Proceedings of 5th International Conference on Structural Safety and Reliability, August 1989.

STATISTICAL SYSTEM IDENTIFICATION OF STRUCTURES

James L. Beck

Associate Professor of Civil Engineering
California Institute of Technology, Pasadena, CA 91125

ABSTRACT

A general unifying approach to system identification is presented within a Bayesian statistical framework to explicitly treat the inherent uncertainties. It is shown that selecting the most probable model from a class of models for a structure based on its measured input and output leads to a rational and computationally feasible approach for response prediction. It is also asymptotically correct as the sample size is increased. The methodology is illustrated using an output-error formulation which has been successfully applied to recorded seismic motions from structures.

KEYWORDS

System identification; statistical inference; Bayesian probability; structural dynamics; earthquake records.

1. INTRODUCTION

The objective of system identification in structural dynamics is to improve mathematical models of the dynamics of a structure by using its measured response and possibly its measured excitation [1]. The improved models are most often used for response predictions for possible future loads on the structure or for developing control strategies for the structure. In some cases, the improved models are used for damage detection and assessment, although this application has not yet fully matured [2].

In this paper, we present a general methodology for treating the response prediction problem where we choose an appropriate class of models for the input-output behavior of a system under study, whose input and output have been measured over some time interval, and then use this information to predict the output of the system to other input. The methodology has been developed as a result of our interest in the analysis of earthquake motions measured in structures [3], but it is completely general. It treats system identification as an application of Bayesian statistical inference [4, 5] to explicitly consider the uncertainties involved and to provide a coherent framework for interpreting different methods. For illustration, it is shown to provide a justification for the standard output-error least-squares method in which a single "optimal" model is chosen from the class of appropriate models in order to make response predictions.

2. A METHODOLOGY FOR STATISTICAL INFERENCE

Probability Logic

In the theory which follows, it is important to keep in mind that probability is not being used in the "classical" sense as a relative frequency of occurrences of an event in the long run. Instead, a type of "Bayesian" approach is used in which probability is treated as a multi-valued logic subject to certain axioms [6,7]. Specifically, $P(a|b)$, the probability of a given b, denotes a measure of the plausibility of the proposition a given the information stated in proposition b. The propositions may refer to observations or measurements, or they may refer to hypotheses about probability models, for example.

The calculus of probability logic is defined by the axioms of mathematical (Boolean) logic together with three additional axioms:

1) $0 \leq P(a|b) \leq 1$ and $P(a|a) = 1$, 2) $P(a|b) + P(\text{not } a|b) = 1$,

and 3) $P(a,b|c) = P(a|b,c)P(b|c)$

where "," represents the propositional conjunction "and". Cox has shown that the content of axioms 2) and 3) is a necessary consequence of the requirement of consistency with mathematical logic, although the form of all three axioms is conventional [8]. The axioms lead to the same calculus as the Kolmogorov axioms of "mathematical" probability, except that all probabilities are "conditional" in the sense of Kolmogorov because the plausibility of a proposition clearly depends on the information available.

Statistical Inference Problem

Let $Y_1^M = \{\underline{y}_n \in R^{N_0} : n = 1,2,\ldots,M\}$ denote a sequence of M observable vector quantities of interest and $\hat{Y}_1^N = \{\underline{\hat{y}}_n : n = 1,2,\ldots,N\}$ a sample of the first N $\underline{y}_n$. Let P denote a class of probability models for Y_1^M parameterized by $\underline{\theta} \in H \subset R^{N_P}$, that is, P prescribes a function f_M so that:

$$p\left(Y_1^M | \underline{\theta}, P\right) = f_M\left(Y_1^M; \underline{\theta}\right) = f_M(\underline{y}_1, \underline{y}_2, \ldots, \underline{y}_M; \underline{\theta}) \,. \tag{1}$$

For notational convenience, Y_1^M on the left side of Eq. (1) denotes the proposition: "The sequence of M vector quantities takes on the values $\{\underline{y}_n : n = 1,2,\ldots,M\}$ when they

are observed;" and for mathematical convenience, continuous variables are assumed, so p denotes the probability density function.

The problem of interest is to derive an improved probability model for Y_{N+1}^{M}, the remainder of the sequence, by utilizing the data $\hat{Y}_1^N$ and the class of models P. Class P is assumed to be *identifiable*, that is, there is a one-to-one mapping between $\underline{\theta} \in H$ and the probability models $f_M(Y_1^M; \underline{\theta})$, so $\underline{\theta}$ serves to represent a specific probability model. Also, as part of the choice of the class P, suppose that the user chooses a probability model $\pi(\underline{\theta})$ so that:

$$p(\underline{\theta}|P) = \pi(\underline{\theta}) \tag{2}$$

where the function π is based on the user's judgement regarding the relative plausibilities of different models in the class P, that is, of different values of $\underline{\theta}$.

Before utilizing the data $\hat{Y}_1^N$, the best choice of probability function for making predictions about Y_1^M would be $p(Y_1^M|P)$, which, based on the axioms of probability logic, can be expressed as:

$$\begin{aligned} p(Y_1^M|P) &= \int_{\underline{\theta}\in H} p(Y_1^M|\underline{\theta},P)\, p(\underline{\theta}|P) d\underline{\theta} \\ &= \int_{\underline{\theta}\in H} f_M(Y_1^M;\underline{\theta})\pi(\underline{\theta}) d\underline{\theta}\ . \end{aligned} \tag{3}$$

After utilizing the data $\hat{Y}_1^N$, the best choice of probability function for Y_{N+1}^M would be:

$$\begin{aligned} p(Y_{N+1}^M|\hat{Y}_1^N,P) &= \int_{\underline{\theta}\in H} p(Y_{N+1}^M|\underline{\theta},\hat{Y}_1^N,P)\, p(\underline{\theta}|\hat{Y}_1^N,P) d\underline{\theta} \\ &= k\int_{\underline{\theta}\in H} p(Y_{N+1}^M,\hat{Y}_1^N|\underline{\theta},P)\, p(\underline{\theta}|P) d\underline{\theta} \\ &= k\int_{\underline{\theta}\in H} f_M(\underline{\hat{y}}_1,\underline{\hat{y}}_2,\ldots,\underline{\hat{y}}_N,\underline{y}_{N+1},\ldots,\underline{y}_M;\underline{\theta})\,\pi(\underline{\theta}) d\underline{\theta}\ , \end{aligned} \tag{4}$$

since

$$p(\underline{\theta}|\hat{Y}_1^N,P) = kp(\hat{Y}_1^N|\underline{\theta},P)p(\underline{\theta}|P) = kf_N(\hat{Y}_1^N;\underline{\theta})\,\pi(\underline{\theta}) \tag{5}$$

where constant k satisfies:

$$k^{-1} = p(\hat{Y}_1^N|P) = \int_{\underline{\theta}\in H} f_N(\hat{Y}_1^N;\underline{\theta})\,\pi(\underline{\theta}) d\underline{\theta}\ . \tag{6}$$

The difficulty with this rigorous approach is that the integrals are, in general, over a high-dimensional space ($H \subset R^{N_P}$ where N_P is the number of parameters) and they usually cannot be done analytically. This gives rise to computationally expensive, or even

prohibitive, numerical calculations. If a single-model approximation is desired in order to avoid the integrations, then the most rational choice would be the most probable model within the class of models P based on the data $\hat{Y}_1^N$, that is, choose $\hat{\underline{\theta}}$ as the value of $\underline{\theta}$ which uniquely maximizes $p(\underline{\theta}|\hat{Y}_1^N, P)$ given by Eq. (5), and then take the probability function:

$$p(Y_{N+1}^M|\hat{\underline{\theta}}, \hat{Y}_1^N, P) = \frac{p(Y_{N+1}^M, \hat{Y}_1^N|\hat{\underline{\theta}}, P)}{p(\hat{Y}_1^N|\hat{\underline{\theta}}, P)} = \frac{f_M(\hat{\underline{y}}_1, \hat{\underline{y}}_2, \ldots, \hat{\underline{y}}_N, \underline{y}_{N+1}, \ldots, \underline{y}_M; \hat{\underline{\theta}})}{f_N(\hat{\underline{y}}_1, \hat{\underline{y}}_2, \ldots, \hat{\underline{y}}_N; \hat{\underline{\theta}})} \quad (7)$$

for making predictions about Y_{N+1}^M. We shall call this probability function the "optimal predictive probability model" to distinguish it from the "optimal probability model" in P given by $\hat{\underline{\theta}}$.

In most situations, for large sample size N, $\log f_N(\hat{Y}_1^N; \underline{\theta}) = O(N)$ and so $\pi(\underline{\theta})$ has little effect on the determination of the optimal probability model. In this case, $\pi(\underline{\theta})$ can be ignored and the optimal model can be found by maximizing $f_N(\hat{Y}_1^N; \underline{\theta})$. The values $\hat{\underline{\theta}}$ for the most probable model are then identical to the classical maximum likelihood estimates, although the conceptual framework is quite different for these two approaches.

The choice of a single model (the most probable model) in class P raises the question of what error is introduced compared with the correct predictive probability function for Y_{N+1}^M given by Eq. (4). It can be shown that in fact the approximation is asymptotically exact as the sample size N grows. This result can be derived by expanding $\log p(\underline{\theta}|\hat{Y}_1^N, P)$ in a Taylor series about its maximizing value $\hat{\underline{\theta}}$. This gives the following asymptotic expansion for the integral for $p(Y_{N+1}^M|\hat{Y}_1^N, P)$:

$$p\left(Y_{N+1}^M|\hat{Y}_1^N, P\right) = p\left(Y_{N+1}^M|\hat{\underline{\theta}}, \hat{Y}_1^N, P\right) + O(N^{-1}) \quad (8)$$

Large sample sizes are typical in system identification studies in structural dynamics. This means that response predictions can be made with confidence using the optimal predictive probability model in Eq. (7) without the need to evaluate the multiple integrals in Eq. (4).

Any well-tested optimization algorithm can be used to determine $\hat{\underline{\theta}}$ by maximizing $p(\underline{\theta}|\hat{Y}_1^N, P)$ given by Eq. (5). One caveat is that there is often a problem of whether the calculated vector $\hat{\underline{\theta}}$ gives the global maximum, which is guaranteed to be unique when the class of models P is identifiable, or whether only a local maximum has been found.

The assumption of identifiability of the class of models P is not necessary for making predictions about Y_{N+1}^M using the exact predictive probability function in Eq. (4). Its importance is due to the selection of a single optimal model from the class P. The method can, however, be adapted if the class P is only locally identifiable in the sense that there are K values of $\underline{\theta}$, say $\hat{\underline{\theta}}_k$, $k = 1, 2, \ldots, K$, all maximizing $f_N(\hat{Y}_1^N; \underline{\theta})$. In this case, we can use the fact that for large sample sizes N:

$$p\left(Y_{N+1}^M|\hat{Y}_1^N, P\right) = \sum_{k=1}^{K} w_k p\left(Y_{N+1}^M|\hat{\underline{\theta}}_k, \hat{Y}_1^N, P\right) + O(N^{-1}) \quad (9)$$

where

$$w_k = \pi(\hat{\underline{\theta}}_k) \Big/ \sum_{j=1}^{K} \pi(\hat{\underline{\theta}}_j) \ , \tag{10}$$

although determining this weighted average involves the difficult computational problem of finding all K global maxima of $f_N(\hat{Y}_1^N; \theta)$. Note that in this case the prior probability model $\pi(\underline{\theta})$ has some influence even for large sample sizes.

Finally, a study of the equations involved (Eq. (5) and (7)) shows that the usual concern about the reliability of the parameter estimates is not an issue for the prediction problem. Basically, a parameter θ_i which is not precisely pinned down by the data $\hat{Y}_1^N$ is at the same time not a critical parameter in controlling the value of the predictive probability for Y_{N+1}^M.

3. STATISTICAL FRAMEWORK FOR SYSTEM IDENTIFICATION

Prediction Problem

$$\text{Let } \hat{Z}_1^N = \{\hat{\underline{z}}_n \in R^{N_I} : n = 1, 2, \ldots, N\} \text{ and } \hat{Y}_1^N = \{\hat{\underline{y}}_n \in R^{N_O} : n = 1, 2, \ldots, N\} \tag{11}$$

be sampled input and output histories from a system. Suppose an appropriate class of identifiable input-output models $M = \{M(\underline{\theta}) : \underline{\theta} \in H \subset R^{N_M}\}$ has been selected for the system where, for each $\underline{\theta} \in H$, M provides a functional relationship between the model output $\underline{m}_n$ and the system input Z_1^n, that is, $\underline{m}_n = \underline{g}_n(Z_1^n; \underline{\theta})$. For example, a state-space formulation or a class of ARMAX models may be chosen. Also, any unknown initial conditions required to generate $\underline{m}_n$ are included in the parameter vector $\underline{\theta}$. The prediction problem is to use this class M and the available data to predict the output of the system Y_{N+1}^M, for additional prescribed input to the system Z_{N+1}^M.

Before the statistical inference approach of the previous section can be used, a probability model must be set up for the output which is to be predicted since $M(\underline{\theta})$ only prescribes the deterministic model output $\underline{m}_n$ corresponding to the system output $\underline{y}_n$; it does not account for uncertainties arising because the class of models is never perfect ("model error") and measurement noise is always present. Appropriate choices for the class of probability models lead to various standard methods, such as the equation-error and output-error least-squares identification procedures. Also, recursive procedures can be formulated if sequential calculation of the most probable parameters is desired. For example, one can take a combined equation-error/output-error approach to obtain the extended Kalman filter. Here, we illustrate only the output-error approach.

Output-Error Approach

The output error $\underline{e}_n$ is defined to be the difference between the system output and the model output, so:

$$\underline{y}_n = \underline{m}_n(Z_1^n; \underline{\theta}) + \underline{e}_n \tag{12}$$

Although the model output is needed only at discrete times, the underlying class of models M can be in any form, such as continuous-time differential equations, which can be converted to treat discrete input and output. A class of joint probability models P is selected so that:

$$p(\underline{e}_1, \underline{e}_2, \ldots, \underline{e}_M | \underline{\sigma}, P) = h_M(\underline{e}_1, \underline{e}_2, \ldots, \underline{e}_M; \underline{\sigma}) \tag{13}$$

where $\underline{\sigma} \in \sum \subset R^{N_P}$. The desired probability model in Eq. (1) is then given by

$$\begin{aligned} f_M(Y_1^M; \underline{\theta}, \underline{\sigma}) &= p(\underline{y}_1, \underline{y}_2, \ldots, \underline{y}_M | \underline{\theta}, \underline{\sigma}, Z_1^M, M, P) \\ &= p(\underline{e}_1 + \underline{m}_1, \underline{e}_2 + \underline{m}_2, \ldots, \underline{e}_M + \underline{m}_M | \underline{\theta}, \underline{\sigma}, Z_1^M, M, P)|_{\underline{e}_n = \underline{y}_n - \underline{m}_n} \\ &= p(\underline{e}_1, \underline{e}_2, \ldots, \underline{e}_M | \underline{\theta}, \underline{\sigma}, Z_1^M, M, P)|_{\underline{e}_n = \underline{y}_n - \underline{m}_n} \\ &= h_M(\underline{y}_1 - \underline{m}_1(\underline{z}_1; \underline{\theta}), \ldots, \underline{y}_M - \underline{m}_M(Z_1^M; \underline{\theta}); \underline{\sigma}) \end{aligned} \tag{14}$$

Output predictions are made using the optimal predictive probability model in Eq. (7) obtained by substituting Eq. (14) evaluated at the optimal parameters $\hat{\underline{\theta}}$ and $\hat{\underline{\sigma}}$. These most probable values are calculated by maximizing:

$$\begin{aligned} &p(\underline{\theta}, \underline{\sigma} | \hat{Y}_1^N, \hat{Z}_1^N, M, P) \\ &\quad = k h_N(\hat{\underline{y}}_1 - \underline{m}_1(\hat{\underline{z}}_1; \underline{\theta}), \ldots, \hat{\underline{y}}_N - \underline{m}_N(\hat{Z}_1^N; \underline{\theta}); \underline{\sigma})\, \pi(\underline{\theta}, \underline{\sigma}) \end{aligned} \tag{15}$$

The choice of the function $\pi(\underline{\theta}, \underline{\sigma})$ may be based on experience and theoretical results. For example, the parameters of a finite-element model for a structure may be taken as the most plausible values a priori for the parameters of a structural model. As already mentioned, for an identifiable class of models and for large sample sizes N, $\pi(\underline{\theta}, \underline{\sigma})$ has no effect if it is nonzero and continuous, so it can be ignored.

As a special case, the standard least-squares output-error method can be derived by choosing the class P for the output-error joint probabilities as:

$$h_M(\underline{e}_1, \underline{e}_2, \ldots, \underline{e}_M; \underline{\sigma}) = (2\pi\sigma^2)^{-\frac{1}{2}MN_0} \exp\left(-\frac{1}{2\sigma^2} \sum_{n=1}^{M} \underline{e}_n^T \underline{e}_n\right) \tag{16}$$

This is equivalent to saying that the output error is modelled as a Gaussian white-noise sequence with zero mean and covariance matrix $\sigma^2 I_{N_0}$ (I_{N_0} = identity matrix of order N_0). For a structural system, it assumes both temporal and spatial independence of the output errors, that is, it asserts that the user feels that knowing the output errors at other times, or other locations within the structure, does not influence his uncertainty concerning the value of the output error at a specified time and location. In this case, employing the previous theory, the optimal parameter $\hat{\underline{\theta}}$ is determined by minimizing:

$$J(\underline{\theta}) = \sum_{n=1}^{N} \| \hat{\underline{y}}_n - \underline{m}_n(\hat{Z}_1^n; \underline{\theta}) \|^2 \tag{17}$$

and then:

$$\hat{\sigma}^2 = \frac{1}{NN_0} J(\hat{\underline{\theta}}) \tag{18}$$

The optimal predictive probability model given by Eq. (7) is Gaussian, even if the class of models M is nonlinear. In fact, each predicted output $\underline{y}_n$ is independently distributed as Gaussian with mean $\underline{m}_n(\hat{Z}_1^n; \hat{\underline{\theta}})$ and covariance matrix $\hat{\sigma}^2 I_{N_0}$. The most probable output is $\underline{m}_n(\hat{Z}_1^n; \hat{\underline{\theta}})$, $n = N+1, \ldots, M$, which is the output of the most probable deterministic model, $M(\hat{\underline{\theta}})$ (the "optimal" model). The predictive probability is therefore easy to interpret in the output-error case. However, the method requires substantial computational effort because the optimization in determining the most probable model is a "nonlinear" least-squares calculation which must be done iteratively. This is because $\underline{m}_n$ in Eq. (17) is a nonlinear function of $\underline{\theta}$ even if the model is linear in the parameters.

4. CONCLUSION

The methodology proposed herein for the response prediction problem in system identification uses a Bayesian statistical framework to treat the inherent uncertainties. To allow a computationally-feasible Bayesian implementation, a guiding principle is proposed, namely, from a class of appropriate models for a system, choose the most probable model based on measured input and output of the system. This criterion is not only the most rational one for choosing a single model from the class, from the Bayesian point of view it is also asymptotically correct as the sample size is increased. Output prediction is made using the optimal predictive probability, which describes the uncertainty remaining when all the prescribed information has been utilized. If a single best prediction is desired for the output, the most probable output can be taken which is given by the maximum of the optimal predictive probability.

The described output-error method has been successfully applied to earthquake response records by the author and his co-workers. Most of the applications have employed linear multiple-input multiple-output models [9] which are based on the superposition of normal modes of vibration, and include applications to the earthquake response of buildings [10, 11], a bridge [9], and an off-shore platform [12]. Nonlinear hysteretic models have also been employed using pseudodynamic earthquake data from a six-story steel-frame structure tested in Japan [13]. Of course, a structure always behaves nonlinearly to some degree during an earthquake. However, since no model is perfect, the guiding philosophy is to choose a class of deterministic models, such as the linear modal models, model the associated uncertainty with a probability distribution, and then employ the proposed methodology to investigate how much residual uncertainty remains when the optimal deterministic model is used for response predictions. In this spirit, we have concluded from our studies that linear dynamic models perform well for seismic response until the onset of significant structural damage.

REFERENCES

1. NATKE, H. G. (Ed.), Identification of Vibrating Structures, Springer-Verlag, New York, 1982.

2. NATKE, H. G. and J. T. P. YAO (Eds.), Structural Safety Evaluation Based on System Identification Approaches, Vieweg & Son, Wiesbaden, 1988.

3. BECK, J. L., "System identification applied to strong motion records from structures," in Earthquake Ground Motion and its Effects on Structures, S. K. Datta (Ed.), AMD-53, ASME, New York, 1982.

4. BOX, G. E. P. and G. C. TIAO, Bayesian Inference in Statistical Analysis, Addison-Wesley, Reading, Massachusetts, 1973.

5. PETERKA, V., "Bayesian approach to system identification," in Trends and Progress in System Identification, P. Eykhoff (Ed.), Pergamon Press, New York, 1981.

6. JEFFREYS, H., Theory of Probability, Oxford Clarendon Press, 3rd Ed., 1961.

7. JAYNES, E. T., "Prior probabilities," *IEEE Trans. Syst. Sci. & Cybernetics*, SSC-4, 227–241, 1968.

8. COX, R. T., The Algebra of Probable Inference, Johns Hopkins Press, Baltimore, 1961.

9. WERNER, S.D., J.L. BECK and M.B. LEVINE, "Seismic response evaluation of Meloland Road Overpass using 1979 Imperial Valley Earthquake Records," *Int. J. of Earthquake Eng. and Struc. Dyn.*, **15**, 249-274, February 1987.

10. BECK, J.L. and P.C. JENNINGS, "Structural identification using linear models and earthquake records," *Int. J. of Earthquake Eng. and Struc. Dyn.*, 8, 145-160, April 1980.

11. McVERRY, G.H. and J.L. BECK, Structural Identification of JPL Building 180 Using Optimally Synchronized Earthquake Records, Report No. EERL 83-01, California Institute of Technology, Pasadena, 1983.

12. MASON, A.B., J.L. BECK, J. CHEN and R.R. ULLMAN, "Modal parameter identification of an offshore platform from earthquake response records," *Sessions Related to Seismic Eng.*, 217-226, Structures Congress, ASCE, New York, May 1989.

13. JAYAKUMAR, P. and BECK, J.L., "System identification using nonlinear structural models," Structural Safety Evaluation Based on System Identification Approaches, H. G. Natke and J.T.P. Yao (eds.), Vieweg & Son, Wiesbaden, 1988.

Errata for "Statistical System Identification of Structures" by J.L. Beck

After submission of the paper at the ICOSSAR conference in August 1989, the author found an error in Eq. (10), which should be replaced by:

$$w_k = {w_k'}\Big/{\sum_{j=1}^{K} w_j'}$$

where

$$w_k' = \pi(\underline{\hat{\theta}}_k)|A_N(\underline{\hat{\theta}}_k)|^{-1/2}$$

involves the determinant of the Hessian matrix:

$$[A_N(\underline{\theta})]_{ij} = -\frac{\partial^2}{\partial\theta_i\,\partial\theta_j}\left[\ln f_N(\underline{Y}_1^N;\underline{\theta}) + \ln\pi(\underline{\theta})\right]$$

The correct equation also appears in the conference paper:

"Updating of a Model and its Uncertainties Utilizing Dynamic Test Data," J.L. Beck and L.S. Katafygiotis, *Proceedings First International Conference on Computational Stochastic Mechanics*, 125-136, Computational Mechanics Publications, Boston, September 1991.