Bayesian inference of the spatial distributions of material properties

A. Vigliotti[a,b], G. Csányi[a] and V.S. Deshpande[a*]

[a]*Department of Engineering, University of Cambridge,
Trumpington Street, Cambridge CB2 1PZ, UK.*
[b]*Innovative Materials Laboratory, Italian Aerospace Research Centre, Capua 81043, Italy.*

Abstract
The inverse problem of estimating the spatial distributions of elastic material properties from noisy strain measurements is ill-posed. However, it is still typically treated as an optimisation problem to maximise a likelihood function that measures the agreement between the measured and theoretically predicted strains. Here we propose an alternative approach employing Bayesian inference with Nested Sampling used to explore parameter space and compute Bayesian evidence. This approach not only aids in identifying the basis function set (referred to here as a model) that best describes the spatial material property distribution but also allows us to estimate the uncertainty in the predictions. Increasingly complex models with more parameters generate very high likelihood solutions and thus are favoured by a maximum likelihood approach. However, these models give poor predictions of the material property distributions with a large associated uncertainty as they overfit the noisy data. On the other hand, the Bayes' factor peaks for a relatively simple model and indicates that this model is most appropriate even though its likelihood is comparatively low. Intriguingly, even for the appropriate model that has a unique maximum likelihood solution, the measurement noise is amplified to give large errors in the predictions of the maximum likelihood solution. By contrast, the mean of the posterior probability distribution reduces the effect of noise in the data and predicts the material properties with significantly higher fidelity. Simpler model selection criteria such as the Bayesian information criterion are shown to fail due to the non-Gaussian nature of the posterior distribution of the parameters. This makes accurate evaluation of the posterior distribution and the associated Bayesian evidence integral (by Nested Sampling or other means) imperative for this class of problems. The output of the Nested Sampling algorithm is also used to construct likelihood landscapes. These landscapes show the existence of multiple likelihood maxima when there is paucity of data and/or for overly complex models. They thus graphically illustrate the pitfalls in using optimisation methods to search for maximum likelihood solutions in such inverse problems.

Keywords: *Occam's razor; inverse elasticity problems; model selection; Nested Sampling*

*Corresponding author. E-mail address: vsd@eng.cam.ac.uk.

1. Introduction

Estimating the spatial distributions of mechanical properties in a heterogeneous solid body from the measurements of strains or displacements fields has wide ranging applications, including material characterisation, medical diagnosis and civil infrastructure monitoring. For example, there has been a recent burst of activity in the development of smart civil infrastructure and this includes application of strain measurement technologies like fibre optic sensing to monitor the structural health of tunnels (Gue et al., 2015), bridges (Ko and Ni, 2005) and concrete sleepers (Butler et al., 2017) to name a few. In these applications strains are measured at a small number of discrete locations. In some other applications full field measurements of the displacement fields within specimens are available. For example, a technique known as "displacements under applied loading by Magnetic Resonance Imaging" (dualMRI) has been developed to perform in vivo measurements of displacements and strain in musculoskeletal tissues (Chan et al, 2012). In all these cases the spatial distributions of material properties is the information of primary interest but displacement and strain information does not directly describe these material property distributions. The reconstruction of material property maps from noisy (and sometimes sparsely spaced) strain measurements is an ill-posed inverse problem that requires complex modelling approaches.

A number of methods have been proposed to identify constitutive parameters from strain/displacement measurements; readers are referred to Arvil et al. (2008) for an overview. Two commonly used approaches are: (i) the finite element model updating (FEMU) approach and (ii) the virtual field method (VFM). The former strategy is an optimization method and involves adjusting the parameters in order to minimize the difference between computed and measured strains as measured by a likelihood function (Rouger et al., 1990; Molimard et al., 2005). This approach can be used for either full field or discrete strain data. On the other hand, VFM (Grédiac, 1989) is a direct identification method that does not require any model iteration and hence is computationally less expensive but is best suited for full field data. Moreover, in practice, it requires a well-chosen mechanical test that excites of all strain components under the known boundary conditions. Examples of proposed specimens include a T-shaped specimen under tensile loading (Grédiac and Pierron, 1998) and thick laminated composite tubes under compression (Pierron et al. 2000). Nevertheless, there is a strong similarity between these two identification methods, at-least in the context of linear elasticity, as discussed by Avril and Pierron (2007).

The FEMU and VFM strategies discussed above are best suited for constitutive parameter identification problems in homogenous media. In order to use these methods to determine spatial property distributions (for example in heterogeneous materials such as the articular cartilage or a structurally deteriorating bridge structure), the unknown constitutive parameters need to be cast as weights of basis functions used to describe the spatial variations of the material properties. This is in fact what is done in the so-called "equilibrium gap" method and is equivalent to VFM with piecewise fields. The resulting inverse problem of identifying these weights is typically ill-posed with the outcome not only depending on the solution strategy, but also on the choice of the basis functions. Various types of regularizations have been proposed to reduce the intrinsic instability of these solutions, but the reconstructed results are inevitably strongly dependent on the choice of the regularization parameters (Richards et al., 2009). Moreover, the instabilities are aggravated by the presence of measurement noise.

Adjoint-weighted and gradient-based variational methods have recently been proposed in an attempt to stabilize the solutions in the presence of measurement noise; see for example Bal and Uhlmann (2013) and Bal et al. (2014). By contrast, Bellis (2017) proposed a reconstruction method based on an integral formulation of the linear elasticity problem whereby a given strain field is expressed as the solution of the Lippmann–Schwinger equation. This approach circumvents the underlying instability issues but is necessarily restricted to linear elasticity and elastic moduli distributions with a small contrast. By contrast, Nguyen et al (2015) proposed a multiscale statistical inverse method for performing the experimental identification of the elastic properties of materials at macro and mesoscales. Their method allows for identification of both the mean component and the statistical fluctuations of a stochastic model of the elasticity field of a heterogeneous microstructure using experimental data from a single specimen

Statistical and probabilistic methods are in fact being increasingly used for solutions of ill-posed inverse problems. In particular, the Bayesian approach allows for a full characterization of all possible solutions, and their relative probabilities, whilst simultaneously addressing the problems associated with the ill-posed problem in a clear and precise fashion. Readers are referred to Stuart (2010) and also the monographs by Kiapio et al. (2005) and Tarantola (2005) for a detailed discussion of the mathematical basis of Bayesian inference methods. Bayesian approaches, although computationally expensive to implement, are starting to lie within the range of the available computational resources especially given that they allow for the quantification of uncertainty in inverse problems. Most of current Bayesian approaches use Markov Chain Monte Carlo methods and/or filtering to identify maximum posterior solutions with the uncertainty quantified by evaluating the Hessian. Examples in solid mechanics include the works of Bui-Thanh et al. (2013) to derive the material property distributions via Monte Carlo sampling from seismic measurements and that of Thurin et al. (2017) who used Kalman filtering in a two-dimensional Marmousi model. The complete Bayesian calculation of evaluating the posterior distribution and computing the Bayesian evidence is avoided in these studies, presumably due to the numerical difficulties associated with such a task. Consequently, the fidelity of the implicit approximations in these Bayesian approaches remains unclear for problems in solid mechanics.

In this study follow a Bayesian approach to the inverse problem of determining spatial material property distributions from strain measurements. In particular we do not use Laplacian or other such approximations for the posterior as in the majority of previous studies. Rather we propose a method to estimate the entire posterior probability distribution of the material properties and thereby quantify the Bayesian evidence in support of particular choices of basis functions. The outline of the paper is as follows. We first present an overview of the Bayesian inference technique as applied to the inverse problem. Next, we describe the inverse elasticity problem and the Nested Sampling technique used to evaluate posterior probability distributions and Bayesian evidence for this elasticity problem. Finally, we discuss the computational results by comparing solutions based on Bayesian inference and maximal likelihood.
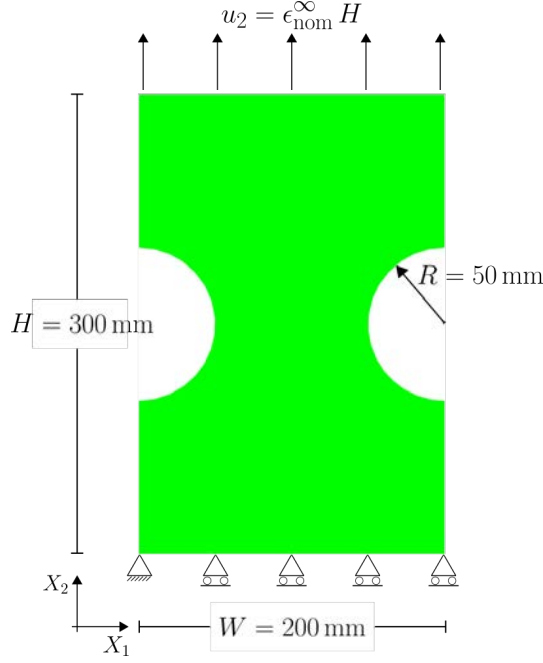
Figure 1: Sketch of the problem for predicting the distribution of elastic constants in a double-notched tensile specimen. The specimen is subjected to a remote tensile strain $\varepsilon_{\text{nom}}^{\infty}$ and the resulting strains within the specimen are measured at locations indicated in Fig. 2. The leading dimensions of the specimen are indicated along with the Cartesian co-ordinate system X_i.

2.   Inverse problems: optimisation versus Bayesian inference

The focus of the present study is the quantitative reconstruction of the elastic property maps that characterize a spatially heterogeneous, isotropic, linear elastic medium from measurements of strains within the domain. These strains may either be provided by the measurements of deformation at discrete locations within the domain via strain gauges (as might be the case for structural health monitoring applications) or obtained from strain maps furnished by full-field kinematic measurements (Grédiac and Hild, 2013) such as digital image correlation (DIC). We discuss the basic principle for the case of a limited number of discrete strain measurements with the understanding that full-field data is the limiting case of a large number of spatially distributed strain measurements. The reconstruction problem from DIC measurements, which often have larger noise, will be discussed in future studies.

Consider the representative problem shown in Fig. 1 of a double-notched specimen subjected to uniaxial tension in the X_2-direction. With x_i denoting the Cartesian co-ordinates of a point, the specimen has a spatial distribution of the Young's modulus $E(x_i)$ and Poisson's ratio $v(x_i)$ while the strains $\varepsilon_j^{(k)}$ are measured at locations $k = 1, \dots, N$ as indicated in Fig. 2. Here the subscript (j) denotes a component of the strain (e.g. $j = 0, 45$ and 90 denote the axial strains at orientations $\phi = 0°$, $45°$ and $90°$, respectively with respect to the X_1-axis). For a two-dimensional (2D) plane strain situation, three independent strain measurements fully specify the strain state at a given location and, in the problem, considered here we assume that three strain measurements in the $X_1 - X_2$ plane are made at each of the N locations within the specimen. These $3N$ strain measurements at a given applied remote strain $\varepsilon_{\text{nom}}^{\infty}$ are the available data $\mathcal{D}$ from which we aim to estimate $E(x_i)$ and $v(x_i)$. Full field data where displacements

4

are measured on a fine grid (the grid size is typically small but finite) corresponds to the case of a large N such that the spacing between the strain measurement locations in Fig. 2 is small.

2.1 *Estimation based on optimisation*

For clarity, we begin by briefly describing the generic optimisation problem as used for example in the FEMU approach. For a given guess of the property distributions $\hat{E}(x_i)$ and $\hat{v}(x_i)$ we can solve the boundary value problem sketched in Fig. 1 by any appropriate numerical method (e.g. by the finite element method (FEM)) and calculate the strains $\hat{\varepsilon}_j^{(k)}$ that correspond to the measurements $\varepsilon_j^{(k)}$. The aim of the optimisation is to choose the distributions $\hat{E}(x_i)$ and $\hat{v}(x_i)$ that minimise the differences between $\hat{\varepsilon}_j^{(k)}$ and $\varepsilon_j^{(k)}$: this choice is then thought of as the best representative of the true distributions $E(x_i)$ and $v(x_i)$. In mathematical terms this first involves defining a likelihood function $\mathcal{L}$, which reaches its maximum value when the difference between the measured and calculated strains is zero, and then optimising the distributions $\hat{E}(x_i)$ and $\hat{v}(x_i)$ so as to maximize $\mathcal{L}$. For example, with the standard deviation σ_ε representing the accuracy of the strain measurements, and the measurement $\varepsilon_j^{(k)}$ assumed to equal the mean from a large number independent measurements, the Gaussian distribution $\mathcal{N}(\varepsilon_j^{(k)}, \sigma_\varepsilon)$ is the maximum entropy distribution of the component j of the strain at location (k) (i.e. the Gaussian distributions makes the least claim to be informed beyond the available statistical information on the measurements). Then assuming that the accuracy of all the strain measurements are equal and that the errors are uncorrelated, an appropriate definition of the likelihood is a multivariate Gaussian given by

$$\mathcal{L} \equiv \prod_k \prod_j \frac{1}{\sigma_\varepsilon \sqrt{2\pi}} \exp\left[-\frac{1}{2}\left(\frac{\hat{\varepsilon}_j^{(k)} - \varepsilon_j^{(k)}}{\sigma_\varepsilon}\right)^2\right]. \quad (2.1)$$

This likelihood has a maximum value

$$\mathcal{L}_0 \equiv \prod_k \prod_j \frac{1}{\sigma_\varepsilon \sqrt{2\pi}}, \quad (2.2)$$

which is attained when all the estimated values $\hat{\varepsilon}_j^{(k)}$ are exactly equal to the measurements $\varepsilon_i^{(k)}$. Recalling that the calculated strains $\hat{\varepsilon}_j^{(k)}$ are functions of $\hat{E}(x_i)$ and $\hat{v}(x_i)$, the optimisation problem is then stated as

$$\max_{\hat{E},\hat{v}} \mathcal{L}, \quad (2.3)$$

where $\mathcal{L}(\hat{E},\hat{v})$ is calculated by solving a standard solid mechanics problem, i.e. given the geometry and the experimental boundary conditions, along with the trial material property distributions $\hat{E}(x_i)$ and $\hat{v}(x_i)$, we calculate $\hat{\varepsilon}_j^{(k)}$ that satisfy the usual equilibrium and compatibility constraints. Note that while the solution to the optimisation problem (2.3) is independent of σ_ε, it will be an important parameter for Bayesian inference as discussed in Section 4. We emphasize that throughout this study, the definition of the likelihood function is that given in (2.1).

In practical terms, the numerical solution of the above defined optimisation problem

involves devising a discrete representation of the continuous distributions $\hat{E}(x_i)$ and $\hat{v}(x_i)$ so as to reduce the optimisation problem to a finite number of variables. Of course, the choice of basis functions for this representation is not unique. For example, the values $\tilde{E}_r$ and $\tilde{v}_r$ at the $r = 1, \dots, M$ nodes of the FE mesh can be treated as the optimisation variables, with $\hat{E}(x_i)$ and $\hat{v}(x_i)$ then obtained from the finite element shape (interpolation) functions. However, even this definition of the shape functions is non-unique (e.g. the number and spatial distribution of nodes as well as the order of the shape functions needs to be chosen) and this adds an extra set of unknown parameters to the problem. We label each choice of the discrete representation as a separate model $\mathcal{H}_\ell(\boldsymbol{w})$ where $\boldsymbol{w} \equiv (\tilde{E}_r, \tilde{v}_r)$ denotes a vector of the $2M$ parameters of model $\mathcal{H}_\ell$. The optimisation procedure described above determines the set of parameters for the model that best fit the given data, but gives no guidance about the choice of the model. As demonstrated below in Section 4, simply choosing the model that yields the highest likelihood is misleading. This is because increasingly complex models with more parameters can always fit the data better and better. Thus, choosing the model with the highest likelihood will inevitably lead us to implausible, over-parameterized models. Therefore, another level of inference is necessary to choose an appropriate model.

2.2 The mechanism of the Bayesian razor

Two levels of inference can be distinguished in the process of data modeling. At the first level of inference, we assume that a particular model is true, and fit that model to the data, i.e., we infer what values its free parameters should plausibly take, given the data. This analysis is repeated for each model. The second level of inference is the task of model selection. Here we wish to compare the models in the light of the data, and assign some sort of preference or ranking to the alternatives. Bayesian methods are able consistently and quantitatively solve both the inference tasks.

We proceed to illustrate how Bayesian inference applies at each level of inference, where each model $\mathcal{H}_\ell$ has a vector $\boldsymbol{w}$ of parameters. Each model is defined by (i) its prior probability distribution $P(\boldsymbol{w}|\mathcal{H}_\ell)$ that states what values the model's parameters might be expected to take; and (ii) a set of conditional distributions, one for each value of $\boldsymbol{w}$, defining the prediction $P(\mathcal{D}|\boldsymbol{w}, \mathcal{H}_\ell)$ that the model makes about data $\mathcal{D}$, i.e. the likelihood of the data given the parameters $\boldsymbol{w}$ is $\mathcal{L} \equiv P(\mathcal{D}|\boldsymbol{w}, \mathcal{H}_\ell)$.

2.2.1 Model fitting

At the first level of inference we assume that model $\mathcal{H}_\ell$ is correct and the task is to infer its parameters $\boldsymbol{w}$ given the data $\mathcal{D}$. Using Bayes' theorem, the posterior probability of the parameters $\boldsymbol{w}$ is

$$P(\boldsymbol{w}|\mathcal{D}, \mathcal{H}_\ell) = \frac{P(\mathcal{D}|\boldsymbol{w}, \mathcal{H}_\ell)P(\boldsymbol{w}|\mathcal{H}_\ell)}{P(\mathcal{D}|\mathcal{H}_\ell)}. \qquad (2.4)$$

Then recalling that the sum of the posterior probability over the entire space of $\boldsymbol{w}$ equals unity, the normalizing constant $\mathcal{Z}_\ell \equiv P(\mathcal{D}|\mathcal{H}_\ell)$ follows as

$$\mathcal{Z}_\ell \equiv P(\mathcal{D}|\mathcal{H}_\ell) = \int P(\mathcal{D}|\boldsymbol{w}, \mathcal{H}_\ell)P(\boldsymbol{w}|\mathcal{H}_\ell)d\boldsymbol{w}. \quad (2.5)$$

In most Bayesian analyses, $\mathcal{Z}_\ell$ plays no essential role at this level of inference (and is typically not computed when performing the model fitting exercise). This is because it is common practice to find the maximum posterior which defines the most probable set of parameters $\boldsymbol{w}_{\mathrm{MP}}$ and $\mathcal{Z}_\ell$ has no direct dependence on $\boldsymbol{w}$ with respect to which the

maximisation is performed. In such cases the uncertainty in the prediction of $\boldsymbol{w}$ is estimated from the Hessian by assuming that the posterior is Gaussian centred on $\boldsymbol{w}_{\mathrm{MP}}$. We emphasize that the maximum of $P(\boldsymbol{w}|\mathcal{D}, \mathcal{H}_\ell)$ has no fundamental status in Bayesian inference and in fact we shall show that for the problem considered here, that the mean of the posterior is a more meaningful estimate of $\boldsymbol{w}$. Of course, determining the mean requires the entire posterior and $\mathcal{Z}_\ell$ to be computed even for this first level of inference.

Bayesian fitting can be contrasted to the usual optimisation approach where the best guess of $\boldsymbol{w}$ is simply taken as the set of parameters $\boldsymbol{w}_{\mathrm{OPT}}$[1] that maximise the likelihood $\mathcal{L} \equiv P(\mathcal{D}|\boldsymbol{w}, \mathcal{H}_\ell)$, independent of the prior. This approach not only fails to quantify the inherent uncertainty of the prediction, but also can result in a poor estimate of the model parameters especially in the presence of data noise; see Section 4.3. Of course, the maximum likelihood approach also gives no guidance on model selection as indicated above.

2.2.2   Model selection and Occam's razor
At the second level of inference, we wish to infer which model is most plausible given the data. If several models are compatible with a set of observations (i.e. several models give sufficiently high likelihoods), Occam's razor advises us to choose the simplest. This principle is typically advocated for one of two reasons. The first is aesthetic: "A theory with mathematical beauty is more likely to be correct than an ugly one that fits some experimental data" (Kragh (1990) quoting Paul Dirac). The second reason is the past empirical success of Occam's razor. However, there is an additional, different justification for Occam's razor, viz. coherent inference as embodied by Bayesian probability, which intrinsically includes Occam's razor in a *quantitative* manner.

We wish to compare the plausibility of two alternative models $\mathcal{H}_1$ and $\mathcal{H}_2$ in the light of data $\mathcal{D}$, i.e. compare $P(\mathcal{H}_1|\mathcal{D})$ and $P(\mathcal{H}_2|\mathcal{D})$. To do so we relate $P(\mathcal{H}_1|\mathcal{D})$ to the predictions made by the model about the data $P(\mathcal{D}|\mathcal{H}_1)$ via Bayes' theorem as

$$P(\mathcal{H}_1|\mathcal{D}) = \frac{P(\mathcal{D}|\mathcal{H}_1)P(\mathcal{H}_1)}{P(\mathcal{D})}, \qquad (2.6)$$

where the probability $P(\mathcal{H}_1)$ represents our prior belief in model $\mathcal{H}_1$ and the probability $P(\mathcal{D})$ of observing the data $\mathcal{D}$ is simply the sum over all alternative models $\mathcal{H}_\ell$, i.e. $P(\mathcal{D}) = \sum_\ell P(\mathcal{D}|\mathcal{H}_\ell)P(\mathcal{H}_\ell)$. (Note that $P(\mathcal{D}|\mathcal{H}_1)$ is defined in (2.5) and represents the model likelihood integrated over the entire parameter space.) By computing the ratio of their posteriors, we can therefore assess the relative merits of any two models $\mathcal{H}_1$ and $\mathcal{H}_2$ without reference to further alternatives,

$$\frac{P(\mathcal{H}_1|\mathcal{D})}{P(\mathcal{H}_2|\mathcal{D})} = \frac{P(\mathcal{H}_1)}{P(\mathcal{H}_2)} \frac{P(\mathcal{D}|\mathcal{H}_1)}{P(\mathcal{D}|\mathcal{H}_2)}. \qquad (2.7)$$

We now proceed to show how (2.7) embodies the principle of Occam's razor. To aid the discussion we assume, without loss of generality, that $\mathcal{H}_1$ is a simpler model compared to $\mathcal{H}_2$. The ratio $P(\mathcal{H}_1)/P(\mathcal{H}_2)$ gives us the opportunity to introduce a prior bias in favour of $\mathcal{H}_1$ on the basis of experience. This would correspond to the aesthetic or empirical motivations for Occam's razor. However, such a prior bias is not necessary as the second ratio, which is the data-dependent factor, inherently embodies Occam's razor. Simple models tend to make precise predictions while complex models, by their

---

[1] The maximum likelihood solution $\boldsymbol{w}_{\mathrm{OPT}}$ and maximum posterior solution $\boldsymbol{w}_{\mathrm{MP}}$ coincide for the special case of a uniform prior distribution.

nature, are more flexible and therefore capable of making a greater variety of predictions. Thus, because $\mathcal{H}_2$ is a more complex model, it will spread its predictive probability $P(\mathcal{D}|\mathcal{H}_2)$ more thinly over the data space than $\mathcal{H}_1$. Therefore, in the case where the data are compatible with both models, the simpler $\mathcal{H}_1$ will have a higher likelihood in (2.6) compared to $\mathcal{H}_2$, without us having to express any subjective dislike for complex models. Notice that while the probability $P(\mathcal{D}|\mathcal{H}_\ell)$ plays the role of likelihood in (2.6), it is simultaneously the denominator in (2.4) in the first level of inference. Because it embodies the concept of Occam's razor in terms of model comparison it is commonly referred to as the evidence $\mathcal{Z}_\ell \equiv P(\mathcal{D}|\mathcal{H}_\ell)$ of model $\mathcal{H}_\ell$ (MacKay, 2003).

From the above discussion it follows that if we choose to assign equal priors $P(\mathcal{H}_\ell)$ to all the alternative models, we can rank alternative models $\mathcal{H}_\ell$ by evaluating the evidence $\mathcal{Z}_\ell$. The idea is formalised by defining the Bayes' factor

$$\Delta \mathcal{K}_\ell \equiv 10 \log_{10}\left(\frac{\mathcal{Z}_\ell}{\mathcal{Z}_{\mathrm{ref}}}\right), \qquad (2.8)$$

that quantifies the support for model $\mathcal{H}_\ell$ over a reference model $\mathcal{H}_{\mathrm{ref}}$ (with evidence $\mathcal{Z}_{\mathrm{ref}}$). In (2.8), $\Delta \mathcal{K}_\ell$ is in units of decihartleys (dHart)[2] with $\Delta \mathcal{K}_\ell > 0$ meaning that $\mathcal{H}_\ell$ is more strongly supported by the data than $\mathcal{H}_{\mathrm{ref}}$. Decihartley is a particularly useful unit as a measure of information in Bayes' factor, with 10 decihartleys corresponding to odds of 10:1, 20 decihartleys to 100:1 odds etc. In fact, Good (1979) suggested that a change in $\Delta \mathcal{K}$ of 1 dHart (i.e. a change in the odds from evens to about 5:4) is about as finely as humans can reasonably be expected to quantify their degree of belief in a hypothesis. We note in passing that the minimum description length (MDL) principle (Rissanen, 1978) can always be interpreted as Bayesian model comparison with MDL having no apparent advantages over the direct probabilistic approach; see MacKay (2003).

Jeffreys (1961) gave a scale for the interpretation of $\Delta \mathcal{K}$ (Table 1). For example, $\Delta \mathcal{K}_\ell < -5$ suggests at-least substantial support for model $\mathcal{H}_{\mathrm{ref}}$ over $\mathcal{H}_\ell$ (note that the definition of the Bayes' factor is symmetric such that $\Delta \mathcal{K}_\ell > 5$ would imply at a minimum substantial support for model $\mathcal{H}_\ell$ over $\mathcal{H}_{\mathrm{ref}}$).

$\Delta \mathcal{K}_\ell$ (dHart)	Support of evidence for model $\mathcal{H}_{\mathrm{ref}}$ over $\mathcal{H}_\ell$
> 0	negative (supports $\mathcal{H}_\ell$ over $\mathcal{H}_{\mathrm{ref}}$)
0 to -5	barely worth mentioning
-5 to -10	substantial
-10 to -15	strong
-15 to -20	very strong
< -20	decisive

Table 1: Interpretation of the Bayes' factor as proposed by Jeffreys (1961).

---

[2] Hartley (symbol Hart and also called a ban or a dit) is a unit that measures information or entropy using base 10 logarithms rather than base 2 logarithms that define the Shannon (symbol Sh) or bit.
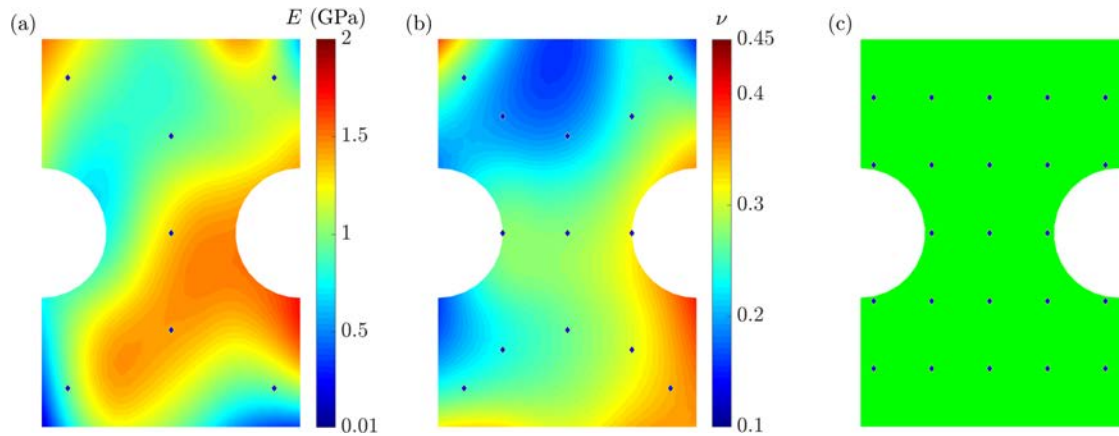
Figure 2: The location of strain gauge rosettes in synthetic experiments with (a) $N = 7$, (b) $N = 13$ and (c) $N = 23$ strain gauge rosettes. The rosettes measure the strains at orientations $\phi = 0^o$, 45° and 90°, respectively with respect to the X_1-axis. The actual spatial distributions of the Young's modulus E and Poisson's ratio ν within the specimen are indicated in (a) and (b), respectively.

3.   Estimation of spatial distributions of material properties

As an illustration of the Bayesian inference methodology for prediction of the spatial distribution of material properties given strain data, here we consider a double-notched tensile specimen with spatially varying isotropic elastic constants. The specimen of height $H = 300$ mm and width $W = 200$ mm has two semi-circular notches of radius $R = 50$ mm at mid-height as shown in Fig. 1. The specimen is subjected to a nominal remote tensile strain $\varepsilon_{nom}^{\infty}$ and strains $\varepsilon_0^{(k)}, \varepsilon_{45}^{(k)}, \varepsilon_{90}^{(k)}$ are measured at N locations $k = 1, ..., N$ via strain gauge rosettes. The locations of symmetrically and regularly positioned strain gauges are indicated in Fig. 2 for $N = 7, 13$ and 23.

3.1   Problem description

Since the aim of this study is to investigate the fidelity of the Bayesian inference technique for the prediction of material property distributions, we use numerically generated "synthetic experimental data". This enables us to compare the inferred properties with the known true distributions and thereby make judgements with regards the predictive capabilities of the Bayesian approach. We thus first explain the procedure for numerically generating the required input data.

Spatial distributions of E and ν in the $X_1 - X_2$ plane were generated via 2D non-uniform rational B-spline (NURBS) functions (Piegl and Tiller, 1997), i.e. no variations of properties in the X_3-direction. The NURBS are specified by B-splines of order q and a grid of $p \times p$ control weights over the $W \times H$ rectangular domain that circumscribes the specimen. Open uniform knot vectors with $q + 1$ equal knot values at each end were used in both the X_1 and X_2 directions. The "synthetic experimental specimen" was generated using quadratic ($q = 2$) B-splines with a $p = 5$ control weight grid, i.e. 25 control weights $\widetilde{E}$ and 25 control weights $\widetilde{\nu}$ for the distributions of E and ν, respectively. To generate the experimental specimen, the control weights were assumed to be random variables such that $\widetilde{E} \sim \mathcal{U}(0.01, 2.0)$ GPa and $\widetilde{\nu} \sim \mathcal{U}(0.1, 0.45)$, respectively where $\mathcal{U}(a, b)$ denotes a uniform distribution with support $[a, b]$. One

9

such randomly generated distribution was chosen to be the synthetic experimental specimen and the distributions of E and v within this experimental specimen are shown in Figs. 2a and 2b, respectively (Young's modulus distributions in a range of specimens generated with $p = 3, 5$ and 8 are shown in Supplementary Information Fig. S1 to give a sense of the wavelengths associated with different control weight grids). The synthetic experimental measurements of the strains $\varepsilon_j^{(k)}$ were then generated via a small-strain, plane strain FE analysis. The specimen was discretised into 1015 constant strain triangles of side ~ 1.1 mm and the elastic constants (E, v) of each element were set equal to the value obtained by interrogating the NURBS functions at the centroid of triangles. The FE analysis was then conducted by subjecting the specimen to displacements $u_2 = \varepsilon_{\text{nom}}^\infty H$ on the top surface and $u_2 = 0$ on the bottom surface with the tractions $T_1 = 0$ on both these surfaces; see Fig. 1. The experimental measurements of the strains $\varepsilon_j^{(k)}$ at the locations indicated in Fig. 2 were then directly ascertained from the FE calculations. The linearity of the problem described here implies that the strains $\varepsilon_j^{(k)}$ scale linearly with $\varepsilon_{\text{nom}}^\infty$. Therefore, without loss of generality, we present numerical simulations for the choice $\varepsilon_{\text{nom}}^\infty = 0.1$.

In order to better mimic real experimental data we included some measurement noise/error in the values of $\varepsilon_j^{(k)}$ generated via the FE calculations. This noise was added to $\varepsilon_j^{(k)}$ in the form of independent Gaussian noise with a zero mean and a standard deviation σ_{err}. We consider three cases: (i) no noise with $\sigma_{\text{err}} = 0$; (ii) low noise with $\sigma_{\text{err}} = 0.016\varepsilon_{\text{nom}}^\infty$ and (iii) high noise with $\sigma_{\text{err}} = 0.05\varepsilon_{\text{nom}}^\infty$. The values of $\varepsilon_j^{(k)}$ with this superimposed noise and the applied strain $\varepsilon_{\text{nom}}^\infty$ were then treated as the only available experimental data to be used in the inference of the spatial material property distributions. We emphasize that since the added noise is random, it is conceivable that there exists no spatial distribution of (E, v) that yields a solution of the elastic boundary value problem that exactly matches the strain data $\varepsilon_j^{(k)}$.

3.2 *Nested Sampling for phase space exploration and evaluation of evidence*

The aim here is to predict the spatial distribution of the elastic constants shown in Figs. 2a and 2b given the strains $\varepsilon_j^{(k)}$, with the specimen subjected to a remote strain $\varepsilon_{\text{nom}}^\infty = 0.1$. As discussed in Section 2.1, the inverse problem is made numerically tractable by choosing a discrete representation for the estimated distribution $(\hat{E}, \hat{v})$: this reduces the inverse problem to a finite number of variables. Here we choose a discrete representation based on the NURBS functions described in Section 3.1 and vary the number of control weights: a representation with a $p \times p$ control weight grid and therefore $2p^2$ parameters $\boldsymbol{w} \equiv (\widetilde{\boldsymbol{E}}, \widetilde{\boldsymbol{v}})$ is labelled as model $\mathcal{H}_p$. Thus, in model $\mathcal{H}_p$ the p^2 weights $\widetilde{\boldsymbol{E}}$ give the estimated spatial distribution $\hat{E}(x_i)$ while the p^2 weights $\widetilde{\boldsymbol{v}}$ give the estimated spatial distribution $\hat{v}(x_i)$. The models are nested in the sense that model $\mathcal{H}_{p'}$ produces a set of distributions that is a subset of the distributions of model $\mathcal{H}_p$ where $p > p'$. Non-nested models where we keep p fixed but vary the order q of the B-spline functions are discussed in Appendix A.

Calculation of the posterior probability distribution and evaluation of the evidence $\mathcal{Z}_p$ of model $\mathcal{H}_p$ in this $2p^2$ dimensional space represents a significant numerical challenge. Nested Sampling is an algorithm specifically designed to sample high

dimensional spaces for systems where the bulk of the high likelihood mass is contained in an exponentially small fraction of the parameter phase space. For example, Baldock et al. (2016) used it to sample the phase space of metals and construct pressure-temperature phase diagrams with the likelihood given by the Boltzmann factor. The algorithm outputs a set of samples and associated weights from which we can not only estimate the evidence and calculate the posterior probability but also derive information on the structure of the phase space. Here we briefly describe the Nested Sampling algorithm; readers are referred to Skilling (2006) for further details of the theoretical basis and to Pártay et al. (2010) and Pártay et al. (2014) for the numerical methods used. We begin by defining the prior mass $X(\lambda)$ of model $\mathcal{H}_p$ with likelihood $\mathcal{L} > \lambda$ as

$$X(\lambda) = \int_{\mathcal{L}(w)>\lambda} P(w|\mathcal{H}_p)dw, \qquad (3.1)$$

where the subscript p is omitted on X for the sake of brevity of notation. In the case of using a uniform probability distribution for the prior $P(w|\mathcal{H}_p)$, $X(\lambda)$ can be thought of as the fraction of parameter space corresponding to a likelihood $\mathcal{L} > \lambda$. Then noting that $dX \equiv P(w|\mathcal{H}_p)dw$, the evidence follows as

$$\mathcal{Z}_p = \int P(\mathcal{D}|w, \mathcal{H}_p)P(w|\mathcal{H}_p)dw = \int_0^1 \mathcal{L}(X)dX. \qquad (3.2)$$

The Nested Sampling algorithm is an iterative procedure which generates a sequence of likelihood levels $\mathcal{L}_1 < \mathcal{L}_2 < \mathcal{L}_3 \ldots$ where for each iteration s, $\mathcal{L}_s$ is chosen such that

$$\frac{\int_w \mathbb{I}[\mathcal{L}(w) > \mathcal{L}_s]dw}{\int_w \mathbb{I}[\mathcal{L}(w) > \mathcal{L}_{s-1}]dw} \approx \alpha, \qquad (3.3)$$

with $0 < \alpha < 1$ being a fixed fraction and $\mathbb{I}$ the indicator function. Hence the algorithm takes equidistant steps in the logarithm of the phase space volume, with $\Delta X_s = \omega_s = \alpha^{s-1} - \alpha^s$ being the portion of the prior mass corresponding to distributions of material properties with likelihood between $\mathcal{L}_{s-1}$ and $\mathcal{L}_s$. Numerical integration of (3.2) then gives

$$\mathcal{Z}_p \approx \sum_s \omega_s \mathcal{L}_s. \qquad (3.4)$$

The sequence of likelihood levels $\mathcal{L}_s$ is calculated by maintaining an active set of K samples, uniformly distributed over the set of configurations[3] with likelihood above the current level. The set is initialised with samples uniformly distributed throughout the whole phase space. At the first iteration the lowest likelihood configuration in the active set is chosen to assign the first likelihood level $\mathcal{L}_1$ and the corresponding configuration, Ω_1, is removed from the active set. Now a single new configuration is required to replace the one that was removed. The key to the algorithm is to perform this replacement while ensuring that the samples in the active set are kept uniformly distributed over the phase space with likelihood $\mathcal{L} > \mathcal{L}_1$. The original algorithm does not prescribe a specific method to perform this replacement and here we use the following Monte Carlo scheme. First, a randomly selected member of the remaining $K-1$ samples in the active is cloned. Next, this member is used to initialise a Markov chain with equilibrium distribution given by $P(\Omega) \propto \mathbb{I}[\mathcal{L}(\Omega) > \mathcal{L}_1]$. Finally, the end

---

[3] Each possible spatial distribution of the elastic constants within the specimen is referred to as a configuration.

configuration of the Markov chain is used to replenish the active set. The next likelihood level $\mathcal{L}_2$ is then taken to be given by lowest likelihood configuration Ω_2 within the replenished active set and the procedure repeats, generating a sequence of configurations with increasing likelihood $\mathcal{L}_3, \mathcal{L}_4, \dots$. Additional details on the numerical procedure are given in Appendix B with the numerical codes used in this study available in Vigliotti and Deshpande (2018).
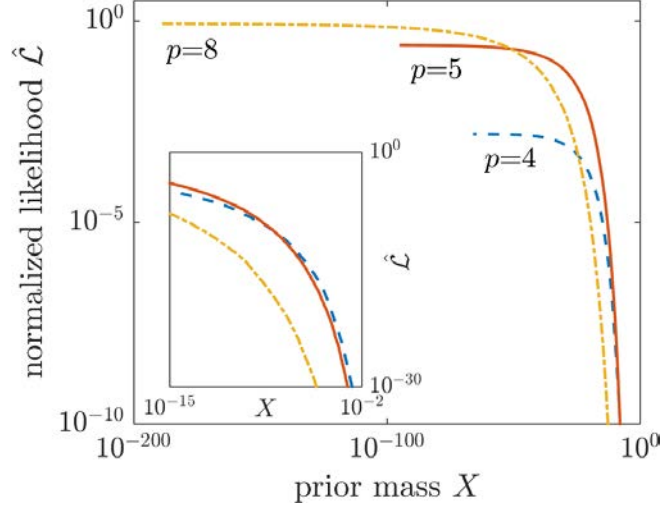
Figure 3: Predictions of the variation in the normalised likelihood $\hat{\mathcal{L}}$ with the prior mass X for models $\mathcal{H}_4$, $\mathcal{H}_5$ and $\mathcal{H}_8$ used to analyse the reference case of the experiment with $N = 23$ strain gauges and experimental noise $\sigma_{\mathrm{err}} = 0.016\varepsilon_{\mathrm{nom}}^{\infty}$. A zoom-in of the variation of $\hat{\mathcal{L}}$ at relatively high values of X is shown in the inset.

At each iteration s, the fraction of the configuration space with likelihood higher than $\mathcal{L}_s$ is proportional to the Beta distribution $\mathcal{B}e(K + 1,1)$ (Skilling, 2006) and therefore α is approximately the mean of this distribution, i.e. $\alpha \approx K/(K + 1)$. Thus, after iteration s,

$$X_s = \alpha^s = \left[\frac{K}{K + 1}\right]^s, \qquad (3.5)$$

of the initial unit prior mass remains to be explored. The Nested Sampling algorithm does not prescribe a specific terminating condition, other than running until the estimators for the observables of interest have satisfactorily converged. Here we terminate the simulation at iteration J when

$$\frac{2(\mathcal{L}_{K,J}^{\mathrm{max}} - \mathcal{L}_{K,J}^{\mathrm{min}})}{\mathcal{L}_{K,J}^{\mathrm{max}} + \mathcal{L}_{K,J}^{\mathrm{min}}} < \epsilon, \qquad (3.6)$$

where $\mathcal{L}_{K,J}^{\mathrm{min}}$ and $\mathcal{L}_{K,J}^{\mathrm{max}}$ are the minimum and maximum values of the likelihoods of the K samples in the active set at iteration $s = J$ and ϵ a user defined tolerance. The likelihood of the sample removed at $s = J$ is labelled $\mathcal{L}_{\mathrm{end}}$. We note in passing that given an appropriately small choice of ϵ, $\mathcal{L}_{\mathrm{end}}$ is a good approximation of the maximum likelihood that can be attained by the model $\mathcal{H}_p$ for the given dataset. Moreover, as discussed by Skilling (2006) the sequence $\{\Omega_1, \Omega_2, \dots\}$ gives a set of posterior representatives such that the posterior probability of the parameter set $\boldsymbol{w}_s$ of the sample Ω_s removed in iteration s is

$$P(\boldsymbol{w}_s|\mathcal{D},\mathcal{H}_p) = \frac{\omega_s \mathcal{L}_s}{\mathcal{Z}_p}. \qquad (3.7)$$

All calculations presented in this study were performed with a $K = 256$ active set and a tolerance $\epsilon = 10^{-3}$. Convergence tests confirmed that increasing K and decreasing ϵ resulted in no appreciable change in the estimation of the evidence and posterior distributions. Typical calculations reported here required between $J = 40 \times 10^3$ and 110×10^3 iterations with a Markov chain of $\mathcal{T} = 512$ steps was used to generate the sample to replenish the active set. Therefore, the evidence calculation for each model $\mathcal{H}_p$ involved between $J\mathcal{T} \approx 20 \times 10^6$ and 56×10^6 likelihood evaluations, where each likelihood evaluation required a FE calculation to determine the strains $\hat{\varepsilon}_i^{(k)}$ for the assumed spatial distribution of elastic constants $(\hat{E}, \hat{v})$. The priors in all the calculations were chosen to be the same as those used in generating the elastic constant distributions within the synthetic experimental specimen. Thus, the control weights for the Young's modulus for each step in the Markov chain were $\widetilde{\boldsymbol{E}} \sim \mathcal{U}(E_{\min}, E_{\max})$, while the weights for the Poisson's ratio $\tilde{\boldsymbol{v}} \sim \mathcal{U}(v_{\min}, v_{\max})$, where $(E_{\min}, E_{\max}) = (0.01, 2.0)$ GPa and $(v_{\min}, v_{\max}) = (0.1, 0.45)$. Unless otherwise specified, calculations presented here employ a standard deviation of the measurements $\sigma_\varepsilon = 0.005 = 0.05\varepsilon_{\mathrm{nom}}^{\infty}$ to evaluate the likelihood via (2.1). The parameter σ_ε characterises the error in the data[4] and is typically not known precisely but rather estimated from considerations of the accuracy of the experimental measurements. For example, if strains are measured using strain gauges, we expect a $\pm 5\%$ scatter in the gauge factor that is usually not corrected for. In this case, an appropriate choice for σ_ε is therefore 5% of a representative strain such as the applied nominal strain $\varepsilon_{\mathrm{nom}}^{\infty}$.

4.   Predictions based on Bayesian inference

Unless otherwise specified, all results are presented for the reference case of the synthetic experiment with $N = 23$ strain gauges and an experimental noise $\sigma_{\mathrm{err}} = 0.016\varepsilon_{\mathrm{nom}}^{\infty}$. Further, since the elastic constant distributions in the specimen were generated with a 5×5 grid of control weights, we shall use model $\mathcal{H}_5$ as the reference model $\mathcal{H}_{\mathrm{ref}}$ in calculating the Bayes' factor (2.8), i.e. if Bayesian inference is successful, the Bayes' factor will peak for model $\mathcal{H}_5$ with $\Delta\mathcal{K} = 0$. Recall that Bayesian inference is performed on two levels: (i) model fitting to determine the posterior parameter distributions and (ii) using these distributions to evaluate evidence for model selection. However, it is more convenient to discuss the results in the reverse order, i.e. first discuss the selection of the appropriate model and then present the posterior predictions of that model. We proceed here in this manner.

---

[4] We emphasize that the parameter σ_{err} characterizing the experimental noise is an artefact of the synthetic experimental data used here and will not be present when dealing with real experimental data. On the other hand, σ_ε used in the likelihood function is our estimate of the error in the experimental data and is required for the Bayesian inference analysis.
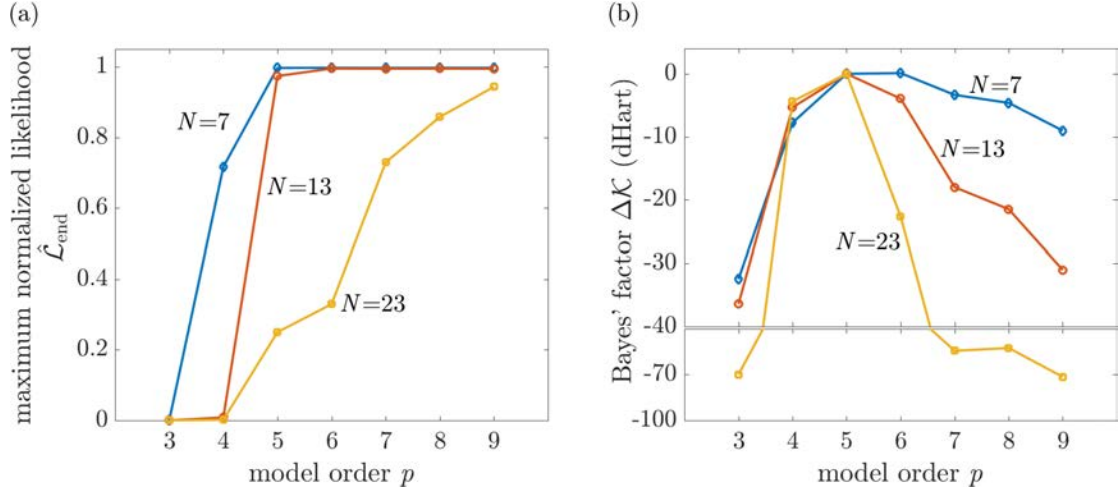
Figure 4: Predictions of the (a) maximum normalised likelihood $\hat{\mathcal{L}}_{end}$ and (b) Bayes' factor $\Delta\mathcal{K}$ as a function of the model order p. The predictions are for the case of synthetic experiments with a measurement noise of $\sigma_{err} = 0.016\varepsilon_{nom}^{\infty}$ and $N = 7, 13$ and 23 strain gauge locations. The Bayes' factor is calculated using model $\mathcal{H}_5$ as the reference.

4.1 Bayesian evidence versus maximum likelihood model selection

Predictions of the variation of the normalised likelihood $\hat{\mathcal{L}} \equiv \mathcal{L}/\mathcal{L}_0$ with the prior mass X are included in Fig. 3 for models $\mathcal{H}_4$, $\mathcal{H}_5$ and $\mathcal{H}_8$ (i.e. models with a $p \times p$ control weight grid and $p = 4, 5$ and 8, respectively). The normalisation of $\mathcal{L}$ is chosen such that $\hat{\mathcal{L}} \leq 1$ and Fig. 3 is interpreted as follows: for any given $\hat{\mathcal{L}}^*$ the corresponding value $X = X^*$ is the prior mass with likelihood higher than $\hat{\mathcal{L}}^*$. At high X, $\hat{\mathcal{L}}$ is the highest for model $\mathcal{H}_4$ but with decreasing X, the more complex models (i.e. $\mathcal{H}_5$ and $\mathcal{H}_8$) show a continuing increase in likelihood while the likelihood of model $\mathcal{H}_4$ begins to plateau. In fact, the likelihood of model $\mathcal{H}_5$ overtakes model $\mathcal{H}_4$ at $X \approx 1.0 \times 10^{-8}$ (see inset of Fig. 3) and similarly the likelihood of model $\mathcal{H}_8$ overtakes model $\mathcal{H}_5$ at $X \approx 2.0 \times 10^{-51}$. These results are understood by recognising that a significantly larger number of configurations are available for complex models compared to simpler models, but a larger fraction of the configurations of the complex models result in low likelihoods, i.e. $\hat{\mathcal{L}}$ remains relatively low in the complex models at high X. However, the greater flexibility offered by complex models such as $\mathcal{H}_8$ implies that a small fraction of the available configurations attain higher likelihoods than any configuration that a simpler model can generate. As a consequence $\hat{\mathcal{L}}_{end}$ increases with increasing number of control weights in the model, i.e. $\hat{\mathcal{L}}_{end}$ increases in the order $\mathcal{H}_4 \rightarrow \mathcal{H}_5 \rightarrow \mathcal{H}_8$.

Predictions of $\hat{\mathcal{L}}_{end}$ as a function of the model order as parameterised by p (recall that model $\mathcal{H}_p$ has $2p^2$ degrees of freedom) are included in Fig. 4a (here we focus on the $N = 23$ case and subsequently in Section 4.4 elaborate on the effect of the quantity of data). Clearly, $\hat{\mathcal{L}}_{end}$ increases with p suggesting that a likelihood measure favours the more complex models. However, it is clear that this inference is misleading: in the case of $N = 23$, i.e. the densest grid of gauges, $\hat{\mathcal{L}}_{end}$ for model $\mathcal{H}_5$ is a nearly a factor of four lower than for model $\mathcal{H}_8$ even though the underlying synthetic experimental data was generated using NURBS with $p = 5$. This discrepancy arises because the input data used included a 1.6% measurement noise. This results in the "correct" model $\mathcal{H}_5$ being unable to closely fit the measured strain values. On the other hand, the additional

flexibility of model $\mathcal{H}_8$ (i.e. its capability of generating spatial variations in $\hat{E}$ and $\hat{v}$ with a smaller wavelength), allows it to achieve a higher likelihood by fitting to the noise in the data. We shall subsequently show that this overfitting results in poor predictions of the spatial distributions of the elastic constants.
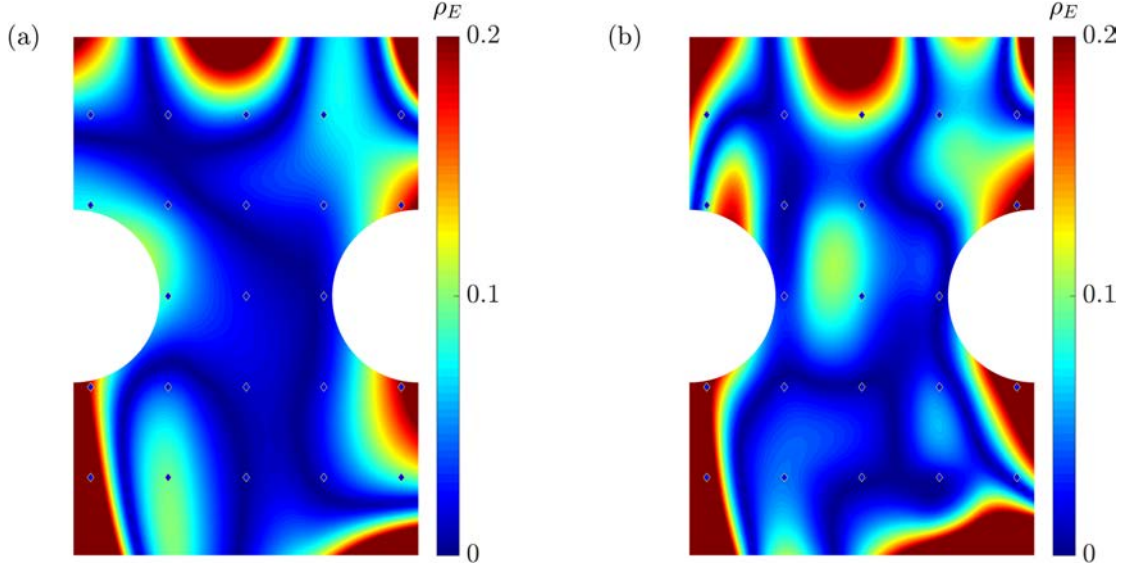
Figure 5: Spatial distributions of the error ρ_E in predictions of the Young's modulus using the mean of the posterior distribution for (a) model $\mathcal{H}_5$ and (b) model $\mathcal{H}_8$. The predictions are for the reference case of the synthetic experiment with $N = 23$ strain gauges and measurement noise $\sigma_{\mathrm{err}} = 0.016\varepsilon_{\mathrm{nom}}^{\infty}$. The strain measurement locations are indicated in both (a) and (b).

The corresponding estimates of the Bayes' factor $\Delta\mathcal{K}$ are in Fig. 4b and show that it is maximal for model $\mathcal{H}_5$ (recall that the model $\mathcal{H}_5$ is set as the reference $\mathcal{H}_{\mathrm{ref}}$ and hence $\Delta\mathcal{K} = 0$ for $p = 5$). Thus, Bayes' factor peaks for model $\mathcal{H}_5$ even though, as seen in Fig. 4a, the maximum likelihood of this model is relatively low. The next highest value of the Bayes' factor is $\Delta\mathcal{K} \approx -5$ for model $\mathcal{H}_4$: the interpretation in Table 1 therefore indicates substantial evidence in support of model $\mathcal{H}_5$ over $\mathcal{H}_4$. Moreover, since $\Delta\mathcal{K} < -20$ for all other models there is decisive evidence for $\mathcal{H}_5$ over these models. Thus, Bayesian inference correctly predicts that the underlying spatial distribution of elastic constants is best represented by NURBS with $p = 5$ even though the likelihood $\hat{\mathcal{L}}_{\mathrm{end}}$ of this model is relatively low.

4.2   Predictions of the spatial distributions of material properties
While Bayesian evidence correctly predicts that model $\mathcal{H}_5$ is the appropriate model to estimate the material property distributions, it remains to verify the accuracy with which model $\mathcal{H}_5$ predicts the distributions of E and v. While such verification is not possible under normal circumstances where the true material property distributions are unknown, the advantage of the synthetic experiment used here is that we can precisely quantify the accuracy of the predictions. The statistics of the posterior such as the mean or the mode of the posterior distribution have no fundamental status in Bayesian inference but are still useful to summarise the predictions. Here we choose to use the mean and standard deviation of the posterior distribution as representative of the Bayesian estimate of the material properties and the associated uncertainty in the predictions: the advantages of this choice are further discussed in Section 4.3 in the context of maximum likelihood/posterior predictions. To define these quantities recall that the sample Ω_s removed in iteration s of the Nested Sampling algorithm

corresponds to a configuration with spatial distributions $\hat{E}_s(x_i)$ and $\hat{v}_s(x_i)$ of the Young's modulus and Poisson's ratio, respectively. We then obtain the mean of the spatial distribution of the Young's modulus via (3.7) as

$$\langle \hat{E}(x_i)\rangle \equiv \int P(\boldsymbol{w}|\mathcal{D}, \mathcal{H}_p)\hat{E}(x_i)d\boldsymbol{w}$$

$$= \frac{1}{Z_p}\sum_{s=1}^{J}(\omega_s \mathcal{L}_s)\hat{E}_s(x_i), \tag{4.1}$$

while the standard deviation of $\hat{E}$ is given by

$$\Delta_{\hat{E}}(x_i) \equiv \sqrt{\frac{1}{Z_p}\sum_{s=1}^{J}\omega_s \mathcal{L}_s\left[\hat{E}_s(x_i) - \langle\hat{E}(x_i)\rangle\right]^2}. \tag{4.2}$$

An error measure quantifying the difference between the predicted mean and actual distribution of the Young's modulus is then defined as

$$\rho_E(x_i) \equiv \frac{\left|E(x_i) - \langle\hat{E}(x_i)\rangle\right|}{E(x_i)}. \tag{4.3}$$

Predictions of ρ_E within the specimen are shown for models $\mathcal{H}_5$ and $\mathcal{H}_8$ in Figs. 5a and 5b, respectively. While relatively large errors are evident for both models near some of the straight edges of the specimen it is clear that the errors are small in the specimen interior especially for model $\mathcal{H}_5$. The large errors near straight edges stem from the fact that the stress states in those regions are nearly uniaxial and this makes the set of elastic constant distributions that are compatible with the strain data highly non-unique in those locations. Nevertheless, the errors in both models $\mathcal{H}_5$ and $\mathcal{H}_8$ are reasonably similar and we shall proceed to demonstrate that the key advantage of model $\mathcal{H}_5$ (as quantified by its higher evidence) is the lower level of uncertainty in its predictions.
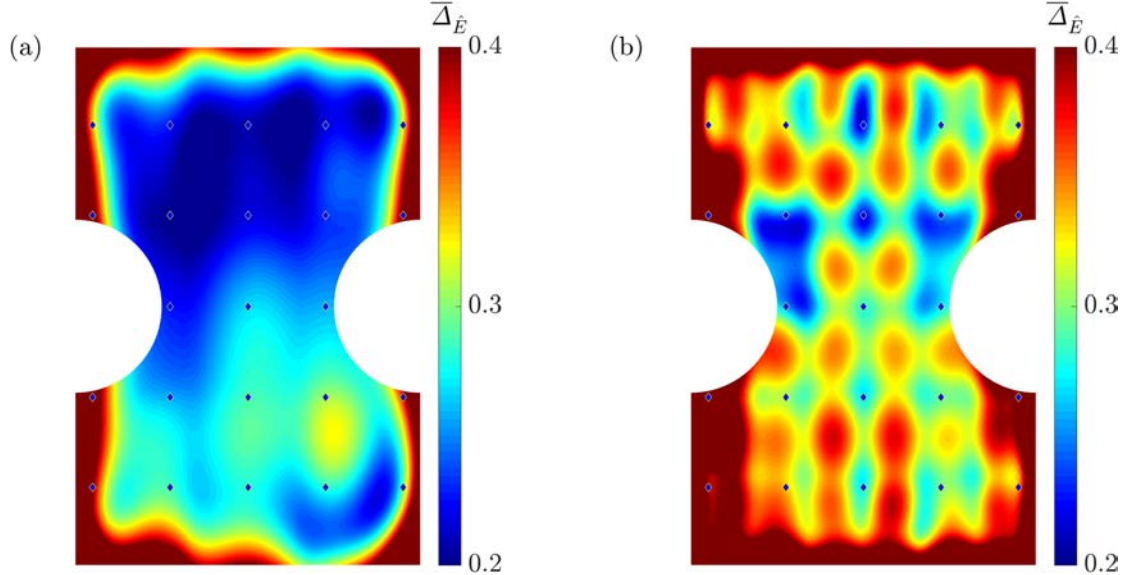
Figure 6: Spatial distributions of the uncertainty in the predictions as quantified by the normalised standard deviation $\bar{\Delta}_{\hat{E}}$ of the predicted Young's modulus for (a) model $\mathcal{H}_5$ and (b) model $\mathcal{H}_8$. The predictions are for the reference case of the synthetic experiment with $N = 23$ strain gauges and measurement noise $\sigma_{\text{err}} = 0.016\varepsilon_{\text{nom}}^{\infty}$. The strain measurement locations are indicated in both (a) and (b).

One of the main strengths of the Bayesian approach is that it can estimate the inevitable uncertainty associated with making predictions for this inherently ill-posed inverse problem. We quantify this uncertainty via the standard deviation of the posterior distribution, with a larger standard deviation indicative of a higher uncertainty in the predictions. Before the arrival of the data, we had a prior belief that the Young's modulus was uniformly distributed over the range $(E_{\min}, E_{\max})$ and thus there was a spatially uniform uncertainty $(E_{\max} - E_{\min})/(2\sqrt{3})$ as given by the standard deviation of the uniform distribution. We therefore define a normalised measure of the uncertainty

$$\bar{\Delta}_{\hat{E}}(x_i) \equiv \frac{2\sqrt{3}\Delta_{\hat{E}}(x_i)}{E_{\max} - E_{\min}}, \qquad (4.4)$$

to quantify the reduction in the level of uncertainty of the Young's modulus after the arrival of the data: $\bar{\Delta}_{\hat{E}}(x_i) = 1$ indicates no reduction in the level of uncertainty while $\bar{\Delta}_{\hat{E}}(x_i) = 0$ implies that we know the Young's modulus at x_i precisely. Predictions of $\bar{\Delta}_{\hat{E}}(x_i)$ are shown for models $\mathcal{H}_5$ and $\mathcal{H}_8$ in Figs. 6a and 6b, respectively. It is clear that $\bar{\Delta}_{\hat{E}}$ is significantly lower for model $\mathcal{H}_5$ compared to $\mathcal{H}_8$ with the largest uncertainties along the straight edges of the specimen where the errors ρ_E were also the highest. On the other hand, large uncertainties are seen throughout the specimen for model $\mathcal{H}_8$ with $\bar{\Delta}_{\hat{E}}$ only reducing in the vicinity of the strain measurement locations. The excess parameters in model $\mathcal{H}_8$ aid in closely matching the available data but the associated overfitting results in a large variability in between the data locations. In summary, Bayesian inference correctly predicts the most appropriate model, and this model gives predictions with the highest fidelity and the lowest uncertainty. In fact, spatial distributions of the uncertainty also give an indication of where the errors in the estimated properties are likely to be the highest. We note that analogous quantities $\langle \hat{v}(x_i) \rangle$, $\bar{\Delta}_{\hat{v}}(x_i)$ and $\rho_v(x_i)$ for the Poisson's ratio can also be defined via $\hat{v}_s(x_i)$; see Supplementary Information Figs. S2 and S3.

4.3   Bayesian versus maximum likelihood predictions
An alternative to Bayesian inference in such inverse problems is an optimisation approach where a search is conducted for the maximum likelihood solution (for the case of a uniform prior as used here this is equivalent to the maximum posterior solution). While it is clear that this approach cannot be used to distinguish between models as more complex models inevitably give higher likelihoods, here we examine the fidelity of this approach in selecting the parameters of a given model that is a priori assumed to be appropriate.

Maximum likelihood configurations are usually not uniquely defined with multiple different configurations giving equal or approximately equal local maxima of the likelihood function; see discussion in Section 5.2. Nevertheless, to illustrate predictions of this approach we obtain a maximum likelihood solution as follows. We first select the configuration with the highest likelihood in the active set after the termination of the Nested Sampling algorithm and then use it to seed a Nelder-Mead (Nelder and Mead, 1965) optimisation calculation to maximize $\hat{\mathcal{L}}$. The configuration that the optimisation outputs as the local maximum in $\hat{\mathcal{L}}$ is treated as the best representation of the maximum likelihood solution. We then define an error $\rho_E(x_i)$ analogous to (4.3) except that $\langle \hat{E}(x_i) \rangle$ is replaced by the spatial distribution $\hat{E}_{\mathrm{OPT}}(x_i)$ of the Young's modulus in the maximum likelihood configuration. Predictions of ρ_E are shown for models $\mathcal{H}_5$ and $\mathcal{H}_8$ in Figs. 7a and 7b, respectively (also see Supplementary Fig. S4 for

a replot of Fig. 7 using a larger scale for ρ_E). Not only are the magnitudes of the errors larger for model $\mathcal{H}_8$ compared to model $\mathcal{H}_5$ (compare Figs. S4a and S4b), the more significant finding is that the errors of these maximum likelihood solutions are significantly higher compared to the solutions based on the mean of the posterior distributions given in Fig. 5. This suggests that even if the true underlying model was known a priori (model $\mathcal{H}_5$ in this case), an estimate of the material properties based on the mean of the posterior distributions is more accurate than a maximum likelihood solution.
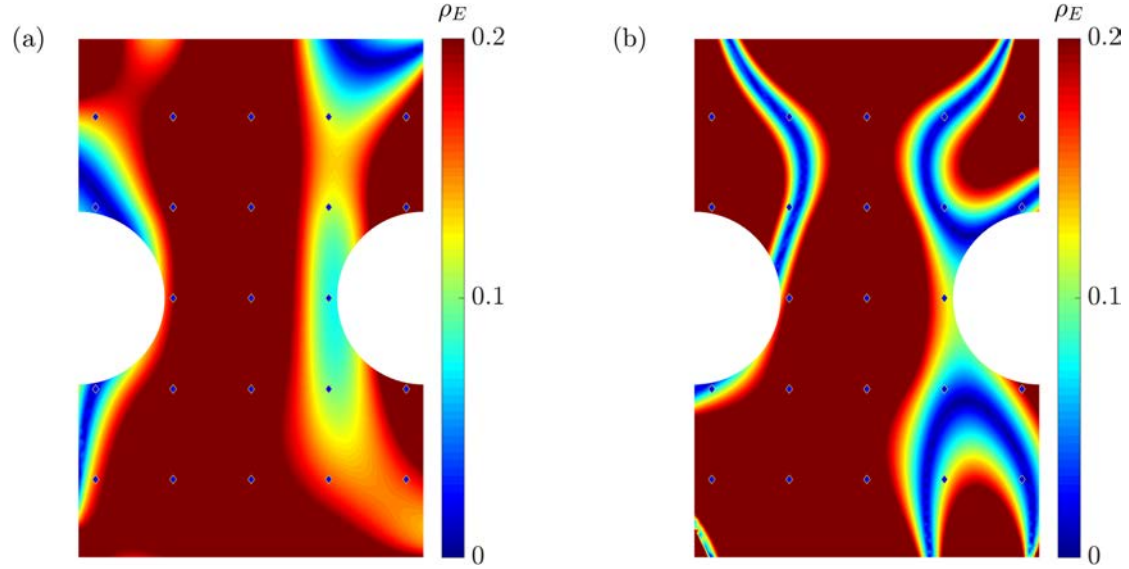
Figure 7: Spatial distributions of the error ρ_E in predictions of the Young's modulus using maximum likelihood solutions for (a) model $\mathcal{H}_5$ and (b) model $\mathcal{H}_8$. The predictions are for the reference case of the synthetic experiment with $N = 23$ strain gauges and measurement noise $\sigma_{\mathrm{err}} = 0.016\varepsilon_{\mathrm{nom}}^{\infty}$. Note that a colour scale that is the same as in Fig. 5 has been chosen to highlight the contrast between the maximum likelihood and mean posterior predictions. (A replot of this data using a larger scale for ρ_E is included in Supplementary Information Fig. S4 and better indicates the true values of the errors.)

This intriguing result is linked to the 1.6% measurement noise in the available data and is understood as follows. For the true model $\mathcal{H}_5$, there exists a near unique maximum likelihood configuration (see Section 5.2 and Fig. 14a) and thus the large errors in Fig. 7a are not a result of the presence of multiple maximum likelihood solutions (i.e. not directly related to the non-unique nature of the optimisation problem). Rather, the poor solution is an outcome of "overfitting" in the sense that the optimisation procedure converges to a configuration that matches the available data as closely as possible irrespective of the fact that there is noise in the strain measurements. In fact in the absence of measurement noise, the maximum likelihood solution of model $\mathcal{H}_5$ very accurately predicts the material property distributions with errors ρ_E and ρ_ν less than about 3% over nearly the entire specimen (Supplementary Information Fig. S5). Thus, remarkably the overfitting inherent in the maximum likelihood solution results in an amplification of the 1.6% measurement noise in the strain data to an error in excess of 30% in the predicted Young's modulus (Fig. S4a).
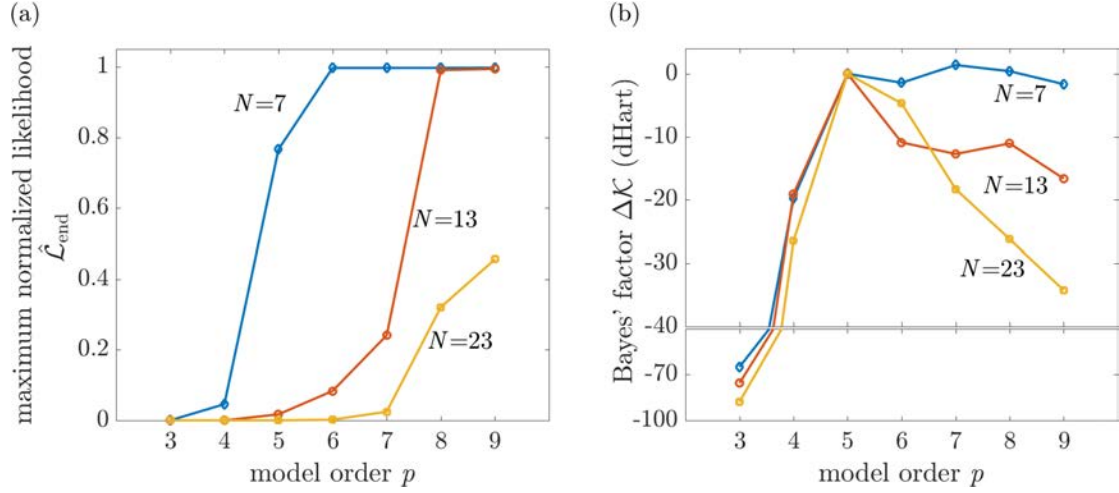
Figure 8: Predictions of the (a) maximum normalised likelihood $\hat{\mathcal{L}}_{\text{end}}$ and (b) Bayes' factor $\Delta\mathcal{K}$ as a function of the model order p. The predictions are for the case of synthetic experiments with a measurement noise $\sigma_{\text{err}} = 0.05\varepsilon_{\text{nom}}^{\infty}$ and $N = 7, 13$ and 23 strain gauge locations. The Bayes' factor is calculated using model $\mathcal{H}_5$ used as the reference.

By contrast, the Bayesian prediction using the mean of the posterior smooths out the random measurement noise and reduces the errors in the Young's modulus to less than 5% over most of the specimen (Fig. 5a). To understand the lower error compared to the maximum likelihood solution, recall that the likelihood function (2.1) recognises the presence of measurement noise and weights deviations between the measured and predicted strain values via an inputted measurement accuracy σ_{ε}. The mean of the posterior (which is weighted sum of the likelihood) therefore includes contributions from all configurations that give strain measurements that agree with the data within the specified accuracy σ_{ε}. This diminishes the effect of measurement noise of the predicted configuration and automatically avoids overfitting. (As a corollary, reducing σ_{ε} to below the noise level in the measurements will decrease the fidelity of the Bayesian predictions based on the mean of the posterior; see Section 4.5.) This ability of Bayesian inference through the mean of the posterior to reduce the effect of the unknown measurement noise is one the key strengths of the Bayesian approach.

The maximum likelihood predictions discussed above are obtained by solving what seems like a non-regularised ill-posed optimisation problem. Often, a regularised version of such optimisation problems is posed in an attempt to avoid overfitting of the data and reduce the amplification of data noise via the solution of the inverse problem. This is typically, but not exclusively, done by regularisations such as Tikhonov regularisation (Tikhonov and Arsenin, 1977), where a regularization term is added to the likelihood function in order to give preference to a particular solution with desirable properties (e.g. smoothness). However, the problem under investigation here is already highly regularized with smooth NURBS describing the spatial distribution of material properties. Moreover, the comparison between the Bayesian and maximum likelihood solutions discussed here was restricted to the true model $\mathcal{H}_5$. This bypasses the model selection problem with attention restricted to the model that has the correct wavelengths of the material property distributions. Thus, further regularisations are expected to yield no improvements in the maximum likelihood predictions and this is demonstrated explicitly in Supplementary Section S5 where we show numerical results for predictions using Tikhonov-type regularisation.

(a) (b)
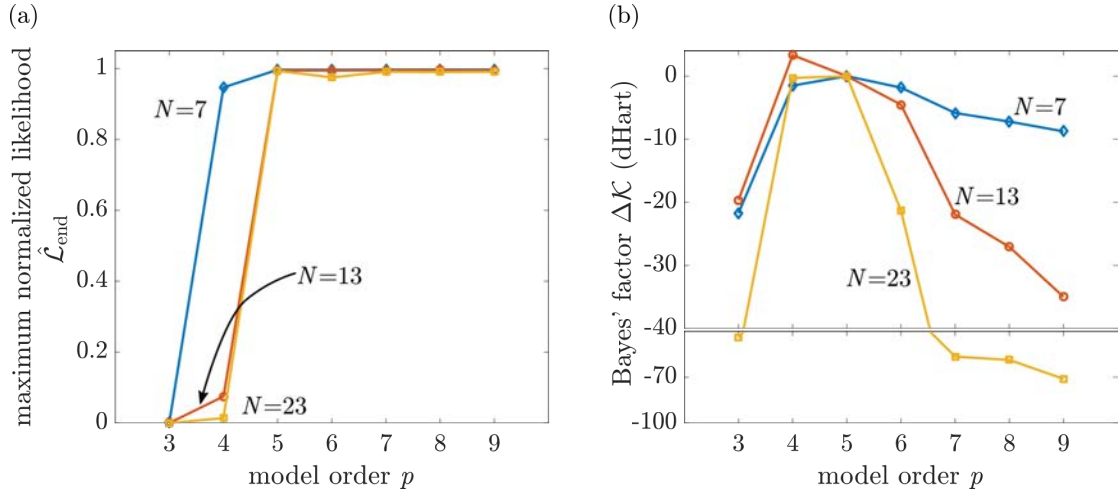
Figure 9: Predictions of the (a) maximum normalised likelihood $\hat{\mathcal{L}}_{\text{end}}$ and (b) Bayes' factor $\Delta\mathcal{K}$ as a function of the model order p. The predictions are for the case of synthetic experiments with no measurement noise ($\sigma_{\text{err}} = 0$) and $N = 7, 13$ and 23 strain gauge locations. The Bayes' factor is calculated using model $\mathcal{H}_5$ used as the reference.

4.4   Model selection: quantity of data and measurement noise

We now return to the model selection problem and discuss the sensitivity of the Bayes' factor to the quantity of data and measurement noise. Predictions of $\hat{\mathcal{L}}_{\text{end}}$ and $\Delta\mathcal{K}$ as a function of the model order p are included in Figs. 4a and 4b, respectively for $N = 7$ and 13 strain gauge locations (measurement noise $\sigma_{\text{err}} = 0.016\varepsilon_{\text{nom}}^{\infty}$). The key difference with less data is that unlike the $N = 23$ case, $\hat{\mathcal{L}}_{\text{end}} \approx 1$ for $p \geq 5$ for both the $N = 7$ and 13 cases. This is because the models are required to fit to less data, which in turn results in higher likelihoods. Thus, given less data the likelihood estimates suggest that all models with $p \geq 5$ are equally acceptable. On the other hand, even with less available data $\Delta\mathcal{K}$ still correctly attains its maximum value for model $\mathcal{H}_5$. However, we observe that the $\Delta\mathcal{K}$ versus p curve is less peaked as the amount of data used in the estimation process reduces. For example, the Bayes' factor for model $\mathcal{H}_8$ is $\Delta\mathcal{K} \approx -4$ and -53 for estimation with $N = 7$ and 23, respectively. Thus, while the evidence for model $\mathcal{H}_5$ over $\mathcal{H}_8$ is *barely worth mentioning* (Table 1) when strain data from only 7 locations is available, the evidence is *decisive* when strain data from 23 locations is used. Bayesian evidence is therefore seen to quantify what we would have intuitively expected, i.e. our ability to differentiate between models reduces with decreasing quantity of data because we gain less information about the specimen. (The information gain from the available data can be quantified via the Kullback-Leibler divergence (Kullback and Leibler, 1951); see Supplementary Information Fig. S7.) On the other hand, inferences based on only the likelihood erroneously suggest that we are obtaining higher fidelity predictions, with multiple models being equally suitable, as the amount of available data decreases.

The effect of measurement noise on the likelihood and Bayes' factor estimates is illustrated in Figs. 8 and 9 for $\sigma_{\text{err}} = 0.05\varepsilon_{\text{nom}}^{\infty}$ and $\sigma_{\text{err}} = 0$, respectively. First consider the $\sigma_{\text{err}} = 0.05\varepsilon_{\text{nom}}^{\infty}$ case. A consequence of the higher noise level is that increasingly complex models are necessary in order to attain high levels of likelihood. For example, with $N = 13$ strain measurement locations, the likelihood only attains its maximum value of $\hat{\mathcal{L}}_{\text{end}} = 1$ for $p \geq 8$ when $\sigma_{\text{err}} = 0.05\varepsilon_{\text{nom}}^{\infty}$, while with $\sigma_{\text{err}} =$

$0.016\varepsilon_{\mathrm{nom}}^{\infty}$ the maximum likelihood was attained for all models with $p \geq 6$. Furthermore, comparing Fig 4b and Fig 8b, we observe that the $\Delta\mathcal{K}$ versus p curve is less peaked with increasing measurement noise. In fact, when data is available from only 7 locations ($N = 7$) with $\sigma_{\mathrm{err}} = 0.05\varepsilon_{\mathrm{nom}}^{\infty}$, the Bayes' factors are unable to differentiate between models with $p \geq 5$ (Fig. 7b). Thus, even without any a priori knowledge of the measurement's noise level, Bayesian evidence appropriately indicates that with increasing measurement noise more data is required to achieve the same strength of evidence for model $\mathcal{H}_5$ over competing models.

Next, consider the case of measurements with no noise (Fig. 9). Now all models with $p \geq 5$, irrespective of the quantity of data, attain $\hat{\mathcal{L}}_{\mathrm{end}} = 1$. This is because the underlying material property distributions were generated using NURBS with a 5×5 grid of control weights and therefore all the (nested) models with $p \geq 5$ are capable of reproducing the noiseless data. However, again the Bayes' factor correctly peaks for model $\mathcal{H}_5$ and correctly identifies the appropriate model for the material property distributions. It is worth emphasizing here that the Bayes' factor predictions in Fig. 9b suggest that when measurement noise is absent, the available data is consistent with both models $\mathcal{H}_4$ and $\mathcal{H}_5$. This is mainly a consequence of the fact that while the data had no measurement noise, we had a used an estimated accuracy of $\sigma_{\varepsilon} = 0.05\varepsilon_{\mathrm{nom}}^{\infty}$ in the likelihood function: this large discrepancy between the actual and estimated measurement noise decreases the fidelity of the model selection process as we now proceed to discuss in Section 4.5.

4.5   Choice of measurement accuracy in the likelihood function

The choice of the standard deviation σ_{ε} in the likelihood definition (2.1) is motivated by considerations of the accuracy of the experimental measurements. Nevertheless, this accuracy is rarely known with precision and here we illustrate its influence on both model selection and the Bayesian predictions of the material property distributions. Predictions of $\Delta\mathcal{K}$ as a function of the model order p for the reference synthetic experiment with $N = 23$ strain measurement locations and an experimental noise $\sigma_{\mathrm{err}} = 0.016\varepsilon_{\mathrm{nom}}^{\infty}$ are included in Fig. 10a for selected values of σ_{ε}. The uncertainty in the model selection is increased when σ_{ε} is increased to $\sigma_{\varepsilon} = 0.1\varepsilon_{\mathrm{nom}}^{\infty}$ (i.e. $\sigma_{\varepsilon} \gg \sigma_{\mathrm{err}}$) and equivalently the confidence with which the Bayes' factor predicts that model $\mathcal{H}_5$ is most appropriate is enhanced when σ_{ε} is brought closer to σ_{err} with the choice $\sigma_{\varepsilon} = 0.025\varepsilon_{\mathrm{nom}}^{\infty}$. However, in this regime with $\sigma_{\varepsilon} > \sigma_{\mathrm{err}}$ the Bayes' factor always correctly estimates that model $\mathcal{H}_5$ is at least one of the appropriate models. Now consider the case when $\sigma_{\varepsilon} < \sigma_{\mathrm{err}}$ with $\sigma_{\varepsilon} = 0.007\varepsilon_{\mathrm{nom}}^{\infty}$. Here there $\Delta\mathcal{K}$ versus p relation is sophisticated but the key fact is that the Bayes' factor now favours the most complex model (i.e. model $\mathcal{H}_9$ in this case). This is because by using such a small value of σ_{ε} we are stating that the data is extremely accurate. This then means that fitting the data closely is not overfitting but a true representation of the underlying material property distributions.

To illustrate the effect of σ_{ε} on the Bayesian predictions of the material property distributions we proceed to present predictions of ρ_E calculated from the mean of the posterior for model $\mathcal{H}_5$ for selected choices of σ_{ε}. Spatial distributions of ρ_E with $\sigma_{\varepsilon} = 0.025\varepsilon_{\mathrm{nom}}^{\infty}$ and $\sigma_{\varepsilon} = 0.007\varepsilon_{\mathrm{nom}}^{\infty}$ are included in Figs. 10b and 10c, respectively. These figures are analogous to Fig. 5a where $\sigma_{\varepsilon} = 0.05\varepsilon_{\mathrm{nom}}^{\infty}$ was employed. The errors in Fig. 10b are similar to those in Fig. 5a suggesting that not anticipating the measurement

noise accurately does not significantly affect the predictions. However, reducing σ_ε to much below σ_{err} (Fig. 10c) significantly increases the error as the mean of the posterior now averages over a very small regime around the data and thereby excludes configurations that have strain predictions that lie within the noise of the data. In fact, reducing σ_ε to much below the actual measurement noise in effect reduces the Bayesian inference process to the maximum likelihood approach with more complex models being favoured as seen in Fig. 10a and also bringing the mean of the posterior closer to the maximum likelihood solution. The discussion presented here suggests that for both model selection and property predictions it is better to err on the side of using a σ_ε value that is slightly higher than what we might anticipate the measurement noise to be and thereby avoid overfitting. Thus, for the Bayesian inference methodology presented here it suffices to have a good upper bound estimate of the measurement noise to set σ_ε.
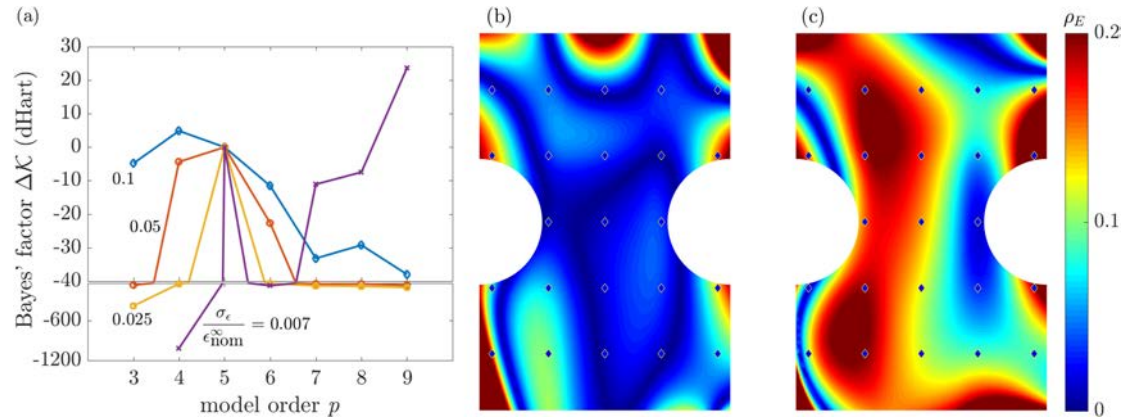
Figure 10: (a) Predictions of the Bayes' factor $\Delta\mathcal{K}$ as a function of the model order p. Predictions are included for four choices of the standard deviation σ_ε used in the likelihood function definition. Spatial distributions of the error ρ_E in predictions of the Young's modulus using the mean of the posterior distribution for model $\mathcal{H}_5$ with estimated measurement accuracies (b) $\sigma_\varepsilon = 0.025\varepsilon_{\mathrm{nom}}^\infty$ and (c) $\sigma_\varepsilon = 0.007\varepsilon_{\mathrm{nom}}^\infty$. The dataset used here is the reference case with $N = 23$ strain gauges and experimental noise $\sigma_{\mathrm{err}} = 0.016\varepsilon_{\mathrm{nom}}^\infty$.

4.6   *Approximation of the Bayes' factor*

While the Bayes' factor is shown to predict the appropriate model with high fidelity, it is computationally expensive to evaluate. Often, other criteria that are simpler to compute, but still include considerations of Occam's razor, are used for model selection. The two most commonly used criteria are the Bayesian information criterion (BIC) proposed by Schwarz (1978) and Akaike information criterion (AIC) introduced by Akaike (1974). The BIC is derived using Laplace's method to approximate the Bayesian evidence integral (2.5) and thus assumes that the parameter posterior is approximately Gaussian centred on the maximum a posteriori probability (MAP) estimate of the parameters. This allows the Occam factor and uncertainties in the model parameters to be directly evaluated from the Hessian (MacKay, 2003). In many circumstances, this Gaussian approximation becomes increasingly accurate as the amount of data increases, with the probability of selecting the true model by the BIC approaching unity in the limit of an infinite dataset. This consistency feature has resulted in the wide use of the BIC and thus here we only consider the BIC. With the available data being the $3N$ strain measurements, the BIC for model $\mathcal{H}_p$ with $2p^2$ parameters is given by

22

$$(\text{BIC})_p = 2\left[p^2 \ln\left(\frac{3N}{2\pi}\right) - \ln(\hat{\mathcal{L}}_{\max})\right], \qquad (4.5)$$

where $\hat{\mathcal{L}}_{\max} \approx \hat{\mathcal{L}}_{\text{end}}$ is the maximum normalised likelihood of model $\mathcal{H}_p$ (the 2π factor in (4.5) is often is dropped as typically $N \gg 2\pi$). The BIC is defined such that a model with the lowest BIC is preferred: Kass and Raftery (1995) suggest that the strength of evidence against model $\mathcal{H}_p$ compared to a reference model $\mathcal{H}_{\text{ref}}$ is *very strong* if $\Delta\text{BIC} \equiv (\text{BIC})_p - (\text{BIC})_{\text{ref}} > 10$.
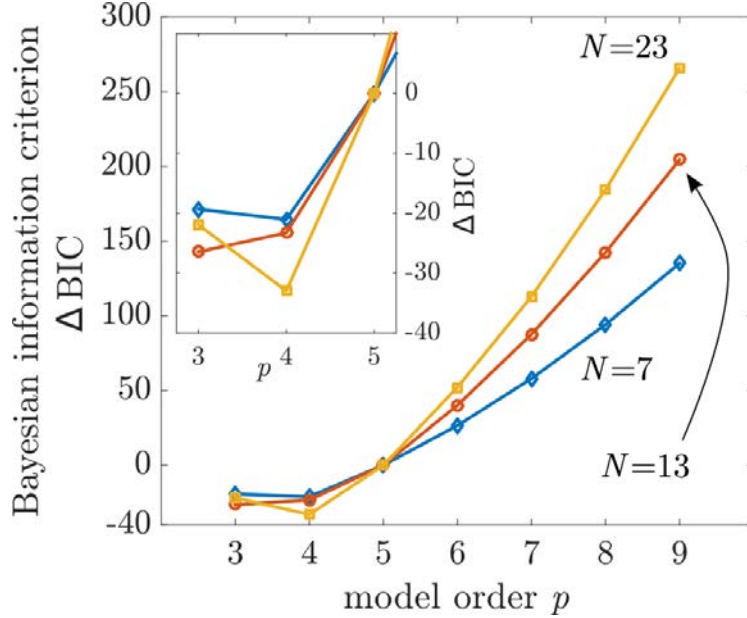
Figure 11: Predictions of the differential Bayesian information criterion ΔBIC as a function of the model order p. The predictions are for the case of synthetic experiments with a measurement noise $\sigma_{\text{err}} = 0.016\varepsilon_{\text{nom}}^{\infty}$ and $N = 7, 13$ and 23 strain gauge locations. The ΔBIC is calculated using model $\mathcal{H}_5$ used as the reference. The inset shows a zoom-in for $3 \leq p \leq 5$.

Predictions of ΔBIC are included in Fig. 11 for the case of the experiment with a measurement noise of $\sigma_{\text{err}} = 0.016\varepsilon_{\text{nom}}^{\infty}$ and model $\mathcal{H}_5$ set as $\mathcal{H}_{\text{ref}}$. It is clear that unlike the accurately computed Bayes' factors, the BIC suggests a preference for overly simple models with the preference for $\mathcal{H}_4$ and even $\mathcal{H}_3$ over $\mathcal{H}_5$ being *very strong* ($\Delta\text{BIC} < -10$ for models $\mathcal{H}_4$ and $\mathcal{H}_3$ for all data cases; see inset of Fig. 11). The primary reason for the failure of this criterion in this inverse elasticity problem is that for a given compatible strain field, multiple distributions of elastic constants can satisfy equilibrium depending on the precise boundary value problem. This directly violates the Gaussian posterior assumption on which the BIC is based (also see the likelihood landscapes in Section 5 that clearly show the non-Gaussian nature of the likelihood/posterior functions). Therefore, model selection for this problem cannot be performed using BIC. We emphasize here that model selection based on AIC was also similarly unsuccessful: for the sake of brevity these results are not explicitly discussed here.
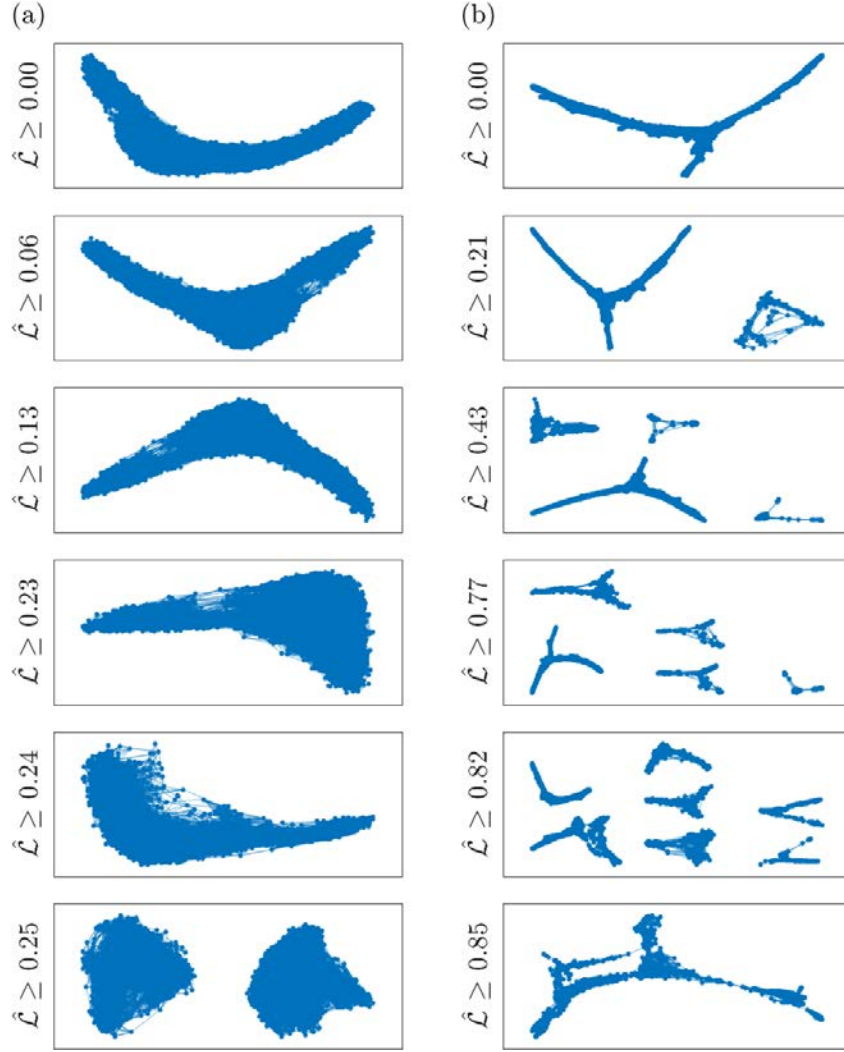
Figure 12: Connectivity graphs for model (a) $\mathcal{H}_5$ and (b) $\mathcal{H}_8$ with increasing normalised likelihood $\hat{\mathcal{L}}$ for the reference case of $N = 23$ strain gauges and experimental noise $\sigma_{\text{err}} = 0.016\varepsilon_{\text{nom}}^{\infty}$.

5. The structure of parameter phase space

Visualising the likelihood function can greatly enhance the understanding of the models as it shows the multiple likelihood maxima and the associated prior masses. However, the graphical representation of the $2p^2 -$ dimensional likelihood function is a challenging task. A similar problem is faced in statistical mechanics where it is useful to visualise the $3n -$dimensional potential energy function of system comprising n atoms. One way of depicting the topology of the energy basins and transition states is a disconnectivity graph (Wales et al., 1998) or the scaled disconnectivity graph (Wales and Bogdan, 2006) where in the latter case the width of the graph is made proportional to the number of minima. However, the disconnectivity graph does not provide information on the number of configurations available for a given energy level, i.e. does not contain entropy information. Here, to visualize the likelihood landscapes, we shall employ a method introduced by Pártay et al. (2010) that identifies the large-scale basins of the likelihood landscape by postprocessing the output sample set $\{\Omega_1, \Omega_2, \dots\}$ produced by Nested Sampling algorithm. The key point is that while we do not discover all the local maxima of $\hat{\mathcal{L}}$, we get a broad-brush view of the landscape using relatively

few samples. The procedure comprises two steps: (i) construction of connectivity graphs of the samples at a series of increasing levels of $\hat{\mathcal{L}}$ and (ii) using these graphs to build likelihood landscapes where the width of a basin is proportional to the prior mass of the basin at the given likelihood level.
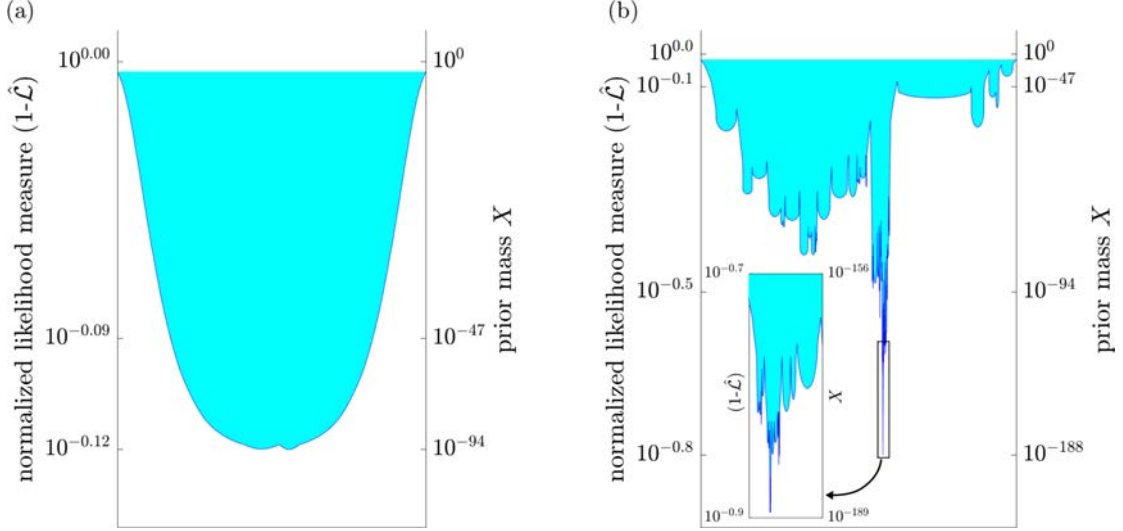
Figure 13: Likelihood landscapes with $(1 - \hat{\mathcal{L}})$ on the vertical axis and the normalised as well as rescaled phase space volumes of the basins on the horizontal axis. The right-hand axis shows the prior mass available below the given level $(1 - \hat{\mathcal{L}})$. Predictions are shown for models (a) $\mathcal{H}_5$ and (b) $\mathcal{H}_8$ for the reference case of the experiment with $N = 23$ strain gauges and a measurement error $\sigma_{\mathrm{err}} = 0.016 \varepsilon_{\mathrm{nom}}^{\infty}$. Note the differences in the likelihood scales in (a) and (b). The inset in (b) shows a magnified view of the deep basin to emphasize the splitting that occurs within that basin.

5.1 Connectivity graphs and splitting of the phase space with increasing likelihood
The output of the Nested Sampling for model $\mathcal{H}_p$ is a sequence $\{\Omega_1, \Omega_2, \dots\}$ of configurations with increasing likelihood $\hat{\mathcal{L}}$. To construct the connectivity graph, a metric for the distance between two configurations is required and here we use a measure of the separation between configuration (i) and (j) defined as

$$r^{(ij)} \equiv \sqrt{\left(\frac{\left\|\widetilde{\boldsymbol{E}}_i - \widetilde{\boldsymbol{E}}_j\right\|_2}{\Delta E}\right)^2 + \left(\frac{\left\|\widetilde{\boldsymbol{\nu}}_i - \widetilde{\boldsymbol{\nu}}_j\right\|_2}{\Delta \nu}\right)^2}, \quad (5.1)$$

where $(\widetilde{\boldsymbol{E}}_i, \widetilde{\boldsymbol{\nu}}_i)$ are the $2p^2$ control weights that define the configuration and $\|\cdot\|_2$ denotes the Euclidean norm. The Young's modulus and Poisson's ratio components are normalized by $\Delta E \equiv (E_{\max} - E_{\min})$ and $\Delta \nu \equiv (\nu_{\max} - \nu_{\min})$ respectively so as to appropriately weight their relative contributions. We begin with constructing a graph $\mathcal{G}_1$ whose nodes are all the configurations collected by the Nested Sampling algorithm. Each of these nodes is connected via edges to η nearest nodes (using the above distance measure) that have a lower likelihood. Thus, a connection in the graph implies proximity of two configurations in phase space, as we shall clarify subsequently.

We then construct the connectivity graphs as follows. The lowest likelihood of the nodes in graph $\mathcal{G}_1$ is $\hat{\mathcal{L}}_1$ and this node is removed and all edges connected to this node eliminated to create a new graph $\mathcal{G}_2$. This process of removing the node with the lowest likelihood is continued and at each stage r a new graph $\mathcal{G}_r$ is created. In this process

the initial graph $\mathcal{G}_1$ will split into two or more disconnected subgraphs with increasing $\hat{\mathcal{L}}$ as connections are lost between nodes. We illustrate the outcome of this procedure for the reference case of the synthetic experiment with $N = 23$ strain gauges, experimental noise $\sigma_{\text{err}} = 0.016\varepsilon_{\text{nom}}^{\infty}$ and the choice of connectivity $\eta = 10$. Graphs[5] with increasing $\hat{\mathcal{L}}$ are plotted in Figs. 12a and 12b for models $\mathcal{H}_5$ and $\mathcal{H}_8$, respectively. With increasing $\hat{\mathcal{L}}$ there is no splitting of the graph for model $\mathcal{H}_5$ expect at $\hat{\mathcal{L}} \approx \hat{\mathcal{L}}_{\text{end}} \approx 0.25$ where the graphs splits into two basins with approximately equal number of nodes. This is in contrast with model $\mathcal{H}_8$ where significant splitting occurs at $\hat{\mathcal{L}} \approx 0.5\hat{\mathcal{L}}_{\text{end}}$. The physical interpretation of these split subgraphs is that a path in configuration space cannot be found connecting a configuration in one of the disconnected subgraphs to a configuration within another subgraph without passing through lower likelihood configurations. Note that in Fig. 12b, some subgraphs are eliminated with increasing likelihood when all nodes within those subgraphs have likelihood lower than the current level. This represents a local extremal point in the likelihood function. We shall now proceed use this understanding to construct likelihood landscapes.

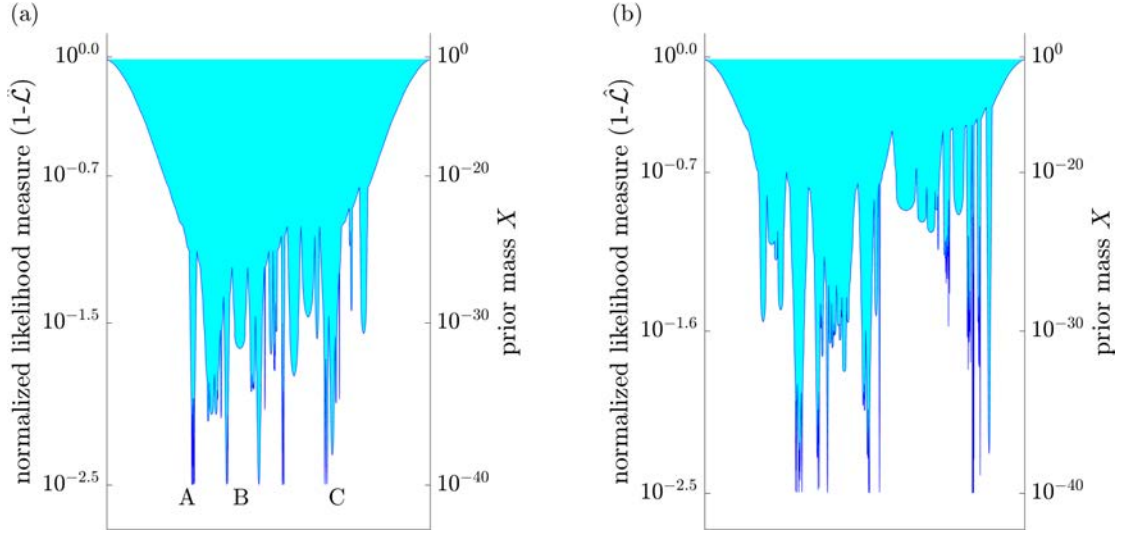### 5.2   Likelihood landscape charts

The information included in the connectivity graphs can be effectively portrayed in a single figure in the form of likelihood landscape charts, where we plot $(1 - \hat{\mathcal{L}})$ on the vertical axis and the width of the chart proportional to the prior mass available above the given $\hat{\mathcal{L}}$. Here we choose to plot $(1 - \hat{\mathcal{L}})$ on the vertical axis so that the high likelihoods appear as minima in the chart – this makes the chart more intuitive as it is analogous to the energy landscapes used to visualize the phase space of chemical systems with the basins depicted as valleys.

Recall that each node in the graph in Fig. 12 represents a fraction ω_s of the prior mass that is associated with the likelihood range $\hat{\mathcal{L}}_{s-1} < \hat{\mathcal{L}} < \hat{\mathcal{L}}_s$. Here the subscript s is the iteration number at which that node (configuration) was extracted during the Nested Sampling algorithm. At low likelihoods there are no disconnected subgraphs indicating the existence of a single basin. The prior mass X_s of this single basin at a likelihood $\hat{\mathcal{L}}_s$ is

$$X_s = \left(\frac{K}{K+1}\right)^s \approx \sum_{i=s}^{J} \omega_i , \qquad (5.2)$$

where the Nested Sampling was terminated at iteration $i = J$, i.e. X_s is the sum of the weights of all nodes with a likelihood greater than $\hat{\mathcal{L}}_s$ and the approximate sign emphasizes that since J is finite we have not fully explored the phase space in the Nested Sampling algorithm. In Fig. 12 as the likelihood is increased and the graph splits into two or more disconnected subgraphs, the nodes in the subgraphs are grouped within newly formed basins. The prior mass $X_s^{(m)}$ of each basin m at likelihood level $\hat{\mathcal{L}}_s$ is $X_s^{(m)} = \sum_{i=s}^{J} \omega_i^{(m)}$ where $\omega_i^{(m)}$ are the weights of the nodes within basin m and of course the total prior mass at $\hat{\mathcal{L}}_s$ is given by the sum $X_s = \sum_m X_s^{(m)}$ over all basins m. If at a given $\hat{\mathcal{L}}_s$ a subgraph is eliminated without further splitting, it represents a basin associated with a local minimum in $(1 - \hat{\mathcal{L}})$ (corresponding to local maxima in $\hat{\mathcal{L}}$). The output of the algorithm is a hierarchical nested tree of basins, with known prior masses.

---

[5] The graphs were drawn using subspace embedding (Koren, 2005) by first plotting the graph nodes in a 100-dimensional subspace and then projecting back into a 2D space.

Figure 14: Likelihood landscapes with $(1 - \hat{\mathcal{L}})$ on the vertical axis and the normalised as well as rescaled phase space volumes of the basins on the horizontal axis. The right-hand axis shows the prior mass available below the given level $(1 - \hat{\mathcal{L}})$. Predictions are shown for models (a) $\mathcal{H}_5$ and (b) $\mathcal{H}_8$ for the synthetic experiment with $N = 7$ strain gauges and an experimental noise $\sigma_{\mathrm{err}} = 0.016\varepsilon_{\mathrm{nom}}^{\infty}$. In (a) we have labelled three basins as A, B and C with configurations corresponding to the maximum likelihood in these basins illustrated in Fig. 15.

The rapid shrinking of the available prior mass with decreasing $(1 - \hat{\mathcal{L}})$ presents a challenge to visualize the landscapes. In order to make the visualisation feasible we recursively scale the horizontal phase volume axis as follows:

(i)   We begin by first defining logarithmically normalized prior masses of the basins as

$$\hat{X}_s^{(m)} = \log_{10}\left(\frac{X_s^{(m)}}{X_J}\right), \qquad (5.3)$$

where $X_J \equiv [K/(K + 1)]^J$ with Nested Sampling terminated at iteration J.

(ii)   The process of constructing the landscapes begins at a low likelihood when only one basin exists (Fig. 12). Let this basin split into daughter basins at a likelihood level $\hat{\mathcal{L}}_p$ with $\hat{Y}_p = \hat{X}_p$ denoting the normalized prior mass at $\hat{\mathcal{L}}_p$. We then define a scaled prior mass for daughter basin m at $\hat{\mathcal{L}}_s$ as $\tilde{X}_s^{(m)} \equiv \hat{X}_s^{(m)}/\hat{Y}_p$.

(iii)   This rescaling is carried out recursively such that when the daughter basin in (ii) splits at a scaled prior mass $\tilde{Y}_q^{(m)} = \tilde{X}_q^{(m)}$, the rescaled prior mass of its daughter n is defined as $\tilde{X}_s^{(n)} \equiv \hat{X}_s^{(n)}/\tilde{Y}_q^{(m)}$. This recursive rescaling is equivalent to redefining the horizontal axis scale for each basin when it forms, i.e. the prior mass of the basins depicted in the landscapes are relative to the prior mass of the basins at their formation. Such a rescaling makes it complicated to work out from the landscapes the absolute prior masses of basins that recursively split but is required to visualize the prior masses of the basins that can span more than 150 orders of magnitude.

The likelihood landscapes of models $\mathcal{H}_5$ and $\mathcal{H}_8$ for the reference case of the synthetic experiment with $N = 23$ strain gauges and an experimental noise $\sigma_{\mathrm{err}} = 0.016\varepsilon_{\mathrm{nom}}^{\infty}$ are plotted in Fig. 13. In this plot, the width of each basin at a given likelihood level $(1 - \tilde{\mathcal{L}}_s)$ is the normalized and rescaled prior masses $\tilde{X}_s^{(m)}$ while the horizontal

ordering of the basins is arbitrary. The likelihood level $(1 - \hat{\mathcal{L}})$ is indicated on the left-hand axis while the right-hand axis indicates the corresponding prior mass with likelihood greater than the current level. The landscape for model $\mathcal{H}_5$ (Fig. 13a) has two local maxima at approximately the same likelihood. This is contrasted with the landscape for model $\mathcal{H}_8$ in Fig. 13b where the landscape first splits into multiple basins at a relatively low likelihood. Thereafter, some of the basins undergo further repeated splits with the landscape containing multiple local maxima in $\hat{\mathcal{L}}$. This is especially true for the deep basin in the model as shown in the inset in Fig. 13b. Model $\mathcal{H}_5$ therefore predicts a near unique maximum likelihood solution while model $\mathcal{H}_8$ suggests the existence of multiple approximately equally probable maximum likelihood solutions.
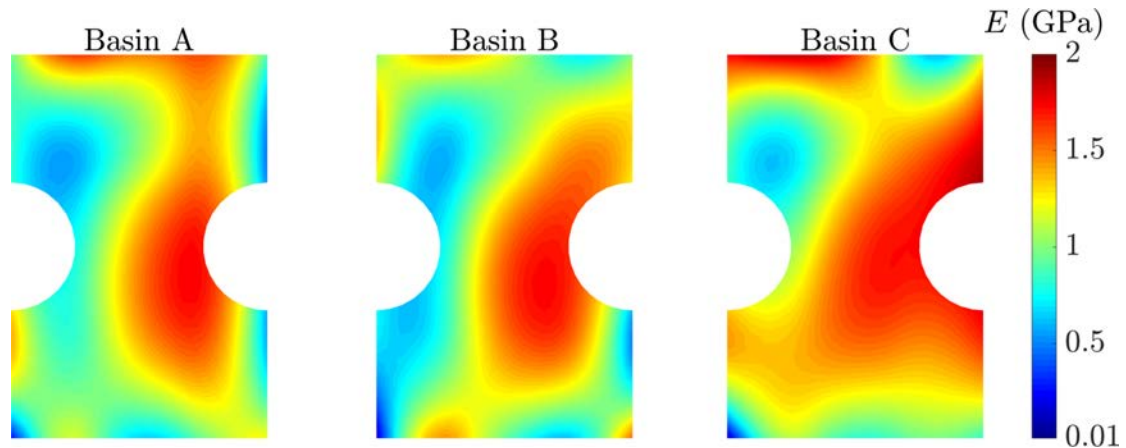
Figure 15: The predicted spatial distributions of the Young's modulus $\hat{E}$ for configurations corresponding to the maximum likelihood in basins A, B and C marked in Fig. 14a. The colour scale is set to be equal to the prior distribution of the Young's modulus.

Paucity of available data was shown in Section 4.1 to decrease the strength of the evidence for model $\mathcal{H}_5$ over more complex models (Fig. 4). To better understand this, we show in Fig. 14 likelihood landscapes for the synthetic experiment with $N = 7$ strain gauge locations and an experimental noise $\sigma_{\text{err}} = 0.016\varepsilon_{\text{nom}}^{\infty}$. Now even model $\mathcal{H}_5$ (Fig. 14a) has multiple basins with a large number of local likelihood maxima similar to model $\mathcal{H}_8$ (Fig. 14b). The main difference between the landscapes of models $\mathcal{H}_5$ and $\mathcal{H}_8$ is that the landscape of model $\mathcal{H}_8$ splits into multiple basins starting at a lower likelihood. The multiple basins with only $N = 7$ strain gauge locations are consequence of the scarcity of data, which allows both models to make a greater variety of predictions that are compatible with the available data. This results in the strength of evidence for model $\mathcal{H}_5$ over $\mathcal{H}_8$ being "barely worth mentioning" with the slight preference for model $\mathcal{H}_5$ over $\mathcal{H}_8$ being due to the fact that the landscape of model $\mathcal{H}_8$ splits into multiple basins at a lower likelihood.

With a deficiency of available data ($N = 7$) multiple equal likelihood maxima are predicted (Fig. 14). Each of these basins is associated with configurations that differ significantly in terms of the spatial distributions of the elastic constants. To illustrate this we consider the three basins labelled A, B and C in Fig. 14a. The spatial distributions $\hat{E}(x_i)$ of the Young's modulus for the highest likelihood configuration in each of these basins are plotted in Fig. 15. Clearly, each of these distributions is significantly different even though they all have approximately the same likelihood. As mentioned in the Introduction, the search for high likelihood solutions is usually posed

as an optimization problem, and solved using algorithms such as gradient searches, simulated annealing etc. These algorithms deliver a single solution associated with a local or global minimum depending on the solution strategy and the initialisation procedure for the calculation. The discussion here illustrates the pitfalls of such an approach for this ill-posed problem with multiple equally likely solutions: an optimisation approach would not only arbitrarily pick one of the high likelihood solutions but also does not give an indication of the inherent uncertainty associated with the proposed solution.

6.   Discussion

The Bayesian approach used here tackles both the problem of model selection as well as estimating the model parameters using the mean of the posterior distribution. The numerical results presented above suggest clear advantages of this Bayesian approach over traditional optimisation methods based on maximum likelihood. However, these gains come at the price of considerably higher computational costs and finding ways to reduce these costs is of interest. In the majority of the Bayesian inference literature model selection is performed using approximations of the Bayes' factor with BIC and AIC being the most common criteria. However, here we have shown that the highly non-Gaussian nature of the posterior distribution of the problem analysed in this study implies that these more traditional approaches to reduce computational costs fail. Rather, we foresee that the development of computationally more efficient strategies for Bayesian inference in this class of problems will involve improving the performance of the Nested Sampling algorithm. This might include two approaches:

(i) Use of more sophisticated Monte Carlo schemes than the simple Gibbs type sampler employed here. For example, Hamiltonian Monte Carlo or Galilean Monte Carlo has recently been used (Baldock et al., 2017) to gain a speedup factor equal to the dimensionality of atomistic systems. Similar gains might very well be possible in elasticity problems too.

(ii) Development of fast surrogate models to be used for evaluating the likelihood functions instead of FE solvers. In this initial investigation, we wanted to study the application of Bayesian inference for inverse elasticity problems, and not end up mixing together errors due to noise and sampling with errors due to the use of surrogate or reduced models. However, recent advances in machine learning schemes for differential equation solutions (Raissi et al. 2017a, 2017b) suggest such approaches are worth investigating.

The Bayesian formalism allows for educated choices of experimental designs (e.g. number and location of the strain gauge rosettes in this case) to be made by maximising the expected utility of an experiment. In particular, the utility is defined as the prior-posterior gain in Shannon information and thus the goal in Bayesian experimental design is to maximise the expected Kullback–Leibler divergence D_{KL} between the prior and the posterior distributions. The Bayesian calculations using Nested Sampling immediately provide D_{KL}. Such calculations to gauge the relative utility of increasing number N of measurement locations are presented in the Supplementary Information section S6 and suggest that increasing the utility begins to plateau around $N = 23$, justifying our choice of the reference number of gauges used here. However, we emphasise that in this study, the precise location of the strain gauge rosettes was not chosen using a Bayesian approach. Rather, similar to most current experimental studies the given number of strain gauges were distributed as uniformly as possible over the

specimen. The improvement in the predictions by using a Bayesian approach to select the location of the strain gauges remains a topic for future investigations.

7.   Concluding remarks

The inverse problem of estimating the spatial distribution of elastic constants from the measurement of strains within the domain is an ill-posed problem in the following sense: (i) the solution is dependent on the choice of the model (i.e. set of basis functions) used to represent the distribution of elastic constants and (ii) multiple parameter sets for a given model might be compatible with the available data. Here we have investigated the application of Bayesian inference to both identify the appropriate model as well as to estimate the model parameters and their uncertainty, given the available strain data. Nested Sampling, which is an algorithm specifically designed to sample high dimensional phase spaces where the bulk of the probability mass is contained in an exponentially small fraction of the phase space volume, was used to determine the posterior distributions of the parameters and calculate the Bayesian evidence of the models.

In order to illustrate the method we numerically generated a synthetic specimen with a random spatial distribution of isotropic elastic constants using NURBS functions with grid of 5×5 control weights. The synthetic strain data was then calculated via the finite element (FE) method. These strains, with and without added measurement noise, were then used as the only available data from which to estimate the spatial distribution of material properties using models with a grid of $p \times p$ control weights. In the presence of measurement noise, the maximum likelihoods increased with increasing p (i.e. increasing model complexity). However, this was a consequence of the complex models overfitting the experimental data, which ultimately resulted in poor predictions of the actual spatial distributions of the material properties. On the other hand, the Bayes' factor correctly estimated that a model with a grid of 5×5 control weights best describes the spatial distribution of elastic constants within the specimen and with the lowest uncertainty. Moreover, inference based on the Bayes' factor appropriately indicates an increasing uncertainty in model selection with decreasing data and increasing measurement noise. By contrast, inference based purely on likelihood is misleading with the maximum likelihood increasing as the amount of available data decreases. Simpler model selection criteria such as Bayesian information criterion (BIC) were shown to fail for the inverse elasticity problem investigated here and this makes accurate evaluation of the posterior distribution and the associated Bayesian evidence integral (by Nested Sampling or other means) imperative for this class of problems.

The output of the Nested Sampling algorithm was also used to construct likelihood landscape charts. These charts show that when there is sufficient quantity of data, the correct model with $p = 5$ model has a near unique maximum likelihood solution. On the other hand, with increasing model complexity (i.e. $p > 5$) or with a paucity of data, the landscapes show multiple local maxima in the likelihood with a wide variety of equally probable distributions of elastic constants. While an optimisation approach will undoubtedly mislead when the likelihood landscape has multiple maxima, intriguingly even for the correct $p = 5$ model with a near unique maxima, the mean of the posterior distribution was seen to predict material properties with significantly higher fidelity compared to the maximum likelihood solution. This is because the maximum likelihood

solution amplifies the measurement noise by overfitting to the noisy data while the mean of the posterior, at-least to some extent, cancels out the random noise. In summary, there are clear advantages of Bayesian inference over optimisation approaches based on maximum likelihood. The downside of Bayesian inference is of course the considerably higher computational cost – but we believe that in many real world and safety critical applications, the availability of cheap computational power would tip the balance in favour of the better inference method.

While Bayesian inference using Nested Sampling has been demonstrated here with the focus on an inverse elasticity problem, the ideas are general and can be readily used for a wide range of high dimensionality problems in mechanics. This includes selection of constitutive models given available (noisy) material data as well as the corresponding estimation of parameters and the associated uncertainty. The key advantages of Bayesian inference over optimisation procedures typically used in these problems include: (i) an in-built Occam's razor that aids in avoiding overfitting both for model selection and parameter estimation; and (ii) a quantification of the uncertainty associated with the estimated parameter values.

Appendix A: Effect of the order of the B-spline functions

All numerical results presented in the main text employ nested models where the size p of the control weight grid is varied between models $\mathcal{H}_p$ but the NURBS always use quadratic ($q = 2$) B-spline functions. Here we explore the Bayes' factors for non-nested models where we keep $p = 5$ fixed and vary the order q of the B-spline functions. Knot vectors similar to models $\mathcal{H}_p$ are employed in all these models, i.e. open and uniform with $q + 1$ equal knot values at both ends. Consistent with the remainder of the study, here we set $\mathcal{H}_{\text{ref}}$ to be model $\mathcal{H}_5$ (i.e. a model with a $p = 5$ control weight grid and quadratic B-spline functions) to define $\Delta\mathcal{K}$.

Predictions of the evolution of the normalised likelihood $\hat{\mathcal{L}}$ with the prior mass X are included in Fig. A1a. When nearly all the initial parameter space is available (i.e. at high X), $\hat{\mathcal{L}}$ rises faster for models with higher q. This is because; models with higher order B-splines are more non-local such that a perturbation of a control weight affects the elastic constant distributions in a larger portion of the domain. This implies that larger fractions of the initial parameter space attain relatively high values of $\hat{\mathcal{L}}$. However, the maximum likelihoods attained by the model peak for the model with quadratic B-splines as the distributions within the synthetic experimental specimen were generated using the same NURBS functions. These predictions of the maximum likelihoods, as parameterised by $\hat{\mathcal{L}}_{\text{end}}$, are summarised in Fig. A1b as a function of q. Clearly $\hat{\mathcal{L}}_{\text{end}}$ peaks at $q = 2$ although the cubic B-spline model also gives reasonably high maximum likelihoods. The corresponding predictions of $\Delta\mathcal{K}$ are also included in Fig. A1b and indicate that Bayesian inference procedure is unable to distinguish between models with $q = 2$ and $q = 3$ ($\Delta\mathcal{K} \approx 5$ for the $q = 3$ model which as per Table 1 is interpreted as "not worth mentioning" in terms of the strength of evidence). This ambiguity arises because while the $q = 2$ model attains a higher likelihood, a greater fraction of the possible configurations of the $q = 3$ model have a higher likelihood a high X (see inset in Fig. A1a). Thus, while the Bayes' factor can unambiguously predict that model $\mathcal{H}_5$ is most appropriate among the nested models $\mathcal{H}_p$, there is some ambiguity in the predictions using the same dataset for these non-

nested models. Nevertheless, the Bayes' factor correctly indicates that the $q = 2$ model is one of the two approximately equally probable candidates.
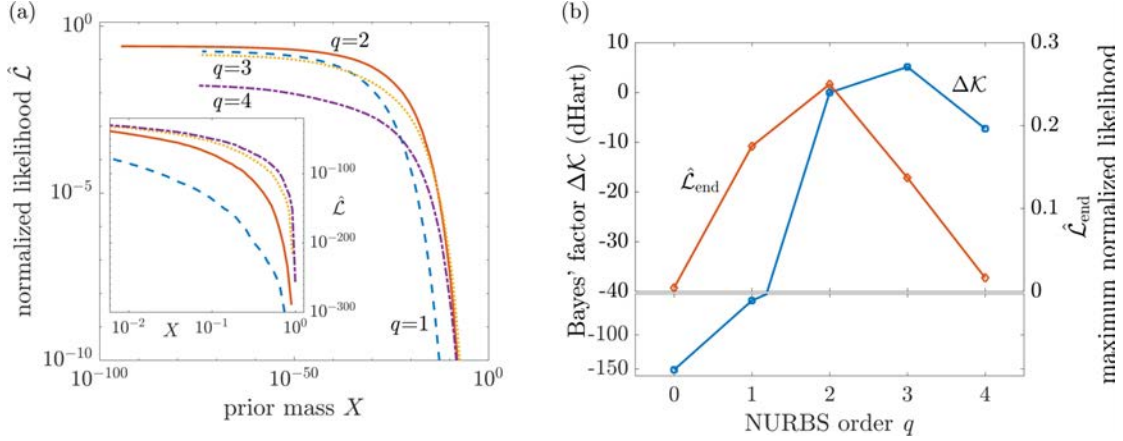
Figure A1: (a) Predictions of the variation in the normalised likelihood $\hat{\mathcal{L}}$ with the prior mass X for models with order q of the B-spline functions used in the NURBS. (b) The corresponding predictions of the Bayes' factor and normalised maximum likelihood $\hat{\mathcal{L}}_{\mathrm{end}}$. The dataset used here is the reference case with $N = 23$ strain gauges and experimental noise $\sigma_{\mathrm{err}} = 0.016\varepsilon_{\mathrm{nom}}^{\infty}$. All models have a control weight grid with $p = 5$.

Appendix B: Description of the Monte Carlo procedure used to regenerate the active set in Nested Sampling

In Nested Sampling, when the sample with the lowest likelihood is removed from the active set, a single new configuration is required to replace the one that was removed. This new sample needs to be chosen such that the regenerated active set with K samples remains uniformly distributed over the phase space with likelihood above the current level. However, new samples cannot be generated in the sth iteration by rejection sampling (whereby random configurations are proposed and only samples with $\hat{\mathcal{L}} > \hat{\mathcal{L}}_{s-1}$ are accepted). This is because the volume of the allowed region in configuration space collapses exponentially as the Nested Sampling calculation progresses: if we used rejection sampling, then after $100K$ iterations the probability of accepting proposed configurations would be e^{-100} the probability at the first iteration. Instead, new samples are generated from the prior distribution by Markov Chain Monte Carlo (MCMC) with equilibrium probability distribution $P(\Omega) \propto \mathbb{I}[\mathcal{L}(\Omega) > \mathcal{L}_{s-1}]$. The Markov chain is initialised by cloning a randomly selected existing sample from the region $\hat{\mathcal{L}} > \hat{\mathcal{L}}_{s-1}$ (i.e. from the remaining active set with $K - 1$ samples) with the chain being sufficiently long so that the final configuration has lost correlation with its starting point. This Markov chain is generated using the Metropolis et al. (1953) algorithm as follows:

(i) Randomly pick one control weight from $\widetilde{\boldsymbol{E}}$ and one control weight from $\tilde{\boldsymbol{\nu}}$. These weights are then perturbed by two independently drawn random numbers with probability density functions $\mathcal{U}(-\varsigma\Delta E, \varsigma\Delta E)$ and $\mathcal{U}(-\varsigma\Delta\nu, \varsigma\Delta\nu)$, where $0 \leq \varsigma \leq 1$ is the step size while $\Delta E \equiv (E_{\max} - E_{\min})$ and $\Delta\nu \equiv (\nu_{\max} - \nu_{\min})$. If the perturbed weights lie outside the prior intervals $(E_{\min}, E_{\max})$ and $(\nu_{\min}, \nu_{\max})$ then the perturbed state is rejected and the process repeated until a perturbed state within the prior is obtained.

(ii) The likelihood $\hat{\mathcal{L}}$ of this new configuration is computed via a FE calculation.

(iii) The Metropolis criterion is used to decide whether or not to accept this perturbed configuration so as to obtain a chain with equilibrium probability distribution $P(\Omega) \propto \mathbb{I}[\mathcal{L}(\Omega) > \mathcal{L}_{s-1}]$, i.e. if $\hat{\mathcal{L}} > \hat{\mathcal{L}}_{s-1}$ the perturbed state is accepted, otherwise the configuration prior to step (i) is repeated in the Markov chain.

(iv) A Markov chain with $\mathcal{T}$ steps is generated and the active set replenished with the configuration from the last step of Markov chain.

In line with typical MCMC calculations, we attempt to achieve an acceptance rate of about 35% in the Metropolis criterion and periodically adjust ς in order to ensure that we stayed with $\pm 5\%$ of this target acceptance rate. Thus, the Nested Sampling algorithm has only two numerical parameters: the walk length $\mathcal{T}$ and active set size K. The parameter $\mathcal{T}$ should be set large enough so that the decorrelated sample is equally likely to have originated from each of the other samples while the error in calculating the evidence $\mathcal{Z}_p$ is proportional to $1/\sqrt{K}$ (Skilling, 2006).

The Markov chain in Nested Sampling involves K independent copies of the system, which evolve quasi-simultaneously, and it is only in this replicated space that we expect the sampler to converge. Individual walkers are not expected to mix well within a single iteration, and traditional measures of convergence such as trace plots are not informative. In fact, it is one of the advantages of the Nested Sampling algorithm that it starts with completely independent walkers, and so relatively short MCMC trajectories are enough at each iteration to maintain sufficient uniformity of the sampling distribution for the evidence and final solution to converge adequately. We ascertained convergence of the entire Nested Sampling procedure by varying the walk length $\mathcal{T}$ over two orders of magnitude and K over the range 150 to 5000. These convergence studies demonstrated that the choice $K = 256$ and $\mathcal{T} = 512$ is adequate and hence used for all the calculations presented here. In order to reduce computational time, a parallel version of the above algorithm proposed by Baldock (2014) was employed. Given $\mathcal{P}$ processors, the active set comprising K samples was divided equally among the processors. At each iteration, the sample with the lowest likelihood in the entire active set (over all processors) was removed. A clone was then generated (from the entire active set) to replace the sample that was removed. Now not only was this cloned sample decorrelated using the Markov chain as described above, but also one randomly selected sample on each of the remaining processors also underwent the MCMC process. However, in this parallel version a Markov chain of length $\mathcal{T}/\mathcal{P}$ was employed on each processor. This thereby decreases computational time by a maximum of a factor of $\mathcal{P}$ (i.e. linear scaling with number of processors). All calculations reported here were performed on $\mathcal{P} = 16$ processors and the accuracy of the parallel algorithm was confirmed by comparing with a few selected single processor calculations.

References

Akaike, H., (1974). *A new look at the statistical model identification*. **IEEE Transactions on Automatic Control** 19:716-723.

Avril, S., Pierron, F., (2007). *General framework for the identification of constitutive parameters from full-field measurements in linear elasticity*. **International Journal of Solids and Structures** 44: 4978-5002.

Avril, S., Bonnet, M., Bretelle, A-S, Grédiac, M., Hild, F., Ienny, P., Latourte, F., Lemosse, D., Pagano, S., Pagnacco, E., Pierron, F., (2008). *Overview of identification methods of mechanical parameters based on full-field measurements*. **Experimental Mechanics** 48: 381.

Bal, G., Uhlmann, G., (2013). *Reconstruction of Coefficients in Scalar Second Order Elliptic Equations from Knowledge of Their Solutions*. **Communications on Pure and Applied Mathematics** 66: 1629-1652.

Bal, G., Bellis, C., Imperiale, S., Monard, F., (2014). *Reconstruction of constitutive parameters in isotropic linear elasticity from noisy full-field measurements*. **Inverse Problems** 30: 125004.

Baldock, R.J.N., (2014). *Classical statistical mechanics with nested sampling*. PhD. Dissertation University of Cambridge (also available in Springer Theses series, 2017).

Baldock, R.J.N., Pártay, L.B., Bartók, A.P, Payne, M.C., Csányi, G., (2016). *Determining pressure-temperature phase diagrams of materials*. **Physical Review B** 93: 174108.

Baldock, R.J.N., Bernstein, N., Salerno, K.M., Pártay, L.B., Csányi, G., (2017). *Constant-pressure nested sampling with atomistic dynamics*. **Physical Review E** 96: 043311.

Bui-Thanh, T., Ghattas, O., Martin, J., Stadler, G., (2013). *A computational framework for infinite-dimensional Bayesian inverse problems. Part I: The linearized case, with applications to global seismic inversion*. **SIAM Journal on Scientific Computing** 35: A2494-A2523.

Bellis, C., (2017). *Converting strain maps into elasticity maps for materials with small contrast*. **Journal of the Mechanics and Physics of Solids** 98: 411-428.

Butler, L.J., Xu, J., He, P., Gibbons, N., Dirar, S., Middleton, C.R., Elshafie, M.Z.E.B., (2017). *Robust fibre optic sensor arrays for monitoring early-age performance of mass-produced concrete sleepers*. **Structural Health Monitoring: An International Journal** https://doi.org/10.1177/1475921717714615.

Chan, D.D., Neu, C.P., (2012). *Transient and microscale deformations and strains measured under exogenous loading by noninvasive magnetic resonance*. **PloS One** 7:e33463.

Good, I.J., (1979). *Studies in the history of probability and statistics. XXXVII AM Turing's statistical work in World War II*. **Biometrika** 66: 393-396.

Grédiac, M., (1989). *Principe des travaux virtuels et identification*. **Comptes rendus de l'Académie des sciences. Série 2, Mécanique, Physique, Chimie, Sciences de l'univers, Sciences de la Terre** 309: 1-5.

Grédiac, M., Pierron, F., (1998). *A T-shaped specimen for the direct characterization of orthotropic materials*. **International Journal for Numerical Methods in Engineering** 41: 293-309.

Grédiac, M., Hild, F., (2012). *Full-field measurements and identification in solid mechanics*. John Wiley & Sons.

Gue, C.Y., Wilcock, M., Alhaddad, M.M. Elshafie, M.Z.E.B., Soga, K., Mair, R.J., (2015). *The monitoring of an existing cast iron tunnel with distributed fibre optic sensing (DFOS)*. **Journal of Civil Structural Health Monitoring** 5:573-586.

Jeffreys, H., (1961). *The theory of probability (3 ed.)*. Oxford: Oxford University Press.

Kiapio, J., Somersalo, E., (2005). *Statistical and computational inverse problems*. Springer-Verlag, New York.

Kass, R.E, Raftery, A.E., (1995). *Bayes factors*. **Journal of the American Statistical Association**, 90:773-795.

Ko, J.M., Ni, Y.Q., (2005). *Technology developments in structural health monitoring of large-scale bridges*. **Engineering Structures** 27:1715-1725.

Koren, Y., (2005). *Drawing graphs by eigenvectors: theory and practice*. **Computers & Mathematics with Applications** 49:1867-1888.

Kragh, H., (1990). *Dirac: a scientific biography*. Cambridge: Cambridge University Press.

Kullback, S., Leibler, R.A., (1951). *On information and sufficiency*. **Annals of Mathematical Statistics** 22:79-86.

MacKay, D.J.C., (2003). *Information theory, inference and learning algorithms*. Cambridge: Cambridge University Press.

Molimard, J., Riche, R.Le, Vautrin, A., Lee, J.R., (2005). *Identification of the four orthotropic plate stiffnesses using a single open-hole tensile test*. **Experimental Mechanics** 45: 404-411.

Nelder, J.A., Mead, R., (1965). *A simplex method for function minimization*. **Computer Journal** 7: 308-313.

Nguyen, M. T., Desceliers, C., Soize, C., Allain, J. M., Gharbi, H., (2015). *Multiscale identification of the random elasticity field at mesoscale of a heterogeneous microstructure using multiscale experimental observations*. **International Journal for Multiscale Computational Engineering** 13: 281-295.

Pártay, L.B., Bartók, A.P, Csányi, G., (2010). *Efficient sampling of atomic configurational spaces*. **The Journal of Physical Chemistry B** 114:10502-10512.

Pártay, L.B., Bartók, A.P, Csányi, G., (2014). *Nested sampling for materials: The case of hard spheres*. **Physical Review E** 89:022302.

Piegl, L., Tiller, W., (1997). *The NURBS Book (2nd ed.)*. Springer-Verlag New York, Inc. New York, NY, USA.

Pierron, F., Zhavoronok, S., Grédiac, M., (2000). *Identification of the through-thickness properties of thick laminated tubes using the virtual fields method*. **International Journal of Solids and Structures** 37:4437-4453.

Raissi, M., Perdikaris, P., Karniadakis, G.E., (2017a). *Inferring solutions of differential equations using noisy multi-fidelity data*. **Journal of Computational Physics** 335, 736-746.

Raissi, M., Karniadakis, G.E., (2017b). *Machine learning of linear differential equations using Gaussian processes*. **arXiv preprint** arXiv:1701.02440.

Richards, M.S., Barbone, P.E., Oberai, A.A., (2009). *Quantitative three-dimensional elasticity imaging from quasi-static deformation: a phantom study*. **Physics in Medicine and Biology** 54:757-779.

Rissanen, J., (1978). *Modeling by shortest data description*. **Automatica** 14:465-471.

Rouger, F., Khebibeche, M., Govic, C.Le., (1991). *Non determined tests as a way to identify wood elastic parameters the finite element approach*. **Mechanical Identification of Composites**. Springer Netherlands, 82-90.

Schwarz, G.G., (1978). *Estimating the dimension of a model*. **The Annals of Statistics** 6: 461-464.

Skilling, J., (2006). *Nested sampling for general Bayesian computation*. **Bayesian Analysis** 1:833-859.

Stuart, A. M., (2010). *Inverse problems: a Bayesian perspective*. **Acta Numerica** 19, 451-559.

Tarantola, A., (2005). *Inverse problem theory and methods for model parameter estimation*. SIAM.

Thurin, J., Brossier, R., Métivier, L., (2017). *An ensemble-transform Kalman filter - full waveform inversion scheme for uncertainty estimation*. **Proceedings of the SEG International Exposition and 87th Annual Meeting** 1307-1313.

Tikhonov, A. N., Arsenin, V.Y., (1977). *Solution of ill-posed problems*. Washington: Winston & Sons.

Vigliotti, A., Deshpande, V., (2018). *Matlab scripts implementing the Nested Sampling algorithm for the estimation of the mechanical property distribution of solids starting from limited and noisy experimental measurements*. Mendeley Data, v1 http://dx.doi.org/10.17632/zm2zgk537f.1.

Wales, D.J., Miller, M.A., Walsh, T.R., (1998). *Archetypal energy landscapes*. **Nature** 394:758-760.

Wales, D.J., Bogdan, T.V., (2006). *Potential energy and free energy landscapes*. **The Journal of Physical Chemistry B** 110: 20765-20776.

Supplementary Information

S1.   Material property distributions available to the different models

One randomly selected distribution of material properties with a $p = 5$ control weight grid is shown in Fig. 2 and that is used as the synthetic experimental specimen in the whole of the study. To give a sense of the wavelengths associated with the different control weight grids here we show in Fig. S1 randomly selected distributions of the Young's modulus with $p = 3, 5$ and 8 control weight grids. These distributions were generated in a manner identical to that used to generate configurations for models $\mathcal{H}_3, \mathcal{H}_5$ and $\mathcal{H}_8$ and thus give a sense of the configurations achievable with these models. Clearly the configurations have longer wavelengths for the variation of E with decreasing p.
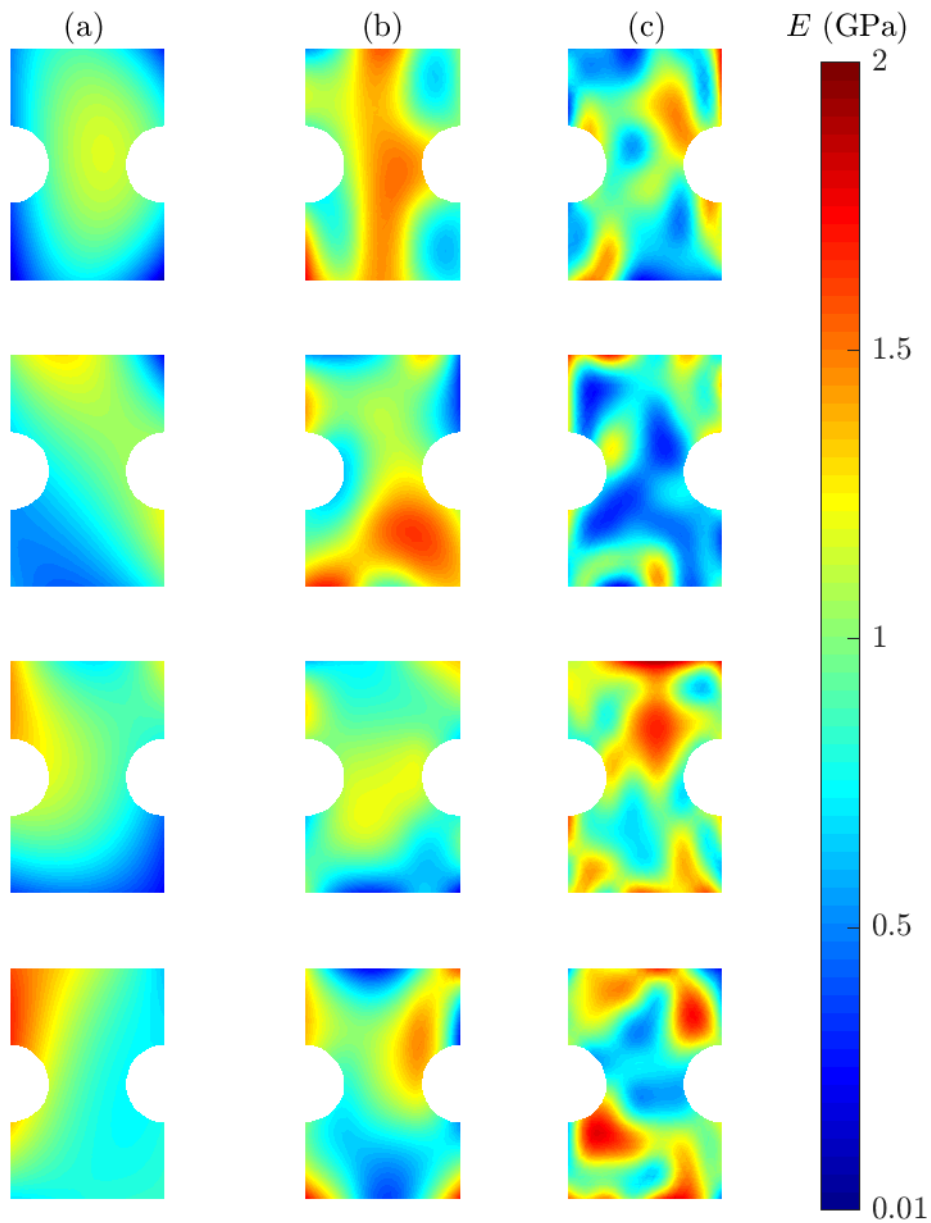
Figure S1: Spatial distributions of Young's modulus E in the specimen generated via control weight grids with (a) $p = 3$, (b) $p = 5$ and (b) $p = 8$. In each case four randomly generated specimens are shown.

S2.   Bayesian predictions of Poisson's ratio distributions

Bayesian predictions for the Poisson's ratio analogous to the Young's modulus distributions shown in Figs. 5 and 6 are presented here. We define the mean of the posterior distribution of the Poisson's ratio as

$$\langle \hat{v}(x_i) \rangle \equiv \frac{1}{\mathcal{Z}_p} \sum_{s=1}^{J} (\omega_s \mathcal{L}_s) \hat{v}_s(x_i), \tag{S1}$$

where $\hat{v}_s(x_i)$ is the spatial distribution of Poisson's ratio in the sample removed in iteration s of the Nested Sampling. Similarly, the standard deviation of the posterior distribution of $\hat{v}$ is given by

$$\Delta_{\hat{v}}(x_i) \equiv \sqrt{\frac{1}{\mathcal{Z}_p} \sum_{s=1}^{J} \omega_s \mathcal{L}_s [\hat{v}_s(x_i) - \langle \hat{v}(x_i) \rangle]^2}, \tag{S2}$$

with the normalised standard deviation defined as

$$\bar{\Delta}_{\hat{v}}(x_i) \equiv \frac{2\sqrt{3}\Delta_{\hat{v}}(x_i)}{v_{\max} - v_{\min}}. \tag{S3}$$

The error

$$\rho_v(x_i) \equiv \frac{|v(x_i) - \langle \hat{v}(x_i) \rangle|}{v(x_i)}, \tag{S4}$$

in the prediction based on the mean of the posterior of the Poisson's ratio is plotted in Fig. S2 while predictions of $\bar{\Delta}_{\hat{v}}$ are included in Fig. S3 for models $\mathcal{H}_5$ and $\mathcal{H}_8$. The overall conclusions discussed for the Young's modulus in the context of Figs. 5 and 6 hold for the Poisson's ratio too.
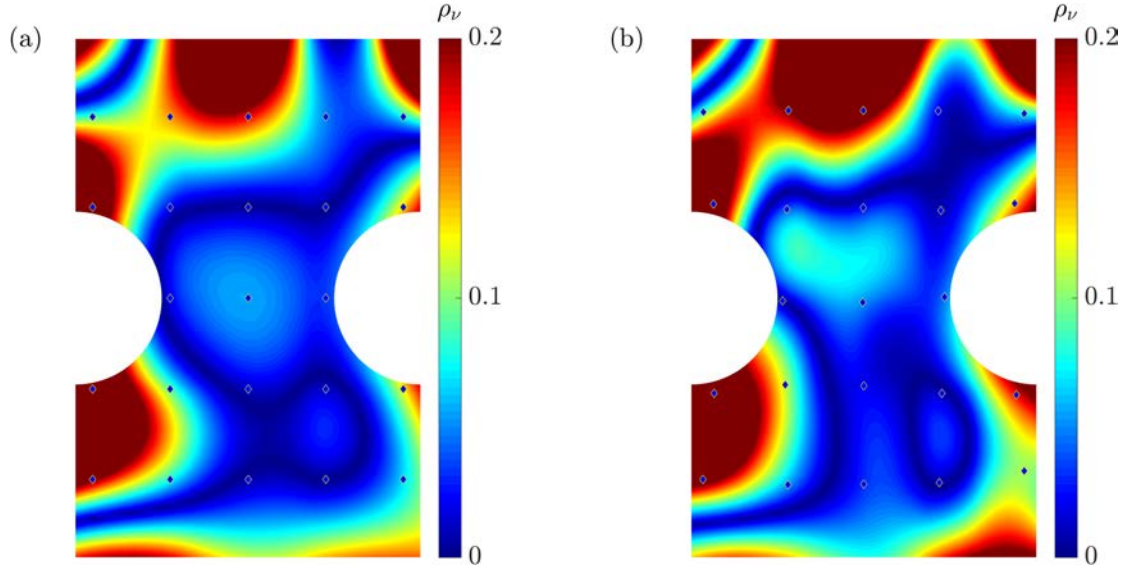
Figure S2: Spatial distributions of the error ρ_v in predictions of the Poisson's ratio using the mean of the posterior distribution for (a) model $\mathcal{H}_5$ and (b) model $\mathcal{H}_8$. The predictions are for the reference case of the experiment with $N = 23$ strain gauges and an experimental noise $\sigma_{\mathrm{err}} = 0.016\varepsilon_{\mathrm{nom}}^{\infty}$. The strain measurement locations are indicated in both (a) and (b).
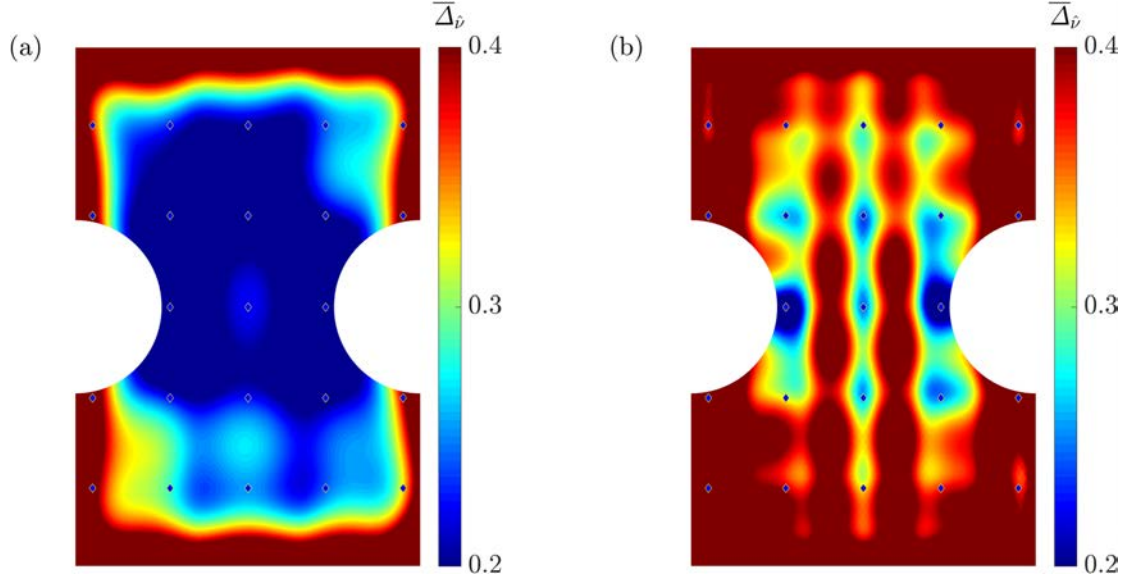
Figure S3: Spatial distributions of the uncertainty in the predictions as quantified normalised standard deviation $\bar{\Delta}_{\hat{\nu}}$ of the predictions of the Poisson's ratio for (a) model $\mathcal{H}_5$ and (b) model $\mathcal{H}_8$. The predictions are for the reference case of the experiment with $N = 23$ strain gauges and an experimental noise $\sigma_{\mathrm{err}} = 0.016\varepsilon_{\mathrm{nom}}^{\infty}$. The strain measurement locations are indicated in both (a) and (b).
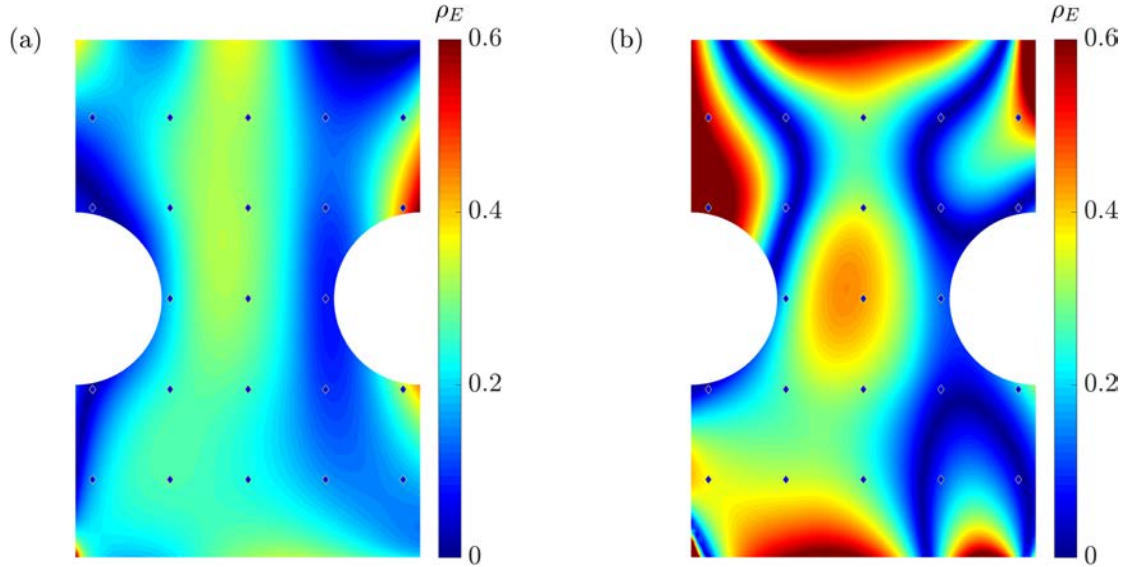
Figure S4: Spatial distributions of the error ρ_E in predictions of the Young's modulus using maximum likelihood solutions for (a) model $\mathcal{H}_5$ and (b) model $\mathcal{H}_8$. The predictions are for the reference case of the synthetic experiment with $N = 23$ strain gauges and measurement noise $\sigma_{\mathrm{err}} = 0.016\varepsilon_{\mathrm{nom}}^{\infty}$. This is the same data as in Fig. 7 but re-plotted on a different scale.

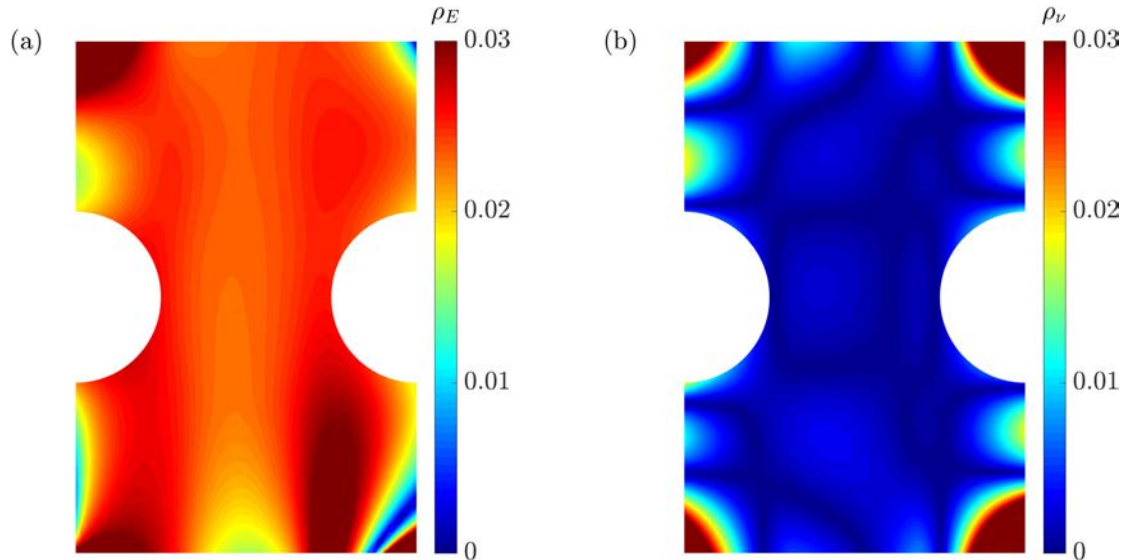## S3. Errors in the maximum likelihood predictions of the Young's modulus

The error distributions ρ_E in Fig. 7 are shown on a scale equal to that in Fig. 5 to illustrate the differences in the magnitudes of the errors between predictions based on the mean of the posterior and the maximum likelihood. This obscures information on ρ_E in Fig. 7 as $\rho_E > 0.2$ over most of the domain. Here in Fig. S4 we include distributions of ρ_E using a larger scale (i.e. same data as in Fig. 7 but re-plotted on a

different scale). The error ρ_E from a maximum likelihood solution is approximately three times larger than the corresponding Bayesian predictions (Fig. 5).

S4. Maximum likelihood predictions in the absence of measurement noise

Predictions of the spatial distributions of the error ρ_E in the Young's modulus and Poisson's ratio ρ_ν for the maximum likelihood solution of model $\mathcal{H}_5$ using the $N = 23$ dataset with no measurement noise ($\sigma_{\mathrm{err}} = 0$) are included in Figs. S5a and S5b, respectively. In contrast to the large errors when 1.6% measurement noise is present (Fig. S4a) the errors in both the Young's modulus and Poisson's ratio are now less than about 3% throughout the specimen. This clearly shows that the measurement noise is amplified due to overfitting of the maximum likelihood solution (even with the correct model) to the noisy data. Of course, even with no measurement noise the errors can be high in the maximum likelihood solutions that employ overly complex models (These models will give multiple solutions with equal maximum likelihoods and while the errors will be low for at-least of those of those solutions, they can be very high for the other solutions.)

Figure S5: Spatial distributions of the errors (a) ρ_E and (b) ρ_ν of the maximum likelihood predictions of the Young's modulus and Poisson's ratio, respectively for model $\mathcal{H}_5$. The dataset with $N = 23$ gauges and no measurement noise ($\sigma_{\mathrm{err}} = 0$) was used in these predictions. Note the significantly smaller scale in (a) and (b) compared to other equivalent figures.

S5. Tikhonov-type regularisation

Tikhonov regularisation is the most commonly used method for regularisation of ill-posed optimisation problems. In the context of the elasticity problem analysed here, the forward problem of predicting strains from known material properties is equivalent to a relatively low pass filter such that errors in the material properties are not significantly amplified in the predicted strains. However, the inverse mapping operates as a high-pass filter that has the effect of amplifying noise in the strain measurements as seen in Section 4.3. The goal of regularisation is to add a term to the likelihood function so as to give preference to particular types of solutions and in Tikhonov regularisation typically preference is given to solutions with smaller norms (e.g. Euclidean norms,

variance etc.). Given that in the problem analysed here, our prior belief that the Young's modulus and Poisson's ratio are uniformly distributed over the ranges $(E_{\min}, E_{\max})$ and $(\nu_{\min}, \nu_{\max})$, respectively we define a regularised likelihood function

$$\hat{\mathcal{L}}^* = \hat{\mathcal{L}} - \varsigma \hat{\mathcal{L}}^T, \tag{S5}$$

where

$$\hat{\mathcal{L}}^T = \frac{1}{2\sqrt{p}} \left(\frac{\left\| \tilde{E} - \bar{E} \right\|_2}{\bar{E}} + \frac{\left\| \tilde{\nu} - \bar{\nu} \right\|_2}{\bar{\nu}} \right), \tag{S6}$$

and $\|\cdot\|_2$ denotes the Euclidean norm. Here, $\bar{E} \equiv (E_{\min} + E_{\max})/2$ and $\bar{\nu} \equiv (\nu_{\min} + \nu_{\max})/2$ are the means of the priors and ς is the Tikhonov regularisation parameter. The regularised optimisation problem then reduces to maximising $\hat{\mathcal{L}}^*$ and here this was performed using a Nelder-Mead (Nelder and Mead, 1965) scheme with the solution that maximises $\hat{\mathcal{L}}$ as the starting guess to the optimisation.

The predictions of this optimisation are included in Fig. S6 where distributions of the Young's modulus (model $\mathcal{H}_5$ with $N = 23$ strain gauges and a measurement noise $\sigma_{\mathrm{err}} = 0.016\varepsilon_{\mathrm{nom}}^\infty$) plotted for selected values of ς. In Fig. S6 the left-hand column shows spatial distributions of the error ρ_E in the predictions while the right-hand column shows the modulus distributions. For small levels of regularisations with $\varsigma \le 1$ the solution is largely unaffected by the regularisation term. By contrast, at larger values of ς the predicted modulus distributions are uniform as $\hat{\mathcal{L}}^*$ is dominated by the regularisation term $\hat{\mathcal{L}}^T$ and hence the solution is insensitive to the strain measurements. In fact, there is no intermediate value of ς where the regularisation term improves the quality of predictions. This is also true for the predictions of Poisson's ratio and these results are omitted here for the sake of brevity. We also attempted a variety of other forms of $\hat{\mathcal{L}}^T$ but were unable to find a regularisation term that improved the maximum likelihood predictions. This is traced back to the fact that the underlying model is already highly regularised and has the appropriate spatial wavelengths of the material property distributions. Thus, further smoothing/regularisation has no beneficial effect.
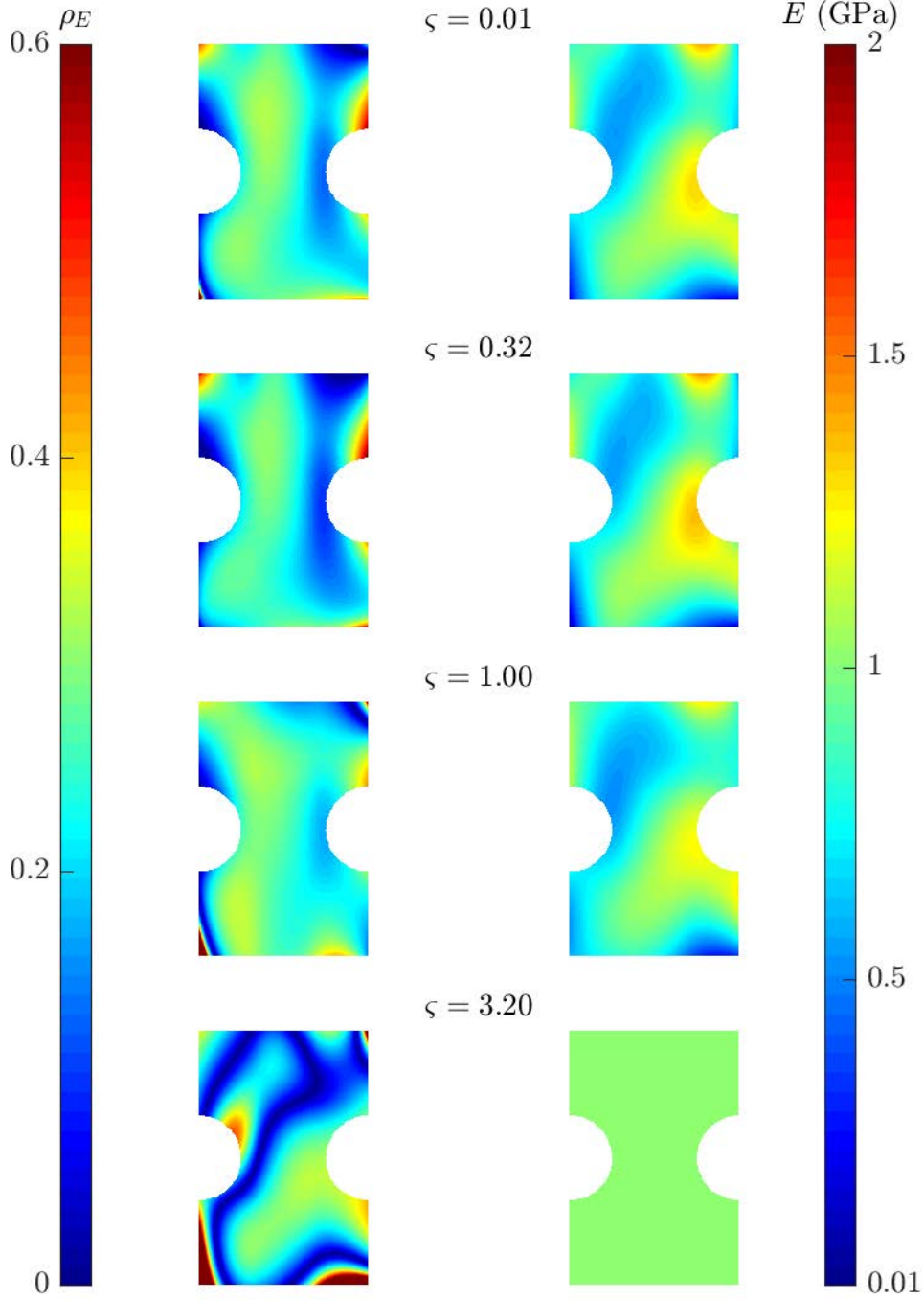
Figure S6: Optimisation predictions of the error ρ_E (left column) and Young's modulus distributions (right column) using a Tikhonov-type regularised likelihood function. Predictions are shown for selected values of the Tikhonov regularisation parameter ς for model $\mathcal{H}_5$ with $N = 23$ strain gauges and a measurement noise $\sigma_{\text{err}} = 0.016\varepsilon_{\text{nom}}^{\infty}$.

S6.   Information gained from the strain measurements

In Bayesian inference, the Kullback-Leibler divergence D_{KL} is a measure of the information gained when we revise our belief from the prior probability distribution $P(\boldsymbol{w}|\mathcal{H}_p)$ to the posterior distribution $P(\boldsymbol{w}|\mathcal{D}, \mathcal{H}_p)$ after the arrival of the data $\mathcal{D}$. The Kullback-Leibler divergence or information gain in units of Hart is defined as

$$D_{KL} = \int P(\boldsymbol{w}|\mathcal{D}, \mathcal{H}_p) \log_{10} \left[\frac{P(\boldsymbol{w}|\mathcal{D}, \mathcal{H}_p)}{P(\boldsymbol{w}|\mathcal{H}_p)} \right] d\boldsymbol{w}, \qquad \text{(S7)}$$

which upon using (2.4) and (2.5) reduces to

$$D_{KL} = -\log_{10}(\mathcal{Z}_p) + \int P(\boldsymbol{w}|\mathcal{D}, \mathcal{H}_p) \log_{10}[P(\mathcal{D}|\boldsymbol{w}, \mathcal{H}_p)] d\boldsymbol{w}. \qquad \text{(S8)}$$

Recalling that the likelihood $\mathcal{L} \equiv P(\mathcal{D}|\boldsymbol{w}, \mathcal{H}_p)$, D_{KL} as given by (S6) is obtained from the output of the Nested Sampling via the sum

$$D_{KL} = -\log_{10}(\mathcal{Z}_p) + \frac{1}{\mathcal{Z}_p} \sum_{s=1}^{J} (\omega_s \mathcal{L}_s) \log_{10}(\mathcal{L}_s). \qquad \text{(S9)}$$

The information gain in units of dHart is plotted in Fig. S7 for models $\mathcal{H}_3, \mathcal{H}_5$ and $\mathcal{H}_8$ as a function of the number N of strain measurement locations for the case of the data with experimental noise $\sigma_{\text{err}} = 0.016\varepsilon_{\text{nom}}^{\infty}$. Of course, the information gain increases with increasing amount of data N but the rate of increase in D_{KL} decreases with increasing N as additional strain data adds increasingly smaller amounts of information to determine the model parameters. This decrease in $\partial D_{KL}/\partial N$ with increasing N is most pronounced for the models with the least number of parameters.

The plateauing out of D_{KL} with increasing N suggests that obtaining further data by measuring strains on increasingly finer grids may not add significant information for solving the inverse problem investigated here. This additional data might in fact be detrimental if it were obtained by changing to a measurement system with higher noise (measurement noise is amplified while solving the inverse problem; see Section 4.3).
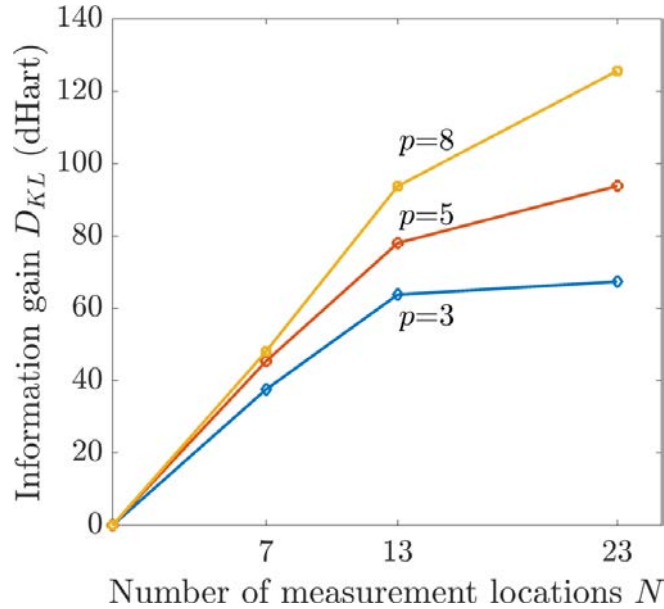
Figure S7: Predictions of information gain D_{KL} as a function of the number N of strain measurement locations for models $\mathcal{H}_3, \mathcal{H}_5$ and $\mathcal{H}_8$ (strain data with $\sigma_{\text{err}} = 0.016\varepsilon_{\text{nom}}^{\infty}$). The lines are labelled by the model order p.