

Probabilistic Methods in Model Validation

Thomas L. Paez

Validation and Uncertainty Quantification Department

Laura Swiler

Optimization and Uncertainty Quantification Department

Sandia National Laboratories

PO Box 5800

Albuquerque, New Mexico 87185

ABSTRACT

Extensive experimentation over the past decade has shown that fabricated physical systems that are intended to be identical, and are nominally identical, in fact, differ from one another, and sometimes substantially. This fact makes it difficult to validate a mathematical model for any system and results in the requirement to characterize physical system behavior using the tools of uncertainty quantification. Further, because of the existence of system, component, and material uncertainty, the mathematical models of these elements sometimes seek to reflect the uncertainty. This presentation introduces some of the methods of probability and statistics, and shows how they can be applied in engineering modeling and data analysis. The ideas of randomness and some basic means for measuring and modeling it are presented. The ideas of random experiment, random variable, mean, variance and standard deviation, and probability distribution are introduced. The ideas are introduced in the framework of a practical, yet simple, example; measured data are included. This presentation is the third in a sequence of tutorial discussions on mathematical model validation. The example introduced here is also used in later presentations.

Nomenclature

E	Modulus of elasticity
G	Shear modulus of elasticity
L, U	Interval limits
X	Random variable
s	Sample standard deviation
w	Measure of structural dynamic response
x, y	Arbitrary data
s^2	Sample variance
\bar{x}	Sample mean

Introduction

The use of probability theory to characterize structural dynamics was effectively introduced in the United States in 1958 (Crandall, 1958, Paez, 2006). The application was random vibrations. (However, the ideas used in random vibrations were put into use in the field of communications as early as 1925. See Carson, 1925.) The ideas of random vibration were generalized into the field of random shock and vibration, and the field is known as probabilistic structural dynamics. The theory of probabilistic structural dynamics seeks to characterize structural dynamic response probabilistically, by considering excitations applied to structures as random, but considering the structures themselves as deterministic.

Over the past decade, or so, the field of probabilistic structural dynamics has been extended to consider the potential for randomness in structures as well as their excitations. When used in connection with the finite element method, the theory is known as stochastic finite elements (Ghanem and Spanos, 1991). The technique has been applied to structural dynamic systems (Ghanem, Doostan, Red-Horse, 2008), and other systems. But the recognition that structural systems vary randomly is not recent. In 1969 Collins (1969) considered the eigenvalue problem for structures with randomly varying parameters. Experimentation over several decades (and, particularly

over the past decade) has shown that physical systems that are intended to be identical, and are, in fact, nominally identical, differ from one another, sometimes substantially (Smallwood, Gregory, Coleman, 2000).

The study of the validation of mathematical models was introduced into engineering about ten years ago. It is defined as the “process of determining the degree to which a computer model is an accurate representation of the real world from the perspective of the intended model applications.” (ASME, 2006, U.S. DOE, 2000, AIAA, 1998) The objective of validation is to compare the responses predicted with a mathematical model to the responses realized by a physical system during an experiment, and, perhaps, judge the adequacy of the model. The strictness of the criterion used to judge model adequacy depends on the intended application of the model. The goal of model validation is to assure that the mathematical models of physical systems and phenomena are adequate for use in a pre-established application. In structural dynamics, validation has focused on attempts to assure the adequacy of finite element models. (Though it is equally important, practically no emphasis has been placed on validation of mathematical models of random structural dynamic excitations. See Field, Paez, Smallwood, 2008).

If both mechanical structures and their excitations could be accurately treated as deterministic, then their behaviors would reflect no uncertainty, and model validation would involve deterministic comparisons of model-predicted behavior to physical system behavior measured during experiments. However, as mentioned above, experiments have shown that structures – even those fabricated to the highest standards of quality control – reflect uncertain behavior. In view of this, the uncertainty in structural system characteristics must be considered in validation analyses.

This paper presents some of the most fundamental ideas from the theory of probability, including the modeling of individual random sources, the modeling of pairs of correlated random sources, the propagation of randomness through a deterministic model, and the comparison of random model outputs to random experimental outputs.

Fundamental Tools of Probabilistic Uncertainty Quantification, Example

The literature on probability and statistics – under development for more than three centuries – is filled with tools and techniques that are useful for pursuing analyses critical to model validation. Some of these tools permit (1) characterization of the randomness in experimental data, (2) creation of models to simulate random phenomena, and (3) comparison of model predictions to experimental data to judge whether the model used to make the predictions should be rejected as representative of the experiments. Some of the tools and techniques are relatively straightforward and easy to understand, and others are more esoteric. We summarize, here, some of the former tools.

We consider, first, elements (1) and (2), listed above, and we do so in the context of a specific example. We start by describing a technique for modeling a single, or univariate, random source, and follow that by describing a technique for modeling two or more random sources, simultaneously. (A random source that has two quantities of interest is called bivariate.) We then show how the uncertainty associated with an element or system subcomponent can be propagated through another model – usually a finite element model. Finally, we consider element (3), listed above, namely, the comparison of probability predictions from a model to the probabilistic character of experimental response.

Modeling and Simulation of a Univariate Random Source

We are interested in characterizing the properties of a bonding material, i.e., a material used to adhesively bond two shells together. To accomplish this goal, we might construct multiple, nominally identical replicates of an experimental system that can be used to infer the desired material properties. (For the details of such an experiment carried out at Sandia National Laboratories, see Appendix A.)

In this example, one of the material properties of interest is shear modulus of elasticity, G . During experiments performed on five separate structures, realizations of the shear modulus of elasticity were inferred, and they are listed in Figure 1a.

Test Index, i	$G_i (psi)$
1	3151.9
2	3334.5
3	3297.9
4	3201.8
5	3141.2

Figure 1a. Bond material shear modulus of elasticity values inferred from experiments on five nominally identical test structures.

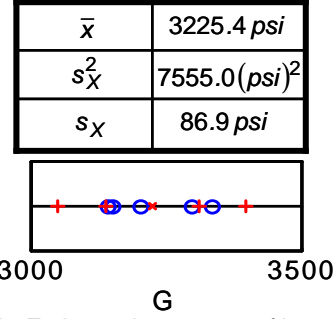


Figure 1b. Estimated moments of bond material shear modulus of elasticity. Number line with data, sample mean, and mean ± 1 and 2 sample standard deviations.

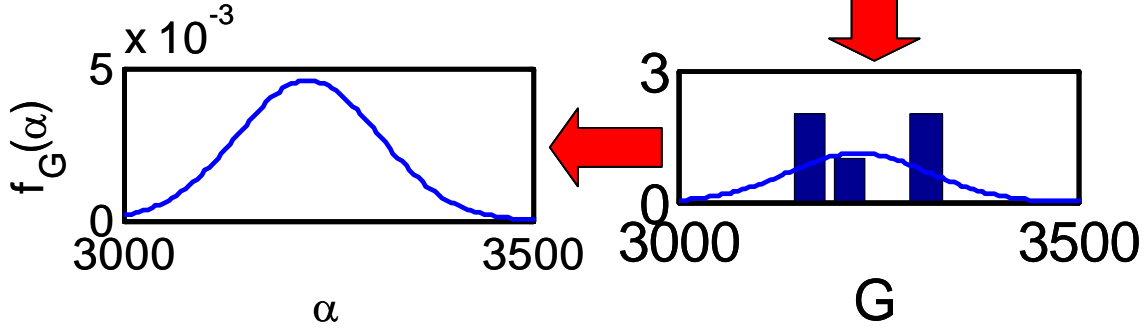


Figure 1d. Best fit normal probability density function of bond material shear modulus.

Figure 1c. Histogram of inferred bond material shear modulus data, and assumed shape of distribution.

Denote the sample values of the shear modulus of elasticity, $x_i, i = 1, \dots, n$, where, here, $n = 5$. The sample mean of the experimental data is denoted \bar{x} , and is defined as

$$\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i \quad (1)$$

This quantity is the average of the data and an estimate of the mean or expected value of the random source of the experimental data. The mean is a “central value” of all values that might be drawn at random. The source of random data is called a random variable. (It is assumed that the random variable, X , has a specific mean, μ_X , that can never be known except through measurement of infinite samples.) The sample mean, \bar{x} , is a statistic of the data. That is, it is a quantity dependent on the measurements, only. It has the same units as the data. The sample mean of the data in Figure 1a is provided in Figure 1b.

Another statistic that characterizes a collection of measured data and, it is hoped, the random source, X , of the data is the sample variance. It is denoted s_X^2 and is defined as

$$s_X^2 = \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2 \quad (2)$$

It is the average of the squares of the deviations of the data from the sample mean, and it is an estimator of the variance, σ_X^2 , of the random variable, X . The sample variance has units that equal the square of the units of the data. The sample variance of the data in Figure 1a is listed in Figure 1b.

An alternate measure of the spread in data that has units of the data, themselves, is the sample standard deviation. It is denoted s_X and defined

$$s_X = \sqrt{s_X^2} \quad (3)$$

That is, it is simply the square root of the sample variance. It is the estimator of the standard deviation, σ_X , of the random variable, X . The sample standard deviation of the data in Figure 1a is listed in Figure 1b.

The sample mean and variance are called moments because they satisfy the mathematical definition of a moment. The sample mean is the first normalized moment of the data about the origin; the sample variance is the second normalized moment about the sample mean. (Montgomery et al., 2006)

Figure 1b shows a number line with the data (o), the sample mean (x), and the sample mean minus/plus one and two sample standard deviations (+). Clearly, the sample mean lies at the centroid of the measured data, and the sample standard deviation is a measure of the spread of the data.

A statistical tool for characterizing the distribution of values in the measured data (and, again, it is hoped, the random source of the data) is the histogram. A histogram is a bar chart indicating the frequency of occurrence of data values within specific, quantitative intervals. The histogram of the data in Figure 1a is shown by the stair-step function in Figure 1c. Five data are rather minimal for learning much about data distribution and the distribution of their source. More data might indicate a particular probability distribution. However, when data are minimal, but it is known that the source is random, a convenient form for the distribution might be specified. The smooth curve superposed over the histogram in Figure 1c has the shape of the probability density function (PDF) of a random variable that is normally distributed. It points to one of the goals of modeling, and that is to generalize the information in the data into a probability model that can be used to characterize the distribution of the source of the data.

A normal PDF is shown in Figure 1d. It is the familiar “bell-shaped” curve. The PDF of random variable X is denoted $f_X(x)$, $-\infty < x < \infty$, and a normal PDF has the form

$$f_X(x) = \frac{1}{\sqrt{2\pi}\sigma_X} \exp\left[-\frac{1}{2\sigma_X^2}(x - \mu_X)^2\right] \quad -\infty < x < \infty \quad (4)$$

where $\mu_X = \bar{x}$ and $\sigma_X^2 = s_X^2$. It can be shown (Ang and Tang, 1975) that the mean of the random variable X is the centroid of the PDF curve, and its variance is the second moment of PDF curve about the mean. In order to be a valid PDF a function must satisfy the requirements

$$\begin{aligned} f_X(x) &\geq 0 \quad -\infty < x < \infty \\ \int_{-\infty}^{\infty} f_X(x) dx &= 1 \end{aligned} \quad (5)$$

The normal PDF of Eq. (4) clearly satisfies the first requirement, and it can be shown, directly, that it satisfies the second requirement.

Further, the PDF of a random variable is related to the probability that the sample of a random variable obtained during a random experiment falls within an interval. Specifically

$$P(a < X \leq b) = \int_a^b f_X(x) dx \quad -\infty < a < b < \infty \quad (6)$$

The quantity $P(a < X \leq b)$, $-\infty < a < b < \infty$, is the probability that when a single random experiment is performed (i.e., the random source associated with X is sampled) the result will lie in the interval $(a, b]$. Probability is the relative chance of occurrence of an event.

Once the probability distribution of a random variable is known samples from the random source can be generated. There is a general means for obtaining samples from an arbitrarily distributed source, but for the present, we simply point out that when the source is normally distributed the means for obtaining one or more random samples is very direct. Many software packages like MATLAB, EXCEL, and others can be used to generate random samples from a normally distributed source with zero-mean and unit-variance. (These are samples from the so-called standard normal random variable.) Let r denote such a sample. Then a sample from a normal random source with mean μ_X and variance σ_X^2 is given by

$$X = \mu_X + \sigma_X r \quad (7)$$

Any number of random samples can be generated.

Consider the five values of inferred bond material shear modulus, listed in Figure 1a. If we assume they come from a normal random source with mean and variance equal to the sample mean and variance in Figure 1b, then we can generate more samples as described above. Fifty samples were generated, and the histogram of the generated data is shown in Figure 2. The original data values are shown by dots along the abscissa.

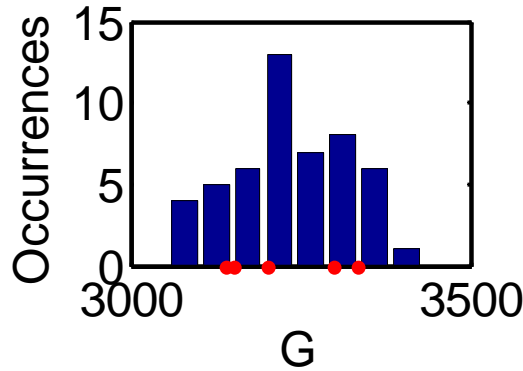


Figure 2. Histogram of generated data, and the original data (dots).

The approach to modeling and simulation of a random source described in Figures 1a through 1d and Eqs. (1) through (7) can be very simply modified to develop an alternate (and, perhaps, more suitable) model. When it is required that a collection of data have a distribution of values where all members are non-negative or positive, it may be preferable to analyze and model the logarithm of the data. For example, here, instead of computing the sample mean and variance of the data $x_i, i = 1, \dots, n$, we might compute those moments for $\log(x_i), i = 1, \dots, n$, instead. We might use those sample moments to create a model for the random variable $\log(X)$, instead of X . We might generate samples from the random source $\log(X)$, and once done, we might invert the generated samples to obtain $x = 10^{\log(x)}$. The effort associated with this approach is essentially the same as expended on the approach previously described.

Modeling and Simulation of a Bivariate Random Source

The tools developed to this point show how to model univariate data and generate samples from the model. But it is sometimes also necessary to model and generate multivariate data.

Consider, again, the data provided in Figure 1a – the inferred shear modulus of elasticity of the bond material – and, in addition, the inferred modulus of elasticity. Both sets of experimental data are listed in Figure 3a. The figure on the top of Figure 3b plots the shear modulus data versus the modulus of elasticity data. This is known as a scatter-plot. The line included with the data shows that the data fall almost on a straight line, and that fact simplifies representation of the data. Before considering the simplification, though, consider the general case.

Test Index, i	G_i (psi)	E_i (psi)
1	3151.9	9235
2	3334.5	9770
3	3297.9	9630
4	3201.8	9362
5	3141.2	9235

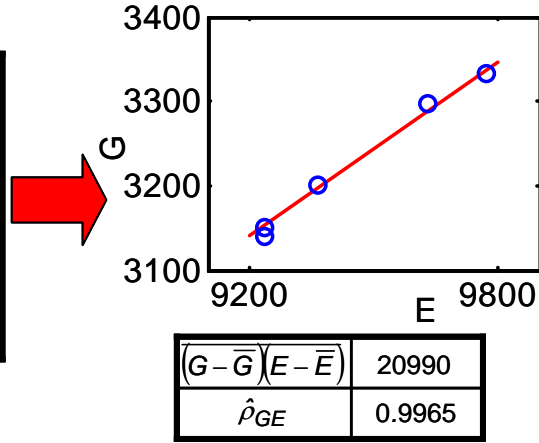


Figure 3a. Bond material shear modulus of elasticity and modulus of elasticity values inferred from experiments on five nominally identical test structures.

Figure 3b. Scatter-plot of data, and sample covariance and sample correlation coefficient.

There are two fundamental measures of the degree of linear relation between a pair of random sources where realizations (samples) of the random sources are simultaneously, or jointly, realized. The first is the sample covariance. Let two sets of data, simultaneously realized, be denoted $(x_i, y_i), i = 1, \dots, n$. Their sample covariance is denoted $\overline{(X - \bar{x})(Y - \bar{y})}$, and is defined as

$$\overline{(X - \bar{x})(Y - \bar{y})} = \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y}) \quad (8)$$

The sample covariance is the degree of linear relation between a pair of data sets; it is the average of the product of the mean-normalized data. The magnitude of the sample covariance is not normalized, but it can be shown that the magnitude can never surpass the product of the sample standard deviation of the x-data times the sample standard deviation of the y-data. The normalized form of the sample covariance is the sample correlation coefficient. It is denoted $\hat{\rho}_{XY}$, and it is defined

$$\hat{\rho}_{XY} = \frac{\overline{(X - \bar{x})(Y - \bar{y})}}{s_X s_Y} \quad (9)$$

The sample correlation coefficient is in the interval $[-1, 1]$, and its magnitude indicates the normalized degree of linear correlation between a pair of data sets. The sample covariance and sample correlation coefficient between the shear modulus and modulus of elasticity data are listed in Figure 3b with the scatter-plot. The latter quantity makes it clear that the correlation is very high. When this is so, the two variables are said to be nearly linearly correlated or almost perfectly correlated, and one variable is approximately a linear function of the other.

We now establish the parameters of that linear function using an approach called linear regression (Ang and Tang, 1975). Consider, again, the two sets of data denoted $(x_i, y_i), i = 1, \dots, n$. We wish to identify a linear model with the form

$$y_i = \hat{\alpha}x_i + \varepsilon_i \quad i = 1, \dots, n \quad (10)$$

where $\hat{\alpha}$ is a constant, and $\varepsilon_i, i = 1, \dots, n$, is a sequence of realizations from a mean-zero random variable with sample variance, s^2 . The constant $\hat{\alpha}$ is identified by (1) solving Eq. (10) for ε_i , (2) squaring both sides of the resulting expression, (3) summing both sides of the equation, and minimizing the sum with respect to $\hat{\alpha}$ by taking the partial derivative on both sides with respect to $\hat{\alpha}$, equating the resulting expression to zero, and solving the resulting equation. The result is

$$\hat{\alpha} = \frac{\sum_{i=1}^n x_i y_i}{\sum_{i=1}^n x_i^2} \quad (11)$$

The straight line in Figure 3b is the one with slope $\hat{\alpha} = 0.3415$, obtained by using the modulus of elasticity data in place of the $x_i, i = 1, \dots, n$, and the shear modulus data in place of the $y_i, i = 1, \dots, n$.

The sample variance s^2 can be obtained by evaluating the model predictions

$$y_i^{(mod)} = \hat{\alpha}x_i \quad i = 1, \dots, n \quad (12)$$

then using the results to evaluate

$$s^2 = \frac{1}{n-1} \sum_{i=1}^n (y_i - y_i^{(mod)})^2 \quad (13)$$

In general, to simulate joint realizations of the x - and y -data modeled, we assume or develop a distribution for the x -data as in the previous section. We generate a realization from the random source of x -data; that operation yields the first generated variable. Next, we generate a sample from a normal random variable with zero-mean and variance s^2 (from Eq. (13)). We use that generated datum along with the generated x -value in Eq. (10), to obtain the generated y -value. Finally, repeat this operation until the desired amount of data is obtained.

In general, the relation of Eq. (10) may contain a constant, and that constant can be evaluated using the same general approach and formulas described here. (See Ang and Tang, 1975).

The situation is somewhat simplified in the case where the sample correlation coefficient is near plus or minus one, as it is in Figure 3b. In that case we go through all the operations listed above, but will find that the sample variance, s^2 , of Eq. (13) will be approximately zero. In that case, to generate x - and y -data, we start by generating an x -datum, then use it in Eq. (12) to generate the y -datum. Repeat the operations until the desired number of data pairs is obtained.

We followed the steps listed above to generate samples of the modulus of elasticity and the shear modulus from the random source that was modeled. The results are shown in Figure 4. The red points represent the generated data, and the blue circles represent the original data listed in Figure 3a. The regression line shown in Figure 3b is repeated, here. The fit of the model to the data is not perfect, and more sophisticated methods could be used to make the model match the experimental data a little more closely, but that hardly seems worthwhile in this application.

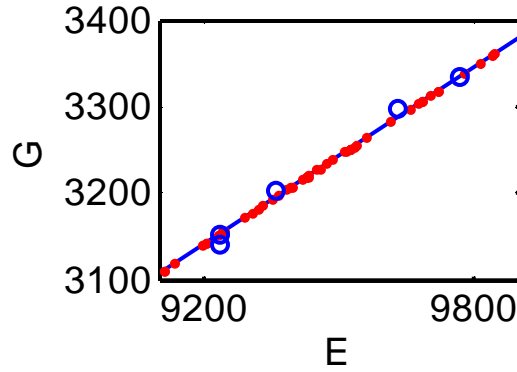


Figure 4. Experimental realizations of shear modulus versus modulus of elasticity (blue circles), versus modulus of elasticity (red dots) over regression line (blue line).

Propagation of Randomness through a Model

The previous two sub-sections show how to compute the statistics of data and how to model univariate and bivariate data. In the example started earlier, the objective is to characterize the probability model of a bonding material. The reason for creating the model of the bonding material is to establish the effect of its randomness on the behavior of a structural dynamic system. We will not go into details about how to obtain the model of the structural dynamic system, except to say that it would be obtained in a finite element framework. Specifically, a deterministic finite element model (FEM) would be constructed – one that can be used to compute one or more critical measures of response of the system under consideration. A deterministic finite element model is one whose input data – geometry, material parameters, boundary conditions, initial conditions, etc. – are all non-random quantities. The critical measure of response may be anything computable with an FEM, for example, peak motion at one or more locations, peak stress or strain at one or more points, etc. We can express the FEM computation in functional form.

$$w = g_{FE}(p, E, G) \quad (14)$$

where w is the response measure of interest, $g_{FE}(\cdot)$ embodies the FEM operations, p is the input data for the model (except for E and G), and E and G are the bond material modulus of elasticity and shear modulus. The latter two quantities are random variables, but when they take on specific values – like their generated values – they can be used in the FEM to compute specific results. We take the response measure of interest, w , to be a scalar, here.

The Monte Carlo method (see Ang and Tang, Volume II, 1984) uses a deterministic model, like the one in Eq. (14), to develop probabilistic results. The approach is straightforward. The analyst uses generated values of E and G , like those shown in Figure 4, one pair at a time, to compute the response measure of interest. Denote the generated pairs of E and G as $(E_i, G_i), i = 1, \dots, n_{gen}$. When the individual pairs of input parameters are used in Eq. (14) to compute the response, the model yields the set of outputs $w_i, i = 1, \dots, n_{gen}$. The outputs can be analyzed and modeled as described in the previous section, entitled “Modeling and Simulation of a Univariate Random Source,” to obtain a probability model.

For example, suppose that the response measure of interest, w , denotes peak displacement response at a point on a structure, in a particular direction, and is modeled in an FEM. Also suppose that the FEM is exercised over a bivariate range of values of the values (E, G) . Let the shaded surface in Figure 5 represent the response values, w , excited by the input values, (E, G) . The shaded surface is a response surface. (The (E, G) plane in Figure 5 above which the response surface is defined includes other values in addition to the points included in $(E_i, G_i), i = 1, \dots, n_{gen}$.)

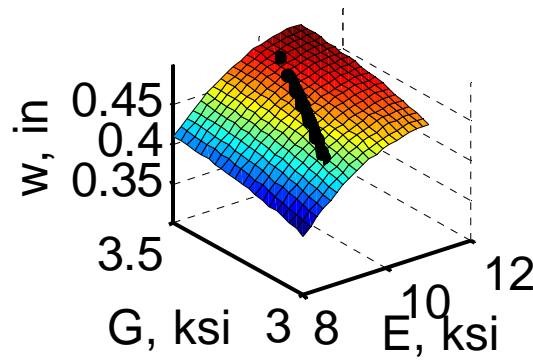


Figure 5. Peak displacement response of the structure for E in $[8000, 11,000]$ psi and G in $[3000, 3500]$ psi (shaded surface). Realized values of response measure, $w_i, i = 1, \dots, n_{gen}$ (black dots).

The Monte Carlo-computed response values, $w_i, i = 1, \dots, n_{gen}$, are also shown on the response surface in Figure 5. These are the realizations of random output corresponding to the random input pairs $(E_i, G_i), i = 1, \dots, n_{gen}$. The sample mean, sample variance, and sample standard deviation of the output values – the black dots in Figure 5 – can be computed using Eqs. (1), (2), and (3). They are

$$\bar{w} = 0.4621 \text{ in} \quad s_W^2 = 1.4513 \times 10^{-4} \text{ in}^2 \quad s_W = 0.0120 \text{ in} \quad (15)$$

The histogram of the output data is shown in Figure 6.

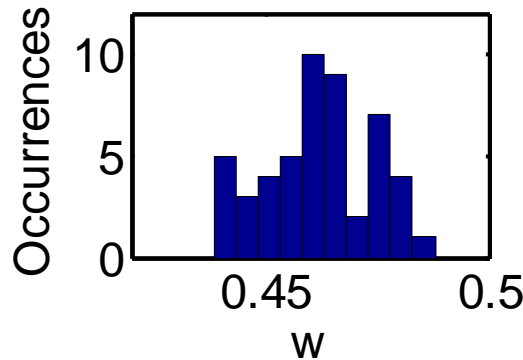


Figure 6. Histogram of the Monte Carlo-computed response.

Based on the appearance of the histogram of the outputs, Figure 6, we might model the response as a normal random variable with the moments in Eq. (15).

Comparison of Model Predictions to Experimental Response

The preceding sections described methods to (1) use univariate experimental data to develop a probability model and obtain samples from the model, (2) use bivariate experimental data to develop a probability model and obtain samples from the model, and (3) propagate data from a random source through an FEM. As noted in the first paper in this tutorial sequence (Paez, 2009), the objective of model validation is to compare predictions from a model – usually an FEM – to experimental results. We have emphasized in this paper that experimental structures and phenomena are random, and that randomness should be reflected in their models. Four comparison situations can occur.

- Though the experimental system is random, only one realization of the system may be available, and only one experiment might be performed on the system. (The reason may relate to expense, availability, time constraints, etc.) Further, the model of the system may be deterministic. That is, all the input data for the model may be non-random quantities. (The reasons for developing a deterministic model usually relate to

expense, time constraints, in-house expertise, etc.) Validation comparisons for this case involve requirements of arbitrary agreement between model predictions and experimental results and will not be considered here.

- As above, only one realization (or, a small number of realizations) of the system may be available, and only one experiment might be performed on the system. However, through experiments separate from the validation experiments, a probability model for the parameters of the structural model may be built. The validation comparisons, in this case, involve comparison of the single experimental output to the distribution of random outputs from the system model.
- The third comparison reverses the situation of the second. Here, multiple experimental results are available, but the model is deterministic. The validation comparisons, in this case, involve comparison of the single output from the system model to the distribution of experimental outputs.
- Finally, multiple experimental outputs may be available (from multiple experimental systems and/or multiple experiments on each system) and a probabilistic model may be available. In this case, the validation comparisons try to establish how well the model-predicted moments of response or probability distribution of response predict the corresponding quantities from the experiment.

We consider, briefly, the latter three types of comparisons. The comparisons in bullets 2 and 3, are similar, in the sense that one part (experiment or model prediction) is deterministic and the other can be modeled as random. Two relatively straightforward means for comparison will be outlined.

First, consider the case where the models developed in the previous subsections characterize the FEM-based predictions, and there is one structure-level experimental result, only. That is, the moments listed in Eq. (15) and the histogram shown in Figure 6, characterize the model prediction of peak response at a point. (Those results were propagated from a probability model involving material modulus of elasticity, E , and shear modulus, G . The probability model for E and G was obtained from experiments separate from the validation experiment.) Further, one structure has been fabricated, and that structure has been tested one time, to obtain the single measurement of peak response at the point of interest, and that measured response is $w^{(exp)} = 0.4210 \text{ in}$.

A simple validation criterion requires that the experimental result, $w^{(exp)}$, lie within the interval associated with model-predicted results and defined as

$$[\bar{w} - 1.96s_w, \bar{w} + 1.96s_w] \quad (16)$$

where the quantities defining the limits are from Eq. (15). The rationale behind this criterion is that if the peak response random source has a normal distribution, then about ninety-five percent of the model-predicted results will fall within the interval. In specifying this criterion, we assume that the experimental result is not an outlier, and we accept a five percent probability that our conclusion regarding validity of the model will be incorrect, i.e., that we will reject a model when, in fact, we should not. In this case, the quantitative interval is $[0.4385, 0.4857] \text{ in}$, so we reject the model, though it is close to the experimental result, and it is conservative, i.e., it tends to over-predict the structural peak response.

The validation criterion might be modified to

$$[\alpha(\bar{w} - 1.96s_w), \beta(\bar{w} + 1.96s_w)] \quad (16a)$$

where α and β are multiplicative factors chosen prior to the comparison. Normally, α would be a positive constant chosen equal to or smaller than one, and β would be a positive constant chosen equal to or greater than one. In the comparison considered, here, a value of α smaller than one permits the validation of models that produce conservative results, i.e., results that are greater than the experimental results. A value of β greater than one permits the validation of models that produce under-conservative results. In this example, if α had been chosen to equal a value of 0.96 or less, and β had been chosen as one, the model would have been judged valid.

It must be mentioned that there is nothing magical about the factor of 1.96 used in expressions (16) and (16a). It is obtained from a table of the standard normal distribution (Ang and Tang, 1975) and reflects a particular level of

probability, 0.95, in this case. Another factor could be used to develop a “looser” or “tighter” criterion. But validation of model predictions using a “loose” criterion says much less about model accuracy than validation of model predictions using a “tight” criterion.

A similar approach to validation comparisons defines an interval, not on the basis of moments, but on the basis of the data. The empirical cumulative distribution function (CDF) of a collection of data is defined as a stair-step function with increments of $1/n_{gen}$ plotted as a function of the sorted version of the data, $w_i, i = 1, \dots, n_{gen}$. The function is denoted $F_W(w), -\infty < w < \infty$, and the height of the function at any abscissa value, w , is interpreted as an estimate of the probability

$$F_W(w) = P(W \leq w) \quad -\infty < w < \infty \quad (17)$$

The probability on the right-hand side is the relative chance that when a random experiment is performed, the random variable W will assume a value (realization) equal to or lower than w . The empirical CDF of the model-predicted peak structural responses, $w_i, i = 1, \dots, n_{gen}$, is plotted in Figure 7.

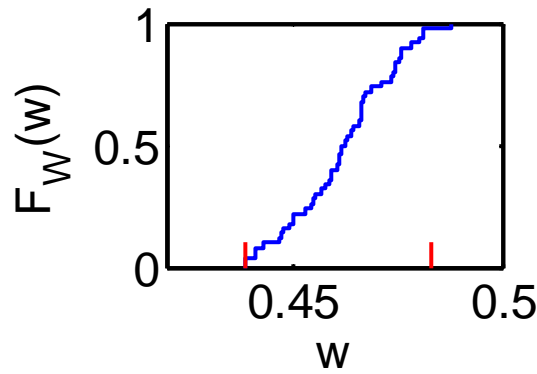


Figure 7. Empirical CDF of the model-predicted peak responses (blue). Estimated ninety-five percent probability interval on peak responses.

The empirical CDF can be used to obtain a probability interval of the random variable, W . A p -valued probability interval is the interval within which there is a probability, p , that a model-prediction lies. The quantity p must lie within the interval (0,1). The interval is obtained by inverting the empirical CDF. The limits are defined

$$L = F_W^{-1}\left(\frac{1-p}{2}\right) \quad U = F_W^{-1}\left(\frac{1+p}{2}\right) \quad (18)$$

and the interval is $[L, U]$.

A validation criterion might require that the experimental result, $w^{(exp)}$, lie within the interval $[L, U]$. The rationale behind this criterion is that the interval provides a good estimate of the range within which ninety-five percent of the results will lie, without regard to whether or not the random variable, W , is normally distributed. In specifying this criterion, as before, we assume that the experimental result is not an outlier, and we accept a $(1-p) \times 100$ percent probability that our conclusion regarding validity of the model will be incorrect.

We choose to use a ninety-five percent probability interval, here. In this case, the quantitative interval is $[0.4387, 0.4829]$ in, and it is shown by the red lines in Figure 7. The interval does not contain the experimental result, so we reject the model, though it is close to the experimental result, and it is conservative. Using the same rationale as in expression (16a) we can also define the validation criterion as the requirement that $w^{(exp)}$ lie within the interval $[\alpha L, \beta U]$ where the constants α and β are chosen as above.

When the situation is the opposite of that analyzed here, i.e., a probability model for the experimental results is available, and a deterministic model-prediction, only, is available, validation comparisons are completely analogous to the ones described above.

The fourth bullet listed above describes a situation in which sufficient data are available to create probability models for both the experimental outcomes and the model predictions. We already have model-predicted peak responses summarized in the moments of Eq. (15), the histogram of Figure 6, and the empirical CDF of Figure 8. Suppose that, in addition, we have multiple replicates of the experimental data. Let the experimentally measured peak responses be denoted $w_i^{(exp)}, i = 1, \dots, n_{exp}$, and let their empirical CDF be that shown in Figure 8a (blue). (The plot reflects $n_{exp} = 12$ experiments.) In addition, the empirical CDF of the model-predicted peak responses is repeated in the figure (red). It should be clear that if the two empirical CDFs are “close enough” the model is a valid representation of the physical system. The question is: How close do the curves need to be? The answer can be found in many ways, but a classical approach is embodied in the two-sample Kolmogorov-Smirnov test.

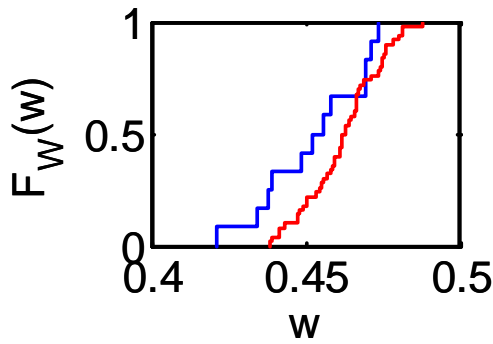


Figure 8a. Empirical CDF of the experimentally measured peak responses (blue). Empirical CDF of the model-predicted responses (red).

Effective Sample Size	Critical Statistic, Δ
5	0.56
10	0.41
20	0.34
25	0.29
30	0.26
40	0.21
Large n	$1.36/\sqrt{n}$

Figure 8b. Limits for the two-sample Kolmogorov-Smirnov test. Level of significance is 0.05.

The two-sample Kolmogorov-Smirnov test requires that the greatest vertical distance between the empirical CDFs be equal to or less than a critical statistic, Δ , in order for the model to be considered “equivalent” to the physical system in a statistical sense. Some values of the critical statistic are listed in Figure 8b. To use the two-sample Kolmogorov-Smirnov test, we need to compute an effective sample size from the actual sample sizes of the experimental data and the model-predicted data. The former quantity is n_{exp} , and the latter quantity is n_{gen} . The effective sample size is defined

$$n_{eff} = \frac{n_{exp}n_{gen}}{n_{exp} + n_{gen}} \quad (19)$$

Because $n_{exp} = 12$ and $n_{gen} = 50$, the effective sample size is $n_{eff} = 9.68$. The effective sample size is approximately equal to 10, therefore, the critical statistic, $\Delta = 0.41$ is used. The greatest vertical distance between the two empirical CDFs in Figure 8a is 0.32, therefore, the hypothesis that the model predictions form a satisfactory representation of the actual system cannot be rejected.

It must be noted that the critical statistics listed in Figure 8b are for the case in which the significance of the test is five percent. That is, when this criterion is applied, there is a five percent chance that the conclusion drawn from the comparison is incorrect. If it is desired to develop a comparison based on another level of significance, the critical values must be obtained from a suitable table. Benjamin and Cornell (1970) contains such a table and an explanation of how to use it.

The validation comparison developed, here, shows why, if at all possible, a comparison of probability models should be pursued. Note from Figure 8a that the smallest value of the experimentally measured peak response is 0.4210 in. That was the value used in the previous two validation comparisons, described earlier. Though the quantity is not an outlier, it falls at the lower extreme of the measured data. Whenever a deterministic quantity – whether from a single experimental result or as a parameter in a model – is used as the basis for a comparison we run the risk that the quantity will represent an extreme value. This can lead to rejection of a model that should, in fact, be accepted, though it does not necessarily occur in all cases.

Conclusions

This paper provides an elementary introduction to some of the tools useful for probabilistic uncertainty quantification, and, in particular, probabilistic tools useful for operations required in model validation. An example – based on real laboratory experiments – was used to show how to

- Develop models for random, univariate data
- Generate samples from the univariate model
- Develop models for random bivariate data
- Generate samples from the bivariate model
- Propagate randomness through a deterministic finite element model (FEM)
- Compare predictions from an FEM to experimental results in three cases
 - One experimental result is available and a probability model is available for the FEM outputs
 - Multiple experimental results are available and one FEM output is available
 - Probability models are available for both the experimental results and the FEM

Though many techniques are available for probabilistic data modeling, development of random FEM, and comparison of model predictions to experimental results, the tools and approaches presented here provide a fundamental basis for performing these operations.

Acknowledgment

Sandia is a multi-program 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.

References

- AIAA (American Institute of Aeronautics and Astronautics), (1998), *Guide for the Verification and Validation of Computational Fluid Dynamics Simulations*, AIAA-G-077-1998, Reston, VA, American Institute of Aeronautics and Astronautics.
- Ang, A., Tang, W., (1975), *Probability Concepts in Engineering Planning and Design, Volume 1 – Basic Principles*, John Wiley and Sons, New York.
- Ang, A., Tang, W., (1984), *Probability Concepts in Engineering Planning and Design, Volume 2 – Decision, Risk, and Reliability*, John Wiley and Sons, New York.
- ASME, (2006), *Guide for Verification and Validation in Computational Solid Mechanics*, The American Society of Mechanical Engineers, ASME V&V 10-2006.
- Benjamin, J. R., Cornell, C. A., (1970), *Probability, Statistics, and Decision for Civil Engineers*, McGraw-Hill Book company, New York.
- Carson, J. R., (1925), "Selective Circuits and Static Interference," *Bell System Technical Journal*, pp. 265-279.
- Collins, J. D., Thompson, W. T., (1969), "The Eigenvalue Problem for Structural Systems with Uncertain Parameters," *AIAA Journal*, Vol. 7, No. 4, pp. 642-648.
- Crandall, S., (Ed.), (1958), *Random Vibration*, Technology Press of MIT and John Wiley and Sons, New York.
- Field, Jr., R. V., Paez, T. L., Smallwood, D. O., (2007), "Validation of Random Vibration Environments," *Proceedings of IMAC XXVI, SEM, Orlando, (February)*, accepted for publication.
- Ghanem, R. G., Spanos, P. D., (1991), *Stochastic finite Elements: A Spectral Approach*, Springer-Verlag, New York.

- Ghanem, R. G., Doostan, A., Red-Horse, J. R., (2008), "A Probabilistic Construction of Model Validation," *Computer Methods in Applied Mechanics and Engineering*, V. 197, Issues 29-32, pp. 2585-2595.
- Montgomery, D. C., Runger, G. C., Hubele, N. F., (2006), *Engineering Statistics*, Fourth Edition, Wiley, New York.
- Paez, T. L., (2006), "The History of Random Vibrations," *MSSP* (Mechanical Systems and Signal Processing), Number 8, 2006.
- Paez, T. L., (2009), "Introduction to Model Validation," *Proceedings of the 2009 IMAC*, SEM, Orlando, FL.
- Smallwood, D., Gregory, D., Coleman, R., (2000), "Damping Investigations of a Simplified Frictional Shear Joint," *Proceedings of the 71st Shock and Vibration Symposium*, SAVIAC, The Shock and Vibration Information Analysis Center.
- U.S. Department of Energy, (2000), "Advanced Simulation and Computing (ASCI) Program Plan," 01-ASCI-Prog-01, Sandia National Laboratories, Albuquerque, New Mexico.

Appendix A – Experiments that Characterize the Parameters of a Bonding Material

Experiments were performed at Sandia National Laboratories to establish the parameters of some samples of a bonding material. Specifically, models for the modulus of elasticity, E , and the shear modulus, G , were sought. The bonding material was, eventually, to be used to bond together some shells of an aerospace structure. In order to obtain data for use in specification of a probability model, five nominally identical structures were built to infer the bonding material parameters. A schematic of the structure is shown in Figure a.



Figure a. Schematic of the structure used to infer bonding material parameters.

The structure consists of two steel masses, shown in black in Figure a, bonded together by a layer of the bonding material, shown in gray in Figure a. The diameter of all the disks in Figure a is 1.125 in. The thickness of both steel disks is 0.375 in. Five structures were fabricated and the average thicknesses of the bonding disks in the five structures are [0.506, 0.507, 0.507, 0.505, 0.506] in.

Each of the five structures was instrumented with tri-axial accelerometers on both steel disks. Dynamic experiments were performed on all five structures in which one of the steel disks was impacted with a modal hammer, and responses were measured. A modal analysis was performed on each structure, and the modal frequencies were identified. Observation of the mode shapes made is clear that, among many other modes, there existed modes that are primarily shear modes, and modes that are primarily axial modes. For all five experimental structures the identified modal frequencies are listed in Table A.1. Denote the experimentally obtained shear modes $f_i^{(sh)}, i = 1, \dots, 5$, and the axial modes $f_i^{(ax)}, i = 1, \dots, 5$.

Table A.1. Experimentally identified modal frequencies of the five structures.

Structure index	Shear mode frequency (Hz)	Axial mode frequency (Hz)
1	1750	2210
2	1800	2275
3	1787	2240
4	1762	2215
5	1750	2230

A finite element model (FEM) of the nominal structure was constructed. The model has known geometry, and known properties for the steel disks. The modulus of elasticity, E , and shear modulus, G , of the bonding material are unknown, but nominal values E_0 and G_0 can be used in their places. The nominal model can be used to

perform a modal analysis, and the shear mode and axial mode frequencies can be obtained. Denote these modal frequencies $f_{FE}^{(sh)}, f_{FE}^{(ax)}$.

In addition, the FEM can be used to evaluate the shear and axial modal frequencies at other values of the parameters E and G . The results of those analyses can be used to obtain an approximation to the sensitivity matrix of modal frequencies to material parameters. The sensitivity matrix is defined

$$\mathbf{S} = \begin{bmatrix} \frac{\partial f_{FE}^{(sh)}}{\partial E} & \frac{\partial f_{FE}^{(sh)}}{\partial G} \\ \frac{\partial f_{FE}^{(ax)}}{\partial E} & \frac{\partial f_{FE}^{(ax)}}{\partial G} \end{bmatrix} \quad (\text{A.1})$$

Given the sensitivity matrix, we can write a first order Taylor's series approximation to the experimental shear and axial modal frequencies in terms of the modal frequencies of the FEM.

$$\begin{Bmatrix} f_i^{(sh)} \\ f_i^{(ax)} \end{Bmatrix} \cong \begin{Bmatrix} f_{FE}^{(sh)} \\ f_{FE}^{(ax)} \end{Bmatrix} + \mathbf{S} \begin{Bmatrix} (E_i - E_0) \\ (G_i - G_0) \end{Bmatrix} \quad i = 1, \dots, 5 \quad (\text{A.2})$$

where $(E_i, G_i), i = 1, \dots, 5$, are the bonding material modulus of elasticity and shear modulus for the bonding material in the five structures. All the elements in Eq. (A.2) are known except for E_i and G_i , therefore, they can be identified. This relation was used to compute the sample E and G values provided in Figure 3a.