Epistemic Uncertainty Quantification Tutorial

Laura P. Swiler, Thomas L. Paez, Randall L. Mayes
Sandia National Laboratories, New Mexico
PO Box 5800
Albuquerque, NM 87185-1318
USA
lpswile@sandia.gov, tlpaez@sandia.gov, rlmayes@sandia.gov

ABSTRACT

This paper presents a basic tutorial on epistemic uncertainty quantification methods. 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). 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.

Keywords: Epistemic uncertainty, interval analysis, evidence theory.

1. Introduction

The treatment of uncertainty in the analysis of computer models is essential for understanding possible ranges of outputs or scenario implications. 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, Type B, or reducible uncertainty, meaning that the uncertainty can be reduced through increased understanding (research), or increased and more relevant data. [7,8] Epistemic quantities are sometimes referred to as quantities which have a fixed value in an analysis, but we do not know that fixed value. For example, the elastic modulus for the material in a specific component is presumably fixed but unknown or poorly known. In contrast, uncertainty characterized by inherent randomness which cannot be reduced by further data is called aleatory uncertainty. Some examples of aleatory uncertainty are weather or the height of individuals in a population: these cannot be reduced by gathering further information. Aleatory uncertainty is also called stochastic, variability, irreducible and type A uncertainty. Aleatory uncertainties are usually modeled with probability distributions, but epistemic uncertainty may or may not be 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 [1].

There are many ways of representing epistemic uncertainty, including probability theory, fuzzy sets, possibility theory, and imprecise probability. The problem of selecting an appropriate mathematical structure to represent epistemic uncertainties can be challenging. 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. Section 2 presents a structural dynamics example that will be used to demonstrate the various methods. Section 3 discusses interval analysis and shows results, Section 4 discusses evidence theory and shows results, and Section 5 discusses second-order probability and associated results. Section 6 summarizes the paper.

2. Motivating 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 1 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 2. 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 [11,12], 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.

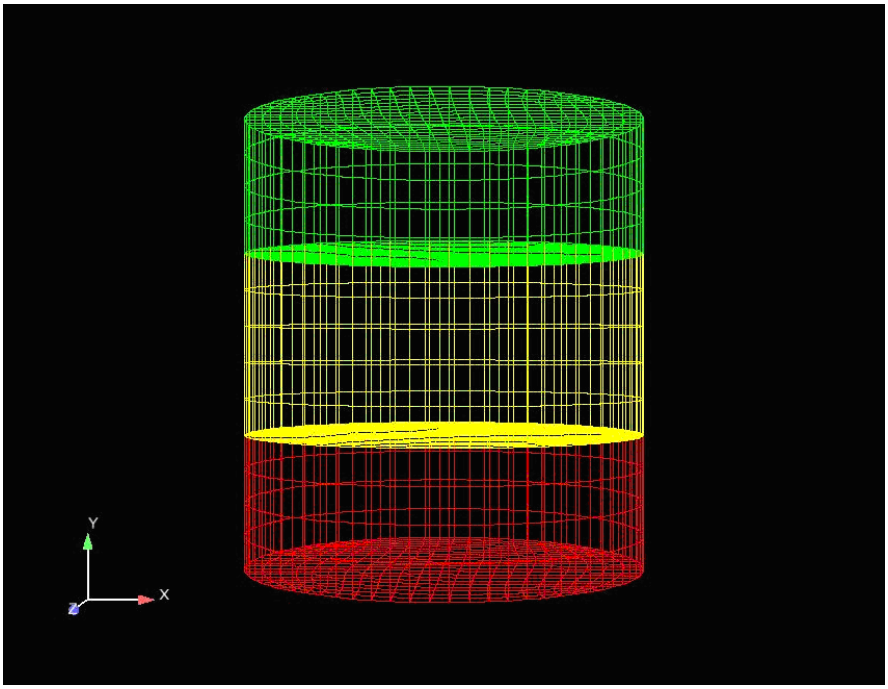

Figure 1. 3 disc model with filled rubber as the middle disc (in yellow)

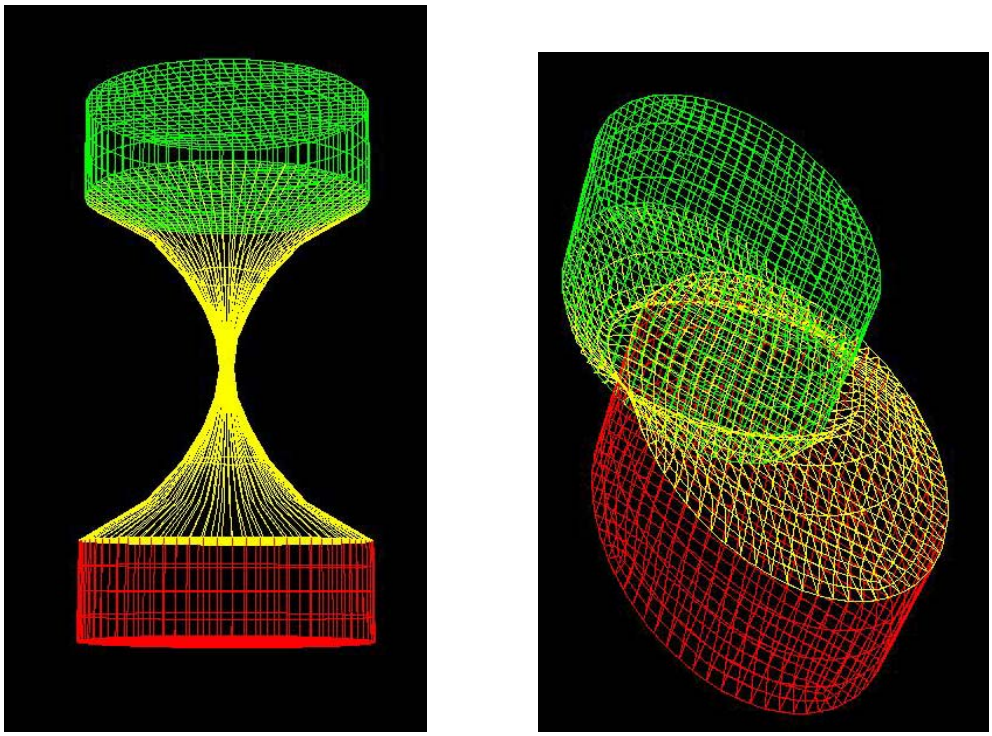

Figure 2. Axial Mode (left) and Shear Mode (right) for 3-disc model

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 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].

We used DAKOTA [2,3], a software framework that allows one to perform uncertainty quantification, optimization, and parameter studies (see: http://www.cs.sandia.gov/DAKOTA/) to perform the computational runs of the Salinas model presented in subsequent sections. DAKOTA was configured to drive the analysis for 3 case studies: pure interval analysis, Dempster-Shafer evidence theory, and second-order probability analysis.

3. 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 [6,8]. That is, there is no particular structure on the possible values for the epistemic uncertain variables except that they lie within bounds. 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 is 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.

In this paper, a single input variable is represented as x_i, **x** represents the vector of m uncertain variables, and the output y is a function of **x**: $y = F(\mathbf{x})$. Figure 3 shows a Monte Carlo sampling approach that is often used to propagate aleatory uncertainty. In this figure, there are 3 input parameter distributions ($m = 3$) represented on the left side. Five samples are taken from each (N=5), and the simulation model is run five times with these sets of input, resulting in 5 realizations of the output y shown on the right. In the case of aleatory uncertainty propagation, one can interpret the resulting output samples probabilistically and fit an appropriate distribution.

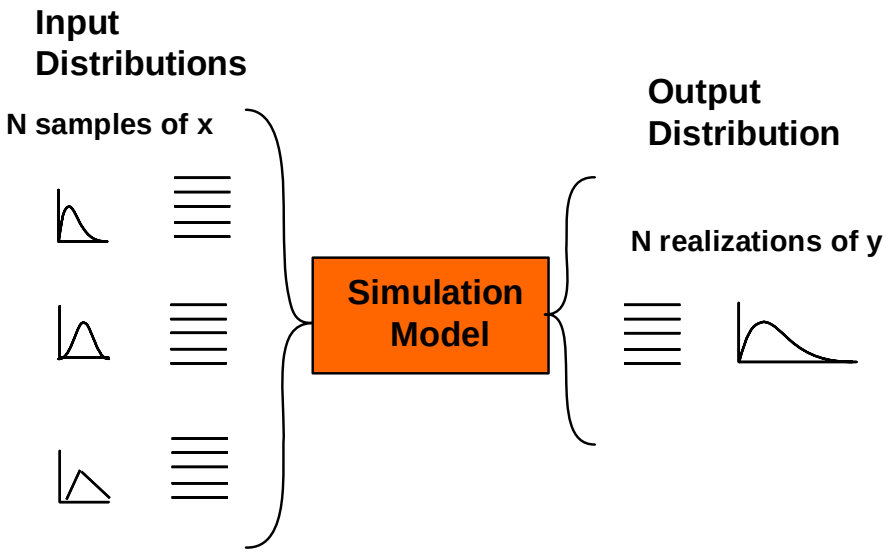

Figure 3. Monte Carlo Sampling used for Aleatory Uncertainty Propagation

Figure 4 shows how Monte Carlo sampling may be used to propagate epistemic uncertainty. Note that the input distributions are all represented by intervals, and so is the output. As mentioned above, one must be careful not to interpret the result with any type of structure other than an interval on the output. Also, while sampling is easy to implement, it may underestimate the true output interval.

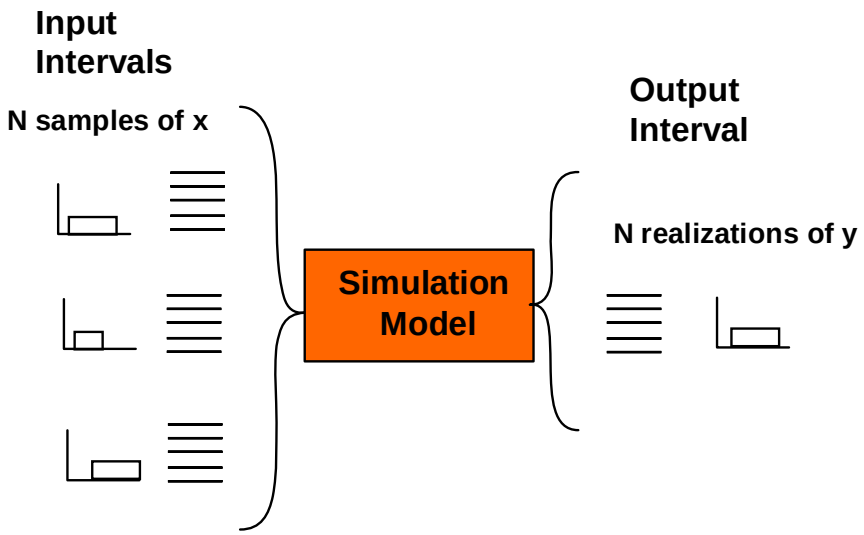

Figure 4. Monte Carlo Sampling used for Epistemic Interval Propagation

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.

3.1. Interval Results: Sampling

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 5 and 6. 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

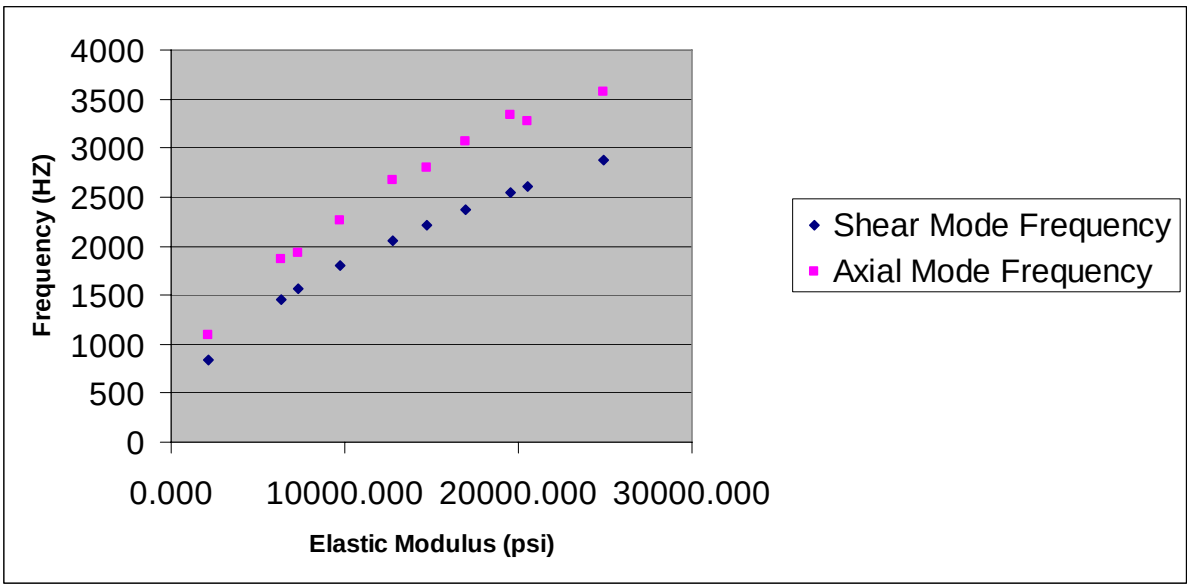

Figure 5. Shear and Axial Mode Frequencies as a function of E for 10-sample case

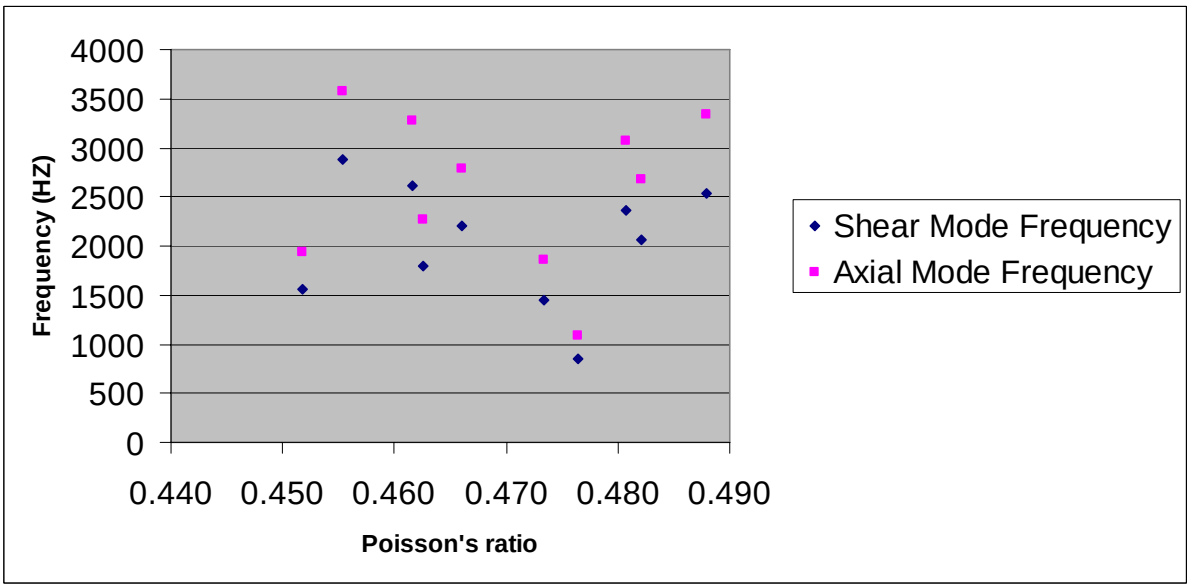

Figure 6. Shear and Axial Mode Frequencies as a function of ν for 10-sample case

Figures 5 shows that both the axial and shear mode frequencies are almost perfectly linearly correlated with the elastic modulus, E. However, Figure 6 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 analysis. Note that an initial interval analysis may also be used for sensitivity analysis and to identify issues and perform further iterations, which is what we demonstrate here.

To further examine the effects of E and ν on the axial and shear mode, and to determine if there were some interactions between E and ν, we performed a 36 sample orthogonal array study where we had six levels of E and 6 levels of ν, so we had 36 sample points. Orthogonal arrays allow one to calculate "main effects." That is, with an orthogonal array, you can calculate the mean of the shear mode frequency (for example) with E being fixed as 2000psi as ν varies from 0.45 to 0.495. If this mean is nearly that same as the mean shear mode frequency when E is fixed at 10000psi or 25000 psi (again, averaging over ν), we say that E does not have a strong influence. However, if the mean shear mode frequency with E fixed at 2000psi and the mean shear mode frequency with E fixed at 10000psi or 25000psi are statistically significantly different, than E has a strong main effect. Figures 7

and 8 show that as E is varied between 2000 and 25000, the mean frequency response of both the axial and shear modes varies significantly. However, as ν is varied between 0.45 and 0.49, the mean response variation is NOT statistically significant. The significance tests were very strong for both cases: p-values of 0 for E and p-values of 1.0 for ν.

Note that the output intervals for the shear mode frequency based on the orthogonal array results is [813.0,2884.0] Hz, and the output interval for the axial mode frequency is [1008.1,3831.9] Hz which are wider than those obtained by the initial ten samples.

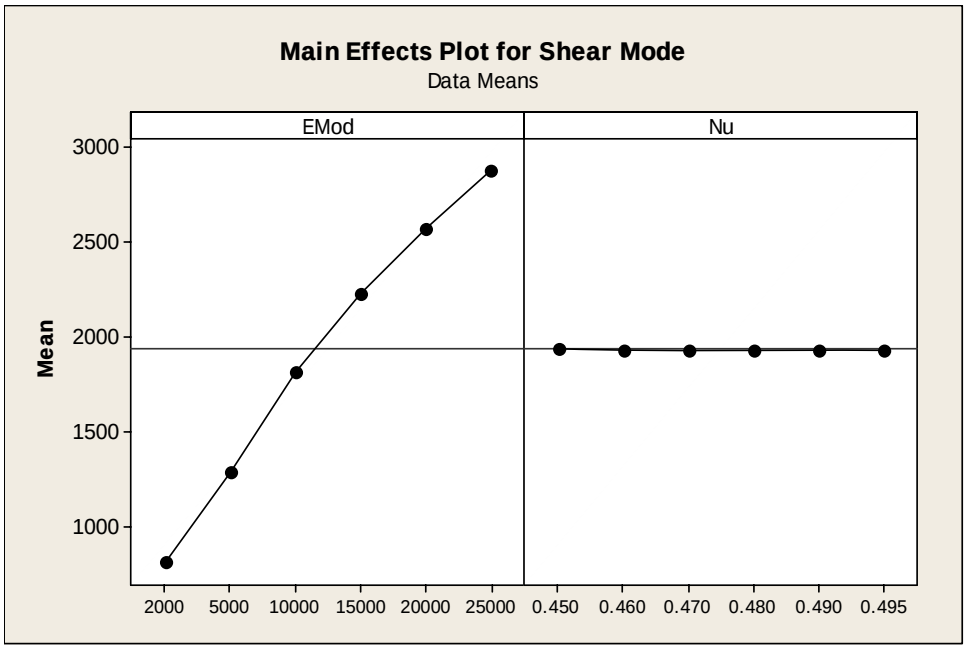

Figure 7. Main Effects for Shear Mode

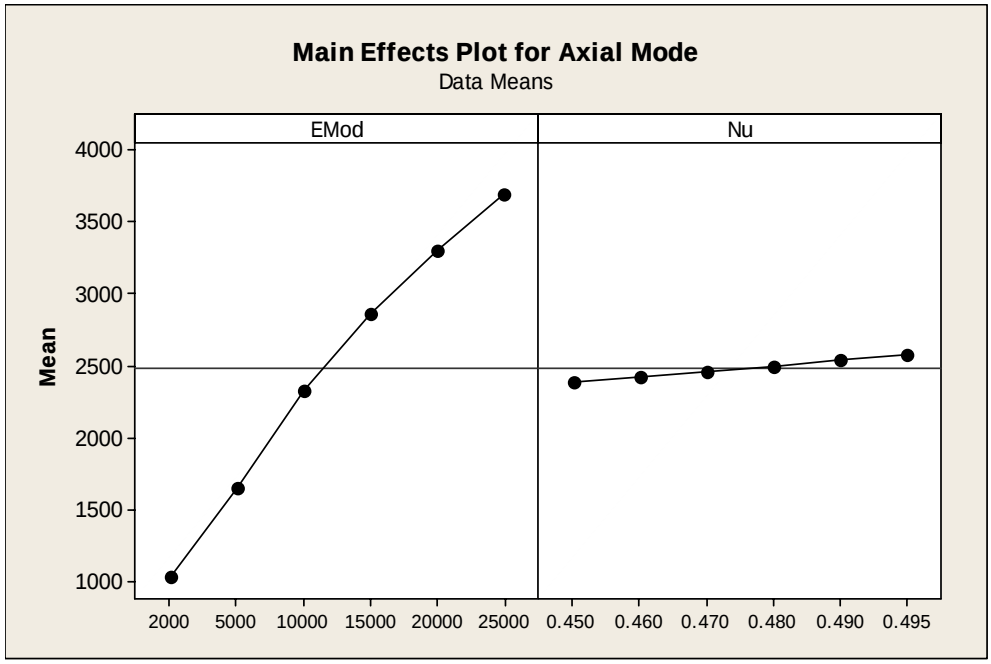

Figure 8. Main Effects for Axial Mode

Figures 9 and 10 show the interaction effects. That is, they show as you keep E fixed (for example) at 2000 and vary ν (the black line in the upper right box), the shear mode frequency does not change: it is around 813 Hz. There is a similar pattern for the other levels of E as ν varies. However, in the lower left box, as Emod varies, the

value of ν does not matter much: all the lines are on top of each other, meaning that at E of 10000 (for example), the shear mode frequency is about 1820Hz no matter what the value of ν. The interaction plot for the axial mode is similar, though we do see a slight influence of ν. There is no significant interaction between E and ν, at least in this model. If you look at the shear mode frequency for a particular E value, as ν is increased from 0.45 to 0.495, the shear frequency does decrease, but the decrease is around 5 Hz, which is insignificant when measured against the change related to increasing E from 2000 to 25000.

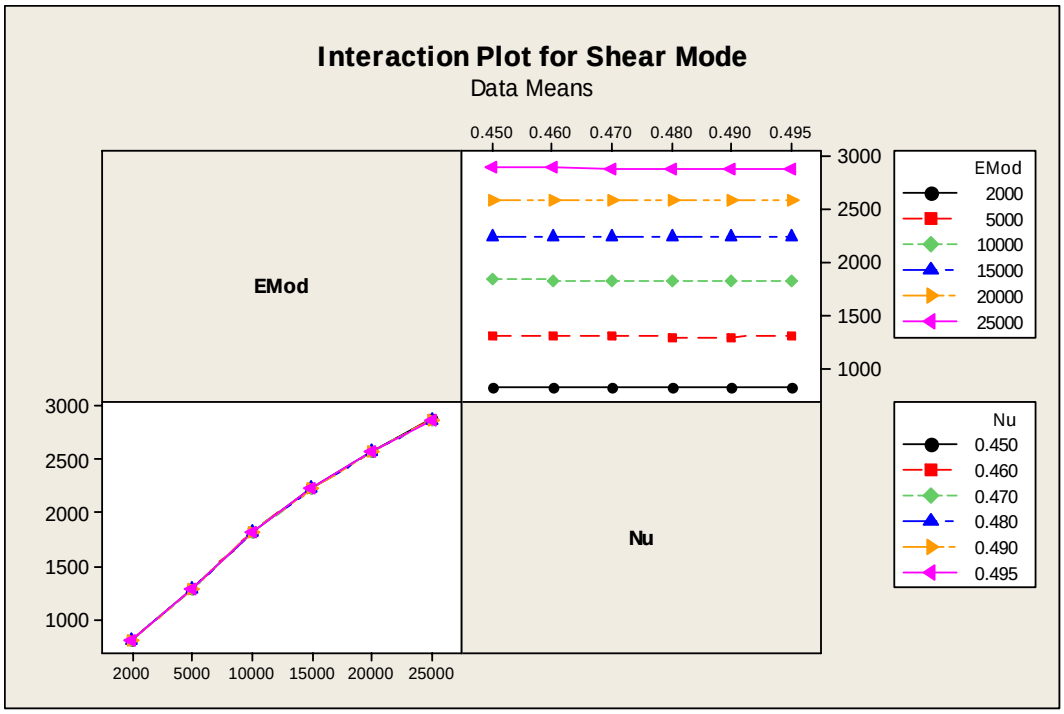

Figure 9. Interaction Effects for Shear Mode

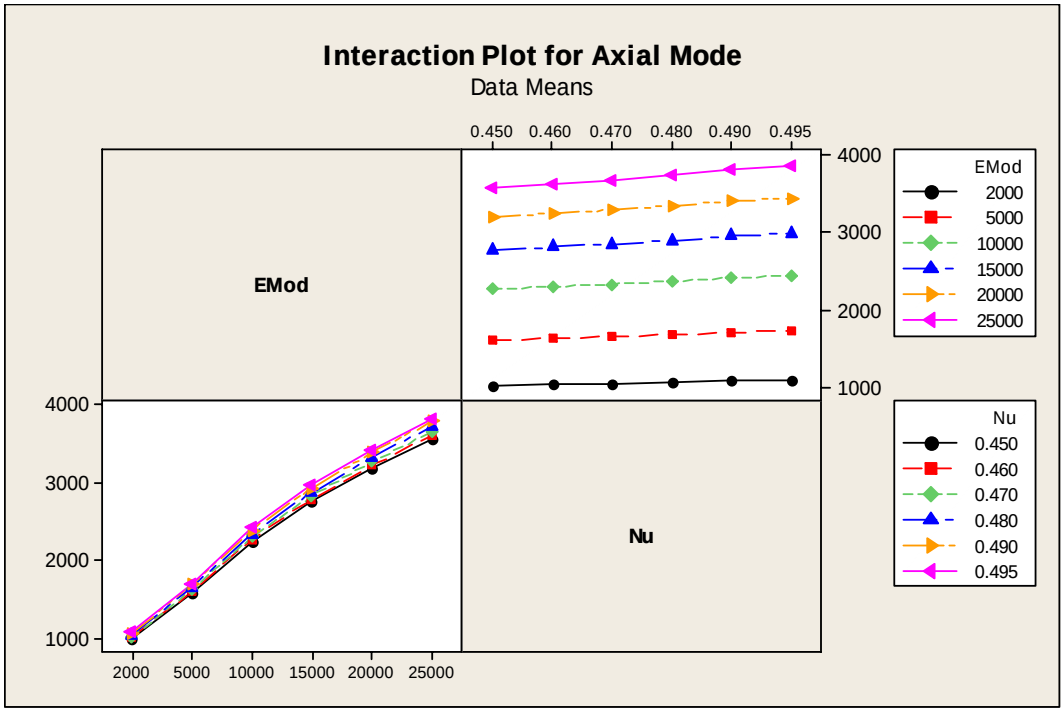

Figure 10. Interaction Effects for Axial Mode

Based on all of this analysis, we 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.

3.2. Interval Results: Surrogate-based Methods

This section discusses the use of surrogates to determine interval output bounds. 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 [5,15,16]. 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 [4]. 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.

We first combined the samples generated above (the ten sample, 30 sample, and 36 sample orthogonal array) to create a full data set with 76 sample points. Then, we constructed a few different surrogate models based on these points: a quadratic regression model, a MARS model (multivariate adaptive splines), and a neural network. 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.

	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

Table 2. Interval bounds according to sampling a surrogate model

Finally, we used an optimization method on the surrogate to determine the upper and lower bounds. 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.

Note that 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 interpreting statistical measures based on surrogate builds extremely accurately [5,16].

	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
	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

Table 3. Interval bounds obtained according to optimizing a surrogate model

4. Dempster-Shafer Evidence Theory

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. [7]

Dempster-Shafer Theory of Evidence may be used to perform epistemic analysis [6,8,13,14,18]. It relaxes the assumptions of probability theory in situations where there is little information on which to evaluate a probability or when the information is nonspecific, ambiguous, or conflicting. For example, if an expert believes that a system may fail due to a particular component with a likelihood of 0.3, does that necessarily mean that the expert believes the system will not fail due to that component with a probability of 0.7? There may be certain pieces of evidence, which when considered in combination, lend more or less credence to the likelihood of an event. The Dempster-Shafer theory can account for evidence that can be assigned to multiple possible events (sets of events) whereas in probability theory, evidence is associated with only one possible event. Additionally, Dempster-Shafer theory can handle conflicting evidence. For example, if two people report that they saw a tree branch fall on your car, you would have a higher degree of belief that a tree limb did in fact land on your car than if one of the individuals said it fell on your car and the other individual reported it did not fall on your car.

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, not a single probability distribution.

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 a lot of interval cells and few samples, the estimates for the minimum and maximum function evaluations is likely to be poor. Surrogate methods may also be used in Dempster-Shafer, either global surrogates or separate surrogates within each cell.

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 11 below. The resulting cumulative distribution functions of belief and plausibility for the shear mode frequency are shown in Figure 12 and for the axial mode frequency are shown in Figure 13. 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: 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: Pl($y^* \leq y$). For example, in Figure 12, 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 13 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.

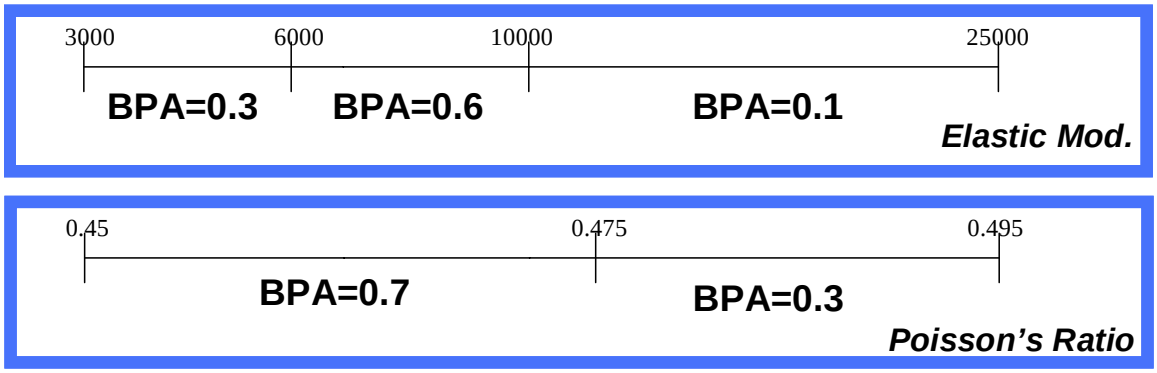

Figure 11. Intervals and associated BPAs for Dempster-Shafer analysis

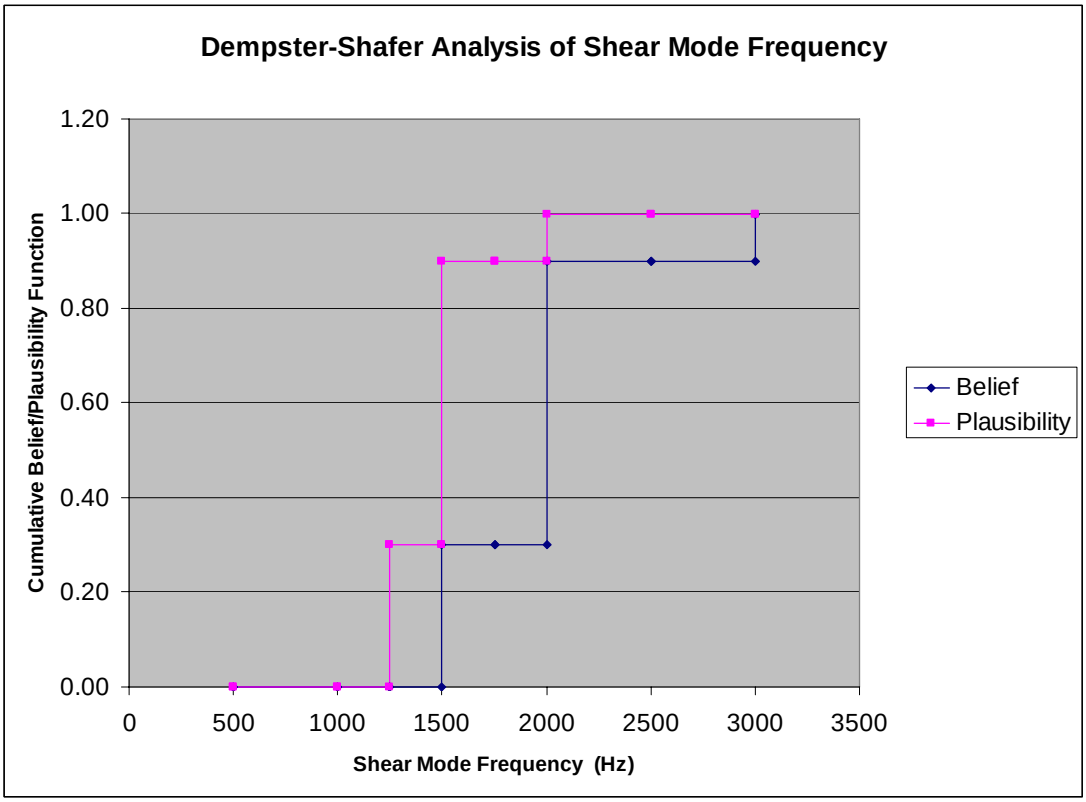

Figure 12. Cumulative Belief and Plausibility Distributions for Shear Mode

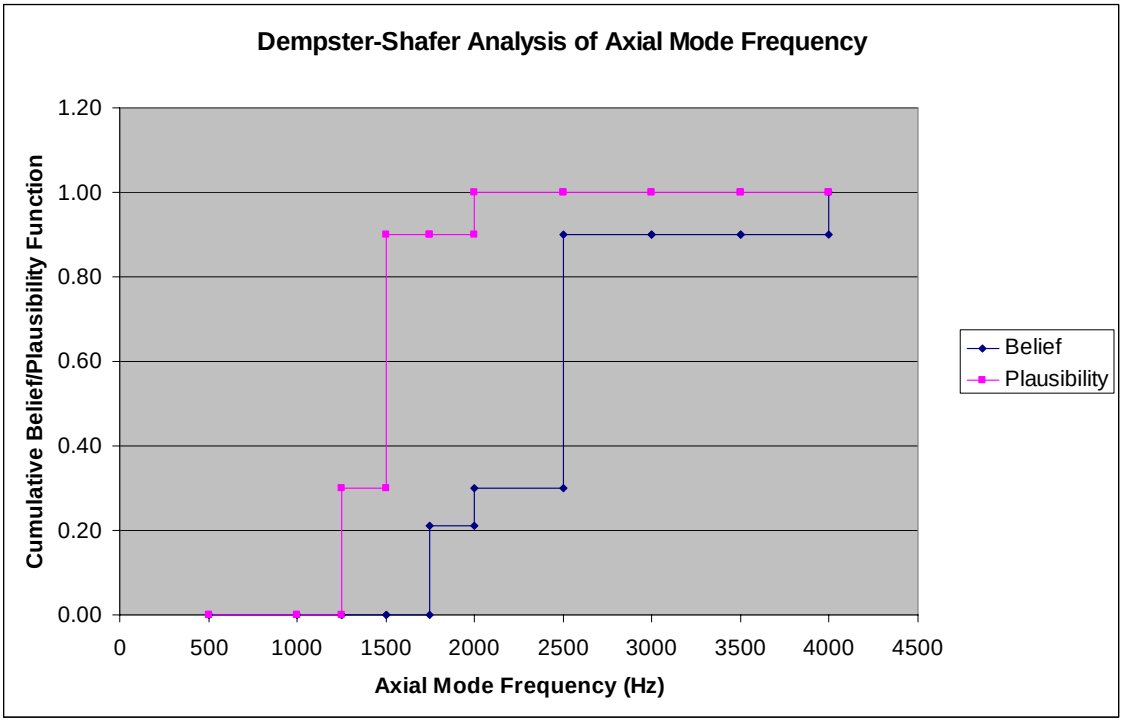

Figure 13. Cumulative Belief and Plausibility Distributions for Axial Mode

5. 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 14 shows the sampling structure of second-order probability. Second-order probability approaches have been used extensively in the performance assessment for nuclear waste repositories [9] and in nuclear reactor safety assessments [10]. There is a strong regulatory precedent for using this approach.

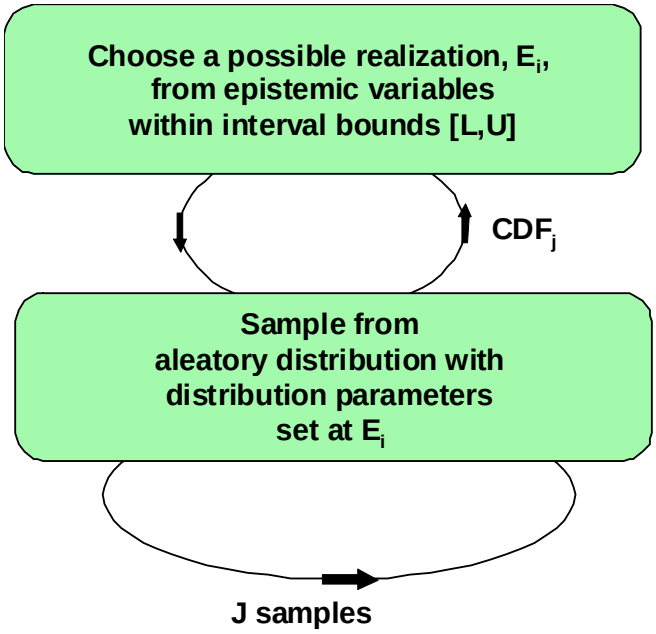

Figure 14. Second-order Probability

Second-order probability may be expensive since we have 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. When you plot the 50 CDFs, you get the upper and lower bound on the family. 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.

Continuing with our 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. 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 15 and 16. 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 15 and 16. 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.

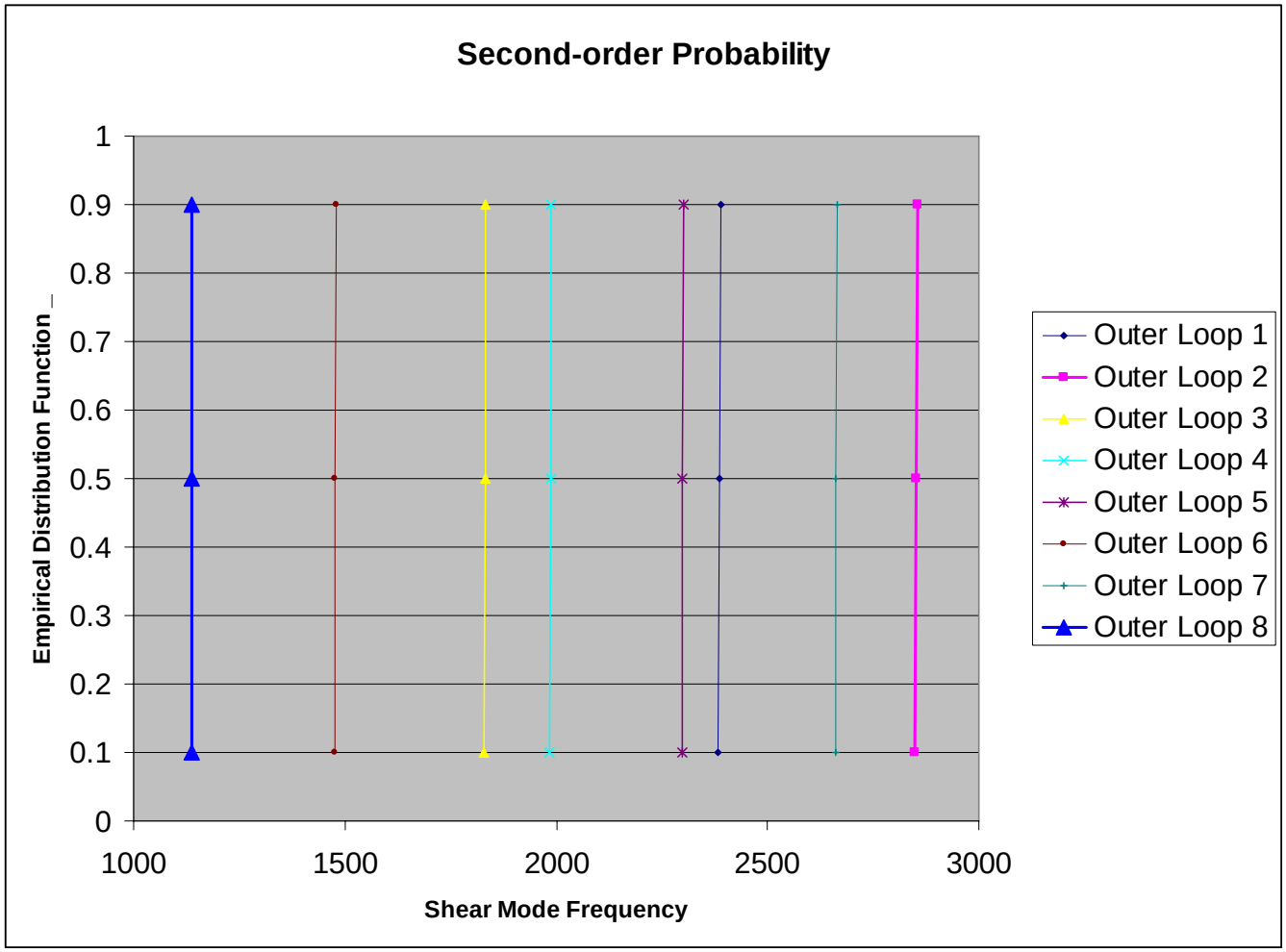

Figure 15. Second-order Probability Analysis for Shear Mode

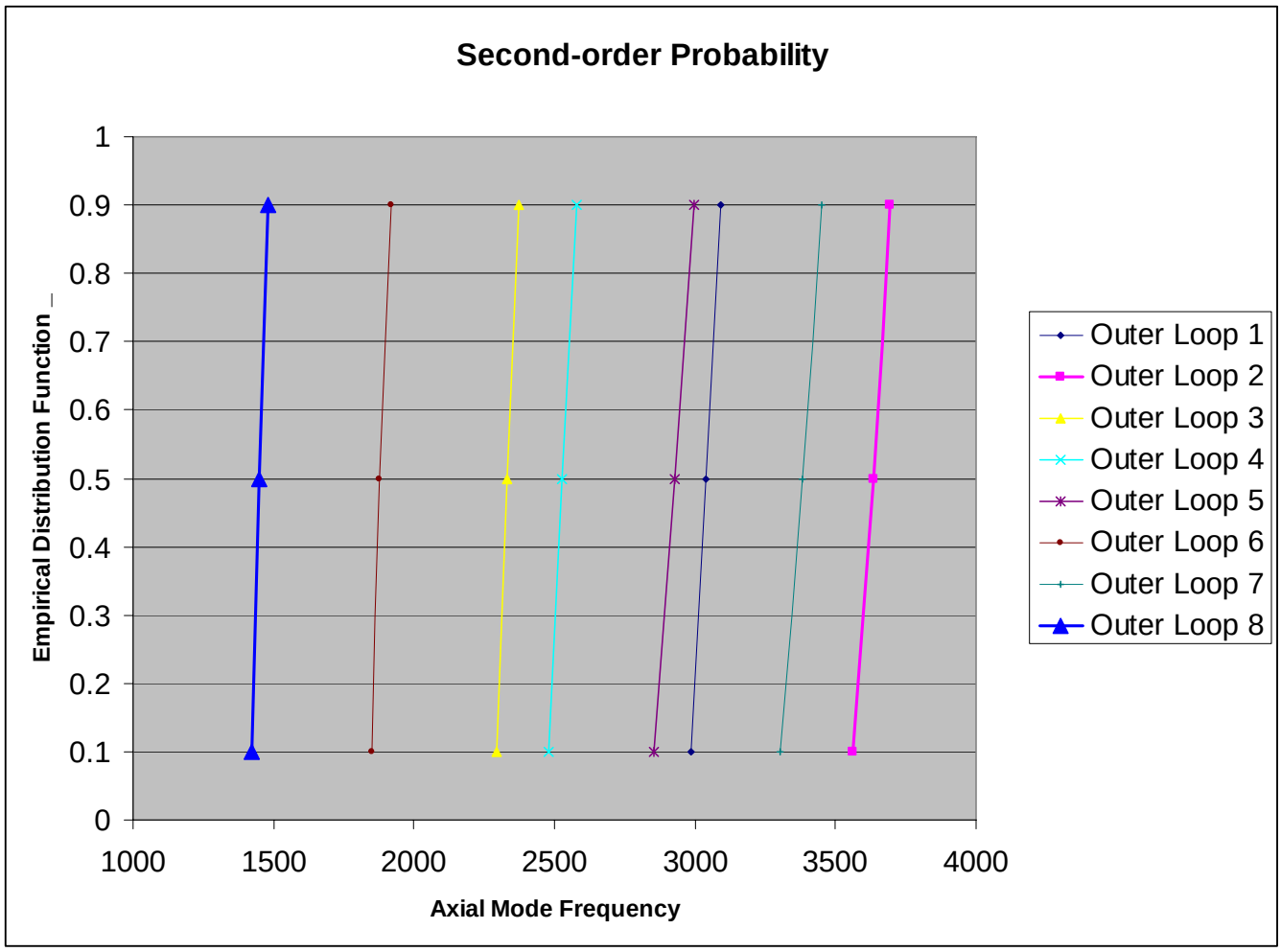

Figure 16. Second-order Probability Analysis for Axial Mode

6. Summary

This paper has presented a basic overview of three methods that are often used to quantify and propagate epistemic uncertainty in uncertainty analyses. Epistemic uncertainty, characterizing lack-of-knowledge, is often prevalent in engineering applications, but it is often treated (incorrectly) probabilistically as aleatory information. We outlined and demonstrated three methods used in propagating epistemic uncertainties: interval analysis, Dempster-Shafer evidence theory, and second-order probability. The structural dynamics problem provided a realistic example of the epistemic treatment of material properties (elastic modulus and Poisson's ratio) to understand how the lack of knowledge about these properties affects the shear and axial mode frequencies.

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

[1] 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.

[2] 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.

[3] 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.

[4] 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.

[5] 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.

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

[7] Helton, J.C., Johnson, 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.

[8] 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.

[9] Helton, J.C., Anderson, D. R., Jow, H.-N., Marietta, M.G., and G. Basabilvazo. “Performance Assessment in Support of the 1996 Compliance Certification Application for the Waste Isolation Pilot Plant.” *Risk Analysis*, Vol. 19, No. 5, 1999.

[10] Helton, J.C., and R. J. Breeding. “Calculation of reactor accident safety goals.” *Reliability Engineering and System Safety* Vol. 39, pp. 129-158, 1993.

[11] 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.

[12] Reese, G., Bhardwaj, M. K., and T. Walsh. Salinas-Theory Manual. Sandia National Laboratories, Aug. 2008. Salinas Release 2.7.0.

[13] Sentz, K. and S. Ferson. “Combination of Evidence in Dempster-Shafer Theory” Sandia National Laboratories Technical Report SAND 2002-0835.

[14] Shafer, G. *A Mathematical Theory of Evidence.* Princeton, NJ: Princeton University Press, 1976.

[15] Simpson, T. W., Toropov, V., Balabanov, V. and F. A. C. Viana. "Design and Analysis of Computer Experiments in Multidisciplinary Design Optimization: A review of how far we have come – or not." *12th AIAA/ISSMO Multidisciplinary Analysis and Optimization Conference*, Sept. 2008, Victoria Canada.

[16] 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.

[17] 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.

[18] Yager, R. R., & Liu, L. *Classic works of the Dempster-Shafer theory of belief functions*. Studies in fuzziness and soft computing series, v. 219. Berlin: Springer, 2008.