STATISTICAL VALIDATION OF PHYSICAL SYSTEM MODELS

Thomas L. Paez
Patrick Barney
Experimental Structural Dynamics Department
Sandia National Laboratories
Albuquerque, New Mexico
(505) 844-7052

Norman F. Hunter
Engineering Science and Analysis Division
Los Alamos National Laboratory
Los Alamos, New Mexico
(505) 667-2099

Carlos Ferregut
Luis E. Perez
FAST Center for Structural Integrity of Aerospace Systems
The University of Texas at El Paso
El Paso, Texas
(915) 747-6921

It is common practice in applied mechanics to develop mathematical models for mechanical system behavior. Frequently, the actual physical system being modeled is also available for testing, 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 physical systems when data taken during operation of the physical 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 physical 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 physical 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. A numerical example is presented to demonstrate the application of the technique.

Introduction

Mathematical models are often used in engineering practice to connect system behavior to a specific phenomenology, to serve as an ordered framework for experimentally generated information, or to provide a structure for predicting future responses to inputs. Examples of mathematical models are: finite element models, finite difference models, modal models, and artificial neural network models. Sometimes test data for the modeled system are also available. In both phenomenological and non-phenomenological applications these data are often used to identify the parameters of the mathematical model. Though ad hoc approaches are often developed to characterize the accuracy of mathematical models, there are no generally applicable formal procedures for judging the quality of the mathematical models.

DISCLAIMER

DISCLAIMER

Portions of this document may be illegible
in electronic image products. Images are
produced from the best available original
document.

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 structural dynamics framework to assess the accuracy of measures of structural response and characteristics of mechanical systems, for example, response spectral density, cross-spectral density, frequency response function, modal parameters, etc. 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 physical system models when experimental data are available. The procedure includes the following steps. First, identify one or more measures of physical system character as the basis for validation of the mathematical model, for example, the first three modal frequencies of a linear system. Second, using the bootstrap and the experimental data from the physical system, estimate the confidence intervals on the modal frequencies, say at the 99% level. (The 99% confidence intervals define the region within which we believe with 99% certainty the actual modal frequencies lie.) Third, evaluate the modal frequencies from the mathematical model and locate them relative to the confidence intervals from the bootstrap analysis. Now make a statistical hypothesis: the modal frequencies from the mathematical model are accurate representations of the corresponding modal frequencies from the actual system. Perform a statistical test of hypothesis. If the modal frequencies from the mathematical model fall within the confidence intervals on the modal frequencies from bootstrap analysis then the hypothesis is accepted at the one percent level of significance. Otherwise, the hypothesis is rejected. The mathematical statistics of this framework are developed in this paper.

A numerical example is presented to demonstrate use of the technique.

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. (At this point, the data are considered arbitrary. Later, we will interpret the data as time series.) 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)*

Creation of bootstrap sample is accomplished through random selection among elements of X.

$$\downarrow$$

For example, let $X = (x_1, ..., x_{16})$. A potential bootstrap sample is shown below. (The sample contains 16 elements.)
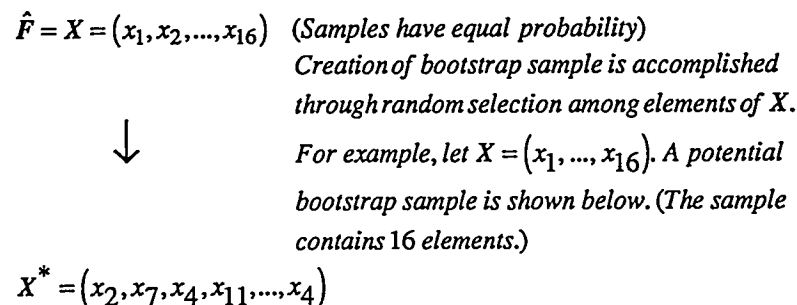
$$X^* = (x_2, x_7, x_4, x_{11}, ..., x_4)$$

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. (Here the statistic of interest is arbitrary. Later, in the numerical example, we will take the statistics to be modal frequencies of a linear system.) 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

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

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

The first expression is the standard deviation estimator of the bootstrap replicates; the second is their sample mean. In one type of bootstrap analysis, the two-sided, $(1-\alpha) \times 100\%$ confidence interval is obtained by sorting the bootstrap replicates of the statistic of interest, and identifying (or interpolating) the $(\alpha/2) \times 100\%$ percentile value and the $(1-\alpha/2) \times 100\%$ 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.
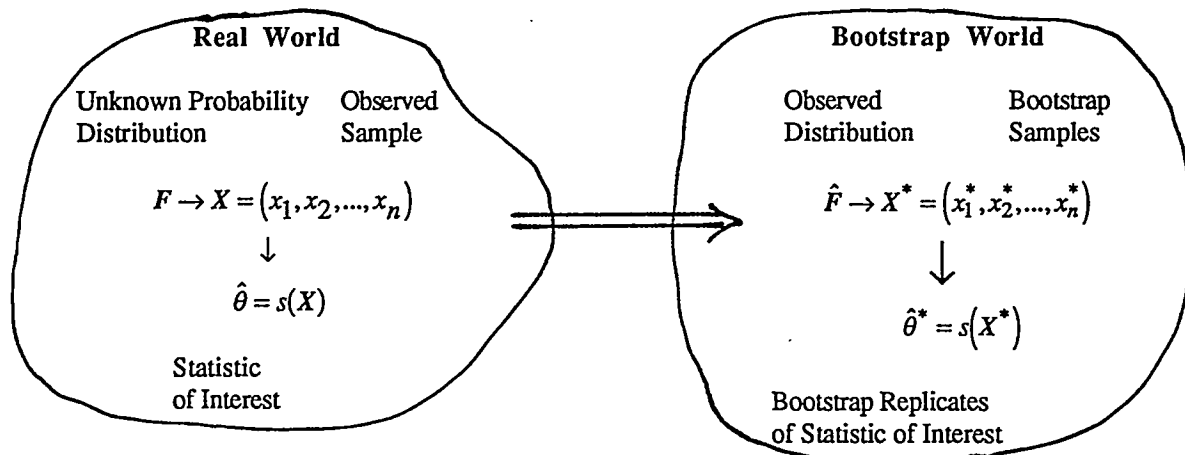

Figure 2. The bootstrap approximation to the real world. The observed distribution is our best estimate of the true distribution. The observed sample is X, and the statistic of interest $\hat{\theta} = s(X)$ can be computed based on this. In the bootstrap world the observed data are used to generate as many bootstrap samples X^* as we wish. Each bootstrap sample is used in the formula $\hat{\theta}^* = s(X^*)$ 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 mechanical system to be characterized are available. Denote these $(X,Y) = (x_1,...,x_n,y_1,...,y_n)$ 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 $s(.)$ may yield 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 many examples of what these parameters might represent. In general, for example, they might be:

> Measures of dynamic response such as auto and cross-spectral density, response shock spectra, etc.
> Measures of linear system character such as frequency response functions, modal frequencies, mode dampings, mode shapes, state space matrices, etc.
> Identified parameters associated with finite element models.
> 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

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

where the (X^*,Y^*) 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 $\{\hat{\theta}_b^*\}$, $b = 1,...,B$, 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 $\{\hat{\theta}\}$) or the joint confidence region for multiple parameters in $\{\hat{\theta}\}$. 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 $\{\hat{\theta}\}$ 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 $\left\{\hat{\theta}_b^*\right\}$, $b = 1,...,B$, occupies a point in the space whose coordinates are defined by the elements of $\left\{\hat{\theta}\right\}$. That is, if the vector $\left\{\hat{\theta}\right\}$ 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 $\left\{\hat{\theta}\right\}$. When the ensemble of generated bootstrap replicates is large enough it can be used to empirically infer the characteristics of the sampling distribution of $\left\{\hat{\theta}\right\}$. 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

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

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 $\left\{\hat{\theta}\right\}$. The bootstrap replicates of the mean estimator and the skewness estimator are the bootstrap replicates $\left\{\hat{\theta}_b^*\right\}$, $b = 1,...,1000$. 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.
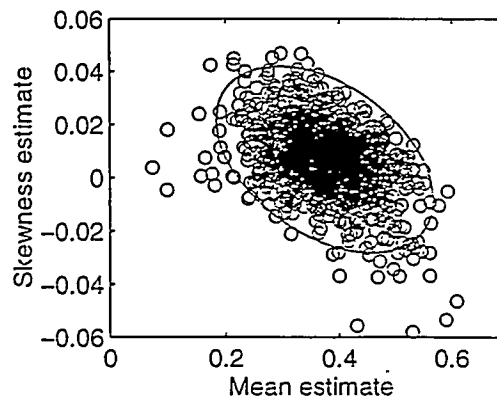

Figure 3. Joint bootstrap replicates (shown by circles) of the mean and skewness estimators based on 20 realizations of the random variable defined in Eq. (a), and the estimated 95% confidence region.

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 $\left\{\hat{\theta}_{mod}\right\}$. These model parameters must be obtained in a manner compatible with the specification of the model. For example, modal frequencies and mode shapes 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 $(1-\alpha)\times100\%$ confidence region for $\left\{\hat{\theta}\right\}$. We can use the bootstrap to accomplish this. Denote this region $R_{1-\alpha}\left(\hat{\theta}\right)$. If

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

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.

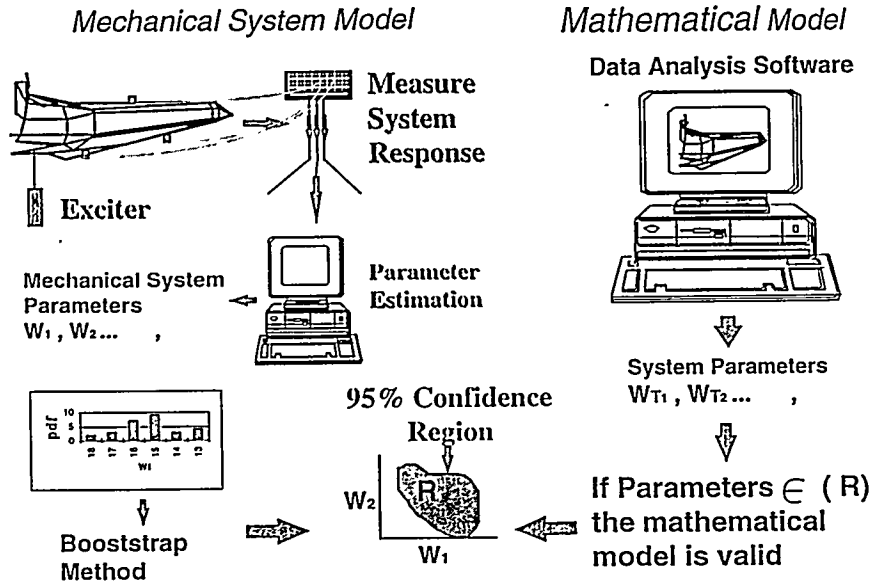## Mathematical Model Validation

Figure 4.

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, and in particular, its 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.
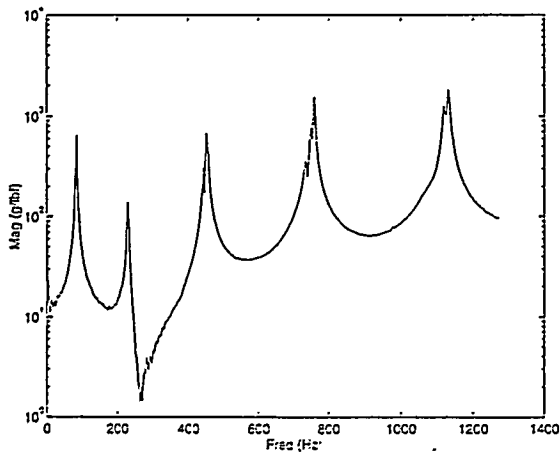

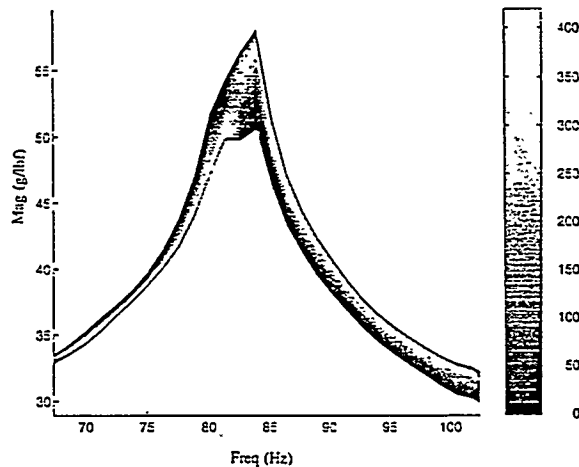Figure 5. Frequency response function of the beam used in the experiment.

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

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

$$(81.58, 83.27)\ Hz, \quad (228.12, 231.37)\ Hz, \quad (446.85, 455.30)\ Hz \tag{5}$$
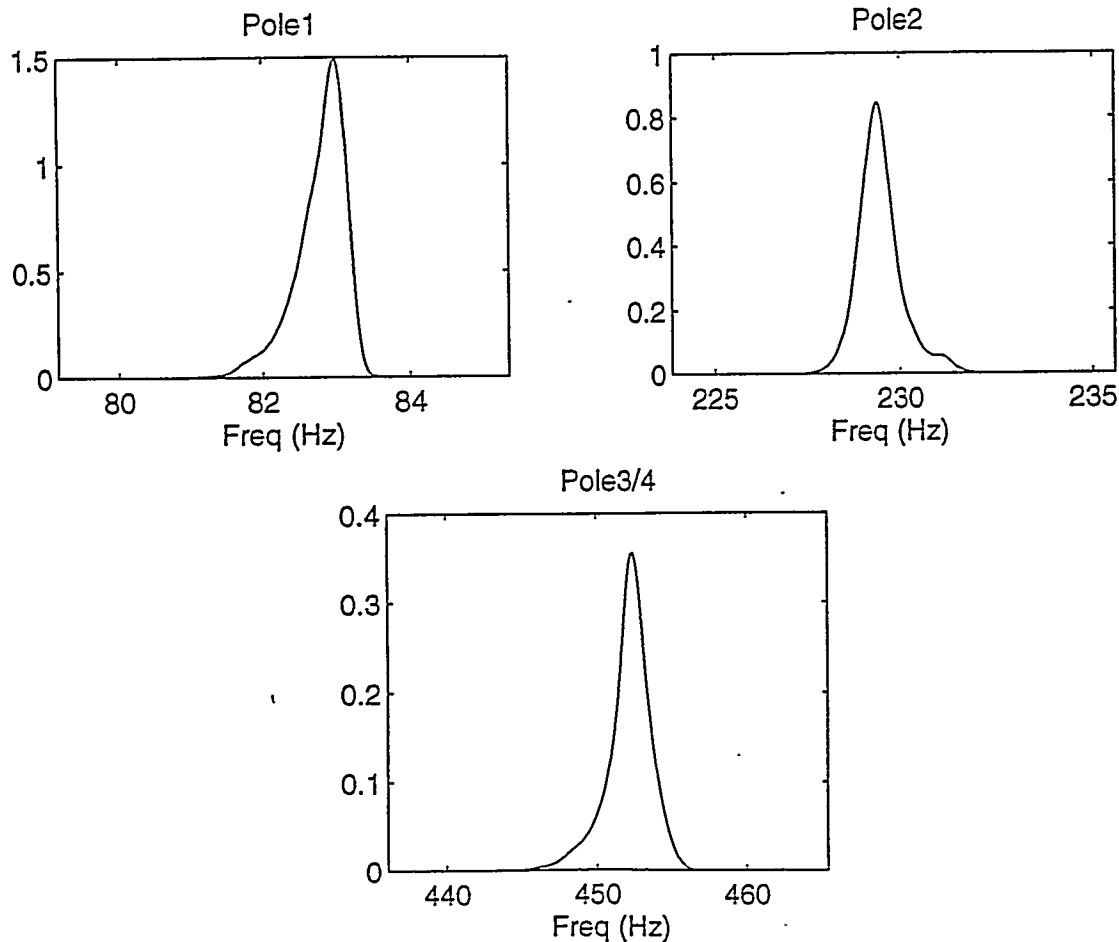

Figure 7. Kernel density estimators for the sampling distributions of the first, second, and third modal frequencies of the experimental system.

The "first cut" mathematical model of the beam used in this experiment is a finite element model created in the computer code ALGOR using three dimensional Euler 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 in developing this "first cut" model, no attempt was made to reconcile the finite element model to the experimentally measured data or to the experimental modal frequencies. However, in this particular case and most other cases, the analyst can 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.

Such a reconciliation was performed on the Euler beam model. The modification made was to move the extra mass from the middle of the beam to node location 23. For this model the first three modal frequencies are

$$82.58\ Hz,\quad 229.29\ Hz,\quad 451.88\ Hz \tag{7}$$

This model is validated with respect to all three modal frequencies at the one percent level of significance.

Based on this latter model, a 24 element Timoshenko beam element finite element model was developed with the same geometry as the Euler beam model. Its modal frequencies were analyzed, and the first three are

$$82.93\ Hz,\quad 230.93\ Hz,\quad 456.26\ Hz, \tag{8}$$

By comparison with the marginal confidence intervals of Eq. (5), this model is validated with respect to its first two modal frequencies at the one percent level of significance, and it is not validated with respect to the third modal frequency.

A third finite element model made of 24 brick elements was also developed. This model used one brick through the thickness of the beam, and each brick was one inch long. Its modal frequencies were analyzed, and the first three are

$$83.55\ Hz,\quad 233.56\ Hz,\quad 463.04\ Hz, \tag{9}$$

This model is not validated with respect to any of the modal frequencies at the one percent level of significance. Clearly refinement, and perhaps other modifications to the model, could be performed to cause the model to be validated with respect to at least some measures.

Under certain circumstances it may be desirable or necessary to simultaneously validate a pair of measures of system character, for example, modal frequencies. This can be done with, say, the second and third modal frequencies by constructing the confidence region for these frequencies, based on the bootstrap replicates of these quantities. Figure 8 shows the 99% joint confidence region for the second and third modal frequencies; it was constructed using the elliptic parametric model for the shape of the region as in the example above. The circles in the graph represent joint realizations of bootstrap replicates of the second and third modal frequencies. The solid line is the boundary of the joint confidence region. It is clear from Figure 8 that the Euler beam model is validated with respect to the joint behavior of its second and third modal frequencies; the frequencies jointly lie well within the ellipse. The Timoshenko beam model is also validated with respect to the joint behavior of its second and third modal frequencies; the frequencies jointly lie just within the upper right boundary of the 99% joint confidence region. Why is the joint behavior of the Timoshenko beam model validated when the marginal value of the third modal frequency is not? This occurs because the third modal frequency from the model is near the edge of the bootstrap replicates, and we have chosen to use the elliptic parametric model for the boundary of the confidence region. Though the ellipse appears fairly accurate, it may not quite be the correct model. This result points to the need for care in the specification of the boundary of the confidence region.
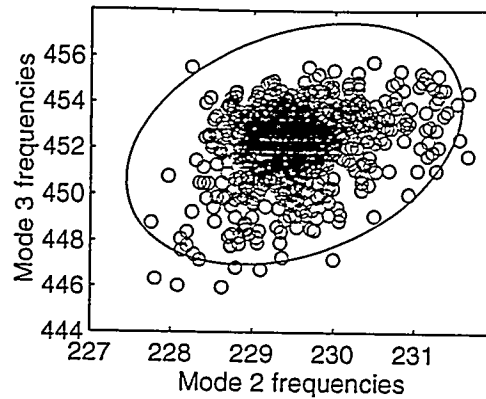
Figure 8. Bootstrap replicates (circles) and 99% joint confidence region (solid line) of the second and third modal frequencies of the experimental structure.

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. It can also 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

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.

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.