A Comparison of Sampling Techniques for Uncertainty Quantification

Steven Holman\textsuperscript{1}, Cory Rupp\textsuperscript{2}, Ryan Maupin\textsuperscript{3}

\textsuperscript{1}Graduate Research Assistant, University of Illinois, Urbana-Champaign, IL, 61820
\textsuperscript{2}Graduate Research Assistant, University of Colorado, Boulder, CO, 80302
\textsuperscript{3}Weapon System Analysis Team Leader, ESA-WR, Los Alamos National Laboratory, Los Alamos, NM, 87545

NOMENCLATURE

\begin{align*}
f_1 & \quad \text{solution based on fine mesh} \\
f_2 & \quad \text{solution based on medium mesh} \\
f_3 & \quad \text{solution based on coarse mesh} \\
g & \quad \text{error coefficient} \\
h & \quad \text{discretization parameter (element length)} \\
F_s & \quad \text{factor of safety} \\
r & \quad \text{grid refinement ratio} \\
p & \quad \text{order of convergence} \\
n & \quad \text{number of input parameters} \\
m & \quad \text{number of probability levels} \\
\eta & \quad \text{response surface output variable} \\
\beta & \quad \text{response surface fit coefficient} \\
x & \quad \text{response surface input variable}
\end{align*}

ABSTRACT

Uncertainty quantification in numerical simulations generally relies on applying sampling techniques to select input parameters for a suite of deterministic calculations and then combining the results of the calculations into output distributions. Several approaches, including metamodeling and Latin Hypercube Sampling, are prevalent today. This paper will focus on the comparison of several common sampling techniques as applied to the validation of a transient dynamics finite element calculation on a spherical marine float drop test.

1 INTRODUCTION

Probabilistic analysis and design is a powerful tool for uncertainty quantification that can be used in a number of applications. Existing designs can be analyzed for probability of failure for a given load case, and new designs can be optimized for strength, life, weight, or another parameter based on the results of probabilistic analysis. The procedure for performing such an analysis requires the use of sampling or other techniques to determine the effects of uncertainty in one or more parameters of interest. A comparison of these different techniques in order to determine the best technique has previously not been performed at Los Alamos National Laboratory on engineering problems.

This study compares a number of methods used in uncertainty quantification. Two sampling methods as well as three analytical methods are investigated and tested for accuracy, numerical efficiency, and other properties. The
methods are evaluated using the results of finite element analyses of a spherical steel marine float undergoing a drop test. A number of parameters are selected as input variables into the uncertainty methods and finite element model. The results from the methods are then compared to experimental data to determine which method most accurately captures the actual uncertainty.

2.0 FINITE ELEMENT MODEL

The finite element model developed for this study consists of a spherical shell connected to a hollow cylinder protruding through the shell (Figure 1). The float geometry is meshed with Belytschko-Lin-Tsay shell elements [1] using I-DEAS finite element software. The elements are computationally efficient shell elements with five thickness integration points and controlled with a Flanagan-Belytschko exact volume stiffness form hourglass stabilization method. Material properties for stainless steel are used in the kinematic/isotropic elastic-plastic element formulation. A solid, rigid element block is used as the impact surface and is held using nodal constraints. The contact constraint between the float and the block is a node-surface sliding surface type with iterative Lagrange constraint enforcement. A one-quarter model with symmetry boundary conditions was used to reduce computation time. Dyna3D, an explicit structural dynamics finite element code, is used as the solver for the numerical experiments. The numerical experiments involve the impact of the float with the rigid block, and measuring the strain response at a position on the sphere located 45° (between the pole and the equator) from the impact location. Post-processing is performed on a group of elements selected to represent the strain gage attached to the float using GRIZ (for visualization), THUG (for time-history data extraction), and MATLAB (for data processing).

![Figure 1 – Finite element model of the marine float drop test.](image)

2.1 Discretization and the GCI

The process of creating a finite element model usually involves a trade-off between computational efficiency and model accuracy. When a model needs to accurately replicate the behavior of a real object, it requires a fine enough mesh to closely approximate the real object. However, a fine mesh and more accurate solution require an exponential increase in computational time. Therefore, it is necessary to determine the numerical accuracy of a given mesh, such that an informed compromise between discretization uncertainty and computational efficiency can be made. The grid convergence index (GCI) is a useful tool for providing an estimate of whether or not a given mesh is dense enough to provide a reliable result. The GCI, based on Richardson Extrapolation, provides a band of reasonable assurance that the solution lies somewhere in or nearby the band.

Richardson Extrapolation and the GCI are derived from series representations of discrete solutions, which assume that the discrete solution \( f \) can be represented by an exact solution \( f_{\text{exact}} \) and a series of error terms depending on the discretization parameter \( h \) and influence coefficients \( g \), as described in Equation 1. Two discrete solutions \( f_1 \) and \( f_2 \), obtained from meshes defined by element lengths \( h_1 \) and \( h_2 \), are used to extrapolate an approximated exact solution \( f_{\text{exact}} \). The refinement ratio, \( r \), is defined by the ratio of \( h_1 \) and \( h_2 \). This equation (2) can be generalized for a method of order \( p \), which must be known a priori, approximated, or calculated.
\[ f = f_{\text{exact}} + g_1 h + g_2 h^2 + g_3 h^3 + \cdots \]  

(1)

\[ f_{\text{exact}} \approx f_1 + \frac{f_1 - f_2}{r^p - 1} \]  

(2)

By extending this process to the discrete solutions presented by various mesh densities in finite element modeling problems, this method provides some insight into how well the meshes are converging to a stable solution as grid refinement progresses. When three or more meshes are solved in a simulation, the actual order of convergence of the series of solutions can be calculated directly, assuming that all of the solutions are in the region of convergence. By requiring a constant grid refinement, \( r \), the following equation will solve the order of convergence where \( f_1, f_2, \) and \( f_3 \) are the fine, medium, and coarse meshes, respectively.

\[ p = \ln \left( \frac{f_3 - f_2}{f_2 - f_1} \right) / \ln(r) \]  

(3)

The coarse grid GCI is then defined by the discrete solutions, grid refinement ratio, order of convergence, and a factor of safety \( F_s \) equal to a value of 3 as determined by Roache [2]. This value is an estimate on the numerical uncertainty of the coarse grid.

\[ GCI = F_s \left( \frac{\text{abs}(f_{\text{fine}} - f_{\text{coarse}})}{f_{\text{fine}}} \right) \cdot \frac{r^p}{r^p - 1} \]  

(4)

2.2 Model Convergence

Results of the GCI calculation on the marine float drop test can be seen in Table 1. Convergence can be observed by the decreasing value of the GCI. Another model verification issue is the use of a one-quarter model with symmetry boundary conditions and its comparison with a full model. The difference in peak absolute strain was found to be less than one percent between the full and quarter model, indicating that the one-quarter mesh is an adequate representation of the problem.

<table>
<thead>
<tr>
<th>Mesh Density</th>
<th>Maximum Strain</th>
<th>GCI</th>
</tr>
</thead>
<tbody>
<tr>
<td>Low</td>
<td>223.5</td>
<td>0.0709</td>
</tr>
<tr>
<td>medium</td>
<td>228.7</td>
<td>0.0027</td>
</tr>
<tr>
<td>high</td>
<td>228.9</td>
<td>(N/A)</td>
</tr>
</tbody>
</table>

Table 1: GCI results

Based on the results of this convergence study, the medium mesh density model was judged sufficiently refined for the purposes of the uncertainty study. This mesh is therefore used for the remainder of the study.

3.0 SAMPLING TECHNIQUES

Five commonly used sampling techniques are compared in this study, including: Latin Hypercube Sampling (LHS), Response Surface Modeling using a Central Composite Design (CCD), Mean Value method (MV), Advanced Mean Value method (AMV), and Advanced Mean Value method with p-level iterations (AMV+). These sampling techniques are used as an alternative to the computationally expensive Monte Carlo method for uncertainty quantification. Whereas the Monte Carlo method may require millions of sample points to provide a Cumulative Distribution Function (CDF), the alternative sampling methods can provide, at the expense of some accuracy, a CDF in fewer than 100 sample points. It is helpful, therefore, to determine which method provides the best representation of the CDF at the least computational expense.
3.1 Model Inputs and Variations

The variational parameters used to perform probabilistic analyses of the marine float included the spherical outer diameter, tube inner diameter, tube length, sphere wall thickness, tube wall thickness, and impact velocity. The input parameter distributions were found by performing a statistical measurement survey of the floats purchased from McMaster-Carr. The floats were made of 20 gauge, type 304 stainless steel, with a 5” nominal diameter, and a 0.5” nominal tube inner diameter. All floats were used to measure the major dimensions and two spheres were cut apart so that thicknesses could be measured. The distribution for the impact velocity was set to a mean of 5.5 m/sec corresponding to a free fall height of 1.543 meters or about 5 feet. The standard deviation was based on free fall calculations using a reasonable error of 2 cm in measuring the drop height. Goodness of fit tests were used to determine which statistical distribution best represented the measurements. The distribution data can be seen in Table 2.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Mean</th>
<th>Std. Dev.</th>
<th>Distribution</th>
</tr>
</thead>
<tbody>
<tr>
<td>Spherical outer diameter (mm)</td>
<td>127.108</td>
<td>0.337</td>
<td>Weibull</td>
</tr>
<tr>
<td>Tube inner diameter (mm)</td>
<td>13.163</td>
<td>0.0794</td>
<td>Weibull</td>
</tr>
<tr>
<td>Tube length (mm)</td>
<td>133.120</td>
<td>0.296</td>
<td>Normal</td>
</tr>
<tr>
<td>Sphere wall thickness (mm)</td>
<td>0.813</td>
<td>0.0393</td>
<td>Weibull</td>
</tr>
<tr>
<td>Tube wall thickness (mm)</td>
<td>1.320</td>
<td>0.0465</td>
<td>Weibull</td>
</tr>
<tr>
<td>Impact Velocity (mm/sec)</td>
<td>5500</td>
<td>35.6</td>
<td>Normal</td>
</tr>
</tbody>
</table>

Several potentially important input variables, including material properties (elastic modulus and yield strength), were also considered for the study. However, time constraints precluded characterization of additional input variables.

3.2 Latin Hypercubes Sampling

The basic idea of Latin Hypercube Sampling (LHS) is to drastically reduce the number of sampling points in the parameter domain so that finite element calculations of the response variable can be performed efficiently. LHS is considered a stratified sampling method that partitions the sampling domain and samples in such a manner that redundant sampling points are excluded, which is an inherent problem with the Monte Carlo sampling technique. One parameter that needs to be specified by the user when using LHS is how many finite element solutions, or sample points, a model needs. Unfortunately, there is no explicit rule for knowing how many runs to use. (Some studies have been performed which look at the influence of sample size and variation in multiple LHS runs [3].) Once the number of sampling point evaluations, n, has been determined, each input parameter’s range is divided into n regions of equal probability. One parameter value is then randomly selected from within each interval for each parameter. This produces a set of n^(number of parameters) possible sampling points. From this set, n final sampling points are randomly selected without repetition for evaluation.

In this study of the marine floats, 77 sample points were used to determine the LHS produced cumulative distribution function of the maximum strain observed by the strain gage. This number of points was selected because it provides for direct comparison to a full factorial central composite design sampling set that uses 77 sampling points. The LHS algorithm was executed within the uncertainty analysis software package NESSUS, developed at the Southwest Research Institute.

3.3 Response Surface Modeling using a Central Composite Design

Response surface methods, also known as metamodeling, under which the Central Composite Design is categorized, are typically a powerful and efficient way of estimating responses based on various inputs. This, however, is only true if the surface or hypersurface is relatively smooth. Even if the phenomena that is being characterized is known to be a fairly smooth and continuous function, noise from measurements or processing can drastically change the predictions generated by the response surface.
The CCD consists of sampling points located at predetermined locations at the center, axes, and factorial points in the distributed parameter space [4]. The factorial points are the only points that contribute to the estimation of the interaction terms. Axial points contribute primarily to the estimation of quadratic terms. Without the axial points, only the sum of the quadratic terms could be estimated. Axial points do not contribute to the estimation of interaction terms. The center points provide an internal estimate of error (pure error) and contribute toward the estimation of quadratic terms.

The finite element solutions of the set of sampling points are used to create a hypersurface using the method of least squares. The full factorial CCD allows for a second-order hypersurface, although a lower order surface can be used in its place. The equations for these surfaces are as follows:

First-order model:

$$\eta = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_k x_k$$  \hspace{1cm} (5)

Second-order model:

$$\eta = \beta_0 + \sum_{j=1}^{k} \beta_j x_j + \sum_{j=1}^{k} \beta_{j,j} x_j^2 + \sum_{i<j} \beta_{i,j} x_i x_j$$  \hspace{1cm} (6)

Once the hypersurface is created, the Monte Carlo technique is used to sample the hypersurface. This can be done because solutions are taken from an easily evaluated polynomial function rather than a full model. The results of the Monte Carlo simulation are then used to create the cumulative distribution function of the output.

The full factorial Central Composite Design for a six-parameter model results in a set of 77 sampling points. This number also dictated how many points to use in the LHS algorithm. The specific points are determined by the layout of the CCD and the distributions of each input parameter. One shortcoming of the CCD method is that, as a result of its quadratic nature, probability levels in the tails of the solution distribution function are poorly estimated. The area at the center portion of the distribution function, however, is usually well represented.

3.4 Mean Value, Advanced Mean Value, and Advanced Mean Value with p-level iterations

Included in uncertainty quantification and probabilistic analysis methods is a family of analytical methods that include the Mean Value method, Advanced Mean Value method, and Advanced Mean Value method with p-level iterations [5]. Analytical methods differ from sampling methods such as Monte Carlo, LHS, and CCD in that rather than calculating probability levels, the solution parameter is instead found for a given probability level. With the ability to select certain probability levels it is possible to take a closer and more accurate look at the tails of a solution distribution. The number of probability levels is not limited, allowing for a smaller number of levels in non-important regions while a larger number can be used in more important regions of the distribution function. Probability levels can also be added at any time so if more definition is needed in a region then the whole analysis does not need to be started from scratch.

The Mean Value method assumes a linearized response of the solution parameter and uses the Most-Probable-Point (MPP) values at each probability level on a normalized joint probability distribution function of the input parameters to calculate the solution parameter. This calculation takes only 2*n+1 function evaluations (finite element solutions) to complete for a central differencing scheme, where n is the number of input parameters. The Advanced Mean Value method is an expansion of the Mean Value method in which the solution parameter is updated at each probability level using the Most-Probable-Point-Locus (MPPL). The AMV method significantly improves the accuracy of the cumulative distribution function over the MV method with only a small computational cost. The method takes a total of 2*n+1+m function evaluations to finish, where m is the number of probability levels. Lastly, the AMV plus p-level iteration method further improves the CDF by updating the sensitivity values used for calculating the MPP and MPPL for each probability level. The number of extra function evaluations needed to carry out the iterations depends on how long it takes each probability level to converge, which it is not guaranteed to do. The extra iterations may significantly increase the needed number of finite element runs, but still usually provides a solution for fewer runs than any of the sampling methods, with the caveat that the number of input parameters is relatively small (six or fewer).
4 EXPERIMENTAL SETUP

For use in model validation efforts and for use as a guide in determination of the best sampling technique, the marine float drop test was performed experimentally as well as numerically. Eight marine floats (Figure 2) were instrumented with hemi-spherically opposed strain gages. A block of hardened steel was used as the impact surface. The float was dropped using a 0.5 inch steel rod as a guide so that rotations off the tube axis would be minimized at impact. The rod guided the float vertically until six inches above the impact surface after which it would be in pure free fall. The drop height was made as close as possible to the prescribed 5 feet used to create the velocity uncertainty distribution for the model. Because of the small sample size and the need for curve definition, multiple drops were performed on each float. Although questions arise as to the effect of subsequent drops on the data, the experiment is not the focus of the paper and is only used as a comparison to the sensitivity analysis methods. Data was acquired and stored using the NOMADD data acquisition system developed at Los Alamos National Laboratory.

5.0 RESULTS

The maximum absolute strain recorded from the strain gauges on both the model and the experiment was used as the uncertainty response variable. The time-series solutions of the model were filtered using a Butterworth filter with a filtering frequency of 50kHz, which is the hardware filtering frequency for the NOMADD data acquisition system used to obtain the experimental data. From the filtered data, the maximum absolute value was recorded. Comparison of the numerical time-series data with the time-series data from an experimental test (Figure 3) shows a large discrepancy in not only shape but also the time at which the peak strain occurred. This, as well as other differences in the plot, indicates a discrepancy in wave speed between the model and experiment. The inconsistency in wave speeds may be explained by an incorrect value and uncertainty in elastic modulus, a variable that was not included in the uncertainty analysis. A look at the time-series uncertainty caused by a change in the elastic modulus can be seen in Figure 4. Another source of error could be that the density of the stainless steel was incorrectly modeled. Density, however, has very little variance from sample to sample in steels, so is not likely the root of error. Unfortunately, time constraints and the large number of finite element runs required to include this extra degree of uncertainty did not allow the model to be corrected, nor was the statistical sampling of the elastic modulus of the floats an available option.

Figure 2: A marine float used for an experimental drop test.
5.1 Comparison of Cumulative Distribution Functions

The maximum absolute strain data from the LHS, CCD, and experimental data sets were used to create empirical cumulative distribution functions (CDFs) for each. The CDFs for the mean value methods were automatically generated in NESSUS with the response variables corresponding to probabilities. Plots of the numerical CDFs alone and the numerical CDFs compared with the experimental CDF are shown in Figures 5 and 6, respectively.

Figure 5 reveals many differences between the sampling methods and their capabilities. Relative to the other methods, the mean value method overestimates the distribution. The AMV and AMV+ methods as expected can then be seen to adjust the distribution to more closely match the LHS and CCD distributions. Also seen in the mean value methods is their lack of definition in their CDFs, which could be a setback in highly nonlinear problems. An example of a problem with nonlinearity occurred when using AMV+ in which one point of the CDF at a probability of $2.53 \times 10^{-4}$ did not converge in the iteration step.
All the distributions compare well with each other at high cumulative probabilities (P>0.2) but are significantly dissimilar in lower cumulative probabilities. While the CCD curve is relatively smooth, the LHS curve shows a striking nonlinearity at a probability of about 0.12. This illustrates one of the major failings of response surface methods such as CCD in which nonlinearities in the response surface are extremely difficult if not impossible to resolve. The AMV and AMV+ methods were to some degree able to pick up this nonlinearity.

Although the numerical methods are all relatively similar, they differ drastically from the experimental distribution as can be seen in Figure 6. The mean of the distributions differs by about 50 microstrain, and a factor of two in the standard deviations is also easily noticeable. Including the uncertainty in elastic modulus cannot alone account for all of the error. Because the finite element model uses elastic-plastic elements, uncertainty can also be found in the value of the yield stress. Tabular values of the yield stress were used as input values for the model, but the value of yield stress in metals is well known to be highly variable and depends on a number of factors including heat treatment and manufacturing. If the yield stress in the model is too low then plastic deformation will prevent the correct amount of strain energy from reaching the strain gage, thereby causing a lower maximum absolute strain. The floats, which were most likely punch formed, may have undergone strain strengthening due to the forming process, which would result in a larger yield stress. This is the most probable cause of the vast differences between the numerical and experimental CDFs. Unfortunately, the large differences provide for negative results in a model validation study. The results of the sampling techniques are still significant, however, because they show differences between the methods.

5.2 Computational Efficiency of Sampling Methods

Computational efficiency is one of the largest issues concerning uncertainty quantification and can vary significantly between different techniques. The efficiency of a method is measured by the number of finite element evaluations that must be performed to complete the analysis. The number of finite element evaluations is used as the efficiency metric because one solution takes orders of magnitude longer to complete than any other calculation in any given technique. Table 3 provides the number of finite element runs each method took to generate its respective CDF. Although the mean value methods are significantly more efficient (with the exception of AMV+) than the sampling methods (LHS and CCD), what they gain in efficiency they lose in definition. The converse is true of the sampling methods.

<table>
<thead>
<tr>
<th>Method</th>
<th># of FE runs</th>
<th># of CDF points</th>
</tr>
</thead>
<tbody>
<tr>
<td>LHS</td>
<td>77</td>
<td>77</td>
</tr>
<tr>
<td>CCD</td>
<td>77</td>
<td>77+</td>
</tr>
<tr>
<td>MV</td>
<td>13</td>
<td>10</td>
</tr>
<tr>
<td>AMV</td>
<td>23</td>
<td>10</td>
</tr>
<tr>
<td>AMV+</td>
<td>65</td>
<td>10</td>
</tr>
</tbody>
</table>

6 CONCLUSIONS

Each uncertainty quantification method discussed in this paper has its strengths and weaknesses. The best method to use in a given situation depends on the type of problem and what qualities are desired in the cumulative distribution function. The mean value methods (mean value (MV), advanced mean value (AMV), and advanced mean value with p-level iteration (AMV+)) are good for an efficient general CDF curve where definition is not important. The AMV method provides a substantial increase in accuracy compared to the MV for a small increase in the number of finite element evaluations. For this reason, AMV should always be used over MV. The convergence issues with AMV+ make its use questionable, as well as the possible significant increase in number of FE runs. These methods in general provide good accuracy at low computational cost and are designed to be especially effective in extremely low probability regions.

The Latin Hypercube Sampling (LHS) and Central Composite Design (CCD) methods are more suitable when a more detailed CDF is desired. The differences between them, however, are significant and the results of this paper lean toward the exclusive use of LHS over CCD. Because CCD is basically a multidimensional quadratic
curve-fitting technique, it is restricted in its use to continuous response functions of second-order behavior for a good representation of uncertainty. The quadratic restriction also causes problems in approximating low probability levels. A higher order response behavior or presence of a discontinuity will provide less accurate results. Also of concern is the number of uncertainty parameters. For a full factorial response surface method such as CCD, it takes $2^n+2n+1$ finite element evaluations where $n$ is the number of uncertainty parameters. A large number of parameters requires reduction of the method using partial factorial designs which in turn increases the error in the method.

Rather than use a response surface to produce a detailed CDF, this study suggests the use of LHS. Nonlinear response functions are not an issue, and the number of uncertainty parameters is not constrained. LHS can be used for any number of parameters and with any number of sampling points. The number of sampling points does, however, influence the accuracy of the CDF. The larger the number of points, the closer the method gets to a Monte Carlo simulation. A study by Helton and Davis [3] has shown that increasing the number of sampling points can reduce uncertainty in the CDF. It is also argued that multiple LHS simulations can band the true CDF solution making multiple LHS simulations with fewer sampling points each a desirable sampling method. Overall, LHS does not have the limitations of CCD and is a better approximation of the Monte Carlo method, making it more suitable for uncertainty analysis.

ACKNOWLEDGEMENTS

First of all we would like to thank all the Los Alamos National Laboratory staff members that have enabled us to complete this project. We would like to thank Renida Carter for her help with the experimental side of the project, including the data acquisition with the NOMADD software and the coordination of other lab supporting staff members. These additional staff members include Laurel Conner, Tom Harlow, Larry Berkbigler, Don Ortiz, and team leader Jim Lake, who enabled these people to help out so much. We would also like to thank Amanda Rutherford and Jason Pepin for their time and guidance throughout the length of this project. Finally, we would like to express our gratitude to Dr. Charles “Chuck” Farrar, for his continued support of summer student programs at Los Alamos.

REFERENCES


