Available online at www.sciencedirect.com
ScienceDirect
Comput. Methods Appl. Mech. Engrg. 291 (2015) 102–122
Computer methods in applied mechanics and engineering
www.elsevier.com/locate/cma

A Bayesian approach to selecting hyperelastic constitutive models of soft tissue

Sandeep Madireddy, Bhargava Sista, Kumar Vemaganti*

Department of Mechanical and Materials Engineering, University of Cincinnati, Cincinnati, OH 45221-0072, United States

Received 17 November 2014; received in revised form 11 March 2015; accepted 22 March 2015
Available online 6 April 2015

Abstract

Hyperelastic constitutive models of soft tissue mechanical behavior are extensively used in applications like computer-aided surgery, injury modeling, etc. While numerous constitutive models have been proposed in the literature, an objective method is needed to select a parsimonious model that represents the experimental data well and has good predictive capability. This is an important problem given the large variability in the data inherent to soft tissue mechanical testing.

In this work, we discuss a Bayesian approach to this problem based on Bayes factors. We propose a holistic framework for model selection, wherein we consider four different factors to reliably choose a parsimonious model from the candidate set of models. These are the qualitative fit of the model to the experimental data, evidence values, maximum likelihood values, and the landscape of the likelihood function. We consider three hyperelastic constitutive models that are widely used in soft tissue mechanics: Mooney–Rivlin, Ogden and exponential. Three sets of mechanical testing data from the literature for agarose hydrogel, bovine liver tissue, porcine brain tissue are used to calculate the model selection statistics. A nested sampling approach is used to evaluate the evidence integrals. In our results, we highlight the robustness of the proposed Bayesian approach to model selection compared to the likelihood ratio, and discuss the use of the four factors to draw a complete picture of the model selection problem.

Keywords: Bayesian; Uncertainty quantification; Soft tissue constitutive model; Hyperelastic; Nested sampling; Model selection

1. Introduction

1.1. Motivation

Hyperelastic constitutive models of soft tissue mechanical behavior find widespread use in applications like traumatic brain injury simulation, computer-aided surgery and functional tissue engineering. These models are generally phenomenological in nature and are derived from a strain energy density function that is postulated to depend on invariants of the deformation field in the elastic body. The model parameters are usually determined by a

* Corresponding author.
E-mail addresses: madiresy@mail.uc.edu (S. Madireddy), sistasa@mail.uc.edu (B. Sista), Kumar.Vemaganti@uc.edu (K. Vemaganti).

http://dx.doi.org/10.1016/j.cma.2015.03.012
0045-7825/

least squares fit of the constitutive model to mechanical testing data from experiments such as compression, tension [1–3], indentation [4,5], aspiration [6,7], etc.

For most problems, the choice of the constitutive model is driven by the quality of the fit (e.g., R^2 value) and the analyst's experience. Commonly used models include those based on strain invariants (e.g., Neo-Hookean, Mooney–Rivlin, exponential, Pucci–Saccomandi, Arruda–Boyce, etc.), or on the principal stretches (e.g., Ogden, Peng–Landel, etc.). See [8,9] for a comprehensive list. Given this wide range of available models, an objective approach is needed to ensure that the chosen constitutive model strikes a balance between fit quality, predictive capability, and complexity. A simple way to measure the model's complexity is through the dimensionality of its parameter space. A model that is too simple and has too few parameters may not faithfully represent the experimental mechanical response, while a model that is overly complex may be of limited use or have so many parameters that they cannot be easily identified from the available data. Thus the issue of parameter estimation is an integral part of the model selection process since the model is complete only after the parameters are estimated.

Another issue of concern in soft tissue mechanics is that the experimental data almost always have high variability, which leads to large uncertainties in the constitutive model parameters. This large variability is due to many factors: (i) soft tissue is a complex material that can be highly heterogeneous; (ii) patient to patient differences can be large; (iii) soft tissue is difficult to procure, handle and preserve, and obtaining good specimens is challenging; and (iv) experimental protocols vary widely between laboratories. These difficulties are exacerbated when the problem of interest involves irreversible and discontinuous phenomena like tissue fracture and damage, a common occurrence in injury biomechanics [10]. This variability cannot be ignored and should be systematically included in the parameter estimation and model selection process in order to obtain a reliable constitutive model.

1.2. Model selection strategies

Model selection procedures often rank the candidate models based on some widely accepted criteria. The selected model should obey the principle of parsimony or Occam's razor, which states that the simplest model that can explain the data should be accepted. That is, a balance between goodness of fit and model complexity has to be achieved. This property of the model is essential to avoid overfitting and to render the model more testable. Most model selection methods can be categorized into three groups: (a) Bayesian methods using the Bayes factor [11,12], (b) frequentist methods that include p-value method, Mallows's C_p, etc., and (c) Information-theoretic methods like the Akaike Information Criterion (AIC), Bayesian Information Criterion (BIC), and variants [13].

The Bayesian approach has some advantages over frequentist methods [14]. First, it is easier to interpret the posterior probabilities of the model and the Bayes factor as the odds of one model over the other. Second, the Bayesian approach is consistent in the sense that it guarantees the selection of the true model if it is part of the candidate model set under very mild conditions. If the true model is not present in the set, then the model closest to the true model in terms of the Kullback–Leibler divergence is chosen [15,16].

The Bayesian approach mainly involves updating prior knowledge of the model with new experimental observations (represented by a likelihood function) to obtain the posterior distribution of the model parameters. The marginal likelihood or evidence is then obtained by marginalizing the likelihood function over the space of its existence; this is discussed in more detail in Section 3.

Bayesian model selection is a natural Occam's razor. The use of the marginal likelihood automatically penalizes overly complex models because complex models spread their probability mass very widely and predict that everything is possible. Thus the probability of the actual data is small. This is illustrated in Fig. 1 (adapted from [17]), where the model space $\mathfrak{M}$ consists of three models $\boldsymbol{m}_1, \boldsymbol{m}_2, \boldsymbol{m}_3$ with equal probabilities, each with a single parameter θ. The data space $\mathfrak{D}$ is assumed to be one-dimensional, and $\boldsymbol{d} \in \mathfrak{D}$ is the dataset used to estimate the parameter θ, which is assumed to be a single value. Model $\boldsymbol{m}_3$ is a more complex model in the sense that it can predict more data, while model $\boldsymbol{m}_1$ is the simplest. The dashed straight line corresponds to a particular observed dataset $\boldsymbol{d}$. The quantity $P(\boldsymbol{d}|\boldsymbol{m}_i)$ represents the probability of obtaining the data given a particular model $\boldsymbol{m}_i$; this is the marginal likelihood or evidence. The quantity $P(\theta|\boldsymbol{d}, \boldsymbol{m}_i)$ represents the posterior probability of the parameter given the dataset and model. For the example shown here, model $\boldsymbol{m}_1$ is unable to predict the data, model $\boldsymbol{m}_2$ predicts the data and has the highest evidence value, while model $\boldsymbol{m}_3$ also predicts the data but with a lower evidence value. Thus model $\boldsymbol{m}_2$ is the model of choice.

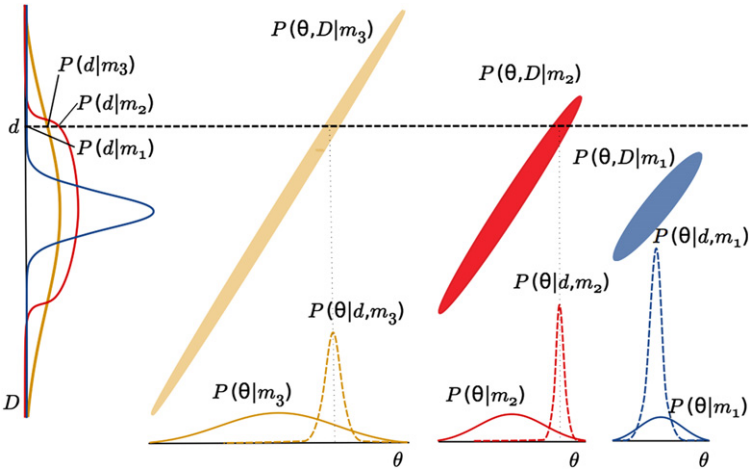

Fig. 1. Occam's Razor: Figure adapted from [17]. Competing models $\boldsymbol{m}_1$ (simplest) through $\boldsymbol{m}_3$ (most complex), each with a single parameter θ, attempt to explain observed data $\boldsymbol{d}$. The evidence or the probability of obtaining the data given the model, $P(d|m_k)$, is highest for model $\boldsymbol{m}_2$, and the more complex $\boldsymbol{m}_3$ is not needed.

The Bayesian approach is also more general in the sense that it does not require nested models, standard distributions or any asymptotic assumptions while the other methods require at least one of these.

1.3. Choice of the prior

Another important aspect of model selection and perhaps the most criticized aspect of the Bayesian framework is the specification of the prior probability distribution on the parameters. For the case of parameter estimation, the selection of the prior is not as critical when the data is informative. But in the case of model selection, the prior indeed influences the evidence integral and needs to be well justified. However this influence can be expected to decrease when the data is more informative [18].

In general, the selection of the prior is done through a sensitivity analysis based on a few appropriate priors. In the situation where the prior distribution is not exactly known, one often chooses a non-informative prior, i.e., one whose contribution to the posterior is minimal.

One approach to constructing a non-informative prior is based on the principle of indifference and the principle of insufficient reason. Here, one assumes uniform distributions on all the parameters to span an infinite range $[-\infty, \infty]$. This covers all possible true values of the parameters and gives them equal probability. The problem with this is that in order for the probability to sum to one, infinitely small probability has to be placed on the parameters, which when multiplied by a likelihood obtained from a low information data can lead to negligible evidence. One way to get around this is to constrain the parameter space to the bounds in which the likelihood is estimated to be appreciable. This can be done from expert opinion or by inspection in several regions of the parameter space.

Other approaches to choosing non-informative priors include:

(a) Constructing a prior based on the invariance structure in the problem using the right Haar measure;
(b) Constructing a prior that is invariant under reparametrization – e.g., Jeffreys prior [19];
(c) Reference priors: Choosing a prior that maximizes the Kullback–Leibler distance between posterior and the prior [20];
(d) Maxent priors: Choosing a prior that maximizes the entropy [21]; and
(e) Other methods such as data-translated likelihoods, probability matching [22].

These priors generally turn out to be improper and do not integrate to one, which can be problematic [22]. One way to get around improper priors is by using partial or alternative Bayes factors [23], wherein a part of the data called the training set is used to make the prior proper. This can then be used as an informative prior for the rest of the data. Some of the ways to obtain partial Bayes factors include pseudo-Bayes factor [24], posterior Bayes factor [25], intrinsic Bayes factor [26], and fractional Bayes factor [27].

1.4. Evidence integral evaluation

Although Bayesian model selection has many advantages over other approaches, one of its major drawbacks is the cost of evaluating the multidimensional evidence integral. The evidence integral can be evaluated analytically only for a few cases in which the closed form of the posterior is available. For example, this is possible for the exponential family distributions with conjugate priors [11]. Under the assumption of large data, asymptotic approximation methods (Laplace method, Schwarz method) can be used.

In the case of more complicated distributions, sampling-based methods like simple Monte Carlo, importance sampling, Markov Chain Monte Carlo (MCMC) and its variants like delayed rejection adaptive metropolis (DRAM) [28], or differential evolution adaptive metropolis (DREAM) [29] can be used. These methods are computationally expensive, and tend to have problems with sampling efficiency for high dimensional multimodal distributions with large degeneracies between parameters. Also the proposal distribution needs to be carefully tuned for efficiency, and checking for convergence can be problematic [30]. Adaptive quadrature-based methods [31] are suitable for low dimensions. Some fast approximating methods include thermodynamic integration, Gaussian approximation of each mode, Savage–Dickey ratio, and bridge sampling, but they are only suitable for a few special situations [32]. One of the more recent and efficient sampling-based methods is the nested sampling approach [33]. This along with its implementation MULTINEST [30] has been shown to be more efficient than the aforementioned methods, and can be used to obtain the evidence value and the parameter estimates simultaneously. This is the approach employed in our study.

1.5. Bayesian model selection for phenomenological models

The Bayesian model selection procedure has not been explored in depth in the context of soft tissue constitutive models. For phenomenological models in general, Prudencio et al. [34] used this approach to select between two competing continuum damage models to predict the failure in a composite material wherein a parallel adaptive multilevel sampling algorithm [35] is used to calculate the evidence integral. Oden et al. [36] used a similar procedure to select between two phenomenological models of tumor growth in living tissues using synthetic data. For structural dynamics problems, Beck and Yuen [37], Beck [38], and Yuen [39] advocated the use of asymptotic approximations of the evidence integral to calculate Bayes factors for globally identifiable cases (large datasets) and MCMC-based methods for globally unidentifiable cases (small datasets).

Elsheikh et al. [40] used a nested sampling-based Bayesian model selection approach to choose the number of terms in the Karhunen–Loève (KL) expansion to build a surrogate for subsurface flow models. Hadjidoukas et al. [41] used the transitional MCMC (TMCMC) method to calculate the evidence integral to select among three discrete element method models used to simulate a granular material. Sandhu et al. [42] calculated the evidence integral for 60 different nested models that describe the response of nonlinear aeroelastic oscillatory systems. Chib–Jeliazkov estimator [43] calculated using parallel adaptive MH based MCMC algorithm is used to evaluate the evidence integral.

Evidence integral calculation based on MULTINEST algorithm is more widely used in the field of cosmology [18]. Martin et al. [44] used this procedure to choose between two different classes of models that describe the cosmological inflationary potential. Feroz et al. [32] used it to select among two variants of constrained minimal supersymmetric standard model. In the field of systems biology, Pullen and Morris [45] used nested sampling based on MULTINEST for model selection among four different models.

1.6. Current research

In this article, we address the question of model selection and ranking among competing soft tissue constitutive models based on the Bayesian framework. Specifically, three phenomenological constitutive models (Ogden, Mooney–Rivlin and exponential) that describe the hyperelastic response of soft tissue are considered. For the multiterm Ogden model, we consider $N = 1, 2, 3$. This allows us to perform two types of comparisons among the models: models of the same form but different levels of complexity, and models of the same complexity but different forms. Three sets of mechanical testing data from the literature are used: agarose hydrogel, bovine liver tissue, and porcine brain tissue. For each dataset, we calculate the model parameters for each of the aforementioned phenomenological models using Bayesian inference, and evaluate evidence values and Bayes factors. We then rank the models using two scales: Bayes factors and likelihood ratios. We use the nested sampling algorithm MULTINEST [30] to evaluate the

evidence integrals. Two distinct priors are considered: uniform prior and Gaussian prior. This allows us to study the effect of the prior on the model selection statistics and determine Bayesian robustness.

We propose a holistic approach for model selection, wherein we consider four important factors to reliably choose a parsimonious model from the candidate set of models. These are: the fit of the model with the maximum likelihood parameters to the data, evidence values, maximum likelihood values, and the landscape of the log-likelihood function. Each of these factors provides a different piece of information about the model. A good fit ensures that the model is adequate to capture the physics of the problem, while the log-likelihood value quantifies the goodness of the fit. A high evidence value provides a balance between goodness-of fit and model complexity, while the landscape of the likelihood function provides clues to any overfitting issues. These four pieces of information have to be considered together in order to find a parsimonious model in the candidate set (provided such a model exists among the candidates).

We compare the model rankings that results from the two model selection criteria (Bayes factors and likelihood ratios) for each of the datasets. Our results show that a model selection approach based on the Bayes factor is a robust way to carry out model selection when the evidence integral evaluation and the sensitivity to the prior probability are properly addressed.

The paper is structured as follows. Following this Introduction, Section 2 describes the characterization of the material behavior of soft tissues and the constitutive models used in this study. In Section 3 we introduce the Bayesian model selection framework and its different components adopted in this work. In Section 4, we present the experimental data and the model selection statistics obtained for various constitutive models and datasets. This is followed by a discussion of the results in Section 5 and some concluding remarks in Section 6.

2. Materials and methods

In this study, the deformation of soft tissue is modeled using the theory of continuum mechanics. We assume that soft tissue is an isotropic incompressible hyperelastic medium. Though soft tissue is often viscoelastic and anisotropic [46,47], we make the assumption of hyperelasticity to simplify the analysis. The validity of this assumption is discussed further in Section 4.1.

2.1. Continuum mechanics

Let $\mathbf{X}$ and $\mathbf{x}$ be the reference and current coordinates of a particle in the body undergoing deformation $\mathbf{x} = \chi(\mathbf{X})$. The deformation gradient defined as $\mathbf{F} = \frac{\partial \mathbf{x}}{\partial \mathbf{X}}$, and the right Cauchy–Green deformation tensor is $\mathbf{C} = \mathbf{F}^{\mathrm{T}}\mathbf{F}$. The eigenvalues of $\mathbf{F}$ are the principal stretches and denoted λ_k, $k = 1, 2, 3$. Then the eigenvalues of $\mathbf{C}$ are given by λ_k^2. The principal invariants of $\mathbf{C}$ are

$$I_1(\mathbf{C}) = \mathrm{trace}(\mathbf{C}), \qquad I_2(\mathbf{C}) = \frac{1}{2}((I_1(\mathbf{C}))^2 - \mathrm{trace}(\mathbf{C}^2)), \qquad I_3(\mathbf{C}) = \det(\mathbf{C}). \tag{1}$$

For a hyperelastic material, the strain energy density function Ψ can be expressed in terms of the right Cauchy–Green deformation tensor $\mathbf{C}$, or principal stretches λ_i

$$\Psi = \Psi(\mathbf{C}) = \Psi(\lambda_1, \lambda_2, \lambda_3). \tag{2}$$

For an incompressible isotropic material, the second Piola–Kirchhoff stress tensor $\mathbf{S}$ can be obtained from Ψ using the expression

$$\mathbf{S} = -p\mathbf{C}^{-1} + 2\frac{\partial \Psi}{\partial \mathbf{C}} \tag{3}$$

where p is a Lagrange multiplier term that arises from the assumption of incompressibility. This means that the stress in the material cannot be determined from the strain energy density alone since a hydrostatic term can do no work. Thus it must be found using the boundary conditions of the problem at hand. For instance, in the case of uniaxial tension in say direction 1, we set $S_{22} = S_{33} = 0$ to determine the unknown term p. Subsequently the Cauchy stress tensor can be obtained from $\mathbf{S}$ as

$$\boldsymbol{\sigma} = \mathbf{F}\mathbf{S}\mathbf{F}^{\mathrm{T}} = -p\mathbf{I} + 2\mathbf{F}\frac{\partial \Psi}{\partial \mathbf{C}}\mathbf{F}^{\mathrm{T}}. \tag{4}$$

In this study, the hyperelastic response is characterized using three different phenomenological constitutive models. They are the Mooney–Rivlin model, the Ogden model, and the exponential model.

The strain energy density function and the corresponding uniaxial stress for the Mooney–Rivlin model Ψ are given by

$$\Psi = C_1(I_1 - 3) + C_2(I_2 - 3) \tag{5}$$

$$\sigma_{11}(\lambda) = \left(2C_1 + \frac{2C_2}{\lambda}\right)\left(\lambda - \frac{1}{\lambda^2}\right) \tag{6}$$

where I_1 and I_2 are the first and the second principal invariants of $\mathbf{C}$, λ is the principal stretch in the direction of loading, and C_1 and C_2 are the model parameters. The shear modulus of the material is $\mu = 2(C_1 + C_2)$.

For the Ogden model, the strain energy density function and the uniaxial stress are given by

$$\Psi = \sum_{i=1}^{N} \frac{2\mu_i}{{\alpha_i}^2}(\lambda_1^{\alpha_i} + \lambda_2^{\alpha_i} + \lambda_3^{\alpha_i} - 3) \tag{7}$$

$$\sigma_{11}(\lambda) = \sum_{i=1}^{N} \frac{2\mu_i}{\alpha_i}(\lambda^{\alpha_i - 1} - \lambda^{-\frac{\alpha_i}{2} - 1}) \tag{8}$$

where μ_i and α_i are model parameters and the value of N specifies the order of the model. The shear modulus of the material is given by $\mu = \Sigma_{i=1}^{N} \mu_i$.

Finally, for the exponential model, the strain energy density function and the resulting uniaxial stress are given by

$$\Psi = B_1(\exp(B_2(I_1 - 3)) - 1) \tag{9}$$

where B_1 and B_2 are now the model parameters. The uniaxial Cauchy stress is

$$\sigma_{11}(\lambda) = 2\left(\lambda - \frac{1}{\lambda^2}\right) B_1 B_2 \exp\left(B_2\left(\lambda^2 + \frac{2}{\lambda} - 3\right)\right) \tag{10}$$

where B_1 and B_2 are the model parameters and the shear modulus is proportional to the product of the parameters: $\mu = K B_1 B_2$.

The Mooney–Rivlin model, the exponential model and the Ogden model with $N = 1$ have the same complexity in the sense that they are all characterized by two material parameters. On the other hand, the Ogden model with $N = 1$, $N = 2$ and $N = 3$ leads to models of different complexities characterized by two, four, and six parameters, respectively.

3. Model selection

3.1. Model selection framework

Assume $\mathfrak{S}$ to be the physical system under study, and model space $\mathfrak{M}$ represents the space of all models that characterize the system. Every model $\boldsymbol{m}_k$ from the space $\mathfrak{M}$ is a functional form $f(\boldsymbol{\theta}|\boldsymbol{m}_k)$ with a fixed number of parameters $\boldsymbol{\theta}$. Let $\mathfrak{D}$ be the data space which is the space of all possible instrument responses.

We hypothesize that the model and the data are related through

$$f(\boldsymbol{\theta}|\boldsymbol{m}_k) = \boldsymbol{d} + \boldsymbol{\eta} \tag{11}$$

where $\boldsymbol{d} \in \mathfrak{D}$ is the hypothetical true data, and $\boldsymbol{\eta}$ is the noise that contributes uncertainty to any experiment or observation that involves measurements. For soft tissue, this noise generally arises from a combination of the inherent variability in the material, its environment, and the measurement itself. Note that $\boldsymbol{d} + \boldsymbol{\eta}$ is the actual system response obtained from the experiment. We assume for simplicity that the data $\boldsymbol{d}$ is known in the form of a mean vector $\boldsymbol{D}$ and a vector of standard deviations $\boldsymbol{\Delta}$. Let the cardinality of the candidate set of models be N, the prior probability of each

model to be $P(\boldsymbol{m}_k|I)$. Assuming that $\sum_{k=1}^{N} P(\boldsymbol{m}_k|I) = 1$, from Bayes theorem it follows that

$$P(\boldsymbol{m}_k|\boldsymbol{d}, I) = \frac{P(\boldsymbol{d}|\boldsymbol{m}_k, I)\,P(\boldsymbol{m}_k|I)}{\sum_{k=1}^{N} P(\boldsymbol{d}|\boldsymbol{m}_k, I)\,P(\boldsymbol{m}_k|I)}. \tag{12}$$

To ascertain the relative merits of a pair of models $\boldsymbol{m}_k, \boldsymbol{m}_j$ the ratio of their posteriors in light of the experimental data is calculated as

$$\frac{P(\boldsymbol{m}_k|\boldsymbol{d}, I)}{P(\boldsymbol{m}_j|\boldsymbol{d}, I)} = \frac{P(\boldsymbol{d}|\boldsymbol{m}_k, I)}{P(\boldsymbol{d}|\boldsymbol{m}_j, I)} \frac{P(\boldsymbol{m}_k|I)}{P(\boldsymbol{m}_j|I)} \tag{13}$$

or

$$\text{Posterior odds} = BF_{kj} \times \text{Prior odds} \tag{14}$$

where BF_{kj} is the Bayes factor, which transforms the prior odds into the posterior odds in light of the data $\boldsymbol{d}$. Here, $P(\boldsymbol{d}|\boldsymbol{m}_k, I)$ is called the marginal likelihood or evidence, and is obtained by marginalizing the posterior of the parameters for a given model $P(\boldsymbol{\theta}_k|\mathbf{d}, m_k, I)$ over its parameter space $\boldsymbol{\theta}_k$.

$$P(\mathbf{d}|m_k, I) = \int_{\boldsymbol{\theta}_k} P(\mathbf{d}|\boldsymbol{\theta}_k, m_k, I)\,P(\boldsymbol{\theta}_k|m_k, I)\,\mathrm{d}\boldsymbol{\theta}_k. \tag{15}$$

The quantity $P(\mathbf{d}|\boldsymbol{\theta}_k, m_k, I)$ is called the likelihood function, which gives the likelihood of observing the data $\mathbf{d}$ given the parameters $\boldsymbol{\theta}_k$, and $P(\boldsymbol{\theta}_k|m_k, I)$ is the prior knowledge about the parameters before the experimental data are obtained.

The evidence can be seen as the probability mass (volume) contained under the posterior probability distribution over the parameter space for a given model. The value of the evidence would be higher for a model if more of its parameter space is at a higher likelihood than the model for which the likelihood distribution is highly peaked for a small parameter space and the rest of it has low likelihood values.

3.2. Choice of the likelihood function

The data in this study is obtained from experimental measurements on soft tissue and hydrogels. The noise associated with the experimental measurements, $\boldsymbol{\eta}$, can reasonably be represented as a Gaussian process based on maximum entropy considerations [48]. The likelihood function, which is a joint distribution of this experimental data assuming the data are independent and identically distributed (i.i.d) can be written as

$$P(\mathbf{d}|\boldsymbol{\theta}_k, m_k, I) = \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi\,{\Delta_i}^2}} \exp\left[-\frac{(f(\boldsymbol{\theta}_k|\boldsymbol{m}_k) - D_i)^2}{2{\Delta_i}^2}\right] \tag{16}$$

where n is the number of points at which the data are measured. It is better to work with the logarithm of the likelihood than with the likelihood function itself since it varies slowly with respect to the parameters. The logarithm of the likelihood is given by

$$\log\left[P(\mathbf{d}|\boldsymbol{\theta}_k, m_k, I)\right] = \sum_{i=1}^{n} -\frac{1}{2}\log(2\pi\,{\Delta_i}^2) - \sum_{i=i}^{n}\left[\frac{(f(\boldsymbol{\theta}_k|\boldsymbol{m}_k) - D_i)^2}{2{\Delta_i}^2}\right]. \tag{17}$$

3.3. Algorithm to evaluate the evidence integral

As mentioned earlier, we use Nested Monte Carlo sampling based algorithm, MULTINEST to evaluate the evidence integral. It is a Monte Carlo technique used to estimate Bayesian evidence, wherein the parameter estimates are obtained as a by-product. The evidence is calculated by transforming the multi-dimensional evidence integral over the parameter space into a one-dimensional integral over the likelihood space. This is illustrated in Fig. 2.

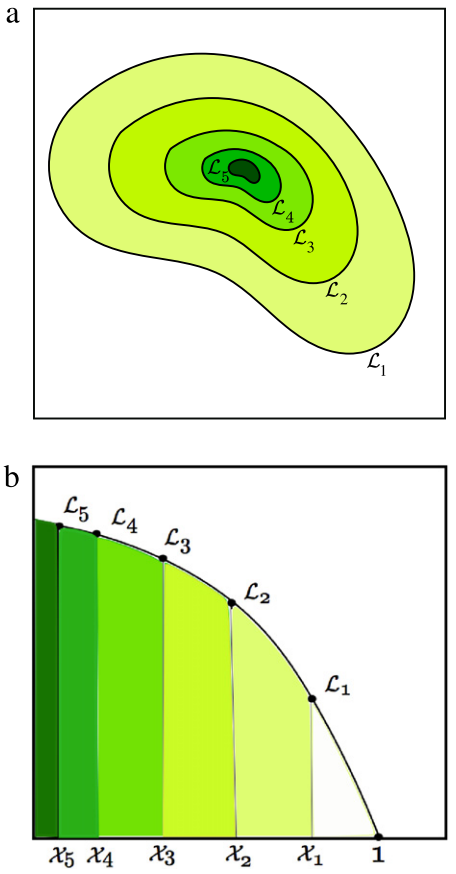

Fig. 2. Schematic of the nested sampling approach: (a) Contours of likelihood function, and (b) likelihood as a function of the prior volume X.

Adopting the terminology of [30], a differential element of the prior "mass" is defined by $\mathrm{d}X = \pi(\Theta)\mathrm{d}\Theta$, where $\pi(\Theta)$ is the assumed prior distribution. The prior volume is obtained from the expression

$$X(\gamma) = \int_{\{\theta:\mathcal{L}(\Theta)>\gamma\}} \pi(\Theta)\mathrm{d}\Theta \tag{18}$$

which is the proportion of the prior with likelihood greater than γ. Subsequently, the evidence integral can be calculated using the prior volume to be

$$\mathcal{Z} = \int_0^\infty X(\gamma)\mathrm{d}\gamma.$$

When $\mathcal{L}(X)$, the inverse of $X(\gamma)$ exists, i.e., $\mathcal{L}(X(\gamma)) = \gamma$, the evidence can be rewritten as

$$\mathcal{Z} = \int_0^1 \mathcal{L}(X)\mathrm{d}X.$$

This is the area under the X—$\mathcal{L}$ curve, as seen in Fig. 2.

The overall algorithm can be written as [30]:

1. Sample the prior on parameters n times to get a set of n sample points (active points)
2. Calculate likelihood for each point
3. Sort the points based on the likelihood
4. Calculate the evidence for the set
5. Remove the point with the lowest likelihood ($\mathcal{L}^*$)

Table 1
Jeffreys Scale, from [11].

$2\log_e(B_{ab})$	B_{ab}	Evidence against model M_b	
0–2	1–3	Not worth more than a bare mention	Class 1
2–6	3–20	Substantial	Class 2
6–10	20–150	Strong	Class 3
>10	>150	Decisive	Class 4

6. Generate a new sample point from the prior subjected to the constraint $\mathcal{L}(\Theta) > \mathcal{L}^*$
7. Calculate the evidence for the new set
8. Check if the difference between the current and the previous evidence is within a tolerance
9. If yes, exit. Else go to step 3.

Generating unbiased samples efficiently from the prior subject to the constraint in step 6 is a major challenge in the nested sampling algorithm. MULTINEST uses the simultaneous ellipsoidal rejection sampling scheme to address this challenge. In this sampling scheme, the set of active points at every iteration is partitioned into clusters, and ellipsoids are constructed around them using the expectation–minimization (EM) algorithm discussed in [30] to approximate each cluster. A new sample is generated from the union of these ellipsoids such that it satisfies the constraint in step 6. This algorithm reliably isolates the multiple modes and evaluates the statistics for each mode separately. We set the number of live points to 4000; the convergence tolerance for Step 8 in the algorithm above is 0.01 [32]. We use a modified version of the MULTINEST algorithm implementation by Feroz et al. [30] in our computations. The modifications include updated likelihood functions and an extra level of parallelization to allow the simultaneous computation of statistics for multiple models. The latter is accomplished using the message passing interface (MPI) [49]. All the computations are carried out on an in-house cluster.

3.4. Choice of prior probability distribution function

In this study, we use two different priors: the uniform prior $\mathcal{U}(-100, 100)$ and the Gaussian prior $\mathcal{N}(0, 100)$ over all the parameters to study the sensitivity of the maximum likelihood and evidence to the prior choice. The range of the uniform prior is carefully chosen to be large enough to encompass the parameter space in which the likelihood is appreciable. Also it should be noted that nested sampling algorithm generates samples from the parameter space of the prior, so improper priors over long ranges cannot be handled efficiently. This is one of the main drawbacks of the nested sampling approach, wherein choosing a large range prior could have a significant impact on the convergence rate and computational cost. For this reason, we do not use priors such as the reference prior or the maxent prior discussed earlier in Section 1.3.

3.5. Interpretation of the Bayes factor

The Bayes factor represents the evidence provided by the data in favor of one hypothesis or model over the other. The values of the Bayes factors can be interpreted based on Jeffreys scale, which is shown in Table 1 [11].

4. Results

4.1. Experimental data

The three different datasets used in this study are obtained from the literature. These datasets correspond to agarose hydrogel [50], bovine liver tissue [52], and porcine brain tissue [51]. The details of the experimental testing are summarized in Table 2. For each dataset, we obtain parameter estimates for the three hyperelastic material models and study the model selection aspects discussed earlier.

The assumption of hyperelasticity is generally valid when the strain rates under consideration are low. The three sets of mechanical testing data used in our work all involve slow strain rates (0.01/s–0.04/s), which allows us to ignore any viscoelastic effects and idealize the materials as hyperelastic.

Table 2
Experimental data used in the study.

	Data A	Data B	Data C
Material	Agarose hydrogel	Bovine liver	Porcine brain
Testing mode	Unconfined compression	No-slip uniaxial compression	Uniaxial tension
Strain range	0%–50%	0%–30%	0%–30%
Strain rate	0.01/s	0.04/s	0.01/s
Noise level	Low	Low	High
Source	Billade [50]	Roan & Vemaganti [1]	Rashid et al. [51]

Table 3
The effect of the prior: Maximum of the log-likelihood function determined by the Bayesian approach.

Material model	Data A		Data B		Data C	
	Uniform prior	Gaussian prior	Uniform prior	Gaussian prior	Uniform prior	Gaussian prior
Mooney–Rivlin	−184.537	−184.537	−40.256	−40.256	54.508	54.508
Exponential	−136.631	−136.631	19.415	19.415	53.968	53.968
Ogden ($N = 1$)	−942.329	−942.329	18.332	18.332	53.029	53.029
Ogden ($N = 2$)	30.931	30.929	19.376	19.403	54.897	54.898
Ogden ($N = 3$)	31.280	31.083	19.407	19.411	54.899	54.899

The other important assumption we make is that of material isotropy. Agarose, one of the materials considered here, is a polysaccharide with an isotropic microstructure consisting of a porous interconnected network [50]. Liver tissue is also generally considered isotropic [1], though some studies assume transverse isotropy [53]. Brain tissue is generally considered anisotropic, but the significance of this anisotropy is still an open question. Some researchers have found considerable directional dependence of material properties [54], while others have found the opposite [55]. For the purpose of our study, we take brain tissue to be isotropic as it somewhat simplifies the constitutive model form and therefore reduces the number of model parameters that have to be estimated. Moreover, since the experimental data at hand are from a single mode of deformation, an isotropic model is more appropriate. We plan to extend our results to the case of anisotropic constitutive models as part of our future work. This step depends on the availability of data from mechanical testing in multiple directions.

4.2. Model fits

The maximum likelihood estimate (MLE) is obtained for each of the candidate models (Mooney–Rivlin, exponential and Ogden ($N = 1, 2, 3$)) for the three sets of experimental data using MULTINEST. The resulting fits to data A using these MLE are shown in Fig. 4. Qualitatively, the fits for Mooney–Rivlin, exponential, and Ogden ($N = 1$) models are poor, whereas the Ogden ($N = 2, 3$) models provide good fits.

Similarly the model fits for Data B are shown in Fig. 5. The Mooney–Rivlin model provides a poor fit, whereas the exponential and Ogden ($N = 1, 2, 3$) models result in good fits.

The results for Data C are shown in Fig. 6. Mooney–Rivlin, Ogden ($N = 1$), and exponential models lead to poor fits. Ogden ($N = 2, 3$) models provide qualitatively good fits.

4.3. Log-likelihood function, log-evidence, and Bayes factors

The values of the log-likelihood function corresponding to the MLEs for each of the models for the three datasets are shown in Table 3. For each case, these values are obtained using two priors: uniform prior and Gaussian prior. There is no discernible difference in the values obtained from the two priors. For Data A, the Mooney–Rivlin, exponential and Ogden ($N = 1$) models all show very low log-likelihood values, which is not surprising given the poor fits obtained from these models. A similar situation exists with Data B, where good visual fits correspond with high log-likelihood values.

For Data C, however, the situation is different. All models have comparable log-likelihood values even though the fits from Ogden ($N = 2, 3$) seem clearly better (qualitatively speaking) than the rest.

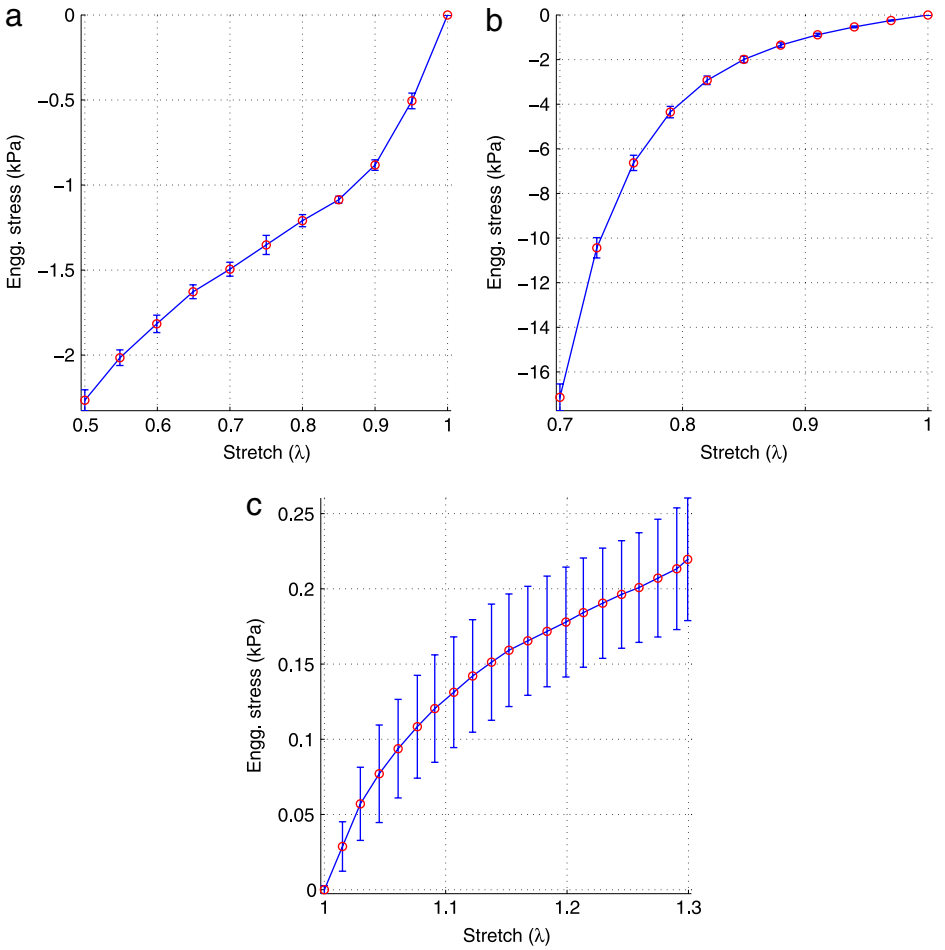

Fig. 3. Experimental stress–strain data used in the study: (a) Data A: Unconfined compression of agarose, (b) Data B: Uniaxial compression of bovine liver, and (c) Data C: Uniaxial extension of porcine brain tissue.

In Table 4, we list the values of the evidence integral for all the cases. The evidence integral is calculated using MULTINEST, with the tolerance value for log-evidence set at 0.01. Additionally, the software estimates the error in the integral, which is also shown in the table.

Bayes factors are calculated for all the models with respect to the model that has the highest value of evidence. For data A, the model with the highest evidence value is the Ogden ($N = 2$) model, while the Ogden ($N = 1$) model has the highest evidence for Data B and Data C. The resulting values of the Bayes factors are classified based on the Jeffreys scale, which is shown in Table 5.

Next, we rank the candidate models based on the Bayes factors and the classical log-likelihood ratios. The ranks for Data A, Data B, and Data C are shown in Tables 6–8, respectively. In all cases, the two criteria lead to different rankings of the models. In particular, for Data C, the difference is dramatic and the rankings of the models by the two criteria are completely reversed. The model considered the best by the Bayes factor criterion is the worst one according to the log-likelihood ratio and vice versa.

Additionally we look at the landscape of the likelihood function (based on the work by [56]) as a crucial criterion to consider in the model selection process. For Data A, the log-likelihood function shows distinct peaks for the Mooney–Rivlin, exponential, and Ogden ($N = 1, 2$), but shows a relatively flat likelihood surface for Ogden ($N = 3$). For Data B, distinct peaks are obtained for Mooney–Rivlin, exponential, and Ogden ($N = 1$) but the log-likelihood surface is flat for Ogden ($N = 2, 3$). The landscape of the log-likelihood functions for the five models for Data C is similar to that of Data B.

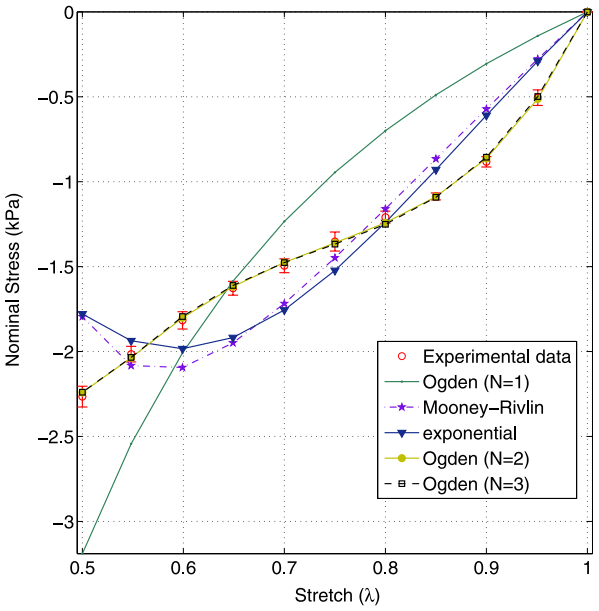

Fig. 4. Data A: Comparison of Mooney–Rivlin, exponential and Ogden model fits.

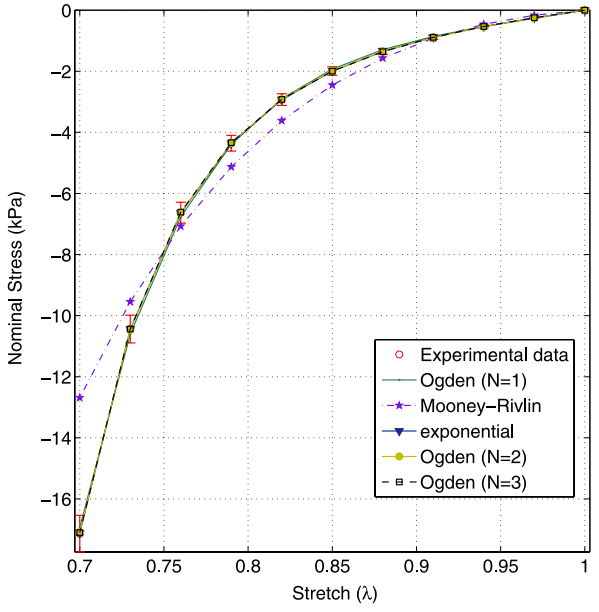

Fig. 5. Data B: Comparison of Mooney–Rivlin, exponential and Ogden model fits.

5. Discussion

5.1. Dependence on the prior

For the data and models considered here, the maximum value (peak) of the log-likelihood function is insensitive to the change in the prior (Table 3), and so are the corresponding values of the MLE. This means the data is informative enough that the obtained values of the MLE are data driven rather than prior driven. Thus one can have more confidence in the values of MLE and the maximum log-likelihood.

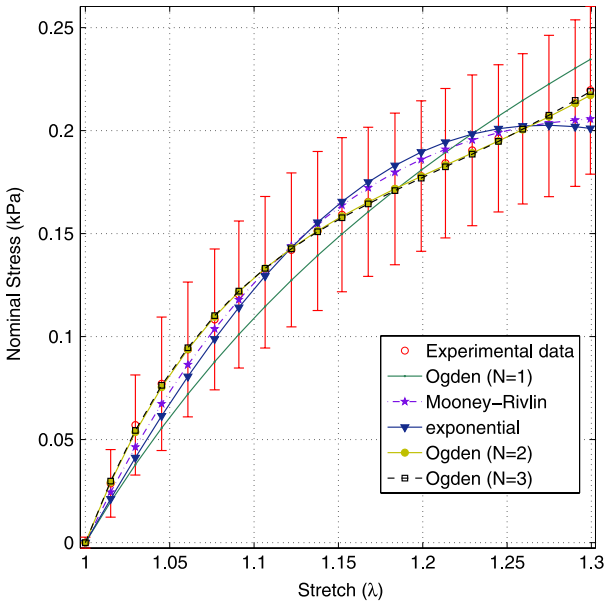

Fig. 6. Data C: Comparison of Mooney–Rivlin, exponential and Ogden model fits.

Table 4
The effect of the prior: Marginal likelihood (Evidence) value determined by the Bayesian approach.

Material model	Data A		Data B		Data C	
	Uniform prior	Gaussian prior	Uniform prior	Gaussian prior	Uniform prior	Gaussian prior
Mooney–Rivlin	-202.836 ± 0.066	-203.261 ± 0.067	-53.109 ± 0.054	-53.558 ± 0.055	39.385 ± 0.056	39.046 ± 0.060
Exponential	-153.915 ± 0.064	-154.398 ± 0.065	3.525 ± 0.061	3.234 ± 0.062	39.844 ± 0.056	39.844 ± 0.055
Ogden ($N = 1$)	-958.146 ± 0.061	-958.567 ± 0.062	5.778 ± 0.053	5.315 ± 0.056	41.721 ± 0.051	41.021 ± 0.053
Ogden ($N = 2$)	1.046 ± 0.083	1.411 ± 0.085	1.740 ± 0.062	0.902 ± 0.064	37.827 ± 0.062	36.973 ± 0.064
Ogden ($N = 3$)	-3.534 ± 0.086	-4.358 ± 0.085	-0.771 ± 0.067	-0.22 ± 0.069	34.951 ± 0.068	33.466 ± 0.070

Table 5
Bayes factors for the various datasets (MR: Mooney–Rivlin, EX: exponential, O1: Ogden ($N = 1$), O2: Ogden ($N = 2$), O3: Ogden ($N = 3$)).

Data A		Data B		Data C	
Bayes factor	Class	Bayes factor	Class	Bayes factor	Class
$\log_e(B_{O2-MR}) = 203.882$	Class 4	$\log_e(B_{O1-MR}) = 58.887$	Class 4	$\log_e(B_{O1-MR}) = 2.3360$	Class 2
$\log_e(B_{O2-EX}) = 154.961$	Class 4	$\log_e(B_{O1-EX}) = 2.2530$	Class 2	$\log_e(B_{O1-EX}) = 1.877$	Class 2
$\log_e(B_{O2-O1}) = 959.192$	Class 4	$\log_e(B_{O1-O2}) = 4.0380$	Class 3	$\log_e(B_{O1-O2}) = 3.8940$	Class 3
$\log_e(B_{O2-O3}) = 4.580$	Class 3	$\log_e(B_{O1-O3}) = 6.5490$	Class 4	$\log_e(B_{O1-O3}) = 6.770$	Class 4

Table 6
Data A: Model ranks based on Bayes factors and maximum log-likelihood ratios.

	Data A		Remarks
	Bayes factor	Log-likelihood	
Rank 1	Ogden ($N = 2$)	Ogden ($N = 3$)	(a) Fits to data good for Ogden ($N = 2, 3$)
Rank 2	Ogden ($N = 3$)	Ogden ($N = 2$)	(b) The likelihood surface is flat for Ogden ($N = 3$),
Rank 3	Exponential	Exponential	indicating over-fitting
Rank 4	Mooney–Rivlin	Mooney–Rivlin	
Rank 5	Ogden ($N = 1$)	Ogden ($N = 1$)	

Table 7
Data B: Model ranks based on Bayes factors and maximum log-likelihood ratios.

	Data B		Remarks
	Bayes factor	Log-likelihood	
Rank 1	Ogden ($N = 1$)	Exponential	(a) Fits to data good for exponential, Ogden ($N = 1, 2, 3$)
Rank 2	Exponential	Ogden ($N = 3$)	(b) The likelihood surface is flat for Ogden ($N = 2, 3$)
Rank 3	Ogden ($N = 2$)	Ogden ($N = 2$)	(c) exponential: probability mass with high likelihood
Rank 4	Ogden ($N = 3$)	Ogden ($N = 1$)	is smaller than with Ogden ($N = 1$)
Rank 5	Mooney–Rivlin	Mooney–Rivlin	

Table 8
Data C: Model ranks based on Bayes factors and maximum log-likelihood ratios.

	Data C		Remarks
	Bayes factor	Log-likelihood	
Rank 1	Ogden ($N = 1$)	Ogden ($N = 3$)	(a) Fits to data good for Ogden ($N = 2, 3$)
Rank 2	Exponential	Ogden ($N = 2$)	(b) The likelihood surface is flat for Ogden ($N = 2, 3$)
Rank 3	Mooney–Rivlin	Mooney–Rivlin	(c) Under-fitting model preferred over over-fitting model
Rank 4	Ogden ($N = 2$)	Exponential	(d) The right model balancing goodness of fit and
Rank 5	Ogden ($N = 3$)	Ogden ($N = 1$)	model complexity not in the candidate set

The values of the marginal likelihood (evidence) for the various models do not change significantly with the prior, as seen from Table 4. For a given dataset, the small differences in the evidence values from the two priors do not change the evidence ratios and thus the Bayes factors (Table 5) and ranks remain unchanged.

5.2. Model selection: Four factors

There are four factors that should be considered in the Bayesian framework for model selection to find a parsimonious model that balances model complexity and goodness of fit:

1. The qualitative fit of the model (with the MLE parameters) to the data. The fit "quality" is generally subject to the analyst's judgment.
2. The maximum log-likelihood values. This is a quantitative measure of the goodness of fit. This is not an absolute measure but can be used to compare models.
3. The log-evidence values. This quantifies the likelihood of the data given the model, and a higher value of evidence means that the data are more likely to be predicted by the model. The ratio of evidences is the Bayes factor.
4. The landscape of the log-likelihood function. A flat log-likelihood function is undesirable because the model parameters are more difficult to identify uniquely.

Note that the likelihood and evidence values are quantitative factors for model selection, while the model fit and the landscape of the likelihood are qualitative in nature. The latter are particularly useful since the quantitative factors alone cannot guarantee a parsimonious model. Consider, for instance, a situation where one model underfits the data while a second model overfits the data. The latter could have a higher evidence value than the former, but the overfitting will not be obvious from this comparison alone. However, by examining the landscape of the likelihood function, it will become possible to infer this and consider an alternate (third) model that is parsimonious.

Ideally, a parsimonious model should have a good fit to data, high values of log-likelihood and log-evidence, and a log-likelihood function that is not flat. For some problems, it may not be possible to satisfy all of these requirements. In such cases, the Bayes factor can be used as a rational basis for making model selection decisions. We discuss these issues next.

The models used in this study can be grouped into two classes:

1. Models with the same functional form but increasing complexity with increasing number of terms. This is the case with the Ogden ($N = 1, 2, 3$) models, where a model of order N has $2N$ parameters.
2. Models with same complexity but different functional forms. For this we consider Mooney–Rivlin, exponential, and Ogden ($N = 1$), all of which have two parameters.

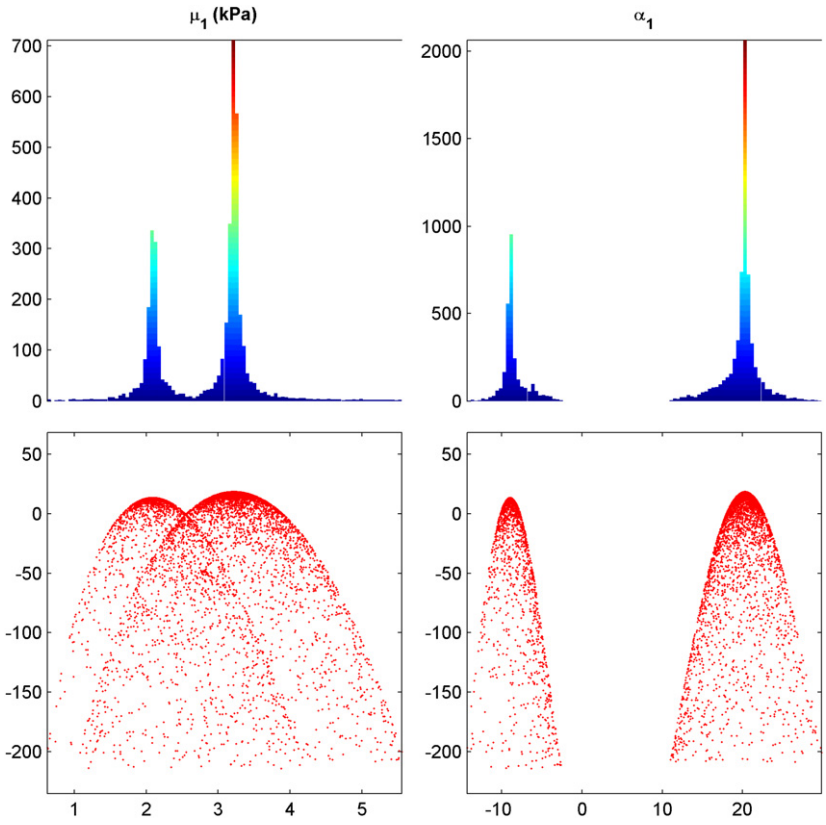

Fig. 7. Data B: Log-likelihood distributions for the Ogden ($N = 1$) model parameters.

5.2.1. Models with increasing complexity

For the cases considered here and in general, the maximum log-likelihood value increases with an increase in the complexity of the model. This can be seen for all the three datasets for Ogden ($N = 1, 2, 3$) in Table 3. But the log-likelihood does not address the question of "overfitting", i.e., is the model more complex (and thus has more parameters) than necessary. Addressing this issue is essential to selecting a parsimonious model. One way to avoid overfitting is through the use of terms that penalize overly complex models. Current approaches in the literature that use this technique assume large datasets and asymptotics [13], and are of limited use for the problem at hand. In contrast to the log-likelihood, the marginal likelihood or evidence is a natural Occam's razor as discussed in the Introduction. Put another way, the evidence value increases till sufficient complexity is reached and decreases beyond that.

For Data B, Ogden ($N = 1, 2, 3$) all provide good fits. However, the evidence value decreases from Ogden ($N = 1$) to Ogden ($N = 2$), and decreases further for Ogden ($N = 3$). Meanwhile the corresponding log-likelihood values increase with N. This indicates that Ogden ($N = 1$) satisfies the principle of parsimony for Data B, and the other two lead to overfitting. This behavior is clearly seen in the landscape of the log-likelihood functions for these models. In Fig. 7, the likelihood plots for Ogden ($N = 1$) parameters have distinct peaks, whereas the model parameters for Ogden ($N = 2, 3$) (seen in Figs. 8 and 9) have flat likelihoods. This means these models overfit the data, which makes parameter identification difficult, especially when a local optimization method is used to estimate them (rather than the global MULTINEST approach employed here).

For Data A, the Ogden ($N = 1$) model performs poorly, while Ogden ($N = 2, 3$) both provide good fits (see Fig. 4). The evidence value increases from Ogden ($N = 1$) to Ogden ($N = 2$) and then decreases for Ogden ($N = 3$), indicating that the latter overfits the data. As with Data B discussed above, the likelihood plots confirm the overfitting. These are however not shown here for the sake of brevity, but may be found in [56].

For Data C, Ogden ($N = 1$) does not provide a good fit, but Ogden ($N = 2, 3$) do. However, the evidence value for Ogden ($N = 1$) is higher than Ogden ($N = 2, 3$), which means that the latter models overfit the data. (This is

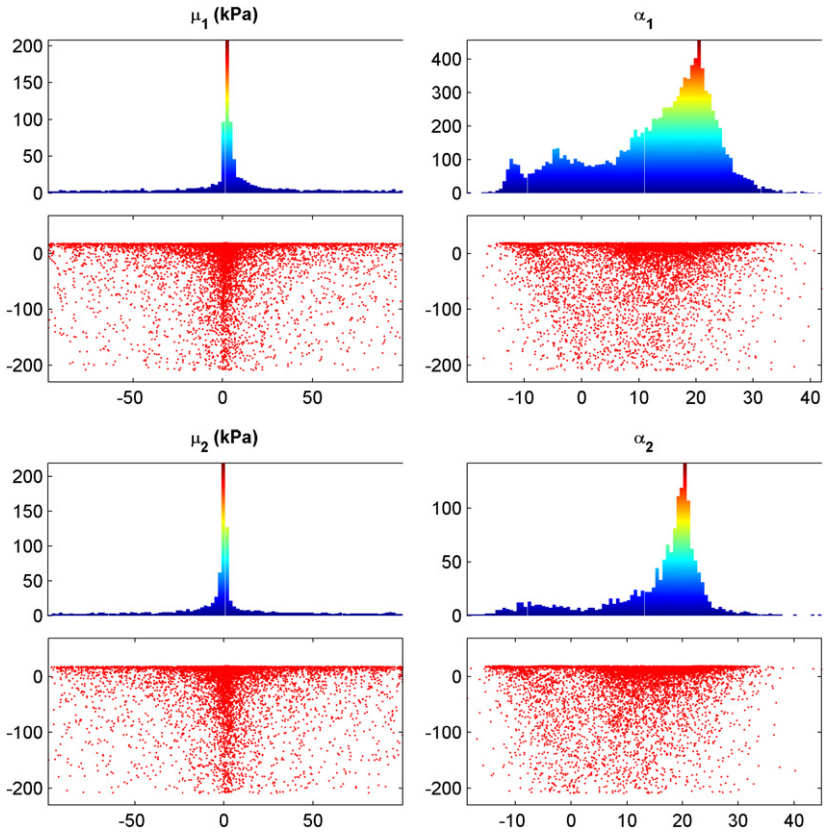

Fig. 8. Data B: Log-likelihood distributions for the Ogden ($N = 2$) model parameters.

Table 9
Maximum of the log-likelihood function and evidence with experimental variance one-sixth of the original for Data C.

Material model	Max log-likelihood	log-evidence
Mooney–Rivlin	77.979	59.409 ± 0.066
Exponential	58.528	40.776 ± 0.064
Ogden ($N = 1$)	24.733	9.843 ± 0.058
Ogden ($N = 2$)	92.068	69.093 ± 0.073
Ogden ($N = 3$)	92.085	66.589 ± 0.077

confirmed by the flat log-likelihood plots for Ogden ($N = 2, 3$) in [56].) Thus, there is no parsimonious model among the candidates; they either under-fit the data (as seen from the model fit in Fig. 5) or overfit the data as indicated by the decreasing evidence values and flat log-likelihoods. Indeed, for this particular dataset with the given levels of noise, the ideal model would lie (in complexity) between Ogden $N = 1$ and Ogden $N = 2$.

To illustrate the role of the noise on model selection and ranking, we consider the same dataset but with one-sixth the original experimental variance. The resulting values of maximum log-likelihood and evidence are shown in Table 9. This time, Ogden ($N = 2$) has all four properties described earlier, and emerges as the best among the candidates. The corresponding fits to the experimental data are shown in Fig. 10. The exponential, Mooney–Rivlin and Ogden ($N = 1$) models all have poor fits and low evidence values, while the more complex Ogden ($N = 3$) provides a good fit but shows a lower evidence value than the Ogden ($N = 2$).

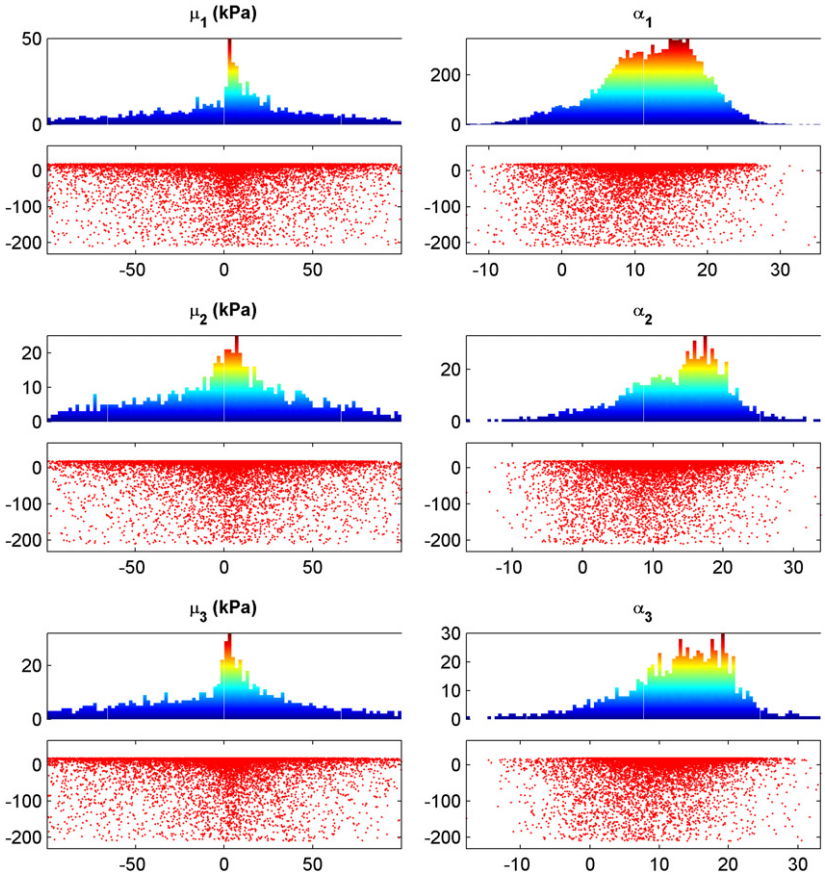

Fig. 9. Data B: Log-likelihood distributions for the Ogden ($N = 3$) model parameters.

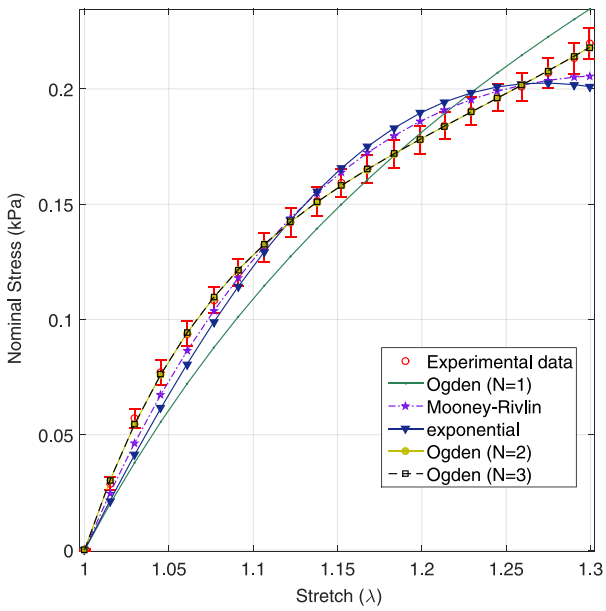

Fig. 10. Data C with reduced noise: Comparison of Mooney–Rivlin, exponential and Ogden models fits.

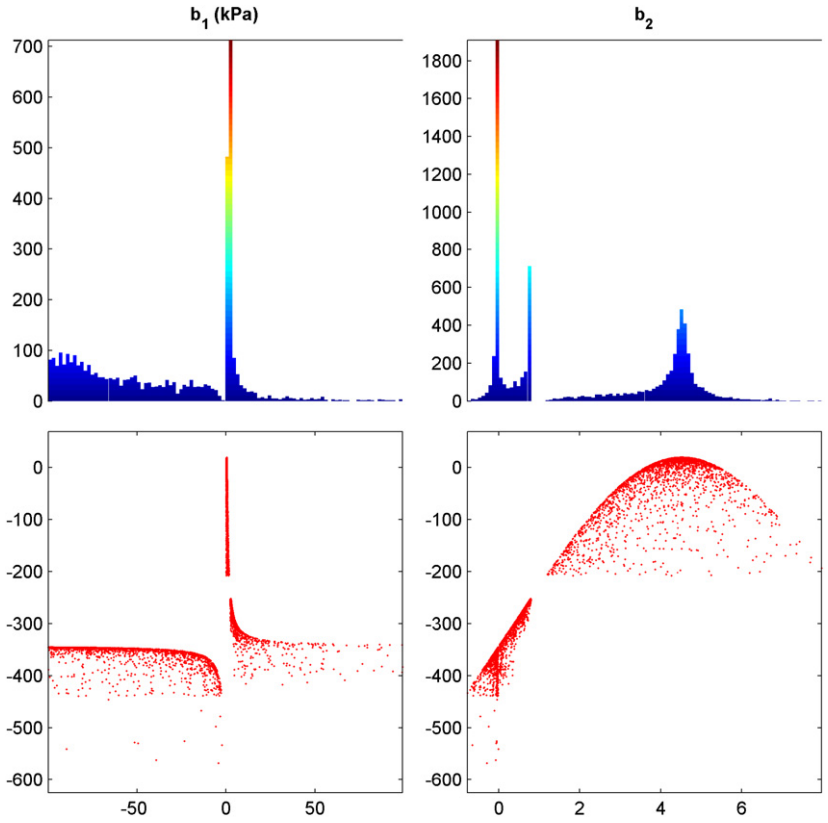

Fig. 11. Data B: Log-likelihood distributions for the exponential model parameters.

5.2.2. *Models with the same complexity*

For Data A, none of the candidate models (Mooney–Rivlin, exponential, and Ogden ($N = 1$)) provide a good fit to the data. This results in very low log-likelihood and evidence values—see Tables 3 and 4.

For Data B, the Mooney–Rivlin model provides a poor fit to the data with a low value of log-likelihood and is not considered further. For the other two models, the parameter distributions are shown in Figs. 7 and 11. Comparing these two, the exponential model has a higher log-likelihood value, while the Ogden ($N = 1$) model has a higher value of evidence. This is because the likelihood function for the exponential model is highly peaked near a small area around the MLE in the parameter space, but has low values in the rest of the parameter space. Therefore, its marginal likelihood, which is the integral of the probability mass in parameter space, is small. On the other hand, for Ogden ($N = 1$), the region of the log-likelihood function with appreciable log-likelihood values is larger, and leads to a higher marginal likelihood value. This shows that maximum likelihood ratio is by itself not an objective criterion for ranking or selecting models, even when the models are of the same complexity.

For data C, Mooney–Rivlin, exponential, and Ogden ($N = 1$) models all under-fit the data. Their maximum log-likelihoods are lower than the higher complexity models, but the evidence values are higher. This means the Bayesian approach based on evidence prefers lower complexity models over higher complexity models in the absence of an optimal parsimonious model.

5.3. *Model ranking*

As we move from class 1 to class 4 based on the Jeffreys scale, the investigator has more confidence in the chosen model over the alternatives. In this study, for Data A, Ogden ($N = 2$) is the best among the candidates with the odds of $\exp(4.580) = 97.5144$ compared to the next best model. It satisfies all the criteria described earlier: good fit, high likelihood and evidence values, and a non-flat likelihood landscape. This gives us high confidence in this model.

For Data B, the Ogden ($N = 1$) has an odds of $\exp(2.2530) = 9.5162$ over the exponential model. It too satisfies the criteria laid out earlier and is thus ranked the best among the candidate models.

However for Data C, the situation is more complex mainly because of the large amount of noise in the experimental data. The Ogden ($N = 1$) and exponential models under-fit the data, but their fits still lie within the error bars in Fig. 3. The Ogden ($N = 2, 3$) models represent the trend in the data more faithfully and capture the curvature accurately. However, they over-fit the data and have low evidence values, and are therefore ranked lower on the Jeffreys scale. For this dataset, none of the candidate models satisfies all the criteria laid out earlier.

6. Conclusions

In this work, we study the Bayesian approach to model selection and ranking among a set of five different models for the hyperelastic response of soft tissues. These include models of varying complexities: the Mooney–Rivlin model, the exponential model, and the Ogden model ($N = 1, 2, 3$). Three different sets of mechanical testing data are considered that describe the uniaxial deformation of agarose gel, bovine liver tissue, and porcine brain tissue.

Model selection and parameter estimation should not be considered as separate steps but should be considered as a single framework used to find a reliable model that can be used to predict the quantities of interest (QOI). We advocate the Bayesian approach to model selection using the Bayes factors because of its numerous advantages over the traditional methods along with its inherent capacity to handle the experimental noise. In the Bayesian framework we advocate the use of nested sampling based algorithm MULTINEST for its inherent global optimization and efficient evidence evaluation (for a detailed discussion on the global optimization of MULTINEST in the context of parameter estimation, refer to [56]).

Another important aspect is the sensitivity of the likelihood function to the choice of the prior (Bayesian robustness) to ascertain that posterior distribution is data dominated rather than prior dominated. In our study, we used two priors namely, uniform prior and gaussian prior and found that the values of maximum log-likelihood, MLE and log-evidence are insensitive to the prior increasing the confidence in our results.

In this study, four important factors are considered simultaneously to reliably choose a parsimonious model from the candidate set of models. These are the fit of the model (with the MLE parameters) to the data, log-evidence values, maximum log-likelihood values, and the landscape of the log-likelihood function. While the importance of the model fit is obvious, log-evidence values help pick a parsimonious model that balances goodness of fit and model complexity, while further insight into the overfitting of a model is obtained by looking at the landscape of the log-likelihood function. This is flat when the model overfits the data and peaked when not. Also maximum log-likelihood values give information about the goodness of fit. This information can be effectively used to determine whether a parsimonious model is present in the candidate model set (as in the case of Data A and Data B), or if none of the models in the set are parsimonious (as in the case of Data C).

Our results show that the Bayes factor is a more reliable criterion than the likelihood ratio for choosing between models and is a robust method to do model selection when the evidence integral evaluation and the sensitivity to the prior probability are properly addressed. The Bayes factors can be used to conveniently rank the models in the order of usability to provide further insight to the investigator. This method is general and can be used for any combination of data and constitutive model.

Acknowledgment

We are grateful to the University of Cincinnati Simulation Center for providing financial support for this work.

Appendix A. Supplementary data

Supplementary material related to this article can be found online at http://dx.doi.org/10.1016/j.cma.2015.03.012.

References

[1] E. Roan, K. Vemaganti, The nonlinear material properties of liver tissue determined from no-slip uniaxial compression experiments, J. Biomech. Eng. 129 (3) (2007) 450–456.

[2] A. Brunon, K. Bruyere-Garnier, M. Coret, Mechanical characterization of liver capsule through uniaxial quasi-static tensile tests until failure, J. Biomech. 43 (11) (2010) 2221–2227.

[3] C. Chui, E. Kobayashi, X. Chen, T. Hisada, I. Sakuma, Combined compression and elongation experiments and non-linear modelling of liver tissue for surgical simulation, Med. Biol. Eng. Comput. 42 (6) (2004) 787–798.
[4] C.-K. Chai, A.C. Akyildiz, L. Speelman, F.J. Gijsen, C.W. Oomens, M.R. van Sambeek, A. van der Lugt, F. Baaijens, Local axial compressive mechanical properties of human carotid atherosclerotic plaques—characterisation by indentation test and inverse finite element analysis, J. Biomech. 46 (10) (2013) 1759–1766.
[5] E. Samur, M. Sedef, C. Basdogan, L. Avtan, O. Duzgun, A robotic indenter for minimally invasive measurement and characterization of soft tissue response, Med. Image Anal. 11 (4) (2007) 361–373.
[6] M. Kauer, V. Vuskovic, J. Dual, G. Székely, M. Bajka, Inverse finite element characterization of soft tissues, Med. Image Anal. 6 (3) (2002) 275–287.
[7] A. Nava, E. Mazza, M. Furrer, P. Villiger, W. Reinhart, In vivo mechanical characterization of human liver, Med. Image Anal. 12 (2) (2008) 203–216.
[8] T. Beda, Modeling hyperelastic behavior of rubber: A novel invariant-based and a review of constitutive models, J. Polym. Sci. B 45 (13) (2007) 1713–1732.
[9] L. Hoss, R.J. Marczak, A new constitutive model for rubber-like materials, Mec. Comput. 29 (2010) 2759–2773.
[10] T. Bhattacharjee, M. Barlingay, H. Tasneem, E. Roan, K. Vemaganti, Cohesive zone modeling of mode i tearing in thin soft materials, J. Mech. Behav. Biomed. Mater. 28 (2013) 37–46.
[11] R.E. Kass, A.E. Raftery, Bayes factors, J. Amer. Statist. Assoc. 90 (430) (1995) 773–795.
[12] L. Wasserman, Bayesian model selection and model averaging, J. Math. Psychol. 44 (1) (2000) 92–107.
[13] K.P. Burnham, D.R. Anderson, Model Selection and Multimodel Inference: A Practical Information-Theoretic Approach, Springer, 2002.
[14] J.O. Berger, L.R. Pericchi, J. Ghosh, T. Samanta, F. De Santis, J. Berger, L. Pericchi, Objective Bayesian methods for model selection: introduction and comparison, in: Lecture Notes-Monograph Series, 2001, pp. 135–207.
[15] R.H. Berk, et al., Limiting behavior of posterior distributions when the model is incorrect, Ann. Math. Stat. 37 (1) (1966) 51–58.
[16] J. Dmochowski, Intrinsic priors via Kullback–Leibler geometry, Bayesian Stat. 5 (1996) 543–549.
[17] D.J.C. MacKay, Information Theory, Inference, and Learning Algorithms, Cambridge University Press, New York, 2003.
[18] R. Trotta, Bayes in the sky: Bayesian inference and model selection in cosmology, Contemp. Phys. 49 (2) (2008) 71–104.
[19] C. Robert, The Bayesian Choice: From Decision-Theoretic Foundations to Computational Implementation, Springer, 2007.
[20] J.O. Berger, J.M. Bernardo, On the development of reference priors, Bayesian Stat. 4 (4) (1992) 35–60.
[21] E.T. Jaynes, Probability Theory: The Logic of Science, Cambridge University Press, 2003.
[22] A.R. Syversveen, Noninformative Bayesian priors. Interpretation and problems with construction and applications, Statistics 3 (1998). Preprint.
[23] A.E. Gelfand, D.K. Dey, Bayesian model choice: asymptotics and exact calculations, J. R. Stat. Soc. Ser. B Stat. Methodol. (1994) 501–514.
[24] S. Geisser, W.F. Eddy, A predictive approach to model selection, J. Amer. Statist. Assoc. 74 (365) (1979) 153–160.
[25] M. Aitkin, Posterior bayes factors, J. R. Stat. Soc. Ser. B Stat. Methodol. (1991) 111–142.
[26] J.O. Berger, L.R. Pericchi, The intrinsic Bayes factor for model selection and prediction, J. Amer. Statist. Assoc. 91 (433) (1996) 109–122.
[27] A. O'Hagan, Fractional bayes factors for model comparison, J. R. Stat. Soc. Ser. B Stat. Methodol. (1995) 99–138.
[28] H. Haario, M. Laine, A. Mira, E. Saksman, Dram: efficient adaptive mcmc, Stat. Comput. 16 (4) (2006) 339–354.
[29] J.A. Vrugt, C. Ter Braak, C. Diks, B.A. Robinson, J.M. Hyman, D. Higdon, Accelerating markov chain Monte Carlo simulation by differential evolution with self-adaptive randomized subspace sampling, Int. J. Nonlinear Sci. Numer. Simul. 10 (3) (2009) 273–290.
[30] F. Feroz, M. Hobson, M. Bridges, Multinest: an efficient and robust Bayesian inference tool for cosmology and particle physics, Mon. Not. R. Astron. Soc. 398 (4) (2009) 1601–1614.
[31] A.F. Smith, Bayesian computational methods, Philos. Trans. R. Soc. Lond. Ser. A Math. Phys. Eng. Sci. 337 (1647) (1991) 369–386.
[32] F. Feroz, B.C. Allanach, M. Hobson, S.S. AbdusSalam, R. Trotta, A.M. Weber, Bayesian selection of sign μ within msugra in global fits including wmap5 results, J. High Energy Phys. 2008 (10) (2008) 064.
[33] J. Skilling, et al., Nested sampling for general Bayesian computation, Bayesian Anal. 1 (4) (2006) 833–859.
[34] E.E. Prudencio, P.T. Bauman, D. Faghihi, K. Ravi-Chandar, J.T. Oden, A computational framework for dynamic data driven material damage control, based on Bayesian inference and model selection, Internat. J. Numer. Methods Engrg. (2014) http://dx.doi.org/10.1002/nme.4669.
[35] E. Prudencio, S.H. Cheung, Parallel adaptive multilevel sampling algorithms for the Bayesian analysis of mathematical models, Int. J. Uncertain. Quantif. 2 (3) (2012).
[36] J.T. Oden, E.E. Prudencio, A. Hawkins-Daarud, Selection and assessment of phenomenological models of tumor growth, Math. Models Methods Appl. Sci. 23 (07) (2013) 1309–1338.
[37] J.L. Beck, K.-V. Yuen, Model selection using response measurements: Bayesian probabilistic approach, J. Eng. Mech. 130 (2) (2004) 192–203.
[38] J.L. Beck, Bayesian system identification based on probability logic, Struct. Control Health Monit. 17 (7) (2010) 825–847. http://dx.doi.org/10.1002/stc.424.
[39] K.-V. Yuen, Recent developments of Bayesian model class selection and applications in civil engineering, Struct. Saf. 32 (5) (2010) 338–346.
[40] A.H. Elsheikh, I. Hoteit, M.F. Wheeler, Efficient Bayesian inference of subsurface flow models using nested sampling and sparse polynomial chaos surrogates, Comput. Methods Appl. Mech. Engrg. 269 (2014) 515–537. http://dx.doi.org/10.1016/j.cma.2013.11.001.
[41] P. Hadjidoukas, P. Angelikopoulos, D. Rossinelli, D. Alexeev, C. Papadimitriou, P. Koumoutsakos, Bayesian uncertainty quantification and propagation for discrete element simulations of granular materials, Comput. Methods Appl. Mech. Engrg. 282 (2014) 218–238. http://dx.doi.org/10.1016/j.cma.2014.07.017.
[42] R. Sandhu, M. Khalil, A. Sarkar, D. Poirel, Bayesian model selection for nonlinear aeroelastic systems using wind-tunnel data, Comput. Methods Appl. Mech. Engrg. 282 (2014) 161–183. http://dx.doi.org/10.1016/j.cma.2014.06.013.
[43] S. Chib, I. Jeliazkov, Marginal likelihood from the Metropolis–Hastings output, J. Amer. Statist. Assoc. 96 (453) (2001) 270–281. http://dx.doi.org/10.1198/016214501750332848.

[44] J. Martin, C. Ringeval, R. Trotta, Hunting down the best model of inflation with Bayesian evidence, Phys. Rev. D 83 (6) (2011) 063524.
[45] N. Pullen, R.J. Morris, Bayesian model comparison and parameter inference in systems biology using nested sampling, PLoS One 9 (2) (2014) e88419.
[46] G.A. Holzapfel, Biomechanics of soft tissue, Handb. Mater. Behav. 3 (2001) 1049–1063.
[47] Y.-C. Fung, Biomechanics, Springer, 1990.
[48] D.S. Sivia, Data Analysis: A Bayesian Tutorial, Oxford University Press, 1996.
[49] W. Gropp, E. Lusk, A. Skjellum, Using MPI: Portable Parallel Programming with the Message-Passing Interface, vol. 1, MIT Press, 1999.
[50] N.S. Billade, Mechanical characterization, computational modeling and biological considerations for carbon nanomaterial–agarose composites for tissue engineering applications (Ph.D. thesis), University of Cincinnati, 2009.
[51] B. Rashid, M. Destrade, M.D. Gilchrist, Experimental characterisation of neural tissue at collision speeds, in: 2012 IRCOBI Conference Proceedings, 12–14 September 2012-Dublin (Ireland): IRC-12-49, International Research Council on the Biomechanics of Injury, 2012, pp. 405–416.
[52] E. Roan, K. Vemaganti, Strain rate-dependent viscohyperelastic constitutive modeling of bovine liver tissue, Med. Biol. Eng. 49 (4) (2011) 497–506.
[53] C. Chui, E. Kobayashi, X. Chen, T. Hisada, I. Sakuma, Transversely isotropic properties of porcine liver tissue: Experiments and constitutive modelling, Med. Biol. Eng. Comput. 45 (1) (2007) 99–106.
[54] X. Jin, F. Zhu, H. Mao, M. Shen, K.H. Yang, A comprehensive experimental study on material properties of human brain tissue, J. Biomech. 46 (16) (2013) 2795–2801.
[55] S. Nicolle, M. Lounis, R. Willinger, Shear properties of brain tissue over a frequency range relevant for automotive impact situations: New experimental results, Stapp Car Crash J. 48 (2004) 239–258.
[56] S. Madireddy, B. Sista, K. Vemaganti, Bayesian calibration of hyperelastic constitutive models of soft tissue, Int. J. Uncertain. Quantif. (2014). in review.