

Model Selection Using Response Measurements: Bayesian Probabilistic Approach

James L. Beck¹ and Ka-Veng Yuen²

Abstract: A Bayesian probabilistic approach is presented for selecting the most plausible class of models for a structural or mechanical system within some specified set of model classes, based on system response data. The crux of the approach is to rank the classes of models based on their probabilities conditional on the response data which can be calculated based on Bayes' theorem and an asymptotic expansion for the evidence for each model class. The approach provides a quantitative expression of a principle of model parsimony or of Ockham's razor which in this context can be stated as "simpler models are to be preferred over unnecessarily complicated ones." Examples are presented to illustrate the method using a single-degree-of-freedom bilinear hysteretic system, a linear two-story frame, and a ten-story shear building, all of which are subjected to seismic excitation.

DOI: 10.1061/(ASCE)0733-9399(2004)130:2(192)

CE Database subject headings: Bayesian analysis; Model studies; Time series analysis; Probabilistic methods; Mechanical systems, structural; Measurement; Excitation.

Introduction

The problem of model identification using dynamic data from a structural or mechanical system is an active area of research in aerospace, civil, and mechanical engineering, because of its importance in model updating, active control, and health/condition monitoring (Gersch et al. 1976; Beck 1978; Hoshiya and Saito 1984; Pi and Mickleborough 1989; Hjelmstad et al. 1992; Mottershead and Friswell 1993; Beck and Katafygiotis 1998; Katafygiotis and Beck 1998; Katafygiotis et al. 1998; Quek et al. 1999; Sanayei et al. 1999; Shi et al. 2000; Pappa et al. 2000; Vanik et al. 2000).

The uncertainties in the values of the model parameters can be updated using Bayesian inference (Jeffreys 1961; Box and Tiao 1973; Sivia 1996; Jaynes 2003). A Bayesian system identification framework has been presented for both linear and nonlinear dynamic models for the case of measured input and output (Beck and Katafygiotis 1998; Katafygiotis et al. 1998; Beck and Au 2002; Yuen and Katafygiotis 2002) and for the case of output-only measurements (Katafygiotis and Yuen 2001; Yuen and Katafygiotis 2001; Yuen et al. 2002; Yuen and Beck 2003).

The usual approach in system identification is to find the best (optimal) model in a specified class of models, e.g., a class of shear building models or a class of bilinear hysteretic models. The more general problem of model class selection has not been well

explored in system identification. It is obvious that a more complicated model can "fit" the data better than a less complicated one which has fewer adjustable (uncertain) parameters. Therefore, if the optimal model class is chosen by minimizing some norm of the error between the output data and the corresponding predictions of the optimal model in each class, the optimal model class will always be the most complicated one. For example, in modal identification, using a 20-mode model would always be better than using a 10-mode model because the former would fit the data better, although the improvement might be negligible. This approach is therefore likely to lead to overfitting the data. When an overfitted model is used for future prediction, it will very likely lead to poor results because the model will depend too much on the details of the data, and the noise in the data might have an important role in the data fitting. Therefore, in model class selection, it is necessary to penalize a complicated model. This point was also noted for structural mechanics problems by Grigoriu et al. (1979).

The same point was recognized much earlier by Jeffreys, who did pioneering work on the application of Bayesian methods (Jeffreys 1961). He pointed out the need for a quantitative expression of the very old philosophy expounded by William of Ockham (or Occam in Latin) and known as "Ockham's razor," which is roughly translated from Latin as "It is vain to do with more what can be done with fewer" (Sivia 1996). In the present context, this philosophy implies that simpler models are preferable to unnecessarily complicated ones, that is, the selected class of models should agree closely with the observed behavior of the system but otherwise be as simple as possible. Box and Jenkins (1970) also emphasize the same principle when they refer to the need for parsimonious models in time-series forecasting, although they do not give a quantitative expression of their principle of parsimony. Akaike recognized that maximum likelihood estimation is insufficient for model order selection in time-series forecasting using autoregressive moving average (ARMA) models and came up with another term to be added to the logarithm of the likelihood function that penalizes parameterization of the models (Akaike

¹Professor, Division of Engineering and Applied Science, MC 104-44, California Institute of Technology, Pasadena, CA 91125 (corresponding author). E-mail: jimbeck@caltech.edu

²Assistant Professor, Department of Civil and Environmental Engineering, Univ. of Macau, Macau, China.

Note. Associate Editor: Roger G. Ghanem. Discussion open until July 1, 2004. Separate discussions must be submitted for individual papers. To extend the closing date by one month, a written request must be filed with the ASCE Managing Editor. The manuscript for this paper was submitted for review and possible publication on April 8, 2002; approved on July 8, 2003. This paper is part of the *Journal of Engineering Mechanics*, Vol. 130, No. 2, February 1, 2004. ©ASCE, ISSN 0733-9399/2004/2-192-203/\$18.00.

1974). This was later modified by Akaike (1976) and by Schwarz (1978).

In recent years, there has been a reappraisal of the work of Jeffreys on the application of Bayesian methods (Jeffreys 1961), especially due to the expository publications of Jaynes (Jaynes 1983, 2003). In particular, the Bayesian approach to model selection has been further developed by showing that the *evidence* for each model class provided by the data (that is, the probability of getting the data based on the whole model class) automatically enforces a quantitative expression of a principle of model parsimony or of Ockham's razor (Gull 1988; Mackay 1992; Sivia 1996). There is no need to introduce ad hoc penalty terms as was done in some of the earlier work on model class selection.

In this paper, the Bayesian approach is expounded and applied to select the most plausible class of dynamic models representing a structural or mechanical system (from within some specified set of model classes) by using its response measurements. In the next section, the model class selection procedure is explained. Then two Bayesian system identification techniques are introduced using input-output data and output-only data, respectively. Finally, illustrative examples are presented using a single-degree-of-freedom bilinear hysteretic system, a linear two-story frame, and a linear ten-story shear building, all of which are subjected to seismic excitation.

Model Class Selection

Let \mathcal{D} denote the input-output or output-only dynamical data from a structural or mechanical system. The goal is to use \mathcal{D} to select the most plausible class of models representing the system out of N_M given classes of models $\mathcal{M}_1, \mathcal{M}_2, \dots, \mathcal{M}_{N_M}$. Since probability may be interpreted as a measure of plausibility based on specified information (Cox 1961), the probability of a class of models conditional on the set of dynamic data \mathcal{D} is required. This can be obtained by using Bayes' theorem as follows:

$$P(\mathcal{M}_j|\mathcal{D},\mathcal{U}) = \frac{p(\mathcal{D}|\mathcal{M}_j,\mathcal{U})P(\mathcal{M}_j|\mathcal{U})}{p(\mathcal{D}|\mathcal{U})}, \quad j=1,2,\dots,N_M \quad (1)$$

where $p(\mathcal{D}|\mathcal{U}) = \sum_{j=1}^{N_M} p(\mathcal{D}|\mathcal{M}_j,\mathcal{U})P(\mathcal{M}_j|\mathcal{U})$ by the theorem of total probability and \mathcal{U} expresses the user's judgement on the initial plausibility of the model classes, expressed as a prior probability $P(\mathcal{M}_j|\mathcal{U})$ on the model classes $\mathcal{M}_j, j=1, \dots, N_M$, where $\sum_{j=1}^{N_M} P(\mathcal{M}_j|\mathcal{U}) = 1$. The factor $p(\mathcal{D}|\mathcal{M}_j,\mathcal{U})$ is called the *evidence* for the model class \mathcal{M}_j provided by the data \mathcal{D} . Note that \mathcal{U} is irrelevant in $p(\mathcal{D}|\mathcal{M}_j,\mathcal{U})$ and so it can be dropped in the notation because it is assumed that \mathcal{M}_j alone specifies the probability density function (PDF) for the data, that is, it specifies not only a class of deterministic dynamic models but also the probability descriptions for the prediction error and initial plausibility for each model in the class \mathcal{M}_j (Beck and Katafygiotis 1998). Eq. (1) shows that the most plausible model class is the one that maximizes $p(\mathcal{D}|\mathcal{M}_j)P(\mathcal{M}_j|\mathcal{U})$ with respect to j .

Note that $P(\mathcal{M}_j|\mathcal{D},\mathcal{U})$ can be used not only for selection of the most probable class of models, but also for response prediction based on all the model classes. Let u denote a quantity to be predicted, e.g., first-story drift. Then the PDF of u given the data \mathcal{D} can be calculated from the theorem of total probability as follows: $p(u|\mathcal{D},\mathcal{U}) = \sum_{j=1}^{N_M} p(u|\mathcal{D},\mathcal{M}_j)P(\mathcal{M}_j|\mathcal{D},\mathcal{U})$, rather than using only the best model for prediction. However, if the probability $P(\mathcal{M}_{\text{best}}|\mathcal{D},\mathcal{U})$ for the best model class is much larger than the probability of the others, then the above expression is approxi-

mated by $p(u|\mathcal{D},\mathcal{U}) = p(u|\mathcal{D},\mathcal{M}_{\text{best}})$, and it is sufficient to just use the best model class for predicting the system behavior.

The evidence for \mathcal{M}_j provided by the data \mathcal{D} is given by the theorem of total probability

$$p(\mathcal{D}|\mathcal{M}_j) = \int_{\Theta_j} p(\mathcal{D}|\boldsymbol{\theta}_j,\mathcal{M}_j)p(\boldsymbol{\theta}_j|\mathcal{M}_j)d\boldsymbol{\theta}_j, \quad j=1,2,\dots,N_M \quad (2)$$

where $\boldsymbol{\theta}_j$ is the parameter vector in a parameter space $\Theta_j \subset \mathbb{R}^{N_j}$ that defines each model in \mathcal{M}_j , the prior PDF $p(\boldsymbol{\theta}_j|\mathcal{M}_j)$ is specified by the user and the likelihood $p(\mathcal{D}|\boldsymbol{\theta}_j,\mathcal{M}_j)$ is evaluated using the methods presented in the next section.

In *globally identifiable* cases (Beck and Katafygiotis 1998), the updated (posterior) PDF for $\boldsymbol{\theta}_j$ given a large amount of data \mathcal{D} may be approximated accurately by a Gaussian distribution, so $p(\mathcal{D}|\mathcal{M}_j)$ can be approximated by using Laplace's method for asymptotic approximation (Papadimitriou et al. 1997)

$$p(\mathcal{D}|\mathcal{M}_j) \approx p(\mathcal{D}|\hat{\boldsymbol{\theta}}_j,\mathcal{M}_j)p(\hat{\boldsymbol{\theta}}_j|\mathcal{M}_j)(2\pi)^{N_j/2}|\mathbf{H}_j(\hat{\boldsymbol{\theta}}_j)|^{-1/2}, \quad j=1,2,\dots,N_M \quad (3)$$

where N_j is the number of uncertain parameters for the model class \mathcal{M}_j , the optimal parameter vector $\hat{\boldsymbol{\theta}}_j$ is the most probable value [it is assumed to maximize $p(\boldsymbol{\theta}_j|\mathcal{D},\mathcal{M}_j)$ in the interior of Θ_j], and $\mathbf{H}_j(\hat{\boldsymbol{\theta}}_j)$ is the Hessian matrix of $-\ln[p(\mathcal{D}|\boldsymbol{\theta}_j,\mathcal{M}_j)p(\boldsymbol{\theta}_j|\mathcal{M}_j)]$ with respect to $\boldsymbol{\theta}_j$ evaluated at $\hat{\boldsymbol{\theta}}_j$. For *unidentifiable* cases (Beck and Katafygiotis 1998), the evidence $p(\mathcal{D}|\mathcal{M}_j)$ can be calculated by using an extension of the asymptotic expansion used in Eq. (3) (Beck and Katafygiotis 1998; Katafygiotis et al. 1998) or by using a Markov chain Monte Carlo simulation technique (Beck and Au 2002) on Eq. (2). The discussion here will focus on the *globally identifiable* case.

The likelihood factor $p(\mathcal{D}|\hat{\boldsymbol{\theta}}_j,\mathcal{M}_j)$ in Eq. (3) will be higher for those model classes \mathcal{M}_j that make the probability of the data \mathcal{D} higher, that is, that give a better "fit" to the data. For example, if the likelihood function is Gaussian, then the highest value of $p(\mathcal{D}|\hat{\boldsymbol{\theta}}_j,\mathcal{M}_j)$ will be given by the model class \mathcal{M}_j that gives the smallest least-squares fit to the data. As mentioned earlier, this likelihood factor favors model classes with more uncertain parameters. If the number of data points N in \mathcal{D} is large, the likelihood factor will be the dominant one in Eq. (3) because it increases exponentially with N , while the other factors behave as N^{-1} , as shown below.

The remaining factors $p(\hat{\boldsymbol{\theta}}_j|\mathcal{M}_j)(2\pi)^{N_j/2}|\mathbf{H}_j(\hat{\boldsymbol{\theta}}_j)|^{-1/2}$ in Eq. (3) are called the *Ockham factor* by Gull (1988). The Ockham factor represents a penalty against parameterization (Gull 1988; Mackay 1992), as we demonstrate in the following discussion.

We wish to show that the Ockham factor decreases exponentially with the number of uncertain parameters in the model class. For this purpose, consider an alternative expression for it, derived as follows. It is known that for a large number N of data points in \mathcal{D} , the updated (posterior) PDF $p(\boldsymbol{\theta}_j|\mathcal{D},\mathcal{M}_j)$ is well approximated by a Gaussian PDF with mean $\hat{\boldsymbol{\theta}}_j$ and covariance matrix given by the inverse of the Hessian matrix $\mathbf{H}_j(\hat{\boldsymbol{\theta}}_j)$. The principal posterior variances for $\boldsymbol{\theta}_j$, denoted by $\sigma_{j,i}^2$ with $i=1,2,\dots,N_j$ are therefore the inverse of the eigenvalues of this Hessian matrix (Beck and Katafygiotis 1998). The determinant factor $|\mathbf{H}_j(\hat{\boldsymbol{\theta}}_j)|^{-1/2}$ in the Ockham factor can therefore be expressed as the product of all the $\sigma_{j,i}$ for $i=1,2,\dots,N_j$. Assume that the prior PDF $p(\boldsymbol{\theta}_j|\mathcal{M}_j)$ is Gaussian with mean (most probable value a priori) $\boldsymbol{\theta}_j$ and a diagonal covariance matrix with variances $\rho_{j,i}^2$ with $i=1,2,\dots,N_j$. The logarithm of the Ockham factor for the

model class \mathcal{M}_j , denoted by β_j , can therefore be expressed as

$$\beta_j = \ln[p(\hat{\theta}_j|\mathcal{M}_j)(2\pi)^{N_j/2}|\mathbf{H}_j(\hat{\theta}_j)|^{-1/2}] \approx -\sum_{i=1}^{N_j} \ln \frac{\rho_{j,i}}{\sigma_{j,i}} - \frac{1}{2} \sum_{i=1}^{N_j} \left(\frac{\hat{\theta}_{j,i} - \bar{\theta}_{j,i}}{\rho_{j,i}} \right)^2 \quad (4)$$

Since the prior variances will always be greater than the posterior variances if the data provide any information about the model parameters in the model class \mathcal{M}_j , all the terms in the first sum in Eq. (4) will be positive and so will the terms in the second sum unless the posterior most probable value $\hat{\theta}_{j,i}$ just happens to coincide with the prior most probable value $\bar{\theta}_{j,i}$. Thus, one might expect that the log Ockham factor β_j will decrease if the number of parameters N_j for the model class \mathcal{M}_j is increased. This expectation is confirmed by noting that the posterior variances are inversely proportional to the number of data points N in \mathcal{D} , so the dependence of the log Ockham factor on N is

$$\beta_j = -\frac{1}{2} N_j \ln N + R_j \quad (5)$$

where the remainder R_j depends primarily on the choice of prior PDF and is $\mathcal{O}(1)$ for large N . It is not difficult to show that this result holds for even more general forms of the prior PDF than the Gaussian PDF used here.

It follows from Bayes' theorem that we have the exact relationship

$$p(\mathcal{D}|\mathcal{M}_j) = p(\mathcal{D}|\hat{\theta}_j, \mathcal{M}_j)p(\hat{\theta}_j|\mathcal{M}_j)/p(\hat{\theta}_j|\mathcal{D}, \mathcal{M}_j) \quad (6)$$

A comparison of this equation and Eq. (3) shows that the Ockham factor is approximately equal to the ratio $p(\hat{\theta}_j|\mathcal{M}_j)/p(\hat{\theta}_j|\mathcal{D}, \mathcal{M}_j)$ which is always less than unity if the data provide any information about the model parameters in the model class \mathcal{M}_j . Indeed, for large N , the negative of the logarithm of this ratio is an asymptotic approximation of the information about θ_j provided by data \mathcal{D} (Kullback 1968). Therefore, the log Ockham factor β_j removes the amount of information about θ_j provided by \mathcal{D} from the log likelihood $\ln p(\mathcal{D}|\hat{\theta}_j, \mathcal{M}_j)$ to give the log evidence $\ln p(\mathcal{D}|\mathcal{M}_j)$.

The Ockham factor may also be interpreted as a measure of robustness of the model class \mathcal{M}_j . If the updated PDF for the model parameters for the given model class is very peaked, then the ratio $p(\hat{\theta}_j|\mathcal{M}_j)/p(\hat{\theta}_j|\mathcal{D}, \mathcal{M}_j)$, and so the Ockham factor, is very small. But a narrow peak implies that response predictions using this model class will depend too sensitively on the optimal parameters $\hat{\theta}_j$. Small errors in the parameter estimation will lead to large errors in the response predictions. Therefore, a class of models with a small Ockham factor will not be robust to noise in the data during parameter estimation, that is, during selection of the optimal model within the class.

To summarize, in the Bayesian approach to model selection, the model classes are ranked according to $p(\mathcal{D}|\mathcal{M}_j)P(\mathcal{M}_j|\mathcal{U})$ for $j=1, \dots, N_M$, where the best class of models representing the system is the one which gives the largest value of this quantity. The evidence $p(\mathcal{D}|\mathcal{M}_j)$ may be calculated for each class of models using Eq. (3) where the likelihood $p(\mathcal{D}|\hat{\theta}_j, \mathcal{M}_j)$ is evaluated using the methods presented in the next section. The prior distribution $P(\mathcal{M}_j|\mathcal{U})$ over all the model classes \mathcal{M}_j , $j=1, \dots, N_M$, must be specified. In this work, a uniform prior distribution is chosen, leaving the Ockham factor alone to penalize model classes with increased numbers of parameters.

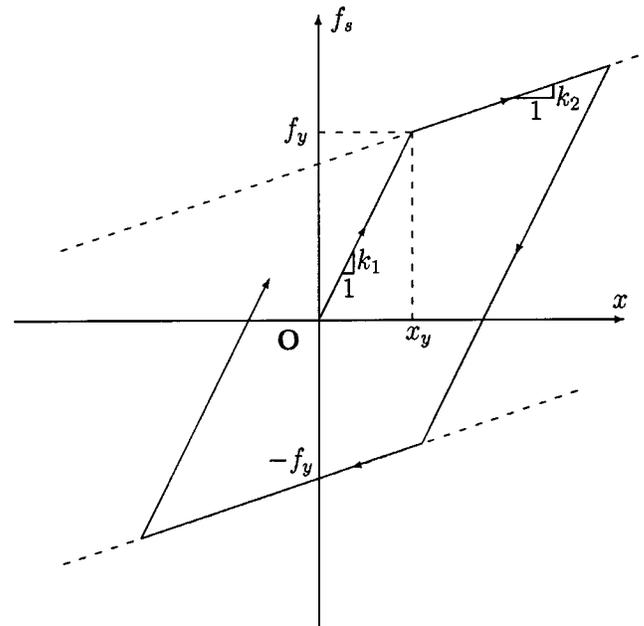


Fig. 1. Relationship between restoring force and displacement of bilinear hysteretic system (Example 1)

Comparison with Akaike's Approach

In the case of Akaike's information criterion (Akaike 1974), the best model class among the \mathcal{M}_j for $j=1, 2, \dots, N_M$ is chosen by maximizing an objective function $AIC(\mathcal{M}_j|\mathcal{D})$ over j that is defined by

$$AIC(\mathcal{M}_j|\mathcal{D}) = \ln p(\mathcal{D}|\hat{\theta}_j, \mathcal{M}_j) - N_j \quad (7)$$

where the log likelihood function is roughly proportional to the number of data points N in \mathcal{D} , while the penalty term is taken to be N_j , the number of adjustable parameters in the model class \mathcal{M}_j . [Akaike actually stated his criterion as minimizing $-2(AIC)$ but the equivalent form is more appropriate here.] When the number of data points is large, the first term will dominate. Akaike (1976) and Schwarz (1978) later developed independently another version of the objective function, denoted BIC, that is defined by

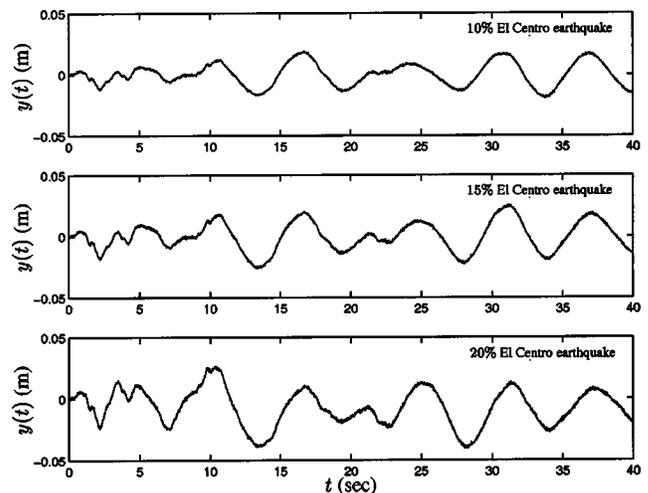


Fig. 2. Response measurements of oscillator for three levels of excitation (Example 1)

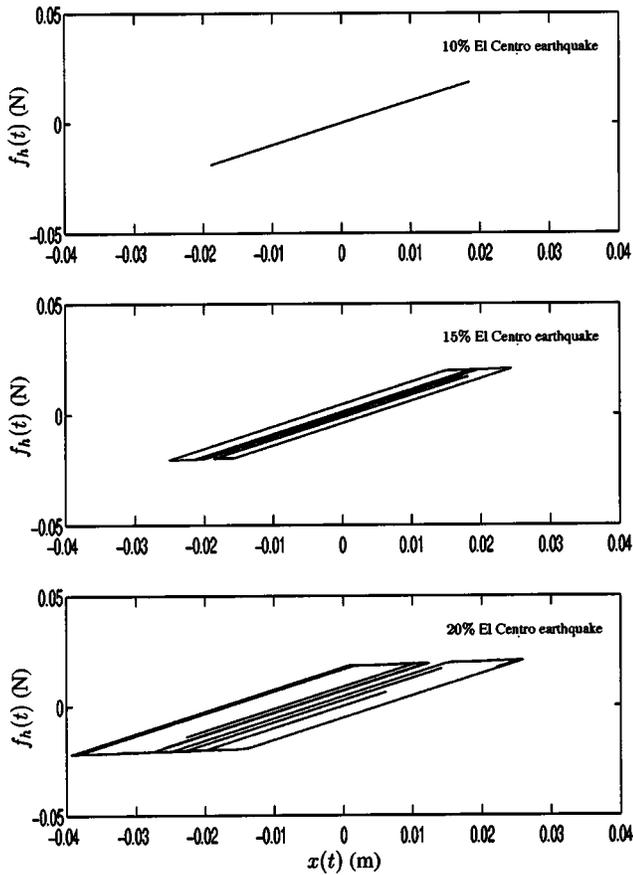


Fig. 3. Hysteresis loops of oscillator for three levels of excitation (Example 1)

$$\text{BIC}(\mathcal{M}_j|\mathcal{D}) = \ln p(\mathcal{D}|\hat{\boldsymbol{\theta}}_j, \mathcal{M}_j) - \frac{1}{2} N_j \ln N \quad (8)$$

where now the penalty term increases with the number of data points N .

BIC can be compared directly with the logarithm of the evidence from Eq. (3)

$$\ln p(\mathcal{D}|\mathcal{M}_j) \approx \ln p(\mathcal{D}|\hat{\boldsymbol{\theta}}_j, \mathcal{M}_j) + \beta_j \quad (9)$$

where the logarithm of the Ockham factor β_j is given by Eq. (5). This shows that for large N , the BIC agrees with the leading order terms in the logarithm of the evidence and so in this case it is equivalent to the Bayesian approach using equal priors for all of the $P(\mathcal{M}_j|\mathcal{U})$.

Model Updating Using a Bayesian Framework

A general Bayesian framework for structural model updating was proposed in (Beck and Katafygiotis 1998). It was originally presented using input-output measurements but it was recently extended for output-only measurements (Katafygiotis and Yuen 2001; Yuen and Katafygiotis 2001; Yuen et al. 2002; Yuen and Beck 2003). In this section, two of these methods are presented for input-output data and for output-only data, respectively.

Input-Output Measurements

In this section, a Bayesian approach for linear/nonlinear model updating using input-output measurements is presented. For details, see Beck and Katafygiotis (1998).

Table 1. Optimal Parameter Values for Each Model Class Representing the Oscillator (Example 1)

Excitation level	Model class	c	k_1	k_2	x_y	σ_η
10% El Centro earthquake	1	0.0204	1.000	—	—	0.0005
	2	—	1.019	—	UN	0.0013
	3	—	1.019	UN	UN	0.0013
15% El Centro earthquake	1	0.0902	0.989	—	—	0.0020
	2	—	1.0179	—	0.0214	0.0017
	3	—	1.001	0.108	0.0197	0.0007
20% El Centro earthquake	1	0.1928	0.956	—	—	0.0098
	2	—	0.9936	—	0.0211	0.0051
	3	—	0.9942	0.0924	0.0200	0.0011

Consider a system with N_d degrees of freedom (DOFs) and equation of motion

$$\mathbf{M}\ddot{\mathbf{x}} + \mathbf{f}_s(\mathbf{x}, \dot{\mathbf{x}}; \boldsymbol{\theta}_s) = \mathbf{T}\mathbf{f}(t) \quad (10)$$

where $\mathbf{M} \in \mathbb{R}^{N_d \times N_d}$ is the mass matrix, $\mathbf{f}_s \in \mathbb{R}^{N_d}$ is the nonlinear restoring force characterized by the structural parameters $\boldsymbol{\theta}_s$, $\mathbf{T} \in \mathbb{R}^{N_d \times N_f}$ is a force distributing matrix, and $\mathbf{f}(t) \in \mathbb{R}^{N_f}$ is an external excitation, e.g., force or ground acceleration, which is assumed to be measured.

Assume now that discrete response data are available at N_0 ($\leq N_d$) observed DOFs. Let Δt denote the sampling time step. Because of measurement noise and modeling error, referred to hereafter as prediction error, the measured response $\mathbf{y}(n) \in \mathbb{R}^{N_0}$ (at time $t = n\Delta t$) will differ from the model response $\mathbf{L}_0\mathbf{x}(n)$ corresponding to the measured degrees of freedom where \mathbf{L}_0 denotes an $N_0 \times N_d$ observation matrix, comprised of zeros and ones. Herein, the prediction error is modeled as discrete zero-mean Gaussian white noise vector process $\boldsymbol{\eta}(n) \in \mathbb{R}^{N_0}$:

$$\mathbf{y}(n) = \mathbf{L}_0\mathbf{x}(n) + \boldsymbol{\eta}(n) \quad (11)$$

where the discrete process $\boldsymbol{\eta}$ satisfies

$$E[\boldsymbol{\eta}(n)\boldsymbol{\eta}^T(p)] = \boldsymbol{\Sigma}_\eta \delta_{np} \quad (12)$$

where $E[\cdot]$ denotes expectation, δ_{np} denotes the Kronecker delta function, and $\boldsymbol{\Sigma}_\eta$ denotes the $N_0 \times N_0$ covariance matrix of the prediction error process $\boldsymbol{\eta}$.

Let $\boldsymbol{\theta}$ denote the parameter vector for identification. It may include the following parameters: (1) The structural parameters $\boldsymbol{\theta}_s$; (2) parameters defining the structural mass distribution; (3) the elements of the force distributing matrix \mathbf{T} ; and (4) the elements of the upper right triangular part of the prediction-error covariance matrix $\boldsymbol{\Sigma}_\eta$ (symmetry defines the lower triangular part of this matrix). Herein, it is assumed that the mass distribution can be modeled sufficiently accurately from structural drawings and so it is not part of the model parameters to be identified.

Let the dynamic data \mathcal{D} consist of the measured time histories at N discrete times of the excitation and observed response. If the covariance matrix for the prediction errors is $\boldsymbol{\Sigma}_\eta = \sigma_\eta^2 \mathbf{I}_{N_0}$, implying equal variances and stochastic independence for the predic-

Table 2. Probabilities of Different Model Classes Based on Data (Example 1)

Excitation level	$P(\mathcal{M}_1 \mathcal{D}, \mathcal{U})$	$P(\mathcal{M}_2 \mathcal{D}, \mathcal{U})$	$P(\mathcal{M}_3 \mathcal{D}, \mathcal{U})$
10% El Centro earthquake	1.0	3.1×10^{-1217}	3.1×10^{-1217}
15% El Centro earthquake	4.4×10^{-1174}	3.2×10^{-957}	1.0
20% El Centro earthquake	6.4×10^{-2303}	5.7×10^{-1609}	1.0

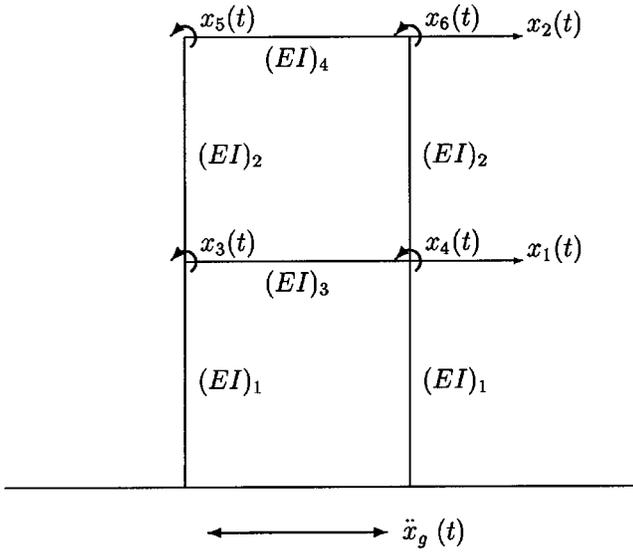


Fig. 4. Linear two-story structural frame (Example 2)

tion errors for different channels of measurements, then the updated PDF of the model parameters θ given the data \mathcal{D} and model class \mathcal{M} may be expressed as

$$p(\theta|\mathcal{D}, \mathcal{M}) = c_1 p(\theta|\mathcal{M}) (2\pi)^{-NN_0/2} \sigma_\eta^{-NN_0} \times \exp\left(-\frac{NN_0}{2\sigma_\eta^2} J_1(\theta|\mathcal{D}, \mathcal{M})\right) \quad (13)$$

where c_1 is a normalizing constant and $p(\theta|\mathcal{M})$ is the prior PDF of the model parameters θ expressing the user's judgement about the relative plausibility of the values of the model parameters before the data are used. The objective function $J_1(\theta|\mathcal{D}, \mathcal{M})$ is given by

$$J_1(\theta|\mathcal{D}, \mathcal{M}) = \frac{1}{NN_0} \sum_{k=1}^N \|\mathbf{L}_0 \mathbf{x}(k\Delta t; \theta, \mathcal{M}) - \mathbf{y}(k\Delta t)\|^2 \quad (14)$$

where $\mathbf{x}(k\Delta t; \theta, \mathcal{M})$ is the calculated response based on the assumed class of models and the parameter vector θ while $\mathbf{y}(k\Delta t)$ is the measured response at time $k\Delta t$. Furthermore, $\|\cdot\|$ denotes the Euclidean norm of a vector. The most probable model parameters $\hat{\theta}$ are obtained by maximizing $p(\theta|\mathcal{D}, \mathcal{M})$ in Eq. (13). For large N , this is equivalent to minimizing $J_1(\theta|\mathcal{D}, \mathcal{M})$ in Eq. (14) over all parameters in θ that it depends on, because this factor dominates in Eq. (13). The most probable value of the prediction-error variance in $\hat{\theta}$ is $\hat{\sigma}_\eta^2 = \min J_1(\theta|\mathcal{D}, \mathcal{M})$. In the globally identifiable case (Beck and Katafygiotis 1998), it turns out that $p(\theta|\mathcal{D}, \mathcal{M})$ is well approximated by a Gaussian distribution with mean $\hat{\theta}$ and covariance matrix equal to the inverse of the Hessian of $-\ln[p(\theta|\mathcal{D}, \mathcal{M})]$ at $\hat{\theta}$.

Output-Only Measurements

In this section, a Bayesian approach for linear/nonlinear model updating using output measurements is presented. For details, see Katafygiotis and Yuen (2001) and Yuen and Beck (2003).

Consider the same equation of motion as in Eq. (10). However, now the external excitation $\mathbf{f}(t) \in \mathbb{R}^{N_f}$ is modeled by a Gaussian process with zero mean and a spectral density matrix function characterized by the excitation parameters θ_f :

Table 3. Optimal Parameter Values for Each Model Class Representing the Structural Frame (Example 2)

Parameter	ϕ_1	ϕ_2	θ_1	θ_2	θ_3	S_{n1}	S_{n2}
Class 1	1.131	1.007	0.913	0.879	—	0.158	0.159
Class 2	1.057	1.536	1.130	1.130	—	0.150	0.063
Class 3	1.027	1.093	0.988	1.001	1.024	0.085	0.080

$$\mathbf{S}_f(\omega) = \mathbf{S}_f(\omega; \theta_f) \quad (15)$$

Assume response data are available at N discrete times for N_0 ($\leq N_d$) observed DOFs with prediction error modeled as in Eqs. (11) and (12).

Let θ denote the parameter vector for identification. It may include the following parameters: (1) The structural parameters θ_s ; (2) the excitation parameters θ_f ; and (3) the elements of the upper right triangular part of the prediction-error covariance matrix Σ_η (symmetry defines the lower triangular part of this matrix).

Spectral Density Estimator and Its Statistical Properties

Consider the stochastic vector process $\mathbf{y}(t)$ and a finite number of discrete data $\mathbf{Y}_N = \{\mathbf{y}(n), n = 1, \dots, N\}$. Based on \mathbf{Y}_N , we introduce the following discrete estimator of the spectral density matrix of the stochastic process $\mathbf{y}(t)$:

$$\mathbf{S}_{y,N}(\omega_k) = \mathcal{F}_N(\omega_k) \bar{\mathcal{F}}_N^T(\omega_k) \quad (16)$$

where \bar{z} denotes the complex conjugate of a complex variable z and $\mathcal{F}_N(\omega_k)$ denotes the (scaled) discrete Fourier transform of the vector process \mathbf{y} at frequency ω_k , as follows:

$$\mathcal{F}_N(\omega_k) = \sqrt{\frac{\Delta t}{2\pi N}} \sum_{n=0}^{N-1} \mathbf{y}(n) e^{-i\omega_k n \Delta t} \quad (17)$$

where $\omega_k = k\Delta\omega$, $k = 0, \dots, N_1 - 1$ with $N_1 = \text{INT}[(N+1)/2]$, $\Delta\omega = 2\pi/T$, and $T = N\Delta t$.

Using Eq. (11) and taking the expectation of Eq. (16) (noting that \mathbf{x} and $\boldsymbol{\eta}$ are taken as stochastically independent) yields

$$E[\mathbf{S}_{y,N}(\omega_k)|\theta] = \mathbf{L}_0 E[\mathbf{S}_{x,N}(\omega_k)|\theta] \mathbf{L}_0^T + E[\mathbf{S}_{\eta,N}(\omega_k)] \quad (18)$$

where $\mathbf{S}_{x,N}(\omega_k)$ and $\mathbf{S}_{\eta,N}(\omega_k)$ are defined in a manner similar to that described by Eqs. (16) and (17). It easily follows from Eqs. (12) and (16) that

$$E[\mathbf{S}_{\eta,N}(\omega_k)] = \frac{\Delta t}{2\pi} \Sigma_\eta \equiv \mathbf{S}_{\eta 0} \quad (19)$$

The term $E[\mathbf{S}_{x,N}(\omega_k)|\theta]$ in Eq. (18) can also be easily calculated by noting that $\mathbf{S}_{x,N}(\omega_k)$ has elements

$$S_{x,N}^{(j,l)}(\omega_k) = \frac{\Delta t}{2\pi N} \sum_{n,p=0}^{N-1} x_j(n) x_l(p) e^{-i\omega_k(n-p)\Delta t} \quad (20)$$

Grouping together terms having the same value of $(p-n)$ in Eq. (20) and taking the expectation we obtain the following expression:

$$E[S_{x,N}^{(j,l)}(\omega_k)] = \frac{\Delta t}{4\pi N} \sum_{n=0}^{N-1} \gamma_n [R_x^{(j,l)}(n\Delta t) e^{-i\omega_k n \Delta t} + R_x^{(j,l)}(-n\Delta t) e^{i\omega_k n \Delta t}] \quad (21)$$

where γ_n is given by $\gamma_0 = N$ and $\gamma_n = 2(N-n)$, $n \geq 1$. $R_x^{(j,l)}$ are the cross-correlation functions between the j th and l th component of the response \mathbf{x} . However, it is usually not possible to obtain the

Table 4. Natural Frequencies (in Hz) of Optimal Model in Each Class (Example 2)

Mode	1	2
Actual	2.000	5.144
Class 1	2.048	5.009
Class 2	2.000	5.323
Class 3	1.995	5.142

correlation functions theoretically for nonlinear systems. In this case, we can generate samples of the response and hence samples of their spectral density estimators. Then, the expected values of the spectral estimates can be approximated by the average of the spectral density estimators obtained from the samples.

Assume that there is a set of independent, identically distributed, time histories $\mathbf{Y}_N^{(1)}, \dots, \mathbf{Y}_N^{(M)}$. As $N \rightarrow \infty$, the corresponding discrete Fourier transforms $\mathcal{Y}_N^{(n)}(\omega_k)$, $n = 1, \dots, M$ are independent and follow an identical complex N_0 -variate normal distribution with zero mean. Then, if $M \geq N_0$, according to Krishnaiah (1976), the average spectral density estimate

$$\mathbf{S}_{y,N}^M(\omega_k) = \frac{1}{M} \sum_{n=1}^M \mathbf{S}_{y,N}^{(n)}(\omega_k) = \frac{1}{M} \sum_{n=1}^M \mathcal{Y}_N^{(n)}(\omega_k) \overline{\mathcal{Y}_N^{(n)}(\omega_k)}^T \quad (22)$$

follows a central complex Wishart distribution of dimension N_0 with M DOFs and mean $E[\mathbf{S}_{y,N}^M(\omega_k)]$. The PDF of this distribution is given by

$$p(\mathbf{S}_{y,N}^M(\omega_k)) = c_0 \frac{|\mathbf{S}_{y,N}^M(\omega_k)|^{M-N_0}}{|E[\mathbf{S}_{y,N}(\omega_k)]|^M} \times \exp(-M \operatorname{tr}\{E[\mathbf{S}_{y,N}(\omega_k)]^{-1} \mathbf{S}_{y,N}^M(\omega_k)\}) \quad (23)$$

where $c_0 = M^{M-N_0} \pi^{-N_s(N_s-1)/2} / [\prod_{p=1}^{N_s} (M-p)!]$ is a normalizing constant, and $|\mathbf{A}|$ and $\operatorname{tr}\{\mathbf{A}\}$ denote the determinant and the trace, respectively, of a matrix \mathbf{A} . Note that this approximation is very accurate for large N even if $\mathbf{y}(n\Delta t)$, $n = 1, \dots, N$, is not Gaussian. This is due to the robustness of the probability distribution of the discrete Fourier transform with respect to the probability distribution of the response signal.

Furthermore, it can be shown that when $N \rightarrow \infty$, the complex vectors $\mathcal{Y}_N(\omega_k)$ and $\mathcal{Y}_N(\omega_l)$ are independent for $k \neq l$ (Yuen et al. 2002). As a result, the matrices $\mathbf{S}_{y,N}^M(\omega_k)$ and $\mathbf{S}_{y,N}^M(\omega_l)$ are independently Wishart distributed for $k \neq l$, that is

$$p[\mathbf{S}_{y,N}^M(\omega_k), \mathbf{S}_{y,N}^M(\omega_l)] = p[\mathbf{S}_{y,N}^M(\omega_k)] p[\mathbf{S}_{y,N}^M(\omega_l)] \quad (24)$$

Although Eqs. (23) and (24) are correct only asymptotically as $N \rightarrow \infty$, it was shown by simulations that these are indeed very accurate approximations in a certain bandwidth of frequencies for the case where N is finite but large (Yuen 2002). In the case of displacements (or accelerations), this frequency bandwidth corresponds to the lower (or higher) frequency range $\omega_k \in [\omega_1, \omega_K]$ (or ω_K, ω_{N_1-1}) for some cutoff frequency ω_K . The frequency range increases as the level of prediction error increases.

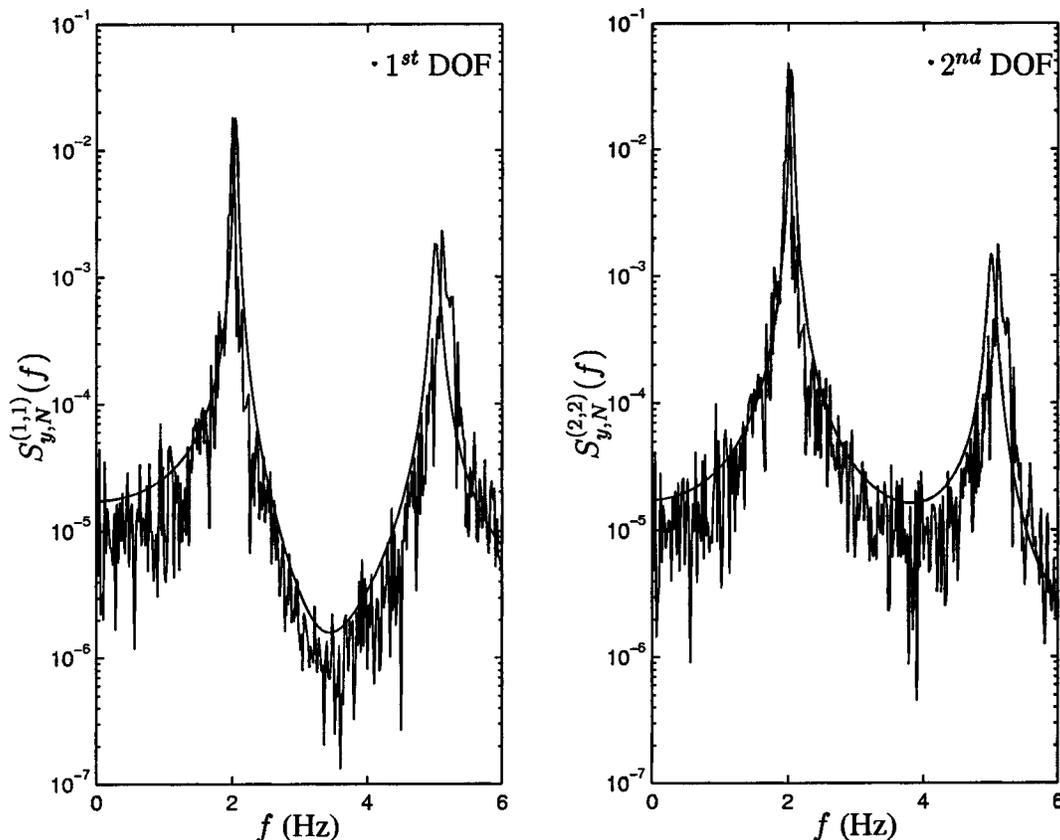


Fig. 5. Spectra estimated directly from measurements and from optimal model in Model Class 1 (Example 2)

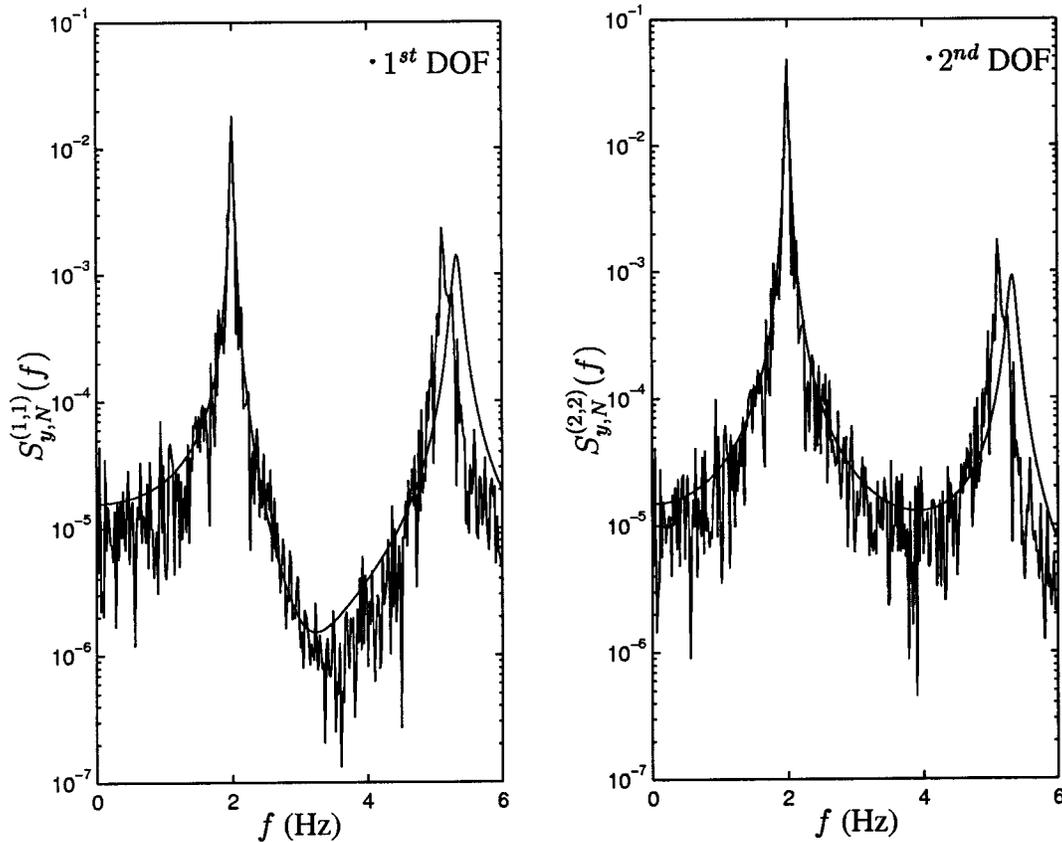


Fig. 6. Spectra estimated directly from measurements and from optimal model in Model Class 2 (Example 2)

Identification Based on Spectral Density Estimates

Based on the above discussion regarding the statistical properties of the average spectral estimator $\hat{\mathbf{S}}_{y,N}^M(\omega_k)$, a Bayesian approach for updating the PDF of the uncertain parameter vector $\boldsymbol{\theta}$ is proposed as follows. Given $M \geq N_0$ independent sets of observed data $\mathcal{D} = \{\hat{\mathbf{Y}}_N^{(n)}, n = 1, \dots, M\}$, one may calculate the corresponding observed spectral estimate matrices $\hat{\mathbf{S}}_{y,N}^{(n)}, n = 1, \dots, M$ using Eqs. (16) and (17). Next, one can calculate the average matrix estimates $\hat{\mathbf{S}}_{y,N}^M(\omega_k)$ using Eq. (22) and then form the set $\hat{\mathbf{S}}_{y,N}^{M,K} = \{\hat{\mathbf{S}}_{y,N}^M(k\Delta\omega), k = 1, \dots, K\}$. Using Bayes' theorem, the updated PDF of the model parameters $\boldsymbol{\theta}$ given the data $\hat{\mathbf{S}}_{y,N}^{M,K}$ is then given by

$$p(\boldsymbol{\theta}|\mathcal{D}, \mathcal{M}) \approx p(\boldsymbol{\theta}|\hat{\mathbf{S}}_{y,N}^{M,K}, \mathcal{M}) = c_2 p(\boldsymbol{\theta}|\mathcal{M}) p(\hat{\mathbf{S}}_{y,N}^{M,K}|\boldsymbol{\theta}, \mathcal{M}) \quad (25)$$

where c_2 is a normalizing constant such that the integral on the right hand side of Eq. (25) over the domain of $\boldsymbol{\theta}$ is equal to 1. The factor $p(\boldsymbol{\theta}|\mathcal{M})$ in the above equation represents the prior PDF, which expresses the relative plausibilities of different values of $\boldsymbol{\theta}$ based on prior information and engineering judgment. The likelihood $p(\hat{\mathbf{S}}_{y,N}^{M,K}|\boldsymbol{\theta}, \mathcal{M})$ expresses the contribution of the measured data. Based on Eqs. (23) and (24), the likelihood can be expressed as follows:

$$p(\hat{\mathbf{S}}_{y,N}^{M,K}|\boldsymbol{\theta}, \mathcal{M}) \approx c_0^K \prod_{k=1}^K \frac{|\hat{\mathbf{S}}_{y,N}^M(\omega_k)|^{M-N_0}}{|E[\mathbf{S}_{y,N}(\omega_k)]|^M} \times \exp(-M \text{tr}\{E[\mathbf{S}_{y,N}(\omega_k)]^{-1} \hat{\mathbf{S}}_{y,N}^M(\omega_k)\}) \quad (26)$$

The most probable parameters $\hat{\boldsymbol{\theta}}$ are obtained by minimizing the objective function $J_2(\boldsymbol{\theta}) = -\ln[p(\boldsymbol{\theta}|\mathcal{M})p(\hat{\mathbf{S}}_{y,N}^{M,K}|\boldsymbol{\theta}, \mathcal{M})]$. Furthermore, the updated PDF $p(\boldsymbol{\theta}|\mathcal{D}, \mathcal{M})$ can be well approximated by a Gaussian distribution centered at the optimal point $\hat{\boldsymbol{\theta}}$ if $\boldsymbol{\theta}$ is globally identifiable. The covariance matrix $\boldsymbol{\Sigma}_\theta$ is equal to the inverse of the Hessian of the function $J_2(\boldsymbol{\theta})$ evaluated at $\boldsymbol{\theta} = \hat{\boldsymbol{\theta}}$. This property provides a very efficient way for the quantification of the uncertainty for the model parameters without evaluating high dimensional integrals.

Illustrative Examples

Example 1: Single-Degree-of-Freedom Nonlinear Oscillator under Seismic Excitation

In this example, a bilinear hysteretic oscillator with linear viscous damping is considered

$$m\ddot{x} + c\dot{x} + f_h(x; k_1, k_2, x_y) = f(t) \quad (27)$$

where m = mass; c = damping coefficient; and $f_h(x; k_1, k_2, x_y)$ = hysteretic restoring force, whose behavior is shown in Fig. 1. Here, $m = 1$ kg is assumed known. The parameters $\tilde{\boldsymbol{\theta}} = [\tilde{c}, \tilde{k}_1, \tilde{k}_2, \tilde{x}_y]^T$ used to generate the data are $\tilde{c} = 0.02$ N·s/m, $\tilde{k}_1 = 1.0$ N/m, $\tilde{k}_2 = 0.1$ N/m, $\tilde{x}_y = 0.02$ m, which gives a small-amplitude natural frequency of $1/2\pi$ Hz.

The oscillator is assumed to be excited by 10, 15, and 20% scaling of the 1940 El Centro earthquake record. The duration of measurement is $T = 40$ s with sampling frequency 60 Hz, so that the number of data points is $N = 2,400$. It is assumed that the earthquake excitation and response displacement are measured to

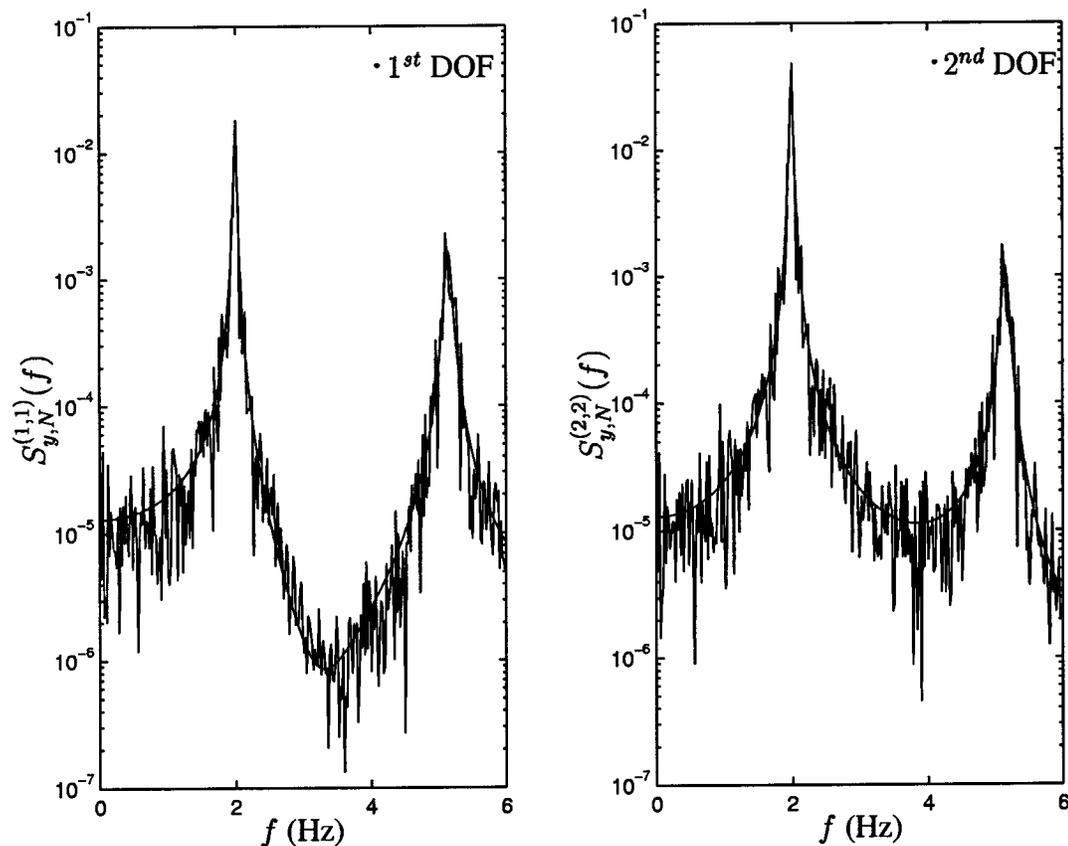


Fig. 7. Spectra estimated directly from measurements and from optimal model in Model Class 3 (Example 2)

give the data \mathcal{D} where 5% rms noise is imposed on the structural response measurements, i.e., the measurement noise is 5% of the rms of the noise-free response. Fig. 2 shows the measurements for the three levels of excitation and Fig. 3 shows the corresponding hysteresis loops. It can be seen that the oscillator behaved linearly (did not yield) when subjected to 10% of the El Centro earthquake record. Three classes of models are considered. They all use zero-mean Gaussian discrete white noise as the prediction-error model.

Model Class 1 (\mathcal{M}_1): Linear oscillators with damping coefficient $c > 0$, stiffness parameter $k_1 > 0$, and prediction-error standard deviation σ_η .

Model Class 2 (\mathcal{M}_2): Elastoplastic oscillators, i.e., bilinear hysteretic but with $k_2 = 0$, with stiffness parameter $k_1 > 0$, yielding level x_y , and prediction-error standard deviation σ_η ; and no viscous damping.

Model Class 3 (\mathcal{M}_3): Bilinear hysteretic oscillators with preyield stiffness $k_1 > 0$, postyielding stiffness $k_2 > 0$, yielding level x_y , and prediction-error standard deviation σ_η . Note that this class of models does not include the exact model since linear viscous damping is not included.

Independent uniform prior distributions are assumed for the parameters c , k_1 , k_2 , x_y , and σ_η over the ranges (0,0.5) N·s/m, (0,2) N/m, (0,0.5) N/m, (0,0.1) m, and (0,0.01) m, respectively.

Table 5. Probabilities of Different Model Classes Based on Data (Example 2)

$P(\mathcal{M}_1 \mathcal{D},\mathcal{U})$	$P(\mathcal{M}_2 \mathcal{D},\mathcal{U})$	$P(\mathcal{M}_3 \mathcal{D},\mathcal{U})$
2.6×10^{-23}	1.7×10^{-15}	1.0

Table 1 shows the optimal parameter values for each class of models for the three levels of excitation using the method presented in the section "Input-Output Measurements." "UN" indicates that the parameter is unidentifiable. For example, in \mathcal{M}_2 with 10% El Centro earthquake, x_y is unidentifiable because the oscillator behaves perfectly linearly (Fig. 3). In fact, the optimal parameter values for \mathcal{M}_1 are very close to their target values at this level of excitation. For higher levels of excitation, the optimal linear model in \mathcal{M}_1 has lower stiffness and higher values of its damping coefficient to represent the increased flexibility and energy dissipation due to yielding.

Table 2 shows the values of $P(\mathcal{M}_j|\mathcal{D},\mathcal{U})$, $j=1,2,3$, for the three levels of excitation that are calculated from Eq. (1) using the evidence for each model from Eq. (3) and equal priors $P(\mathcal{M}_j|\mathcal{U})=1/3$. Note that in all three cases the model classes other than the most probable model class have essentially zero probability, implying that these model classes can be discarded for response prediction. In the case of 10% scaling of the El Centro earthquake record, it is not surprising that $P(\mathcal{M}_1|\mathcal{D},\mathcal{U})$ is the largest since the oscillator behaves linearly (Fig. 3). However, for higher levels of excitation, $P(\mathcal{M}_3|\mathcal{D},\mathcal{U})$ is the largest. Although \mathcal{M}_3 does not include linear viscous damping, the hysteretic behavior can be captured well by this model. More interestingly, \mathcal{M}_2 outperforms \mathcal{M}_1 at these two levels of excitation. Although \mathcal{M}_2 cannot capture the viscous damping mechanism, the energy dissipated by the hysteretic behavior for 15 and 20% scaling of the El Centro earthquake record is much more significant than the contribution from the viscous damping, as can be seen by the large increase in the optimal damping ratio for the corresponding "equivalent" linear systems \mathcal{M}_1 in Table 1. Fur-

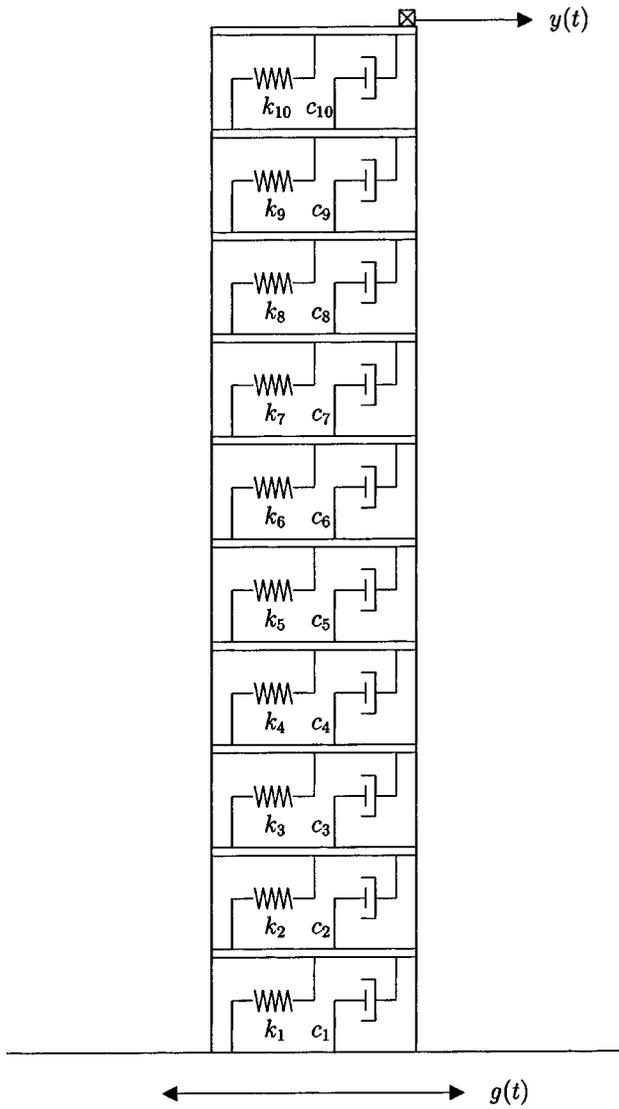


Fig. 8. Ten-story shear building (Example 3)

thermore, the restoring force behavior for \mathcal{M}_2 is more correct than for \mathcal{M}_1 , although it is still not exact.

This example illustrates an important point in system identification. In reality, there is no exact class of models for a real structure and the best class depends on the system data that is used. Response predictions should only be made for excitation levels that do not differ greatly from those used in the system identification. If we wish to select between the linear models (\mathcal{M}_1) and the elastoplastic models without viscous damping (\mathcal{M}_2), then \mathcal{M}_2 is better for high levels of excitation while \mathcal{M}_1 is better for lower levels of excitation.

Table 6. Optimal Natural Frequencies (rad/s) of Building (Example 3)

Number of modes	ω_1	ω_2	ω_3	ω_4	ω_5	ω_6	ω_7	ω_8
Exact	5.789	17.24	28.30	38.73	48.30	56.78	64.00	69.79
1	6.946	—	—	—	—	—	—	—
2	5.799	20.68	—	—	—	—	—	—
3	5.814	17.16	33.96	—	—	—	—	—
4	5.842	17.18	27.94	43.82	—	—	—	—
5	5.848	17.19	27.97	38.06	50.58	—	—	—
6	5.849	17.19	27.97	38.09	48.10	56.72	—	—
7	5.849	17.19	27.97	38.09	48.13	56.34	64.18	—
8	5.849	17.19	27.97	38.09	48.13	56.34	64.18	69.41

Example 2: Linear Two-Story Frame under Seismic Excitation

The second example refers to a 6-DOF two-story structural frame with story height $H=2.5$ m and width $W=4.0$ m, as shown in Fig. 4. All the chosen model classes are linear. All members are assumed to be rigid in their axial direction. For each member, the mass is uniformly distributed along its length. The rigidity-to-mass ratio is chosen to be $\tilde{EI}_1/m = \tilde{EI}_2/m = \tilde{EI}_3/10m = \tilde{EI}_4/10m = 2,252 \text{ m}^4 \cdot \text{s}^{-2}$, where m denotes the mass per unit length of all members. As a result, the first two natural frequencies of this structure are 2.000 and 5.144 Hz. Furthermore, a Rayleigh damping model is assumed, i.e., the damping matrix $\mathbf{C} = \alpha\mathbf{M} + \beta\mathbf{K}$, where \mathbf{M} and \mathbf{K} are the mass and stiffness matrices, respectively. In this case, the nominal values of the damping coefficients $\tilde{\alpha}$ and $\tilde{\beta}$ are chosen to be 0.182 s^{-1} and $0.442 \times 10^{-3} \text{ s}$ so that the damping ratios for the first two modes are 1.00%.

Three classes of structural models are considered. In each case, independent zero-mean discrete Gaussian white noise is used for the prediction-error model, with spectral intensity $S_{n1} = 0.027 \text{ m}^2 \cdot \text{s}^{-3}$ and $S_{n2} = 0.059 \text{ m}^2 \cdot \text{s}^{-3}$ at the two observed degrees of freedom. In order to have better scaling, the damping parameters are parametrized as follows: $\alpha = \phi_1 \tilde{\alpha}$ and $\beta = \phi_2 \tilde{\beta}$.

Model Class 1 (\mathcal{M}_1) assumes a class of two-story shear buildings with nominal interstory stiffness $\bar{k}_1 = \bar{k}_2 = 2 \times 12 \tilde{EI}_1 / H^3$. In order to have better scaling, the stiffness is parameterized as follows: $k_j = \theta_j \bar{k}_j$, $j=1,2$. Therefore, the uncertain parameters are $\theta_j, \phi_j, S_{nj}, j=1,2$.

Model Class 2 (\mathcal{M}_2) assumes the actual class of models except that due to modeling error, $EI_1 = \theta_1 \tilde{EI}_1$, $EI_2 = \theta_2 \tilde{EI}_2$, $EI_3 = 0.5 \theta_1 \tilde{EI}_3$, and $EI_4 = 0.5 \theta_2 \tilde{EI}_4$, where the nominal values were given earlier. Therefore, the uncertain parameters are θ_j, ϕ_j and $S_{nj}, j=1,2$.

Model Class 3 (\mathcal{M}_3) assumes that $EI_1 = \theta_1 \tilde{EI}_1$, $EI_2 = \theta_2 \tilde{EI}_2$, and $EI_j = \theta_3 \tilde{EI}_j$, $j=3,4$. Therefore, the uncertain pa-

Table 7. Probabilities of Models with Different Numbers of Modes Based on Data (Example 3)

Number of modes m	1	2	3	4	5	6	7	8
ln likelihood	1894	2251	2511	2619	2682	2714	2723	2723
ln Ockham factor	-43.7	-56.4	-68.9	-69.2	-75.9	-91.2	-109	-121
ln evidence	1850	2195	2442	2550	2606	2623	2614	2602
$P(\mathcal{M}_m \mathcal{D}, \mathcal{U})$	3.0×10^{-336}	2.2×10^{-186}	6.4×10^{-79}	2.4×10^{-32}	1.0×10^{-7}	1.0	1.7×10^{-4}	1.3×10^{-9}

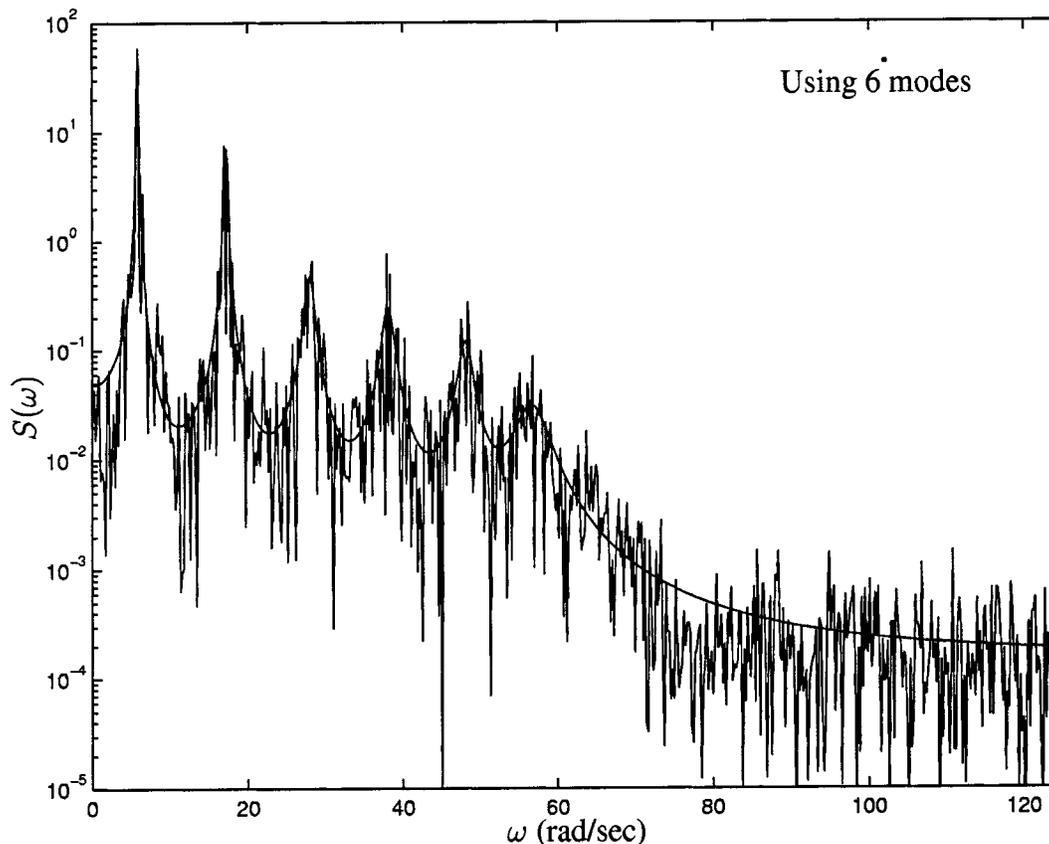


Fig. 9. Spectrum estimated directly from measurements and from optimal model with six modes (Example 3)

parameters are $\theta_1, \theta_2, \theta_3, \phi_1, \phi_2, S_{n1}$, and S_{n2} . Note that the true model lies in this set.

The structure is assumed to be excited by a white noise ground motion, which is not measured. The spectral intensity of the ground motion is taken to be $S_0 = 1.0 \times 10^{-5} \text{ m}^2 \cdot \text{s}^{-3}$. The data \mathcal{D} consist of the absolute accelerations with 10% measurement noise at the first and second DOFs over a time interval of 100 s, using a sampling interval of 0.01 s. Identification was performed using the Bayesian spectral density approach of the section "Output-Only Measurements" with the same set of data for each of the three classes of models. The number of data points N is taken to be 600 because only the estimated spectrum up to 20.0 Hz is used.

The prior distributions $p(\theta_j | \mathcal{M}_j)$, $j = 1, 2, 3$, are assumed to be an independent uniform distribution over the interval (0,2) for $\theta_1, \theta_2, \theta_3, \phi_1, \phi_2$ and over the interval (0,0.5) $\text{m}^2 \cdot \text{s}^{-3}$ for S_{n1} and S_{n2} .

Table 3 shows the optimal parameter values for each class of models. It is not surprising that both θ_1 and θ_2 in Case 1 are less than unity because the shear building models assume rigid floors but the floors of the actual structure are not rigid. Table 4 shows the associated natural frequencies with the actual frame and the optimal models. Note that the optimal model in \mathcal{M}_3 can fit both frequencies very well since the exact model is in this class. On the other hand, \mathcal{M}_1 and \mathcal{M}_2 cannot fit the frequency of the second mode as well as \mathcal{M}_3 . Figs. 5–7 show the estimated spectrum using the measurements (zigzag curve) and the optimal model spectrum (smoother curve) for the three classes of models, respectively. One can see that the optimal model in \mathcal{M}_2 provides a better fit to the first mode than \mathcal{M}_1 , but it is the opposite for the

second mode. The optimal model in \mathcal{M}_3 gives excellent matching with the estimated spectrum for both modes.

Table 5 shows the values of $P(\mathcal{M}_j | \mathcal{D}, \mathcal{U})$ for $j = 1, 2, 3$, calculated from Eq. (1) using the evidence for each model from Eq. (3) and equal priors $P(\mathcal{M}_j | \mathcal{U}) = 1/3$. As expected, $P(\mathcal{M}_3 | \mathcal{D}, \mathcal{U})$ is the largest among the three classes of models because it contains the actual model. On the other hand, $P(\mathcal{M}_1 | \mathcal{D}, \mathcal{U})$ is the smallest one. Although the optimal model in \mathcal{M}_1 gives a better fit for the second mode than \mathcal{M}_2 , it does not fit the first mode as well as the optimal model in \mathcal{M}_2 and the contribution of the first mode to the structural response is one order of magnitude larger than the second mode. This implies that although \mathcal{M}_2 has significant modeling error for the beams (about 50%), it is still a better class of models than the shear building models.

Example 3: Ten-Story Shear Building under Seismic Excitation

The third example uses response measurements from the ten-story building shown in Fig. 8. The Bayesian approach is applied to select the optimal number of modes for a linear model. It is assumed that this building has a uniformly distributed floor mass and story stiffness over its height. The stiffness to mass ratios \tilde{k}_j/m_j , $j = 1, \dots, 4$, are chosen to be $1,500 \text{ s}^{-2}$ so that the fundamental frequency of the building is 0.9213 Hz. Rayleigh damping is assumed, i.e., the damping matrix \mathbf{C} is given by $\mathbf{C} = \alpha \mathbf{M} + \beta \mathbf{K}$, where $\alpha = 0.0866 \text{ s}^{-1}$ and $\beta = 0.0009 \text{ s}$, so that the damping ratios for the first two modes are 1.00%. The structure is assumed to be subjected to a wideband random ground motion, which can be adequately modeled as Gaussian white noise with spectral in-

tensity $S_{f_0} = 0.02 \text{ m}^2 \cdot \text{s}^{-3}$. Note that the matrix \mathbf{T} in Eq. (10) is equal to the matrix $-[m_1, \dots, m_{10}]^T$ in this case. Each model class \mathcal{M}_j ($j=1, \dots, 8$) consists of a linear modal model with j modes and the uncertain parameters are the natural frequency, damping ratio and modal participation factor for each mode; and the spectral intensity S_n of the prediction error at the observed degree of freedom.

The data \mathcal{D} consist of the (simulated) absolute accelerations at the top floor with 5% noise added over a time interval $T=30$ s, using a sampling interval $\Delta t=0.01$ s. The added noise is simulated using a spectral intensity $S_n = 1.94 \times 10^{-4} \text{ m}^2 \cdot \text{s}^{-3}$. The Bayesian spectral density approach of the section "Output-Only Measurements" is used for the identification. The number of data points N is taken to be 600 because only the estimated spectrum up to 20.0 Hz is used.

Independent prior distributions for the parameters are taken as follows: Gaussian distribution for the natural frequencies with mean $5.5(2j-1)$ rad/s and coefficient of variation 0.05 for the j th mode. Furthermore, the damping ratios, modal participation factor, and spectral intensity of the prediction error are assumed to be uniformly distributed over the ranges (0,0.05), (0,2), and (0,0.01) $\text{m}^2 \cdot \text{s}^{-3}$, respectively.

Table 6 shows the optimal natural frequencies for model classes with one to eight modes. Table 7 shows the values of the log likelihood, the log Ockham factor, the log evidence, and the probability of each model class $[P(\mathcal{M}_m|\mathcal{D},\mathcal{U})]$ for model classes with one to eight modes ($m=1, \dots, 8$), calculated from Eq. (1) using the evidence for each model from Eq. (3) and equal priors $P(\mathcal{M}_j|\mathcal{U})=1/8$. It implies that using six modes is optimal. Fig. 9 shows the spectrum estimated directly from the data (zigzag curve) and the optimal model spectrum using six modes (smoother curve). One can see that the optimal model using six modes can fit the measured spectrum very well. Furthermore, all the six optimal natural frequencies are very close to their target values, which is not the case for using two to five modes. It was found that if AIC is used, eight modes is optimal because the penalty term is too small compared to the changing of the log likelihood term in Eq. (7). On the other hand, if BIC in Eq. (8) is used, then six modes are optimal, agreeing with the Bayesian approach using the evidence for the various modal models.

Conclusion

A Bayesian probabilistic approach for model class selection is presented and numerical examples are given to illustrate the method. The best class of models is taken to be the one which is the most plausible based on the data, that is, it possesses the largest probability of any of the model classes conditional on the data. This probability depends on the evidence for the model class provided by the data and the user's choice of prior probability distribution over the classes of models. The methodology can handle input-output and output-only data for linear and nonlinear dynamical systems.

References

Akaike, H. (1974). "A new look at the statistical identification model." *IEEE Trans. Autom. Control*, 19, 716–723.

Akaike, H. (1976). "On entropy maximization principle." *Applications of statistics*, P. R. Krishnaiah, ed., North-Holland, Amsterdam, The Netherlands, 27–41.

Beck, J. L. (1978). "Determining models of structures from earthquake records." *Technical Rep. No. EERL 78-01*, California Institute of

Technology, Earthquake Engineering Research Laboratory, Pasadena, Calif.

Beck, J. L., and Au, S. K. (2002). "Bayesian updating of structural models and reliability using Markov Chain Monte Carlo simulation." *J. Eng. Mech.*, 128(4), 380–391.

Beck, J. L., and Katafygiotis, L. S. (1998). "Updating models and their uncertainties. I: Bayesian statistical framework." *J. Eng. Mech.*, 124(4), 455–461.

Box, G. E. P., and Jenkins, G. M. (1970). *Time series analysis, forecasting and control*, Holden-Day, San Francisco.

Box, G. E. P., and Tiao, G. C. (1973). *Bayesian inference in statistical analysis*, Addison-Wesley, Reading, Mass.

Cox, R. T. (1961). *The algebra of probable inference*, Johns Hopkins University Press, Baltimore.

Gersch, W., Taoka, G. T., and Liu, R. (1976). "Structural system parameter estimation by two-stage least squares method." *J. Eng. Mech.*, 102(5), 883–899.

Grigoriu, M., Cornell, C., and Veneziano, D. (1979). "Probabilistic modelling as decision making." *J. Eng. Mech. Div.*, 105(4), 585–596.

Gull, S. F. (1988). "Bayesian inductive inference and maximum entropy." *Maximum entropy and Bayesian methods*, J. Skilling, ed., Kluwer, Boston, 53–74.

Hjelmstad, K. D., Wood, S. L., and Clark, S. J. (1992). "Mutual residual energy method for parameter estimation in structures." *J. Struct. Eng.*, 118(1), 223–242.

Hoshiya, M., and Saito, E. (1984). "Structural identification by extended Kalman filter." *J. Eng. Mech.*, 110(12), 1757–1770.

Jaynes, E. T. (1983). *Papers on probability statistics and statistical physics*, R. Rosenkrantz, ed., Reidel, Dordrecht, The Netherlands.

Jaynes, E. T. (2003). *Probability theory: The logic of science*, L. Bretthorst, ed., Cambridge University Press, Cambridge, U.K.

Jeffreys, H. (1961). *Theory of probability*, 3rd Ed., Clarendon, Oxford, U.K.

Katafygiotis, L. S., and Beck, J. L. (1998). "Updating models and their uncertainties. II: Model identifiability." *J. Eng. Mech.*, 124(4), 463–467.

Katafygiotis, L. S., Papadimitriou, C., and Lam, H. F. (1998). "A probabilistic approach to structural model updating." *Soil Dyn. Earthquake Eng.*, 17(7–8), 495–507.

Katafygiotis, L. S., and Yuen, K.-V. (2001). "Bayesian spectral density approach for modal updating using ambient data." *Earthquake Eng. Struct. Dyn.*, 30(8), 1103–1123.

Krishnaiah, P. R. (1976). "Some recent developments on complex multivariate distributions." *J. Multivariate Anal.*, 6, 1–30.

Kullback, S. (1968). *Information theory and statistics*, Dover, Mineola, N.Y.

Mackay, D. J. C. (1992). "Bayesian interpolation." *Neural Comput.*, 4(3), 415–447.

Mottershead, J. E., and Friswell, M. I. (1993). "Model updating in structural dynamics, a survey." *J. Sound Vib.*, 167, 347–375.

Papadimitriou, C., Beck, J. L., and Katafygiotis, L. S. (1997). "Asymptotic expansions for reliability and moments of uncertain systems." *J. Eng. Mech.*, 123(12), 1219–1229.

Pappa, R. S., Doebling, S. W., and Kholwad, T. D. (2000). "On-line database of vibration-based damage detection experiments." *J. Sound Vib.*, 34(1), 28–33.

Pi, Y. L., and Mickleborough, N. C. (1989). "Modal identification of vibrating structures using ARMA model." *J. Eng. Mech.*, 115(10), 2232–2250.

Quek, S. T., Wang, W. P., and Koh, C. G. (1999). "System identification of linear MDOF structures under ambient excitation." *Earthquake Eng. Struct. Dyn.*, 28, 61–77.

Sanayei, M., McClain, J. A. S., Wadia-Fascetti, S., and Santini, E. M. (1999). "Parameter estimation incorporating modal data and boundary conditions." *J. Struct. Eng.*, 125(9), 1048–1055.

Schwarz, G. (1978). "Estimating the dimension of a model." *Ann. Stat.*, 6(2), 461–464.

- Shi, T., Jones, N. P., and Ellis, J. H. (2000). "Simultaneous estimation of system and input parameters from output measurements." *J. Eng. Mech.*, 126(7), 746–753.
- Sivia, D. S. (1996). *Data analysis: A Bayesian tutorial*, Oxford Science Publications, Oxford, U.K.
- Vanik, M. W., Beck, J. L., and Au, S. K. (2000). "A Bayesian probabilistic approach to structural health monitoring." *J. Eng. Mech.*, 126(7), 738–745.
- Yuen, K.-V. (2002). "Model selection, identification and robust control for dynamical systems." PhD thesis, *Technical Rep. No. EERL 2002-03*, California Institute of Technology, Pasadena, Calif.
- Yuen, K.-V., Beck, J. L., and Katafygiotis, L. S. (2002). "Probabilistic approach for modal identification using non-stationary noisy response measurements only." *Earthquake Eng. Struct. Dyn.*, 31(4), 1007–1023.
- Yuen, K.-V., and Beck, J. L. (2003). "Updating properties of nonlinear dynamical systems with uncertain input." *J. Eng. Mech.*, 129(1), 9–20.
- Yuen, K.-V., and Katafygiotis, L. S. (2001). "Bayesian time-domain approach for modal updating using ambient data." *Probab. Eng. Mech.*, 16(3), 219–231.
- Yuen, K.-V., and Katafygiotis, L. S. (2002). "Bayesian modal updating using complete input and incomplete response noisy measurements." *J. Eng. Mech.*, 128(3), 340–350.
- Yuen, K.-V., Katafygiotis, L. S., and Beck, J. L. (2002). "Spectral density estimation of stochastic vector processes." *Probab. Eng. Mech.*, 17(3), 265–272.