Mechanical Systems and Signal Processing 25 (2011) 2399–2412

Contents lists available at ScienceDirect

Mechanical Systems and Signal Processing

journal homepage: www.elsevier.com/locate/jnlabr/ymssp

Model selection in finite element model updating using the Bayesian evidence statistic

Linda Mthembu [a,*], Tshilidzi Marwala [b], Michael I. Friswell [c], Sondipon Adhikari [d]

[a] Department of Mechanical Engineering Technology, Faculty of Engineering and Built Environment, University of Johannesburg, PO Box 17011, Doornfontein 2028, South Africa
[b] Faculty of Engineering and Built Environment, University of Johannesburg, PO Box 17011, Doornfontein 2028, South Africa
[c] Aerospace Structures, College of Engineering, Swansea University, Singleton Park, Swansea SA2 8PP, United Kingdom
[d] Aerospace Engineering, College of Engineering, Swansea University, Singleton Park, Swansea SA2 8PP, United Kingdom

ARTICLE INFO

Article history:
Received 16 August 2009
Received in revised form
25 March 2011
Accepted 2 April 2011
Available online 4 May 2011

Keywords:
Finite element model updating
Nested sampling
Model selection
Bayesian evidence
Monte Carlo

ABSTRACT

This paper considers the problem of finite element model (FEM) updating in the context of model selection. The FEM updating problem arises from the need to update the initial FE model that does not match the measured real system outputs. This inverse system identification-problem is made even more complex by the uncertainties in modeling some of the structural parameters. Such uncertainty often results in a number of competing forms of FE models being proposed which leads to lack of consensus in the field. A model can be formulated in a number of ways; by the number, the location and the form of the updating parameters. We propose the use of a Bayesian evidence statistic to help decide on the best model from any given set of models. This statistic uses the recently developed stochastic nested sampling algorithm whose by-product is the posterior samples of the updated model parameters. Two examples of real structures are each modeled by a number of competing finite element models. The individual model evidences are compared using the Bayes factor, which is the ratio of evidences. Jeffrey's scale is then used to determine the significance of the model differences obtained through the Bayes factor.

1. Introduction

System identification [17] forms an important stage of many scientific modeling problems and is mainly concerned with the derivation of mathematical models of a system from its measured dynamics. The results from the system identification can then be used to understand and predict the system responses in future designs or different environments. Therefore the analyst is interested in the accuracy, confidence range, and more critically the correctness of the assumed mathematical model.

In this paper, the systems considered are structural and the model domain is that of finite element models (FEMs). In this context, these models are used to approximate the structural dynamics of systems such as train chassis, aircraft fuselages, bicycle frames, or civil structures. It is often the case that the finite element predictions do not match the measured structure's dynamic response [10,12]. This inconsistency may be due to the following: the form of the FE model, the identity and magnitude of the uncertain parameters, the noise and/or errors in the measurements. Furthermore the

* Corresponding author. Tel.: +27 11 559 6136; fax: +27 11 559 6137.
E-mail address: lmthembu@uj.ac.za (L. Mthembu).

0888-3270/$ - see front matter
doi:10.1016/j.ymssp.2011.04.001

measured data is incomplete due to the impracticality of capturing the full dynamics of the structure at every degree of freedom and over the complete frequency range, and this renders the problem ill-conditioned [9,10,12]. Consensus on this inverse problem is made more difficult by the fact that a multitude of mathematical models of the structure can be developed from engineering judgment. Moreover, these models can have varying levels of complexity, which leads to non-unique solutions for a particular modeled system. This situation makes the identification of the best FEM of the system of primary importance.

A number of pertinent questions arise in system identification: (1) Which aspects of the models do we need to update [14,24]? (2) How are we to update these models? [6,13,15,18,22,23] (3) Is the updated model the best one [16]? In this paper we propose approaching the finite element model updating problem by answering the third question. This can be recast as; what *evidence* do we have that our updated model is the correct one given that a number of plausible models of the real system can be generated from answering the first two questions? Question one is normally addressed from engineering judgment, often complemented by some form of parameter sensitivity measures/studies [9,10,12,24,33]. Question two is guided by the relevance of current available methods/approaches in the field. The model evidence approach proposed in this paper has received little attention in the FEM updating field and we believe it goes to the core of the model updating problem (FEMUP). By addressing the third question, we can get a better understanding of the given structural model design, the qualitative significance of the difference between proposed models and which model(s) is (are) worth considering for further analysis. This evidence measure, we argue, ought to be an essential and necessary measure to establish before any model updating problem is settled. We thus propose that the FEMUP should be approached from a model selection perspective. This is perhaps best approached using Bayesian inference, which allows one to deal with models in an intuitive and systematic way [3,4,17,19,20].

In the next section, we formally present the finite element model background. In Section 3, we introduce parameter estimation and model selection in the Bayesian framework. The Bayesian definition of model evidence is also presented. Section 4 introduces the evidence calculation algorithm. Section 5 presents two example implementations of the evidence concept: firstly on a simple unsymmetrical H-beam and secondly on a more complex Garteur SM-AG19 structure. At the end of each example we discuss the findings and Section 6 concludes the paper.

2. Finite element background

In finite element modeling, dynamic structures are analyzed by discretizing the structure into constituent elements. When assembled these elements constitute a system described by a second-order matrix differential equation of the form [9,10,12]

$$\mathbf{M}\ddot{\mathbf{x}}(t)+\mathbf{C}\dot{\mathbf{x}}(t)+\mathbf{K}\mathbf{x}(t)=\mathbf{f}(t) \tag{1}$$

where $\mathbf{M}$, $\mathbf{C}$, and $\mathbf{K}$ are of equal size and are called the mass, damping and stiffness matrices, or alternatively the system matrices, $\boldsymbol{f}(t)$ is the input force and $\boldsymbol{x}(t)$ is the response vector. If Eq. (1) is transformed to the modal domain and the structure considered is lightly damped or undamped, $\mathbf{C}=0$, the corresponding eigenvalue equation for the jth mode becomes

$$[-(\omega_j^m)^2\mathbf{M}+\mathbf{K}]\phi_j^m=\mathbf{0} \tag{2}$$

where ω_j^m and ϕ_j^m are the jth measured system natural frequency and its corresponding mode shapes, together known as modal properties. If the natural frequency and mode shape are replaced by analytical quantities then the eigenvalue equation is not exactly satisfied, but

$$[-(\omega_j^a)^2\mathbf{M}+\mathbf{K}]\phi_j^a=\varepsilon_j \tag{3}$$

where ε_j is the residual vector for the jth analytical frequency, ω_j^a, and the corresponding mode shape ϕ_j^a. In this setting, given a set of measured system modal data, D, the finite element model problem is to determine the initial model that realistically approximates the mass and stiffness matrices. Thus the predicted modal data will be as close to the measured modal data as possible (so-called data-match). If these predictions are not sufficiently accurate, then the residual vector will be non-zero and some of the model parameters will need to be modified, giving rise to the FEMUP.

In the next section, we introduce the proposed Bayesian inference approach in the context of finite element model updating.

3. Bayesian inference

Bayesian inference allows one to quantify uncertainties in quantities of interest in a formal way. Bayesian inference is often implemented in two settings; parameter estimation and model selection [4,19]. Parameter estimation is concerned with the plausibility of a given model's parameters based on some observed measurements and this is often carried out using standard Bayes theorem and/or sampling methods e.g. Markov Chain Monte Carlo (MCMC) [7,19,20,27]. In contrast, model selection deals with the mathematical hypothesis of the ability of the model(s) to approximate a particular observed/measured quantity.

3.1. Parameter estimation

In parameter estimation the mathematical model (in our case a particular FE model) is assumed to be true. The model is then allowed to approximate the measured data and the plausibility of the model parameters can be inferred from their posterior probability. This probability is calculated via Bayes theorem as follows:

$$P(\theta|D,H) = \frac{P(D|\theta,H)P(\theta|H)}{P(D|H)} \tag{4}$$

where the left-hand side is the posterior probability of the updating parameters for the true model H, given some data D, the prior probability of the model parameters is $P(\theta|H)$ and $P(D|\theta,H)$ is the likelihood of the model. The denominator $P(D|H)$ is called the marginal likelihood, or the *evidence*, of the model where the parameters have been marginalized out. Bayes theorem automatically incorporates the updating of the parameters by definition; it updates the prior probability distribution of the model parameter values with the likelihood (under the assumed parameters) of the model approximating the measured data. The likelihood function is the difference between the model (using the assumed updating parameters) and the real system data. The updated parameter values are then revealed in the posterior probability distribution. This probability quantifies the plausibility of the chosen parameters to approximate the observed data.

Bayes theorem assumes we have some a priori knowledge of the model parameter value distribution or the possible variation of the parameter values. In [21], the parameter estimation context of the Bayesian framework was implemented to obtain the posterior probability distribution of model parameters. By obtaining the posterior probability of the updating parameters the probability distribution of the modal properties may be calculated. This gives the researcher a quantitative measure of the confidence intervals of the modal data and updating parameters of the model.

3.2. Model comparison/selection

In model selection, a number of candidate finite element models for a system are formulated and compared to determine which model would best approximate the measured data. In general, Bayesian inference provides a platform to evaluate which model(s) is (are) the most probable for given measured dataset(s). In this paper, one measured dataset is observed and the finite element model that best approximates this data is determined. The posterior probability of each model within a set of plausible models is given by Bayes theorem as

$$P(H_i|D) = \frac{P(D|H_i)P(H_i)}{P(D)} \tag{5}$$

where $P(H_i)$ is the prior probability of each model and $P(D)$ is the probability of the data. Since the denominator is independent of the models, we may neglect its influence so that Bayes theorem reduces to

$$P(H_i|D) \propto P(D|H_i)P(H_i) \tag{6}$$

The first term on the right-hand side of Eq. (6) is the denominator term (or evidence) from Eq. (4). On the assumption that each designed model is equally likely to match the data, the evidence term is then the deciding factor on which model is the most probable for a particular observed dataset. Model selection for finite element models has been presented in [3,29]. In [29] the posterior distribution of the model was assumed to be a normal distribution, which is known to work for certain types of models [3,4,19]. The updating method proposed in this paper does not place any prior assumption on the form of the posterior distribution of the model but has recently [8] been shown to perform better for normally distributed functions. In [29], a recently proposed MCMC type posterior probability sampling algorithm (TMCMC from [7]) was implemented. This algorithm estimates the model evidence by sampling the posterior probability distribution of the model by a sequence of non-normalized intermediate probability functions. This algorithm uses a number of free tuning parameters; namely a variable to balance the sampling steps, the number of intermediate probability distribution functions (PDF), the tempering parameter and the control parameter [7,29]. The algorithm proposed in this work is believed to be simpler to implement.

In the next section we explain the concept of Bayesian evidence. We then introduce an algorithm called nested sampling to efficiently estimate the model evidence.

3.3. Bayesian evidence

Eq. (4) shows that the evidence term is a normalizing factor and can be written as

$$\text{evidence} = P(D|H_i) = \int P(D|\theta,H_i)P(\theta|H_i)\,d\theta \tag{7}$$

This equation may be interpreted as follows: given a unit of parameter space $d\theta$ with a model having a prior parameter probability of $P(\theta|H_i)$ over this space, the posterior probability distribution of the model over the same space will depend on how well the model with these parameters approximates the data $P(D|\theta,H_i)$. For the posterior distribution, the evidence is proportional to the two terms of the integrand in Eq. (7). The first term is the data- approximation (likelihood) component $P(D|\theta_{mp}, H_i)$ due to the most probable model parameter values, θ_{mp}.

3.3.1. The likelihood function

We assume the likelihood function to be a normal function, with the function exponential E given by the difference between the measured and analytical natural frequencies at the measured modes. Thus

$$P(D|\theta,H_i) = \frac{1}{(2\pi\sigma^2)^{(N_p N_m)/2}} \exp\left[-\frac{1}{2\sigma^2}\sum_i^{N_p}\sum_j^{N_m}(E^2)\right] \tag{8}$$

where E is the error matrix. Subscript j represents the jth modal property and i represents the ith measurement position on the structure. N_p represents the number of measured mode-shape coordinates and N_m is the number of measured modes. The uncertain error variance at each measured position on the structure is σ, and represents the uncertainty between the measured and analytical outputs.

3.3.2. Information gain

The second term in Eq. (7), $P(\theta|H_i)\,d\theta$, is the model complexity penalty term (or the information gain). This information gain relates the prior probability of the model updating parameter values to the posterior distribution of the updated parameter values. In the posterior distribution, the most probable updating parameter values occupy a smaller peaked region than in their initial prior distribution. This change factor is equal to the fraction of the posterior parameter space to the prior parameter space [4,19] and effectively measures how much information the model has extracted from the data. The information gain measure can be calculated as

$$I = -\int \ln\left[\frac{p(\theta_j|D,H_i)}{p(\theta_j|H_i)}\right] p(\theta_j|D,H_i)\,d\theta_j \tag{9}$$

Eq. (9) clearly shows that if the posterior parameter distribution is a substantially small fraction of an initially large prior distribution space, as is the case for complex models, then the natural logarithm of that ratio becomes large. This is the factor that penalizes the model complexity in the evidence calculation. The bigger the value of this factor the more information needs to be extracted from the measured data to update the suggested model parameters. The complexity of the model influences the value of the information gain and therefore affects the value of the Bayesian evidence.

3.3.3. The evidence reformulated

The evidence formulation of Eq. (7) can be re-written as

$$Z = \int L(\theta)\pi(\theta)\,d\theta \tag{10}$$

where $L(\theta)$ is the likelihood and $\pi(\theta)$ is the prior distribution. Analytically evaluating this integral may be difficult or impossible if the product of the prior and likelihood is not simple. This often happens when, for example, the parameter space has high dimension, which requires the calculation of multidimensional likelihoods. The most popular approach to approximate such integrals is to apply numerical techniques such as importance sampling and thermodynamic integration [19,27] although these encounter the problem that the prior parameters are distributed in regions where the likelihood function is not highly concentrated. Other sampling techniques, for example TMCMC [7,29], that are based on the Monte Carlo Markov Chain (MCMC) paradigm can be used, but these tend to only sample the peaks of the posterior distribution which can under-sample most of the narrow best-approximation regions [19,27]. Recently Skilling [30–32] proposed the nested sampling algorithm, which is able to efficiently estimate integrals of the form shown in Eq. (10). The algorithm works by transforming the multidimensional parameter space integral into a one dimensional integral where classical numerical approximation techniques to estimate the area under a function can be applied. The algorithm has been successfully applied and further modified in a number of recent astronomy and cosmology papers [11,25,26,34].

4. Sampling

4.1. Nested sampling

Nested sampling (NS) is a Monte Carlo, but not a Markov Chain, sampling method. The main idea behind the method is to divide the prior parameter space into 'equal mass' units and to order these by model likelihood. The total prior mass is denoted by X and each unit in this prior mass is $P(\theta|H)\ d\theta = \pi(\theta)\,d\theta = dX$.

The likelihood function is written as

$$P(D|\theta,H) = L(\theta) = L(X) \tag{11}$$

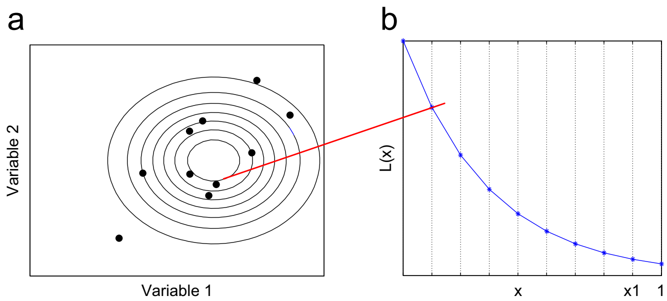

Fig. 1. (a) Samples ($N=10$) in 2D iso-contour parameter space and (b) the ten samples ordered by likelihood in 1D during nested sampling. The line shows the second highest likelihood sample in both spaces.

in this space. This formulation transforms the likelihood into a function of a single parameter. The evidence integral from Eq. (9) then becomes

$$\text{evidence} = Z = \int L(\theta)\pi(\theta)\,d\theta = \int_0^1 L\,dX \quad (12)$$

The algorithm supposes that the likelihood can be evaluated at all X_i such that $L_i = L(X_i)$, where X_i is a sequence of values that decreases from 1 to 0 such that $0 < X_N < \cdots < X_2 < X_1 < X_0 = 1$ as illustrated on the right image of Fig. 1. The left image shows a number of samples and the likelihood iso-contours in the 2-dimensional posterior probability parameter space.

The one dimensional integral function in Eq. (12) is easily estimated by any numerical method. For example using the trapezoid rule gives

$$Z = \sum_{i=1}^{N} Z_i \quad \text{where } Z_i = L_i b_i, \quad b_i = \frac{1}{2}(X_i - X_{i+1}) \quad (13)$$

where L_i is the likelihood at that sample. In the context of finite element model updating, the algorithm achieves this approximation in the following manner:

1. Sample N updating parameter values (e.g. Young's modulus values) from some prior probability distribution (based on the assumed distribution of Young's modulus values in the structure). Evaluate their likelihoods.
2. From the N samples select the sample that results in the lowest likelihood, say $L_{i=1}$.
3. Increment the evidence by some summation rule, e.g. $Z_i = (L_i/2)(X_i - X_{i+1})$ for the trapezoid rule.
4. Discard the sample with the lowest likelihood (L_1) in the original N samples. Replace it with a new sample drawn uniformly from the prior parameter distribution within the remaining prior volume [0, X_1]. This new sample must satisfy the constraint that its likelihood is larger than the discarded sample's likelihood ($L_{new} > L_1$).
5. Repeat steps 2–4 until some stopping criterion is reached. This could be the desired precision on the evidence or some iteration count.

For further details on nested sampling see [30–32]. The next section presents the experiments performed on two structures and introduces the candidate finite element models, together with the evaluation of their evidences.

4.2. Model definition and comparison

In FEM a mathematical model (H_i in Section 3) is defined by the identity and location of the updating parameters. In the first example models were designed to have a different *number* of free (updating) parameters at different *positions* along the same structure. The number of free-parameters in a model is one measure of the complexity of that mathematical model, although this has recently been questioned in the context of the evidence of a complex model [28]. Classically, complex models tend to over-approximate the data but in the Bayesian context this is not the necessarily the case. In this paper we also wish to determine the relationship between the complexity of the finite element model and how well it matches the measured data, and, according to Occam's razor, the model with fewer parameters is preferred. In the Bayesian formulation of the model updating problem this condition is implicit. Jeffrey's scale [27,34] is often used to determine the significance of the difference between model evidences. This measure is calculated from the ratio of model evidences, also known as Bayes factors. Table 1 shows the interpretation of this measure used in this paper.

Table 1
The significance ratios used in Jeffrey's scale.

Z_j/Z_i	$\log_2(Z_j/Z_i)$	$\log_e(Z_j/Z_i)$	$\log_{10}(Z_j/Z_i)$	Evidence against model H_i
1–3.2	0–1.7	0–1.2	0–0.5	Weak
3.2–10	1.7–3.3	1.2–2.3	0.5–1	Substantial
10–100	3.3–6.6	2.3–4.6	1–2	Strong
>100	>6.6	>4.6	>2	Decisive
>1000	>10	>7	>3	Beyond reasonable doubt

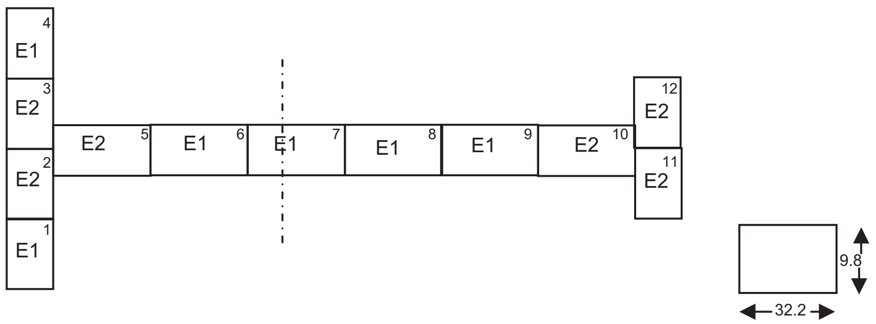

Fig. 2. Model H_2 of a 12 element unsymmetrical H-beam.

5. Applications of Bayesian evidence

To investigate the applicability of the Bayesian evidence statistic to quantify different models we provide two examples. The first example considers a simple unsymmetrical H-beam structure previously modeled in [23] and the second example considers the more complex Garteur SM-AG19 structure that has previously been modeled in [5,13,16]. Both structures presented are modeled using Version 6.3 of the Structural Dynamics Toolbox (SDT®) for MATLAB. In both examples we use measured data from real physical experiments.

5.1. *Example 1: unsymmetrical H-beam*

A simple unsymmetrical aluminum H-beam structure, shown in Fig. 2, is modeled. The structure is divided into 12 Euler–Bernoulli beam elements. All beam sections used standard isotropic material properties. The beam is free to move in all six degrees of freedom. The measured natural frequencies of interest of this structure are 53.9, 117.3, 208.4, 254, and 445 Hz, which correspond to modes 7, 8, 10, 11, and 13, respectively. The unsymmetrical H-beam had three thin cuts of 1 mm, introduced into elements 3, 4, and 5 that went half-way through the cross-section of the beam. The structure with these cuts was used so that the initial FE model gives data that are far from the measured data and thereby test the proposed procedure on a difficult FE updating process. The structure was suspended using elastic rubber bands (see [23] for more details on the structure and experimental set-up) and has the geometric and standard material properties shown in Table 2.

5.1.1. *Finite element model*

To validate that the model evidence calculation can reveal the most plausible finite element model, eight illustrative models of the beam are developed and the evidence of each is calculated. The objective is then to determine from the evidence, the most probable model from this set.

5.1.2. *Candidate mathematical models*

All models in this example assume the only uncertain property of the beam is its Young's modulus (E) value. In the model updating context Young's modulus of an element is an equivalent parameter that represents the local stiffness of the beam sections. For example, slots cut in the beams will reduce the stiffness locally and this is modeled by reducing Young's modulus of the element containing the slot. Beam elements are grouped differently to form a particular number of parameters for each model. The beam is modeled by eight competing models, H_i, $i=1, \ldots, 8$. Model H_1 assumes the whole beam's Young's modulus is the variable to be updated from the initial standard material value (see Table 2). This means model H_1 has one parameter, E. Model H_2 has two parameters, E_1 and E_2; the elements numbered 1, 4, 6–9 are grouped together to form parameter E_1, while elements numbered 2, 3, 5, 10–12 form parameter E_2 (see Fig. 2 for the element

Table 2
The standard material and measured geometric properties of the unsymmetrical H-beam.

Material (aluminum)	Young's modulus 7.2×10^{10} N m^{-2}
Length	600 mm
Width	32.2 mm
Section thickness	9.8 mm
Left edge length	400 mm
Right edge length	200 mm
Density	2790kg/m^3

Table 3
Model parameters and labels of the eight H-beam FE models.

Model	No. parameters	Parameter labels	Element grouping
H_1	1	E	{1–12}
H_2	2	E_1 and E_2	{1, 4, 6–9} and {2, 3, 5, 10–12}
H_3	3	E_1, E_2, and E_3	{1, 4, 6–9}, {2, 3, 11, 12} and {5, 10}
H_4	4	E_1, E_2, E_3, and E_4	{1, 4, 6–9}, {2, 3} {11, 12} and {5, 10}
H_5	5	E_1, E_2, E_3, E_4 and E_5	{1, 4, 6–9}, {2, 3}, {11, 12}, {5} and {10}
H_6	2	E_1 and E_2	{1, 2, 3, 4} and {5–12}
H_7	2	E_1 and E_2	{1–6} and {7–12}
H_8	3	E_1, E_2 and E_3	{1–5}, {6–9} and (10, 11, 12}

numbering and grouping for model H_2). Basically H_2 models the elements connected near the joints as one parameter and those away from the joints as another.

The design of model H_6 assumes the dynamics of the whole left edge of the beam is different from the rest of the structure and this is modeled by grouping the first four elements together (to form E_1) and grouping the last eight in another parameter (E_2). Model H_8 assumes the left edge together with the first horizontal element, the horizontal section and right edge together with the last horizontal element are different and thus the model has three parameters. Table 3 shows the rest of the models and their parameterizations.

5.1.3. Prior model and nested sampling parameters

Each model was run through the stochastic nested sampling algorithm three times. The number of Young's modulus updating parameters sampled for each model was set to $N=250$ and the nested sampling algorithm's stopping criterion was set to a maximum of 300 iterations. The exploration for a new sample, point number 4 in Section 4.1, in the nested sampling algorithm had 20 MCMC steps. The updating variables were assumed to be identical but independent and were sampled from a normal prior distribution $P(\theta|H_i)$ with a mean of 7.2×10^{10} N m^{-2} and a variance of approximately 0.55×10^{20}. The inverse standard deviation of the error probability or likelihood function, $P(D|\theta,H_i)$, is the error between successive measured and initial model modal properties. It is set to uniformly vary between approximately 1 and 0.10.

5.1.4. Results and discussions

Fig. 3 (a, b) plots the log likelihood values for the 250 samples for model H_1 before and after the implementation of nested sampling. Similar results are observed in Fig. 3 (c, d), where 250 log likelihood values for samples from model H_5 before and after nested sampling algorithm are plotted. In each model the lowest log-likelihood after nested sampling is much higher than before the algorithm was applied. Fig. 3 shows that, for both models, the nested sampling algorithm was able to sample approximately 70% more updating parameter values that minimize the difference between the measured data and the initial finite element model using the hard log-likelihood constraint as set out in the nested sampling algorithm point number 4.

Tables 4 and 5 present the results of the stochastic simulations of the eight finite element beam models. These results are the posterior most probable parameter values based on the normal likelihood probability simulation from the three runs of the nested algorithm of each model in the set.

From the results in Table 4, model H_3 has, on average, the widest posterior distribution of all of Young's modulus values and model H_1 has the smallest posterior distribution. Model H_5's Young's Modulus value of 7.16×10^{10} N m^{-2} is the closest to the standard material modulus while model H_1's modulus is the most varied at 6.82×10^{10} N m^{-2}. In all of the models, the E_1 updating parameter has the lowest Young's modulus value and with quite small variances from the mean.

Not only is this the parameter that changes most, but also the parameter updating algorithm is confident in these E_1 posterior parameter updating values. The significance of this parameter's value is further supported by the relatively high averages for the E_2 updating parameters (which occupy the same location as E_1) in models H_6, H_7, and H_8. This suggests that in order for the finite element model to better approximate the real unsymmetrical H-beam the left-hand side of the

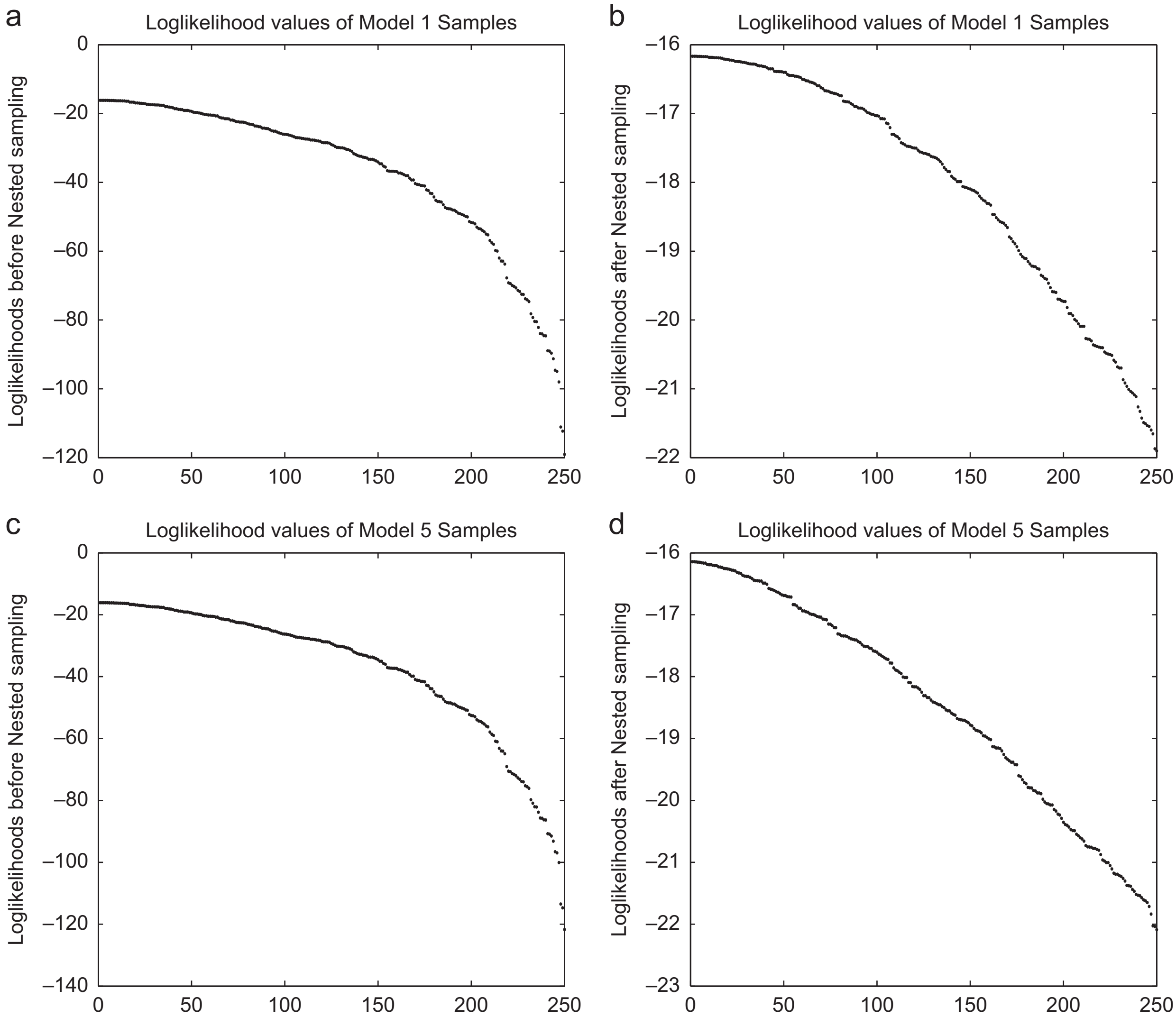

Fig. 3. The plot of log likelihood values for 250 samples from models H_1 and H_5. Subplots (a, c) are for *before* the nested sampling implementation and subplots (b, d) *after* nested sampling implementation.

beam model should be less stiff than the standard aluminum beam. Table 5 shows the Bayesian log-evidence, data-match and the information gain results of the nested sampling algorithm.

Table 5 shows that model H_2 has the highest log-evidence and model H_3 has the lowest log-evidence for the set of models. Model H_3 extracted the largest amount of information from the data to update its three parameters and was accordingly penalized for that. Surprisingly model H_5, which is the most complex model, extracted less information to update its five parameters and has a higher evidence measure than less complex models H_1, H_3, H_4, H_7, and H_8. Similarly models H_3, H_7, and H_8 extracted more information than the more complex model H_4. This demonstrates that the Bayesian evidence measure not only penalizes the complexity of a model but also the amount of information the model extracts from the data to update its extra parameters. Models H_6 and H_7, which are similar, have a poorer data-match and they extract more information from the data than their complex equivalent model H_2, which has a higher evidence measure. In the similarly arranged models H_3, H_4, and H_5, the simplest model extracts the most information to update its parameters. Systematically splitting E_2 in model H_3 to become E_2 and E_3 in model H_4 and E_4 in model H_4 to become E_4 and E_5 in model H_5 has resulted in less information being extracted from the data.

The results demonstrate a common phenomenon that occurs when an extra variable is introduced whose effect on the model is similar to an existing parameter. This can often lead to over-fitting in classic regression but in Bayesian analysis this effect is counter-acted by a penalty that is imposed on any extra parameter through the information the model requires to update it. In order to evaluate/interpret the significance of the different Bayesian evidence measures we use Jeffrey's scale, with the results shown in Table 6. Note that the evidence for model H_2 against model H_3, which only has one

Table 4
The posterior mean and standard deviations of Young's modulus updating parameter values for each model over three runs of nested sampling.

	MODEL 1		**MODEL 2**		**MODEL 3**		**MODEL 4**		**MODEL 5**		**MODEL 6**		**MODEL 7**		**MODEL 8**	
	MEAN	STD	MEAN	STD	MEAN	STD	MEAN	STD	MEAN	STD	MEAN	STD	MEAN	STD	MEAN	STD
E1	6.82*E*+10	2.36*E*+09	6.80*E*+10	2.69*E*+09	6.64*E*+10	4.95*E*+09	6.80*E*+10	3.14*E*+09	6.72*E*+10	3.96*E*+09	6.85*E*+10	1.42*E*+09	6.83*E*+10	2.09*E*+09	6.82*E*+10	2.50*E*+09
E2	–	–	7.09*E*+10	6.03*E*+09	7.45*E*+10	6.98*E*+09	7.31*E*+10	6.81*E*+09	7.25*E*+10	5.90*E*+09	7.35*E*+10	7.09*E*+09	7.30*E*+10	5.89*E*+09	7.09*E*+10	7.55*E*+09
E3	–	–	–	–	7.04*E*+10	5.26*E*+09	7.08*E*+10	4.90*E*+09	7.38*E*+10	6.61*E*+09	–	–	–	–	7.23*E*+10	7.23*E*+09
E4	–	–	–	–	–	–	7.20*E*+10	7.03*E*+09	7.24*E*+10	5.98*E*+09	–	–	–	–	–	–
E5	–	–	–	–	–	–	–	–	7.25*E*+10	5.57*E*+09	–	–	–	–	–	–
$1/\sigma$ (rad)	0.13		0.11		0.18		0.14		0.14		0.13		0.15		0.14	

Table 5
FEM log evidences, data-match and information gain results of the NS evidence statistic for the unsymmetrical H-beam models.

Model	Log-evidence	Data-match	Information gain
H_1	−25.430	−22.891	2.540
H_2	−23.394	−21.228	2.166
H_3	−31.224	−28.117	3.107
H_4	−25.935	−23.355	2.580
H_5	−25.377	−22.928	2.449
H_6	−25.219	−22.665	2.554
H_7	−27.361	−24.625	2.737
H_8	−25.951	−23.196	2.755

Table 6
Jeffrey's scale results for a selection of the models. This table shows the significance of the difference between Bayesian evidence measures.

Model	Z_j/Z_i	$\log_e(Z_j/Z_i)$	Evidence against model H_i
H_2/H_3	2515	7.83	Beyond reasonable doubt
H_8/H_3	195	5.27	Decisive
H_2/H_1	7.66	2.03	Substantial
H_2/H_6	6.21	1.83	Substantial
H_2/H_7	52.58	3.96	Strong
H_6/H_4	2.05	0.72	Weak

Table 7
FE models and updating parameters for the Garteur SM-AG19 structure.

Model name and type	**Fuselage**	**Wing**	***R*-wing**	***L*-wing**	**Wing/fuselage**
P_2 (plate)	E_1G_1	E_6G_6			E_7 G_7 (steel)
B_1 (beam)			$I_{6MIN}I_{6MAX}J_6$	$I_{5MIN}I_{5MAX}$ J_5	
B_2 (beam)	E_1I_{1min}	E_5 $I_{5MIN}E_5I_{5MAX}G_5J_5\rho_5$			

Model name and type	**Vertical tail**	**Wing thickness**	**Right drum**	**Left drum**	**Overall density**	**Residual type**
P_2 (plate)	E_2G_2	W_T	Mass (M_L)	Mass (M_R)	ρ	Freq
B_1 (beam)	I_{2min}				ρ	Freq
B_2 (beam)	E_2 I_{2min}					Freq

$EI_{min,max}$: minimum and maximum bending stiffness; GJ: torsional stiffness; E: Young's modulus; G: shear modulus; ρ: mass density; M_R/M_L: right and left mass; W_T: wing thickness; Freq: natural frequency.

extra parameter, is 'beyond reasonable doubt'. On the other hand model H_3's equivalent, model H_8, is decisively better than model H_3. Tables 4 and 5 show that model H_2, as measured by the Bayesian evidence statistic, is substantially better than its closest competition model H_6, which interestingly has the same number of parameters.

Thus the Bayesian evidence statistic has highlighted that an increased number of parameters will not necessarily result in a significantly better model. Indeed the top three models all have a small number of parameters. The information extracted from the measured data by the updating parameters is a critical factor of the Bayesian evidence measure. In the next section, we present the second application example of Bayesian model updating using the nested sampling algorithm.

5.2. Example 2: Garteur SM-AG19 testbed

In 1995–1996, this structure was used as a benchmark study by 12 members of the Garteur Structures and Materials Action Group 19. One of the aims of the study was to compare the different computational model updating procedures on a single common test structure [1,2,5,13,16]. A number of papers have presented the results of this benchmark study and no conclusive best computational model updating procedure was found. In this paper we compare three of the seven models, as presented in [16], using our proposed Bayesian evidence statistic. Table 7 shows the three chosen models and the types of updating parameters they used.

5.2.1. Measured data

The benchmark study also allowed participants to test a single representative structure using their own test equipment. In this paper we only consider two test results; those from DLR in Gottingen, Germany and those by the Imperial College of

Science, Technology and Medicine (IC) in the UK. The natural frequencies measured by these two institutes are shown in Table 8.

5.2.2. Candidate mathematical models

The three models (shown in Table 7) used for comparison consist of one plate element model and two beam element models for the same Garteur SM-AG19 test structure. The models are named according to their finite element types, e.g. model B_2 is a Euler–Bernoulli beam element model, and the model numbering is arbitrary. The lowest number of updating parameters in the three model set is from model B_2 with 6 parameters and the maximum is model P_2 and B_1 with 8 parameters each. As can be seen from Table 7 the updating parameters are quite varied across the three models. The updating parameter numbering is according to *what property* and *which part* of the test structure is updated e.g. E_1I_{1min}, means Young's modulus and minimum second moment of area of the fuselage are updated.

5.2.3. Standard properties

In order to quantify the model updating capabilities of the proposed nested sampling algorithm, we ran the three models with standard material property and measured geometric values for the uncertain parameters. Table 9 shows the FEM natural frequency, percentage frequency difference, and average error results for each model compared to the two measured data sets.

In both sets of measured data and model comparisons, model B_2 has marginally lower average errors to B_1 and model P_2 has the highest average error. Clearly before updating the plate model P_1 is a poor model of the given Garteur structure using this measured data. The main difference between B_1 and B_2 is that model B_1 has a split main-wing which uses separate updating parameters for the left ($I_{5MIN}I_{5MAX}J_5$) and right wing ($I_{6MIN}I_{6MAX}J_6$) while B_2 updates the whole wing under one parameter ($I_{5MIN}I_{5MAX}J_5$). Model B_1 only has one material property (ρ-overall density) to update and the rest of the parameters are geometric in nature.

Table 8
Measured natural frequency data from two institutes, DLR and IC.

Mode no.	**DLR**	**IC**	Difference (%) (DLR/IC)
	Natural frequency (Hz)		
1	6.38	6.54	−2.50
2	16.10	16.55	−2.79
3	33.13	34.86	−5.22
4	33.53	35.30	−5.27
5	35.65	36.53	−2.46
6	48.38	49.81	−2.95
7	49.43	50.63	−2.42
8	55.08	56.39	−2.37
9	63.04	64.96	−3.04

Table 9
Natural frequency, percentage differences and average error for three models using standard material properties and geometric measurements.

P_2			**B_1**			**B_2**		
STD_{wn}	DLR Δf (%)	IC Δf (%)	**STD_{wn}**	DLR Δf (%)	IC Δf (%)	**STD_{wn}**	DLR Δf (%)	IC Δf (%)
6.55	−2.61	−0.10	5.77	9.64	11.85	5.77	9.63	11.85
18.59	−15.48	−12.34	15.44	4.10	6.70	15.44	4.09	6.70
39.97	−20.65	−14.66	30.98	6.49	11.13	31.11	6.11	10.77
40.54	−20.91	−14.85	31.49	6.08	10.79	31.66	5.59	10.32
41.16	−15.45	−12.67	33.78	5.25	7.53	33.82	5.13	7.42
52.33	−8.16	−5.05	45.15	6.67	9.35	45.16	6.66	9.34
53.12	−7.47	−4.93	54.82	−10.90	−8.27	54.82	−10.90	−8.28
57.10	−3.66	−1.25	56.14	−1.93	0.44	56.14	−1.93	0.44
69.94	−10.95	−7.67	60.02	4.79	7.60	60.02	4.79	7.60
Avg. error	4.29	3.08		2.23	2.93		2.19	2.89

STD_{wn} is the natural frequency of each model using standard material properties and geometric measurements. Δf (%) is the percentage difference between the STD_{wn} and measured natural frequency data from each institute.
Note: The avg. error is the sum average of the absolute error between the above two natural frequencies.

5.2.4. Prior model and nested sampling parameters

All model updating parameters shown in Table 7 except for the masses (M_R, M_L) and wing thickness (W_T) are sampled from a normal prior distribution. The masses and thickness are sampled from a uniform distribution between 0.15 and 0.2 kg and 9.5 and 11 mm, respectively. Each model's nested sampling algorithm was run at the following settings: number of samples, $N=250$; maximum number of iterations is 1500, and the number of MCMC steps is 15. The mean and variance values for all of the updating parameters are shown in Table 10. The inverse variance of the error probability function $P(D|\theta,M_i)$ is set to vary uniformly between 0.5 and 0.25. These values determine the assumed average variances of the difference between the measured and model natural frequencies for all models.

In the next section, we present the results of the application of the nested sampling algorithm to the Garteur SM-AG19 testbed using the three models.

5.2.5. Results and discussions

The nested sampling (NS) updating algorithm improved all of the models of the given structure. There is a maximum of a 70% and a minimum of an 11% improvement in the average natural frequency errors after NS. The maximum parameter value variations for all models are less than 8% of the standard values. The updated finite element model results are shown in Table 11. After nested sampling model P_1 achieves the lowest average error. This is not surprising as this model has the most updating parameters but practically this model will be computationally expensive to run.

A better method to evaluate the updated models is to look at the model evidence statistic. Table 12 shows the model evidences. According to these evidences, and Jeffrey's scale in Table 1, none of the models has better than weak Bayesian

Table 10
The prior mean and variance values for the updating parameters.

Parameter	Mean	Variance	Parameter	Mean	Variance
E_{1-6}	$7.2e-10$ Pa	$1e-19$	I_{2min}	$8.33e-9$ m^4	$2e-20$
E_7	$21e-10$ Pa	$3e-20$	I_{5min}	$8.33e-9$ m^4	$2e-10$
G_{1-6}	$2.8e-10$ Pa	$1e-19$	I_{5max}	$8.33e-7$ m^4	$2e-15$
G_7	8.17E10 Pa	$1e-19$	I_{6min}	$8.33e-9$ m^4	$2e-20$
ρ	2700 kg/m^3	$2e-4$	I_{6max}	$8.33e-7$ m^4	$2e-15$
ρ_5	2700 kg/m^3	$2e-4$	J_5	$3.12e-8$ m^4	$2e-19$
M_R and M_L	0.15–0.20 kg	Uniform distribution	J_6	$3.12e-8$ m^4	$2e-19$
I_{1min}	$1.56e-6$ m^4	$2e-14$	W_T	9.95–11 mm	Uniform distribution

Table 11
Natural frequency, percentage differences and average error for the three models using nested sampling results for the updating parameters.

DLR measured data	Model performance					
ω_n	P_2 ω_n	DLR Δf (%)	B_1 ω_n	DLR Δf (%)	B_2 ω_n	DLR Δf (%)
6.38	5.75	9.89	5.88	7.84	5.94	6.92
16.10	16.67	−3.56	15.73	2.29	15.84	1.60
33.13	34.50	−4.13	31.50	4.91	32.05	3.25
33.53	34.57	−3.11	32.03	4.48	32.71	2.46
35.65	39.00	−9.40	34.37	3.61	34.69	2.70
48.38	47.42	1.98	46.05	4.81	46.61	3.66
49.43	51.14	−3.46	55.03	−11.32	54.28	−9.81
55.08	54.60	0.88	56.15	−1.95	55.41	−0.60
63.04	64.54	−2.39	60.73	3.66	60.83	3.50
Avg. error		1.29		1.84		1.41
IC measured data	Model performance					
ω_n	P_2 ω_n	IC Δf (%)	B_1 ω_n	IC Δf (%)	B_2 ω_n	IC Δf (%)
6.54	5.86	10.39	5.99	8.48	6.00	8.24
16.55	16.96	−2.47	16.03	3.14	15.96	3.57
34.86	35.23	−1.06	32.10	7.91	32.77	5.99
35.3	35.31	−0.03	32.67	7.45	33.69	4.55
36.53	39.47	−8.05	35.06	4.04	35.23	3.56
49.81	48.17	3.29	46.95	5.74	47.26	5.11
50.63	51.76	−2.23	55.98	−10.57	55.08	−8.78
56.39	55.35	1.85	57.17	−1.38	56.07	0.58
64.96	65.48	−0.81	61.77	4.92	60.98	6.13
Avg. error		0.972		2.24		1.94

Table 12
FEM log evidences, data-match and information gain results of the NS evidence statistic for the Garteur SM-AG19 models.

Model	Log-evidence	Data-match	Information gain
P_2 (DLR)	−3.60	−3.07	0.53
B_1 (DLR)	−1.68	−0.92	0.76
B_2 (DLR)	−1.29	0.10	1.39
P_2 (IC)	−3.17	−2.90	0.27
B_1 (IC)	−3.92	−2.68	1.24
B_2 (IC)	−3.73	−0.83	1.90

evidence over the others for both sets of measured data. Model B_2 has the highest evidence and the highest information gain for the DLR data. This means the model used the prior parameter space efficiently to find the best posterior parameter values. This model's mean posterior values produced the highest model approximation to the measured DLR data; this is shown by the 0.10 data-match value. With regards to the IC data, model P_2 has the highest evidence but the lowest information gain. Again this model did not use the prior space efficiently; this is further confirmed by the data-match value being the worst in the model sets.

It is difficult to assert the best model for the Garteur structure. The Bayesian evidence statistic reveals no clear winner. Model B_2 is attractive in its use of parameter space but model B_1 is attractive in its use of fewer updating parameters for comparable performance. The plate model is not optimal for any of the measured data, especially with the 8 updating parameters, but there is no clear evidence against it. It should also be noted that the three models chosen, and the parameters selected, were the best models in the opinion of the three groups involved in the original benchmark exercise.

6. Conclusion

In conclusion, this paper has argued that model selection should be integrated into the finite element model updating problem. Specifically, a Bayesian model updating perspective to the problem of finite element model updating is introduced. This paradigm is shown to be well established in other areas of science with little work available in the finite element updating literature. Bayesian inference is able to simultaneously update the models whilst also incorporating a statistic to evaluate the evidence for each model. The theory of Bayesian inference is presented together with a method to stochastically approximate the Bayesian evidence statistic. This method, nested sampling, was recently introduced as an efficient way of estimating integrals of the form needed to determine the model evidence.

We demonstrated the applicability of Bayesian model updating and the estimation of the evidence through two structural examples. In each example a number of competing finite element models are proposed to approximate the structure. Each model is unique and has a specific number of parameters and parameter identities. The method is initially demonstrated with a simple unsymmetrical structure, followed by a complex structure modeled by three completely different finite element type models. Bayesian inference is then applied to update each model. The evidence of each model is estimated from the updating parameter priors and the likelihood of the parameters after the data has been taken into consideration. The evidence calculation revealed that model complexity is not necessarily proportional to the model evidence. Two important factors affect the model evidence; these are how the model approximates the data and the amount of information each model extracts from the data to update its parameters. This is largely influenced by the choice of model parameters.

The Bayesian model evidence calculation allows the engineer to determine if there is any advantage in updating certain parameters in his/her model and what evidence one has that the chosen model is better than other models. To put the model evidences in context we further showed the significance of the model Bayesian evidence differences by using Jeffrey's scale. This statistic allows the engineer to determine if the model evidence differences are significant enough to warrant discarding some models.

Acknowledgments

The financial assistance of the National Research Foundation of South Africa is greatly appreciated.

References

[1] E. Balmes, J. Wright, GARTEUR group on ground vibration testing, Results from the test of a single structure by 12 laboratories in Europe, in: International Modal Analysis Conference (IMAC), vol. 3089 (2), 1997, pp. 1246–1352.
[2] E. Balmes, Predicted variability and differences between tests of a single structure, in: International Modal Analysis Conference (IMAC), vol. 3242 (1), 1998, pp. 558–564.
[3] J.L Beck, K.-V. Yuen, Model selection using response measurements: Bayesian probabilistic approach, Journal of Engineering Mechanics 130 (2004) 192–203.
[4] C.M. Bishop, Pattern Recognition and Machine Learning, Springer, 2006.

[5] K. Bohle, C.-P. Fritzen, Results obtained by minimizing natural frequency and Mac-value errors of a plate model, Mechanical Systems and Signal Processing 17 (1) (2003) 55–64.
[6] J. Carvalho, B.N. Datta, A. Gupta, M. Lagadapati, A direct method for model updating with incomplete measured data and without spurious modes, Mechanical Systems and Signal Processing 21 (2007) 2715–2731.
[7] J.Y. Ching, Y.C. Chen, Transitional Markov chain Monte Carlo method for Bayesian model updating, model class selection, and model averaging, Journal of Engineering Mechanics—ASCE 133 (2007) 816–832.
[8] N. Chopin, C.P. Robert, Properties of nested sampling, Biometrika 97 (3) (2010) 741–755.
[9] B.N. Datta, Finite element model updating, eigenstructure assignment and eigenvalue embedding techniques for vibrating systems, Mechanical Systems and Signal Processing 16 (2002) 83–96.
[10] D.J. Ewins, Modal Testing: Theory and Practice, Letchworth, Research Studies Press, 1984.
[11] F. Feroz, M.P. Hobson, Multimodal nested sampling: an efficient and robust alternative to MCMC methods for astronomical data analysis, Monthly Notices of the Royal Astronomical Society 384 (2007) 449–463.
[12] M.I. Friswell, J.E. Mottershead, Finite Element Model Updating in Structural Dynamics, Kluwer Academic Publishers, Boston, 1995.
[13] D. Goge, M Link, Results obtained by minimizing natural frequency and mode shape errors of a beam model, Mechanical Systems and Signal Processing 17 (1) (2003) 21–27.
[14] B. Horton, H. Gurgenci, M. Veidt, M.I. Friswell, Finite element model updating of the welded joints in a hollow section H-frame, in: International Conference on Applications of Modal Analysis: Recent Advances in Modal Analysis Practice, 1999, pp. 309–335.
[15] G.-H. Kim, Y.-S. Park, An automated parameter selection procedure for finite element model updating and its applications, Journal of Sound and Vibration 309 (2008) 778–793.
[16] M. Link, M. Friswell, Working group 1: generation of validated structural dynamics—results of a benchmark study utilizing the Garteur SM-AG19 test-bed, Mechanical Systems and Signal Processing 17 (1) (2003) 9–20.
[17] L. Ljung, System Identification: Theory for the User, Prentice-Hall, Englewood Cliffs, NJ, 1987.
[18] M.I. Levin, N.A.J. Lieven, Dynamic finite element model updating using simulated annealing and genetic algorithms, Mechanical Systems and Signal Processing 12 (1998) 91–120.
[19] D.J.C. MacKay, Information Theory, Inference, and Learning Algorithms, Cambridge University Press, 2003.
[20] C. Mares, B. Dratz, J.E. Mottershead, M.I. Friswell, Model updating using Bayesian Estimation, ISMA 2006, Leuven, Belgium, September 2006, pp. 2607–2616.
[21] T. Marwala, S. Sibisi, Finite element model updating using Bayesian framework and modal properties, Journal of Aircraft 42 (2005) 275–278.
[22] T. Marwala, Finite element model updating using wavelet data and genetic algorithm, Journal of Aircraft 39 (2002) 709–710.
[23] T. Marwala, Finite element model updating using response surface method, in: Proceedings of the 45th AIAA/ASME/ASCE/AHS/ASC Structures, Structural Dynamics and Materials Conference, Palm Springs, California, USA, April 2004, AIAA Paper, 2004–2005, pp. 5165–5173.
[24] J.E. Mottershead, C. Mares, M.I. Friswell, S. James, Selection and updating of parameters for an aluminium space-frame model, Mechanical Systems and Signal Processing 14 (2000) 923–944.
[25] P. Mukherjee, D. Parkinson, A.R. Liddle, A nested sampling algorithm for cosmological model selection, Astrophysical Journal 638 (2006) L51–L54.
[26] D. Parkinson, P. Mukherjee, P Liddle, A Bayesian model selection analysis of WMAP3, Physical Review D 73 (2006) 123523.
[27] I. Murray, Advances in Markov chain Monte Carlo methods, Ph.D. Thesis, Gatsby Computational Neuroscience Unit (GCNU), University College London, 2007.
[28] I. Murray, Z. Ghahramani, A note on the evidence and Bayesian Occam's razor, Technical Report Gatsby Computational Neuroscience Unit GCNU-TR 2005–003 (2005).
[29] M. Muto, L.J. Beck, Bayesian updating and model class selection for hysteretic structural models using stochastic simulation, Journal of Vibration and Control 14 (2008) 7–34.
[30] J. Skilling, Nested sampling, in: R. Fischer, R. Preuss, U. von Toussaint (Eds.), Bayesian Inference and Maximum Entropy Methods in Science and Engineering, AIP Conference Proceedings, vol. 735, 2004, pp. 395–405.
[31] J. Skilling, Nested sampling for general Bayesian computation, Bayesian Analysis 1 (2005) 833–860.
[32] J. Skilling, Nested sampling for Bayesian computations, in: J.M. Bernardo, M.J. Bayarri, J.O. Berger, A.P. Dawid, D. Heckerman, A.F.M. Smith, M. West (Eds.), Bayesian Statistics, vol. 8, Proceedings of the Valencia/ISBA Eighth World Meeting on Bayesian Statistics, Oxford University Press, 2007.
[33] B. Titurus, M.I. Friswell, Regularization in model updating, International Journal for Numerical Methods in Engineering 75 (2008) 440–478.
[34] R. Trotta, Applications of Bayesian model selection to cosmological parameters, Monthly Notices of the Royal Astronomical Society 378 (2007) 72–82.