ARTICLE IN PRESS

Journal of Sound and Vibration ∎ (∎∎∎∎) ∎∎∎–∎∎∎

Contents lists available at ScienceDirect

Journal of Sound and Vibration

journal homepage: www.elsevier.com/locate/jsvi

A Bayesian approach to identifying structural nonlinearity using free-decay response: Application to damage detection in composites

J.M. Nichols [a,*], W.A. Link [c], K.D. Murphy [d], C.C. Olson [b]

[a] U.S. Naval Research Laboratory, 4555 Overlook Avenue, Washington, DC 20375, USA
[b] Naval Research Laboratory, USA
[c] USGS Patuxent Wildlife Research Center, 12100 Beech Forest Road, Laurel, MD 20708, USA
[d] Department of Mechanical and Aeronautical Engineering, University of Connecticut, Storrs, CT 06269-3139, USA

ARTICLE INFO

Article history:
Received 29 June 2009
Received in revised form
14 July 2009
Accepted 3 February 2010
Handling Editor: A.V. Metrikine

ABSTRACT

This work discusses a Bayesian approach to approximating the distribution of parameters governing nonlinear structural systems. Specifically, we use a Markov Chain Monte Carlo method for sampling the posterior parameter distributions thus producing both point and interval estimates for parameters. The method is first used to identify both linear and nonlinear parameters in a multiple degree-of-freedom structural systems using free-decay vibrations. The approach is then applied to the problem of identifying the location, size, and depth of delamination in a model composite beam. The influence of additive Gaussian noise on the response data is explored with respect to the quality of the resulting parameter estimates.

Published by Elsevier Ltd.

1. Introduction

System identification can be loosely defined as the process of estimating parameters associated with a specified model (or models) from acquired data. There are two main schools of thought in estimation problems: the frequentist approach, often based on the method of maximum likelihood (ML), and the Bayesian approach. Both methods seek to provide the best possible parameter estimates in the face of the inevitable uncertainty (e.g. measurement error) present in the observed data. In fact the likelihood function, describing the joint distribution of the data given the model and model parameters, is the central ingredient in both approaches to estimation. However, the two approaches are fundamentally different in how they treat model parameters resulting in different approaches to inference.

Historically, researchers working on nonlinear system identification problems have tended to focus on ML methods. In a recent review paper Kerschen et al. [1, see Sections 6 and 7 and the many references contained therein], cover numerous available techniques for nonlinear parameter estimation. These approaches are based on time domain (e.g. [2,3]), frequency domain (e.g. [4–6]) or higher-order spectral analysis [7,8]. Each of these techniques takes the best estimate to be the one that minimizes the mean square error between data and model. Although not explicitly stated, this choice of cost function yields ML estimates for the parameters, provided that the uncertainty in the model is taken as additive, iid, jointly Gaussian noise. The specific pros and cons associated with these methods are highlighted in [1] and are therefore not discussed here. Rather, this work is focused on the conceptually different Bayesian alternative to nonlinear parameter estimation problems.

* Corresponding author. Tel.: +1 860 486 4109; fax: +1 860 486 5088.
E-mail address: jonathan.nichols@nrl.navy.mil (J.M. Nichols).

0022-460X/$ - see front matter Published by Elsevier Ltd.
doi:10.1016/j.jsv.2010.02.004

Report Documentation Page

Form Approved
OMB No. 0704-0188

Public reporting burden for the collection of information is estimated to average 1 hour per response, including the time for reviewing instructions, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing the collection of information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing this burden, to Washington Headquarters Services, Directorate for Information Operations and Reports, 1215 Jefferson Davis Highway, Suite 1204, Arlington VA 22202-4302. Respondents should be aware that notwithstanding any other provision of law, no person shall be subject to a penalty for failing to comply with a collection of information if it does not display a currently valid OMB control number.

1. REPORT DATE **14 JUL 2009**	2. REPORT TYPE	3. DATES COVERED **00-00-2009 to 00-00-2009**
4. TITLE AND SUBTITLE **A Bayesian approach to identifying structural nonlinearity using free-decay response: Application to damage detection in composites**		5a. CONTRACT NUMBER
		5b. GRANT NUMBER
		5c. PROGRAM ELEMENT NUMBER
6. AUTHOR(S)		5d. PROJECT NUMBER
		5e. TASK NUMBER
		5f. WORK UNIT NUMBER
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) **U.S. Naval Research Laboratory,4555 Overlook Avenue,Washington,DC,20375**		8. PERFORMING ORGANIZATION REPORT NUMBER
9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)		10. SPONSOR/MONITOR'S ACRONYM(S)
		11. SPONSOR/MONITOR'S REPORT NUMBER(S)
12. DISTRIBUTION/AVAILABILITY STATEMENT **Approved for public release; distribution unlimited**		
13. SUPPLEMENTARY NOTES		
14. ABSTRACT **see report**		
15. SUBJECT TERMS		

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT **unclassified**	b. ABSTRACT **unclassified**	c. THIS PAGE **unclassified**	**Same as Report (SAR)**	**13**	

Standard Form 298 (Rev. 8-98)
Prescribed by ANSI Std Z39-18

Perhaps the chief benefit of the Bayesian worldview is that it treats the parameters we seek to estimate, denoted by the vector $\boldsymbol{\theta}$, as random variables with joint distribution $p(\boldsymbol{\theta})$. This allows us to make probabilistic statements about our estimated parameters. For example, let us say our goal is to collect data from a structure and estimate the crack length, denoted " θ_1" . The Bayesian framework allows us to make statements such as the following:

There is a 95 percent chance that the crack length θ_1 lies in the interval $0.1 \leq \theta_1 \leq 0.5$ [cm].

There is no way to make an analogous statement using the frequentist view of the world because the crack length would not be treated as a random variable. Instead we would have to make statements about the repeatability of our estimation procedure i.e. the statistical machinery that produced the interval estimate. Thus we might be able to make the statement "The methods we have used produce random intervals which may or may not include the unknown parameter; they succeed in 95 percent of similar circumstances, but we can't say anything with certainty about [0.1,0.5]" Although in some cases the two methods will produce the same estimate and interval the interpretation of the interval is different, the Bayesian interval being a probability statement about the parameter. Thus, our main reason for gravitating toward the Bayesian viewpoint is that it provides a direct estimate of the information we are after as opposed to related (but different) information. Additional reasons for the Bayesian approach are highlighted in Chapter 1 of [9]. First, the approach is more amenable to more complex data and/or models than is the frequentist (ML) approach (including cases where a frequentist approach does not exist). Secondly, the approach described here does not require asymptotics in approximating the confidence intervals (as does the frequentist approach of using the Fisher Information Matrix to bound the variance of a parameter estimate, see e.g., [10]).

This is certainly not the first paper to explore Bayesian methods in structural dynamics. A general Bayesian approach to structural dynamics problems was put forth in the '90s with works by Beck and Katafygiotis [11,12]. These, and subsequent works (see e.g., [13]), relied on asymptotics in order to solve for the marginal parameter distributions which, as will be shown, often take the form of high-dimensional integrals. At around the same time Sohn and Law [14] also developed a Bayesian approach to the structural identification problem. Their approach, also described in subsequent papers [15,16], circumvented the issue of solving for the desired marginal parameter distributions by using a so-called "branch-and-bound" strategy to instead solve for the most likely damage hypothesis in a damage detection application.

Rather than abandon the goal of parameter estimation or rely on asymptotics, one can make use of a powerful approach for sampling the marginal distributions without having to perform the integration. Such an approach was first proposed by Metropolis et al. [17] in the early 1950's for solving high-dimensional integrals in particle physics. A later work by Hastings [18] extended this general approach, resulted in what is now known as the Markov Chain Monte Carlo (MCMC) approach to approximating probability distributions. MCMC methods have since become extremely popular in implementing Bayesian estimation in a number of fields ranging from ecology [9] to genetics [19] and have recently seen increased use in engineering applications. Beck et al. appear to have pioneered the use of MCMC methods in structural dynamics in Beck and Au [20] and more recently in Cheung and Beck [21]. Additional work by Glaser et al. [22] illustrated the approach in detecting stiffness reduction in beams using static measurements. This method has also been used to estimate failure probabilities in structural reliability problems as part of the "Subset Simulation" approach of Au et al. [23]. In related works, Zhang and Cho [24] used the MCMC approach to help design an evolutionary algorithm for performing system identification, while Kerschen et al. [25] used the MCMC approach to select among competing models for describing the dynamics of a nonlinear mechanical system.

Our goal in this work is to focus on the use of the MCMC approach to Bayesian parameter estimation in nonlinear systems using free-vibration, time-series data. The approach is certainly more general and could be applied to forced structures as well. However, in practice the forcing function is not always obtainable while for the free-decay problem we simply have to include the initial conditions as random variables to be predicted. A different approach that does not require input data would be to perform the analysis using estimated frequencies as a means of comparing model to data as was done in [13]. However here the practitioner has the additional step of estimating the frequencies from observed data. This is compounded by the task of determining the analytical model frequencies which for nonlinear systems can be extremely challenging. Thus, free-decay response data provide for a direct model-to-data comparison and are easily obtainable in experiment. Another practical advantage of this general approach to system identification is that one does not need to measure time-series data from each of the structural degrees-of-freedom (DOF) in order to estimate the associated parameters. For example, a single time-series response from one of the DOFs can, in some cases, be used to estimate model parameters associated with DOFs not observed.

In this work the approach is first used to estimate the parameters associated with a two degree-of-freedom nonlinear structural system. The relationship between the quality of the resulting estimates and the signal-to-noise ratio is explored. We then turn our attention to the difficult problem of estimating and tracking delamination growth in a composite beam model. This model was recently developed in [26] where it was shown to accurately capture the localized buckling that occurs due to the separation of the laminates. Subsequent work by the authors [27] focused on detection of the delamination using a higher-order spectral analysis. Here, the focus is on identifying the damage parameters using only free vibration data, a task to which the Bayesian approach using MCMC to sample the desired parameter distributions is well-suited.

2. Bayesian approach to structural parameter estimation

In structural dynamics the uncertainty that gives rise to the likelihood function is typically the result of measurement error on the observed data. Assume that we have measured a structural response at a specific location giving the N-point time-series

$$z_n = y_n + \eta_n, \quad n = 1 \ldots N \tag{1}$$

where y_n is the noise-free response and η_n is a sequence of independent, identically distributed (iid) samples drawn from a zero-mean Gaussian distribution with variance σ_G^2. In this case we can write the likelihood

$$p_L(\mathbf{z}|\boldsymbol{\theta}, \sigma_G^2) = \frac{1}{(2\pi\sigma_G^2)^{N/2}} \exp\left[-\frac{1}{2\sigma_G^2}\sum_{n=1}^{N}(z_n - y_n(\boldsymbol{\theta}))^2\right] \tag{2}$$

where we have specified the dependence on the model parameters through the model response $y_n(\boldsymbol{\theta})$. The likelihood function $p_L(\mathbf{z}|\boldsymbol{\theta}, \sigma_G^2)$ is a probabilistic statement (model) about the distribution of the observed data $\mathbf{z} \equiv z_n \quad n = 1 \ldots N$ given a model response, determined by the parameter vector $\boldsymbol{\theta}$. The method of maximum likelihood estimates the model parameters to be those that maximize $p_L(\mathbf{z}|\boldsymbol{\theta}, \sigma_G^2)$. Some important and useful properties of MLEs is that, under regularity conditions, they are asymptotically unbiased and possess the minimum possible variance in the resulting parameter estimates. However, the approach suffers from the practical and conceptual drawbacks already mentioned in the introduction.

In the event that data from M structural locations are recorded the likelihood function can be easily altered as

$$p_L(\mathbf{z}|\boldsymbol{\theta}, \sigma_G^2) = \frac{1}{(2\pi\sigma_G^2)^{MN/2}} \exp\left[-\frac{1}{2\sigma_G^2}\sum_{m=1}^{M}\sum_{n=1}^{N}(z_n^{(m)} - y_n^{(m)}(\boldsymbol{\theta}))^2\right] \tag{3}$$

where the superscript (m) will henceforth be used to denote the degree-of-freedom being observed. In this case we have made two important but realistic assumptions. First, we have assumed that the sensor noise is uncorrelated. Secondly we have assumed that the sensor noise has the same variance across all sensors. Neither assumption is critical to the proposed approach—for example if the sensor noise was assumed different for each sensor we would simply identify each noise variance separately in the algorithm to be described in Section 3. This same likelihood function was used in [21]. It should be mentioned that the approach does not require all degrees-of-freedom to be measured, that is to say $M \leq DOF$. Of course the quality of the estimates typically improves with data from multiple measurement points, but is not strictly necessary. Finally, we should also point out here that this framework can accommodate multiple models $\mathcal{M}_i \quad i = 1 \ldots N_M$ such that the likelihood function becomes $p_L(\mathbf{z}|\boldsymbol{\theta}, \sigma_G^2, \mathcal{M}_i)$ (i.e. the likelihood is conditional on the specified model). However, in this work we will only consider a single model and drop the model notation $\mathcal{M}_i$. Additionally, in the following general discussion we consider the parameter vector to include the noise variance σ_G^2 for notational convenience.

Bayesian analysis relies on Bayes Theorem, relating the prior parameter probability density function (PDF) $p_\pi(\boldsymbol{\theta})$ to the posterior $p(\boldsymbol{\theta}|\mathbf{z})$ via

$$p(\boldsymbol{\theta}|\mathbf{z}) = p_L(\mathbf{z}|\boldsymbol{\theta})p_\pi(\boldsymbol{\theta})/p_D(\mathbf{z}). \tag{4}$$

The key ingredient relating the prior to posterior distributions is the likelihood function. Eq. (4) provides us a simple means of relating prior information to the parameter distributions we want. The term in the denominator, $p_D(\mathbf{z})$, is a normalizing constant that can be ignored in the following development. The joint parameter distribution contains the desired marginal distributions of each of the P parameters

$$p(\theta_j|\mathbf{z}) = \int_{\mathbb{R}^{P-1}} p(\boldsymbol{\theta}|\mathbf{z})\,\mathrm{d}\boldsymbol{\theta}_{-j} \propto \int_{\mathbb{R}^{P-1}} p_L(\mathbf{z}|\boldsymbol{\theta})p_\pi(\boldsymbol{\theta})\,\mathrm{d}\boldsymbol{\theta}_{-j} \tag{5}$$

where the notation $\int_{\mathbb{R}^{P-1}} d\boldsymbol{\theta}_{-j}$ denotes the multidimensional integral over all parameters other than θ_j. Thus, we require a means of performing a potentially high-dimensional integral involving a likelihood function for which we will often have a very complicated expression. For example, an analytical expression for y_n for a nonlinear system response is a very challenging problem for even a single DOF system possessing simple nonlinearities. Fortunately there exists a convenient numerical approach to sampling from the marginal parameter distributions.

3. Markov-chain Monte-Carlo methods

The term "Monte Carlo methods" is used to describe simulation techniques for investigating probability distributions. These techniques can be highly efficient, especially when independent samples can be generated. Unfortunately, posterior distributions used in Bayesian inference are often complicated, making it difficult to draw independent samples. Nevertheless it is often easy to draw a *dependent* sequence of samples representing posterior distributions. Over the last 60 years, stochastic algorithms have been developed which sample new values using rules determined by a fixed number of previous observations. The result is a Markov chain; this form of simulation is known as Markov chain Monte Carlo (MCMC). The magic of MCMC is in producing algorithms for which the resultant Markov chains have stationary distributions equal to the distribution we wish to sample.

3.1. Gibbs sampling

Gibbs sampling is a form of MCMC used to sample multivariate distributions. Consider a vector valued parameter $\boldsymbol{\theta}=(\theta_1,\theta_2,\ldots,\theta_P)'$, data $\mathbf{z}$, prior $p_\pi(\boldsymbol{\theta})$, and likelihood $p_L(\mathbf{z}|\boldsymbol{\theta})$. Define $\boldsymbol{\theta}_{-j}$ as the vector obtained by deleting the jth component of $\boldsymbol{\theta}$, viz.,

$$\boldsymbol{\theta}_{-j}=(\theta_1,\theta_2,\ldots,\theta_{j-1},\theta_{j+1},\ldots,\theta_P)'.$$

Gibbs sampling of the posterior distribution

$$p(\boldsymbol{\theta}|\mathbf{z})\propto p_L(\mathbf{z}|\boldsymbol{\theta})p_\pi(\boldsymbol{\theta})$$

produces a sequence (Markov Chain) $\{\boldsymbol{\theta}^{(i)}\}i=1\ldots$ ChainLength with components $\theta_j^{(i)}$ generated sequentially, for j=1,2,…,P. The value $\theta_j^{(i)}$ is sampled from its *full conditional distribution* $p(\theta_j^{(i)}|\mathbf{z},\boldsymbol{\theta}_{-j}=\boldsymbol{\theta}_{-j}^{(i-1)})$, which for simplicity is denoted as $p(\theta_j^{(i)}|\cdot)$. The full conditional distribution of θ_j can be thought of as the posterior distribution of θ_j, if it happened that the other $P-1$ parameters were known, and equal to their present values, i.e., if $\theta_k=\theta_k^{(i-1)}$ for $k\neq j$. Like the posterior distribution $p(\boldsymbol{\theta}|\mathbf{z})$, the full conditional distribution is proportional to the product of likelihood and prior,

$$p(\theta_j^{(i)}|\cdot)\propto p_L(\mathbf{z}|\boldsymbol{\theta})p_\pi(\boldsymbol{\theta});$$

the difference is that $p(\theta_j^{(i)}|\cdot)$ is a function of θ_j alone, rather than of all P variables comprising $\boldsymbol{\theta}$.

To implement Gibbs sampling, then, we only need to be able to sample full conditional distributions. The ease or difficulty of this endeavor depends on the form of the likelihood function and the prior. Among the easy cases is the circumstance of *conditional conjugacy*, when full conditional and prior are both identifiable as members of the same parametric family.

3.2. Conditional conjugacy

To illustrate the notion of conjugacy, consider z, the number of successes in N independent Bernoulli trials, with success parameter θ. A likelihood function for θ is

$$p_L(z|\,\theta)=\theta^z(1-\theta)^{N-z}.$$

A beta prior for θ on [0,1] has density function

$$p_\pi(\theta)\propto\theta^{\alpha-1}(1-\theta)^{\beta-1},$$

for some values $\alpha,\beta>0$. The posterior distribution is proportional to the product of likelihood and prior, hence is clearly a member of the beta family of distributions, with parameters $\alpha'=\alpha+z$ and $\beta'=\beta+N-z$. Beta priors are conjugate to the binomial likelihood.

Suppose that the parameters σ_G^2 and $\boldsymbol{\theta}$ are unknown in model (2) or (3). For reasons that will become clear subsequently, we reparameterize the model using the precision parameter $\tau=1/\sigma_G^2$, and show how conjugacy can be used in describing a full conditional distribution for τ. Given independent priors on $p_\pi(\tau)$ and $p_\pi(\boldsymbol{\theta})$, the posterior distribution is

$$p(\boldsymbol{\theta},\tau|\,\mathbf{z})\propto\tau^{MN/2}\exp\Big(-\frac{\tau}{2}\,Q(\mathbf{z},\boldsymbol{\theta})\Big)p_\pi(\tau)p_\pi(\boldsymbol{\theta}),$$

where

$$Q(\mathbf{z},\boldsymbol{\theta})=\sum_{i=1}^{M}\sum_{n=1}^{N}(z_n^{(i)}-y_n^{(i)}(\boldsymbol{\theta}))^2 \tag{6}$$

is the sum of squares in the likelihood. It follows that the full conditional distribution for τ is

$$p(\tau|\,\mathbf{z},\boldsymbol{\theta})\propto\tau^{MN/2}\exp\Big(-\frac{\tau}{2}\,Q(\mathbf{z},\boldsymbol{\theta})\Big)p_\pi(\tau).$$

The gamma family of distributions is described by density functions $p(x)\propto x^{\alpha-1}\exp(-\beta x)$ for $x>0$, indexed by parameters $\alpha,\beta>0$. Thus if τ has a gamma prior with parameters α and β, the full conditional distribution for τ is also in the gamma family, with parameters $\alpha'=\alpha+MN/2$ and $\beta'=\beta+Q(\mathbf{z},\boldsymbol{\theta})/2$. The gamma prior is conditionally conjugate for τ. We may therefore directly sample τ (hence σ_G^2) at each step in the Markov chain without having to resort to the more computationally intensive sampling approach described in the next section. In this work we use the so-called "diffuse" (uninformative) prior, obtained by setting $\alpha=1,\beta=0$. Conditional conjugacy simplifies Gibbs sampling by allowing reference to standard probability distributions (like the beta or gamma) readily available in standard software packages.

3.3. Metropolis–Hastings algorithm

Unfortunately, conjugate or conditionally conjugate priors are not always available. In such cases, techniques such as rejection sampling can be used for sampling the full conditional distribution. One may also resort to the omnibus *Metropolis–Hastings algorithm*, a simple implementation of which we now describe.

Suppose that we wish to draw samples from a specific distribution $f(\theta)$ (in our case $f(\theta)$ might be the left-hand-side of Eq. 5, $p(\theta_j)$). The Metropolis–Hastings algorithm generates a sequence $\{\theta^{(i)}\}$ by a 2-step procedure. At stage i, a candidate value θ^* is sampled based on the current value $\theta^{(i-1)}$; it is sampled from a distribution function $g(x|\theta^{(i-1)})$. A Bernoulli trial is performed, with success probability $r' = \min\{r, 1\}$, where

$$r = \frac{f(\theta^*)g(\theta^{(i-1)}|\theta^*)}{f(\theta^{(i-1)})g(\theta^*|\theta^{(i-1)})};$$

if the result of the trial is success, we set $\theta^{(i)} = \theta^*$; otherwise, we set $\theta^{(i)} = \theta^{(i-1)}$. The alternatives are often described as "moving" or "staying at the current value", but it is important to note that in the latter case, the same value is recorded twice (once as $\theta^{(i-1)}$, once as $\theta^{(i)}$).

It is often easy to describe a candidate generating distribution which is symmetric in its arguments, hence cancels out of the numerator and denominator of r. For example, if θ^* is uniformly distributed on an interval of length $2A$ centered on $\theta^{(i-1)}$, then its density function is

$$g(x|\theta^{(i-1)}) = \frac{1}{2A}, \quad |x - \theta^{(i-1)}| < A, \tag{7}$$

so that $g(\theta^{(i-1)}|\theta^*) = g(\theta^{(i-1)}|\theta^*)$.

Suppose that we use this uniform candidate generator, with $A = 1$, and apply the technique to generating samples from the standard normal distribution $f(\theta) = \exp(-\theta^2/2)/\sqrt{2\pi}$. Then r simplifies to

$$r = \exp\left(\frac{(\theta^{(i-1)})^2 - (\theta^*)^2}{2}\right).$$

Two features of this remarkable algorithm deserve comment. First, that the first-order Markov chain it produces will have stationary distribution equal to $f(\theta)$, provided only that the candidate generating distribution and the starting value $\theta^{(0)}$ allow the chain to reach all values t for which $f(\theta) > 0$. Thus in generating standard normal variates with the uniform candidate generator, it does not matter if we start the chain at $\theta^{(0)} = 25$, even though such values are exceedingly rare for a standard normal variate. Most of the first 50 or so moves toward zero will all be accepted, while most of the first 50 or so moves away from zero will all be rejected, and the chain will rush in towards the range $(-3,3)$ typical of standard normal variables. These initial values are discarded in MH sampling, described as a "burn-in".

The second important feature of this algorithm is that its implementation depends on the desired distribution only through the ratio in r. Thus in calculating r for the standard normal distribution, the constant terms $\sqrt{2\pi}$ cancel. This is an enormous boon in calculating full conditional distributions, which are defined implicitly as the distribution proportional to product of likelihood and prior; the normalizing constant in Eq. (4), which makes the distribution integrate to 1, need not be calculated.

Practical implementation of the MCMC technique can require some tuning. Large values for the "tuning parameter" A (Eq. (7)) decreases the likelihood of accepting a new value, thus the samples in the chain tend to be highly correlated. Too small a step, however, and the algorithm can take a prohibitively long time to converge. The tradeoff between excessive computation time (small steps) and not generating independent samples (big steps) is well-known. Here we use a simple approach whereby the tuning parameter is adjusted on-the-fly in order to achieve an appropriate acceptance rate of between 30 percent and 50 percent. This range of acceptance probabilities has been demonstrated to produce Markov chains with low auto-correlation (good mixing) [9]. During the burn-in period we simply divide A by the constant value 1.007 after each rejection and multiply by 1.01 after each acceptance. Thus, an acceptance causes us to expand our parameter search while rejection results in smaller "kicks" to the previous value in the chain. The asymmetry in the constants causes a slight bias in the favor of rejection. This simple approach is quite effective at producing acceptance rates of 40 percent, a good target for the MCMC algorithm. Details of this approach are discussed in [9].

Algorithm 1. The MCMC algorithm using Metropolis–Hastings with Gibbs sampling for Gaussian likelihood and Uniform "transition" distribution $g(\theta_j^*, \theta_j(i-1)) = 1/(2A_j)$

Task: Generate the posterior parameter distributions $p(\theta_j)$ for $j = 1 \ldots P$ given the model, the observed data $\mathbf{z} \equiv z_n$ $n = 1 \ldots N$, and prior parameter distributions $p_{\pi_j}(\theta_j)$. Also estimate the noise variance $\theta_{P+1} = \sigma_G^2$

Initialization: Initialize chain $i = 0$

Initial guesses for parameter values (chosen from the priors) $\boldsymbol{\theta}(0) \equiv \theta_j(0)$ $j = 1 \ldots P$

Initial values for tuning parameters A_j $j = 1 \ldots P$

Sample initial variance from inverse Gamma distribution

$\theta_{P+1}(0) = \mathrm{IG}(N/2 + 1, 0.5Q(\mathbf{z}, \boldsymbol{\theta}(0))$

Set number of burn-in iterations B

Main iteration:increment i by 1 and apply

For each parameter j

Generate candidate $\theta_j^* = \theta_j(i-1) + 2A_j \times U(-1, 1)$ where $U(a,b)$ is a uniformly distributed number on $[a,b]$

Compute $r = \exp((-0.5/\theta_{P+1}(i-1)) \times (Q(\mathbf{z}, \boldsymbol{\theta}^*) - Q(\mathbf{z}, \boldsymbol{\theta}(i-1)))) * p_{\pi_j}(\boldsymbol{\theta}^*)/p_{\pi_j}(\boldsymbol{\theta}(i-1))$ where $\boldsymbol{\theta}^* \equiv (\theta_1(i), \cdots, \theta_{j-1}(i), \theta_j^*, \theta_{j+1}(i-1), \cdots, \theta_P(i-1))$ and $\boldsymbol{\theta}(i-1) \equiv (\theta_1(i), \cdots, \theta_{j-1}(i), \theta_j(i-1), \theta_{j+1}(i-1), \cdots, \theta_P(i-1))$

If $U(0,1) < r$ keep the new value, i.e. set $\theta_j(i) = \theta_j^*$, and adjust tuning parameter $A_j = A_j \times 1.01$
Else reject the new value, keeping $\theta_j(i) = \theta_j(i-1)$ and adjust tuning parameter $A_j = A_j/1.007$.
Directly sample the variance posterior $\theta_{P+1}(i) = \mathrm{IG}(N/2+1, 0.5Q(\mathbf{z}, \boldsymbol{\theta}(i))$
After $i > B$ iterations, cease adjusting the A_j and record subsequent
values $\theta_j(i)$ as members of $p(\theta_j)$.

The algorithm for executing this specific implementation of MCMC is given in Algorithm 1. This pseudo-code implicitly assumes (1) a diffuse Gamma prior on the precision parameter $1/\sigma_G^2$, (2) a Gaussian likelihood function, governed by the sum-squared error between data and model and (3) a uniform transition distribution (Eq. (7)). The code is quite simple to implement, the costly step being the determination of the sum-squared error function $Q(\cdot,\cdot)$ (Eq. (6)). For nonlinear systems, for which no analytical solution exists, generating the model output y_n for a given parameter vector requires numerically integrating the equations of motion. Also note that in the above described implementation, as the Gibbs sampler moves through the parameter vector the newly updated values are used. That is to say, for $j=2$ we use $\theta_1(i)$ rather than $\theta_1(i-1)$ as the first element of $\boldsymbol{\theta}(i-1)$. Finally, many statistical packages provide routines that generate samples from a Gamma distribution but not from an inverse Gamma distribution. Fortunately there is a simple relationship between the Gamma $G(\cdot,\cdot)$ and inverse Gamma $IG(\cdot,\cdot)$ distributions

$$IG(a,b) = 1/G(a, 1/b)$$

so that for our problem we may sample the variance by drawing

$$\theta_{P+1}(i) \sim 1.0/[\mathrm{G}(N/2+1, 2/Q(\mathbf{z}, \boldsymbol{\theta}(i))].$$

There are obviously many variations on the above described algorithm, however Algorithm 1 presents a simple but useful implementation that is straightforward to code in software.

4. Example 1: 2-DOF nonlinear spring–mass-damper system

In order to illustrate the above described identification procedure, consider the two degree-of-freedom (DOF) system described by the second order, nonlinear, ordinary differential equations

$$[M]\ddot{\mathbf{y}}_t + [C]\dot{\mathbf{y}}_t + [K]\mathbf{y}_t + \mathbf{g}(\dot{\mathbf{y}}_t, \mathbf{y}_t) = 0 \tag{8}$$

where

$$[M] = \begin{bmatrix} m_1 & 0 \\ 0 & m_2 \end{bmatrix} \quad [C] = \begin{bmatrix} c_1+c_2 & -c_2 \\ -c_2 & c_2 \end{bmatrix} \quad [K] = \begin{bmatrix} k_1+k_2 & -k_2 \\ -k_2 & k_2 \end{bmatrix}$$

are constant coefficient mass, damping and stiffness matrices respectively. The nonlinear function $\mathbf{g}(\cdot)$ provides quadratic coupling between masses. Here we consider a quadratic restoring force between masses 1 and 2 so that

$$\mathbf{g}(\dot{\mathbf{y}}_t, \mathbf{y}_t) = \begin{Bmatrix} -k_{non}(y_t^{(2)} - y_t^{(1)})^2 \\ k_{non}(y_t^{(2)} - y_t^{(1)})^2 \end{Bmatrix}$$

where k_{non} is the nonlinear stiffness coefficient. We consider as the observed data $N=512$ points of a noise-contaminated, free-decay response, obtained by numerically integrating Eq. (8) with a time-step of $\Delta_t = 0.01\,\mathrm{s}$ so that discrete and continuous time are related via $t = n\Delta_t$. The initial conditions were set to the values $y_0^{(1)} = 0.0\,\mathrm{m}$, $v_0^{(1)} \equiv \dot{y}_0^{(1)} = 0.0\,\mathrm{m/s}$, $y_0^{(2)} = 0.15\,\mathrm{m}$, and $v_0^{(2)} \equiv \dot{y}_0^{(2)} = 0.0\,\mathrm{m/s}$. The parameter values used in generating the response data were $k_1=2000\,\mathrm{N/m}$, $k_2=1500\,\mathrm{N/m}$, $c_1 = c_2 = 6\,\mathrm{N\,s/m}$, and $k_{non}=10{,}000\,\mathrm{N/m^2}$. The mass parameters were fixed to the values $m_1=m_2=1\,\mathrm{kg}$ and were assumed known (measured) at the outset. The variance of the additive, Gaussian distributed noise is defined by the signal-to-noise ratio $SNR = \sigma_y^2/\sigma_G^2$ so that the observed data are given by

$$z_n^{(1)} = y_n^{(1)} + \sqrt{\frac{\sigma_y^2}{SNR}}\eta_n$$

where each η_n is an iid Gaussian random variable and σ_y^2 is the variance of the true underlying signal. We are assuming in this example that only a single time-series measurement is available in forming the likelihood (i.e. $M=1$ in Eq. (3)). For this example the additive noise level was set at $-10\,\mathrm{dB}$ down, i.e. the signal variance was $10\times$ greater than the noise variance. The job of the algorithm is to identify c_1,c_2,k_1,k_2 as well as the nonlinear parameter k_{non}, the initial conditions $y_0^{(1,2)}$, $v_0^{(1,2)}$, and the noise variance σ_G^2.

Fig. 1 shows the resulting stationary parameter distributions for each of the identified quantities. Each of the model parameters were identified using the above-described Metropolis–Hastings algorithm with Gibbs sampling and assuming Uniform priors. The exception was the variance σ_G^2 which could be sample directly by assuming a conjugate Gamma prior as described in Section 3.2. In other words, at each step in the chain we choose $\sigma_G^2 \sim 1/G(MN/2, \frac{2}{Q(\mathbf{z},\boldsymbol{\theta})})$ where $G(\cdot)$ is the Gamma distribution. In building the Markov Chains we used a burn in of 150,000 iterations and retained the subsequent

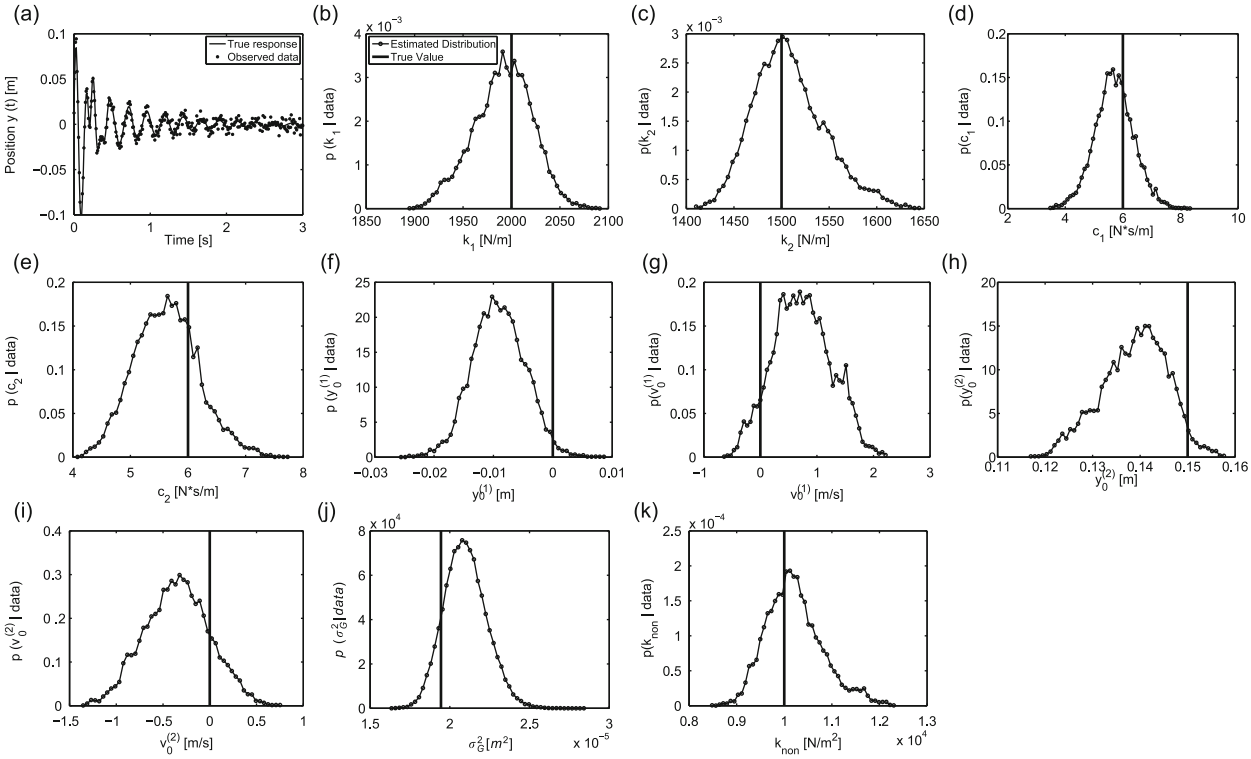

Fig. 1. True model output and corrupted (observed) data (a) followed by the posterior distributions of, respectively (b–k), k_1, k_2, c_1, c_2, $y_0^{(1)}$, $v_0^{(1)}$, $y_0^{(2)}$, $v_0^{(2)}$, σ_G^2, and k_{non}.

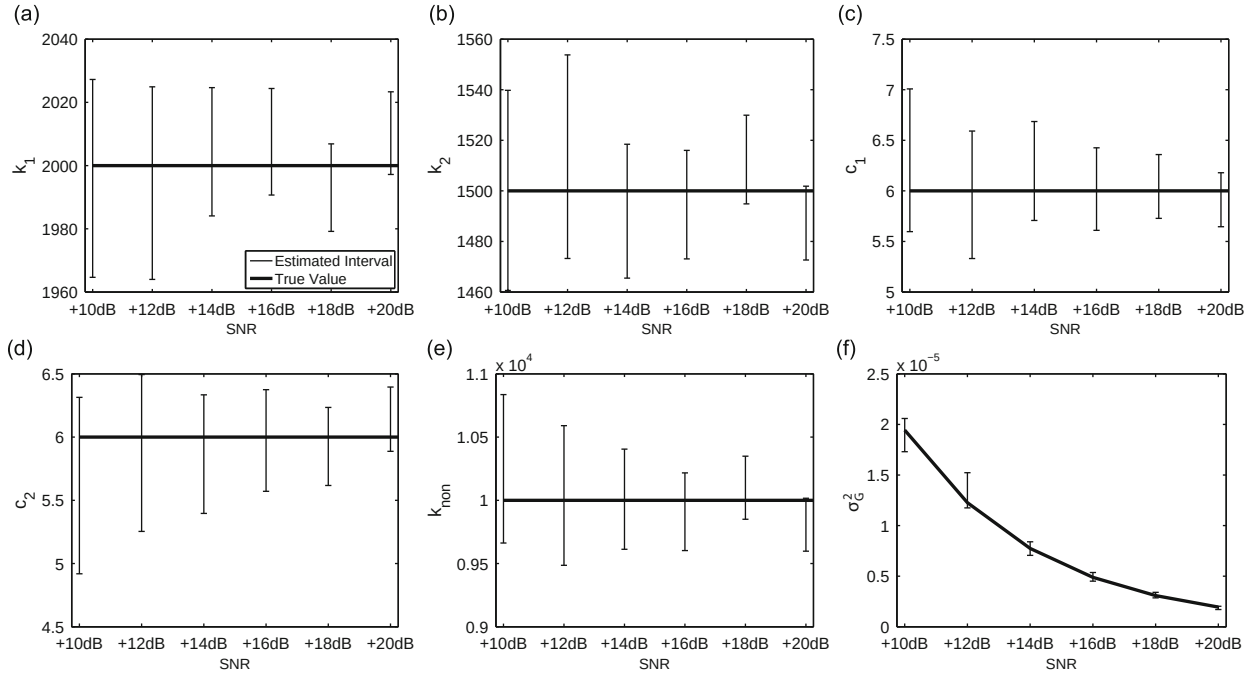

Fig. 2. Progression of the 95 percent confidence interval associated with the parameter estimates as a function of signal-to-noise ratio: (a) linear stiffness, k_1 (b) linear stiffness, k_2 (c) linear damping, c_1 (d) linear damping, c_2 (e) nonlinear stiffness, k_{non} and (f) noise variance, σ_G^2. As expected, a higher SNR tends to result in better estimates.

50,000 iterations as the stationary posterior distributions of interest. Fig. 1 shows the noise-corrupted time-series data used in the identification along with all identified parameters. The algorithm is clearly able to identify both the system parameters and the initial conditions using only a short, noisy free-decay response as the observed data. Certain parameters are identified with more confidence than others. Notoriously difficult to obtain damping estimates have the largest percentage variance of all parameters while the stiffness estimates are better defined. The nonlinearity parameter is similarly estimated to lie within a fairly narrow confidence band.

As expected, the distributions tend to narrow as the signal-to-noise ratio is increased. Fig. 2 shows the progression of the 95 percent confidence intervals, defined as the central 95 percent of the 50,000 values that comprise the stationary Markov chain, as a function of SNR. Using this approach each of the parameters is correctly identified with their associated confidence intervals. There is no need to base the confidence intervals on an assumed Gaussian distribution as is often done in parameter estimation (e.g. assume ± 3 standard deviations from the mean as 95 percent confidence). The MCMC algorithm provides an approximation of the entire posterior PDF, regardless of its underlying form. The above results have demonstrated that we can successfully identify both linear and nonlinear system parameters in a multi DOF structure using only measurements of noisy, free-decay response data. The next example involves the more difficult problem of detecting delamination in composite structures.

5. Delamination identification in a composite beam

As a second example, we seek to identify the location, extent, and depth of delamination in a composite beam structure. The dynamic beam model was derived previously in [27] and is shown schematically in Fig. 3. This model is low dimensional (only three independent coordinates need to be specified), yet was shown experimentally to accurately capture the localized buckling that occurs due to the presence of the delamination under static loading [26]. The global beam motion is assumed to be dictated by the first mode of the response, thus the global displacement assumes the form

$$y_1(x_1,t) = q_1(t)\psi_1(x_1)$$

where

$$\psi_1(x) = \frac{3}{2}\left[\left(\frac{x}{L}\right)^2 - \frac{1}{3}\left(\frac{x}{L}\right)^3\right]$$

is a normalized shape function describing the vertical beam deflection at any point x, measured from the left end of the beam. The other two coordinates, describing the time-dependent motion of the upper (region 2) and lower (region 3) laminates respectively, are assumed to be of the form

$$y_2(x_2,t) = y_1(x_2+x_a,t) + q_2(t)\Psi_2(x_2) + \frac{1}{2}(1-a)h$$

$$y_3(x_3,t) = y_1(x_3+x_a,t) + q_3(t)\Psi_3(x_3) - \frac{ah}{2}$$

$$\Psi_{2,3} = 1-\cos^2\left(\pi\frac{x_{2,3}-x_a}{x_b-x_a}\right) \tag{9}$$

where the shape functions $\Psi_{2,3}$ describe the deflected shape of regions 2 and 3 in Fig. 3. The constant terms in Eq. (9) denote the neutral axis offsets, measured from the global neutral axis, of each of the laminates. Substituting these expressions into Lagrange's equation yields a set of three coupled, nonlinear differential equations in terms of the time-dependent vector $\mathbf{q}(t) \equiv (q_1(t), q_2(t), q_3(t))$

$$[M]\ddot{\mathbf{q}}_t + [C]\dot{\mathbf{q}}_t + [K_L]\mathbf{q}_t + [K_M]q_1(t)\mathbf{q}_t + [K_Q]\mathbf{q}_t^2 + [K_C]\mathbf{q}_t^3 = 0. \tag{10}$$

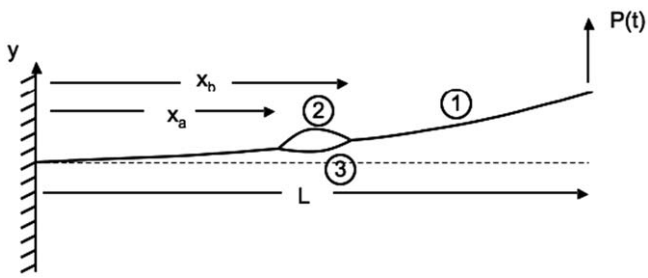

Fig. 3. Schematic of the dynamic delaminated beam model. Region 1 is simply modeled as a linear cantilevered beam whose motion is governed by the first mode of vibration. Regions 2 and 3 are modeled as nonlinear beams where axial stretching is permitted.

The equations of motion are therefore described in terms of the material properties of the beam, the dimensions of the beam, and the parameters associated with the damage. These damage parameters are x_a, the delamination starting location, x_b, the delamination end point, and a, the delamination depth specified as a fraction of the overall beam thickness h. Additionally, we have chosen to add a viscous damping model for both the global vibration of the beam and the local vibrations of the delaminated portions of the beam. Expressions for each of the constant coefficient matrices are provided in Appendix A.

In the following analysis it is assumed that there is no uncertainty in the beam length or material properties. We therefore set L=0.24 m, h=2.25 mm EI=75,889,600,000 GPa, and $\rho = 1234.0$ kg/m. The damage parameters, x_a,x_b,a are assumed to be unknown as are the viscous damping coefficients c_1,$c_{2,3}$ (see Eq. (A6) and the initial global beam deflection $y_1(0)$). Again we assume that we will have access to a single, noise corrupted signal and that the noise is additive, iid Gaussian distributed. Thus the likelihood function is the same as was used in the previous example assuming only one signal (in this case the global motion) can be recorded. Previous work by the authors demonstrated that global, vibration-based detection was not possible unless the measurement was recorded from near the delamination site [27]. This stems from the fact that the only coupling between the laminate vibrations and the rest of the beam is inertial. Thus, for relatively small delaminations (what we are interested in) there is little in the way of influence with respect to the global motion away from the delamination site. The small influence of the delamination on the global response, even at the delamination site, makes this a particularly challenging system identification problem. In what follows we assume that we are able to record the beam response from the delamination site. Again we assume a transient response to an initial displacement and use N=512 sampled points at a sampling interval of $\Delta_t = 0.001$ s. The variance of the additive Gaussian noise was fixed such that the signal-to-noise ratio was SNR=+20 dB.

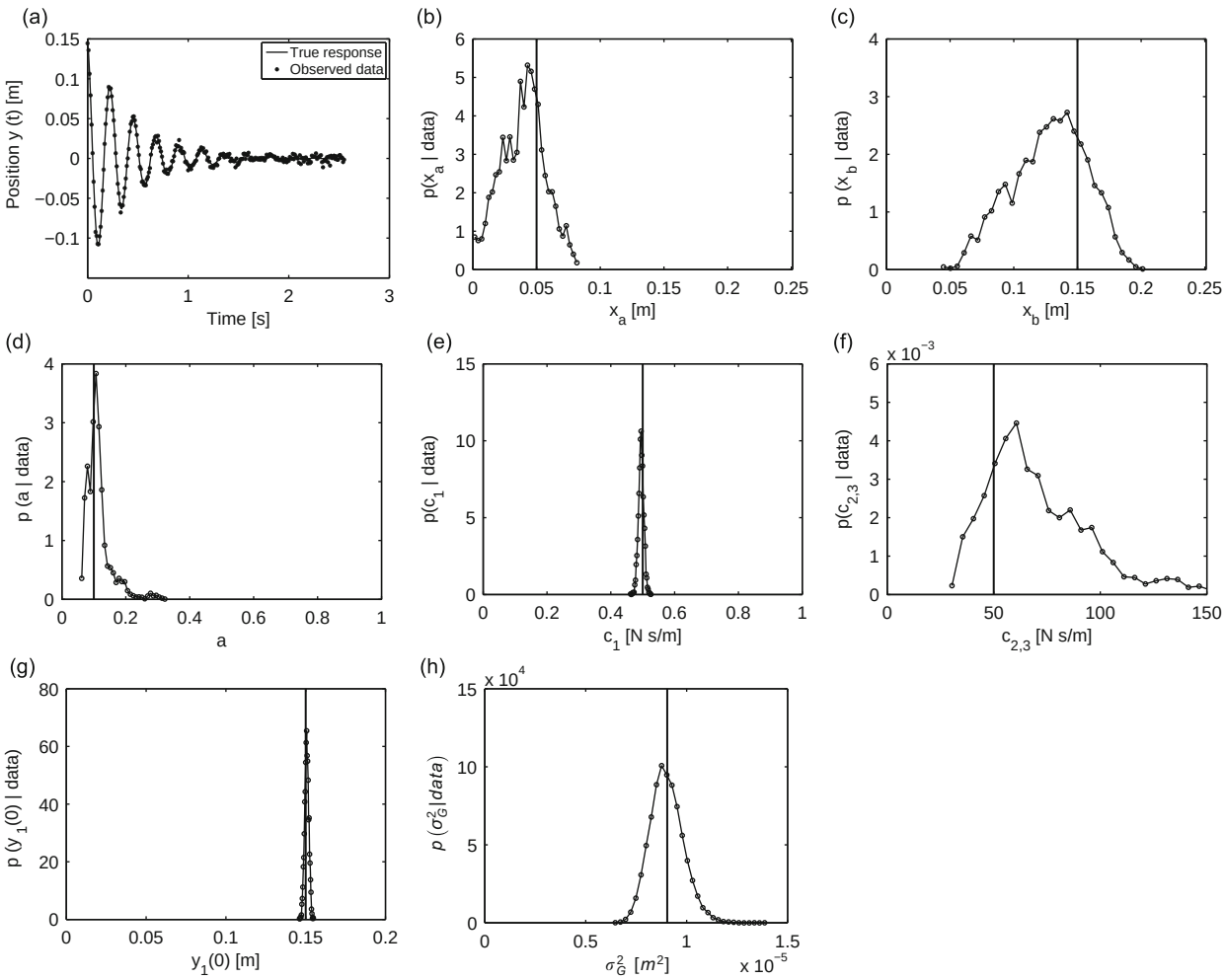

Fig. 4. Actual and observed impulse response data (a) followed by posterior distributions of, respectively (b–h), delamination start point x_a, delamination end point x_b, delamination depth a, global coefficient of viscous damping c_1, local damping coefficient $c_{2,3}$, initial beam tip deflection $y_1(0)$, and noise variance σ_G^2.

Fig. 4 shows the observed (noise corrupted) data along with the true underlying free-decay signal. Also shown are the estimated structure parameter distributions, initial condition distributions, and the distribution of the noise variance along with the true values for each parameter. Despite the *very* small influence of the nonlinearity on the response the algorithm is still able to identify the delamination start and end points as well as the depth. The delamination depth appears easier to identify as evidenced by the very narrow confidence interval around the correct value of $a=0.1$. The delamination start and end points have very little influence on the global response as was illustrated in [27], hence the confidence intervals for these two parameters are large. Thus, the method gives us a clear indication of the degree to which we can trust our estimates. In this case we can be fairly certain of the delamination depth, but much less certain about our ability to estimate both the beginning and end points of the delamination. If our application requires a tighter confidence interval we might need to try a larger excitation (allowing the nonlinearity to more strongly influence the response) or maybe even change to a local damage detection method. Regardless, the information about our faith in the ability to estimate these parameters is clearly valuable information. Both the linear damping c_1, the initial beam deflection $y_1(0)$ and the noise variance σ_G^2 are easily estimated with a high degree of confidence. These parameters clearly have a large influence on the global response thus one might expect them to be easily obtained.

It turns out that shallow delaminations, such as the $a=0.1$ case just presented, are more difficult to identify than thicker ones. Consider the case of a delamination of depth $a=0.2$ with the start and end points of the delamination fixed at $x_a=0.05$ m and $x_b=0.2$ m respectively. As in the previous example the global damage parameters, c_1,$y_1(0)$ are trivial to identify. The difficult to identify parameters, x_a,x_b,a,$c_{2,3}$ are shown in Fig. 5. The variance associated with the estimates of each of the parameters is reduced from the previous case. This example points out that certain combinations of parameters are more easily identified than others. It depends to large extent on the degree to which the estimated model parameters influence the observed data. Parameters with little effect on the observed response will be hard to identify whereas the converse is also true.

As a final example we demonstrate how the approach can be used to track damage in a structure. For this example, the delamination starting point and depth were fixed to the values $x_a=0.05$ m and $a=0.2$ respectively. The delamination end point was slowly varied from $x_b=0.1$ m up to $x_b=0.2$ m. Again, using only the noisy free-decay response the goal was to estimate and track the delamination end-point. Fig. 6 shows the progression of estimated delamination length along with the associated 95 percent confidence interval. Also shown are the "true" values for x_b used in generating the time-series. For each damage case we used a prior generated from the previous case. Specifically, we selected the

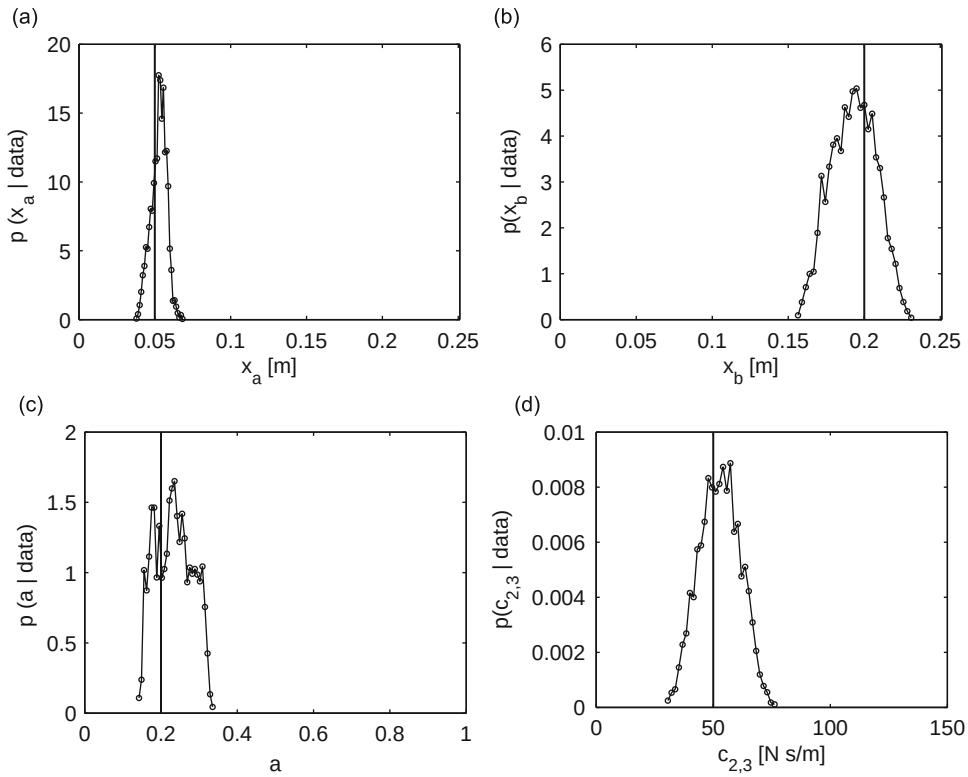

Fig. 5. Calculated posterior distributions of, respectively, (a–d) delamination start point x_a, delamination end point x_b, delamination depth a, and local damping coefficient $c_{2,3}$.

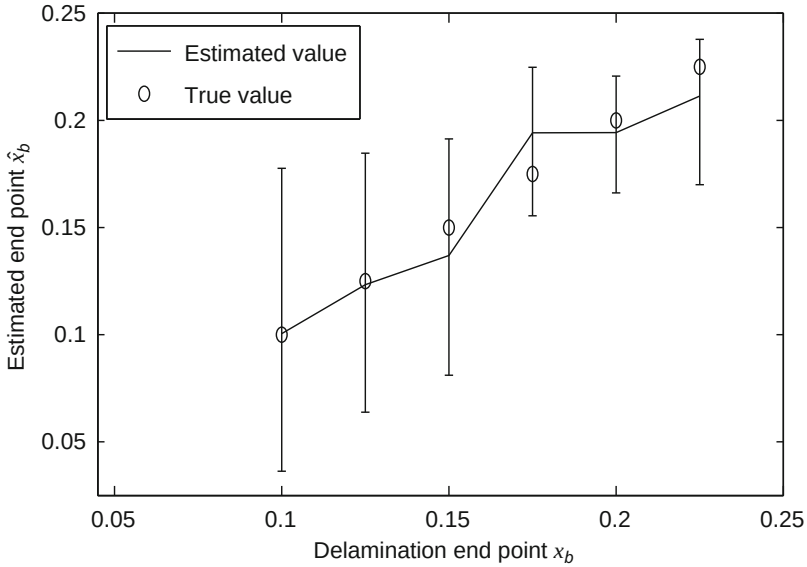

Fig. 6. Estimated delamination end point $\hat{x}_b$ as a function of the actual delamination end point. The intervals of confidence were created as the central 95 percent of the sampled values from $p(x_b)$ obtained using the MCMC algorithm.

prior from a uniform distribution centered at the previous mean value with a width of roughly two standard deviations. For the first damage case (x_b=0.1 m) we also assumed a uniform prior on the range $x_b \sim U(x_a = 0.05, 0.1)$ (i.e. assume an initial delamination length between 0 and 0.05 m). Again, for this particular example the damage parameters x_a,x_b,a have very little influence on the global vibrational response, thus the confidence intervals tend to be relatively large. However the algorithm is clearly tracking the progression of the delamination with reasonable accuracy. By the time the delamination is large with respect to the beam length, the confidence intervals narrow considerably. The large nonlinearities produce a more readily identifiable signature in the global response, hence the associated nonlinearity parameters are more easily identified. The information provided in Fig. 6 is precisely the information that the owner of a structure would be interested in: the estimated damage extent and associated confidence in that estimate. Given this information and the cost associated with the structural damage, one can make optimal decisions regarding how best to maintain that structure.

6. Conclusions

This work has presented an approach for identifying nonlinear, multi-degree-of-freedom structures using only observed free-decay response data. The approach appears to work well for limited, noise corrupted observations that are easily obtainable in experiment. This makes the approach attractive from a practical standpoint with the main drawback being the computational effort required to build the stationary Markov chains. At each step in building process the practitioner is required to numerically integrate the model forward in time for each parameter being identified. Nonetheless, the approach is effective at identifying both linear and nonlinear model parameter distributions for structural models. This is precisely the information that is of interest in damage detection problems. Here we have show the approach can be used to estimate and track the length of delamination in a composite beam using only noisy, free decay response data. When combined with knowledge of the costs associated with structural damage, this approach provides the information necessary for making optimal maintenance decisions.

Acknowledgment

The authors would like to acknowledge the Office of Naval Research under contract No. N00014-09-WX-2-1002 for providing funding for this work.

Appendix A

Eq. (10) was derived using energy methods. The constant coefficient matrices associated with these governing equations are given here. The linear stiffness matrix $[K_L]$ is given by

$$[K_L]=\begin{pmatrix} \frac{bh^3(-4L^3+3(a-1)a(x_b-x_a)^3)E}{16L^6} & 0 & 0 \\ 0 & \frac{a^3bh^3\pi^4E}{6(x_b-x_a)^3} & 0 \\ 0 & 0 & \frac{(1-a)^3bh^3\pi^4E}{6(x_b-x_a)^3} \end{pmatrix}. \tag{A1}$$

The two matrices involving quadratic terms are defined as

$$[K_M]=\begin{pmatrix} 0 & 0 & 0 \\ 0 & -\frac{3(1-a)abh^2\pi^2(2L-x_a-x_b)E}{8L^3(x_b-x_a)} & 0 \\ 0 & 0 & \frac{3(1-a)abh^2\pi^2(2L-x_a-x_b)E}{8L^3(x_b-x_a)} \end{pmatrix}$$

$$[K_Q]=\begin{pmatrix} 0 & -\frac{3(1-a)abh^2\pi^2(2L-x_a-x_b)E}{16L^3(x_b-x_a)} & \frac{3(1-a)abh^2\pi^2(2L-x_a-x_b)E}{8L^3(x_b-x_a)} \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}. \tag{A2}$$

The matrix multiplying the cubic term is given by

$$[K_C]=\begin{pmatrix} 0 & 0 & 0 \\ 0 & \frac{abh\pi^4E}{8(x_b-x_a)^3} & 0 \\ 0 & 0 & \frac{(1-a)bh\pi^4E}{8(x_b-x_a)^3} \end{pmatrix} \tag{A3}$$

The mass matrix is slightly more complicated. Defining

$$\alpha=\frac{-1}{16L^3\pi^2}bh(x_b-x_a)[(x_a+x_b)\rho((-3+\pi^2)x_a^2+6x_ax_b+(-3+\pi^2)x_b^2)-2L((-3+2\pi^2)x_a^2+2(3+\pi^2)x_ax_b+(-3+2\pi^2)x_b^2)] \tag{A4}$$

the mass matrix is given by

$$[M]=\begin{pmatrix} \frac{33}{140}bhL\rho & a\alpha & (1-a)\alpha \\ a\alpha & \frac{3}{8}abh\rho(x_b-x_a) & 0 \\ (1-a)\alpha & 0 & \frac{3}{8}(1-a)bh\rho(x_b-x_a) \end{pmatrix} \tag{A5}$$

The damping for both the global beam motion and laminates is assumed to follow a viscous model such that we have

$$[C]=\begin{pmatrix} c_1 & 0 & 0 \\ 0 & c_{2,3} & 0 \\ 0 & 0 & c_{2,3} \end{pmatrix} \tag{A6}$$

References

[1] G. Kerschen, K. Worden, A.F. Vakakis, J.-C. Golinval, Past, present, and future of nonlinear system identification in structural dynamics, *Mechanical Systems and Signal Processing* 20 (2006) 505–592.
[2] K.S. Mohammad, K. Worden, G.R. Tomlinson, Direct parameter estimation for linear and nonlinear structures, *Journal of Sound and Vibration* 152 (1991) 471–499.
[3] K. Worden, J.R. Wright, M.A. Al-Hadid, Experimental identification of multi-degree-of-freedom nonlinear systems using restoring force methods, *International Journal of Analytical and Experimental Modal Analysis* 9 (1994) 35–55.
[4] M. Feldman, Non-linear system vibration analysis using Hilbert transform—i. Free vibration analysis method "freevib", *Mechanical Systems and Signal Processing* 8 (2) (1994) 119–127.
[5] G. Kerschen, V. Lenaerts, J.C. Golinval, Identification of a continuous structure with a geometrical non-linearity, part i: conditioned reverse path method, *Journal of Sound and Vibration* 262 (2003) 889–906.

[6] D.E. Adams, R.J. Allemang, A frequency domain method for estimating the parameters of a non-linear structural dynamic model through feedback, *Mechanical Systems and Signal Processing* 14 (2000) 637–656.
[7] Z.K. Peng, Z.Q. Lang, S.A. Billings, Nonlinear parameter estimation for multi-degree-of-freedom nonlinear systems using nonlinear output frequency response functions, *Mechanical Systems and Signal Processing* 22 (7) (2008) 1582–1594.
[8] J.M. Nichols, P. Marzocca, A. Milanese, On the use of the auto-bispectral density for detecting quadratic nonlinearity in structural systems, *Journal of Sound and Vibration* 312 (4-5) (2008) 726–735.
[9] W.A. Link, R.J. Barker, *Bayesian Inference with Ecological Examples*, Academic Press, San Diego, CA, 2010.
[10] R.N. McDonough, A.D. Whalen, *Detection of Signals in Noise*, second ed., Academic Press, San Diego, 1995.
[11] J.L. Beck, L.S. Katafygiotis, Updating models and their uncertainties. i: Bayesian statistical framework, *Journal of Engineering Mechanics* 124 (4) (1998) 455–461.
[12] L.S. Katafygiotis, J.L. Beck, Updating models and their uncertainties. ii: Model identifiability, *Journal of Engineering Mechanics* 124 (4) (1998) 463–467.
[13] J.L. Beck, K.-V. Yuen, Model selection using response measurements: Bayesian probabilistic approach, *ASCE Journal of Engineering Mechanics* 130 (2) (2004) 192–203.
[14] H. Sohn, K.H. Law, A Bayesian probabilistic approach for structure damage detection, *Earthquake Engineering and Structural Dynamics* 26 (1997) 1259–1281.
[15] H. Sohn, K.H. Law, Bayesian probabilistic damage detection of a reinforced-concrete bridge column, *Earthquake Engineering and Structural Dynamics* 29 (2000) 1131–1152.
[16] H. Sohn, K.H. Law, Application of load-dependent Ritz vectors to Bayesian probabilistic damage detection, *Probabilistic Engineering Mechanics* 15 (2000) 139–153.
[17] N. Metropolis, A.W. Rosenbluth, M.N. Rosenbluth, A.H. Teller, E. Teller, Equation of state calculations by fast computing machines, *The Journal of Chemical Physics* 21 (6) (1953) 1087–1092.
[18] W.K. Hastings, Monte carlo sampling methods using Markov chains and their applications, *Biometrika* 57 (1) (1970) 97–109.
[19] M. Fang, D. Jiang, H.J. Gao, D.X. Sun, R.Q. Yang, Q. Zhang, A new Bayesian automatic model selection approach for mapping quantitative trait loci under variance component model, *Genetica* 135 (3) (2009) 429–437.
[20] J.L. Beck, S.-K. Au, Bayesian updating of structural models and reliability using Markov chain monte carlo simulation, *Journal of Engineering Mechanics—ASCE* 128 (4) (2002) 380–391.
[21] S.H. Cheung, J.L. Beck, Bayesian model updating using hybrid Monte Carlo simulation with application to structural dynamic models with many uncertain parameters, *Journal of Engineering Mechanics—ASCE* 135 (4) (2009) 243–255.
[22] R.E. Glaser, C.L. Lee, J.J. Nitao, T.L. Hickling, W.G. Hanley, Markov chain Monte Carlo-based method for flaw detection in beams, *ASCE Journal of Engineering Mechanics* 133 (12) (2007) 1258–1267.
[23] S.K. Au, J. Ching, J.L. Beck, Application of subset simulation methods to reliability benchmark problems, *Structural Safety* 29 (2007) 183–193.
[24] B.-T. Zhang, D.-Y. Cho, System identification using evolutionary Markov chain Monte Carlo, *Journal of Systems Architecture* 47 (2001) 587–599.
[25] G. Kerschen, J.-C. Golinval, F.M. Hemez, Bayesian model screening for the identification of nonlinear mechanical structures, *Journal of Vibration and Acoustics* 125 (2003) 389–397.
[26] K.D. Murphy, J.M. Nichols, A low-dimensional model for delamination in composite structures: theory and experiment, *International Journal of Nonlinear Mechanics* 44 (2008) 13–18.
[27] J.M. Nichols, K.D. Murphy, Modeling and detection of delamination in a composite beam: a polyspectral approach, *Mechanical Systems and Signal Processing* 24 (2) (2010) 365–378.