UPDATING MODELS AND THEIR UNCERTAINTIES. I: BAYESIAN STATISTICAL FRAMEWORK

By J. L. Beck,[1] Member, ASCE, and L. S. Katafygiotis,[2] Associate Member, ASCE

ABSTRACT: The problem of updating a structural model and its associated uncertainties by utilizing dynamic response data is addressed using a Bayesian statistical framework that can handle the inherent ill-conditioning and possible nonuniqueness in model updating applications. The objective is not only to give more accurate response predictions for prescribed dynamic loadings but also to provide a quantitative assessment of this accuracy. In the methodology presented, the updated (optimal) models within a chosen class of structural models are the most probable based on the structural data if all the models are equally plausible a priori. The prediction accuracy of the optimal structural models is given by also updating probability models for the prediction error. The precision of the parameter estimates of the optimal structural models, as well as the precision of the optimal prediction-error parameters, can be examined. A large-sample asymptotic expression is given for the updated predictive probability distribution of the uncertain structural response, which is a weighted average of the predictive probability distributions for each optimal model. This predictive distribution can be used to make model predictions despite possible nonuniqueness in the optimal models.

INTRODUCTION

Over the last decade or so, the challenging problem of structural model updating has gained much interest as finite-element modeling capabilities and modal testing have become more mature areas of structural dynamics (Chen and Garba 1980; Torkamani and Ahmadi 1988; Natke 1988; Imregun and Visser 1991; Fritzen and Zhu 1991; Hjelmstad et al. 1992; Mottershead and Friswell 1993; Capecchi and Vestroni 1993). In model updating, a theoretically based finite-element model of a structure is adjusted so that it is somehow "consistent" with dynamic test data from the structure. This updated model should then give more accurate response predictions for prescribed dynamic loadings. In theory, a successful methodology for this process can also be used for global damage detection and assessment by using structural vibration data to continually update the structural model [e.g., *Proc.* (1988)]. In this application, local changes in the stiffness distribution of the model are interpreted as damage at the corresponding location in the structure.

In model updating, linearity is usually assumed and the finite-element model is adjusted so that either the calculated response time histories, frequency response functions, or modal parameters "best" match the corresponding quantities measured or identified from the test data. This type of inverse problem falls within the discipline of system identification (Eykhoff 1974; Goodwin and Payne 1977; *Identification* 1982; Shinozuka et al. 1982; Ljung 1987). There is, however, no well-accepted solution to the model updating problem. This is primarily because of an inherent difficulty: there is a mismatch between the level of information in the detailed theoretical finite-element model, which is of uncertain accuracy, and the relatively sparse information in the "incomplete" set of "noisy" test data. This produces an ill-conditioned and often nonunique inverse problem in updating the theoretical model [e.g., Udwadia (1985); Berman (1989); Janter and Sas (1990)].

For a model updating method to be useful in practice, it should handle the following difficulties:

1. The dynamic test data are incomplete relative to the model complexity needed to produce "physically meaningful" models. For instance, the set of observed degrees of freedom (DOF) is usually a small subset of the set of model DOF due to the limited number of sensors used. Also, the number of identifiable modes of vibration is much less than the number of model DOF because of limited bandwidth in the response.
2. The dynamic test data are contaminated by measurement noise.
3. The chosen class of structural models does not contain the actual structural system; that is, model error always exists and there are no "true" values of the model parameters.
4. The modal parameters controlling the dynamic response are insensitive to changes in the stiffness and mass distributions.
5. To reduce uncertainties and the degree of nonuniqueness, the description of the class of structural models should be able to include as much prior information as possible about the structural system, such as its topological connectivity and experience from dealing with similar systems.

Very few model updating methods can deal successfully with all of these practical problems.

The need for model updating arises because there are always errors associated with the process of constructing a theoretical model of a structure, and this leads to uncertain accuracy in the predicted response. There are many sources of modeling errors, such as variations of the material properties during manufacture; inexact modeling of the material constitutive behavior; uncertainties introduced during the construction process; inexact modeling of the boundary conditions (e.g., there are no exact pinned or fixed joints); errors because of the spatial discretization of the distributed structural system; and unmodeled features such as neglected "nonstructural" components.

Because of these modeling errors, model updating is best tackled as a statistical inference problem. This can be done by embedding the "deterministic" structural models within a class of probability models so that the structural models give a predictable ("systematic") part and the prediction error is modeled as an uncertain ("random") part. Eykhoff (1974), Goodwin and Payne (1977), and Ljung (1987), for example,

[1]Prof. and Executive Ofcr., Depts. of Appl. Mech. and Civ. Engrg., California Inst. of Technol., Pasadena, CA 91125.

[2]Asst. Prof., Dept. of Civ. and Struct. Engrg., Hong Kong Univ. of Sci. and Technol., Clear Water Bay, Kowloon, Hong Kong, China.

Note. Associate Editor: Sami F. Masri. Discussion open until September 1, 1998. Separate discussions should be submitted for the individual papers in this symposium. To extend the closing date one month, a written request must be filed with the ASCE Manager of Journals. The manuscript for this paper was submitted for review and possible publication on May 10, 1996. This paper is part of the ***Journal of Engineering Mechanics***, Vol. 124, No. 4, April, 1998. ©ASCE, ISSN 0733-9399/98/0004-0455–0461/$4.00 + $.50 per page. Paper No. 13233.

emphasize a statistical approach to system identification using a "traditional" statistical framework that could be used to develop model updating methods. Also, powerful statistical techniques have been developed using a Bayesian statistical framework, including the Principle of Maximum (Information) Entropy [e.g., Box and Tiao (1973); Jaynes (1983); Bretthorst (1988); Beck (1989, 1996)]. One important advantage of these Bayesian techniques is that they can be used to develop methods that can handle the nonuniqueness that can arise in model updating because of insufficient data relative to the desired model complexity.

A general Bayesian statistical framework for system identification is presented in this work that can be used for model updating and addresses all of the aforementioned practical difficulties. The objective is not only to give more accurate response predictions for prescribed dynamic loadings but also to provide a quantitative assessment of this accuracy. Therefore, both the structural model and the probability model describing the accuracy of the structural model's predicted response are updated using test data. This is done by applying Bayes' theorem to update probability models that quantify modeling uncertainties.

Since this approach involves the probability of models, which are not "repeatable events," the "traditional" interpretation of probability as a relative frequency of occurrences in the long run is not applicable. However, probability can also be interpreted as a multivalued logic for plausible reasoning under incomplete information (Jeffreys 1961; Cox 1961; Jaynes 1983, 1988). Thus, to quantify the uncertainty within a class of models, a probability distribution is used that gives a measure of how plausible each model is, on the basis of the given information.

In this paper, the Bayesian statistical framework for system identification is first presented and then an asymptotic approximation is described to allow evaluation of the multidimensional integrals that arise for the updated probabilistic predictions of the structural response (Beck 1989, 1996). This theory gives an appropriate statistical framework for properly handling the uncertainties due to ill-conditioning and nonuniqueness associated with model updating.

STATISTICAL SYSTEM IDENTIFICATION

It is assumed that a general mathematical form has been chosen to specify a class of models $\mathcal{M}$ describing the input-output behavior of a structure, but that there are "free" parameters $\mathbf{a} \in S(\mathbf{a}) \subset R^{N_a}$ that need to be assigned values from a region $S(\mathbf{a})$ to choose a particular model $M(\mathbf{a}) \in \mathcal{M}$. The models can be linear or nonlinear, and static or dynamic. They can be expressed as Newtonian equations of motion, in state-space form, or as ARMA models. Beck and Katafygiotis (1992) give an example of a class of models appropriate for updating of a linear finite-element model of a structure for damage detection purposes.

Once the class of structural models is selected, modeling uncertainties of two general types need to be quantified by using probability models. The first type, parameter uncertainty, arises simply because the most appropriate values of the model parameters $\mathbf{a}$ to be used to describe the structure's behavior are uncertain; that is, it is not known a priori which model in the class of models $\mathcal{M}$ is the "best" to describe the structure's behavior. The second type of uncertainty, prediction accuracy, arises because modeling errors lead to an uncertain error in the response predictions given by any model in $\mathcal{M}$. For example, if the class of linear dynamic models is chosen, then there are uncertainties associated with the values of the various parameters that need to be chosen, such as Young's modulus E or the effective moment of inertia of a cracked concrete member. Furthermore, for any given model in the class, the corresponding predicted response will differ from the actual structural response because any mathematical model is only an approximation of the real behavior of a structure, and uncertain modeling errors lead to uncertain accuracy for the model's predictions.

The essence of the statistical system identification approach presented here is to set up probability models describing these two types of uncertainties, which are then updated by applying Bayes' theorem to the available data.

Basic Probability Models

The statistical framework is presented based on using dynamic test data in the time domain, but it can be adapted to frequency-domain or modal-parameter data, or even to static test data. The specified class of models $\mathcal{M}$ must therefore provide a functional relationship between the model output vector $\mathbf{q}(n; \mathbf{a}) \in R^{N_d}$ at time $t_n = n\Delta t$, where Δt is a prescribed sampling interval, and the system input $Z_1^n = \{\mathbf{z}(m) \in R^{N_I}: m = 1, 2, \ldots, n\}$ up to this time

$$\mathbf{q}(n; \mathbf{a}) = \mathbf{q}(n; \mathbf{a}, Z_1^n, \mathcal{M}) \tag{1}$$

For simplicity, it is assumed here that the *complete* system input can be observed; otherwise a stochastic model of the system input must be constructed to cover unobserved or "noisy" inputs. Although the model output is only required at discrete times, the theoretical model that is the underlying basis of the class $\mathcal{M}$ can be in any form, such as a system of continuous-time differential equations. The assumption of discrete-time input is not critical but it is typical of the data usually available, so an interpolation scheme is needed in the case of continuous-time models. In the following, the dependence of $\mathbf{q}(n; \mathbf{a})$ and other model quantities on the input Z_1^n and the theoretical model $\mathcal{M}$ will be suppressed in the notation.

The first step in embedding the deterministic class of models $\mathcal{M}$ in a class of probability models is to define the prediction error $\mathbf{e}(n; \mathbf{a}) \in R^{N_d}$ as the difference between the model output and the system output, so that $\mathbf{q} + \mathbf{e}$ is equal to the system output at the N_d DOF in the structure corresponding to those in the model. Since it is common in practice to not measure all DOF, let $\mathbf{S}_0\mathbf{q} \in R^{N_0}$ and $\mathbf{S}_u\mathbf{q} \in R^{N_d - N_0}$ correspond to the model output at the observed and unobserved DOF, respectively, where $\mathbf{S}_0 \in R^{N_0 \times N_d}$ and $\mathbf{S}_u \in R^{(N_d - N_0) \times N_d}$ are selection matrices with only one nonzero element, equal to unity, in each row. Let $\mathbf{y}(n) \in R^{N_0}$ and $\mathbf{x}(n) \in R^{N_d - N_0}$ denote the system output at time t_n at the observed and unobserved DOF, respectively; then

$$\mathbf{y}(n) = \mathbf{S}_0[\mathbf{q}(n; \mathbf{a}) + \mathbf{e}(n; \mathbf{a})]; \quad \mathbf{x}(n) = \mathbf{S}_u[\mathbf{q}(n; \mathbf{a}) + \mathbf{e}(n; \mathbf{a})] \tag{2a,b}$$

The *observed* prediction error, which is the difference between the model and *measured* system outputs, is a combined effect of measurement noise and prediction error $\mathbf{e}(n; \mathbf{a})$. Although these concepts could be separately modeled, for modern instrumentation the measurement noise is usually negligible compared with the prediction error, and so the latter is approximately equal to its observed values. Therefore, the measured system output is essentially equivalent to the actual system output.

To describe the uncertainty in the prediction error, a class of probability models $\mathcal{P}$ is chosen, parameterized by the prediction-error parameters $\boldsymbol{\sigma} \in R^{N_\sigma}$, which prescribes a function h_M giving the probability density function (PDF) of a sequence of M prediction errors for arbitrary M; that is,

$$p(E_1^M(\mathbf{a})|\boldsymbol{\sigma}, \mathcal{P}) = h_M(\mathbf{e}(1; \mathbf{a}), \ldots, \mathbf{e}(M; \mathbf{a}); \boldsymbol{\sigma}) \tag{3}$$

Here $E_1^M(\mathbf{a})$ denotes the sequence of prediction errors $\{\mathbf{e}(n; \mathbf{a}); n = 1, \ldots, M\}$, for model $M(\mathbf{a})$ in $\mathcal{M}$. Rather than present the case for general h_M, we choose the class $\mathcal{P}$ so that $\mathbf{e}(n; \mathbf{a})$ is

a zero-mean stationary Gaussian white-noise stochastic process with independent components. This choice implies that knowing the prediction errors at other times, or at other locations within the structure, does not influence our uncertainty concerning the value of the prediction error at a specified time and location. This choice for the class $\mathcal{P}$ also corresponds to a probability distribution h_M in (3), which is given by the principle of maximum (information) entropy under the condition of zero means and finite variances (Jaynes 1968, 1983). This means that under the latter conditions, the Gaussian PDF for h_M gives the maximal uncertainty that E_1^M in (3) can have, just as the uniform distribution gives the maximal uncertainty for an uncertain variable with a finite range. Therefore, if any other choice for h_M is made, there should be strong grounds for the implied reduction in uncertainty in the predictions. Finally, the variances of the prediction errors at all DOF are assumed to be equal, so the vector $\boldsymbol{\sigma} \in R^{N_\sigma}$, which needs to be specified to choose a particular probability model from the class $\mathcal{P}$, reduces to a single parameter σ; that is, $N_\sigma = 1$.

The selection of the classes $\mathcal{M}$ and $\mathcal{P}$ allows a class of probability models $\mathcal{M}_P$ to be defined, parameterized by $\boldsymbol{\alpha} = [\boldsymbol{\alpha}^T, \sigma]^T \in S(\mathbf{a}) \subset R^{N_\alpha}$ where $N_\alpha = N_a + 1$, which prescribes a function g_M giving the PDF for the observed and unobserved system output sequences, $Y_1^M = \{\mathbf{y}(n) \in R^{N_0}: n = 1, 2, \ldots, M\}$ and $X_1^M = \{\mathbf{x}(n) \in R^{N_d - N_0}: n = 1, 2, \ldots, M\}$, respectively

$$p(Y_1^M, X_1^M|\boldsymbol{\alpha}, Z_1^M, \mathcal{M}_P) = g_M[\mathbf{y}(1), \ldots, \mathbf{y}(M), \mathbf{x}(1), \ldots, \mathbf{x}(M); \boldsymbol{\alpha}, Z_1^M]$$
$$= \frac{1}{(\sqrt{2\pi}\sigma)^{MN_d}} \exp\left[-\frac{1}{2\sigma^2}\sum_{n=1}^{M} \|\mathbf{y}(n) - \mathbf{S}_0\mathbf{q}(n; \mathbf{a})\|^2\right]$$
$$\times \exp\left[-\frac{1}{2\sigma^2}\sum_{n=1}^{M} \|\mathbf{x}(n) - \mathbf{S}_u\mathbf{q}(n; \mathbf{a})\|^2\right] \tag{4}$$

where g_M is determined using (2) in (3) along with the assumed form of h_M, and $\|.\|$ denotes the standard Euclidean 2-norm.

To account for the uncertainty in the values for the parameters $\boldsymbol{\alpha}$, the specification of $\mathcal{M}_P$ also involves choosing an initial ("prior") PDF $\pi(\boldsymbol{\alpha})$ over the set $S(\boldsymbol{\alpha})$ of possible parameter values, that is

$$p(\boldsymbol{\alpha}|\mathcal{M}_P) = \pi(\boldsymbol{\alpha}) \tag{5}$$

The choice for $\pi(\boldsymbol{\alpha})$ allows engineering judgment about the plausibilities of the different models to be incorporated. It can be chosen as a smooth, slowly varying PDF that is mathematically convenient and roughly reflects the engineer's judgment. Beck and Katafygiotis (1992) give an example that is appropriate for model updating for damage detection purposes. Alternatively, the initial PDF can be chosen by the principle of maximum entropy if the information to be utilized is "testable" (Jaynes 1968). Dependence of the initial PDF on the input Z_1^M can be introduced in (5). For example, this might be done when using linear models to account for the fact that the nonlinearities in structures typically cause them to become more flexible for stronger excitation, although the final results are often insensitive to the choice of the initial PDF.

Summarizing, specification of the class $\mathcal{M}_P$ implies specification of two PDFs, which are the basic probability models for statistical system identification: $g_M(Y_1^M, X_1^M; \boldsymbol{\alpha}, Z_1^M)$ for the system output, which relates to the uncertainty in the prediction accuracy of the models in $\mathcal{M}$; and $\pi(\boldsymbol{\alpha})$ for the model parameters $\mathbf{a}$ and σ, which relates to the uncertainty in the deterministic models in $\mathcal{M}$ and the prediction-error probability models in $\mathcal{P}$.

Initial and Updated Predictive Probability Models

The PDF in (4) gives predictions for the system output for a particular probability model in class $\mathcal{M}_P$, which is specified by the value of the structural model and prediction-error parameters $\boldsymbol{\alpha}$. Prior to utilizing data, the best choice for making predictions of the system output using the assumed class of probability models is the initial predictive PDF

$$p(Y_1^M, X_1^M|Z_1^M, \mathcal{M}_P) = \int_{S(\boldsymbol{\alpha})} p(Y_1^M, X_1^M|\boldsymbol{\alpha}, Z_1^M, \mathcal{M}_P)p(\boldsymbol{\alpha}|\mathcal{M}_P)\, d\boldsymbol{\alpha}$$
$$= \int_{S(\boldsymbol{\alpha})} g_M(Y_1^M, X_1^M; \boldsymbol{\alpha}, Z_1^M)\pi(\boldsymbol{\alpha})\, d\boldsymbol{\alpha} \tag{6}$$

which is readily derived from the axioms of probability (the "total probability theorem"). Eq. (6) gives a predictive PDF using the whole class $\mathcal{M}_P$ as a weighted average of the predictive PDFs for each model in $\mathcal{M}_P$, with weights given by their initial probabilities.

Let $\mathcal{D}_N$ denote a set of observed time history data from the structural system at N discrete times $t_n = n\Delta t$, where Δt is the sampling interval. These data are assumed to consist of the sampled history $\hat{Z}_1^N = \{\hat{\mathbf{z}}(n) \in R^{N_I}: n = 1, 2, \ldots, N\}$ for N_I inputs and the sampled output history $\hat{Y}_1^N = \{\hat{\mathbf{y}}(n) \in R^{N_0}: n = 1, 2, \ldots, N\}$, which is the measured response at the N_0 observed DOF of the structure. Usually, the inputs are exciting forces or boundary accelerations and the measured response quantities are accelerations at the observed DOF.

Based on the new information in the data $\mathcal{D}_N$, the predictive PDF in (6) can be replaced by an updated predictive PDF by using the total probability theorem again

$$p(Y_{N+1}^M, X_1^M|\mathcal{D}_N, Z_{N+1}^M, \mathcal{M}_P)$$
$$= \int_{S(\boldsymbol{\alpha})} p(Y_{N+1}^M, X_1^M|\boldsymbol{\alpha}, \mathcal{D}_N, Z_{N+1}^M, \mathcal{M}_P)p(\boldsymbol{\alpha}|\mathcal{D}_N, \mathcal{M}_P)\, d\boldsymbol{\alpha}$$
$$= c\int_{S(\boldsymbol{\alpha})} p(\hat{Y}_1^N, Y_{N+1}^M, X_1^M|\boldsymbol{\alpha}, \hat{Z}_1^N, Z_{N+1}^M, \mathcal{M}_P)p(\boldsymbol{\alpha}|\mathcal{M}_P)\, d\boldsymbol{\alpha}$$
$$= c\int_{S(\boldsymbol{\alpha})} g_M(\hat{Y}_1^N, Y_{N+1}^M, X_1^M; \boldsymbol{\alpha}, \hat{Z}_1^N, Z_{N+1}^M)\pi(\boldsymbol{\alpha})\, d\boldsymbol{\alpha} \tag{7}$$

where now the system output at the observed DOF is only predicted for the next $(M - N)$ sampling times for a prescribed future input Z_{N+1}^M, since this output is already specified for the first N samples by $\mathcal{D}_N$. In (7), Bayes' theorem is used to calculate the updated ("posterior") PDF from the initial PDF

$$p(\boldsymbol{\alpha}|\mathcal{D}_N, \mathcal{M}_P) = cp(\hat{Y}_1^N|\boldsymbol{\alpha}, \hat{Z}_1^N, \mathcal{M}_P)p(\boldsymbol{\alpha}|\mathcal{M}_P) = cf_N(\hat{Y}_1^N; \boldsymbol{\alpha}, \hat{Z}_1^N)\pi(\boldsymbol{\alpha}) \tag{8}$$

where

$$f_N(\hat{Y}_1^N; \boldsymbol{\alpha}, \hat{Z}_1^N) = \int_{S(X_1^N)} g_N(\hat{Y}_1^N, X_1^N; \boldsymbol{\alpha}, \hat{Z}_1^N)\, d\mathbf{x}(1) \cdots d\mathbf{x}(N)$$
$$= \frac{1}{(\sqrt{2\pi}\sigma)^{NN_0}} \exp\left[-\frac{1}{2\sigma^2}\sum_{n=1}^{N} \|\hat{\mathbf{y}}(n) - \mathbf{S}_0\mathbf{q}(n; \mathbf{a})\|^2\right] \tag{9}$$

and where the normalizing constant c can be evaluated using (8)

$$c^{-1} = p(\hat{Y}_1^N|\hat{Z}_1^N, \mathcal{M}_P) = \int_{S(\boldsymbol{\alpha})} f_N(\hat{Y}_1^N; \boldsymbol{\alpha}, \hat{Z}_1^N)\pi(\boldsymbol{\alpha})\, d\boldsymbol{\alpha} \tag{10}$$

The updated predictive PDF in (7) using the whole class $\mathcal{M}_P$ and the data $\mathcal{D}_N$ can be viewed as a weighted average of the predictive PDFs for each model in $\mathcal{M}_P$, like the initial case in (6), except that now the weights are given by the updated probabilities for each model.

The difficulty with the solution given in (6) for the initial predictive PDF for the system output is that the multidimensional integral can not be evaluated analytically, or numerically if the dimension of the parameters space $S(\boldsymbol{\alpha})$ is too high (say, greater than 5). The situation can be even worse for numerical integration of (7) for the updated predictive PDF, since the function $p(\boldsymbol{\alpha}|\mathcal{D}_N, \mathcal{M}_P)$ is sharply peaked at one or more locations in $S(\boldsymbol{\alpha})$ if the number of sampling items N is not small. This difficulty is overcome by developing an asymptotic approach, which not only provides approximations for the required integrals, but also provides insight into the solution which is not revealed by the integrals themselves.

Optimal Parameters and System Identifiability

Define the optimal parameters $\hat{\boldsymbol{\alpha}} = [\hat{\mathbf{a}}^T, \hat{\sigma}]^T$ for $\mathcal{M}_P$ given data $\mathcal{D}_N$ to be values of the parameters $\boldsymbol{\alpha} \in S(\boldsymbol{\alpha})$ that globally maximize $f_N(\hat{Y}_1^N, \boldsymbol{\alpha}, \hat{Z}_1^N)$ in (9), that is,

$$f_N(\hat{Y}_1^N; \hat{\boldsymbol{\alpha}}, \hat{Z}_1^N) = \max_{\boldsymbol{\alpha} \in S(\boldsymbol{\alpha})} f_N(\hat{Y}_1^N; \boldsymbol{\alpha}, \hat{Z}_1^N) \tag{11}$$

The optimal parameters are equivalent to the maximum likelihood estimates used in "traditional" statistical theory based on the interpretation of probability as relative frequencies of events in the "long run." From (8), they are also the most probable values based on the data in the case that all parameter values are initially equally plausible over a region containing the optimal parameters.

The optimal parameters arise naturally in constructing an asymptotic approximation for the updated predictive distribution in (7), as shown in the next section. In preparation for this derivation, the concept of system identifiability is introduced. First, define an optimal probability model $M_P(\hat{\boldsymbol{\alpha}})$ for given data $\mathcal{D}_N$ to be a model in $\mathcal{M}_P$ given by the optimal parameter $\hat{\boldsymbol{\alpha}} = [\hat{\mathbf{a}}^T, \hat{\sigma}]^T$. Let $S_{\text{opt}}(\mathcal{M}_P; \mathcal{D}_N) \subset \mathcal{M}_P$ denote the set of all optimal models in the class $\mathcal{M}_P$ given $\mathcal{D}_N$, and let $S_{\text{opt}}(\boldsymbol{\alpha}; \mathcal{D}_N) \subset S(\boldsymbol{\alpha})$ denote the set of all corresponding optimal parameters; then:

- A parameter α_j of $\boldsymbol{\alpha}$ is globally system identifiable for class $\mathcal{M}_P$ and data $\mathcal{D}_N$ if $S_{\text{opt}}(\boldsymbol{\alpha}; \mathcal{D}_N)$ contains only one optimal parameter, or, if not, then

$$\hat{\boldsymbol{\alpha}}^{(1)}, \hat{\boldsymbol{\alpha}}^{(2)} \in S_{\text{opt}}(\boldsymbol{\alpha}; \mathcal{D}_N) \Rightarrow \hat{\alpha}_j^{(1)} = \hat{\alpha}_j^{(2)} \tag{12}$$

- A parameter α_j of $\boldsymbol{\alpha}$ is system identifiable for class $\mathcal{M}_P$ and data $\mathcal{D}_N$ if there exists a positive number ε_j such that

$$\hat{\boldsymbol{\alpha}}^{(1)}, \hat{\boldsymbol{\alpha}}^{(2)} \in S_{\text{opt}}(\boldsymbol{\alpha}; \mathcal{D}_N) \Rightarrow |\hat{\alpha}_j^{(1)} - \hat{\alpha}_j^{(2)}| > \varepsilon_j \quad \text{or} \quad \hat{\alpha}_j^{(1)} = \hat{\alpha}_j^{(2)} \tag{13}$$

- A parameter α_j of $\boldsymbol{\alpha}$ is locally system identifiable for class $\mathcal{M}_P$ and data $\mathcal{D}_N$ if it is system identifiable but not globally system identifiable.

The above definitions can be extended; e.g., the model parameter vector $\boldsymbol{\alpha}$ is (globally) system identifiable under $\mathcal{D}_N$ if all of its elements are (globally) system identifiable, whereas $\boldsymbol{\alpha}$ is locally system identifiable if it is system identifiable but at least one of its elements is only locally system identifiable. The class of models $\mathcal{M}_P$ is system identifiable under $\mathcal{D}_N$ if the model parameter vector $\boldsymbol{\alpha}$ is system identifiable, with similar definitions for the locally and globally identifiable cases. If $\boldsymbol{\alpha}$, and so $\mathcal{M}_P$, is system identifiable under $\mathcal{D}_N$, and $S(\boldsymbol{\alpha})$ is a bounded set, then there is only a finite number K of optimal parameters for $\mathcal{M}_P$ given $\mathcal{D}_N$, with $K = 1$ corresponding to global system identifiability of $\mathcal{M}_P$.

Asymptotic Approximation

Suppose there is a finite number K of optimal parameters, and denote them by $\hat{\boldsymbol{\alpha}}^{(k)}$, $k = 1, \ldots, K$. By expanding $\ln f_N(\hat{Y}_1^N; \boldsymbol{\alpha}, \hat{Z}_1^N)$ in a second-order Taylor series about an optimal parameter $\hat{\boldsymbol{\alpha}}$, the following local approximation is derived:

$$f_N(\hat{Y}_1^N; \boldsymbol{\alpha}, \hat{Z}_1^N) \simeq f_N(\hat{Y}_1^N; \hat{\boldsymbol{\alpha}}, \hat{Z}_1^N) \exp\left(-\frac{1}{2}[\boldsymbol{\alpha} - \hat{\boldsymbol{\alpha}}]^T \mathbf{A}_N(\hat{\boldsymbol{\alpha}})[\boldsymbol{\alpha} - \hat{\boldsymbol{\alpha}}]\right) \tag{14}$$

where the elements of the $N_\alpha \times N_\alpha$ Hessian matrix $\mathbf{A}_N(\boldsymbol{\alpha})$ are given by

$$[\mathbf{A}_N(\boldsymbol{\alpha})]_{ij} = -\frac{\partial^2 \ln f_N(\hat{Y}_1^N; \boldsymbol{\alpha}, \hat{Z}_1^N)}{\partial \alpha_i \partial \alpha_j} \tag{15}$$

The order of the elements of $\mathbf{A}_N$ is $\mathbb{O}(N)$ and, therefore, for a large number N of sampling times, which is usually the case with dynamic test data, $f_N(\hat{Y}_1^N; \boldsymbol{\alpha}, \hat{Z}_1^N)$ is very peaked at each optimal parameter $\hat{\boldsymbol{\alpha}}^{(k)}$. Therefore, Laplace's method for asymptotic expansions can be applied to obtain the following approximation for the first integral in (7) (see Appendix I):

$$p(Y_{N+1}^M, X_1^M|\mathcal{D}_N, Z_{N+1}^M, \mathcal{M}_P) = \sum_{k=1}^{K} w_k p(Y_{N+1}^M, X_1^M|\hat{\boldsymbol{\alpha}}^{(k)}, \mathcal{D}_N, Z_{N+1}^M, \mathcal{M}_P)[1 + \mathbb{O}(N^{-1})] \tag{16}$$

where for an optimal parameter $\hat{\boldsymbol{\alpha}}$

$$\begin{aligned} p(Y_{N+1}^M, X_1^M|\hat{\boldsymbol{\alpha}}, \mathcal{D}_N, Z_{N+1}^M, \mathcal{M}_P) &= \frac{p(\hat{Y}_1^N, Y_{N+1}^M, X_1^M|\hat{\boldsymbol{\alpha}}, \hat{Z}_1^N, Z_{N+1}^M, \mathcal{M}_P)}{p(\hat{Y}_1^N|\hat{\boldsymbol{\alpha}}, \hat{Z}_1^N, \mathcal{M}_P)} \\ &= \frac{g_M(\hat{Y}_1^N, Y_{N+1}^M, X_1^M; \hat{\boldsymbol{\alpha}}, \hat{Z}_1^N, Z_{N+1}^M)}{f_N(\hat{Y}_1^N; \hat{\boldsymbol{\alpha}}, \hat{Z}_1^N)} \\ &= \frac{1}{(\sqrt{2\pi}\hat{\sigma})^{(M-N)N_0}} \exp\left[-\frac{1}{2\hat{\sigma}^2}\sum_{n=N+1}^{M} \|\mathbf{y}(n) - \mathbf{S}_0\mathbf{q}(n; \hat{\mathbf{a}})\|^2\right] \\ &\times \frac{1}{(\sqrt{2\pi}\hat{\sigma})^{M(N_d-N_0)}} \exp\left[-\frac{1}{2\hat{\sigma}^2}\sum_{n=1}^{M} \|\mathbf{x}(n) - \mathbf{S}_u\mathbf{q}(n; \hat{\mathbf{a}})\|^2\right] \end{aligned} \tag{17}$$

and the weighting coefficient w_k corresponding to optimal parameter $\hat{\boldsymbol{\alpha}}^{(k)}$ is given by

$$w_k = \frac{w_k'}{\sum_{k=1}^{K} w_k'}, \quad w_k' = \pi(\hat{\boldsymbol{\alpha}}^{(k)})|\mathbf{A}_N(\hat{\boldsymbol{\alpha}}^{(k)})|^{-1/2} \tag{18}$$

The continuous weighted average for the updated predictive PDF in (7) is therefore approximated by a discrete weighted average of the predictive PDFs for each optimal model in class $\mathcal{M}_P$. This asymptotic result is valid when there is a finite number of optimal parameters, so it is necessary that the class $\mathcal{M}_P$ is system identifiable under data $\mathcal{D}_N$. System identifiability is also sufficient if the parameter space $S(\boldsymbol{\alpha})$ is bounded.

Eq. (18) shows that for the asymptotic approximation of the updated predictive PDF given by (16), the initial PDF $\pi(\boldsymbol{\alpha})$ is not required over the whole domain $S(\boldsymbol{\alpha})$. Instead, only the relative values at the optimal parameters $\hat{\boldsymbol{\alpha}}^{(k)}$ need be specified. This allows engineering judgment to be incorporated since some of the optimal models based on the data may not be very plausible from an engineering point of view. If $\mathcal{M}_P$ is globally system identifiable under $\mathcal{D}_N$, that is, $K = 1$ in (18), the initial PDF does not enter at all. This is the well-known Bayesian result that a large amount of data will overwhelm the effect of the chosen initial PDF and one can proceed using the predictive PDF for the unique optimal model (Lindley 1965). The above result shows, however, that this is only true in the case of global system identifiability.

Since $f_N(\hat{Y}_1^N; \boldsymbol{\alpha}, \hat{Z}_1^N)$ is very peaked at each optimal parameter $\hat{\boldsymbol{\alpha}}$, the updated PDF $p(\boldsymbol{\alpha}|\mathcal{D}_N, \mathcal{M}_P)$ is too if the initial PDF

is a smooth and slowly varying function of $\boldsymbol{\alpha}$. It can be deduced from (8) and (14) that in this case $p(\boldsymbol{\alpha}|\mathcal{D}_N, \mathcal{M}_P)$ behaves locally about each optimal parameter $\hat{\boldsymbol{\alpha}}$ like a multidimensional Gaussian distribution with means $\hat{\boldsymbol{\alpha}}$ and an $N_\alpha \times N_\alpha$ covariance matrix $\mathbf{A}_N^{-1}(\hat{\boldsymbol{\alpha}})$. The optimal models may therefore be interpreted as locally most probable models within the class $\mathcal{M}_P$ based on the data $\mathcal{D}_N$, and the covariance matrix can be used to examine how precisely the optimal parameters are identified by the data. Furthermore, it can be deduced from (18) that the predictive PDF in (16) for each of the optimal models is weighted in proportion to the volume of the updated PDF $p(\boldsymbol{\alpha}|\mathcal{D}_N, \mathcal{M}_P)$ under its Gaussian-shaped peak positioned at the corresponding optimal parameter. In fact, a more accurate version of (16) would include all probability models for which $p(\boldsymbol{\alpha}|\mathcal{D}_N, \mathcal{M}_P)$ has a local maximum, but as N increases, the contributions of the optimal models eventually dominate.

The optimal parameters $\hat{\boldsymbol{\alpha}} = [\hat{\mathbf{a}}^T, \hat{\sigma}]^T$ are determined by maximizing f_N with respect to $\boldsymbol{\alpha}$. For fixed model parameters $\mathbf{a}$, maximizing $\ln f_N(\hat{Y}_1^N, \boldsymbol{\alpha}, \hat{Z}_1^N)$ with respect to σ leads to

$$\hat{\sigma}^2(\mathbf{a}) = \frac{1}{N_0 N} \sum_{n=1}^{N} \|\hat{\mathbf{y}}(n) - \mathbf{S}_0 \mathbf{q}(n; \mathbf{a})\|^2 = J(\mathbf{a}) \tag{19}$$

Therefore, the most probable standard deviation $\hat{\sigma}(\mathbf{a})$, for given model parameters $\mathbf{a}$, is equal to the RMS of the prediction errors at all observed DOF. Obviously, the condition for the overall most probable variance is given by (19) when $\mathbf{a} = \hat{\mathbf{a}}$. Substituting (19) into (9),

$$f_N(\hat{Y}_1^N, \mathbf{a}, \hat{\sigma}(\mathbf{a}), \hat{Z}_1^N) = [2\pi e J(\mathbf{a})]^{-N_0 N/2} \tag{20}$$

Therefore, the optimal structural model parameters $\hat{\mathbf{a}}$ are determined by minimizing $J(\underline{a})$ in (19), which corresponds to the usual least-squares output-error method for estimation of model parameters. Although least-squares parameter-estimation methods can be accepted in their own right, by viewing them within the Bayesian statistical framework, one can determine the prediction accuracy of the optimal models, the precision of the parameter estimates of the structural models and prediction-error probability models, and the updated predictive PDF even when there are multiple optimal models.

In structural applications, the model output $\mathbf{q}(n; \mathbf{a})$ involved in (19) is a nonlinear function of the parameters, even if the model has linear dynamics, and so the minimization must be done numerically by an iterative optimization algorithm. There may be multiple optimal model parameters because $J(\mathbf{a})$ might attain its minimum at more than one value $\hat{\mathbf{a}}$ in $S(\mathbf{a})$, the set of permissible values of $\mathbf{a}$. The task of finding all the global minima of the nonconvex function $J(\mathbf{a})$ is nontrivial. Powerful methods have been developed, however, to treat this problem [e.g., Katafygiotis and Beck (1998); Yang and Beck (1998)]. Notice that the optimal prediction-error variance $\hat{\sigma}^2 = \hat{\sigma}^2(\hat{\mathbf{a}}) = J(\hat{\mathbf{a}})$ is uniquely determined even though $\hat{\underline{a}}$ may not be, since by definition $J(\hat{\mathbf{a}})$ is the global minimum of $J(\mathbf{a})$, and so each optimal structural model has the same optimal prediction-error variance.

It is of interest to note that the exact updated predictive PDF in (7) does not require parameter estimation in the usual sense, but that the asymptotic approximation in (16) does. The problem is converted from an integration over all possible probability models in the class $\mathcal{M}_P$, which is computationally prohibitive when there are more than just a handful of parameters, to a nontrivial optimization problem, which can at least be tackled numerically. Since the optimal parameters are also Fisher's MLEs (maximum likelihood estimates), the well-known MLE approach to parameter estimation can be justified in terms of an asymptotic approximation to the exact Bayesian solution for response predictions, but, in addition, the asymptotic result in (16) shows how to proceed when there are multiple MLEs, while it is not clear how to proceed in this situation in "traditional" statistics.

The Hessian matrix $\mathbf{A}_N(\hat{\boldsymbol{\alpha}})$ needed to evaluate the weights in (18) is block diagonal with one block being an $N_a \times N_a$ matrix $\mathbf{B}_N(\hat{\mathbf{a}})$ corresponding to the optimal structural model parameters $\hat{\mathbf{a}}$, and the other block matrix $\mathbf{C}_N(\hat{\sigma})$ being simply a scalar corresponding to the prediction-error parameter $\hat{\sigma}$, which is the same for all optimal parameters $\hat{\mathbf{a}}$. Therefore, the weights in (18) depend only on the determinant of the matrix $\mathbf{B}_N$ with elements given by

$$[\mathbf{B}_N(\hat{\mathbf{a}}^{(k)})]_{ij} = \frac{NN_0}{2\hat{\sigma}^2}\left[\frac{\partial^2 J(\mathbf{a})}{\partial a_i \partial a_j}\right]_{\hat{\mathbf{a}}^{(k)}} \simeq \frac{1}{\hat{\sigma}^2}\sum_{n=1}^{N}\left[\frac{\partial \mathbf{q}(n; \mathbf{a})^T}{\partial a_i} S_0^T S_0 \frac{\partial \mathbf{q}(n; \mathbf{a})}{\partial a_j}\right]_{\hat{\mathbf{a}}^{(k)}} \tag{21}$$

where $J(\mathbf{a})$ is given by (19), $\hat{\sigma}^2 = \hat{\sigma}^2(\hat{\mathbf{a}}) = J(\hat{\mathbf{a}})$ and the last approximation is valid if $J(\hat{\mathbf{a}})$ is sufficiently small. This approximation implies that the weights in (18) are not sensitive to the output data $\hat{Y}_1^N$ other than through their influence on the optimal parameters $\hat{\mathbf{a}}^{(k)}$ since the common factor $1/\hat{\sigma}^2$ cancels in evaluating the w_k. Also, because of the block diagonal structure of the Hessian matrix $\mathbf{A}_N(\hat{\boldsymbol{\alpha}}^{(k)})$, the inverse of $\mathbf{B}_N(\hat{\mathbf{a}}^{(k)})$ is the covariance matrix for the local Gaussian behavior of the updated PDF $p(\mathbf{a}|\mathcal{D}_N, \mathcal{M}_P)$ centered at $\hat{\mathbf{a}}^{(k)}$ in the structural model parameter space.

There is another important simplification proved in Appendix I that can be made to the expression (18) for the weighting coefficients. The submatrix $\mathbf{B}_N(\hat{\mathbf{a}})$ in (21) can be reduced even further if some of the structural model parameters are globally system identifiable. In this case, the elements in (21) containing the partial derivatives involving a globally system identifiable parameter a_j can be dropped. Thus, to determine the weighting coefficients in (18), the only elements of $\mathbf{A}_N(\mathbf{a}, \boldsymbol{\sigma})$ needed in order to evaluate the determinant are those given by (21) that involve partial differentiation only with respect to structural model parameters a_j, which are locally system identifiable. Furthermore, the initial PDF $\pi(\boldsymbol{\alpha})$ in the expression (18) for the weighting coefficients can be replaced by the initial PDF for the locally system identifiable model parameters conditional on the globally identifiable parameters.

To summarize, the asymptotic approximation in (16) can be used for updated predictions of the system output. For prediction of the output at only the unobserved DOF for times t_n, $n = 1, \ldots, M$, (16) can be integrated with respect to Y_{N+1}^M to get the PDF

$$p(X_1^M|\mathcal{D}_N, Z_{N+1}^M, \mathcal{M}_P) = \sum_{k=1}^{K} w_k p(X_1^M|\hat{\boldsymbol{\alpha}}^{(k)}, \mathcal{D}_N, Z_{N+1}^M, \mathcal{M}_P)[1 + \mathcal{O}(N^{-1})] \tag{22}$$

where

$$p(X_1^M|\hat{\boldsymbol{\alpha}}, \mathcal{D}_N, Z_{N+1}^M, \mathcal{M}_P) = \frac{1}{(\sqrt{2\pi}\hat{\sigma})^{M(N_d - N_0)}} \exp\left[-\frac{1}{2\hat{\sigma}^2}\sum_{n=1}^{M}\|\mathbf{x}(n) - \mathbf{S}_u \mathbf{q}(n; \hat{\mathbf{a}})\|^2\right] \tag{23}$$

Of course, if $M < N + 1$, then Z_{N+1}^M is irrelevant and can be ignored in the conditioning information. For prediction of the output at only the observed DOF for times t_n, $n = N + 1, \ldots, M$, (16) can be integrated with respect to X_1^M to get the PDF

$$p(Y_{N+1}^M|\mathcal{D}_N, Z_{N+1}^M, \mathcal{M}_P) = \sum_{k=1}^{K} w_k p(Y_{N+1}^M|\hat{\boldsymbol{\alpha}}^{(k)}, \mathcal{D}_N, Z_{N+1}^M, \mathcal{M}_P)[1 + \mathcal{O}(N^{-1})] \tag{24}$$

where

$$p(Y_{N+1}^M|\hat{\boldsymbol{\alpha}}, \mathcal{D}_N, Z_{N+1}^M, \mathcal{M}_P)$$

$$= \frac{1}{(\sqrt{2\pi}\hat{\sigma})^{(M-N)N_0}} \exp\left[-\frac{1}{2\hat{\sigma}^2}\sum_{n=N+1}^{M} \|\mathbf{y}(n) - \mathbf{S}_0\mathbf{q}(n; \hat{\mathbf{a}})\|^2\right] \quad (25)$$

If $\mathcal{M}_P$ is globally system identifiable under $\mathcal{D}_N$ so that $K = 1$ in these equations, then the predictive PDFs for the output at the observed and unobserved DOF at time t_n are Gaussian distributions. The mean (most probable) value is equal to the output of the most probable model $M(\hat{\mathbf{a}}) \in \mathcal{M}$ at that time and the variance is $\hat{\sigma}^2$, which describes the uncertainty in the predictions of this optimal model. If $\mathcal{M}_P$ is only locally system identifiable under $\mathcal{D}_N$ so that $K > 1$ in these equations, the asymptotic predictive PDF is not Gaussian, even though each term in the sum in (22) and (24) is a Gaussian PDF [see (23) and (25)]. In this case, the mean and variance of the predicted system output $s_i(n)$ at an observed or unobserved DOF i at time t_n is given by

$$E[s_i(n)] \simeq \sum_{k=1}^{K} w_k q_i(n; \hat{\mathbf{a}}^{(k)})$$

$$\text{Var}[s_i(n)] \simeq \hat{\sigma}^2 + \sum_{k=1}^{K} w_k q_i^2(n; \hat{\mathbf{a}}^{(k)}) - E[s_i(n)]^2 \quad (26)$$

Therefore, the mean system output is a weighted sum of the outputs of each optimal structural model and the variance for the system output is the sum of the prediction-error variance and a contribution due to the lack of uniqueness in these models.

CONCLUSIONS

We have presented a general Bayesian statistical framework that can be used for model updating. For a large number of available data points and for a relatively small number of uncertain structural and prediction-error parameters, the updated probability distribution for these model parameters using Bayes' theorem is sharply peaked at isolated optimal values. These parameter values are therefore much more plausible than any other values. This leads to a natural criterion for choosing the "best" model parameters as the optimal ones, which are also maximum likelihood estimates, that is, least-squares estimates for the case presented here. Furthermore, it is asymptotically correct for response predictions to use only the optimal models corresponding to these optimal parameters in an appropriate weighted average of their predictive PDFs. This result is very important, since the high-dimensional integrations that are required to calculate the full Bayesian predictive PDF for the structural response can become computationally prohibitive, but the asymptotic research shows that they can be replaced by a weighted sum over the predictive probability distributions for each optimal model, assuming their number is finite. This produces a computationally feasible, rigorous treatment of the nonuniqueness that can arise in inverse problems such as model updating. It also means that the well-known least-squares (maximum likelihood) approach to parameter estimation can be viewed as the first step of an asymptotic approximation to the full Bayesian solution for response predictions, with the advantage that the asymptotic approximation shows how to proceed when there are multiple maximum likelihood estimates.

The implementation of these asymptotic results requires solution of the problem of finding the set of all optimal parameters for a given class of models. This problem is equivalent to finding all the global maxima of a nonconvex function of the model parameters and so it is a very challenging one. In Katafygiotis and Beck (1998), an important step is made toward solving this problem for a class of linear structural dynamic models by presenting an algorithm to determine all the models that are output-equivalent to a given model. It is also shown that choosing just a single model, as usually done by estimating the model parameters through optimal matching of the model and measured responses at certain degrees of freedom, can lead to unreliable response predictions at the unobserved degrees of freedom, when the model used is not globally identifiable.

ACKNOWLEDGMENTS

This paper was completed while the first writer was on sabbatical leave at the Hong Kong University of Science and Technology. The hospitality of Jay-Chung Chen, Director of the Research Centre, is greatly appreciated.

APPENDIX I.

First, we derive the asymptotic approximation (16) for the first integral in (7) giving the updated predictive PDF for the case of global system identifiability, that is, when there is a unique optimal parameter . Denote the integral in (7) by I; then

$$I = p(Y_{N+1}^M, X_1^M|\mathcal{D}_N, Z_{N+1}^M, \mathcal{M}_P) = c\int_{S(\boldsymbol{\alpha})} h(\boldsymbol{\alpha}) f_N(\hat{Y}_1^N; \boldsymbol{\alpha}, \hat{Z}_1^N)\, d\boldsymbol{\alpha} \quad (27)$$

where

$$h(\boldsymbol{\alpha}) = p(Y_{N+1}^M, X_1^M|\boldsymbol{\alpha}, \mathcal{D}_N, Z_{N+1}^M, \mathcal{M}_P)\pi(\boldsymbol{\alpha}) \quad (28)$$

Substituting (14) into (27)

$$I = cf_N(\hat{Y}_1^N; \hat{\boldsymbol{\alpha}}, \hat{Z}_1^N)$$

$$\cdot \int_{S(\boldsymbol{\alpha})} h(\boldsymbol{\alpha}) \exp E(\boldsymbol{\alpha}) \exp\left(-\frac{1}{2}[\boldsymbol{\alpha} - \hat{\boldsymbol{\alpha}}]^T \mathbf{A}_N(\hat{\boldsymbol{\alpha}})[\boldsymbol{\alpha} - \hat{\boldsymbol{\alpha}}]\right) d\boldsymbol{\alpha} \quad (29)$$

where A_N = Hessian matrix given by (15); and $E(\boldsymbol{\alpha})$ = error in the second-order Taylor series expansion for $\ln f_N(\hat{Y}_1^N, \boldsymbol{\alpha}, \hat{Z}_1^N)$ about $\hat{\boldsymbol{\alpha}}$, so $E(\hat{\boldsymbol{\alpha}}) = 0$. By applying Laplace's method of asymptotic expansion to this integral (Bleistein and Handelsman 1986), an approximation for I is obtained

$$I \simeq (2\pi)^{1/2N_\alpha} c f_N(\hat{Y}_1^N; \hat{\boldsymbol{\alpha}}, \hat{Z}_1^N) h(\hat{\boldsymbol{\alpha}}) |\mathbf{A}_N(\hat{\boldsymbol{\alpha}})|^{-1/2} \quad (30)$$

where $\simeq$ denotes an approximation to $\mathcal{O}(N^{-1})$. Since c is given by the integral in (10), which is the same as (27) except $ch(\boldsymbol{\alpha})$ is replaced by $\pi(\boldsymbol{\alpha})$, it can also be approximated by

$$c^{-1} \simeq (2\pi)^{1/2N_\alpha} f_N(\hat{Y}_1^N; \hat{\boldsymbol{\alpha}}, \hat{Z}_1^N)\pi(\hat{\boldsymbol{\alpha}}) |\mathbf{A}_N(\hat{\boldsymbol{\alpha}})|^{-1/2} \quad (31)$$

Substituting (28) and (31) into (30)

$$I \simeq p(Y_{N+1}^M, X_1^M|\hat{\boldsymbol{\alpha}}, \mathcal{D}_N, Z_{N+1}^M, \mathcal{M}_P) \quad (32)$$

This proves (16) for the special case $K = 1$.

Now consider the case of local system identifiability where $K > 1$. The integral (27) can be decomposed into a sum of K integrals over the K disjoint subregions S_k, $k = 1, \ldots, K$, of a partition of the parameter space $S(\boldsymbol{\alpha})$, where S_k contains a single optimal parameter $\hat{\boldsymbol{\alpha}}^{(k)}$, that is

$$I = \sum_{k=1}^{K} I_k \quad (33)$$

where

$$I_k = c\int_{S_k} h(\boldsymbol{\alpha}) f_N(\hat{Y}_1^N; \boldsymbol{\alpha}, \hat{Z}_1^N)\, d\boldsymbol{\alpha}$$

$$\simeq (2\pi)^{1/2N_\alpha} c f_N(\hat{Y}_1^N; \hat{\boldsymbol{\alpha}}^{(k)}, \hat{Z}_1^N) h(\hat{\boldsymbol{\alpha}}^{(k)}) |\mathbf{A}_N(\hat{\boldsymbol{\alpha}}^{(k)})|^{-1/2} \quad (34)$$

since the asymptotic approximation (30) can be applied to each I_k. Similarly, the integral in (10) for c^{-1} can be approximated by

$$c^{-1} \simeq (2\pi)^{1/2 N_\alpha} \sum_{k=1}^{K} f_N(\hat{Y}_1^N; \hat{\boldsymbol{\alpha}}^{(k)}, \hat{Z}_1^N) \pi(\hat{\boldsymbol{\alpha}}^{(k)}) |\mathbf{A}_N(\hat{\boldsymbol{\alpha}}^{(k)})|^{-1/2} \quad (35)$$

By the definition of an optimal parameter, for each $k = 1, \ldots, K$

$$f_N(\hat{Y}_1^N; \hat{\boldsymbol{\alpha}}^{(k)}, \hat{Z}_1^N) = \max_{\boldsymbol{\alpha} \in S(\boldsymbol{\alpha})} f_N(\hat{Y}_1^N; \boldsymbol{\alpha}, \hat{Z}_1^N) \quad (36)$$

Eq. (16) with the weighting coefficients w_k, $k = 1, \ldots, K$, in (18) follows by substituting (35) and (36) into (34), then substituting (34) into (33).

Finally, we show that the expression for the weighting coefficients in (18) can be simplified when some of the model parameters α_j in $\boldsymbol{\alpha}$ are globally system identifiable. By renumbering if necessary, the model parameters $\boldsymbol{\alpha}$ can be partitioned into $\boldsymbol{\alpha} = [\boldsymbol{\theta}^T, \boldsymbol{\psi}^T]^T$, where $\boldsymbol{\theta} \in R^{N_\theta}$, $\boldsymbol{\psi} \in R^{N_\psi}$ and $N_\theta + N_\psi = N_\alpha$, so that each optimal parameter $\hat{\boldsymbol{\alpha}}^{(k)} = [\hat{\boldsymbol{\theta}}^{(k)T}, \hat{\boldsymbol{\psi}}^T]^T$, $k = 1, \ldots, K$. This implies that each ψ_j and θ_j are respectively globally and locally system identifiable under $\mathcal{D}_N$. Since the parameters $\boldsymbol{\psi}$ are globally system identifiable at $\hat{\boldsymbol{\psi}}$, (32) can be used to write

$$I \simeq p(Y_{N+1}^M, X_1^M | \hat{\boldsymbol{\psi}}, \mathcal{D}_N, Z_{N+1}^M, \mathcal{M}_P)$$

$$= \int_{S(\boldsymbol{\theta}, \hat{\boldsymbol{\psi}})} p(Y_{N+1}^M, X_1^M | \boldsymbol{\theta}, \hat{\boldsymbol{\psi}}, \mathcal{D}_N, Z_{N+1}^M, \mathcal{M}_P) p(\boldsymbol{\theta} | \hat{\boldsymbol{\psi}}, \mathcal{D}_N, \mathcal{M}_P)\, d\boldsymbol{\theta} \quad (37)$$

where the last equation follows from the Total Probability Theorem. Using the previous asymptotic results, the integral in (37) can be approximated by

$$I = \sum_{k=1}^{K} \tilde{w}_k p(Y_{N+1}^M, X_1^M | \hat{\boldsymbol{\theta}}^{(k)}, \hat{\boldsymbol{\psi}}, \mathcal{D}_N, Z_{N+1}^M, \mathcal{M}_P)[1 + \mathcal{O}(N^{-1})] \quad (38)$$

where

$$\tilde{w}_k = \frac{\tilde{w}_k'}{\sum_{k=1}^{K} \tilde{w}_k'}; \quad \tilde{w}_k' = \pi(\hat{\boldsymbol{\theta}}^{(k)} | \hat{\boldsymbol{\psi}}) |C_N(\hat{\boldsymbol{\theta}}^{(k)})|^{-1/2} \quad (39)$$

and the elements of the $N_\theta \times N_\theta$ Hessian matrix $\mathbf{C}_N(\boldsymbol{\theta})$ are given by

$$[\mathbf{C}_N(\boldsymbol{\theta})]_{ij} = -\frac{\partial^2 \ln f_N(\hat{Y}_1^N; \boldsymbol{\theta}, \hat{\boldsymbol{\psi}}, \hat{Z}_1^N)}{\partial \theta_i \partial \theta_j} \quad (40)$$

It follows from (38) and (16) that

$$\tilde{w}_k = w_k; \quad k = 1, \ldots, K \quad (41)$$

that is, the weightings associated with each optimal model can be calculated using (18) where the matrix $\mathbf{A}_N$ is replaced by its submatrix $\mathbf{C}_N$ made up of the elements that involve partial differentiation only with respect to the locally identifiable parameters $\boldsymbol{\theta}$, while the values of the globally identifiable parameters are set equal to their optimal values $\hat{\boldsymbol{\psi}}$. Also, notice that the effect on the weighting coefficients of the initial probability distribution over the parameters can be represented by $\pi(\hat{\boldsymbol{\theta}}^{(k)} | \hat{\boldsymbol{\psi}})$; that is, the initial PDF calculated at the optimal parameters can be replaced by the initial PDF of the locally identifiable parameters conditional on the globally identifiable parameters.

APPENDIX II. REFERENCES

Beck, J. L. (1989). "Statistical system identification of structures." *Proc., 5th Int. Conf. on Struct. Safety and Reliability*, ASCE, 1395–1402.

Beck, J. L. (1996). "System identification methods applied to measured seismic response." *Proc., 11th World Conf. on Earthquake Engrg.*, Elsevier Science Publishing Co., Inc., New York, N.Y.

Beck, J. L., and Katafygiotis, L. S. (1992). "Probabilistic system identification and health monitoring of structures." *Proc., 10th World Conf. on Earthquake Engrg.*, Elsevier Science Publishing Co., Inc., New York, N.Y.

Berman, A. (1989). "Nonunique structural system identification." *Proc., 7th Int. Modal Anal. Conf.*, Society for Experimental Mechanics, Bethel, Conn., 355–359.

Bleistein, N., and Handelsman, R. (1986). *Asymptotic expansions for integrals*. Dover Publications, Inc., New York, N.Y.

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

Bretthorst, G. L. (1988). "Bayesian spectrum analysis and parameter estimation." *Springer lecture notes on physics no. 48*, Springer-Verlag KG, Berlin, Germany.

Capecchi, D., and Vestroni, F. (1993). "Identification of finite-element models in structural dynamics." *Engrg. Struct.*, 15, 21–30.

Chen, J. C., and Garba, J. A. (1980). "Analytical model improvement using modal test results." *AIAA J.*, 18, 684–690.

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

Eykhoff, P. (1974). *System identification: parameter and state estimation*. John Wiley & Sons, Inc., New York, N.Y.

Fritzen, C. P., and Zhu, S. (1991). "Updating of finite-element models by means of measured information." *Comp. and Struct.*, 40, 475–486.

Goodwin, G. C., and Payne, R. L. (1977). *Dynamic system identification, experiment design and data analysis*. Academic Press, Inc., New York, N.Y.

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

Identification of vibrating structures. (1982). H. G. Natke, ed., Springer-Verlag New York, Inc., New York, N.Y.

Imregun, M., and Visser, W. J. (1991). "A review of model updating techniques." *Shock and Vibration Digest*, 23, 9–20.

Janter, T., and Sas, P. (1990). "Uniqueness aspects of model updating procedures." *AIAA J.*, 28, 538–543.

Jaynes, E. T. (1968). "Prior probabilities." *IEEE Trans. Sys. Sci. & Cybernetics*, SSC-4, 227–241.

Jaynes, E. T. (1983). *E. T. Jaynes: papers on probability, statistics and statistical physics*, R. D. Rosenkrantz, ed., Reidel, Dordrecht, The Netherlands. Reprint, 1989, Kluwer Academic Publishers Group, Boston, Mass.

Jaynes, E. T. (1988). "How does the brain do plausible reasoning?" *Maximum-entropy and Bayesian methods in science and engineering. Vol. 1: Foundations*. G. J. Erickson and C. R. Smith, eds., Kluwer Academic Publishers Group, Boston, Mass. (Reprint of *Microwave Lab. Rep. No. 421*, 1957, Stanford University, Stanford, Calif.).

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

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

Lindley, D. V. (1965). *Introduction to probability and statistics from a Bayesian viewpoint*. Cambridge University Press, Cambridge, U.K.

Ljung, L. (1987). *System identification: theory for the user*. Prentice-Hall, Inc., Englewood Cliffs, N.J.

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

Natke, H. G. (1988). "Updating computational models in the frequency domain based on measured data: a survey." *Probabilistic Engrg. Mech.*, Essex, U.K., 8, 28–35.

Proc., Int. Workshop on Struct. Safety Evaluation based on System Identification Approaches. (1988). H. G. Natke and J. T. P. Yao, eds., Vieweg and Sons, Wiesbaden, Germany.

Shinozuka, M., Yun, C. B., and Imai, H. (1982). "Identification of linear structural dynamics systems." *J. Engrg. Mech. Div.*, ASCE, 108, 1371–1390.

Torkamani, M. A. M., and Ahmadi, A. K. (1988). "Stiffness identification of two- and three-dimensional frames." *Earthquake Engrg. and Struct. Dynamics*, 16, 1157–1176.

Udwadia, F. E. (1985). "Some uniqueness results related to soil and building structural identification." *SIAM J. of Appl. Math.*, 45, 674–685.

Yang, C. M., and Beck, J. L. (1998). "Generalized trajectory methods for finding multiple extrema and roots of functions." *J. Optimization Theory and Applications*, 4, 211–227.