AIAA 2009-2249

AIAA paper AIAA-2009-2249, SAND Report: 2008-5360A

Epistemic Uncertainty in the Calculation of Margins

Laura P. Swiler[*], Thomas L. Paez[†], Randall L. Mayes[‡], and Michael S. Eldred[§]
Sandia National Laboratories[**], Albuquerque NM 87185

Epistemic uncertainty, characterizing lack-of-knowledge, is often prevalent in engineering applications. However, the methods we have for analyzing and propagating epistemic uncertainty are not as nearly widely used or well-understood as methods to propagate aleatory uncertainty (e.g. inherent variability characterized by probability distributions). In this paper, we examine three methods used in propagating epistemic uncertainties: interval analysis, Dempster-Shafer evidence theory, and second-order probability. We demonstrate examples of their use on a problem in structural dynamics, specifically in the assessment of margins. In terms of new approaches, we examine the use of surrogate methods in epistemic analysis, both surrogate-based optimization in interval analysis and use of polynomial chaos expansions to provide upper and lower bounding approximations. Although there are pitfalls associated with surrogates, they can be powerful and efficient in the quantification of epistemic uncertainty.

I. Introduction

Most computer models for engineering applications are developed to help assess a design or regulatory requirement. The capability to quantify the impact of uncertainty in the decision context is critical. This paper will focus on situations with epistemic uncertainty, which represents a lack of knowledge about the appropriate value to use for a quantity. Epistemic uncertainty is sometimes referred to as state of knowledge uncertainty, subjective uncertainty, or reducible uncertainty, meaning that the uncertainty can be reduced through increased understanding (research), or increased and more relevant data. [Helton et al.] In contrast, uncertainty characterized by inherent randomness which cannot be reduced by further data is called aleatory uncertainty or variability. Aleatory uncertainties are usually modeled with probability distributions, but epistemic uncertainty often is not modeled probabilistically. Regulatory agencies, design teams, and weapon certification assessments are increasingly being asked to specifically characterize and quantify epistemic uncertainty and separate its effect from that of aleatory uncertainty [Diegert et al.]

There are many ways of representing epistemic uncertainty, including fuzzy sets, possibility theory, and imprecise probability. At Sandia we have chosen to focus on three approaches: interval analysis, Dempster-Shafer evidence theory, and (for mixed aleatory/epistemic uncertainties) second-order probability. The rest of this introductory section outlines these three approaches in more detail. Section 2 discusses surrogate methods in the context of epistemic uncertainty quantification (UQ). Section 3 presents an example, Section 4 provides results, Section 5 discusses margin calculations, and Section 6 summarizes the paper.

[*] Principal Member of Technical Staff, Optimization and Uncertainty Estimation Dept., P.O. Box 5800, Sandia National Laboratories, MS 1318.
[†] Distinguished Member of Technical Staff, Validation and Uncertainty Quantification Dept., P.O. Box 5800, Sandia National Laboratories, MS 0557.
[‡] Principal Member of Technical Staff, Experimental Mechanics, Non-Destructive Evaluation and Modal Validation Dept., P.O. Box 5800, Sandia National Laboratories, MS 0557.
[§] Principal Member of Technical Staff, Optimization and Uncertainty Estimation Dept., P.O. Box 5800, Sandia National Laboratories, MS 1318, Associate Fellow AIAA.
[**] Sandia is a multiprogram laboratory operated by Sandia Corporation, a Lockheed Martin Company, for the United States Department of Energy's National Nuclear Security Administration under Contract DE-AC04-94AL85000.

A. Interval analysis

The simplest way to propagate epistemic uncertainty is by interval analysis. In interval analysis, it is assumed that nothing is known about the uncertain input variables except that they lie within certain intervals. The problem of uncertainty propagation then becomes an interval analysis problem: given inputs that are defined within intervals, what is the corresponding interval on the outputs?

Although interval analysis is conceptually simple, in practice it can be difficult to determine the optimal solution approach. A direct approach is to use optimization to find the maximum and minimum values of the output measure of interest, which correspond to the upper and lower interval bounds on the output, respectively. There are a number of optimization algorithms which solve bound constrained problems, such as bound-constrained Newton methods. In practice, it may require a prohibitively large number of function evaluations to determine these optima, especially if the simulation if very nonlinear with respect to the inputs, has a high number of inputs with interaction effects, exhibits discontinuities, etc. Local optimization solvers will not guarantee finding global optima, and thus to solve this problem properly, one may have to resort to multi-start implementations of local optimization methods or global methods such as genetic algorithms, DIRECT, etc. These approaches can be very expensive.

Another approach to interval analysis is to sample from the uncertain interval inputs, and then take the maximum and minimum output values based on the sampling process as the estimate for the upper and lower output bounds. Usually a uniform distribution is assumed over the input intervals, although this is not necessary. Although uniform distributions may be used to create samples, one cannot assign a probabilistic distribution to them or make a corresponding probabilistic interpretation of the output. That is, one cannot make a CDF of the output: all one can assume is that sample input values were generated, corresponding sample output values were created, and the minimum and maximum of the output are the estimated output interval bounds. This sampling approach is easy to implement, but its accuracy is highly dependent on the number of samples. Often, sampling will generate output bounds which underestimate the true output interval.

Other approaches to interval analysis start with sampling, but then use the samples to create a surrogate model (e.g. a regression model, a neural net, an adaptive spline model, etc.) The surrogate model can then be sampled very extensively (e.g. a million times) to obtain an upper and lower bound estimate. Another approach is to use surrogate-based optimization methods to obtain the upper and lower bounds. We examine the accuracy and feasibility of surrogate approaches in Section 4 below.

B. Dempster-Shafer Theory of Evidence

Dempster-Shafer evidence theory is an attractive approach to propagation of evidence theory when using computational simulations, in part because it is a generalization of classical probability theory which allows the simulation code to remain black-box (it is non-intrusive to the code) and because the Dempster-Shafer calculations use much of the probabilistic framework that exists in most places. [Helton et al.]

Dempster-Shafer Theory of Evidence may be used to perform epistemic analysis. In Dempster-Shafer evidence theory, the epistemic uncertain input variables are modeled as sets of intervals. Note that each variable may be defined by one or more intervals. The user assigns a basic probability assignment (BPA) to each interval, indicating how likely it is that the uncertain input falls within the interval. The BPAs for a particular uncertain input variable must sum to one. The intervals may be overlapping, contiguous, or have gaps. Dempster-Shafer has two measures of uncertainty, belief and plausibility. The intervals are propagated to calculate belief (a lower bound on a probability value that is consistent with the evidence) and plausibility (an upper bound on a probability value that is consistent with the evidence). Together, belief and plausibility define an interval-valued probability distribution on the results, not a single probability distribution. An example of cumulative belief and plausibility distribution functions is shown in Figure 1. Note that they encompass a cumulative distribution function (CDF) that would be obtained by propagation of probability distributions on inputs.

The main method for calculating Dempster-Shafer intervals is computationally very expensive. Many hundreds of thousands of samples are taken over the space. Each combination of input variable intervals defines an input "cell." By interval combination, we mean the first interval of the first variable paired with the first interval for the second variable, etc. Within each interval calculation, it is necessary to find the minimum and maximum function value for that interval "cell." These minimum and maximum values are aggregated to create the belief and plausibility curves. The Dempster-Shafer method may use a surrogate model and/or optimization methods. The accuracy of the Dempster-Shafer results is highly dependent on the number of samples and the number of interval combinations. If one has many interval cells and few samples, the estimates for the minimum and maximum function evaluations are likely to be poor. Surrogate methods may also be used in Dempster-Shafer, either global surrogates or separate surrogates within each cell.

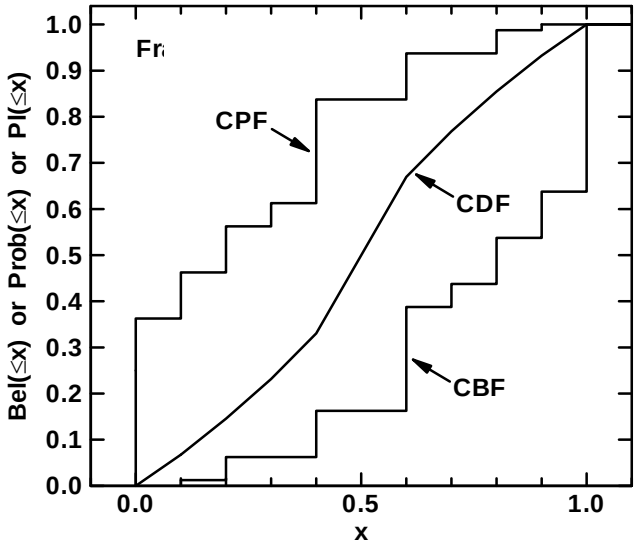

Figure 1. Example of Dempster-Shafer Calculations: Cumulative Distributions for Belief and Plausbility

C. Second-order probability

This section discusses the case where we are trying to propagate both aleatory and epistemic uncertainty. A common situation is where one may know the form of the probability distribution for an uncertain variable (for example, that it is distributed normally or lognormally), but one is not sure of the parameters governing the distribution. In this case, the analysis is done with an outer loop and an inner loop. In the outer loop, the epistemic variables are specified. In this example, the epistemic variables are specified as intervals on parameter values such as means or standard deviations of uncertain variables. A particular value is selected from within the specified intervals. Then, this value is sent to the inner loop. In the inner loop, the values of the distribution parameters are set by particular realizations of the epistemic variables, and the inner loop performs sampling on the aleatory variables in the usual way (e.g., a LHS sample is taken). Figure 2 shows the sampling structure of a second-order probability analysis.

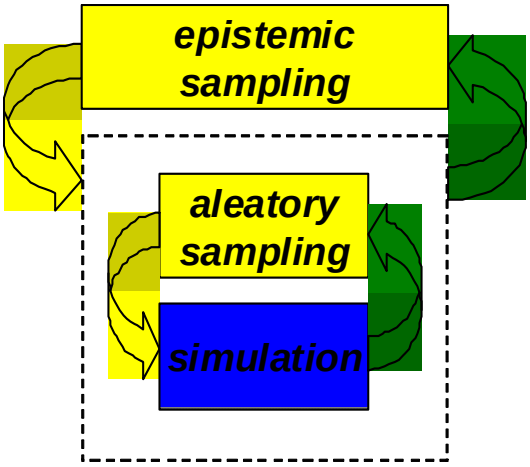

Figure 2. Second-order Probability

Second-order probability may be expensive since it is often implemented with two sampling loops. However, it has the advantage that it is easy to separate and identify the aleatory vs. epistemic uncertainty. Each particular set of epistemic variable values generates an entire CDF for the response quantities based on the aleatory uncertainty. So, for example, if one had 50 values or samples taken of the epistemic variables, one would have 50 CDFs, as seen in the example in Figure 3. When you plot the 50 CDFs, you get the upper and lower bound on the family of

distributions and on the percentiles, as shown by the black line depicting the range on the median in Figure 3. Plots of ensembles or "families" of CDFs generated in second-order probability are sometimes called "horsetail" plots since the CDFs overlaid on each other can look like a horse's tail. Note also that in some situations, second-order probability results can look similar to a Dempster-Shafer analysis but the underlying assumptions are different.

We propose a new approach for performing second-order probability analysis. In this approach, the "inner loop" CDFs will be calculated using a stochastic expansion method, and the outer loop bounds will be performed via interval optimization. The advantages of this can be significant, due to several factors. The first is that the stochastic expansion methods, as explained in more detail in the following section on surrogate models, can be much more efficient than sampling for calculation of a CDF. The second advantage is that stochastic expansion methods allow analytic representation of the moments, so in some cases, the derivatives of the moments with respect to the epistemic variables in the outer loop can be written analytically, and these analytic derivatives can be used with optimization methods to find interval bounds on mean and variance, for example. Finally, the optimization methods in the outer loop may also be more efficient than generating sufficient outer loop samples to get a good estimate of outer loop bounds.

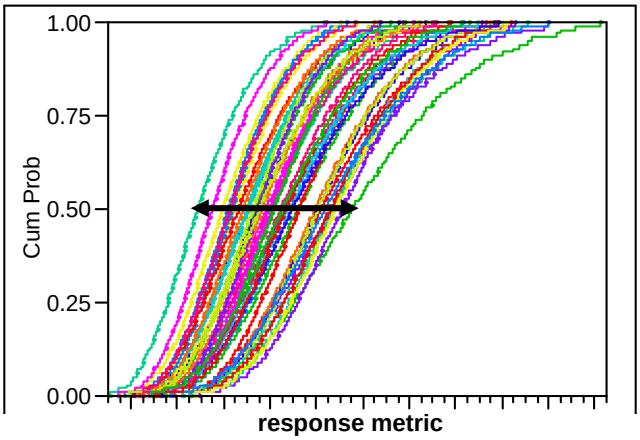

Figure 3. Example Results from Second-order Probability Analysis

II. Surrogate methods in Epistemic UQ

Surrogate methods involve constructing response surface approximations of computationally expensive functions. These surrogates (sometimes called meta-models) are often constructed by taking a set of samples from the function or simulation model of interest, then building a regression or non-parametric interpolation model based on the sample points. However, other surrogate methods exist including multifidelity models (e.g. a low fidelity physics model can be used as a surrogate for a high fidelity one) and reduced order models such as proper orthogonal decomposition or spectral decomposition. In this paper, we limit the discussion to data-fit surrogates, where the surrogate is built or fit to a particular set of sample points.

There is a large literature on the use of surrogate models in optimization. Surrogate-based optimization has become a common approach for solving optimization problems that require the execution of a computationally expensive high-fidelity simulation in order to obtain objective function and constraint values. The essence of such approaches entails constructing low-fidelity models by fitting response surfaces to high-fidelity function values or by reducing the numerical or physical fidelity of the simulation. Optimization methods are then applied to these less expensive low-fidelity functions with periodic corrections from the high-fidelity simulation to ensure convergence to a local minimum of the high-fidelity function. Alexandrov et al. provide the theoretical framework under which convergence for this class of algorithms can be proved. Eldred and Dunlavy explore different constraint handling approaches in the context of that theoretical framework and investigate their effectiveness in practice.

While surrogates are especially beneficial in optimization due to the number of function evaluations that must be performed, surrogates have not been as widely used in uncertainty quantification although UQ also requires large numbers of function evaluations. The main reason is that there is no comparable theory similar to the trust region for optimization. In uncertainty quantification, we are not trying to assess the accuracy of a surrogate only within a trust region, but throughout the entire space if we are to use the surrogate to generate a cumulative distribution

function of the output (for example) or a probability of failure estimate. The surrogate may be locally accurate in some places and inaccurate in other places. There is no general approach for using information about the level of accuracy of the surrogate at specific points to extrapolate to the accuracy of estimates of statistical moments based on the surrogate or to bound various statistics that can be generated from extensively sampling the surrogate over the entire domain. Giunta et al. [2006] investigated the use of response surface approximations for UQ. They concluded that for small sample sizes, generating statistical estimates of moments from the sample points alone was often just as good as or better than using the samples to construct a surrogate, then using the surrogate to generate statistical estimates of moments. This is because surrogates built on a small number of samples tended to be inaccurate and/or biased. They also concluded that surrogates could be useful with larger numbers of samples, and in larger dimensional spaces. Giunta et al. suggest exercising caution when using surrogates for UQ. They suggest using multiple types of surrogate approximations along with generating statistics based purely on the samples (no surrogate approximation) to cross-check the surrogate-based UQ predictions.

The paper by Giunta et al. was focused on using surrogates for propagating aleatory uncertainty. This paper focuses on epistemic uncertainty. Surrogates have not been widely used to propagate and assess epistemic uncertainty. One logical application of surrogates in epistemic uncertainty is in interval analysis: optimization can be performed on a surrogate to determine the maximum and minimum of a function over bounded intervals on the inputs. Many epistemic methods such as Dempster-Shafer evidence theory and possibility theory rely on finding the minimum and maximum of the function on bounded subregions of the space, and so surrogates would be a good fit for this need.

There are many types of surrogates such as polynomial regression models, spline models, and neural networks. We do not present an exhaustive list here; there are a number of papers comparing the performance of surrogates including Wang et al., Swiler et al., and Meckesheimer et al. We will focus on three surrogates for the purposes of propagating epistemic uncertainty for the purposes of demonstrating how surrogates can be used, especially in interval analysis. The first surrogate considered is a quadratic polynomial. The second is a Gaussian process (GP), or kriging model. Gaussian processes are based on spatial statistics, and can be used to fit a wide variety of functional forms. They also provide a direct estimate of the uncertainty associated with their predictions. The basic idea of the GP model is that the response values Y are modeled as a group of multivariate normal random variables. A covariance function is then constructed as a function of the inputs x. The covariance function is based on the idea that when the inputs are close together, the correlation between the outputs will be high. Gaussian processes have recently become widely used as surrogates for computational experiments. A seminal paper on GPs for modeling computer experiments is by Sacks et al.; recent investigations include Martin and Simpson, McFarland et al., and Viana et al.

The third surrogate type that will be investigated for epistemic uncertainty calculations are polynomial chaos expansions (PCE). In PCE, the output is considered a random process which is represented as the sum of orthogonal polynomial basis functions. These basis functions are functions of the input random variables, and the Wiener-Askey scheme provides a framework for choosing orthogonal polynomials specific to the input random variable distribution types (e. g. Hermite polynomials are used for modeling the effects of normals, Legendre polynomials are used for modeling the effects of uniforms, etc.) [Xiu, D. and Karniadakis]. A challenge in using PCE is determining the coefficients of the basis function in the expansion in an efficient way. This may be done in a variety of ways: Eldred et al. present a detailed comparison of sampling, quadrature, cubature, and point collocation methods to determine the expansion coefficients in non-intrusive PCE applications. The interest in using PCE as surrogates for epistemic calculations is that the expansions can be very accurate, especially in the tails, for a given number of samples and thus the calculation of minimum and maximum values based on a PCE may be a feasible and efficient option in some cases.

III. Structural Dynamics Example

We present an example from structural dynamics, where the application of interest is the performance of the bonding material in an aeroshell. In the example we present, the application has been simplified. We have a fairly coarse, 3-D model of 3 discs. The outer 2 discs represent rigid masses (in this case, they are steel) and the inner disc represents a layer of a filled rubber. Figure 4 depicts the geometry of the configuration used in this example. We are interested understanding frequencies of the axial and shear modes for this experimental configuration, shown in Figure 5. There is significant epistemic uncertainty in this example associated with the material properties of the filled rubber. Specifically, we have a wide variety of tests and expert opinion on potential values for the modulus of elasticity in tension and compression, E, and Poisson's ratio, ν. The filled rubber is a rubber material with particles

in it. In this case the particles are glass balloons, which are used to get the density of the material down. A filled rubber softens with increased strain (on other rubbers, we have seen as much as an order of magnitude difference in the modulus, depending on the strain level). In vibration, the strain levels are usually very low, e.g. on the order of 0.1% strain or less.

The simulation code used is Salinas [Reese et al.], which is a finite-element analysis code for modal, vibration, static and shock analysis developed at Sandia National Laboratories for massively parallel implementations (for more information, see: http://jal.sandia.gov/Salinas/). This simulation takes approximately 2 hours to run on a Linux workstation with two Dual-Core Intel® Xeon® 5000 series 64-bit processors and 2Gigabytes of RAM.

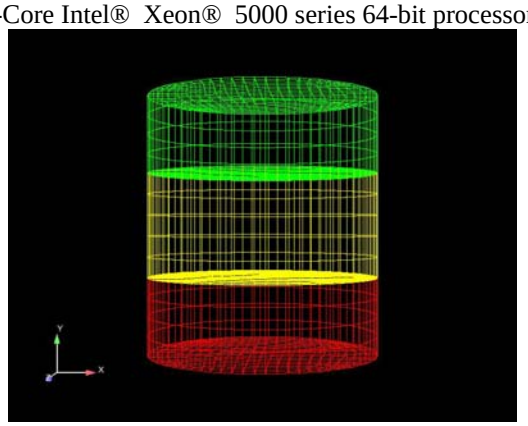

Figure 4. 3 disc model with filled rubber as the middle disc (in yellow)

We have a variety of test data: some dynamic tests, some static, and one ultrasonic. Some of the tests are on the discs and some on the system-level aeroshells. The test data has been taken by several organizations under different conditions and is not very consistent. One of the static tests was taken at strain levels much higher than the small strain of the rubber in vibration, thus invalidating the data for our needs. We don't have much confidence in the ultrasonic test because the filled rubber layer was too thin in comparison to the other layers they had to send the ultrasonic signal through. Also, some of the test data reported to us involves people using test results and calibrating their models to infer values of E and/or ν. For the purposes of this paper, we are not trying to calibrate our finite-element model; we are simply trying to use it to properly propagate epistemic uncertainty. Finally, there is some correlation between of E and ν. To start, based on our assessment of the test data available, we will assume that the value of E falls within the interval of [2000, 25000] psi and the value for ν falls within the interval of [0.45, 0.495].

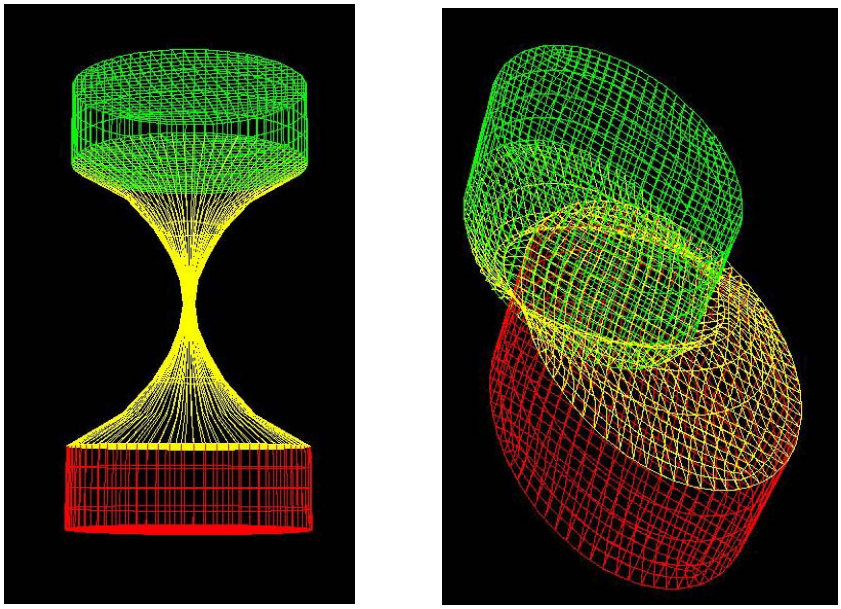

Figure 5. Axial Mode (left) and Shear Mode (right) for 3-disc model

IV. Structural Dynamics Model: Epistemic Results

This section presents the results of various approaches (pure interval analysis, Dempster-Shafer evidence theory, and second-order probability analysis) used to propagate epistemic uncertainty. These results were generated using DAKOTA, a software framework that allows one to perform uncertainty quantification, optimization, and parameter studies (Eldred et al.,: http://www.cs.sandia.gov/DAKOTA/).

A. Interval analysis

For interval analysis, we present 3 approaches. The first analysis is the simplest, based on sampling. The second set of analyses uses samples to construct surrogates, and the surrogates are sampled extensively to determine output bounds. The third set of analyses uses surrogates in conjunction with formal optimization methods to determine output bounds.

1. Sampling-based interval analysis

This section shows the results of applying the Latin Hypercube sampling methodology [17] to the epistemic interval propagation. The input uncertainties in E and ν were defined by the intervals [2000, 25000] and [0.45, 0.49], respectively. Initially, to ensure the DAKOTA and Salinas codes were properly coupled and everything was working correctly, we performed a small, ten-sample study. The results of this study are shown in Table 1 below and in the Figures 6 and 7. Note that based on this small run, the output interval for the shear mode frequency is [845.6, 2878.0] Hz, and the output interval for the axial mode frequency is [1088.1,3580.37] Hz.

Sample	E (Elastic Modulus)	Nu (Poisson's ratio)	Shear Mode Frequency	Axial Mode Frequency
1	6377.50	0.473	1452.47	1858.78
2	24938.67	0.455	2877.98	3580.37
3	9764.92	0.463	1799.41	2263.74
4	20550.80	0.462	2610.35	3277.82
5	14733.46	0.466	2209.13	2793.58
6	19525.95	0.488	2539.35	3333.59
7	12791.57	0.482	2055.63	2670.29
8	16942.52	0.481	2365.74	3065.20
9	7312.58	0.452	1559.74	1931.17
10	2162.54	0.476	845.62	1088.09

Table 1. Initial Interval Sample

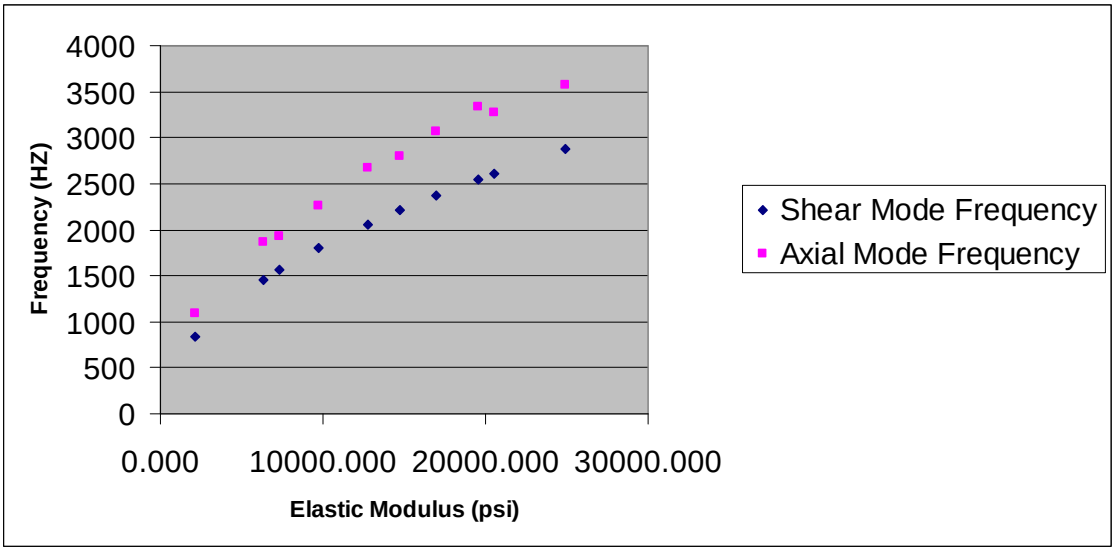

Figure 6. Shear and Axial Mode Frequencies as a function of E for 10-sample case

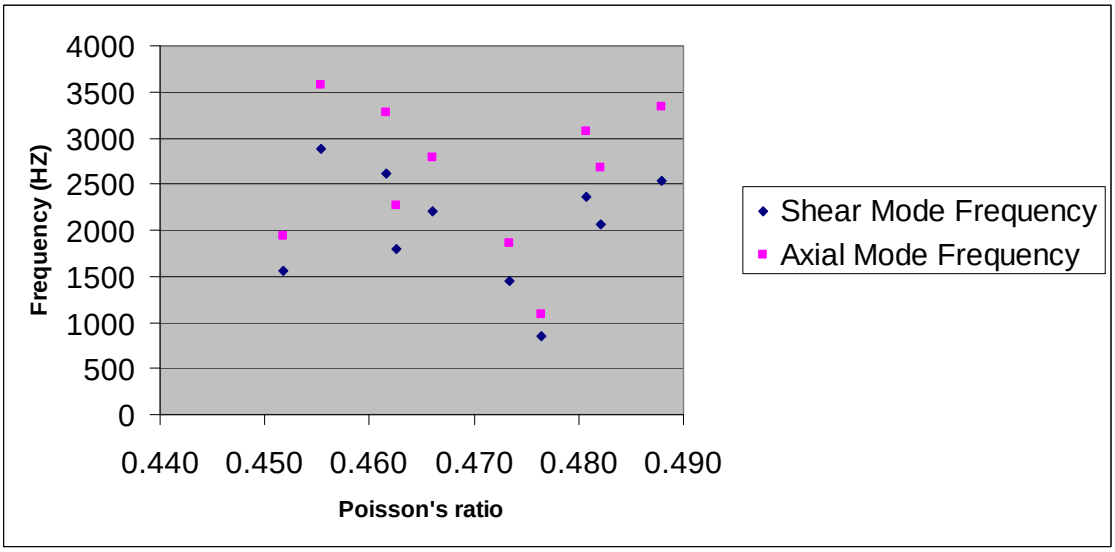

Figure 7. Shear and Axial Mode Frequencies as a function of ν for 10-sample case

Figure 6 shows that both the axial and shear mode frequencies are almost perfectly linearly correlated with the elastic modulus, E. However, Figure 7 shows no significant correlation between Poisson's ratio and the shear and axial modes. This was surprising to us at first glance (we expected some sensitivity between E and ν especially as ν nears the upper end of its interval range), and so we performed some additional analyses [Swiler et al., 2009]. Main effects analyses verified that there was little interaction between E and ν in this problem. We also performed another 30 LHS samples. The 30 sample interval analysis gave similar results to the ten sample study, but with slightly wider intervals: [845.6, 2878.0] for the shear mode frequency and [1088.1, 3696.0] for the axial mode. Going from 10 to 30 samples did not change the output intervals significantly in this example problem since we had one input with a very linear relationship and one variable that was fairly uncorrelated with the output, but in other situations could improve the interval bounds on the output significantly.

2. Surrogate-based Epistemic UQ via Surrogate Sampling

We used a full data set of 76 sample points taken from the initial LHS sampling outlined above as well as samples generated for main effects analysis. Then, we constructed a few different surrogate models based on these points: a quadratic regression model, a MARS model (multivariate adaptive splines), a neural network, and a Gaussian process model. These surrogate models were then sampled with the same set of 1000 points to determine the upper and lower bounds according to the surrogate model. These interval bounds on the output are shown in Table 2 below. Note that the upper and lower bounds are reasonably consistent across the surrogate methods although the underlying surrogates are based on very different models and assumptions. Finally, for comparison purposes, we constructed a polynomial chaos expansion for the shear and axial mode frequencies as a function of E and ν. For this preliminary investigation of PCE, we ran a second-order expansion, for a total of 9 points. Table 2 shows that the upper and lower bounds obtained by sampling the PCE have some error, but these are only based on 9 points instead of the 76 points that the other data fit approximations used. PCE is giving reasonable approximations of the bounds in this problem very efficiently.

	SHEAR MODE FREQUENCY		AXIAL MODE FREQUENCY	
Surrogate Type	Lower Bound	Upper Bound	Lower Bound	Upper Bound
Quadratic Regression	871.13	2849.90	1099.85	3775.50
Mars	816.03	2880.31	1028.04	3812.84
Neural Net	814.49	2893.26	1007.02	3807.57
Gaussian Process	814.10	2876.00	1043.13	3819.63
Polynomial Chaos	925.83	2855.02	1151.12	3775.78

Table 2. Interval bounds according to sampling a surrogate model

3. Surrogate-based Epistemic UQ via Optimization

The previous section discussed the case of creating a surrogate model, then sampling that surrogate model extensively to determine the upper and lower bounds, as indicated by the minimum and maximum of the 1000 samples at which we evaluated the surrogate model. This section discusses the use of a surrogate model, where we optimize the surrogate model instead of sample it to obtain the upper and lower bounds. Note that one could perform optimization directly on the simulation model, but we are assuming that in most cases, the simulation model is computationally expensive and optimization will be more practical to perform on a surrogate. Thus, this section focuses on surrogate-based optimization. Also, note that the calculation of the bounds on the output requires 2 separate optimization procedures: one to find the minimum value of the output and one to find the maximum value of the output. Both optimization procedures are bound-constrained, meaning that the interval bounds on the input variables must be honored.

For the quadratic regression, neural network, and MARS models, the optimization method we used was DIRECT (Dividing Rectangles, see DAKOTA documentation), which is a global optimization method that balances local search in promising regions of the design space with global search in unexplored regions. We used a global optimization method since we are not using a trust region optimization approach: we are constructing one surrogate over the entire [E, ν] input space and optimizing the surrogate. The results are shown in Table 3. Again, we see that there are not huge differences in the interval bounds obtained for the shear and axial mode frequencies, although the neural net seems more inconsistent than the quadratic regression and Mars. Also, the optimum point in input space is often the same, the bounds are different due to the differences in the surrogate estimate of the response at those locations. Since this is a fairly linear problem, we see that the bounds on the shear or axial mode frequencies occur where E is at its minimum or maximum. Due to the difficulty of estimating a significant influence of ν, we see that the optimum locations obtained for ν vary more than for E.

With the Gaussian process surrogate, we used a different optimization approach based on a method called Efficient Global Optimization (EGO) developed in [Jones, Schonlau, and Welch]. EGO was developed to facilitate the unconstrained minimization of expensive implicit response functions. The idea in EGO is to use properties of the Gaussian process (specifically, the predicted variance in the estimate at potential points in the space) to balance "exploitation" of existing good solutions with "exploration" of parts of the domain which are sparsely populated and where a potential optimum could be located. The method builds an initial Gaussian process model as a global surrogate for the response function, then adaptively selects additional samples to be included in the Gaussian process model in subsequent iterations. The new samples are selected based on how much they are expected to improve the current best solution to the optimization problem using a criteria coded into an "expected improvement function." We have taken the EGO concept and implemented in DAKOTA. For the purposes of interval optimization, we modified the existing algorithm so that we first build the GP for function minimization, then we take the existing points generated by that process, change the objective function and expected improvement function to perform function maximization, and warm-start the GP that is used to find the maximum response value. We have found this approach to be very efficient, where the majority of "true" function evaluations of the simulation model are performed in finding the function minimum, and only a few additional samples are added to the GP to find the function maximum. The performance of this EGO-based interval optimization will depend on the nonlinearity of the simulation model and the number of input dimensions. For this structural dynamics problem, it took 16 function evaluations to find the minimum and maximum estimates of shear mode frequency, and 21 function evaluations to find the minimum and maximum estimates of axial mode frequency. In contrast, optimization on the other surrogates (quadratic regression, MARS, and neural networks) using DIRECT involved 76 true function evaluations. For low-dimensional problems, we have seen the EGO-approach perform well using 30-40 function evaluations whereas optimization performed on a surrogated constructed on a fixed sample set may require a hundred samples or more.

A final word of caution: when using surrogate methods, one needs to know something about the appropriateness of the surrogate for a particular function, and be able to evaluate the accuracy of the meta-model. It is possible to use metrics such as cross-validation metrics, root mean squared error, etc. to evaluate the goodness of fit of the surrogate, but these metrics mainly involve the goodness of the surrogate with respect to the training points upon which it was built. The metrics don't necessarily indicate how good the surrogate will be when evaluated at new sample points (for example, when sampling the surrogate extensively to calculate a mean, variance, or percentile).

Thus, while surrogates are a powerful tool, one must be careful of the accuracy of interpreting statistical measures based on surrogate builds [Guinta et al., 2006]

	SHEAR MODE FREQUENCY		AXIAL MODE FREQUENCY	
Surrogate Type	Lower Bound	Upper Bound	Lower Bound	Upper Bound
Quadratic Regression	865.26	2852.54	1088.54	3791.74
Mars	816.03	2882.92	1011.43	3829.90
Neural Net	772.30	2906.90	993.58	3831.86
Gaussian Process (using EGO optim.)	813.32	2884.02	1008.05	3831.89
	Corresponding Bounding inputs [E,ν]	**Corresponding Bounding inputs [E,ν]**	**Corresponding Bounding inputs [E,ν]**	**Corresponding Bounding inputs [E,ν]**
Quadratic Regression	at 2000,0.494	at 25000,0.45	at 2000,0.45	at 25000,0.495
Mars	at 2000,0.468	at 25000,0.45	at 2000,0.45	at 25000,0.495
Neural Net	at 2000, 0.465	at 25000,0.465	at 2000,0.465	at 25000,0.495
Gaussian Process	at 2000, 0.494	at 25000, 0.45	at 2000,0.45	at 25000, 0.495

Table 3. Interval bounds obtained by optimizing a surrogate model

B. Dempster-Shafer Evidence Theory

The previous section discussed interval optimization and some approaches to epistemic uncertainty based on interval calculations. This section presents the same structural dynamics example, but using Dempster-Shafer evidence theory to characterize epistemic uncertainty. In this example, we specified a belief structure on the elastic modulus as follows: BPA of 0.3 on the interval [3000, 6000], BPA of 0.6 on the interval [6000, 10000], and BPA of 0.1 on the interval [10000,25000]. The belief structure on the intervals for ν are as follows: BPA of 0.7 on the interval [0.45,0.475], BPA of 0.3 on the interval [0.475,0.495]. Note that the intervals in this example are defined as contiguous intervals but there is no requirement that they be so: they can be overlapping or disjoint. These intervals are depicted graphically in Figure 8 below. The resulting cumulative distribution functions of belief and plausibility for the shear mode frequency are shown in Figure 9 and for the axial mode frequency are shown in Figure 10. Note that in the context of belief, the cumulative belief function (similar to a cumulative distribution function or CDF) is the cumulative belief that the uncertain quantity y^* is less than a given value y: $\text{Bel}(y^* \leq y)$. Similarly, the cumulative plausibility function is the cumulative plausibility that the uncertain quantity y^* is less than a given value y: $\text{Pl}(y^* \leq y)$. For example, in Figure 9, the cumulative belief that the shear modulus is less than or equal to 1800 Hz is 0.3, while the cumulative plausibility that the shear modular is less than or equal to 1800 Hz is 0.9. Another way of looking at this is the minimum amount of likelihood that could be associated with 1800 Hz is 0.3, while the maximum amount of likelihood that could be associated with 1800 Hz is 0.9. If we were to think in terms of probabilities and CDFs, the belief and plausibility provide an upper and lower bound on the CDF: the cumulative probability that the shear frequency is less than or equal to 1800 Hz is between 0.3 and 0.9. Finally, the "stair-stepping" behavior of these cumulative curves is due to the discrete combinations of intervals on the input variables and the discrete levels of output at which we requested plausibility and belief to be accumulated. It is difficult to represent, but at 1500 Hz, for example, the cumulative belief jumps from 0 to 0.3, and the cumulative plausibility jumps from 0.3 to 0.9. The axial mode plot in Figure 10 is more representative of Dempster-Shafer analyses: it is easy to imagine that the cumulative probability function may lie between the pink (plausibility) and blue (belief) lines in the figure.

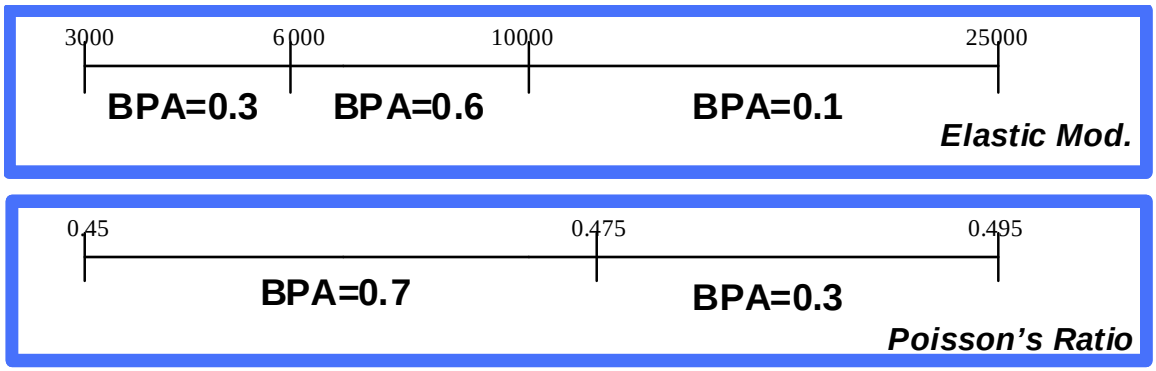

Figure 8. Intervals and associated BPAs for Dempster-Shafer analysis

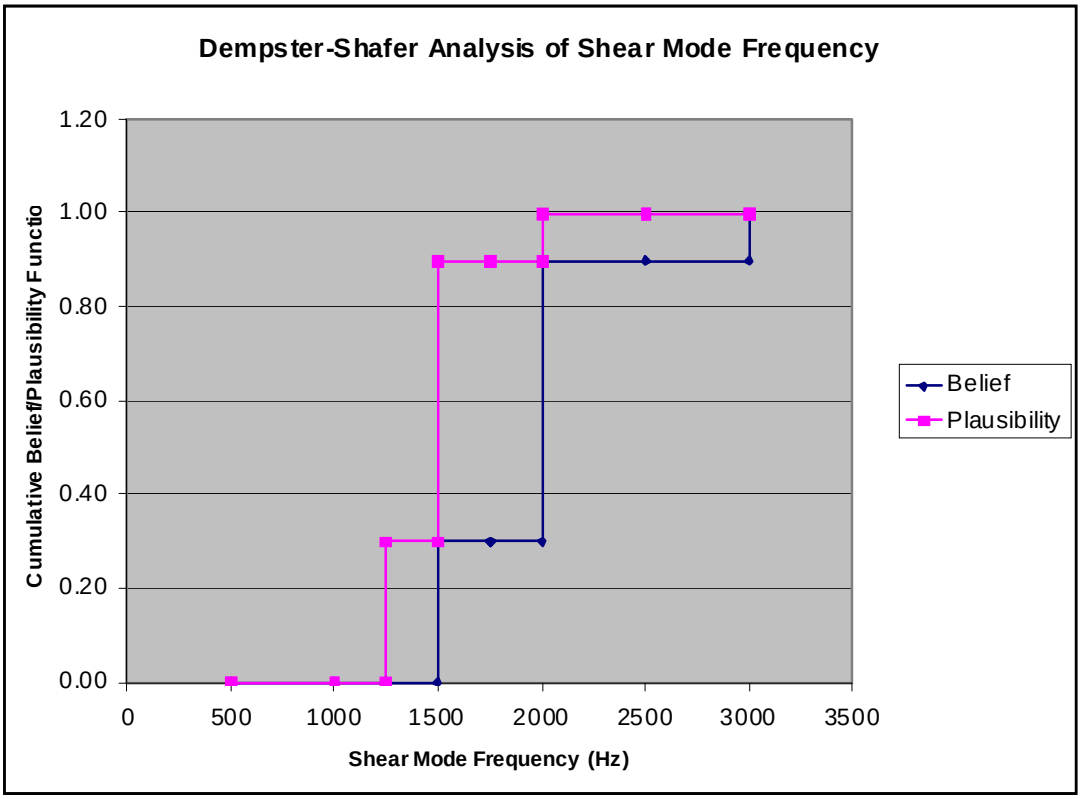

Figure 9. Cumulative Belief and Plausibility Distributions for Shear Mode

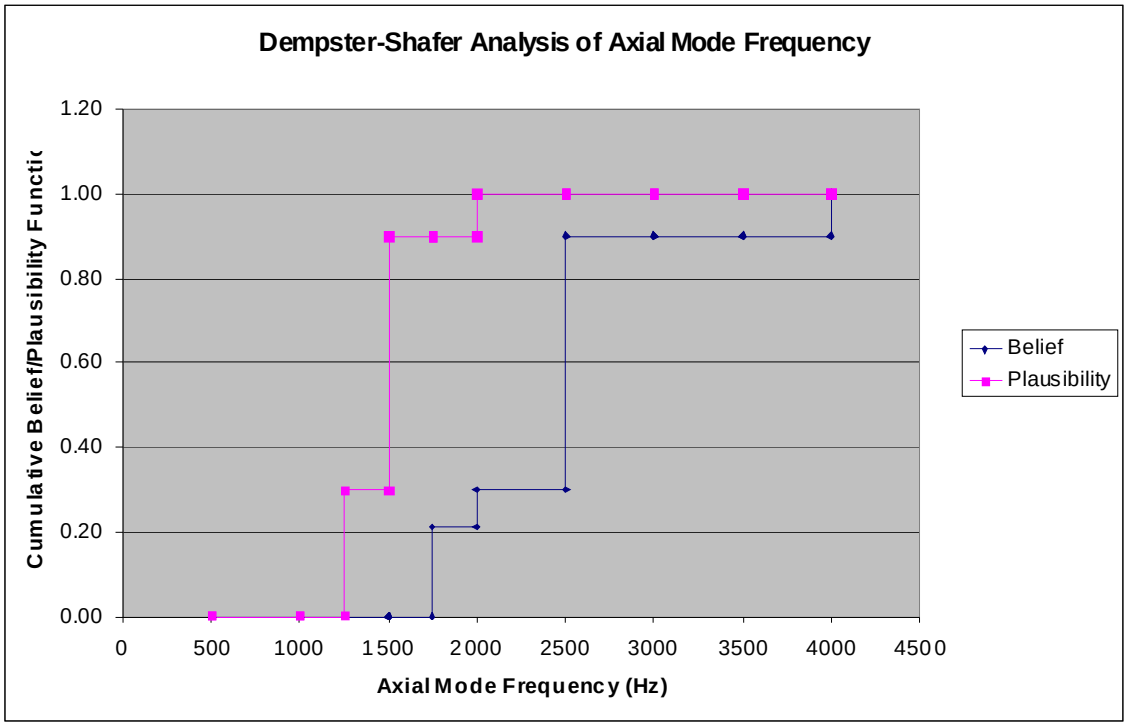

Figure 10. Cumulative Belief and Plausibility Distributions for Axial Mode

C. Second-order Probability

Continuing with the structural dynamics example, we performed a second-order probability analysis where a value for the elastic modulus, E, was taken in the outer loop. We assumed that Poisson's ratio was an aleatory variable, in contrast with the previous analyses in this paper. We performed second-order probability analysis using two approaches: (1) sampling both in the inner and outer loop, and (2) interval optimization in the outer loop using a Gaussian process EGO-type optimization and stochastic expansion in the inner loop.

4. Sampling Approach to Second-order Probability

Conditioned on a particular value of E from the outer loop, 10 samples of ν were taken on the inner loop. Over all outer loops, we then can calculate the minimum and maximum value of the 10^{th} percentile on the inner loop, or the median, or the 90^{th} percentile, etc. Graphically, the results for the second-order probability analysis based on eight outer loops samples of E, with 10 inner loop samples of ν per outer loop sample (80 samples total), are shown in Figures 11 and 12. The blue and pink lines show the minimum and maximum values of the 10^{th}, 50^{th}, and 90^{th} percentiles over all the inner loop empirical distribution functions, respectively. For example, the 10^{th} percentile of the shear mode frequency could lie anywhere between 1137 and 2850 Hz in this example. Note that in a real analysis, one would want to take more samples on both inner and outer loops to obtain more accurate estimates of the minimum and maximum percentiles: the few samples here are shown just for demonstration of the method. In practice, one would want to take at least 30-50 outer loop samples and possibly hundreds of inner loop samples, depending on the inner loop statistic of interest. Also note that the empirical distribution function created for each outer loop based on sampling the inner loop is nearly vertical in Figures 11 and 12. This will not usually be the case: this is due to the fact that varying Poisson's ratio has a very small effect on the mode frequencies relative to varying the elastic modulus, as discussed above.

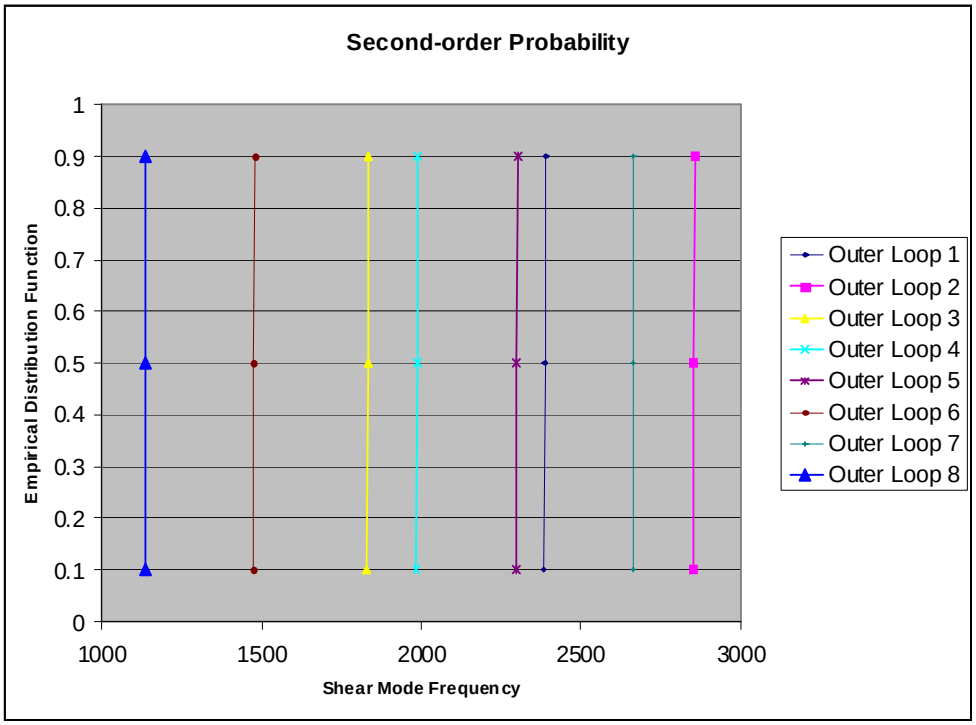

Figure 11. Second-order Probability Analysis for Shear Mode based on Nested Sampling

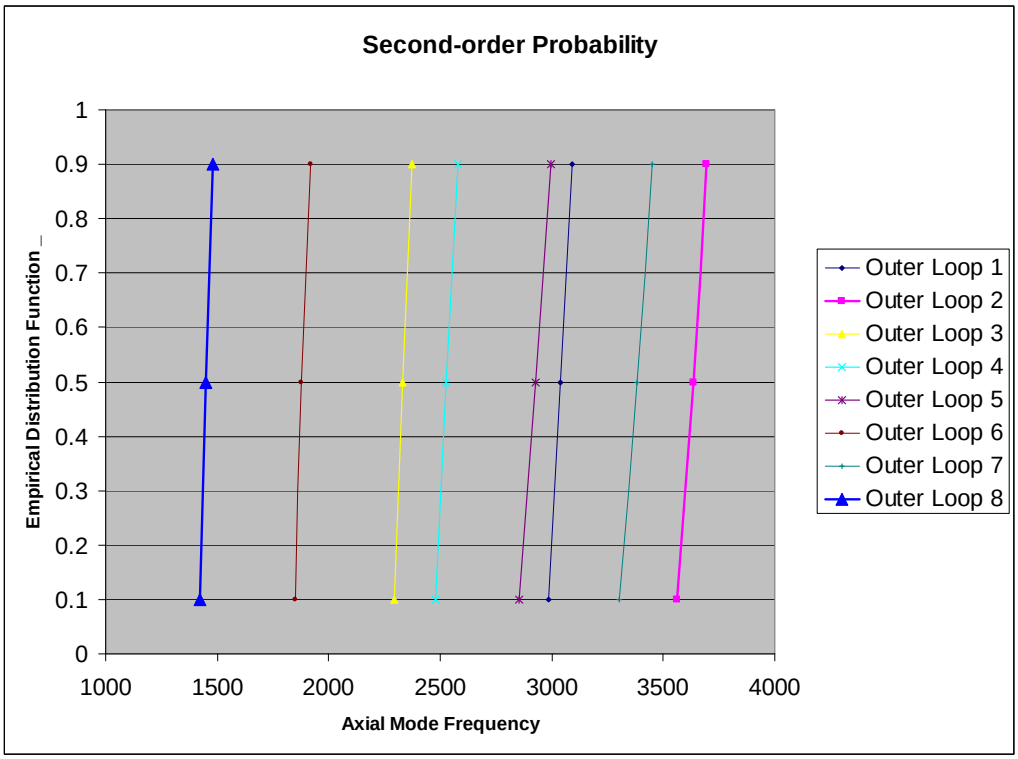

Figure 12. Second-order Probability Analysis for Axial Mode based on Nested Sampling

V. Margin Analysis

Section IV presented a variety of approaches for propagating epistemic uncertainty, demonstrating how epistemic uncertainty structures on the input variables E and ν propagated to epistemic uncertainty structures on the output measures such as shear and axial mode frequency. This section addresses the calculation of margin based on these epistemic output representations.

Helton (2009) provides an excellent reference for various ways to analyze margin, depending on the treatment and characterization of epistemic and aleatory uncertainties. He states: "a margin M is a measure of the difference between a requirement R placed on the performance of a system and the performance P of the system that will actually be realized, with $M \geq 0$ indicating that the requirement is met and $M < 0$ indicating that the requirement is not met." Requirements and performance measures may be vector-valued but for the purposes of this discussion, we are considering a single requirement and single performance measure: the frequency of the axial mode. In the simple case that R and P are single numerical values, the margin M(R,P) = R–P if the performance is required to be less than R, or M(R,P) = P–R if the performance is required to be greater than R. When epistemic uncertainties are involved, the performance is then conditional on these epistemic uncertainties: M(R,P|**e**), where **e** defines a vector of epistemic uncertain variables.

The important thing to remember when analyzing margins under uncertainty is that the *margins are also uncertain and have an uncertainty structure that derives from the uncertainty structure assumed for the uncertain inputs*. We illustrate this below for the cases we presented in Section IV: interval analysis, Dempster-Shafer theory, and second-order probability. Note that uncertainty may also be incorporated in the specification of a requirement (for example, instead of a fixed requirement, the requirement may be bounded in an interval). In the discussion below, we do not consider uncertainty on R, only on P. The reader is encouraged to study the reference by Helton (2009) for more details about translating the input uncertainty structures to margins. Although presenting the uncertainty on the margin may be helpful, Helton argues that it is also important to show the results in "performance" space, not margin space. He states "margin summaries are less informative than performance summaries because they obscure the actual value of the performance measure and the relationship of this measure to its associated requirement." (Helton, 2009) Helton also strongly recommends against reducing margin and uncertainty information down to one number such as margin/uncertainty, where M and uncertainty (U) are reduced one number each: "Bluntly put, "margins/uncertainty" results do not contain enough information to provide a basis for appropriately informed decisions."

The first case of margin analysis we present is based on interval analysis. For the purposes of this illustration, the requirement is that the frequency of the axial mode be no more than 4000 Hz (R=4000). In Section IV, we calculated several estimates of the margin on the axial mode frequency. For this discussion, we take the interval to be that obtained by the Gaussian process surrogate/EGO optimization: [1008 Hz, 3832Hz]. We have an interval on P, so that translates to an interval on M:

$$M = [4000-3832, 4000-1008] = [168, 2992].$$

In this case, the interval on margin is very large: from 168Hz to 2992 Hz. Recall that the wide interval on the elastic modulus translated to a wide interval on axial mode frequency, which translates to a wide margin. If R were within the interval, say R = 3000, the interval on margin would be [-832, 1992]. In this case, since the interval on margin includes negative values, we would say the margin is not met. Note that we cannot put a probability distribution on margin in this case because we only have an interval structure on the axial mode frequency. It is incorrect to say there is approximately a 1/3 chance of negative margin and 2/3 chance of positive margin in the interval [-832, 1992] Hz.

In the second case, we have a Dempster-Shafer representation of uncertainty in the axial mode, as shown in Figure 13. Given the requirement of 4000 Hz, this translates to a cumulative belief and cumulative plausibility function for the margin as shown in Figure 13. At 4000Hz, both the cumulative belief and plausibility that the axial mode is less than or equal to 4000 is 1. That is, the measures of credence associated with the performance set P, where $P_{4000} = \{\tilde{P} : \tilde{P} \in P, \tilde{P} \leq 4000\}$ are given by the cumulative belief and plausibility: $[Bel_P(P_{4000}), Pl_P(P_{4000})] = [1.0,1.0]$. The belief, $Bel_P(P_{4000})$, provides a measure of information that supports the proposition that the appropriate value for the axial mode frequency is contained in P_{4000} and the plausibility $Pl_P(P_{4000})$, provides a measure of the amount of information that does not refute the proposition that the appropriate value for the axial mode frequency is contained in P_{4000}. In this case, the belief and plausibility indicate that margin would be satisfied. However, if the requirement were at 3800Hz instead of 4000, we would have a statement like:

$[Bel_P(P_{3800}), Pl_P(P_{3800})] = [0.9,1.0]$. In this case, there would be a small amount of evidence $Bel_P(P_{3800}^C) = 0.10$ that supports the proposition that the axial mode frequency is greater than 3800.

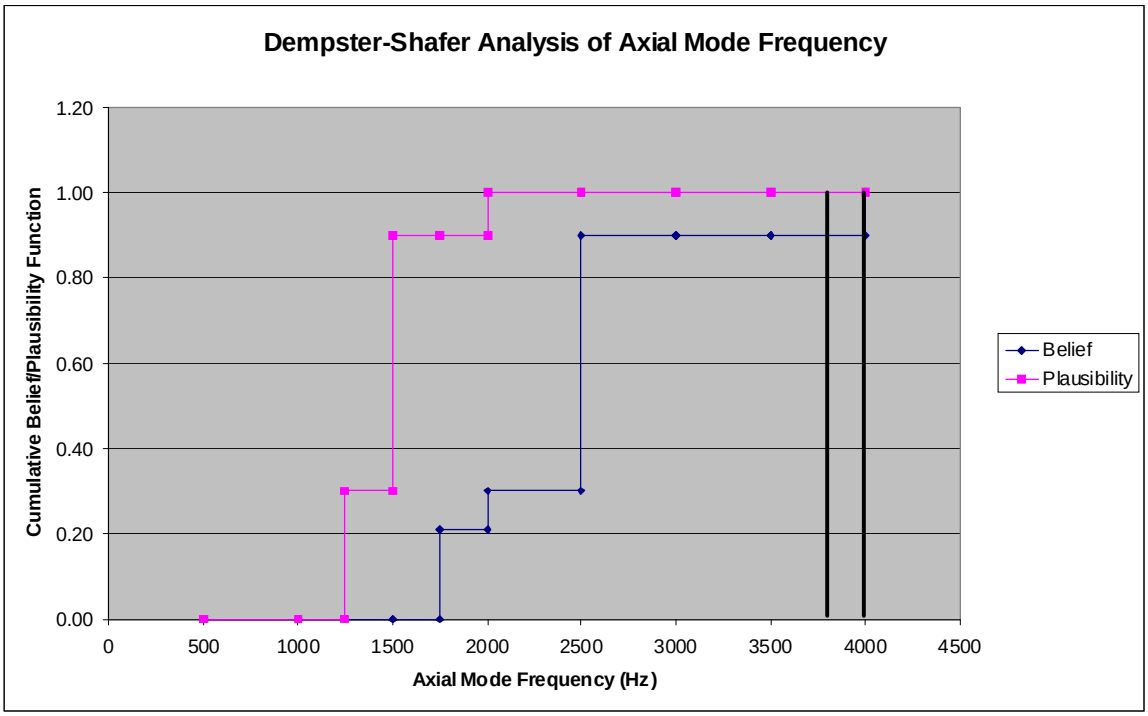

Figure 13. Margins in Dempster-Shafer Analysis.

The third case we present is based on second-order probability analysis. In this case, we have an ensemble or family of CDFs, which provide a bounding interval. There are a variety of ways to define margins. In some cases, margins may be defined for moments or percentiles (e.g. the median response must be less than R), and in this situation, one would take the interval on the median defined by the "envelope" of CDFs, and proceed as in the case of interval analysis to generate an upper and lower bound on the margin for the median requirement. In more complex situations, multiple requirement thresholds may be specified. As an example, the Environmental Protection Agency (EPA) specified requirements on the maximum acceptable normalized radionuclide releases from Waste Isolation Pilot Plant (WIPP), a nuclear waste repository. Performance assessment of this facility (Helton, 09), through modeling and simulation, demonstrated that the WIPP facility will meet the requirements of the maximum acceptable probabilities for exceeding normalized releases of 1 and 10 units (these requirements are given by RL1=0.1 and RL2 = 0.001). Since the "family" or ensemble of CCDF curves shown in Figure 14 is all below the EPA limit, the facility was deemed to have adequate margin.

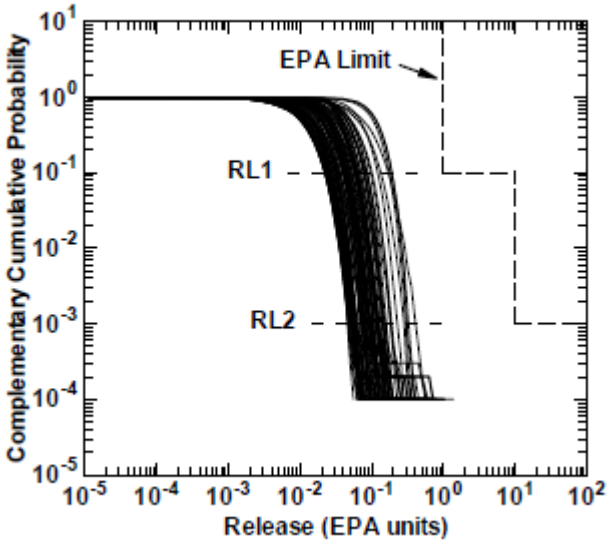

Figure 14. Estimated Complementary Cumulative Distribution Functions for normalized release over 10,000 years generated from a Latin Hypercube sample of size 100 (outer loop) with aleatory samples of 10,000 (inner loop). Taken from Helton, 2009.

In the structural dynamics example, with a fixed requirement of 4000 Hz, margin is also met, as shown in Figure 15. If the margin requirement of 4000 Hz were specified for a particular percentile, such as the 70th percentile, we could calculate the margin at that percentile based on the interval defined by the pink and blue curves shown in Figure 15. For the 70th percentile, for example, the interval on margin is M = [4000-3650, 4000-1460] = [350,2540]Hz.

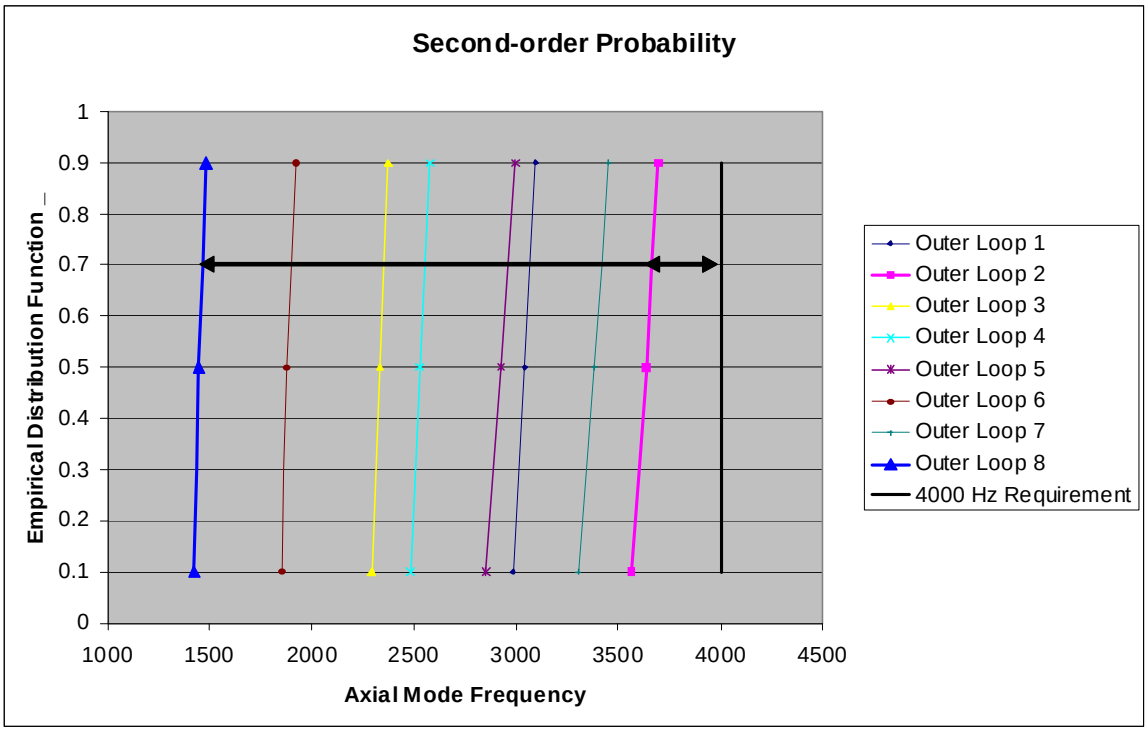

Figure 15. Margin based on 4000 Hz requirement. Note that margin can be calculated relative to a particular percentile, such as the 70th percentile in this example

VI. Summary

This paper has presented several approaches used in characterizing and propagating epistemic uncertainty: interval analysis, Dempster-Shafer evidence theory, and second-order probability. The use of surrogate models, both data-fit surrogate models such as quadratic regression models and Gaussian processes, and stochastic expansions such as polynomial chaos, were discussed and demonstrated. The use of surrogate models was shown to be useful in the propagation of uncertainty. In general, uncertainty quantification methods are computationally intensive. Epistemic methods that require use of optimization (such as finding upper and lower bounds in interval analysis, finding bounds in Dempster-Shafer focal elements, and performing "looped" UQ in second-order probability analysis) all are more efficient when surrogates are able to be used. The paper provided examples of the three types of epistemic uncertainty treatment on a problem in structural dynamics: characterizing the uncertainty in the axial and shear mode frequencies given uncertainties in the material properties of elastic modulus and Poisson's ratio. The paper also demonstrated how the propagation of epistemic uncertainty translated to the assessment of margins.

Acknowledgments

The authors thank the Advanced Simulation and Computing (ASC) Program of the National Nuclear Security Administration (NNSA) at Sandia National Laboratories for funding this research.

References

Alexandrov, N., J. E. Dennis, Jr., R. M. Lewis, and V. Torczon, "A Trust Region Framework for Managing the Use of Approximation Models in Optimization", *Structural Optimization*, 15(1), pp. 6-12, 1998.

Diegert, K., Klenke, S., Novotny, G., Paulsen, R., Pilch, M. and T. Trucano. "Toward a More Rigorous Application of Margins and Uncertainties within the Nuclear Weapons Life Cycle – A Sandia Perspective." Sandia Technical Report SAND2007-6219, October 2007.

Eldred, M.S., Adams, B.M., Gay, D.M., Swiler, L.P., Haskell, K., Bohnhoff, W. J., Eddy, J.P., Hart, W.E., Watson, J.-P., Griffin, J.D., Hough, P.D., Kolda, T.G., Williams, P.J, and Martinez-Canales, M.L., "DAKOTA, A Multilevel Parallel Object-Oriented Framework for Design Optimization, Parameter Estimation, Uncertainty Quantification, and Sensitivity Analysis: Version 4.1 Users Manual," Sandia Technical Report SAND2006-6337, October 2006, updated September 2007.

Eldred, M.S., Adams, B.M., Gay, D.M., Swiler, L.P., Haskell, K., Bohnhoff, W. J., Eddy, J.P., Hart, W.E., Watson, J.-P., Griffin, J.D., Hough, P.D., Kolda, T.G., Williams, P.J, and Martinez-Canales, M.L., "DAKOTA, A Multilevel Parallel Object-Oriented Framework for Design Optimization, Parameter Estimation, Uncertainty Quantification, and Sensitivity Analysis: Version 4.1 Reference Manual," Sandia Technical Report SAND2006-4055, October 2006, updated September 2007.

Eldred, M.S. and Dunlavy, D.M., "Formulations for Surrogate-Based Optimization with Data Fit, Multifidelity, and Reduced-Order Models," paper AIAA-2006-7117 in the *Proceedings of the 11th AIAA/ISSMO Multidisciplinary Analysis and Optimization Conference*, Portsmouth, VA, Sept. 6-8, 2006.

Eldred, M.S., Webster, C.G., and Constantine, P., "Evaluation of Non-Intrusive Approaches for Wiener-Askey Generalized Polynomial Chaos," paper AIAA-2008-1892 in *Proceedings of the 49th AIAA/ASME/ASCE/AHS/ASC Structures, Structural Dynamics, and Materials Conference* (10th AIAA Non-Deterministic Approaches Conference), Schaumburg, IL, April 7-10, 2008.

Giunta, A.A., McFarland, J. M., Swiler, L.P., and Eldred, M.S., "The promise and peril of uncertainty quantification using response surface approximations," *Structure & Infrastructure Engineering: Maintenance, Management, Life-Cycle Design & Performance*, special issue on Uncertainty Quantification and Design under Uncertainty of Aerospace Systems, Vol. 2, Nos. 3-4, Sept.-Dec. 2006, pp. 175-189.

Haldar, A., and S. Mahadevan, *Probability, Reliability, and Statistical Methods in Engineering Design*, John Wiley and Sons, 2000.

Helton, J. C. "Conceptual and Computational Basis for the Quantification of Margins and Uncertainty." SAND Report 2009-XXXX.

Helton, J.C., Johnson, J.D. and W.L. Oberkampf. "An Exploration of Alternative Approaches to the Representation of Uncertainty in Model Predictions." *Reliability Engineering and System Safety* Vol. 85, pp. 39-71, 2004.

Helton, J.C., Johson, J.D., Oberkampf, W.L. and C.B. Storlie. "A sampling-based computational strategy for the representation of epistemic uncertainty in model predictions with evidence theory." Sandia National Laboratories Technical Report SAND2006-5557.

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

Martin, J. D. and T. W. Simpson. "Use of Kriging Models to Approximate Deterministic Computer Models." *AIAA Journal* 2005, Vol.43 no.4 (853-863).

McFarland, J. M., S. Mahadevan, V. J. Romero, and L. P. Swiler. "Calibration and Uncertainty Analysis for Computer Simulations with Multivariate Output." *AIAA Journal* 2008, Vol.46 no.5 (1253-1265).

Meckesheimer, M., A. J. Booker. R. R. Barton, and T. W. Simpson. "Computationally Inexpensive Metamodel Assessment Strategies." *AIAA Journal* 2002, Vol.40 no.10 (2053-2060).

Reese, G., Segalman, D., Bhardwaj, M. K., Alvin, K., Driessen, B., Pierson, K., and T. Walsh. Salinas User's Notes. Sandia National Laboratories, Aug. 2008. Salinas Release 2.7.0.

Sacks, J., Welch, W. J., Mitchell, T. J., and Wynn, H. P., "Design and Analysis of Computer Experiments," *Statistical Science*, Vol. 4, No. 4, 1989, pp. 409–435.

Swiler, L. P., Paez, T. L., and R. L. Mayes. "Epistemic Uncertainty Quantification Tutorial." SAND 2008-6578C, paper 294 in the Proceedings of the IMAC XXVII Conference and Exposition on Structural Dynamics, Society for Structural Mechanics. Orlando, FL, Feb. 2009.

Swiler, L. P, R. Slepoy and A. A. Giunta. "Evaluation of Sampling Methods in Constructing Response Surface Approximations." Paper AIAA-2006-1827 in *Proceedings of the 47th AIAA/ASME/ASCE/AHS/ASC Structures, Structural Dynamics, and Materials Conference*, Newport, Rhode Island, May 1-4, 2006.

Swiler, L.P. and Wyss, G.D., "A User's Guide to Sandia's Latin Hypercube Sampling Software: LHS UNIX Library Standalone Version," Sandia Technical Report SAND2004-2439, July 2004.

Xiu, D. and Karniadakis, G. M., "The Wiener-Askey Polynomial Chaos for Stochastic Differential Equations," *SIAM Journal of Scientific Computing*, Vol. 24, No. 2, 2002, pp. 619–644.

Wang, L., Beeson, D. and G. Wiggs. "A Comparison of Meta-Modeling Methods using Practical Industry Requirements." Paper AIAA-2006-1811 in *Proceedings of the 47th AIAA/ASME/ASCE/AHS/ASC Structures, Structural Dynamics, and Materials Conference*, Newport, Rhode Island, May 1-4, 2006.