LA-UR- 02-2787

Title: BAYESIAN MODEL SCREENING FOR THE IDENTIFICATION OF NON-LINEAR MECHANICAL STRUCTURES (U)

Author(s): Gaëtan Kerschen, University de Liège, Belgium

Jean-Claude Golinval, University de Liège, Belgium

François M. Hemez, Los Alamos National Laboratory, ESA-WR

Submitted to: ASME Journal of Vibration and Acoustics,
Vol. 125, July 2003, pp. 389-397

Bayesian Model Screening for the Identification of Nonlinear Mechanical Structures

Gaëtan Kerschen
e-mail: g.kerschen@ulg.ac.be

Jean-Claude Golinval

Vibrations & Identification des Structures,
Department of Aerospace,
Mechanics and Materials,
University of Liège,
Chemin des Chevreuils 1 (B52),
B-4000 Liege, Belgium

François M. Hemez
Engineering Science & Applications Division,
ESA-WR, Mail Stop P946,
Los Alamos National Laboratory,
Los Alamos, New Mexico 87545
e-mail: hemez@lanl.gov

The development of techniques for identification and updating of nonlinear mechanical structures has received increasing attention in recent years. In practical situations, there is not necessarily a priori knowledge about the nonlinearity. This suggests the need for strategies that allow inference of useful information from the data. The present study proposes an algorithm based on a Bayesian inference approach for giving insight into the form of the nonlinearity. A family of parametric models is defined to represent the nonlinear response of a system and the selection algorithm estimates the likelihood that each member of the family is appropriate. The (unknown) probability density function of the family of models is explored using a simple variant of the Markov Chain Monte Carlo sampling technique. This technique offers the advantage that the nature of the underlying statistical distribution need not be assumed a priori. Enough samples are drawn to guarantee that the empirical distribution approximates the true but unknown distribution to the desired level of accuracy. It provides an indication of which models are the most appropriate to represent the nonlinearity and their respective goodness-of-fit to the data. The methodology is illustrated using two examples, one of which comes from experimental data. [DOI: 10.1115/1.1569947]

1 Introduction

The importance of diagnosing, identifying and modelling nonlinearity has been recognized for a long time, e.g., for the design of shock absorbers and engine mounts. The identification of nonlinear systems began in 1979 with the introduction of the restoring force surface (RFS) method by Masri and Caughey [1]. An equivalent method, referred to as force-state mapping, was proposed independently by Crawley, Aubert and O'Donnel [2,3]. Since then, numerous methods were proposed. It is not our intention to review all the methods available but rather to cite the most popular techniques that have been considered during the last twenty years.

The first application of the Hilbert transform was made in the frequency domain [4]. The time-domain Hilbert transform was also utilized to solve an inverse problem [5,6]. The use of the Volterra series in the field of structural dynamics began in the late 1980s [7]. NARMAX models consist of polynomials that include various linear and nonlinear terms combining the inputs, outputs and past errors and were introduced by Leontaritis and Billings [8,9]. Another area of signal processing that has gained importance in studying nonlinear systems deals with higher-order spectra [10,11]. These are a natural extension of the ordinary linear spectral analysis. For a detailed description of all these techniques, the reader is referred to reference [12].

The development of frequency response function-based approaches has received increasing attention in recent years. The reverse path technique has been proposed by Rice and Fitzpatrick [13] and applied to simulated and experimental data [14,15]. The conditioned reverse path formulation [16] extends the application of the reverse path algorithm to systems characterized by nonlinearities away from the location of the applied force. This method exploits the spectral conditioning techniques introduced by Bendat [17]. A related series of papers by Adams and Allemang also develop the frequency response function-based approaches [18,19].

Finally, it is worth pointing out that there has also been a growth in interest in a particular class of identification techniques based on a finite element model and referred to as finite element model updating techniques [20–22].

The problem of variable selection is one of the common issues in the field of identification of nonlinear systems. The purpose is to model the relationship between the response variable of interest and a subset of predictor variables, possibly with interactions between these latter variables. Generally speaking, there is uncertainty about which subset to use.

A possible means of determining which variables should be included in the model is through least-squares parameter estimation and the use of the significance factor [12]. Cumulative and multiple coherence functions may also be used in conjunction with the conditioned reverse path formulation [23]. The present study investigates an inference technique based on the Bayesian definition of probability—as opposed to the frequentist's point-of-view—for identifying promising subsets of predictors [24–27]. While the frequentist interpretation defines probability strictly as the number of occurrences of an event among a collective of possibilities, the Bayesian approach defines probability as the subjective opinion of the analyst or expert. To stress the difference between the two approaches, consider the simple question ''What is the probability of life on the planet Mars?'' Such question makes no sense in the frequentist framework because observations can obviously not be obtained from a collective of planets similar to the planet Mars. Similarly and even though we might not always be aware of it, many problems occur in structural dynamics that require probability to be defined in terms of our a priori knowledge of the phenomenon studied. The identification of the form of a model is one such problem addressed in this work.

The procedure developed in this work exploits ''priors''—that is, a probability structure that reflects the analyst's a priori opinion about the phenomenon investigated—on the variables of the regression model in order to give the list of all visited models together with their relative posterior probabilities. Models are visited according to their goodness-of-fit to the data, which, in the Bayesian framework, represents the likelihood of predicting the observed response. This implies that models well fitted to the data—that is, more likely models—are visited more often. The marginal probabilities of inclusion of single variables are also

Contributed by the Technical Committee on Vibration and Sound for publication in the JOURNAL OF VIBRATION AND ACOUSTICS. Manuscript received May 2002; Revised January 2003. Associate Editor: M. I. Friswell.

computed. To avoid the overwhelming burden of calculating the posterior probabilities of all models, a Gibbs sampler is considered to perform an efficient stochastic search of the model space. It is emphasized that the main difficulty of this inference problem is that the multi-dimensional probability density function (PDF) of the family of models must be sampled. However, this PDF is unknown, making random walk techniques such as the Markov Chain Monte Carlo (MCMC) sampling the only possible choice. Gibbs sampling has been proposed as a computationally attractive alternative to MCMC, yet, it can explore an unknown PDF [24,26].

While the principle of Bayesian inference has previously been applied to various problems in structural dynamics (e.g., References [28,29]), no attempt has been made, to the best of the authors' knowledge, to adapt the Bayes updating rule to the screening of model form during nonlinear system identification. After a brief discussion of model fitting in Section 2, the Bayesian screening algorithm for model selection is outlined in Section 3. The methodology is illustrated using two examples. Section 4 discusses a numerical simulation intended at demonstrating the overall performance of the screening method. The second example involves experimental data sets collected during the European COST-F3 program (Section 5). The numerical predictability of the identified model is finally assessed in Section 6.

2 Model Fitting

Model fitting generally refers to the calibration of model coefficients β given a sequence of points $(t_k;y_k)$ in the design space. It is assumed that a model is available:

$$y=M(\beta;t) \tag{1}$$

where t denotes the input variables, y denotes the output variables and β denotes the model's coefficients.

For clarity, the discussion will assume that the model form is polynomial-like. Nevertheless, nothing prevents the Bayesian model screening proposed in Section 3 to be applied to other functional forms. Fractional models could be considered, for example, to fit the poles and zeros of frequency response functions. Exponential models could be considered to represent the decay of propagating waves as a function of time or distance. Artificial neural networks are increasingly used in a variety of applications in structural dynamics because they can, depending on their form, approximate any non-linear function [30].

Another notion that must be clarified before proceeding with the discussion is the notion of ''effect.'' The model shown in Eq. (1) depends on inputs t where t does not necessarily refer to time. Functions of the input variables t can be defined that will be referred to as effects and denoted by x in the following. Such functions can assume any form, linear or non-linear. For example, the 2-input, 1-output nonlinear model:

$$y=0.3t_1+2.0\sin(t_2)-1.5e^{-t_1t_2} \tag{2}$$

can equivalently be defined through the three effects $x_1=t_1$, $x_2=\sin(t_2)$ and $x_3=e^{-t_1t_2}$ as:

$$y=0.3x_1+2.0x_2-1.5x_3 \tag{3}$$

While the input variables t_1 and t_2 might be independent, note that the effects x_1, x_2 and x_3 are neither independent nor uncorrelated. The Bayesian model screening discussed in Section 3 does not require the effects to be independent or uncorrelated. With the definition of effects x that can be functions of the input variables t, the polynomial-like model can be simply represented as:

$$y=\sum_{k=1\ldots m} x_k\beta_k=x^T\beta \tag{4}$$

The commonly encountered method of fitting the coefficients β is to define an objective function that represents the prediction error and minimize it. The most straightforward choice is to adopt the Euclidean norm of the prediction error $e_k=y_k-x_k^T\beta$:

$$J(\beta)=\sum_{k=1\ldots N} e_k^Te_k=e^Te \tag{5}$$

in which case the best, linear, unbiased estimator of the coefficients β is provided by:

$$\hat{\beta}=(X^TX)^{-1}X^Ty \tag{6}$$

where the column-vector y collects N observations and the N rows by m columns matrix X evaluates the m effects for each of the N observations:

$$y=\begin{Bmatrix} y_1\\ y_2\\ \vdots\\ y_N\end{Bmatrix};X=\begin{bmatrix} x_{1,1} & x_{1,2} & \cdots & x_{1,m}\\ x_{2,1} & x_{2,2} & \cdots & x_{2,m}\\ \vdots & \vdots & \ddots & \vdots\\ x_{N,1} & x_{N,2} & \cdots & x_{N,m}\end{bmatrix} \tag{7}$$

Clearly, other objective functions yield different estimators. The generalization of the objective function (5) is commonly referred to as the generalized least-squares (GLS) estimator [31]. Weighting matrices are introduced and a regularization term penalizes solutions too distant from the user-defined starting point β_o. Eqs. (8) and (9) show the GLS objective and the corresponding GLS estimator, respectively:

$$J(\beta)=e^TW_{ee}^{-1}e+(\beta-\beta_o)^TW_{bb}^{-1}(\beta-\beta_o) \tag{8}$$

$$\hat{\beta}=(X^TW_{ee}^{-1}X+W_{bb}^{-1})^{-1}X^TW_{ee}^{-1}y \tag{9}$$

In general, weighting matrices are chosen arbitrarily or based on experience, for example, to weight the importance of some observations more than others. When covariance matrices are used, the GLS estimator becomes similar to the Bayesian estimator. Rigorously speaking, other factors should appear in the definition of the Bayesian objective function. Because these additional factors are constant, however, the same estimator as the one shown in Eq. (9) is obtained. An important benefit of Bayesian inference is that it provides a posterior estimate of the covariance matrix:

$$\hat{W}_{bb}^{(\text{posterior})}=(W_{bb}^{-1}+X^TW_{ee}^{-1}X)^{-1} \tag{10}$$

Correlation coefficients of the posterior covariance matrix (10) provide insight into the quality of the estimator. Reference [32] discusses a shock propagation application where significant posterior correlation is obtained between coefficients that have no physical reason to be correlated. The authors conclude that the form of the model is inappropriate. They further demonstrate that it is indeed the case when improved goodness-of-fit and posterior correlation indicators are obtained with a different model.

With the exception of investigating the posterior correlation, however, no practical tool is available to select the appropriate form of a nonlinear model, which is the process we refer to a model screening. Model form—for example, replacing a linear contribution by a cubic stiffness—is usually selected based on experience or empirical observation. Sometimes, several choices seem equally likely and the analyst has to go through the painstaking process of fitting each model and assessing their goodness-of-fit. Because it is based on the concept of goodness-of-fit, such approach leads to over-fitting.

Another subtle but important issue is to estimate the posterior probability of a particular model as opposed to simply relying on the goodness-of-fit. By definition, the posterior probability is conditioned on the evidence available—that is, experimental observations. Posterior probability and goodness-of-fit complement each other because the former indicates if the analyst's prior opinion of the form of the model is consistent with the evidence. In Section 3, a practical tool is proposed for model screening based on the concept of posterior probability.

3 Bayesian Model Screening

In the previous section, the state of the practice in model fitting has been briefly overviewed. Polynomial models have been considered for simplicity. It is emphasized that the Bayesian model screening technique proposed here applies to any model, no matter which functional form it takes. Essentially, the only two general assumptions made are as follows. First, a model $y=M(\beta;x)$ must be available. As previously mentioned, the effects x can be linear or nonlinear functions of the input variables t. Second, an inference is available for calibrating the coefficients β. The inference procedure is usually referred to as ''best-fitting'' with polynomials and ''training'' with neural networks.

Model screening consists in identifying the most probable models based on a family of models defined by the user and reference data that the model's predictions must reproduce with the highest possible fidelity. It is emphasized that model screening does not necessarily identify the best model but rather ranks potential models according to their posterior probability of occurrence.

The procedure starts by, first, defining a family of models. This is achieved by defining various effects x_i and how these effects are allowed to interact to form the population of potential models. Figure 1 illustrates the concept of a family of models by showing two effects x_1 and x_2 that interact with each other. The model-forming rule illustrated in Fig. 1 is that linear and quadratic interactions are allowed between the effects x_1 and x_2. The horizontal plane represents the family of all potential models that must be explored. The vertical dimension represents the likelihood that a particular model is appropriate to represent the data. It is this notion of likelihood that will be employed to guide the search for the most appropriate models. Figure 1 illustrates a hypothetical situation where the model shown with a star symbol, $y=\beta_o+\beta_1x_2+\beta_2x_1^2+\beta_3x_2x_1^2$, is the maximum likelihood model.

The second step of the procedure is to assign the prior probability of occurrence of each effect x_i. The priors can reflect empirical observations, experience or the analyst's knowledge of the system investigated. In the application discussed in Section 5, for example, no specific knowledge of the system can be used to guide a pertinent choice of priors. Probabilities of occurrence are therefore set to a uniform 25% level for all effects.

The next step is to let the Bayesian screening method find the most appropriate models among all possible combinations of effects. To identify the most probable models a measure of goodness-of-fit to the reference data must be defined. This can be assessed using a conventional root mean square (RMS) error between data and predictions. Assuming Gaussian distributions, the RMS error becomes proportional to the likelihood function $L(y|\beta)$ that estimates the likelihood that the model is appropriate given the available data:

$$L(y|\beta)=\sum_{k=1\ldots N}(y_k-x_k^T\beta)^2 \tag{11}$$

Note that the likelihood function (11) is similar to Eq. (5) previously discussed. Other functions can be used, in particular the Bayesian objective (8), as well as the many objective functions commonly used in test-analysis correlation and model updating [33].

Once the likelihood of a particular model has been estimated, the posterior probabilities of the model's effects can be updated according to the Bayes Theorem that states that the posterior probability PDF$(\beta|y)$ is equal to the likelihood function $L(y|\beta)$ multiplied by the prior probability PDF(β) and divided by the probability of the data PDF(y):

$$\mathrm{PDF}(\beta|y)=\frac{L(y|\beta)\mathrm{PDF}(\beta)}{\mathrm{PDF}(y)} \tag{12}$$

The probability of the observed data PDF(y) is generally kept constant and omitted in the updating Eq. (12). Because the procedure is iterative in nature, the Bayes update (12) is repeated and posteriors of the nth iteration become the priors of the $(n+1)$th iteration. All models visited are kept in memory and, once enough samples have been drawn, the probability of occurrence of each model is estimated by the frequency of occurrence—that is, the ratio between the number of times each model is visited and the total number of models visited. The iterative procedure is summarized in Fig. 2.

In summary, Bayesian model screening provides the probabilities of occurrence of the most appropriate members of a user-

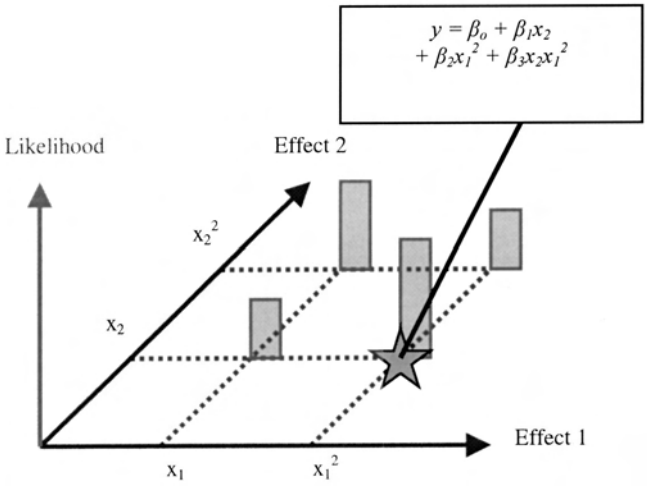

Fig. 1 Concept of "family" of models

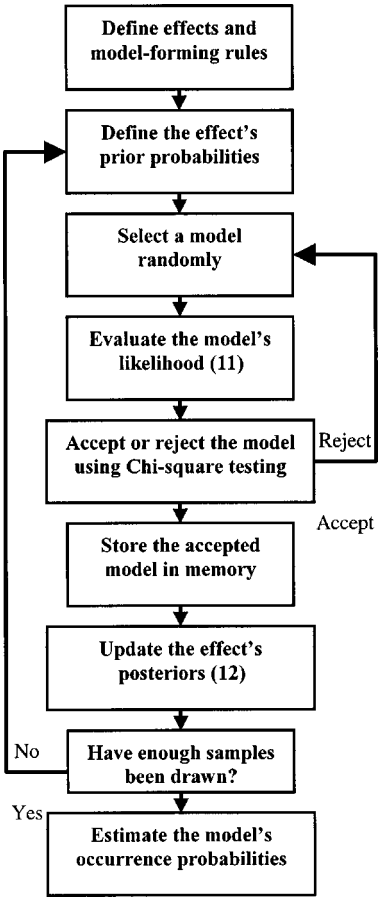

Fig. 2 Simplified flow chart of the Bayesian model screening algorithm

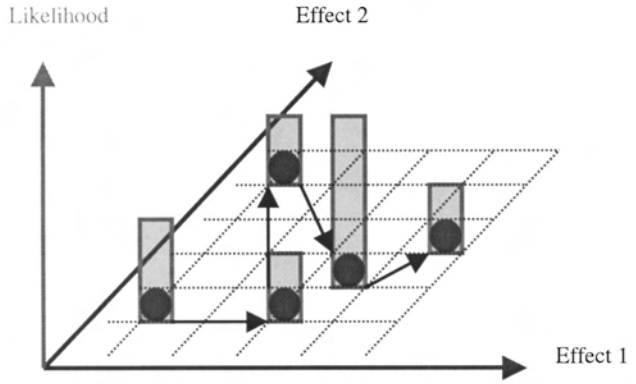

Fig. 3 Concept of random walk optimization

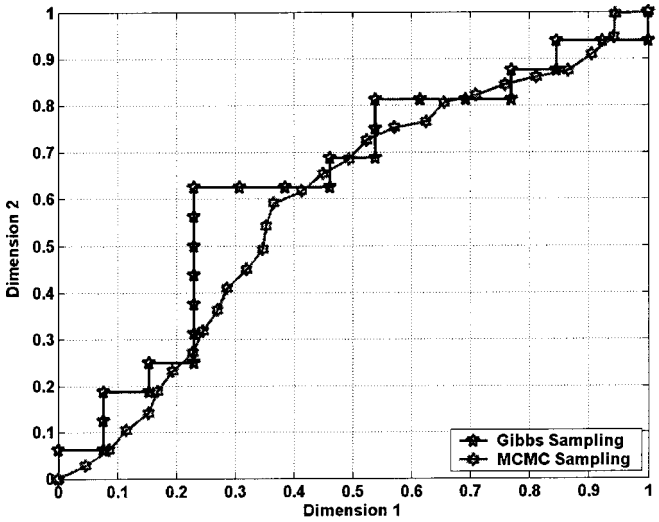

Fig. 4 Illustration of MCMC and Gibbs random walk sampling strategies

defined family of models, their goodness-of-fit indicators and the posterior probabilities PDF($\beta|y$) of effects involved in the most likely models.

To do so, however, the unknown posterior probability function must be sampled. The problem of exploring an unknown PDF is solved with the Markov Chain Monte Carlo algorithm. The MCMC sampling is advantageous in this situation because it can sample any distribution, whether it is Gaussian or not. The MCMC sampling can be viewed conceptually as an optimization solver that performs a random walk through the optimization space. This concept is illustrated in Fig. 3 where points in the optimization space are sequentially visited. More appropriate solutions are guaranteed more frequent visits because the acceptance criterion of a given solution is based on its likelihood function.

Each candidate point in the design space—here, the design space is the horizontal plane of potential models illustrated in Figs. 1 and 3—is accepted or rejected based on its value of the likelihood function (11) and a Chi-square test. This particular acceptance criterion implies that inappropriate models have a small chance of being accepted just like appropriate models have a small chance of rejection. If rejected, a new point is randomly selected in the neighborhood of the last accepted point. The sequence of points accepted is stored to estimate, once the process has been completed, the probability of occurrence of each model.

The sampling procedure used in this work is the Gibbs sampling, the simplest of the many variants of the MCMC algorithm. The main difference between the two is that the Gibbs algorithm samples one direction of the design space at a time, which makes for simpler numerical implementation. Figure 4 illustrates the difference between MCMC and Gibbs sampling. It pictures two random walks from the lower left corner ($x=0;y=0$) to the upper right corner ($x=1;y=1$). A constraint is enforced that prevents the 30 points drawn in both sequences from being repeated and from moving backwards. Pentagram symbols show a sequence of Gibbs samples while hexagram symbols picture a realization of the MCMC chain. In the former case, the solution is advanced in one direction at a time whereas the MCMC chain randomly advances the solution in the two dimensions simultaneously.

4 Numerical Application

The first application presented is extremely simple and aims at illustrating the overall performance of the model screening procedure. Consider an output variable y defined by the following input-output model:

$$y=2\sin(2t)+3\cos(t)-1.5\sin(3t)\cos(2t) \tag{13}$$

where t is an input variable that varies from zero to fifty with increments of $\Delta t=0.05$. It is assumed that the model form shown in Eq. (13) is unknown. Instead, observations $y_k=y(k\Delta t)$, for $k=0\ldots100$, are obtained and the problem consists in identifying the numerical model that best matches the observed data. It is emphasized that, in this numerical simulation, no actual experiment is performed. The continuous solution (13) is shown in Fig. 5 with a solid line. The hexagram symbols represent the discrete samples assumed to be collected.

Next, consider a set of candidate predictors:

$$\begin{cases} x_1=\sin(t) \\ x_2=\cos(t) \\ x_3=\sin(2t) \\ x_4=\cos(2t) \\ x_5=\sin(3t) \\ x_6=\cos(3t) \end{cases} \tag{14}$$

In addition to the six predictors of Eq. (14), six other predictors labeled x_7, x_8, x_9, x_{10}, x_{11} and x_{12} are defined as random functions. It can be observed that, if the functional form of the output variable y were known, it could be written as:

$$y=3x_2+2x_3-1.5x_4x_5 \tag{15}$$

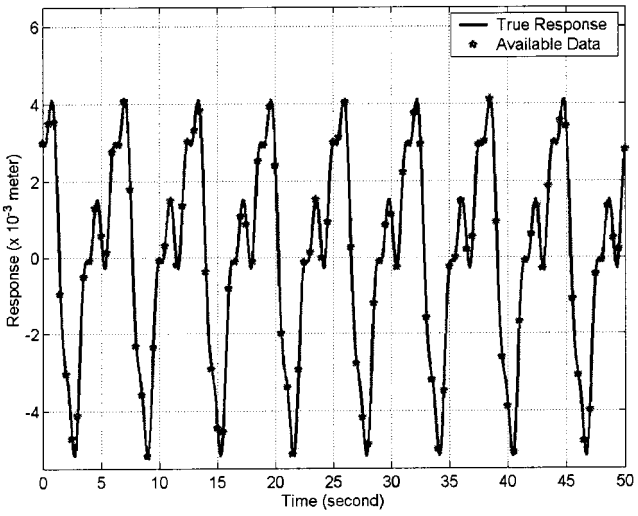

Fig. 5 Simulated non-linear function (13)

Table 1 Top five models and number of appearances

Model	Effects	Posterior Probability
1	x_2, x_3, x_4x_5	52.0%
2	x_2, x_3, x_4x_5, x_{10}	3.0%
3	x_2, x_3, x_4x_5, x_3x_4	2.0%
4	x_2, x_3, x_4x_5, x_2x_{10}	2.0%
5	x_2, x_3, x_4x_5, x_2x_5	2.0%

Clearly, y does not depend on predictors x_1, x_6, x_7, x_8, x_9, x_{10}, x_{11} and x_{12}. The objective of model screening is to identify the model form (15). Equivalently, it can be stated that the objective of model screening is to identify the linear effects x_2, x_3 and the linear interaction effect x_4x_5 from all the potential combinations defined by the family of models considered.

The family of models defined for this illustration is composed of the linear models that include the twelve linear effects x_i and the linear interaction models, defined as the previous models augmented with the 66 interaction effects x_ix_j. The total number of different effects x_i and x_ix_j with twelve predictors is therefore equal to 78. The total number of different models that can be defined belonging to this family by combining the 78 effects is in excess of $3.02\ 10^{+23}$ models, a number that approaches the number of atoms in the known Universe. Clearly, exploring such a large number of combinations without focusing on the models of highest likelihood would not be feasible.

The procedure described in the foregoing section is applied to the data using 50 samples dedicated to the initialization of the Gibbs sampler and 100 samples for the computation. Initializing the Markov chain is referred to as ''burn-in'' and guarantees that the remainder of the chain is not biased due to a particular choice of starting point. The samples drawn during burn-in are disregarded and only the 100 samples drawn during the optimization itself are kept to estimate the final probability of occurrence of each model in the family. The top five models are listed in Table 1. It can be observed that the best model in terms of posterior model probability is the actual model. The mean-square error for the top five models is about 0.003%. This means that it is not necessary to include other terms than the ones present in the best model.

Figure 6 represents the marginal posterior probability of each effect being in a particular model. The prior probabilities—that reflect the prior knowledge—are set to 25% for each linear effect x_i; 10% for the interaction effects x_ix_j if one of the parent effect x_i or x_j is selected in the model; and 1% only for the interaction effects x_ix_j when neither x_i nor x_j are considered in the model.

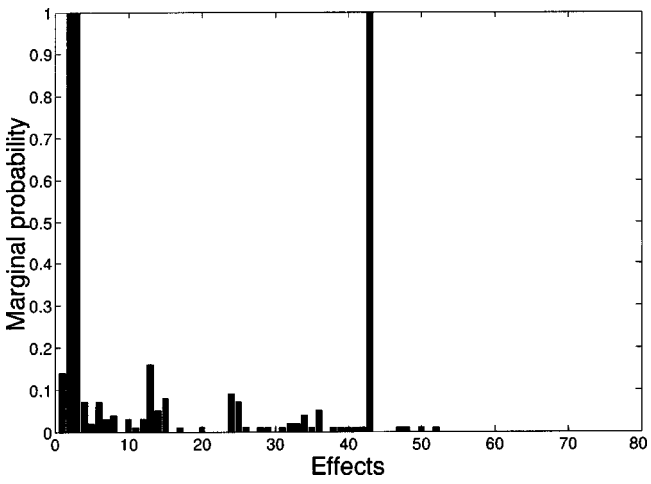

Fig. 6 Marginal posterior probability of each effect included in the family of models

These uniform probabilities reflect the fact that little is known about the form of the model before starting the analysis. It can be observed that effects 2, 3 and 43, namely x_2, x_3 and x_4x_5, are associated with a probability of 100% while the other effects may be ignored because their posterior probabilities are reduced to insignificant levels.

In conclusion, the Bayesian model screening clearly suggests a model that includes the three effects x_2, x_3 and x_4x_5. The identified coefficients corresponding to these effects are equal to 2.99, 2.02 and -1.52, respectively, and they are in good agreement with the actual coefficients shown in Eq. (15). The algorithm is implemented as interpreted Matlab™ functions and it performs the analysis in a few seconds of CPU time with a typical desktop personal computer.

5 Experimental Application

In this Section, Bayesian model screening is applied to the problem of identifying the form of a nonlinear model using real, experimental data. The analyzed data sets are chosen from those proposed by the VTT Technical Research Center of Finland within the framework of the European COST action F3 working group on ''Identification of Nonlinear Systems'' [34].

The structure investigated consists of wire rope isolators mounted between the load mass and the base mass, as shown in Fig. 7. The load mass acts like a free inertial mass. The motion and forces experienced by the isolators are measured. In particular, the acceleration responses $\ddot{x}_2$ and $\ddot{x}_{1b}$ of the load mass and bottom plate, the applied force f and the relative displacement x_{12} between the top and bottom plates are measured. The excitation produced by an electro-dynamic shaker corresponds to a white noise sequence, low-pass filtered at 400 Hertz. What makes this system interesting for identification is that the attenuation of the vibration across the interface is difficult to characterize because the mechanics of the isolators is unknown to a large extent. Significant nonlinear dynamics are expected due to the geometrical nonlinearity—pre-loading in the wire rope isolators changes with the load mass.

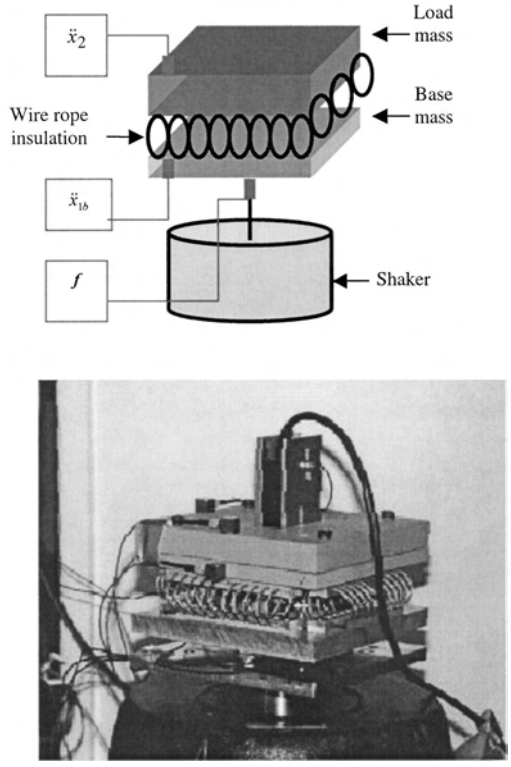

Fig. 7 Wire rope isolators

Table 2 Testing matrix of the VTT benchmark

Forcing Level	Mass 1 (2.2 kg)	Mass 2 (5.8 kg)
Level 1 (0.5 volt)	Test 1	—
Level 2 (2.0 volt)	Test 2	—
Level 3 (4.0 volt)	Test 3	Test 5
Level 4 (8.0 volt)	Test 4	—

Four excitation levels are considered ranging from 0.5 Volt up to 8.0 Volt. A nominal series of four tests are performed with a load mass of 2.2 kg. A fifth test is also carried out with the heavier load mass of 5.8 kg. Table 2 defines the testing matrix from which data sets have been collected.

Reference [35] discusses the identification of the VTT benchmark structure using the RFS. The main idea behind the RFS method is briefly overviewed to explain the system identification approach and the reader is referred to Reference [35] for more details.

The derivation of the main equations of the RFS method starts by writing Newton's second law for the load mass m_2, which yields:

$$m_2\ddot{x}_2 + f_{NL}(x_2 - x_{1b}, \dot{x}_2 - \dot{x}_{1b}) = 0 \tag{16}$$

where f_{NL} denotes the nonlinear internal force. Clearly, the force f_{NL} is unknown but it can be ascertained, as shown in Eq. (16), that its value depends on the displacement and velocity of the load mass relative to those of the bottom plate. Introducing the relative displacement, $x_{12} = x_2 - x_{1b}$, Eq. (16) becomes:

$$f_{NL}(x_{12}, \dot{x}_{12}) = -m_2\ddot{x}_{1b} - m_2\ddot{x}_{12} \tag{17}$$

Equation (17) can be viewed as describing the response of a SDOF system subjected to a base acceleration. Because the acceleration signals shown in the right-hand side of Eq. (17) are measured and the mass is known, it is possible to compute the restoring force f_{NL} at each instant from Eq. (17).

The value of the restoring force is shown in Fig. 8 in the four cases where the load mass is equal to 2.2 kg (Tests 1–4). At low excitation level, the system's behavior is predominantly linear because the restoring force varies linearly with the displacement, as can be observed for the 0.5 Volt and 2.0 Volt levels. As the excitation level is increased, a softening stiffness nonlinearity appears, as can be observed from the 4.0 Volt and 8.0 Volt levels.

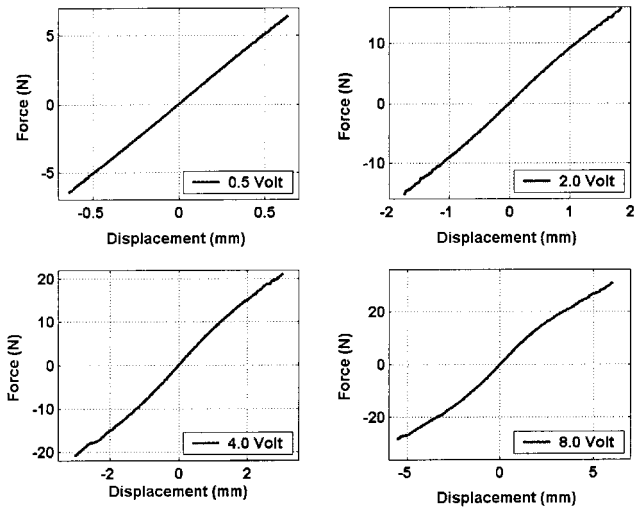

Fig. 8 Estimation of the restoring force at the four levels 0.5 Volt, 2.0 Volt, 4.0 Volt and 8.0 Volt

The next step of the RFS method is to describe the restoring force by means of a mathematical model. This is achieved through model fitting such as described in Section 2. The generic form of models sought is usually given by:

$$f(x, \dot{x}) = \sum_{i=0}^{m} \sum_{j=0}^{n} \alpha_{ij} x^i \dot{x}^j \tag{18}$$

To resolve the problem of order determination, which refers to the identification of the most appropriate dimensions m, n in Eq. (18), an over-determined system of linear equations is formed with the available restoring force data. The singular value decomposition is then used to select the appropriate order. Reference [35] details the identification procedure and shows that the final model includes a linear stiffness term, a viscous damping term and a nonlinear stiffness contribution:

$$f_{NL}(x_{12}, \dot{x}_{12}) = k_l x_{12} + c_l \dot{x}_{12} + k_{nl} |x_{12}|^{\alpha} \text{sign}(x_{12}) \tag{19}$$

where the coefficients k_l, c_l, k_{nl} and α identified with the RFS method and singular value decomposition are listed in Table 3. These results are used in the remainder as the reference through which the performance of the Bayesian model screening is assessed.

The final model features a mean square error (MSE) equal to 2.11%, which indicates an excellent correlation to test data. The MSE indicator is a normalized metric that measures the goodness-of-fit between model predictions and physical observations. It is defined as:

$$\text{MSE} = \frac{100}{N\sigma_y^2} \sum_{k=1 \ldots N} (y_k - x_k^T \hat{\beta})^2 \tag{20}$$

where, to comply with notations introduced in Section 2, y_k represents the available restoring force data and σ_y is the standard deviation of data y_k. The vector β collects the coefficients k_l, c_l and k_{nl}, assuming that the exponent α is known and equal to 1.5, and the vector x_k collects the corresponding effects in Eq. (19).

The exercise of identifying the most appropriate model form is now repeated with the Bayesian model screening. First, three effects are defined in agreement with Eq. (19). They are the linear stiffness x_{12}, linear damping $\dot{x}_{12}$ and nonlinear stiffness $|x_{12}|^{\alpha}\text{sign}(x_{12})$. The corresponding coefficients are denoted by k_l, c_l and k_{nl}, as before. Second, model-forming rules are defined which are that main effects and linear interactions between the main effects are allowed. This means that a total of six effects leading to sixty two different model forms are allowed. Such combinatorial complexity is trivial compared to the example discussed in Section 4. The complexity here stems from the fact that real data sets are analyzed with all the risk of erroneous identification caused by "noisy" measurements and signal conditioning issues.

Because the exponent α is unknown, the Bayesian model screening is repeated for several assumed values of α. The value that leads to the smallest MSE is retained. Repeating model screening could become CPU-time intensive if long MCMC chains are requested for each analysis. For this application, an initial chain of length 50 is dedicated to burn-in and a chain of length 300 is requested for the optimization. It has been verified that requesting more samples does not improve the quality of the final results. Figure 9 shows the evolution of the MSE as a function of the exponent α. The minimum value is obtained for α

Table 3 RFS identification of Eq. (19)

Coefficient	Value	Units
k_l	$1.09\ 10^{+6}$	N/m
c_l	183.44	N.sec/m
k_{nl}	$-8.52\ 10^{+7}$	$\text{N/m}^{1.5}$
α	1.5	Unitless

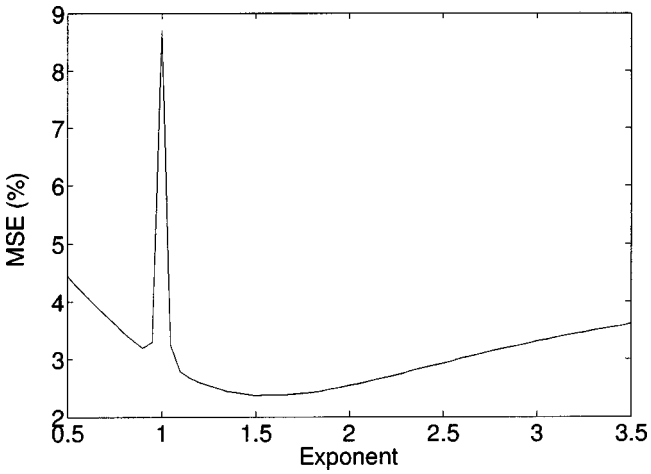

Fig. 9 Evolution of the MSE as a function of the non-linearity exponent α

$=1.5$, the same value as the one identified in Reference [35] with the RFS method and singular value decomposition. It may seem paradoxical that the MSE greatly increases for $\alpha=1$, i.e., for a linear model while its value remains low in the neighborhood of $\alpha=1$ (e.g., $\alpha=0.99$). The reason is that there is still a slight curvature for values of α different from 1 that can be enhanced by taking high values of the corresponding nonlinear parameter k_{nl}.

Table 4 and Figure 10 display the top five models and the marginal posterior probability of each effect, respectively. The mean square error for each of the top five models is around 2.37%, very similar to the RFS results previously reported. Such low MSE values indicate that the agreement with experimental data meets the expected level of accuracy. The most likely model

Table 4 Top five models and number of appearances

Model	Effects	Posterior Probability
1	$x_{12}, \dot{x}_{12}, \lvert x_{12}\rvert^{1.5}\,\mathrm{sign}(x_{12})$	86.0%
2	Model $1+x_{12}\lvert x_{12}\rvert^{1.5}\,\mathrm{sign}(x_{12})$	7.0%
3	Model $1+\dot{x}_{12}\lvert x_{12}\rvert^{1.5}\,\mathrm{sign}(x_{12})$	3.0%
4	Model $1+x_{12}\dot{x}_{12}$	2.6%
5	Model $1+x_{12}\lvert x_{12}\rvert^{1.5}\,\mathrm{sign}(x_{12})+\dot{x}_{12}\lvert x_{12}\rvert^{1.5}\,\mathrm{sign}(x_{12})$	0.3%

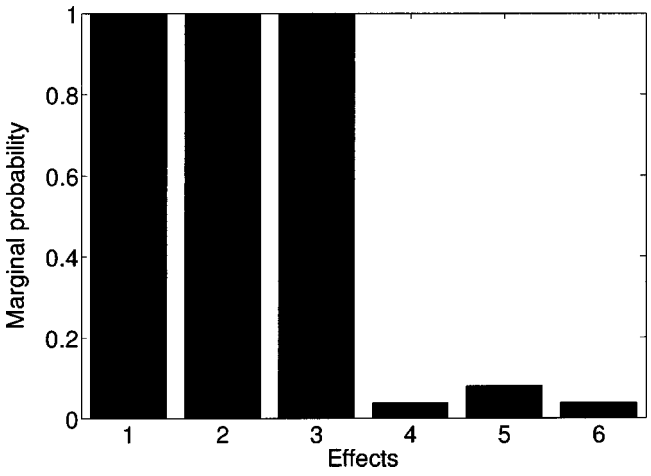

Fig. 10 Marginal posterior probability of each effect included in the family of models

Table 5 Bayesian identification of Eq. (19).

Coefficient	Value	Difference(*)
k_l	$1.12\ 10^{+6}$ N/m	2.75%
c_l	198.19 N.sec/m	8.04%
k_{nl}	$-9.07\ 10^{+7}$ N/m$^{1.5}$	6.46%
α	1.5	0.00%

(*)Difference relative to coefficients in Table 3.

in Table 4 includes only the main effects x_{12}, $\dot{x}_{12}$ and $\lvert x_{12}\rvert^{1.5}\mathrm{sign}(x_{12})$ and appears 86% of the time in the Markov chain.

The main effects, labeled 1–3 in Figure 10, have a posterior probability of 100% while interaction effects, labeled 4–6, may be considered negligible because their posterior probability is below 10%. For this application, the prior probabilities were set to a uniform 20% for the main effects; 10% for an interaction effect when at least one of the "parent" main effect is selected in the model; and 1% only for an interaction effect when none of the parents are selected. The increase in probability for effects 1–3 in Figure 10 and the reduction for effects 4–6 are therefore significant. From these results it can be concluded that a suitable model for the restoring force is given by Eq. (19) with an exponent equal to $\alpha=1.5$.

The coefficients k_l, c_l, k_{nl} and α identified with the Bayesian model screening are listed in Table 5. The last column in Table 5 compares the identification results to those of the RFS method in Reference [35]. To calibrate the model's coefficients, the Bayesian model screening currently relies on the least-squares estimator (6) even though other solvers could be implemented. Although the "true" solution is unknown, it can be stated that both methods provide consistent results because the maximum difference is less than 10%.

The small differences witnessed between the RFS identification and Bayesian model screening may be attributed to the different data sets used. Referring to Table 2, the RFS identification was conducted using the five combinations of input levels and load masses (Tests 1–5). The Bayesian model screening is restricted to four of the five cases, as discussed in Section 6, to provide a validation of the model's predictive accuracy.

6 Validation of the Identified Model

It was pointed out previously that only four of the five data sets are considered during the identification. These are Tests 1, 2, 4 and 5. The remaining data set (Test 3, 2.2 kg load mass, 4.0 Volt level) is exploited to assess the predictive accuracy of the identified model.

Although it might not yet be the state of the practice in structural system identification, many authors, among whom we cite a recent discussion of model validation in Reference [36], have emphasized that identified models should be independently validated. It essentially means that independent experiments or data sets should be used for model screening and parametric calibration, on one hand, and model validation and predictive accuracy assessment, on the other hand. The predictive accuracy of a model cannot be objectively assessed over the operational range of interest as long as the independence between training data and validation data is not met.

Here, data sets collected during Tests 1, 2, 4 and 5 are used for model screening and system identification while the data collected during Test 4 are used for model validation. Essentially, Eq. (19) is evaluated with the coefficients of Table 5 to predict the restoring force. Displacement and velocity time series in Eq. (19) are estimated from numerical integration of the measured acceleration signals. The evolution of the predicted restoring force versus time is then compared to the "true" restoring force measured during Test 3. The true restoring force is estimated directly from acceleration measurements, as shown in Eq. (17). In Fig. 11, the two

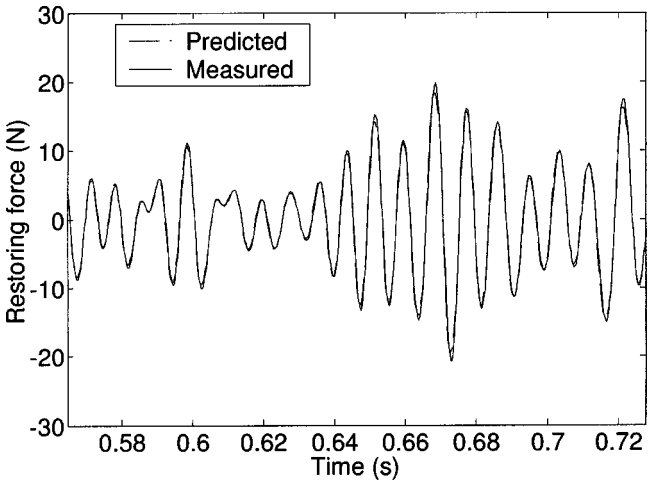

Fig. 11 Comparison between measured and predicted restoring forces (4 Volt level)

time series are compared. It can be observed that the measured and predicted forces agree to the point where the difference between signals is not visible.

The MSE, that assesses the overall fidelity between measurements and model predictions, is equal to 0.77% when the restoring force of Test 3 is predicted based on the identified Eq. (19). Based on the low MSE values obtained during model identification (2.37%) and validation (0.77%), it can be stated that the prediction accuracy seems to be within 97%. This means that a prediction should be within 3% of a measurement, should a hypothetical experiment be conducted. Clearly, the main drawback of our assessment of predictive accuracy is that it is based on a single analysis. Other validation data sets would be required to reach a more quantitative statement of accuracy. Nevertheless, this analysis leads to the conclusion that a reliable identification has been performed over the operational range of interest, that is, within the ranges of 2.2-to-5.8 kg load mass and 0.5-to-8.0 Volt excitation level.

7 Conclusion

Model selection is one of the common issues in the field of identification of nonlinear systems. A Bayesian inference approach for giving insight into the form of the nonlinearity has been proposed in this paper. The key advantage of the method is that a collection of potential models together with their posterior probability is obtained instead of the single best model. It allows for more flexibility in deciding the most appropriate model of the non-linearity. In addition, the marginal posterior probability of each effect being in a particular model may also be evaluated.

The results obtained using two illustrative examples enable us to confirm the performance and the utility of the proposed technique. We believe that Bayesian model screening will become an important addition to the structural dynamicist's toolbox.

Acknowledgments

The authors acknowledge the VTT Technical Research Center of Finland for sharing the ''Dynamic properties of resilient mounts'' benchmark data. The authors are very grateful to Mandy Cundy, technical staff member at Los Alamos National Laboratory, for helping them with the numerical implementation. Dr. Kerschen is supported by a grant from the Belgian National Fund for Scientific Research (FNRS), which is gratefully acknowledged.

References

[1] Masri, S. F., and Caughey, T. K., 1979, ''A Non-parametric Identification Technique for Non-linear Dynamic Problems,'' ASME J. Appl. Mech., **46**, pp. 433–447.
[2] Crawley, E. F., and Aubert, A. C., 1986, ''Identification of Nonlinear Structural Elements by Force-state Mapping,'' AIAA J., **24**, pp. 155–162.
[3] Crawley, E. F., and O'Donnel, 1986, ''Identification of Nonlinear System Parameters in Joints using the Force-state Mapping Technique,'' AIAA Paper 86–1013, pp. 659–667.
[4] Simon, M., and Tomlinson, G. R., 1984, ''Use of the Hilbert Transform in Modal Analysis of Linear and Non-linear Structures,'' J. Sound Vib., **96**, pp. 421–436.
[5] Feldman, M., 1994, ''Nonlinear System Vibration Analysis using the Hilbert Transform-I. Free Vibration Analysis Method 'FREEVIB,' '' Mech. Syst. Signal Process., **8**, pp. 119–127.
[6] Feldman, M., 1994, ''Nonlinear System Vibration Analysis using the Hilbert Transform-II. Forced Vibration Analysis Method 'FORCEVIB,' '' Mech. Syst. Signal Process., **8**, pp. 309–318.
[7] Gifford, S. J., and Tomlinson, G. R., 1989, ''Recent Advances in the Application of Functional Series to Non-linear Structures,'' J. Sound Vib., **135**, pp. 289–317.
[8] Leontaritis, I. J., and Billings, S. A., 1985, ''Input-Output Parametric Models for Non-linear Systems: Part I—Deterministic Non-linear Systems,'' Int. J. Control, **41**, pp. 303–328.
[9] Leontaritis, I. J., and Billings, S. A., 1985, ''Input-Output Parametric Models for Non-linear Systems: Part II—Stochastic Non-linear Systems,'' Int. J. Control, **41**, pp. 329–344.
[10] Bendat, J. S., and Piersol, A. G., 1980, *Engineering Applications of Correlation and Spectral Analysis*, John Wiley & Sons, New-York (U.S.A.).
[11] Choi, D., Chang, J., Stearman, R. O., and Powers, E. J., 1984, ''Bispectral Interaction of Nonlinear Mode Interactions,'' *Proceedings of the 2nd International Modal Analysis Conference*, Orlando, FL, pp. 602–609.
[12] Worden, K., and Tomlinson, G. R., 2001. *Non-linearity in Structural Dynamics: Detection, Identification and Modelling*, Institute of Physics Publishing, Bristol, Philadelphia (PA).
[13] Rice, H. J., and Fitzpatrick, J. A., 1991, ''A Procedure for the Identification of Linear and Non-linear Multi-degree-of-freedom Systems,'' J. Sound Vib., **149**, pp. 397–411.
[14] Rice, H. J., and Fitzpatrick, J. A., 1991, ''The Measurement of Non-linear Damping in Single-degree-of-freedom Systems,'' ASME J. Vibr. Acoust., **113**, pp. 132–140.
[15] Esmonde, H., Fitzpatrick, J. A., Rice, H. J., and Axisa, F., 1992, ''Modelling and Identification of Non-linear Squeeze Film Dynamics,'' J. Fluids Struct., **6**, pp. 223–248.
[16] Richards, C. M., and Singh, R., 1998, ''Identification of Multi-Degree-of-Freedom Non-linear Systems Under Random Excitations by the Reverse-Path Spectral Method,'' J. Sound Vib., **213**, pp. 673–708.
[17] Bendat, J. S., 1990, *Nonlinear System Analysis and Identification from Random Data*, John Wiley & Sons, New-York (U.S.A.).
[18] Adams, D. E., and Allemang, R. J., 1999, ''Characterization of Nonlinear Vibrating Systems using Internal Feeback and Frequency Response Modulation,'' ASME J. Vibr. Acoust., **121**, pp. 495–500.
[19] Adams, D. E., and Allemang, R. J., 2000, ''A Frequency Domain Method for Estimating the Parameters of a Non-Linear Structural Dynamic Model through Feedback,'' Mech. Syst. Signal Process., **14**, pp. 637–656.
[20] Hemez, F. M., and Doebling, S. W., 2001, ''Review and Assessment of Model Updating for Non-linear, Transient Dynamics,'' Mech. Syst. Signal Process., **15**, pp. 45–74.
[21] Lenaerts, V., Kerschen, G., and Golinval, J.-C., 2001, ''Proper Orthogonal Decomposition for Model Updating of Non-linear Mechanical Systems,'' Mech. Syst. Signal Process., **15**, pp. 31–43.
[22] Dippery, K. D., and Smith, S. W., 1998, ''An Optimal Control Approach to Nonlinear System Identification,'' *Proceedings of the 16th International Modal Analysis Conference*, Santa Barbara CA, pp. 637–643.
[23] Bendat, J. S., 1986, *Random Data: Analysis and Measurement Procedures*, John Wiley-Interscience, New York (NY), second edition.
[24] Chipman, H., Hamada, M., and Wu, C. F. J., 1997, ''A Bayesian Variable Selection Approach for Analyzing Designed Experiments With Complex Aliasing,'' Technometrics, **39**, pp. 372–381.
[25] Wu, C. F. J., and Hamada, M., 2000, *Experiments: Planning, Analysis, and Parameter Design Optimization*, Wiley, New-York, NY.
[26] George, E. I., and McCulloch, R. E., 1993, ''Variable Selection via Gibbs Sampling,'' Journal of the American Statistical Society, **88**, pp. 881–889.
[27] Carlin, B. P., and Chib, S., 1995, ''Bayesian Model Choice via Markov Chain Monte Carlo,'' Journal of the Royal Statistical Society Series B, **77**, pp. 473–484.
[28] Kulczycki, P., 2001, ''An Algorithm for Bayes Parameter Estimation,'' ASME J. Dyn. Syst., Meas., Control, **123**, pp. 611–614.
[29] Yuen, K. V., Katafygiotis, L. S., Papadimitriou, C., and Mickleborough, N. C., 2001, ''Optimal Sensor Placement Methodology for Identification With Unmeasured Excitation,'' ASME J. Dyn. Syst., Meas., Control, **123**, pp. 677–686.
[30] Cybenko, G., 1989, ''Approximation by Superpositions of a Sigmoidal Function,'' Mathematics of Control, Signals, and Systems, **2**, pp. 303–314.
[31] Walter, E., and Pronzato, L., 1997, *Identification of Parametric Models From Experimental Data*, Springer-Verlag, Berlin (Germany).

[32] Hasselman, T. K., Anderson, M. C., and Wenshui, G., 1998, ''Principal Components Analysis for Nonlinear Model Correlation, Updating and Uncertainty Evaluation,'' *Proceedings of the 16th International Modal Analysis Conference*, pp. 664–651.
[33] Mottershead, J. E., and Friswell, M. I., 1993, ''Model Updating in Structural Dynamics: A Survey,'' J. Sound Vib., **162**(**2**), pp. 347–375.
[34] http://www.ulg.ac.be/ltas-vis/costf3/costf3.html, portal Web site of the COST-F3 action.
[35] Kerschen, G., Lenaerts, V., Marchesiello, S., and Fasana, A., 2001, ''A Frequency Domain vs. a Time Domain Identification Technique for Non-linear Parameters Applied to Wire Rope Isolators,'' ASME J. Dyn. Syst., Meas., Control, **123**, pp. 645–650.
[36] Doebling, S. W., 2002, ''Structural Dynamics Model Validation: Pushing the Envelope,'' *Proceedings of the International Conference on Structural Dynamics Modelling—Test, Analysis, Correlation and Validation*, Madeira Island (Portugal).