also gratefully acknowledged. The computational resources of Vanderbilt University's ACCRE were used for FEA and CFD analyses.

References

[1] Sankararaman, S., and Mahadevan, S., “Likelihood-Based Approach for Multidisciplinary Analysis,” *ASME Journal of Mechanical Design*, Vol. 134, No.3, 031008-1 -031008-12

[2] Shlens, Jonathon. “A tutorial on principal component analysis.” Systems Neurobiology Laboratory, University of California at San Diego, 2005.

[3] Mahadevan, Sankaran, Ruoxue Zhang, and Natasha Smith. “Bayesian networks for system reliability reassessment.” Structural Safety 23.3 (2001): 231-251.

[4] Hanea, A. M., D. Kurowicka, and R. M. Cooke. “Hybrid method for quantifying and analyzing bayesian belief nets.” Quality and Reliability Engineering International 22.6 (2006): 709-729.

[5] Bedford, Tim, and Roger M. Cooke. "Probability density decomposition for conditionally dependent random variables modeled by vines." Annals of Mathematics and Artificial intelligence 32.1 (2001): 245-268.

[6] Nelsen, Roger B. “An introduction to copulas”. Springer, 2006.

[7] Cherubini, Umberto, Elisa Luciano, and Walter Vecchiato, “Copula methods in finance”, Wiley, 2004.

[8] Valdebenito, M. and Schuller, G., “A survey on approaches for reliability-based optimization,” Structural and Multidisciplinary Optimization, Vol. 42, 2010, pp. 645–663.

[9] Sankararaman, S. and Mahadevan, S., “Likelihood-based representation of epistemic uncertainty due to sparse point data and/or interval data,” Reliability Engineering and System Safety, Vol. 96, No. 7, 2011, pp. 814 – 824.

[10] Rasmussen, C. E. and Williams, C. K. I., Gaussian processes for machine learning, Springer, 2006.

[11] Haldar, A. and Mahadevan, S., Probability, Reliability, and Statistical Methods in Engineering Design, John Wiley and Sons, Inc, 2000.

[12] Rebba, R., Mahadevan, S., and Huang, S., “Validation and error estimation of computational models,” Reliability Engineering and Systems Safety, Vol. 91, 2006, pp. 1390–1397.

[13] Rangavajhala, S., Sura, V., Hombal, V. K., and Mahadevan, S., “Discretization error estimation in multidisciplinary simulations,” AIAA Journal , Vol. 49, No. 12, 2011, pp. 2673–2683.

[14] Richardson, L. F., “The approximate arithmetical solution by finite differences of physical problems involving differential equations, with an application of the stresses in amasonry dam,” Philosophical transactions of the Royal society of London, Vol. 210, 1911, pp. 307 – 357.

[15] Maute, K., Nikbay, M., and Farhat, C., “Coupled analytical sensitivity analysis and optimization of three-dimensional nonlinear aeroelastic systems,” AIAA Journal , Vol. 39,No. 11, 2001, pp. 2051–2061.

[16] Kullback, Solomon, and Richard A. Leibler. “On information and sufficiency.”The Annals of Mathematical Statistics 22.1 (1951): 79-86.