Stochastic Multi-Disciplinary Analysis under Epistemic Uncertainty

Chen Liang, Sankaran Mahadevan*

Vanderbilt University, Nashville, TN - 37235

and Shankar Sankararaman

SGT, Inc., Moffett Field, CA - 94035

Abstract

This paper presents a probabilistic framework to include the effects of both aleatory and epistemic uncertainty sources in coupled multi-disciplinary analysis (MDA). A likelihood-based decoupling approach has been previously developed for probabilistic analysis of muti-disciplinary systems, but only with aleatory uncertainty in the inputs. This paper extends this approach to incorporate the effects of epistemic uncertainty arising from data uncertainty and model errors. Data uncertainty regarding input variables (due to sparse and interval data) is included through parametric or non-parametric distributions using the principle of likelihood. Model error is included in MDA through an auxiliary variable approach based on the probability integral transform. In the presence of natural variability, data uncertainty and model uncertainty, the proposed methodology is employed to estimate the PDF of coupling variables as well as the subsystem and system level outputs that satisfy interdisciplinary compatibility. Global sensitivity analysis, which has previously considered only aleatory inputs and feed-forward or monolithic problems, is extended in this paper to quantify the contribution of model uncertainty in feedback-coupled MDA by exploiting the auxiliary variable approach. The proposed methodology is demonstrated using a mathematical MDA problem and an electronic packaging application example featuring coupled thermal and electrical subsystem analyses. The results indicate that the proposed methodology can effectively quantify the uncertainty in multidisciplinary analysis while maintaining computational efficiency.

*Corresponding Author **Phone** : 615-322-3040 **Email** : sankaran.mahadevan@vanderbilt.edu

1. Introduction

Multi-disciplinary analysis (MDA) and multidisciplinary design optimization (MDO) focus on developing computational methods [1,2] for systems that involve multiple coupled disciplines, analyses or subsystems in various applications such as fluid-structure interaction [3], thermal-structural analysis [4], fluid-thermal-structural analysis [5], etc. The performance of a multidisciplinary system is determined by individual disciplines as well as the interactions between them. The increasing dimensionality with analysis and design variables accumulated from multiple disciplines presents serious computational challenges in MDA and MDO [6].

Based on the direction of information flow, the coupling between two individual disciplinary analyses can be either uni-directional (feed-forward) or bi-directional (feedback). The focus of this paper is on feedback coupling which is more complex due to the iterations between two analyses to achieve interdisciplinary compatibility. Computational methods for feedback-coupled MDA can be classified into three different groups: (1) Field elimination method [7], (2) monolithic method, and (3) partitioned method [8]. The field elimination and monolithic methods tightly couple the disciplinary analyses together, while the partitioned method does not.

An important factor in the analysis and design of multi-disciplinary systems is the presence of uncertainty in the system inputs and the models used for each analysis. In this paper, we consider three sources of uncertainty in multidisciplinary analysis: (1) Natural variability (aleatory), due to the inherent physical variability in all processes; this is irreducible. The variability can be represented by probability distributions, whose distribution type and parameters are estimated from the data. (2) Data uncertainty (epistemic), due to insufficient information, e.g., limited number of samples and/or imprecise data. (3) Model uncertainty (epistemic), caused by the model form assumptions as well as numerical approximations. Epistemic uncertainty can be reduced by gaining more information about the data and the system. The representation and propagation of these types of uncertainty in multi-disciplinary analysis is the focus of this paper.

In recent years, there is increasing emphasis on design optimization under both aleatory and epistemic uncertainty. Methods for the representation and the propagation of aleatory uncertainty in a monolithic or feed-forward system are well established. Aleatory uncertainty has been modeled through random variables with fixed probability distributions and distribution parameters. A variety of approaches such as Monte Carlo Methods, first-order reliability method (FORM), second-order reliability method (SORM), etc. are available for the propagation of aleatory uncertainty through monolithic or feed-forward analysis [9]. However, a small number of studies have addressed the propagation of both aleatory uncertainty and epistemic uncertainty in single disciplinary analysis [10-18]. Studies on uncertainty propagation through feedback-coupled MDA are even fewer [22-27], only addressing aleatory uncertainty.

The input variables in an analysis could be deterministic or stochastic, and epistemic uncertainty may be present regarding both types of inputs (due to lack of information, mentioned here generally as data uncertainty). For example, an input variable may be a fixed but unknown constant, and the information may be available as an interval from an expert. Or an input variable may have natural variability (aleatory), but due to lack of data, its distribution type and/or parameters may be uncertain. Epistemic uncertainty regarding the inputs has been addressed through evidence theory [11,12], possibility theory [13], fuzzy sets [14], imprecise probabilities [15], p-boxes [16], aleatory-alike treatment and conservative treatment [17] likelihood-based probabilistic approaches [15, 18], etc., but has been mostly applied to feed-forward or monolithic problems.

Model errors can generally be categorized into two types [19]: (1) model form error which is due to assumptions about system behavior, boundary conditions, operating conditions and model parameters, and (2) numerical solution errors, which arise from the solution process adopted to solve the mathematical model, and include discretization error, surrogate model error, truncation error (e.g., lower-order approximations), etc.. Liang and Mahadevan [20] considered a detailed treatment of model errors due to both model form and numerical solution method, and developed a methodology to systematically quantify and aggregate the multiple error sources. Kennedy and O'Hagan [21] quantified model errors for monolithic or feed-forward type of models using calibration data within a Bayesian framework. To the best of our knowledge, no work has been reported in uncertainty quantification that includes model errors within feedback-coupled MDA.

The aforementioned sources of uncertainty (variability, data uncertainty and model errors) cause the output of MDA to be uncertain. Non-deterministic MDA in the presence of variability in the inputs can be accomplished by Monte Carlo sampling around a deterministic MDA method, i.e., sampling outside fixed point iteration (SOFPI). However, such analysis is computationally prohibitive. Efficient alternatives, in the presence of aleatory uncertainty alone, have been investigated by many researchers. Gu et al. [22] proposed worst case uncertainty propagation using derivative-based sensitivities. Kokkolaras et al. [23] used the advanced mean value method for uncertainty propagation and reliability analysis. Liu et al. [24] extended the same method by using moment-matching and considering the first two moments. Decoupling methods developed in deterministic MDO have been used for efficient non-deterministic MDA, in the context of reliability analysis. Du and Chen [25] included the disciplinary compatibility constraints in the most probable point (MPP) estimation for reliability analysis. Mahadevan and Smith [26] developed a multi-constraint first-order reliability method (FORM) for MPP estimation.

Review of the previous studies reveals that the existing methods for MDA under uncertainty either require considerable computational effort or introduce several approximations to reduce the computational effort. For example, in the decoupled approach adopted by Du and Chen [25] and Mahadevan and Smith [26], the probability density functions (PDF) of the coupling variables are calculated by Taylor's series-based first-order second moment approximation. These approaches improve the efficiency by trading off the accuracy since they ignore the dependence between the coupling variables. To include dependence between the coupling variables, a likelihood-based MDA (LAMDA) approach was proposed by Sankararaman and Mahadevan [27]. In this method, the probability of satisfying the inter-disciplinary compatibility is calculated using the principle of likelihood, which is then used to estimate the PDF of the coupling variables. This approach requires no coupled system analysis and yet is theoretically exact, thereby preserving the functional dependence between the individual disciplinary analyses.

In Ref. [27], only aleatory uncertainty was considered. In this paper, the LAMDA method is extended to include epistemic uncertainty (i.e., data uncertainty and model errors) in multidisciplinary analysis. A likelihood-based approach is employed to represent the effect of data uncertainty (sparse and/or imprecise data) through either parametric families of distributions [18] or non-parametric distributions for the input variables [10]. The presence of model uncertainty makes the output of analysis uncertain even for a fixed input, and model error usually varies with the input. This presents a serious challenge for non-deterministic MDA, since previously available methods have only considered models with deterministic output for a particular input realization. A novel approach is developed in this paper to include model error in MDA using the concept of an auxiliary variable defined through the probability integral transform.

The system output uncertainty is due to the contribution of different sources of variability, data uncertainty and model uncertainty. The identification of the dominant contributors of uncertainty can be realized using probabilistic sensitivity analysis. A global sensitivity analysis (GSA) approach [28] which explores the entire space of input factors is considered in this paper. However, previous work in GSA has only considered deterministic feed-forward or monolithic models with only *aleatory* inputs; this paper extends GSA to *feedback-coupled* MDA under *both* aleatory and epistemic uncertainty.

The following sections elaborate the aforementioned contributions. Section 2 briefly introduces the basic LAMDA framework. Section 3 proposes a likelihood-based approach to include data uncertainty within the LAMDA framework. Section 4 incorporates model uncertainty in MDA through a novel auxiliary variable approach, based on the probability integral transform. Based on the probability integral transform, section 5 develops a global sensitivity analysis approach for feedback-coupled MDA using the auxiliary variable concept. A mathematical example and an electronic packaging problem are discussed in Section 6 to demonstrate the proposed methodology. Concluding remarks are given in Section 7.

2. Likelihood-based Approach for Multidisciplinary Analysis (LAMDA)

This section briefly introduces the likelihood-based approach for MDA. Figure 1 is a diagram of a multidisciplinary system which consists of three analyses. A feedback analysis is required between analyses 1 and 2. The input vector is $\boldsymbol{x} = \{\boldsymbol{x_1}, \boldsymbol{x_2}, \boldsymbol{x_s}\}$ where $\boldsymbol{x_1}$ and $\boldsymbol{x_2}$ are the input vectors for each individual analysis, and $\boldsymbol{x_s}$ is shared by both analyses. Given a realization of $\boldsymbol{x}$, the interdisciplinary analysis between analyses 1 and 2 is conducted; the coupling variables, i.e. $\boldsymbol{u_{12}}$ and $\boldsymbol{u_{21}}$, will converge to particular values. A simplistic implementation of this iterative analysis is fixed point iteration (FPI). After convergence, each disciplinary analysis releases a subsystem output, i.e., g_1 and g_2, to analysis 3 to evaluate the system level output, i.e., f.

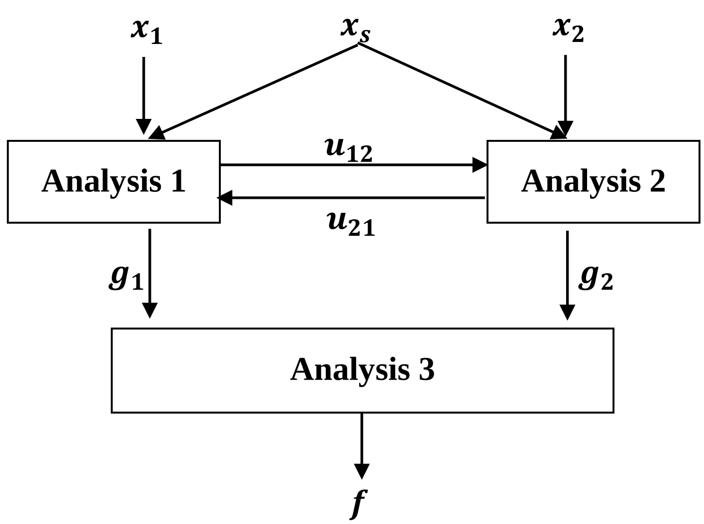


Figure 1: Multidisciplinary system

Figure 2 shows one iteration of the fully coupled analysis. This single iteration is denoted by a function G whose input is u_{12} and output is U_{12}, i.e.,

$$U_{12} = G(u_{12}, \boldsymbol{x}) = A_1(u_{21}, \boldsymbol{x}) \quad (1)$$

where $u_{21} = A_2(u_{12}, \boldsymbol{x})$. The input variable u_{12} is yielded by "Analysis 1" from the previous iteration, and the output U_{12} is the input of "Analysis 2" in the following iteration. Interdisciplinary compatibility is satisfied when $u_{12} = U_{12}$.

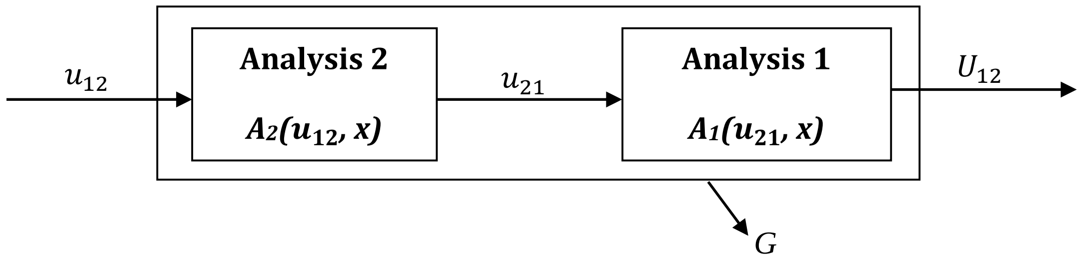


Figure 2: One iteration of coupled analysis

For a given value of u_{12}, when input variability is considered, the output U_{12} can be denoted by a probability density function: $f_{U_{12}}(U_{12}|u_{12})$. It is desired to calculate $P(U_{12} = u_{12} \mid u_{12})$, which is the probability of satisfying the interdisciplinary compatibility conditioned on u_{12}. This is similar to the definition of a likelihood function in parameter estimation problems where $P(y = y_{obs} \,|\theta)$ indicates the probability of observing the output to be equal to some value y_{obs} conditioned on the value of the parameter of interest θ. Thus here the likelihood of u_{12} may be defined as:

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

Note that likelihood is only meaningful up to a proportionality constant. The probability in Eq. (2) can be approximated by integrating the conditional PDF $f_{U_{12}}(U_{12}|u_{12})$ over an infinitesimal window around the conditional value of u_{12}:

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

where δ is the length of the window. In Ref [27], the integration in Eq. (3) is estimated by the first-order reliability method (FORM). FORM calculates the probability that a performance function $H \equiv h(\boldsymbol{x})$ is less than or equal to h_c, given stochastic input variables $\boldsymbol{x}$, which is equivalent to calculating the cumulative probability density (CDF) of H at $H = h_c$ [9]. Using this idea, FORM analyses are applied to calculate the integral in Eq. (3) at the upper and lower bounds, i.e., $h_u = F_{U_{12}}(U_{12} = u_{12} + \frac{\delta}{2}|u_{12})$ and $h_l = F_{U_{12}}(U_{12} = u_{12} - \frac{\delta}{2}|u_{12})$, which are essentially the probability of $(U_{12} < u_{12} + \frac{\delta}{2})$ and $(U_{12} < u_{12} - \frac{\delta}{2})$ respectively. Note that in implementing FORM for each u_{12}, only the feed-forward analysis of Figure 2 is needed to estimate G and its derivatives $\nabla G(\boldsymbol{x})$; i.e., in each iteration of FORM, only $\boldsymbol{x}$ is changing, not u_{12}. The likelihood of u_{12} is approximated by:

$$L(u_{12}) \propto \frac{h_u - h_l}{\delta} \tag{4}$$

The likelihood function only needs to be evaluated at a few points. Then the PDF of u_{12} can be evaluated as:

$$f(u_{12}) = \frac{L(u_{12})}{\int L(u_{12}) du_{12}} \tag{5}$$

A recursive adaptive version of Simpson's quadrature [29] can be used to evaluate the integral in Eq. (5). After evaluating the PDF for a few values of u_{12}, the entire PDF is approximated by interpolation. The LAMDA method is theoretically exact; but approximations are introduced in the numerical implementation by using FORM to calculate the CDF values in Eq. (3). However, the LAMDA framework is not dependent on FORM; if the analysis is nonlinear, SORM or one of several methods of Monte Carlo sampling can be used instead. The key point is that LAMDA only needs a single run through the two analyses for each realization of u_{12} and <u>not</u> an iterative analysis to convergence for each u_{12}.

Once the PDF of the converged value of u_{12} is constructed, the feedback-coupled analysis of Figure 1 can be replaced by a unidirectional coupled analysis as shown in Figure

3. The coupling variable u_{12} is brought to the same level with the input variable $\boldsymbol{x}$; both $\boldsymbol{x}$ and u_{12} can be treated as the input variable to this partially decoupled system. The uncertainty of the subsystem level and system level output can be characterized by sampling $\boldsymbol{x}$ and u_{12}; for each sample of input and u_{12}, only one function evaluation of Analyses 1 and 2 is required to compute g_1 and g_2. The results are then used to calculate f.

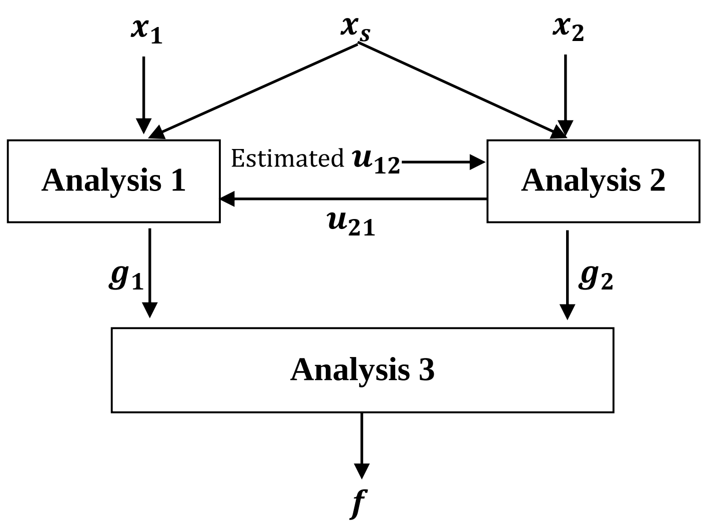


Figure 3: Multidisciplinary system: partially decoupled

Note that the paper's focus is on coupling between disciplines, and the input correlation is not considered. However, the problem of correlated input random variables has been solved long ago, and does not present any new challenge. When input correlation is considered, the input random variables can be transformed into a space of uncorrelated variables, after which FORM can be used to evaluate the likelihood.

3. Inclusion of Data Uncertainty in MDA

This section develops a likelihood-based approach to include data uncertainty regarding the input variables (due to sparse and/or imprecise data) within MDA.

3.1 Likelihood-based Approach for Data Uncertainty Representation

Data uncertainty can be regarding a stochastic or deterministic quantity. This paper focuses on the first type of uncertainty, i.e., only sparse and/or interval data is available on an input random variable. An enhanced LAMDA method that accounts for epistemic uncertainty regarding the input random variables is proposed here. This method combines data uncertainty due to sparse point data and interval data and develops a probabilistic representation for this uncertainty through a nonparametric PDF [10].

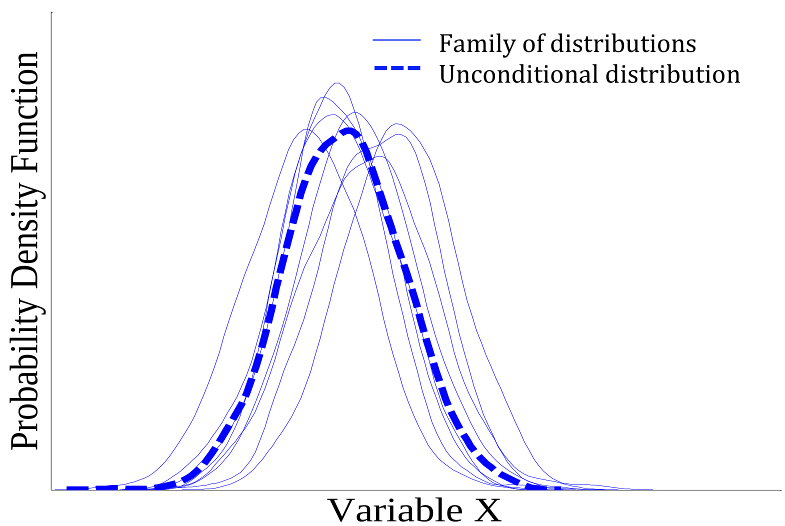


Figure 4: Family of distributions

Suppose the available information for a random variable X is a combination of m data intervals: $\{[a_1, b_1], \ldots [a_m, b_m]\}$, and n data points $\{s_1, \ldots s_n\}$. Based on the principle of likelihood, two approaches can be pursued to represent this type of uncertainty: parametric [18] and non-parametric [10]. In the parametric approach, a discrete random variable D and a random variable vector $\boldsymbol{\theta}$ are assumed to denote the distribution type and the distribution parameters respectively. Randomly sampling D and $\boldsymbol{\theta}$ will result in a family of probability distributions as shown in Figure 4. This family of distributions is used to fit the data interval and data points of $\boldsymbol{X}$, and the likelihood of D and $\boldsymbol{\theta}$ is given by:

$$L(D,\boldsymbol{\theta}) \propto \prod_{j=1}^{m}\left[F_X\left(b_j|D,\boldsymbol{\theta}\right) - F_X(a_j|D,\boldsymbol{\theta})\right]\prod_{i=1}^{n} f_X(s_i|D,\boldsymbol{\theta}) \qquad (6)$$

D and $\boldsymbol{\theta}$ may be estimated by maximizing the likelihood function in Eq. (6) (note that D is discrete). The candidate distribution types can sometimes be selected based on prior knowledge or physical considerations; in other cases, however, the choice of distribution type candidates may be difficult.

To avoid the assumption of distribution type, a non-parametric approach [10] can be adopted. Consider the variable $\boldsymbol{X}$ with m interval and n point data. The maximum and minimum values in this data are used as the upper and lower bounds of *X.* The entire domain is then uniformly discretized by a set of points q_i $(i = \{1, \dots\ Q\})$. Let p_i denote the PDF value at the i^{th} point, i.e., $f_X(x_i = q_i) = p_i$; the PDF over the entire domain can be constructed by interpolating these PDF values. Let $\boldsymbol{p}$ denote the vector of the PDF values, i.e., $\boldsymbol{p} = (p_1, \dots p_Q)$; the likelihood function of $\boldsymbol{p}$, which is defined as the probability of observing the given data (point values and data intervals) given $\boldsymbol{p}$, can be written as:

$$L(\boldsymbol{p}) \propto \prod_{j=1}^{m}\left[F_X\left(b_j|\boldsymbol{p}\right) - F_X(a_j|\boldsymbol{p})\right]\prod_{i=1}^{n} f_X(s_i|\boldsymbol{p}) \qquad (7)$$

The value of $\boldsymbol{p}$ can be estimated by maximizing the likelihood $L(\boldsymbol{p})$ using the optimization problem in Figure 5. The three constraints for the optimization are: (1) the vector $\boldsymbol{p}$ (PDF values at the discretized points) needs to be positive; (2) the PDF value over the entire domain of X needs to be positive; and (3) the integrated area under the PDF curve must be unity. A Gaussian Process (GP) interpolation technique is employed in this paper to fit the entire PDF curve based on $[p_1, \dots p_Q]$; however, other interpolation techniques may also be used.

$$\max_{\boldsymbol{P}} L(\boldsymbol{P})$$

$$\text{s.t. } p_i \geq 0 \text{ for } \forall\, p_i \in \boldsymbol{P}$$

$$f_x(x) \geq 0 \text{ for } \forall\, x$$

$$\int f_x(x)dx = 1$$

Figure 5: Optimization framework for maximum likelihood method

The above likelihood-based method is exploited to fit a non-parametric probability distribution in this paper to include the effect of data uncertainty due to sparse and interval data. It avoids assumptions on the distribution type or distribution parameter. The resulting probability distribution can be easily applied to uncertainty propagation with Monte Carlo sampling or FORM.

4. Inclusion of Model Uncertainty in MDA

4.1 Model Error Quantification

Liang and Mahadevan developed approaches for model error quantification in feed-forward computational models [20]; however, the propagation of model error through multiple models is not straightforward in feedback-coupled MDA. Model errors can be classified into two categories: (1) model form error caused by simplifications or assumptions about the physics of the problem, and (2) numerical errors due to the solution process, such as discretization error, error due to limited sampling, etc. The quantification methods for different types of model errors are distinct from each other. Model form error can be estimated using actual experimental data; and numerical solution error can be calculated using the result of model verification. When input variability and data uncertainty are considered, the model errors need to be quantified at each input realization. This section focuses on the inclusion of model error within MDA in a generalized manner that includes both model form error and numerical solution errors.

A simple way to handle input-dependent model error is to use an additive model discrepancy term and include it in subsequent analysis. Kennedy and O'Hagan [21] used Bayesian calibration to quantify this model discrepancy term. Mahadevan and Rebba [30], and Chen [31] included the additive model error term in reliability-based design optimization. However, when multiple sources contribute to model error, and when these sources do not combine in a simple manner, the additive term approach is not easy to use; Sankararaman et. al [32] used a Bayesian network approach to combine multiple sources of model error. However, complication arises in feedback-coupled MDA if the model error term has to be added after each iteration of individual disciplinary analyses. Also, model error is a function of the input and this function is not generally known; thus, it is not straightforward to include the additive model error term in feedback-coupled MDA.

In many problems, the original disciplinary analyses may be expensive, and may need to be replaced by surrogate models. Many types of surrogate modeling techniques are available (e.g., Gaussian process models [21], polynomial chaos expansion [33], support vector regression [34], artificial neural network [35], etc.). A review of state-of-the-art modeling techniques for solving different types of optimization problems is provided in [35]. Use of a surrogate model introduces error in the prediction, which has two components: bias and variance. A leave-one-out cross validation approach can be used to estimate bias with an existing number of training points [36], and sequential training point selection techniques have been proposed in the literature for bias reduction [36, 37]. Expressions for variance of the surrogate model prediction are also available in the literature (for example, Gaussian process model variance is readily available from the GP equations, and for polynomial chaos expansion see [20]).

Regardless of whatever individual or multiple sources contribute to the model error, the output of a model due to the presence of model error is a probability distribution even for a fixed value of the input. Note that some errors are deterministic (e.g. discretization error) and some are stochastic (e.g. surrogate model error); their combined effect makes the model output stochastic even for a fixed input. This presents an interesting challenge; note that only aleatory uncertainty was considered in the original LAMDA method (Section 2). Therefore, for a given value of u_{12} and $\boldsymbol{x}$, the output U_{12} was a deterministic value. However, in the presence of model error, the output U_{12} becomes a probability distribution. This makes it difficult to evaluate Eq. (2): How can we talk about the probability of a distribution being equal to a particular value? The likelihood in Eq. (2) can only be calculated when the output U_{12} is a deterministic value for a given value of u_{12} and $\boldsymbol{x}$. An auxiliary variable method is proposed below to overcome this challenge.

4.2 Auxiliary Variable Method

For the sake of illustration, consider a normal random variable X with uncertain parameters. Assume that the parameters of X, i.e. μ_X and σ_X, have normal distributions; the uncertainty of X is therefore denoted as: $X\sim N(\mu_X(\mu_\mu,\sigma_\mu),\sigma_X(\mu_\sigma,\sigma_\sigma))$, where $\mu_\mu,\sigma_\mu,\mu_\sigma$ and σ_σ are deterministic values based on sources such as experts opinion. Given a realization of μ_X, and σ_X, X is a distribution. Let P denote an auxiliary variable, defined by the probability integral transform [38] as

$$P=\int_{-\infty}^{x} f_X(X|\mu_X,\sigma_X)dX \tag{8}$$

where x is a generic realization of X. $P\in[0,1]$ is the CDF value. For a realization of μ_X and σ_X, the well-known inverse CDF method of Monte Carlo simulation taken over a realization of P from a uniform distribution gives a fixed value of $\boldsymbol{x}$. Therefore, $\boldsymbol{x}$ can be written as:

$$\boldsymbol{x}=F^{-1}_{X|\mu_X,\sigma_X}(p) \text{ or } \boldsymbol{x}=H'(p,\mu_X,\sigma_X) \tag{9}$$

where p is a realization of P. Thus by introducing the auxiliary variable P, we get a unique value of X for a given value of μ_X and σ_X. The probability integral transform helps to define the auxiliary variable P, and will be used to include the stochastic model error in coupled MDA.

Note that this approach can also be extended to handle the case when a parametric family of distributions is used to represent an input random variable due to data uncertainty. In that case, if a discrete variable D represents the distribution type, and Θ represents the vector of distribution parameters, then a unique value of x can be obtained for a realization of D, θ and P as:

$$x = H'(p, d, \theta) \tag{10}$$

Note that H' is not the same in Eqs. (9) and (10).

4.3 Representation of Model Uncertainty

For a given value of u_{12} and input $\boldsymbol{x}$, the output U_{12} follows a probability distribution due to model error. This distribution can be represented by a conditional PDF $f_{U_{12}}(U_{12}|u_{12}, \boldsymbol{x})$. Let auxiliary variable P denote the conditional CDF at $U_{12} = u_{12}$ given u_{12} and $\boldsymbol{x}$, i.e.,

$$P = \int_{-\infty}^{u_{12}} f_{U_{12}}(U_{12}|u_{12}, \boldsymbol{x}) dU_{12} \tag{11}$$

where $P \in [0,1]$. For a given value of input $\boldsymbol{x}$ and u_{12}, when a single value of P is sampled from $U(0,1)$, a unique value of U_{12} can be obtained through the inverse CDF method. Hence $U_{12} = H'(p, u_{12}, \boldsymbol{x})$, which is deterministic, can now be used to evaluate the likelihood in Eq. (2) using FORM, as shown in Figure 6. With a unique value of U_{12} defined as above, two evaluations of FORM are implemented at $C = u_{12} + \frac{\delta}{2}$ and $C = u_{12} - \frac{\delta}{2}$ to get h_u and h_l respectively. In FORM, $Prob(H' - C \leq 0) = \Phi(-\beta)$.

Given PDFs of $\boldsymbol{x}, P$
Min $\beta = \boldsymbol{u}^T\boldsymbol{u}$
s.t. $H'(\boldsymbol{p}, \boldsymbol{x}, u_{12}) - C = 0$
where standard normal $\boldsymbol{u} = T(p, \boldsymbol{x})$

Figure 6: FORM with auxiliary variable

The PDF of u_{12} can then be obtained using Eq. (4) and Eq. (5).

The auxiliary variable approach to include model error in MDA offers several benefits: (1) the auxiliary variable P represents the overall effect of model error in a generalized manner; no matter how different types of model errors are combined, it considers the overall distribution of the output as a result of these error sources for a fixed input. (2) The use of the auxiliary variable provides an elegant method to include model error in the LAMDA method, and the challenge of accumulating model error through multiple iterations of MDA is bypassed due to the single iteration strategy of LAMDA. (3) The use of the probability integral transform to define the auxiliary variable provides a theoretically exact way to include model error in feedback-coupled MDA. (4) Representation of the model error through a random variable P brings x and P on the same level, and facilitates a single loop approach to implement FORM, thus providing computational efficiency. In contrast, a sampling-based approach to incldue model error would need an additional nested loop of analysis.

5. Global Sensitivity Analysis in Feedback-coupled MDA

Uncertainty propagation analysis is often accompanied by sensitivity analysis to identify the significant contributors to the model output uncertainty. Several benefits are possible such as: (1) reduction of number of uncertainty sources considered in the analysis and design optimization; (2) guidance in resource allocation for data collection; and (3) guidance in model refinement. Global sensitivity analysis has been used to calculate the effect of the variability of an input quantity on the variance of the output quantity [28]. Consider a model given by:

$$Y = G(X_1, X_2 \dots X_n) \tag{12}$$

where X_i and Y are input-output pairs of a generic model. The first-order sensitivity indices are estimated as:

$$S_i^1 = \frac{V_{X_i}(E_{X_{\sim i}}(Y|X_i))}{V(Y)} \tag{13}$$

where the notation $E_{X_{\sim i}}(Y|X_i)$ denotes the expectation of output Y given a particular value of variable X_i and considering the random variations of all other variables except for X_i (denoted by $X_{\sim i}$). The symbol V_{X_i} represents the variance of the aforementioned expectation over multiple samples of X_i. The first-order sensitivity index indicates the contribution of uncertainty due to a particular individual variable, regardless of its interactions with other variables. The evaluation of Eq. 1 can be accomplished by either double-loop or single-loop Monte Carlo sampling. The sum of first-order indices of all variables is always less than or equal to unity.

A total effects index is also calculated to account for the uncertainty contribution of X_i in combination with all other variables:

$$S_i^T = \frac{E_{X_{\sim i}}(V_{X_i}(Y|X_{\sim i}))}{V(Y)} = 1 - \frac{V_{\sim i}\left(E_{X_i}(Y|X_{\sim i})\right)}{V(Y)} \tag{14}$$

where $V_{X_i}(Y|X_{\sim i})$ denotes the variation of output Y at a fixed realization of all variables except X_i , over multiple samples of X_i only; $E_{X_{\sim i}}$ calculates the expectation of this variation over multiple samples of $X_{\sim i}$. The sum of the total effects indices of all variables is always greater than or equal to unity.

Previous work in GSA has only considered aleatory uncertainty in the input variables [20]. When model output uncertainty is caused by input variability, data uncertainty and model errors, the contribution of all the sources needs to be quantified. The sensitivity to the input variable distributions is straightforward to calculate by using sampling techniques. However, when uncertainty caused by stochastic model errors is considered, the GSA cannot be directly implemented due to the lack of a deterministic input-output transfer function.

Consider Eq. 2 and Eq. 3, in which the inner loops of sampling calculate $E_{X_{\sim i}}(Y|X_i)$ and $V_{X_i}(Y|X_{\sim i})$ respectively; both evaluations require deterministic function output. In the presence of model uncertainty, the output $\boldsymbol{Y}$ is a distribution even for a fixed input $\boldsymbol{X}$. Therefore, a new auxiliary variable is introduced to explicitly include model uncertainty in sensitivity analysis. Consider the model in Eq. 4. Suppose model output has a distribution D at a given input $\boldsymbol{X}$; this can be denoted as: $Y \sim D(\mu_{pred}(\boldsymbol{X}), \sigma_{pred}(\boldsymbol{X}))$, where μ_{pred} is the predicted mean function value and σ_{pred} is the standard deviation that represents the uncertainty in the prediction due to model uncertainty. Figure 7 shows the stochastic functional relation between input $\boldsymbol{X}$ and output $\boldsymbol{Y}$.

Figure 7: Output uncertainty due to model errors

Let U_Y denote the auxiliary variable which is defined as $U_Y = \int_{-\infty}^{y} f_Y(Y|\boldsymbol{X})dY$, where $\boldsymbol{X}$ is one realization of the input and f_Y is the PDF of Y conditioned on $\boldsymbol{X}$. The individual and total effects of U_Y are:

$$S_{U_Y}^{1} = \frac{V_{U_Y}(E_X(Y|U_Y))}{V(Y)} \tag{15}$$

$$S_{U_Y}^{T} = \frac{E_X(V_{U_Y}(Y|X))}{V(Y)} = 1 - \frac{V_{U_Y}\left(E_X\left(Y|X_{U_Y}\right)\right)}{V(Y)} \tag{16}$$

Thus, the use of the auxiliary variable method in variance-based GSA provides an explicit means to quantify the contribution of the stochastic model error to the system level output variance. It represents the effect of model uncertainty through the auxiliary random variable U_Y, and brings model uncertainty to the same level of analysis as input uncertainty. The sensitivity of U_Y given in Eq. 5 and Eq. 6 can be regarded as an index of the contribution of model error to the overall output uncertainty.

In summary, Sections 3 to 5 introduced the representation and the propagation of data uncertainty and model uncertainty in coupled MDA. Likelihood-based parametric and non-parametric approaches to handle data uncertainty were presented. Model error sources result in a stochastic model output; and an auxiliary variable method is introduced to account for the model error through a random variable which provides a breakthrough in the implementation of both LAMDA and GSA to feedback coupled MDA. Since variability and data uncertainty of a random variable are together represented by a non-parametric distribution in this paper, the combined effect of aleatory and epistemic uncertainty in each input random variable is identified by a single sensitivity index. However, if a variable has a significant impact on the output uncertainty, and separation of the effects of aleatory and epistemic uncertainty is desired, then refer [39] for details of such analysis.

6. Numerical Examples

A mathematical MDA example is considered in this section first. Two assumptions for model error are made for the sake of illustration and propagated using the enhanced LAMDA approach. Next, an electronic packaging example is used to demonstrate the quantification and propagation of different sources of uncertainty in MDA using the proposed approach.

6.1 Mathematical MDA Problem

The mathematical example shown in Figure 8 consists of 3 analyses. A feedback coupling exists between “Analysis 1” and “Analysis 2”, and the coupling variables are denoted as u_{12} and u_{21}. Then the subsystem output g_1 and g_2 are calculated and used as the inputs to analysis 3 to compute the system level output f. The input variables x_1, x_2 and x_3 are assigned normal distributions: $N(1{,}0.1)$. x_4 is characterized by a lognormal distribution: $LogN(1{,}0.1)$.

6.1.1 Epistemic Uncertainty Due to Insufficient Data

Epistemic uncertainty is assumed for x_5. The data is available in the form of 3 intervals {[0 2], [0.02 1.97], [0.14 1.89]} and 2 point values {0.99, 1.02}. The domain bounded by the maximum and minimum available values of the available data is divided into 10 equally spaced points, with PDF values $P_i, i = 1 \dots 10$. The optimization framework in Figure 5 is then adopted to estimate the optimal P_i that maximizes the likelihood function constructed using Eq. 7. A cubic spline technique is employed to interpolate the likelihood and construct the non-parametric PDF presented in Figure 9.

6.1.2 Epistemic Uncertainty Due to Model Errors

For the sake of illustration, model errors ME_1 and ME_2 are assumed in “Analysis 1” and “Analysis 2” respectively as functions of the input and coupling variables. Two cases of model error are addressed: (1) deterministic model error, and (2) stochastic model error. The assumed mathematical forms of the model errors are listed in Table 1. Additionally, results are also computed for the case with no model error, for the sake of comparison.

6.1.3 Uncertainty Quantification of the Coupling Variables

For each coupling variable in both cases, the entire PDF is estimated by interpolating 15 integration points, each of which has been evaluated by LAMDA (Eq. 7). The propagation of variability and data uncertainty for deterministic model errors can be fulfilled by the original LAMDA method. In the stochastic error scenario, the proposed auxiliary variable method is used to address the model error. Two auxiliary variables h_1 and h_2, representing the CDFs of the model errors respectively, are introduced; both are uniformly distributed from 0 to 1 based on the probability integral transform (Section 4). The resultant PDFs of the coupling variables with deterministic and stochastic model errors and without error are shown in Figure 10. The mean and standard deviation of u_{12} and u_{21} are calculated and listed in Table 2. SOFPI is implemented with 20,000 samples of the input as the benchmark solution.

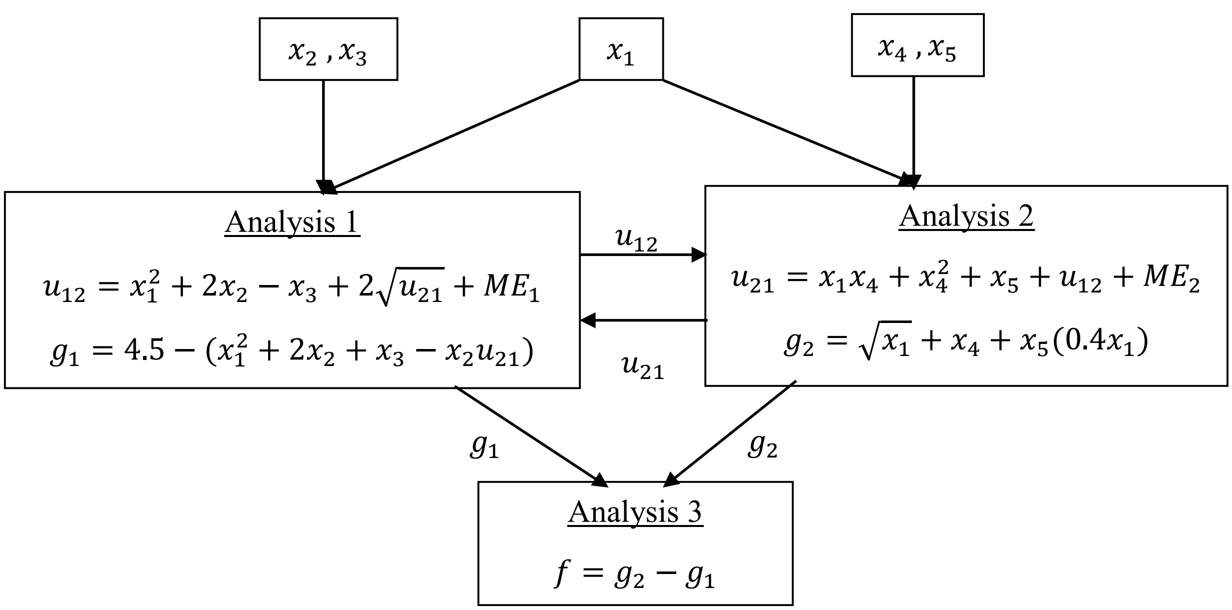


Figure 8: Functional relations of the mathematical MDA model

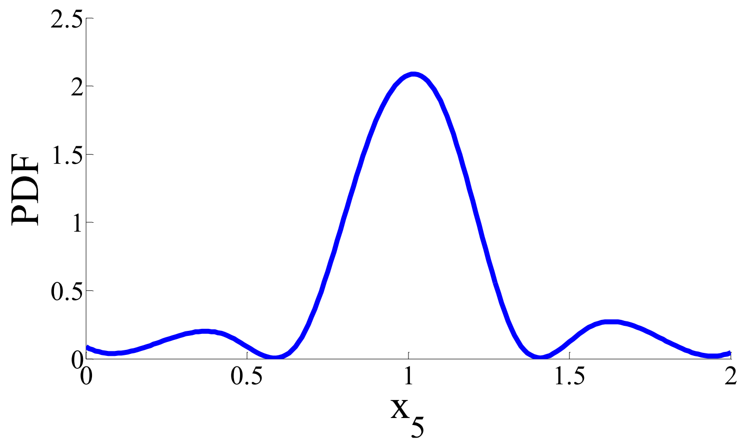


Figure 9: Non-parametric PDF of x_5

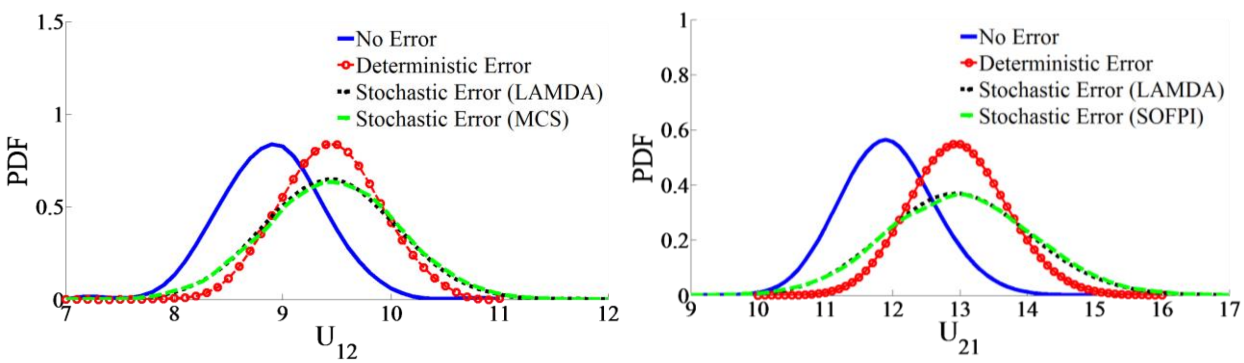


Figure 10: PDFs of coupling variables U_{12} (left) and U_{21} (right)

Table 1: Model errors in coupled analysis

Model Error	**Deterministic Model Errors**	**Stochastic Model Errors**
ME_1	${ME_1}^D = 0.05x_2 + 0.1(u_{21})^{\frac{1}{4}}$	$\mu({ME_1}^S) = {ME_1}^D$ $\sigma({ME_1}^S) = 0.15 * {ME_1}^D$
ME_2	${ME_2}^D = 0.1\sqrt{u_{12}} + 0.2x_5$	$\mu({ME_2}^S) = {ME_2}^D$ $\sigma({ME_2}^S) = 0.15 * {ME_2}^D$

Table 2: Mean and standard deviation of coupling variables

Case No.	Type		u_{12}	u_{21}	No. of Function Evaluations
1	**No model error**	μ_N	8.95	11.94	572
		σ_N	0.49	0.72	
2	**Deterministic model error**	μ_D	9.40	12.99	594
		σ_D	0.48	0.73	
3	**Stochastic model error (LAMDA)**	μ_L	9.45	12.98	768
		σ_L	0.62	1.08	
4	**Stochastic model error (SOFPI)**	μ_S	9.46	13.03	379,246
		σ_S	0.62	1.10	

The following observations are drawn from Table 2:

(1) The PDFs of u_{12} and u_{21} for cases 1 and 2 have almost the same shape and standard deviation, and are only separated by the deterministic model error value.

(2) Including the stochastic model error increases the computational effort by 34.3% comparing with the "No model error" case, and 26.3% with the "Deterministic model error case". This is because the introduced auxiliary variables increase both the dimension and the nonlinearity of the problem. This case can be finished in less than 1 second. On the other hand, 20,000 SOFPI evaluations take 493.1 seconds and 379,246 function evaluations.

(3) Once the PDF of the coupling variable is calculated, the scheme in Figure 3 can be used for uncertainty propagation and estimate the PDF of the individual disciplinary and system outputs: g_1, g_2 and f. Note that this does not require the iterative analysis between Analyses 1 and 2, therefore becomes a simpler uncertainty propagation through a feed-forward analysis. For the sake of illustration, Monte Carlo sampling is used to estimate the PDF of the system output. Following the scheme in Figure 3 the analysis in the other direction is retained. The PDFs of the inputs $\boldsymbol{x}$ and u_{12} are used first

in “Analysis 2” to estimate u_{21} and g_2, and then in “Analysis 1” to estimate g_1, followed by the overall system output f. Table 3 lists the mean values and standard deviations of the outputs in different cases.

Table 3: Mean and standard deviation of g_1, g_2 and f

		$\boldsymbol{g_1}$	$\boldsymbol{g_2}$	$\boldsymbol{f}$
No Model Error	$\boldsymbol{\mu}$	12.50	2.41	-10.1
	$\boldsymbol{\sigma}$	1.2	0.16	1.18
Deterministic Case	$\boldsymbol{\mu}$	13.52	2.41	-11.11
	$\boldsymbol{\sigma}$	1.26	0.16	1.23
Stochastic (LAMDA)	$\boldsymbol{\mu}$	13.60	2.43	-11.17
	$\boldsymbol{\sigma}$	1.42	0.16	1.40
Stochastic (MCS)	$\boldsymbol{\mu}$	13.49	2.41	-11.08
	$\boldsymbol{\sigma}$	1.60	0.15	1.55

(4) It could be argued that in the presence of expensive disciplinary computational models, SOFPI could be used with surrogate models. However, building the surrogate model has computational expense. In the mathematical example, an average of 19 function evaluations is needed for deterministic MDA to converge at each input. To obtain the training points for the surrogate model, such coupled analysis needs to be evaluated at multiple realizations of the input. The number of training points can be very high if the input is highly dimensional and the model is highly nonlinear. Therefore, the total number of function evaluations will still be large. Thus the computational effort in building the surrogate model should also be considered in such comparisons.

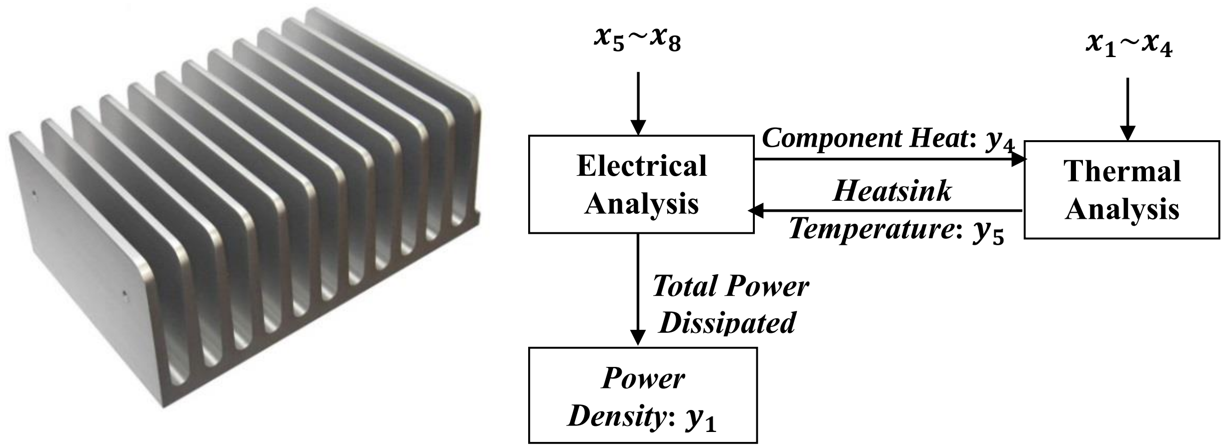
Global sensitivity analysis is conducted to quantify the sensitivity of the system level output f to the uncertain inputs and model errors from both analyses. The auxiliary variables h_1 and h_2 denote the uncertainty introduced by the model errors. The input, coupling and auxiliary variables are sampled, then the decoupled analysis is executed to evaluate the output uncertainty. Therefore, the total number of function evaluations equals the number of variables (input/coupling/ auxiliary) times the sample size (1,000 samples are used), which equals 8,000. The first-order sensitivity indices and the total-effect indices are shown in Table 4.

Table 4: Global sensitivity indices

	$\boldsymbol{x_1}$	$\boldsymbol{x_2}$	$\boldsymbol{x_3}$	$\boldsymbol{x_4}$	$\boldsymbol{x_5}$	$\boldsymbol{h_1}$	$\boldsymbol{h_2}$
First-Order	0.007	0.670	0.019	0.036	0.019	0.039	0.180
Total Effect	0.007	0.681	0.019	0.037	0.021	0.041	0.184

The index for h_2 indicates that the stochastic model error from “Analysis 2” has a large impact on the uncertainty of the final system output, while model error from “Analysis 1” has a small effect. The use of the auxiliary variable method enabled the sensitivity analysis to include uncertainty contributions from model errors. It replaced the double-loop approach with a single-loop calculation, thus greatly reducing the computational effort. According to Table 4, it can be observed that the first-order and total effect sensitivity indices of the corresponding variables are quite similar. Since the total effect reflects the uncertainty significance of a variable from both its individual variation and its interactive effect with other variables, the result indicates that the collaborative effect of the variables is insignificant.

6.2 Electronic Packaging Example



(a) Geometry of a regular heatsink (b) Disciplinary analyses and coupling variables

Figure 11: Electronic packaging problem: feedback-coupled MDA

The electronic packaging problem [40] is a two-discipline analysis with feedback coupling between electrical and thermal analyses. The system is composed of a circuit with a single resistor and a heat sink on which the resistor is mounted. A diagram of the heatsink is shown in Figure 11(a). When the circuit is turned on, the resistor generates heat that is dissipated by the heatsink. The component resistance is affected by the operating temperature, while the temperature depends on the heat produced by the resistor. The interdisciplinary relationships are shown in the Figure 11(b). The deterministic parameters are: Voltage = 10.0 volts and room temperature $T = 20.0$ ℃. The random variables together with their uncertainty are listed in Table 5. The state variables are defined by the relations: $y_1 = \frac{y_5}{y_7}$; $y_2 = x_5\,(1 + x_6\,(y_5 - T))$; $y_3 = \frac{Voltage}{y_2}$; $y_4 = {y_3}^2\,y_2$; $y_5 = Thermal(y_4, x_1, x_2, x_3, x_4)$ is an implicit function of the geometric parameter of the heatsink and the power dissipation in resistor, and $y_6 = x_1\,x_2\,x_3$. The distribution types and parameters are assumed to be precisely known for x_1 to x_5, whereas both are uncertain for x_6 due to the availability of only interval data and sparse point data. A non-parametric PDF is constructed using the likelihood-based approach to combine the aleatory and epistemic uncertainty of x_6. The coupling variables are component heat y_4 (due to power dissipation in resistor) computed in the electrical analysis, and component temperature y_5 estimated in the thermal analysis. The system output power density is the ratio between total power dissipated and the volume of the heatsink. For the purpose of illustration, all the inputs are assumed to be independent variables.

Table 5: Parameters of the electronic packaging system

Parameter		Parameter (Unit)
Input Variables and Associated Uncertainty	x_1	Heat sink width (m) $\sim N(0.1,0.01)$
	x_2	Heat sink length (m) $\sim LogN(0.1,0.01)$
	x_3	Fin length (m) $\sim N(0.05,0.005)$
	x_4	Fin width (m) $\sim N(0.025,0.0025)$
	x_5	Nominal resistance at temperature T° (W) $\sim N(10,0.1)$
	x_6	Temperature coefficient of electrical resistance (°K^{-1}) data Intervals: [0.004,0.009],[0.0043,0.0085],[0.0045,0.0088] data Points: {0.0055, 0.0057}
Thermal and Electrical State Variables	y_1	Power density (watts/m^3)
	y_2	Resistance at temperature T_1 (W)
	y_3	Current in resistor (amps)
	y_4	Power dissipation in resistor (watts)
	y_5	Component temperature (T_1) of resistor (°C)
	y_6	Heat sink volume (m^3)

6.2.1 Model Error Quantification

The two disciplinary analyses (electrical and thermal) are evaluated using two different mathematical models. The electrical analysis is solved algebraically based on electrical circuit analysis and the computational process is straightforward. In the thermal analysis, the component temperature y_5 is retrieved by numerically solving a two-dimensional heat transfer differential equation using a finite difference method. Due to limited computational resources for solving the continuum problem, assume that only a coarse mesh can be used, causing discretization error. Meshes are only required for x and y directions (heat transfer in the thickness direction is ignored for the thin plate). A Gaussian Process-based technique (See Appendix I) is used to estimate discretization error [41] in FDA/FEM analysis as an enhancement of the traditional Richardson extrapolation method. The basic theory of the Gaussian Process (GP) technique is given in Appendix I. The GP approach to quantify discretization error is briefly summarized below:

For a given input x_n, T mesh tests : h_{set}= $\{h^1, ... h^T\}$ are conducted, where h^i denotes a particular mesh size combination. The associated model outputs, i.e. g_{raw} = $\{g_{raw}(x_n, h^1)$,..., $g_{raw}(x_n, h^T)\}$, are then collected. A GP model is constructed using the mesh sizes and the corresponding outputs $\{h_{set};g_{raw}\}$. The corrected estimate of the function value at input x_n is then predicted at h = 0, i.e., the function value is estimated at an infinitesimal mesh size. In the electronic packaging application, the finest affordable mesh size 0.005; Therefore, three mesh tests: h_{set} = {0.007, 0.006, 0.005} (same mesh in both x and y directions) are conducted; the mesh sizes together with the output component temperatures are used to train the GP model; after that, the heatsink temperature at h=0 is predicted using this GP model. Figure 12 is a demonstration of GP getting trained by three data points (circle dot) and prediction at $h = 0$. The red square dot represents the mean prediction by the GP model and the green dashed lines are the 95% bounds. The Richardson extrapolation method is also applied under three mesh tests: h_R = {0.0072, 0.006, 0.005} (the mesh refinement ratios need to be constant), and its result denoted by the blue square dot agrees well with the GP prediction. Even though the true function value is deterministic, the GP prediction quantifies the uncertainty in estimating it. This uncertainty is epistemic uncertainty due to the finite number of training points; as the number of training point increases, this uncertainty will be reduced.

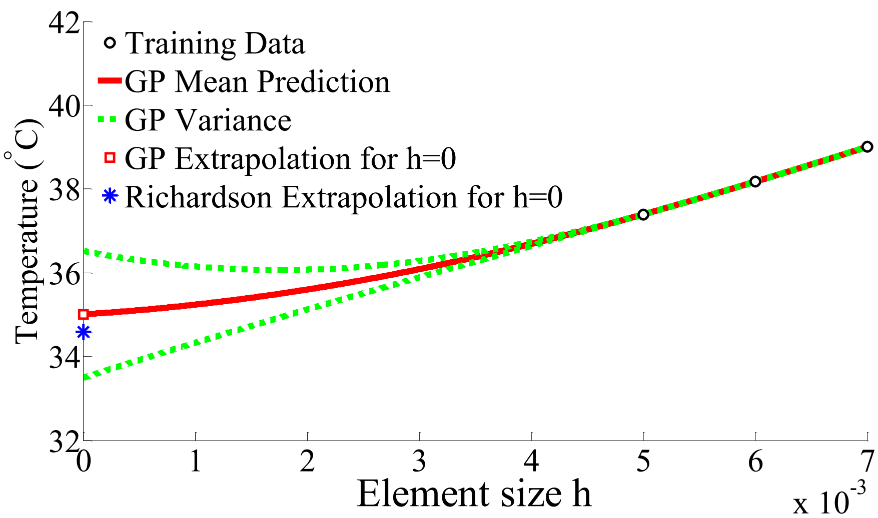


Figure 12: Mean and 95% bound of GP prediction, accounting for discretization error (Thermal analysis)

In stochastic MDA, discretization error needs to be quantified at each input realization using GP extrapolation. As mentioned above, the predictions of GP at mesh size $h = 0$ include a predicted mean value, and variance that indicates the uncertainty in the prediction as shown in Figure 12. The presence of the stochastic model prediction even for a single fixed input value poses a challenge for uncertainty propagation in coupled MDA. Consider one iteration of the coupled analysis in Figure 11(b). Given one realization of the inputs $\boldsymbol{x}$, the output temperature $y_5 = Thermal(y_4, \boldsymbol{x})$ must be deterministic where T denotes the thermal analysis; however, when y_5 is evaluated using a GP, the outcome will be accompanied with variability; the mean and standard deviation of the output, which are determined using Eq. 8 and Eq. 9, are functions of $\boldsymbol{x}$ and y_4:

$$f(y_5|\boldsymbol{x}) \sim N(\mu(\boldsymbol{x}, y_4), \sigma(\boldsymbol{x}, y_4)) \tag{17}$$

For each realization of $\boldsymbol{x}$ and y_4, the epistemic uncertainty due to model error will lead to a family of distributions for y_5. Since FORM requires a deterministic output from the performance function, the stochastic GP prediction cannot be directly used in the LAMDA method.

Therefore an auxiliary variable $P_h \sim U[0,1]$ is defined. With the auxiliary variable P_h, a unique value of the prediction can be determined through inverse CDF. Therefore, the model output becomes a deterministic function of $\boldsymbol{x}, y_4$ and h, and the LAMDA approach can be implemented using FORM as in Figure 6. Two different cases, with different model error assumptions are considered:

Case 1 (No model error): No assumption of model error is made for the electrical analysis; and for the thermal analysis, the temperature value is evaluated using the finest mesh within the limitation of computation resources. The uncertainty sources are only the six input variables; note that x_6 has both aleatory and epistemic uncertainty, whereas x_1 to x_5 have only aleatory uncertainty (fixed distribution type and distribution parameters).

Case 2 (Stochastic model error): The discretization error of thermal analysis is quantified using GP. The resulting uncertainty is then included in LAMDA using an auxiliary variable. The sources of uncertainty being considered are 5 aleatory inputs, 1 input with both aleatory and epistemic uncertainty, and the model prediction uncertainty due to discretization error. Note that the discretization error is actually deterministic, but there is uncertainty in estimating it because of a small number of mesh sizes tested. This uncertainty is expressed by the variance of the GP prediction of temperature at $h = 0$. And in the MDA and sensitivity analysis, this uncertainty is represented by the auxiliary variable P_h.

6.2.2 PDF of the Coupling Variables and System Output

The PDFs of temperature and component heat are estimated for both cases and are shown in Figure 13. The system level output, power density, is calculated as:

$$Power\ density = \frac{Component\ Heat\ (y_4)}{Heatsink\ Volume\ (x_1 \times x_2 \times x_3)} \tag{18}$$

where x_1, x_2, x_3 are the geometric parameters. Its distribution is evaluated using Monte Carlo simulation for illustration. Samples of component heat y_4 together with x_1, x_2 and x_3 are generated independently and used to calculate power density using Eq. (18). Figure 13 compares the marginal PDFs of temperature and component heat under two model error assumptions; the PDFs of power density are compared in Figure 14. The first and second moments of the PDF are compared in Table 6.

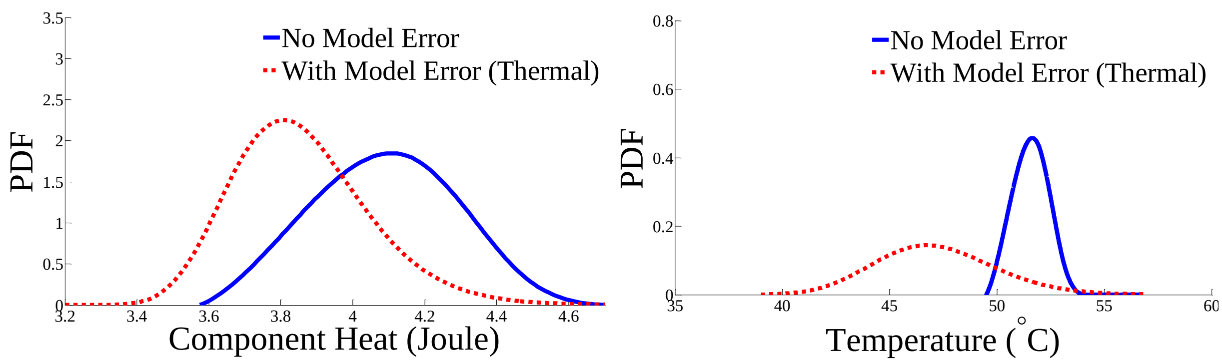


Figure 13: PDF of coupling variables: component heat y_4 (left) and temperature y_5 (right)

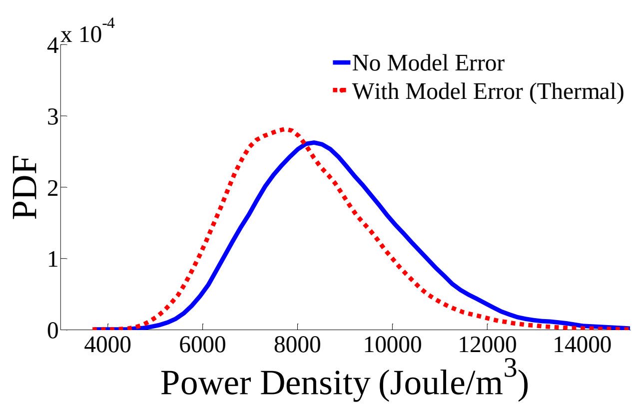


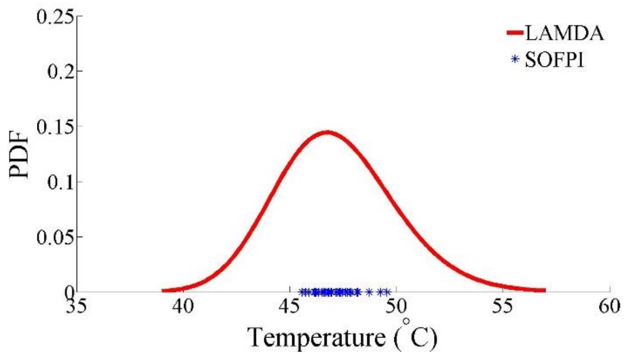
Figure 14: PDF of system output: power density

Table 6: Mean and standard deviation of temperature and power density

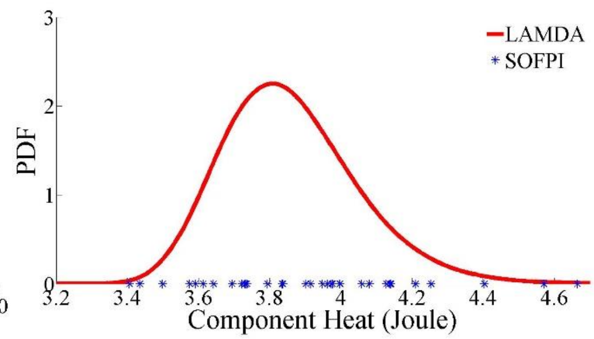
		Temperature	Component Heat	Power Density
No Model Error	**Mean**	52.24	4.13	8757.82
	STD	1.06	0.21	1618.00
Stochastic Model Error	**Mean**	47.13	3.98	8080.00
	STD	2.80	0.23	1507.09

The number of function evaluations in the LAMDA method when considering the model error and stochastic model output is 1219, whereas 910 evaluations are needed when no error is considered. When the disciplinary analyses are computationally cheap, SOFPI can be used to generate the benchmark solution (the entire PDF) for LAMDA to compare with, as shown in the earlier mathematical example. However, when the disciplinary analyses are expensive, it may not be affordable to generate the entire PDF using SOFPI. In such a situation, SOFPI could be run for a few samples of input realizations, and the SOFPI outputs can be compared against the PDF generated by LAMDA.

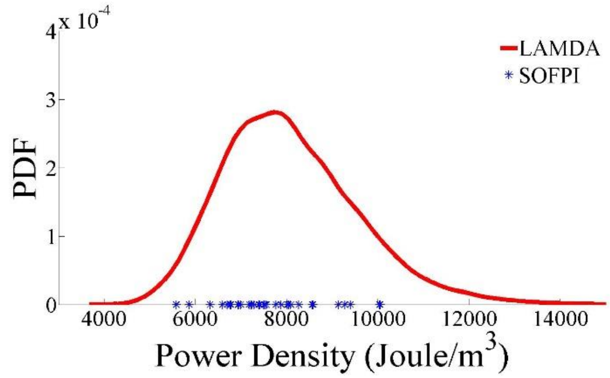
Figure 15 compares SOFPI results for 35 input realizations against the LAMDA-generated PDF for the coupling variables and the system level output. It is seen that the SOFPI results are within the range of the LAMDA-generated PDF. In addition to a graphical comparison, model validation techniques can also be used for a quantitative comparison; several such techniques are well studied in the literature [42, 43].



(a)



(b)



(c)

Figure 15: Comparison of results from LAMDA and SOFPI for (a) Temperature, (b) Component heat, and (c) Power density

6.2.3 Results and discussion

According to Figure 12, for a given input $\{x_1 \dots x_6\}$, the predicted temperature decreases as the meshes become finer. This phenomenon agrees well with the PDFs of the temperature in Figure 13, where the distribution for the stochastic model error case shifts to the left compared with the no model error case. When model error is included, all subsystem outputs have greater variances as expected. In addition, the model error appears to cancel the effect of input variability and data uncertainty and lead to a smaller final output uncertainty. A GSA is implemented to quantify the sensitivity of heat to the input uncertainty and model error. The auxiliary variable P_h represents the uncertainty due to the GP-based estimation of discretization error (i.e., discretization error in thermal analysis). The total number of function evaluations mostly depends on how fast FORM converges. When the number of input, coupling and auxiliary variables is small, and if the analysis is linear, FORM converges quickly and the number of function evaluations is small. On the other hand, if the input and coupling variables are high dimensional, and if more auxiliary variables are used (i.e., more models with stochastic model error), or if the decoupled analysis is highly nonlinear, more function evaluations are expected for FORM to converge. The first-order sensitivity indices and the total-effect indices are given in Table 7.

Table 7: Sensitivity indices of electronic packaging problem

	x_1	x_2	x_3	x_4	x_5	x_6	P_h
First-Order	0.362	0.228	0.327	0.000	0.037	0.046	0.048
Total- Effect	0.371	0.236	0.333	0.000	0.038	0.049	0.053

In Table 7, variables x_1 to x_5 have aleatory uncertainty, x_6 has both aleatory and epistemic uncertainty, and P_h is epistemic uncertainty due to model error. The global sensitivity analysis is able to include both types of uncertainty by using the auxiliary variable approach. It is observed that in this example, three aleatory variables -- length (x_1) and width (x_2) of the heatsink and the length of the fin (x_3) -- have a dominant impact on the output variance, whereas the other uncertainty sources only have a small influence. Similarly to Table 4, the difference between the first-order and total effect is very small, which means the collaborative effect between the variables is negligible.

7. Conclusion

This paper presented a new methodology to systematically include both aleatory and epistemic uncertainty in the input variables, and model errors, within feedback-coupled MDA. A likelihood-based approach is employed to represent both variability and data uncertainty in the input random variables (due to sparse and/or imprecise data) through non-parametric distributions. In the presence of stochastic model error, an auxiliary variable method based on the probability integral transform is proposed to include the effect of model error in coupled MDA. This method brings the epistemic uncertainty to the same level of analysis as input variability such that the propagation of both aleatory and epistemic uncertainty can be implemented in a single loop manner. The proposed methodology provides a general formulation to include both model form error and numerical errors (e.g., discretization error, surrogate model error, etc.) within feedback coupled MDA.

The auxiliary variable approach also provides a breakthrough in global sensitivity analysis, which previously was only used in the context of aleatory uncertainty and for feed-forward problems. A mathematical problem and an electronic packaging application are solved using the proposed methodology. The results indicate good performance and high efficiency of the proposed methodology.

This methodology proposes a comprehensive framework for the representation and propagation of multiple sources of uncertainty through coupled MDA. In reality, many uncertainty sources are correlated with each other; therefore, future research needs to include the correlations between different sources of uncertainty in MDA. Moreover, when the coupling variable is a field-type quantity (e.g., in aeroelastic analysis of an aircraft wing, pressures and displacements at hundreds of nodes are exchanged between CFD and FEA), the feasibility in extending the proposed likelihood-based approach to such high-dimensional multidisciplinary analysis problems needs to be investigated.

8. Acknowledgments

This study was supported by funds from NASA Langley Research Center under a Hypersonics NRA project (Cooperative Agreement No. NNX08AF56A1, Technical Monitor: Lawrence Green). The support is gratefully acknowledged. The computational resources of Vanderbilt University's ACCRE have been used in this paper.

9. References

[1] Cramer, E. J., Dennis, J. E., Frank, P. D., and Shubin, G. R., "Problem Formulation for Multidisciplinary Optimization Problem Formulation for Multidisciplinary Optimization," *Siam Journal Optimization*, vol. 4, pp. 754–776, 1993.

[2] Ilan, K., Steve, A., Robert, B., Peter, G., and Ian, S., "Multidisciplinary optimization methods for aircraft preliminary design," *NASA Technical Report*, 1994.

[3] Belytschko, T., "Fluid-structure interaction," *Computers & Structures*, vol. 12, pp. 459–469, 1980.

[4] Thornton, E. a., "Thermal structures - Four decades of progress," *Journal of Aircraft*, vol. 29, pp. 485–498, 1992.

[5] Wieting, A. R., Dechaumphai, P., Bey, K. S., Thornton, E. a., and Morgan, K., "Application of integrated fluid-thermal-structural analysis methods," *Thin-Walled Structures*, vol. 11, pp. 1–23, 1991.

[6] Sobieszczanski-Sobieski, J., and Haftka, R., "Multidisciplinary aerospace design optimization: survey of recent developments," *Structural optimization*, pp. 1–23, 1997.

[7] Felippa, C. a., Park, K. C., and Farhat, C., “Partitioned analysis of coupled mechanical systems,” *Computer Methods in Applied Mechanics and Engineering*, vol. 190, pp. 3247–3270, 2001.

[8] Michler, C., and Hulshoff, S., “A monolithic approach to fluid–structure interaction,” *Computers & fluids*, 2004.

[9] Haldar, A., and Mahadevan, S., *Probability, reliability, and statistical methods in engineering design*, John Wiley, 2000.

[10] Sankararaman, S., and Mahadevan, S., “Likelihood-based representation of epistemic uncertainty due to sparse point data and/or interval data,” *Reliability Engineering & System Safety*, vol. 96, pp. 814–824, 2011.

[11] Agarwal, H., Renaud, J. E., Preston, E. L., and Padmanabhan, D., “Uncertainty quantification using evidence theory in multidisciplinary design optimization,” *Reliability Engineering & System Safety*, vol. 85, pp. 281–294, 2004.

[12] Helton, J. C., Johnson, J. D., Oberkampf, W. L., and Storlie, C. B., “A sampling-based computational strategy for the representation of epistemic uncertainty in model predictions with evidence theory,” *Computer Methods in Applied Mechanics and Engineering*, vol. 196, pp. 3980–3998, 2007.

[13] Du, L., Choi, K. K., Youn, B. D., and Gorsich, D., “Possibility-Based Design Optimization Method for Design Problems With Both Statistical and Fuzzy Input Data,” *Journal of Mechanical Design*, vol. 128, p. 928, 2006.

[14] Zhang, X., and Huang, H.-Z., “Sequential optimization and reliability assessment for multidisciplinary design optimization under aleatory and epistemic uncertainties,” *Structural and Multidisciplinary Optimization*, vol. 40, pp. 165–175, 2009.

[15] Zadeh, L. A., “Toward a perception-based theory of probabilistic reasoning with imprecise probabilities,” vol. 105, pp. 233–264, 2002.

[16] Ferson, S., Kreinovich, V., Ginzburg, L., Myers, D. S., and Sentz, K., “Constructing Probability Boxes and Dempster-Shafer Structures,” 2003.

[17] Matsumura, T., and Haftka, R. T., “Reliability Based Design Optimization Modeling Future Redesign With Different Epistemic Uncertainty Treatments,” *Journal of Mechanical Design*, vol. 135, p. 091006, 2013.

[18] Zaman, K., Rangavajhala, S., McDonald, M. P., and Mahadevan, S., “A probabilistic approach for representation of interval uncertainty,” *Reliability Engineering & System Safety*, vol. 96, pp. 117–130, 2011.

[19] Rebba, R., Mahadevan, S., and Huang, S., “Validation and error estimation of computational models,” *Reliability Engineering & System Safety*, vol. 91, pp. 1390–1397, 2006.

[20] Mahadevan, S., and Liang, B., “Error and Uncertainty Quantification and Sensitivity Analysis in Mechanics Computational Models,” *International Journal for Uncertainty Quantification*, vol. 1, pp. 147–161, 2011.

[21] Kennedy, M. C., and O’Hagan, A., “Bayesian calibration of computer models,” *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, vol. 63, pp. 425–464, 2001.

[22] Gu, X., Renaud, J. E. E., Batill, S. M. M., Brach, R. M. M., and Budhiraja, a. S., “Worst Case Propagated Uncertainty of Multidisciplinary Systems in Robust Optimization,” *Structural Optimization*, vol. 20, pp. 190–213, 2000.

[23] Kokkolaras, M., Mourelatos, Z. P., and Papalambros, P. Y., “Design Optimization of Hierarchically Decomposed Multilevel Systems Under Uncertainty,” *Journal of Mechanical Design*, vol. 128, p. 503, 2006.

[24] Liu, H., Chen, W., Kokkolaras, M., Papalambros, P. Y., and Kim, H. M., “Probabilistic Analytical Target Cascading: A Moment Matching Formulation for Multilevel Optimization Under Uncertainty,” *Journal of Mechanical Design*, vol. 128, p. 991, 2006.

[25] Du, X., and Chen, W., “Collaborative Reliability Analysis under the Framework of Multidisciplinary Systems Design,” *Optimization and Engineering*, vol. 6, pp. 63–84, 2005.

[26] Mahadevan, S., and Smith, N., “Efficient first-order reliability analysis of multidisciplinary systems,” *International Journal of Reliability and Safety*, vol. 1, pp. 137–154, 2006.

[27] Sankararaman, S., and Mahadevan, S., “Likelihood-Based Approach to Multidisciplinary Analysis Under Uncertainty,” *Journal of Mechanical Design*, vol. 134, p. 031008, 2012.

[28] Saltelli, A., Ratto, M., Andres, T., Campolongo, F., Cariboni, J., Gatelli, D., Saisana, M., and Tarantola, S., *Global sensitivity analysis: the primer*, Wiley-Interscience, 2008.

[29] McKeeman, W. M., “Algorithm 145: Adaptive numerical integration by Simpson’s rule,” *Communications of the ACM*, vol. 5, p. 604, 1962.

[30] Mahadevan, S., and Rebba, R., “Inclusion of Model Errors in Reliability-Based Optimization,” *Journal of Mechanical Design*, vol. 128, p. 936, 2006.

[31] Chen, W., Baghdasaryan, L., Buranathiti, T., and Cao, J., “Model validation via uncertainty propagation and data transformations,” *AIAA journal*, vol. 42, pp. 1403–1415, 2004.

[32] Sankararaman, S., Ling, Y., Shantz, C., and Mahadevan, S., “Inference of equivalent initial flaw size under multiple sources of uncertainty,” *International Journal of Fatigue*, vol. 33, pp. 75–89, 2011.

[33] Huang, S., Mahadevan, S., and Rebba, R., “Collocation-based stochastic finite element analysis for random field problems,” *Probabilistic Engineering Mechanics*, vol. 22, pp. 194–205, 2007.

[34] Clarke, S. M., Griebsch, J. H., and Simpson, T. W., “Analysis of Support Vector Regression for Approximation of Complex Engineering Analyses,” *Journal of Mechanical Design*, vol. 127, p. 1077, 2005.

[35] Wang, G. G., and Shan, S., “Review of Metamodeling Techniques in Support of Engineering Design Optimization,” *Journal of Mechanical Design*, vol. 129, p. 370, 2007.

[36] Hombal, V., and Mahadevan, S., “Bias Minimization in Gaussian Process Surrogate Modeling for Uncertainty Quantification,” *International Journal for Uncertainty Quantification*, vol. 1, pp. 321–349, 2011.

[37] Zhu, P., Zhang, S., and Chen, W., “Multi-point objective-oriented sequential sampling strategy for constrained robust design,” *Engineering Optimization*, pp. 1–21, 2014.

[38] Pearson, E., “The probability integral transformation for testing goodness of fit and combining independent tests of significance,” *Biometrika*, vol. 30, pp. 134–148, 1938.

[39] Sankararaman, S., and Mahadevan, S., “Separating the contributions of variability and parameter uncertainty in probability distributions,” *Reliability Engineering & System Safety*, vol. 112, pp. 187–199, 2012.

[40] Padula, S., Alexandrov, N., and Green, L., “MDO test suite at NASA Langley Research Center,” *AIAA Paper*, pp. 1–13, 1996.

[41] Rangavajhala, S., Sura, V., Hombal, V., and Mahadevan, S., “Discretization Error Estimation in Multidisciplinary Simulations,” *AIAA journal*, vol. 49, pp. 2673–2683.

[42] Liu, Y., Chen, W., Arendt, P., and Huang, H.-Z., “Toward a Better Understanding of Model Validation Metrics,” *Journal of Mechanical Design*, vol. 133, p. 071005, 2011.

[43] Ling, Y., and Mahadevan, S., “Quantitative model validation techniques: New insights,” *Reliability Engineering & System Safety*, vol. 111, pp. 217–231, 2013.

[44] Rangavajhala, S., Liang, C., Mahadevan, S., “Design Optimization Under Aleatory and Epistemic Uncertainties,” *Proceedings of 14TH AIAA/ISSMO Multidisciplinary Analysis and Optimization Conference*, 2012.

Appendix I

Gaussian Process (GP) Modeling

The GP modeling technique has been used in a wide range of applications such as data regression and model calibration. A GP regression or interpolation models the underlying covariance within the data instead of the actual function form. With a set of training points $\boldsymbol{X_T} = \{x_1, x_2, \dots x_n\}$ and the corresponding model outputs: $\boldsymbol{y}_T = \{y_1, y_2, \dots y_n\}$, the GP model estimates the mean and variance at the prediction points $\boldsymbol{X_P}$ as

$$m = K_{PT}(K_{TT} + \sigma_n^2 I)^{-1} y_T \tag{19}$$

$$S = K_{PP} - K_{PT}(K_{TT} + \sigma_n^2 I)^{-1} K_{TP} \tag{20}$$

where $\boldsymbol{K}_{TT} = \left[k\left(x_i, x_j\right)\right]_{i,j}$ is the $t \times t$ matrix of the covariances between $\boldsymbol{X}_T$; $\boldsymbol{K}_{PP}$is the $p \times p$ matrix of the covariances between $\boldsymbol{X}_P$; $\boldsymbol{K}_{PT}$ and $\boldsymbol{K}_{TP}$ are the $p \times t$ matrix of covariances between $\boldsymbol{X}_P$ and $\boldsymbol{X}_T$ and its transpose. Function evaluations with the GP surrogate model are inexpensive; therefore it can be used to replace an expensive high-fidelity computational model in activities such as model calibration [21] and optimization [44].

List of Table Captions

List of Figure Captions