SAND96-1862C
CONF-970112--3

STATISTICAL VALIDATION OF SYSTEM MODELS

Patrick Barney
Experimental Structural Dynamics Department
Sandia National Laboratories
Albuquerque, New Mexico

Carlos Ferregut
Luis E. Perez
FAST Center for Structural Integrity of Aerospace Systems
The University of Texas at El Paso
El Paso, Texas

Norman F. Hunter
Engineering Science and Analysis Division
Los Alamos National Laboratory
Los Alamos, New Mexico

Thomas L. Paez
Experimental Structural Dynamics Department
Sandia National Laboratories
Albuquerque, New Mexico

Abstract

It is common practice in system analysis to develop mathematical models for system behavior. Frequently, the actual system being modeled is also available for testing and observation, and sometimes the test data are used to help identify the parameters of the mathematical model. However, no general-purpose technique exists for formally, statistically judging the quality of a model. This paper suggests a formal statistical procedure for the validation of mathematical models of systems when data taken during operation of the system are available. The statistical validation procedure is based on the bootstrap, and it seeks to build a framework where a statistical test of hypothesis can be run to determine whether or not a mathematical model is an acceptable model of a system with regard to user-specified measures of system behavior. The approach to model validation developed in this study uses experimental data to estimate the marginal and joint confidence intervals of statistics of interest of the system. These same measures of behavior are estimated for the mathematical model. The statistics of interest from the mathematical model

are located relative to the confidence intervals for the statistics obtained from the experimental data. These relative locations are used to judge the accuracy of the mathematical model. An extension of the technique is also suggested, wherein randomness may be included in the mathematical model through the introduction of random variable and random process terms. These terms cause random system behavior that can be compared to the randomness in the bootstrap evaluation of experimental system behavior. In this framework, the stochastic mathematical model can be evaluated. A numerical example is presented to demonstrate the application of the technique.

Introduction

Decisions based on any system analysis depend, naturally, on the mathematical model which is set up to predict the behavior of the system. However, if careful real life decisions are to be made, it is necessary that considerations about the validity of the model itself be taken into consideration. The validity of a model can be assessed by comparisons with other more involved, proven models that exhibit a closer representation of the physical system, or by comparisons with measures of collected data from the field or the laboratory. These so called real data of system behavior, however, contain in one way or another a certain degree of uncertainty, because any data collection and data processing scheme is never error free, and real systems have features that are unaccounted for in measurements and models. Real systems exhibit test-to-test variations, unit-to-unit variations (between units that are nominally identical), and measurement uncertainty.

Development of a system model is commonly guided by a balance between two requirements: (1) the need to represent reality, reflected by the measured data, and (2) the pragmatic need for a relatively simple mathematical model. Therefore, the validation of a model would depend on the degree of uncertainty associated with the measured data of system behavior and the number of basic variables, parameters and complexity of their interrelationships that have been included in the model. It is obvious from this that it is not particularly helpful to try to validate a model by calculating the differences of the results from the measured data. However, any alternative validation scheme should have a level of sophistication which does not alter the pragmatic level of complexity that characterizes the model. Further, it would be convenient if the model validation scheme makes full use of the information provided by the measured data.

Using the concept that the "real" behavior of the system in any measured behavior space is a random realization within the measured space, which is supposed to be represented by the model, a validation methodology based on statistical significance tests may be devised.

The bootstrap is a method for assessing the accuracy of arbitrary statistics of measured data. It was developed by Efron (1979) and is clearly explained in a text by Efron and Tibshirani (1993). It provides a means for estimating the standard error, confidence intervals, and bias in statistical estimates. It was developed for situations in which the underlying data are non-Gaussian, and the statistics of interest are non-Gaussian and not Gaussian-related. It can be used in the system analysis/system modeling framework to assess the accuracy of measures of system response and characteristics of systems, for example, response spectral density, cross-spectral density, frequency response function, modal parameters, and other measures of linear and nonlinear system response. The procedures for using the bootstrap to perform these statistical analyses are described in Hunter and Paez (1995), and Paez and Hunter (1996).

We propose in this paper a framework for statistical validation of system models when experimental data are available. The procedure includes the following steps. First, identify one or more measures of system character as the basis for validation of the mathematical model. (These measures might be quantities to be considered individually, or quantities to be considered jointly. For example, the second eigenfrequency of a linear system might be a quantity to be considered individually. The three individual average values of system response spectral density in three critical frequency ranges might be quantities to be considered jointly.) Next, using the bootstrap and the experimental data from the physical system, estimate the confidence intervals and the joint confidence intervals (as appropriate) for the measures of system character, say at the $(1-\alpha)x100\%$ level. Then evaluate these same measures of system character from the mathematical model. Locate the measures of system character from the mathematical model relative to the confidence intervals of measures of system behavior from the bootstrap analysis. Now make a statistical hypothesis: the measures of system character from the mathematical model are accurate representations of the corresponding measures from the actual system. Perform a statistical test of hypothesis. If the measures of system character from the mathematical model fall within the confidence intervals of the measures of system character from bootstrap analysis, then the hypothesis is accepted at the level of significance α. Otherwise, the hypothesis is rejected. The mathematical statistics of this framework are developed in this paper and a numerical example is presented to demonstrate use of the technique.

The development described above assumes that the model for the system under consideration is deterministic, in the sense that all its parameters are deterministic variables. However, it is clear that under certain circumstances it may be desirable to include parameters in the mathematical model that are random variables and random processes. After all, there are features of the system under consideration that cause the measures of its behavior to display the random variation to be characterized with the bootstrap analysis. In view of this, we describe in the following how mathematical models with random variable and random process parameters might be validated using a simple extension of the present technique.

It is possible to think of situations in which a model may be said to be validated in a "limited sense." Two of these situations may occur: 1) when the model predicts only a certain aspect of the system behavior, and 2) when only a subset of the variables representing that aspect of system behavior fall within the specified confidence intervals. In the first case, the argument can be made that if the predicted behavior is what the engineer was looking for, then the model could be satisfactorily valid under the constraint of this pragmatic condition, and as long as the predicted measurements fall within the defined confidence regions. However, the second situation immediately introduces the need for defining a scalar index that could be used to quantify the degree of validation.

In the following we first introduce the bootstrap - what it is, how it is used, how it is computed. Next, we show how the bootstrap can be used to compute confidence regions for measures of system behavior. We provide a simple example of the application of the bootstrap. Then we develop the framework for statistical validation of mathematical models. Finally, we present the results of an experimental example, demonstrating how the finite element model of an aluminum beam might be validated.

The Bootstrap

The bootstrap is a technique for the assessment of the accuracy of estimates of parameters of probability distributions. These estimates are statistics of measured data and their accuracy is estimated in terms of standard error, confidence intervals, and/or bias. To perform a bootstrap analysis, we measure data from a random source and assume that the observed data represent the source. The source is assumed to generate realizations with an unknown probability distribution. Each observed data point is assigned a probability of occurrence of 1/n, where n is the total number of data points measured. A bootstrap sample of the data is created by selecting at random, with replacement, n elements from the measured data set. This process is illustrated in Figure 1. The procedure is readily implemented using a uniform random number generator which selects, with equal probability, integer values in the range 1 to n. Sampling is done with replacement, so each bootstrap sample may have several occurrences of some data values and other data values may be absent.

$\hat{F} = X = (x_1, x_2, ..., x_{16})$ *(Samples have equal probability)*

$\downarrow$ *Creation of bootstrap sample is accomplished through random selection among elements of X. For example, let $X = (x_1, ..., x_{16})$. A potential bootstrap sample is shown below. (The sample contains* 16 *elements.)*

$X^* = (x_2, x_7, x_4, x_{11}, ..., x_4)$

Figure 1. Obtaining a bootstrap sample.

In a bootstrap analysis, numerous bootstrap samples are created. The statistic of interest is computed from each bootstrap sample; the resulting quantities are known as bootstrap replicates of the statistic of interest. Standard error, confidence intervals, and bias of the statistic of interest are computed using standard techniques and formulas on the bootstrap replicates of the statistic of interest. For example, let B denote the number of bootstrap samples used in an analysis, and let $\hat{\theta}^*(b), b = 1, ..., B$, denote the bootstrap replicates of the statistic of interest. Then the standard error of the statistic of interest is estimated with

$$\hat{se} = \left[\frac{1}{B-1} \sum_{b=1}^{B} \left(\hat{\theta}^*(b) - \hat{\theta}^*(.) \right)^2 \right]^{1/2} \qquad (1)$$

where $\hat{\theta}^*(.) = \frac{1}{B} \sum_{b=1}^{B} \hat{\theta}^*(b)$.

In one type of bootstrap analysis, the two-sided, (1-α)x100% confidence interval is obtained by sorting the bootstrap replicates of the statistic of interest, and identifying (or interpolating) the (α/2)x100% percentile value and the (1-α/2)x100% percentile value in the sorted list, and using the identified values as the limits of the confidence interval. Another more advanced method for confidence interval estimation is discussed in Efron and Tibshirani (1993).

The number of bootstrap samples, B, used in an analysis, ranges from 25 to several thousand. The standard error of a parameter estimate may be computed using 25 to 50 bootstrap samples. Accurate computation of the confidence intervals of an estimated parameter requires analysis of a thousand or more bootstrap samples.

Bootstrap sampling provides an optimal estimate of the probability density function which characterizes the data source given that our knowledge of the source is limited to the measured data. Computation of a statistic from the bootstrap samples simulates computation of the same statistic on samples drawn from the real world distribution. Properties of the "real world" distribution are estimated in the "bootstrap world" as illustrated in Figure 2.

Real World

Unknown Probability Distribution → Observed Sample

$$F \rightarrow X = (x_1, x_2, \ldots, x_n)$$

$$\downarrow$$

$$\hat{\theta} = s(X)$$

Statistic of Interest

Bootstrap World

Observed Distribution → Bootstrap Samples

$$\hat{F} \rightarrow X^* = \left(x_1^*, x_2^*, \ldots, x_n^*\right)$$

$$\downarrow$$

$$\hat{\theta}^* = s\left(X^*\right)$$

Bootstrap Replicates of Statistic of Interest

Figure 2. The bootstrap approximation to the real world. The observed distribution is our best estimate of the true distribution. The observed sample is X, and the statistic of interest $\hat{\theta} = s(X)$ can be computed based on this. In the bootstrap world the observed data are used to generate as many bootstrap samples X^* as we wish. Each bootstrap sample is used in the formula $\hat{\theta}^* = s\left(X^*\right)$ to compute a bootstrap replicate of the statistic of interest. The bootstrap replicates are used to analyze the standard error, confidence intervals and bias of the statistical estimator.

Confidence Regions for Measures of Mechanical System Behavior

We showed in the previous section that the bootstrap is a technique for the accuracy analysis of statistics of random data. Among other things, it can be used to estimate standard error and the confidence intervals of statistical estimators. Figures 1 and 2 and the text in the previous section make it clear that in order to use the bootstrap we need to build up an ensemble of bootstrap replicates of the statistic of interest. In this section we seek to demonstrate that a general approach to the generation of bootstrap replicates can be developed in a very practical framework.

To commence the development we assume that measured inputs (if required) and outputs from the system to be characterized are available. Denote these $(X,Y)=(x_1,\ldots,x_n,y_1,\ldots,y_n)$ where y_i is an output corresponding to input x_i. We assume that one or more statistics of these data are the measures of system behavior or parameter estimates of interest. To keep this discussion general, we denote the statistics of interest as the vector of quantities

$$\{\hat{\theta}\}=s(X,Y) \tag{2}$$

where the function $s(.)$ yields a vector output. These parameter estimates or measures of system behavior can be any quantities that are mathematically describable in terms of the measured input and response data (X,Y). There are hundreds, perhaps thousands, of examples in different fields of interest of what these parameters might represent. In general, for example, they might be:

Constant coefficients or parameters of variable coefficients of linear or nonlinear parametric algebraic equations

Constant coefficients or parameters of variable coefficients of linear or nonlinear parametric ordinary differential equations

Constant coefficients or parameters of functional coefficients of partial differential equations

Measures of behavior that assume a framework for system operation like impulse response functions of transfer functions or frequency response functions

Eigenvalues or eigenfunctions of systems of equations assumed to govern the measured data

Quantities that characterize nonlinear or possibly chaotic systems, like Lyapunov exponents, fractal dimensions of chaotic attractors, and other measures.

Given that the parameters of interest can be estimated using an expression with the form of Eq. (2), they can also be estimated using a bootstrap sample of the data. A bootstrap replicate of the statistics of interest can be denoted

$$\left\{\hat{\theta}^*\right\} = s\left(X^*, Y^*\right) \tag{3}$$

where the $\left(X^*, Y^*\right)$ are bootstrap samples of the measured data (X, Y). To perform the computation in Eq. (2) or (3) using measured data in a practical way may require some imagination in dealing with the data, but it can usually be done directly. (See Hunter and Paez, 1995, or Paez and Hunter, 1996, for specific descriptions of how bootstrap replicates of such quantities as estimates of autospectral density, cross-spectral density, frequency response function, eigenvalues, and eigenvectors can be obtained from measured data.)

Any number of bootstrap replicates $\left\{\hat{\theta}_b^*\right\}$, $b = 1, \ldots, B$, can be generated using the approach and the formulas described above. These replicates can be used to compute the accuracy statistics of interest. The descriptor of special interest in the present application is the confidence interval (if there is only one parameter, or if we are interested only in the marginal behavior of the individual quantities in $\left\{\hat{\theta}\right\}$) or the joint confidence region for multiple parameters in $\left\{\hat{\theta}\right\}$. The reason is that these will be used later as the basis for a test of hypothesis. When there is only one parameter in the vector $\left\{\hat{\theta}\right\}$ then its confidence interval can be obtained as described in the previous section. When there are multiple parameters, their confidence region can be obtained as follows.

Each of the bootstrap replicates $\{\hat{\theta}_b^*\}$, $b = 1,...,B$, occupies a point in the space whose coordinates are defined by the elements of $\{\hat{\theta}\}$. That is, if the vector $\{\hat{\theta}\}$ has N elements, then each of the bootstrap replicates has N corresponding elements, and these replicates are, in general, different. The collection of bootstrap replicates constitutes a measured ensemble from the random process source that has the sampling distribution of the vector $\{\hat{\theta}\}$. When the ensemble of generated bootstrap replicates is large enough it can be used to empirically infer the characteristics of the sampling distribution of $\{\hat{\theta}\}$. Among other things, the limits of the measured ensemble can be used to infer confidence regions for the parameter estimates.

The manner in which the confidence regions are constructed using the ensemble of generated bootstrap replicates is open to the discretion of the analyst. However, there are two general approaches for obtaining confidence intervals. These are the parametric and nonparametric approaches. With nonparametric approaches the analyst seeks to define a confidence region that accurately reflects the shape of the joint probability density function (pdf) of the source of the bootstrap replicates. The methods for accomplishing this are so varied that we will not pursue their description here. The idea behind parametric approaches is that a parametric form for the confidence region that approximately reflects the contours of the joint pdf can be specified and its parameters identified. For example, a multidimensional ellipsoid might be appropriate in many applications for the specification of the confidence region of multiple statistics of measured data.

Example. Let X be a random variable defined as

$$X = U^2 + 0.2Z \qquad \text{(a)}$$

where U is a uniform(0,1) random variable, and Z is a standard normal random variable, independent of U. Create 20 realizations of the random variable X, and from these realizations create 1000 bootstrap samples. From each bootstrap sample create a bootstrap replicate of the mean estimator of X and a bootstrap replicate of the skewness estimator of X. The skewness estimator replicates are plotted versus the corresponding mean estimator replicates in Figure 3, and clearly show some degree of correlation. For this example the mean estimator random variable and the skewness estimator random variable are the two elements of the vector $\{\hat{\theta}\}$. The bootstrap replicates of the mean estimator and the skewness estimator are the bootstrap replicates $\{\hat{\theta}_b^*\}$, $b = 1,...,1000$. We assume that joint confidence regions of the mean and skewness estimators can be defined using ellipsoids. Therefore, to define the 95% joint confidence region for the mean and skewness estimators, we identify the ellipsoid that encompasses 95% of the points in Figure 3; the ellipsoid matches the general shape of the distribution of the replicates. This is the region enclosed by the ellipsoid in Figure 3.

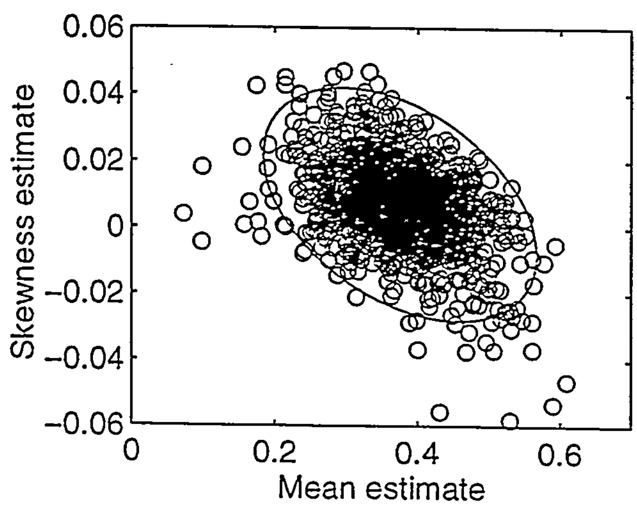


Figure 3. Joint bootstrap replicates (shown by circles) of the mean and skewness estimators based on 20 realizations of the random variable defined in Eq. (a), and the estimated 95% confidence region.

Validation of Mathematical Models

The procedure for validation of mathematical models is based on the concept of statistical hypothesis testing. We assume that a mathematical model for the system under consideration has been constructed, and that characteristics of the mathematical model that correspond to the characteristics of the actual system can be obtained. The characteristics of the mathematical model that correspond to the parameters evaluated for the experimental system will be denoted $\left\{\hat{\theta}_{\text{mod}}\right\}$. These model parameters must be obtained in a manner compatible with the specification of the model. For example, eigenvalues and eigenfunctions of a linear model can be obtained directly from the model. On the other hand, nonlinear mathematical models may require the use of measured inputs to compute simulated outputs, followed by the use of the measured inputs with the simulated outputs in Eq. (2) to evaluate the model parameters.

To perform a model validation, we first make the hypothesis that the mathematical model is a satisfactory representation of the actual system with respect to the parameters in the vector $\left\{\hat{\theta}\right\}$. We test this hypothesis at the $100\alpha\%$ level of significance by estimating the $(1-\alpha)\times 100\%$ confidence region for $\left\{\hat{\theta}\right\}$. We can use the bootstrap to accomplish this. Denote this region $\boldsymbol{R}_{1-\alpha}\left(\hat{\theta}\right)$. If

$$\left\{\hat{\theta}_{\text{mod}}\right\} \in \boldsymbol{R}_{1-\alpha}\left(\hat{\theta}\right) \tag{4}$$

then we accept the hypothesis and consider the model validated with respect to the parameters $\left\{\hat{\theta}\right\}$. Otherwise, we reject the hypothesis and consider the model invalid with respect to the parameters $\left\{\hat{\theta}\right\}$.

Figure 4 shows a schematic representation of the model validation procedure. Recall that model validation is performed from two ends. From one end, a confidence region is specified for statistics of data measured from an experimental system. From the other end, the parameters of interest are computed from the mathematical model. In the middle, the parameters from the mathematical model are located with respect to the confidence region, and the accuracy of the mathematical model is confirmed or rejected.

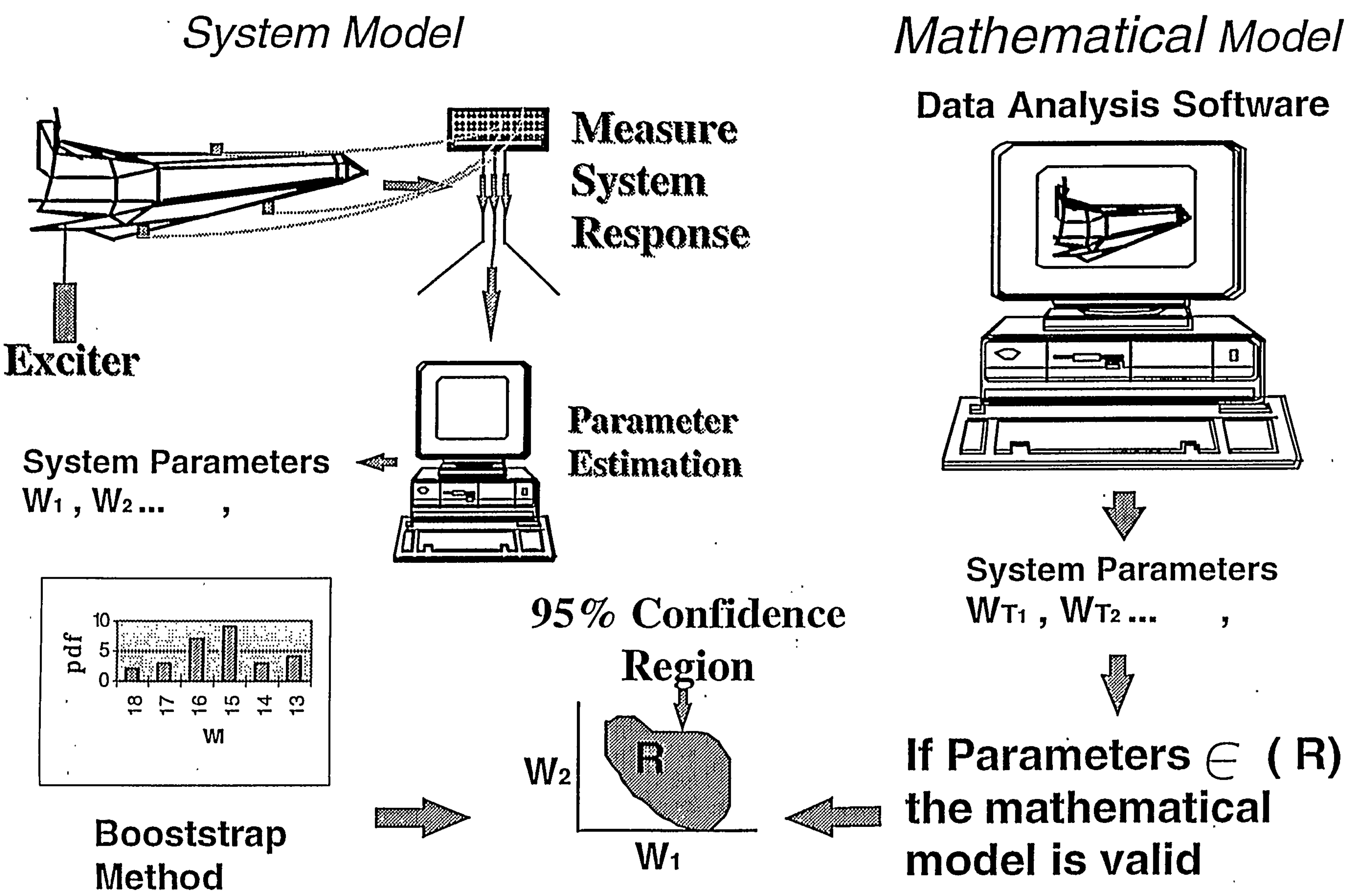
The validation procedure described here can be applied to any individual measure or sets of measures of model characteristics. It is anticipated that in practical applications a good mathematical model will be validated with respect to some measures and will not be validated with respect to other measures. The only way that a mathematical model might be validated with respect to all measures of system performance is that it perfectly incorporates every bit of information in the experimentally measured inputs and responses (X,Y). Even if this could be done, the mathematical model would likely fail if it were tested with reference to other sets of experimentally measured inputs and responses. Therefore, in practical applications it should only be hoped that a mathematical model might be validated with respect to some fundamental set of parameters or measures of system response.

Numerical Example

The system to be considered in this example is a simple, elastic, aluminum beam. Its dimensions are $24 \times 1 \times 0.25$ inches. The beam was suspended from one end by a string to simulate free boundary conditions. A piezoelectric accelerometer was mounted at the end opposite the string attachment point to measure the system's response. An additional mass was added to the beam during each experiment; specifically, an accelerometer was mounted to the beam at a random location. The purpose of this was to simulate the random variation in a complex system. The parameters of interest in the beam are its linear model parameters,

Figure 4.

Mathematical Model Validation



and in particular, its eigenfrequencies or modal frequencies. These can be estimated by first estimating the frequency response function of the beam, then fitting a linear model to the frequency response function, and finally inferring the modal frequencies from the linear model.

The experiments performed on the beam to obtain its dynamic characteristics are impact tests. During each experiment the beam is excited by impacting it with an instrumented hammer near the string attachment end; the input force and the response (near the free end) are measured. The impact experiment was repeated 10 times with one input and one response measurement during each experiment. This time history information was then used in the manner outlined in Hunter and Paez (1995) to estimate the system frequency response function. The first three modal frequencies of the beam were inferred from the frequency response function. The beam frequency response function is shown in Figure 5. This is an average based on the 10 experimental measurements.

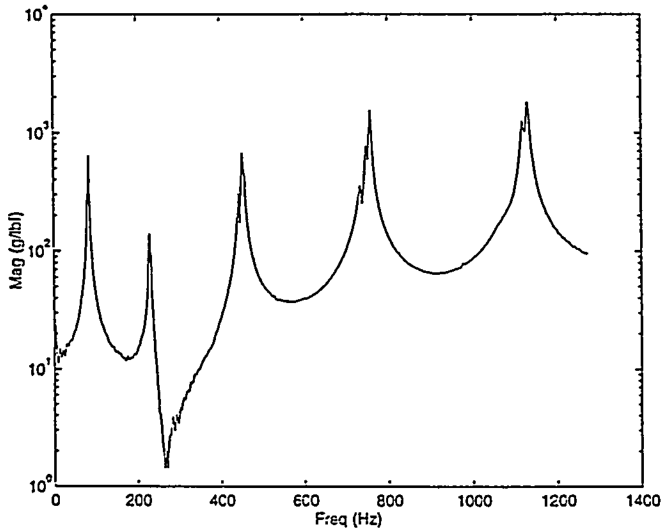


Figure 5. Frequency response function of the beam used in the experiment.

One thousand bootstrap samples of the input/response data were formed by sampling among the 10 input/response pairs, and a bootstrap replicate of the frequency response function was formed from each bootstrap sample. There is substantial variation among the bootstrap replicates of the frequency response function, and this variation is depicted near the first modal frequency in the shade density diagram of Figure 6. The figure is lightly shaded in regions where many replicates of the frequency response function lie, and it is darkly shaded where there are few replicates. The first three modal frequencies of the system were inferred from each frequency response function, thereby creating 1000 bootstrap replicates of the beam modal frequencies. The kernel density estimators (estimators of the pdf's) of the first three beam modal frequencies are shown in Figure 7. (See Silverman, 1986, for a description of the kernel density estimator.) It is apparent from the kernel density estimators that the sampling distributions of the modal frequencies are skewed, and that they are not all skewed in the same direction. Further, the dispersion in modal frequency values increases as the mode number increases. The percentage points of the cumulative distribution function estimators that are the integrals of the kernel density estimators of the beam modal frequencies are used directly to establish the confidence intervals for the modal frequencies. For example, the 99% marginal confidence intervals for the first three modal frequencies are

$$(81.58, 83.27)\ Hz, \quad (228.12, 231.37)\ Hz, \quad (446.85, 455.30)\ Hz \qquad (5)$$

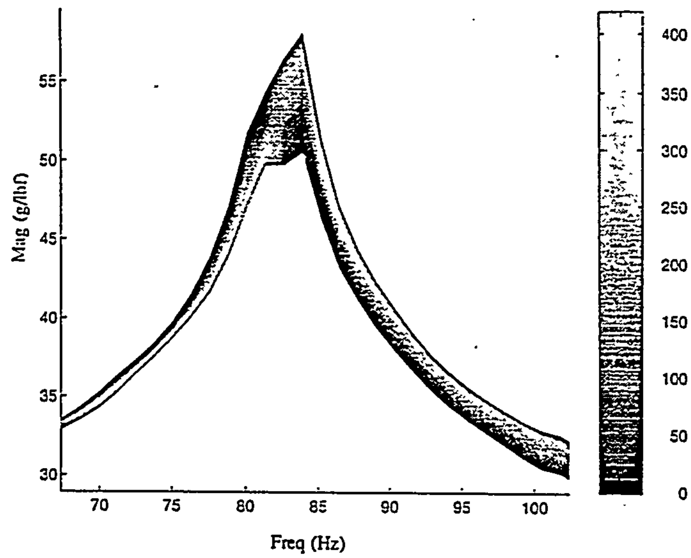


Figure 6. Shade diagram showing the density of bootstrap replicates of frequency response function near the first modal frequency of the system described above.

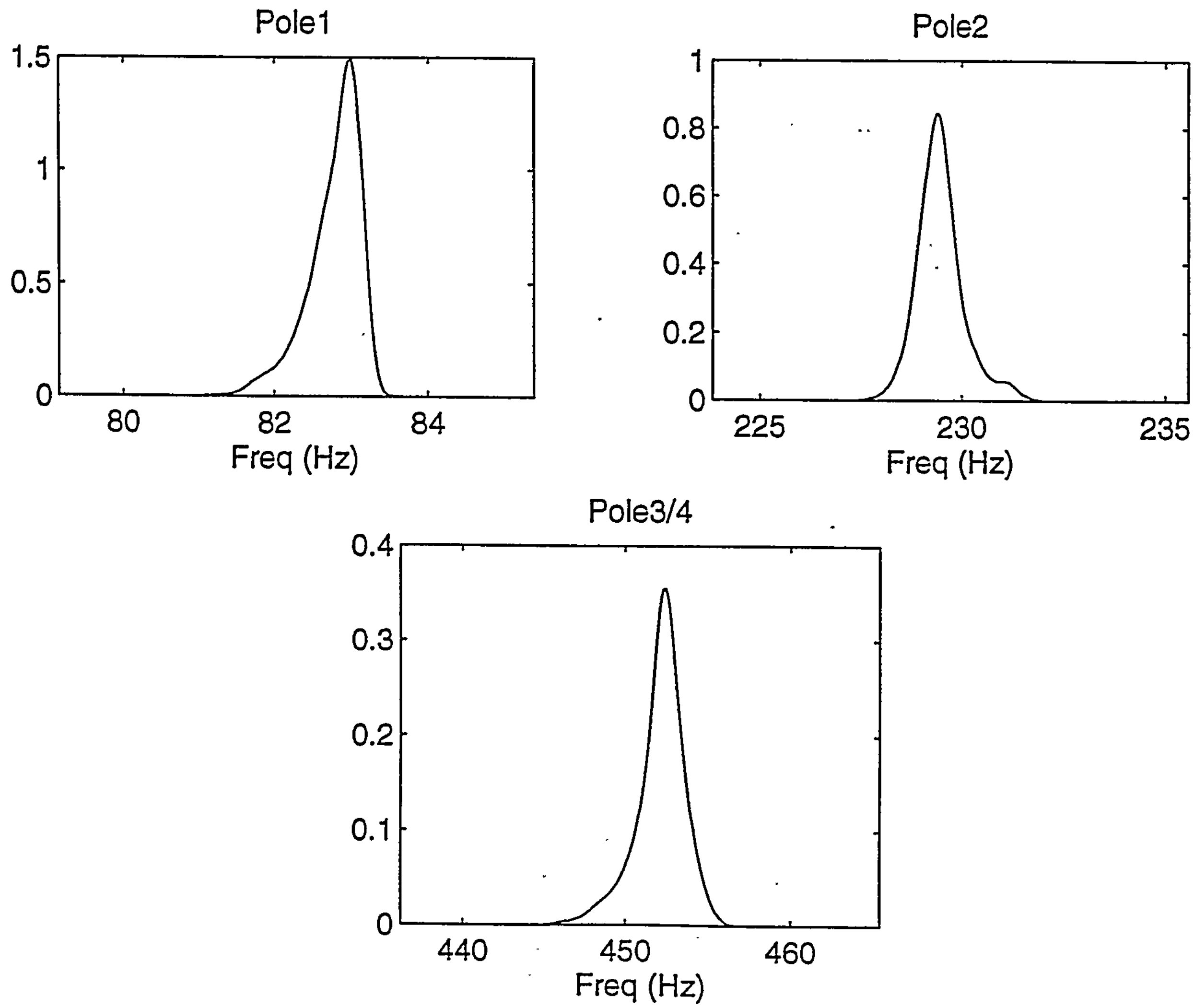


Figure 7. Kernel density estimators for the sampling distributions of the first, second, and third modal frequencies of the experimental system.

The mathematical model of the beam used in this experiment is a finite element model created in the computer code ALGOR using three dimensional beam elements. Twenty-four elements were used to model the beam. Accelerometers were modeled as lumped masses; the second lumped mass was located at the center of the beam. Because the code cannot simulate free-free boundary conditions, the boundary conditions were modeled using rotational and translational springs with very small elastic constants. The modal frequencies of the beam model can be computed directly in the finite element code, and the first three modal frequencies are

$$83.37\ Hz, \quad 232.92\ Hz, \quad 447.60\ Hz \tag{6}$$

Comparing the modal frequencies from the finite element model in Eqs. (6) to the 99% confidence intervals in Eqs. (5) indicates that the finite element model would not be validated with respect to the first and second modal frequencies at the one percent level of significance, and it would be validated with respect to the third modal frequency. Note that no attempt was made to reconcile the finite element model to the experimentally measured data or to the computed modal frequencies. In this particular case, the analyst could reasonably modify parameters in the mathematical model to cause the first, second, and perhaps the third, modal frequencies to match the measured data. This is one of the important points in the model validation framework described here.

Extension of the Validation Concept to Stochastic Mathematical Models

The model validation concept described in the previous sections acknowledges the presence of randomness in statistics of data measured from an experimental system. This is the reason why we can develop confidence regions for parameters and measures of system performance. It is assumed that the mathematical models used to simulate the actual systems are deterministic in the sense that the mathematical form of the model is prescribed, and the parameters of the model are deterministic constants.

To extend this concept we might seek to introduce the potential for randomness in the mathematical model through the introduction of model terms that are random variables or random processes. Such a mathematical model is known as a stochastic model. This requires that we know specifically where randomness might arise in the mathematical model, or that we be able to introduce generic terms whose influence on the model output mimics the behavior of the actual system even though it does not precisely match the system phenomenology. Either way, the effect of the introduction of randomness into the mathematical model is to create random variation in the behavior of the simulated system. This random variation may be quite difficult to analyze in most systems and requires the execution of probabilistic system analysis. A probabilistic system analysis might be performed using a Monte Carlo approach, a semi-analytic approach, or a hybrid approach. For descriptions of these types of analyses, see for example Madsen, Krenk, and Lind (1986). These analytic approaches cannot be described in detail here for lack of space.

However the stochastic model is analyzed, the net objective with regard to model validation is that we seek to identify regions in the space of the vector of parameters $\{\hat{\theta}\}$ where the parameters are most likely to fall. Such regions are equivalent to the confidence region $R_{1-\alpha}(\hat{\theta})$ defined above. Denote the region obtained from the stochastic model where the parameters have a probability of 1-α of occurring as $M_{1-\alpha}(\hat{\theta})$. Our effort in validating the stochastic model is to determine whether the regions $R_{1-\alpha}(\hat{\theta})$ and $M_{1-\alpha}(\hat{\theta})$ are equivalent. We can do this using any of at least three methods.

First, note that the region $M_{1-\alpha}(\hat{\theta})$ can be established based on realizations from the stochastic model in the same way that the region $R_{1-\alpha}(\hat{\theta})$ was established based on the bootstrap replicates $\{\hat{\theta}_b^*\}$, $b = 1,...,B$. Using Monte Carlo techniques we can generate B realizations of the vector $\{\hat{\theta}\}$ from the stochastic mathematical model. The difference between each realization and a corresponding bootstrap replicate from the experimental data can be computed, creating an ensemble of vectors

$$\{D_b\} = \{\hat{\theta}_{\text{mod},b}\} - \{\hat{\theta}_b^*\} \qquad b = 1,...,B \tag{7}$$

where $\{\hat{\theta}_{\text{mod},b}\}$ is the b^{th} realization from the stochastic model. The bootstrap method can then be used to establish whether or not the average of the random vector $\{D_b\}$ can be accepted as equal to zero.

A second method for judging the equivalence between the regions $R_{1-\alpha}(\hat{\theta})$ and $M_{1-\alpha}(\hat{\theta})$ is to use the bootstrap replicates $\{\hat{\theta}_b^*\}$, $b = 1,...,B$, to identify a transformation to a space of uncorrelated, standard normal random variables, use the transformation on realizations from the stochastic model, and use a test of hypothesis to determine whether or not the transformed stochastic model realizations belong to the standard normal space. This is accomplished by first using the bootstrap replicates to approximate the pdf of $\{\hat{\theta}\}$. A good framework for writing the approximate joint pdf of the elements in $\{\hat{\theta}\}$ based on the bootstrap replicates $\{\hat{\theta}_b^*\}$, $b = 1,...,B$, is the kernel density estimator, described in Silverman (1986). From the approximate joint pdf, the conditional pdf's of the elements in

$\{\hat{\theta}\}$ can be written, and based on these, a transformation from the space of $\{\hat{\theta}\}$ to the space of uncorrelated, standard normal random variables can be written. The transformation that accomplishes this is known as the Rosenblatt transform. (See Rosenblatt, 1952.) Now if realizations $\{\hat{\theta}_{\text{mod},b}\}$, $b = 1,...,B$, from the stochastic model are operated on with the Rosenblatt transform, and if the regions $R_{1-\alpha}(\hat{\theta})$ and $M_{1-\alpha}(\hat{\theta})$ are equivalent, then the realizations $\{\hat{\theta}_{\text{mod},b}\}$, $b = 1,...,B$, will appear normal in the transform space. A chi squared test can be used to test the appropriate hypothesis.

A third method for assessing the equivalence between the regions $R_{1-\alpha}(\hat{\theta})$ and $M_{1-\alpha}(\hat{\theta})$ is a simple visual inspection and comparison of projections of the two regions into two dimensional spaces of pairs of coordinates in $\{\hat{\theta}\}$. Even if one of the previous two methods for comparing the regions is adopted this approach is a prudent double check on the results.

Conclusions

We have developed in this paper an approach to statistical model validation that is based on the bootstrap method for statistical analysis. The approach accounts for randomness in real system characteristics and the data measured from real systems, and can be extended to account for randomness in the characteristics of the mathematical model. The approach is formal and systematic in that it is based on a well established statistical analysis procedure, and it provides an objective measure of the interval that a model parameter must occupy in order to be considered representative of the actual system at a particular level of significance. The approach is computer intensive; that is, it is time consuming to generate bootstrap samples and replicates of the statistics of interest. However, its advantage is that it properly accounts for the non-Gaussian nature of arbitrary statistics of interest.

It must be emphasized that the analyst who uses the proposed procedure for statistical model validation must be judicious in his or her choice of the specific measures and the number of measures of model performance used to validate the model. The number of measures should be neither too great nor too small, and should reflect the importance of the application. The specific measures of performance used should reflect the analyst's expectations of the model. Some measures of performance (like average measures of system behavior over a broad region) will be easier to validate than others. However, when detailed model behavior is validated, model performance in the simulation of detailed behavior will be anticipated to be accurate.

Acknowledgment

This work has been supported by the Department of Energy under contract number DE-AC04-94AL85000, by the Air Force Office of Scientific Research under contract number F49620-95-1-051B, and by the University of Texas at El Paso.

References

Efron, B. (1979), "Bootstrap Methods: Another Look at the Jackknife," *Annals of Statistics*, 7, 1-26.

Efron, Bradley and Tibshirani, Robert J.(1993), *An Introduction to the Bootstrap*, Applied Monographs on Statistics and Applied Probability 57, Chapman and Hall, 1993.

Hunter, N., Paez, T., (1995), "Application of the Bootstrap to the Analysis of Vibration Test Data," *Proceedings of the 66th Shock and Vibration Symposium*, Shock & Vibration Information Analysis Center, Biloxi, Mississippi, 99-108.

Madsen, H. O., Krenk, S., Lind, N. C., (1986), *Methods of Structural Safety*, Prentice-Hall, Englewood Cliffs, New Jersey.

Paez, T., Hunter, N., (1996), "Statistical Analysis of Modal Parameters Using the Bootstrap," *Proceedings of the 14th International Modal Analysis Conference*, V. 1, Society for Experimental Mechanics, Inc., Dearborn, Michigan, 240-245.

Rosenblatt, M. (1952), "Remarks on a Multivariate Transformation," *Annals of Mathematical Statistics*, 23, 3, pp. 470-472.

Silverman, B.W.(1986), *Density Estimation for Statistics and Data Analysis*, Chapman and Hall.