CAMBRIDGE
UNIVERSITY PRESS

RESEARCH ARTICLE

A hierarchical Bayesian approach for calibration of stochastic material models

Nikolaos Papadimas[1,*] and Timothy Dodwell[1,2]

[1]Institute of Data Science and AI, University of Exeter, Devon, United Kingdom
[2]The Alan Turing Institute, The British Library, London, United Kingdom
*Corresponding author. E-mail: np380@exeter.ac.uk

Abstract

This article recasts the traditional challenge of calibrating a material constitutive model into a hierarchical probabilistic framework. We consider a Bayesian framework where material parameters are assigned distributions, which are then updated given experimental data. Importantly, in true engineering setting, we are not interested in inferring the parameters for a single experiment, but rather inferring the model parameters over the population of possible experimental samples. In doing so, we seek to also capture the inherent variability of the material from coupon-to-coupon, as well as uncertainties around the repeatability of the test. In this article, we address this problem using a hierarchical Bayesian model. However, a vanilla computational approach is prohibitively expensive. Our strategy marginalizes over each individual experiment, decreasing the dimension of our inference problem to only the hyperparameter—those parameter describing the population statistics of the material model only. Importantly, this marginalization step, requires us to derive an approximate likelihood, for which, we exploit an emulator (built offline prior to sampling) and Bayesian quadrature, allowing us to capture the uncertainty in this numerical approximation. Importantly, our approach renders hierarchical Bayesian calibration of material models computational feasible. The approach is tested in two different examples. The first is a compression test of simple spring model using synthetic data; the second, a more complex example using real experiment data to fit a stochastic elastoplastic model for 3D-printed steel.

Impact Statement

In this article, we provide a hierarchical Bayesian methodology for calibrating mathematical models of complex materials. Importantly, this approach naturally includes experimental data from a set of (potentially different) tests. Using the hierarchical approach, we are able to quantify the population statistics for uncertainties arising from variations of the test, measurement uncertainty, and model misspecification. Our approach is tested on two different test cases, including a real world application for 3D-printed steel, proving both its efficiency and versatility.

1. Introduction

In engineering and across the physical sciences, both in academia and industry, the parameterization of constitutive models for materials from a set of experiments is a very common research question.

CrossMark

Parameterization of constitutive models is often used to classify or compare the response of different materials, or as the starting point for establishing mathematical models of the system, in which that material is used such as finite element analysis. The standard approach is to carry out a series of experiments. For each, the parameters which minimize the squared difference between the experiment's output and the model response are found. This procedure is often called the least squared method (LSM). The population statistics of the derived model parameters over all experimental tests are then computed, and often limited to a simple mean and standard derivation of the samples.

Although very common, the LSM approach requires a sufficient number of experimental tests to establish "good" estimates. This makes it hard to establish good estimates for the distribution of parameters, particularly, if experimental data are limited. Although various contributions have enhanced the LSM method (Charnes et al., 1976), statistical information about the measurement uncertainty in the experiment and the model misspecification (i.e., "all models are wrong") are generally not captured. The simple averaging approach to establishing population statistics, and even more advanced assumptions like taking into account covariances between parameters, also lose much information about the population distribution of the parameters and their often complex interdependencies.

Bayesian model calibration is a popular framework for estimating the values of input parameters into computational simulations in the presence of multiple uncertainties (Kennedy and O'Hagan, 2001; Bayarri et al., 2007), and provides a formalized way to include prior knowledge of parameters to support cases where there are limited measurement data. They have been applied broadly across many area of science in engineering, healthcare, and ecology (Solonen et al., 2012; Thomas et al., 2015; Valderrama-Bahamóndez and Fröhlich, 2019). Although less common in engineering, Bayesian approaches have also been used to calibrate material models from experimental data. The first such use, to the best of the authors' knowledge, was by Isenberg (Isenberg, 1979),who used it to identify elastic material parameters for a simple model and experimental test. It is only relatively recently that Bayesian-based approaches have been used to parameterize models: elastic constants of composite and laminate plates (Gogu et al., 2013), time-dependent material models in viscoelasticity (Rappel et al., 2018), and large strain elasto-plastic models for 3D-printed steel (Dodwell et al., 2021), to give just a few examples. Beyond the straight probabilistic calibration of material parameters based on a single experiment, Bayesian approaches have been used to access the relatively predictive quality over a family of possible models, so-called *Bayesian Model Selection* (Chipman et al., 2001; Johnson and Rossell, 2012). Such approaches have had particular impact in a number of computational material science applications, since there are often many choices of possible material model. A Bayesian model selection approach formalizes the material model selection, balancing both prior knowledge and evidence from experimental data. Notable examples include hyperelastic constitutive models for tissue (Madireddy et al., 2015), phenomenological models for tumour growth (Oden et al., 2013), models of damage progression in composites due to fatigue (Chiachío et al., 2015), and also for fatigue descriptions of metals (Babuška et al., 2016).

The Bayesian approaches in the literature have focused on the calibration of a material model from a single experiment. In many engineering applications, it is the distribution of possible model parameters derived from a set of experiments, rather than individual experiments, that is of interest. The set of experiments may be generated from different samples of material and tested under different conditions, or multiple tests are repeated to explore the variability in the material and/or the measurement uncertainty in the test itself. If we consider a setting where we have performed J experiments of the same material and ψ_j represents the model parameters which describe the jth experiment. It is expected that the parameters ψ_j and parameters for another experiment ψ_i will be positively correlated, since they offer model parameterizations for the same material, albeit in different instances of the same physical experiment. A natural way to build this dependency into a Bayesian setting is to define prior distributions for our material parameters, and see ψ_j is seen as themselves samples from a common population distribution over the set of all possibly observable experiments. This population distribution itself can be parameters ϕ (referred to as *hyper-parameters*)—which might are typical population statistics (e.g., mean, variance, or higher moments). In a Bayesian setting, the hyper-parameters themselves are also equipped with a probabilistic specification, and with that they assign themselves prior distributions, *hyper-priors*. This gives a natural

hierarchical representation of multiexperimental parameterization problem. This framework defines a *Hierarchical Bayesian Model* (HBM) (Gelman et al., 2013; Figure 1). HBMs have been widely used in statistical applications (Kwon et al., 2008; Loredo, 2013; Wolfgang, 2016; Bozorgzadeh et al., 2019). Probably, most notably, they have been used in statistics models for drug trials to determine population statistics in response to a new drug when the data received are unique to an individual patient (Babapulle et al., 2004; Huang et al., 2006).

As far as the authors are aware, this article sets out the first hierarchical Bayesian framework for the parameterization of material models from a set of experimental data. It starts by recapping the calibration of a material model from a single experimental test (Section 2.1), and then extends the approach to a full set of data (Section 2.2). An efficient computational strategy for ensuring the resulting Markov Chain Monte Carlo (MCMC) computations for the HBM remains feasible and is presented. This uses a simple trick to marginalize all realization of the model parameters for a particular experiment in an offline stage. Online, during MCMC sampling of the hyperparameters, an approximate likelihood can be efficiently estimated. The uncertainty in this approximation is captured using Bayesian Quadrature (O'Hagan, 1991; Osborne et al., 2012) to approximate the marginalization integral. The set of approaches is demonstrated for two examples. The first is a simple analytical model for the calibration of a linear spring, and the second an application of the methodology to a more complex nonlinear hyper-elastic model solved with the help of Finite Element Method, in which real experimental data are used (Gardner et al., 2020).

The results demonstrate that a hierarchical Bayesian framework is a natural setting for the calibration of material constitutive in the model from a set of experiment tests. The approach addresses the limitations of standard, nonprobabilistic approaches (e.g., least square) and enables the capture of measurement, model uncertainties, and complex interdependencies between model parameters, including prior knowledge in cases where data are limited. It provides a framework, in which the population statistics of the material parameters can be estimated, rather than snapshots of individual experiments under the assumption, they are independent and correlation between parameters is not significant.

2. A HBM for Estimating Material Parameters from a set of Experiments

We start this section by first recapping the parameterization of a material model from a single experimental test. This single experiment approach is then extended into a hierarchical Bayesian framework (Gelman et al., 2013), in which we infer the population statistics of the material parameters from a set of experimental tests. It then considers computationally efficient strategies for implementing the hierarchical approach.
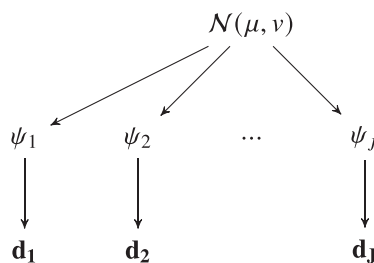

Figure 1. *Model parameters ψ_i can be calibrated to a specific experimental dataset $\mathbf{d}_i$. If repeated over J experiments $\{1, \ldots, J\}$, natural variations in each ψ_i would be observed. In this contribution, a hierarchical Bayesian framework is developed which learns the population distribution of the ψ_i's. If this population distribution is parameterized by some hyper-parameters (e.g., in this figure, we assume $\phi_i \sim \mathcal{N}(\mu, v)$, which are themselves equipped with prior distribution, reflecting uncertainty in their values and representation.*

2.1. Bayesian calibration of a material model given a single experiment

We first set up a Bayesian formulation, from which we can estimate the distribution of material parameters given that we observe the outputs of only a single experiment. Let $\boldsymbol{\psi} = (\psi_1, \ldots, \psi_m)^T \in \boldsymbol{\psi} \subset \mathbb{R}^m$ denote a vector of random parameters which define the parameter space $\boldsymbol{\Psi}$ of a particular material model. It is assumed that the material parameters have a prior distribution defined by the density $\pi_0(\boldsymbol{\psi}) : \boldsymbol{\psi} \to \mathbb{R}^+$.

Let the recorded outputs of an experimental test be defined by the vector $\mathbf{d} = (d_1, \ldots, d_s)^T \in \mathbb{R}^s$. These, for example, could be the load measurements to achieve a given displacement by the test machine, or the strain measurement from strain sensors or digital image correlation software. Let $\mathbf{f}(\boldsymbol{\psi}) : \boldsymbol{\Psi} \to \mathbb{R}^s$ be a *parameter-to-observable* map. The connection between model and data is then established by defining a statistical model:

$$\mathbf{d} = \mathbf{f}(\boldsymbol{\psi}) + \bar{\varepsilon}, \tag{1}$$

where $\bar{\varepsilon}$ represents the measurement or observational noise. We assume that this noise is distributed so that $\bar{\varepsilon} \sim \mathcal{N}(\mathbf{0}, \Sigma_{\bar{\varepsilon}})$, where $\Sigma_{\bar{\varepsilon}} \in \mathbb{R}^{s \times s}$ is a symmetric-positive definite covariance matrix.

The aim in material model calibration is to estimate the posterior distribution of the model parameters, that is the distribution of parameters $\boldsymbol{\psi}$ given the observed experimental data. Under the assumption that $\boldsymbol{\psi}$ and $\bar{\varepsilon}$ are independent random variables, Baye's theorem gives:

$$\pi(\boldsymbol{\psi}|\mathbf{d}) := \frac{1}{Z}\pi(\mathbf{d}|\boldsymbol{\psi})\pi_0(\boldsymbol{\psi}) \quad \text{such that} \quad Z := \int_{\boldsymbol{\psi}} \pi(\mathbf{d}|\boldsymbol{\psi})\pi_0(\boldsymbol{\psi})d\boldsymbol{\psi}. \tag{2}$$

The conditional distribution $\pi(\mathbf{d}|\boldsymbol{\psi})$ defines the likelihood, that is the probability of observing the data d given a set of material parameters $\boldsymbol{\psi}$. Having define the statistical model (1), it follows that:

$$\log \pi(\mathbf{d}|\boldsymbol{\psi}) = \frac{1}{2}(\mathbf{d} - \mathbf{f}(\boldsymbol{\psi}))^T \Sigma_{\bar{\varepsilon}}^{-1}(\mathbf{d} - \mathbf{f}(\boldsymbol{\psi})). \tag{3}$$

In this setting, we have a classic Bayesian inverse problem for which there are well-established methods in computational statistics for solving. Having said this, the estimation of the posterior distribution of model parameters, even for a single experiment, can be extremely challenging. This is particularly the case when: (a) the evaluation of the parameters-to-observable map $\mathbf{f}(\boldsymbol{\psi})$ are computationally expensive, (e.g., a nonlinear finite element model); (b) the material parameter space is high-dimensional (i.e., m is large); or (c) the posterior density has a complex geometry, for example, when multiple parameters have complex nonlinear dependencies. In this article, we use MCMC methods to perform the Bayesian computations and therefore briefly set out the detail here, but they are widely covered in the literature (Liu, 2008; Gelman et al., 2013).

In this contribution, we focus on Metropolis-Hasting-based MCMC algorithms (Roberts and Rosenthal, 2009), where we generate a chain of n samples $\mathcal{T} := \{\boldsymbol{\psi}^{(1)}, \boldsymbol{\psi}^{(2)}, \ldots, \boldsymbol{\psi}^{(n)}\}$ of statistically dependent sets of parameters. The chain is produced from a current state $\boldsymbol{\psi}^{(i)}$, by making a proposal $\boldsymbol{\psi}'$ generated from a known proposal distribution $q(\boldsymbol{\psi}'|\boldsymbol{\psi}^{(i)})$ (Roberts, 2011). This new proposal is accepted as the next state in the chain, that is $\boldsymbol{\psi}^{(i+1)} = \boldsymbol{\psi}'$ with a probability:

$$\alpha\left(\boldsymbol{\psi}'|\boldsymbol{\psi}^{(i)}\right) = \min\left(1, \frac{\pi(\mathbf{d}|\boldsymbol{\psi}')\pi_0(\boldsymbol{\psi}')q(\boldsymbol{\psi}'|\boldsymbol{\psi}^{(i)})}{\pi(\mathbf{d}|\boldsymbol{\psi}^{(i)})\pi_0(\boldsymbol{\psi}^{(i)})q(\boldsymbol{\psi}^{(i)}|\boldsymbol{\psi}')}\right), \tag{4}$$

otherwise it is rejected, and $\boldsymbol{\psi}^{(i+1)} = \boldsymbol{\psi}^{(i)}$. The distribution of the complete chain $\mathcal{T}$ converges to the desired posterior $\pi(\boldsymbol{\psi}|d)$. Although converge can be slow, computations are independent of estimating the normalizing constant Z. In particular, the proposal distributions $q(\cdot|\cdot)$ can be carefully designed to be scalable in high-dimensional parameters spaces, and asymptotically provide an exact (free from bias) approximation of the posterior.

Once a Markov chain $\mathcal{T}$ is generated, the expectations of a function $h(\boldsymbol{\psi})$ can be estimated by Monte Carlo integration:

$$\mathbb{E}_{\pi(\boldsymbol{\psi}|\mathbf{d})}(h) = \int_{\boldsymbol{\Psi}} h(\boldsymbol{\psi}) \pi(\boldsymbol{\psi}|\mathbf{d}) d\boldsymbol{\psi} \approx \widehat{h} := \frac{1}{n} \sum_{i=1}^{n} h\left(\boldsymbol{\psi}^{(i)}\right). \tag{5}$$

The computed value $\widehat{h}$ is an unbiased estimate for the expectation of $h(\boldsymbol{\psi})$ under the posterior distribution, and has a mean square error of:

$$\mathbb{V}\left(\widehat{h}\right) = \frac{\mathbb{V}(h)}{ESS(h)} = \frac{2\mathbb{V}(h)\tau(h)}{n}, \tag{6}$$

where $\tau(h)$ is the integrated autocorrelation of $h(\boldsymbol{\psi})$. The function $ESS(h)$ is the *effective samples size* of the chain, and gives the approximate number of independent samples from the set of correlated samples generated by a Markov chain, and $\mathbb{V}(.)$ represents the variance of distribution, following the notation in (Kong, 1992) and (Kong et al., 1994).

2.2. Hierarchical Bayesian calibration of a material model from a set of experiments

In practice, engineers are seldom interested in finding the material parameters which best fit a single experimental test, but rather in inferring the population distribution of the parameters given a complete experimental dataset. Let us assume J independent experimental tests are performed, which gives the complete dataset:

$$\mathcal{D} := \left\{ \mathbf{d}^{(1)}, \mathbf{d}^{(2)} \dots, \mathbf{d}^{(J)} \right\}. \tag{7}$$

Here, $\mathbf{d}^{(j)} \in \mathbb{R}^{s_j}$ denotes the vector of data for the jth experiment. As for the single experimental formulation, we introduce the *parameter-to-observable* map for each experimental test $\mathscr{F}_j(\boldsymbol{\psi}) : \boldsymbol{\Psi} \to \mathbb{R}^{s_j}$. We note, in particular, that in the formulation which follows, there is no requirement that the experimental tests are the same, this is demonstrated in a computational example (Section 3.2).

We formulate our problem as a HBM (Gelman et al., 2013; Wolfgang, 2016). For each of the J experiments, each experiment has its own data $\mathbf{d}^{(j)}$, and if a single Bayesian calibration was performed (as described above), each could generate its own posterior distribution of material parameters, that is $\pi\left(\boldsymbol{\psi}^{(j)}|\mathbf{d}^{(j)}\right)$. Such an approach neglects the connection between each experiment that they are different experiments on the same material. It is therefore natural to view each $\boldsymbol{\psi}^{(j)}$ as a sample from some population distribution. We therefore define a statistical model for the population distribution, parameterized by additional parameters $\boldsymbol{\phi} \in \mathbb{R}^S$. In a hierarchical Bayesian setting, these are referred as *hyper-parameters*, and in a Bayesian setting, are themselves uncertain and equipped with their own prior densities $\pi(\boldsymbol{\phi})$, termed as *hyper-priors.*

The Bayesian task is, therefore, not to directly infer the distribution of model parameters for each experiment ($\boldsymbol{\psi}_j$), but rather the distribution of hyper-parameter $\boldsymbol{\phi}$ given our complete set of data $\mathcal{D}$, that is to compute samples from $\pi(\boldsymbol{\phi}|\mathcal{D})$. Assuming that each experiment is independent it follows that:

$$\pi(\boldsymbol{\phi}|\mathcal{D}) = \prod_{j=1}^{J} \pi\left(\boldsymbol{\phi}|\mathbf{d}^{(j)}\right) = \prod_{j=1}^{J} \int_{\boldsymbol{\Psi}} \pi\left(\boldsymbol{\phi}, \boldsymbol{\psi}|\mathbf{d}^{(j)}\right) d\boldsymbol{\psi} \propto \prod_{j=1}^{J} \int_{\boldsymbol{\Psi}} \pi\left(\mathbf{d}^{(j)}|\boldsymbol{\phi}, \boldsymbol{\psi}\right) \pi(\boldsymbol{\phi}, \boldsymbol{\psi}) d\boldsymbol{\psi}. \tag{8}$$

For each experiment, we note that the likelihood $\pi\left(\mathbf{d}^{(j)}|\boldsymbol{\phi}, \boldsymbol{\psi}\right)$ is not directly dependent on $\boldsymbol{\phi}$, and hence the dependency is dropped: $\pi\left(\boldsymbol{\psi}|\mathbf{d}^{(j)}\right)$. The connection between model parameters $\boldsymbol{\psi}$ and hyper-parameters $\boldsymbol{\phi}$ is therefore through their joint distribution $\pi(\boldsymbol{\phi}, \boldsymbol{\psi}) = \pi(\boldsymbol{\psi}|\boldsymbol{\phi})\pi(\boldsymbol{\phi})$. It hence follows that:

$$\pi(\boldsymbol{\phi}|\mathcal{D}) \propto \left[\prod_{j=1}^{J} \int_{\boldsymbol{\Psi}} \pi\left(\mathbf{d}^{(j)}|\boldsymbol{\psi}\right) \pi(\boldsymbol{\psi}|\boldsymbol{\phi}) d\boldsymbol{\psi} \right] \pi(\boldsymbol{\phi}). \tag{9}$$

If the expression (9) is compared to Baye's formulae, we therefore identify:

$$\mathcal{L}(\phi|\mathcal{D}) := \prod_{j=1}^{J} \int_{\Psi} \pi\left(\mathbf{d}^{(j)}|\psi\right) \pi(\psi|\phi) d\psi, \tag{10}$$

as being proportional to the likelihood. This is then an integrated likelihood, which marginalizes out model parameters ψ. We can therefore proceed with an MCMC calculation in the space of hyper-parameters ϕ, replacing the likelihood directly with $\mathcal{L}$.

For all but the simplest of models, the likelihood within (9) cannot be analytically calculated, and hence must be approximated. On the face of it, this looks like an extremely expensive task since, in general, for each of the examples considered in this article, the evaluation of likelihood for an experiment $\pi\left(\mathbf{d}^{(j)}|\psi\right)$ is computationally expensive (a nonlinear finite element model) and the number of experiments J maybe large. Yet, a key observation is that the expensive part of the calculation, the evaluation of $\pi\left(\mathbf{d}^{(j)}|\psi\right)$, is independent of ϕ. The component dependent on ϕ, $\pi(\psi|\phi)$, is analytical evaluation of a density (in our examples, a multivariant Gaussian). Therefore, strategies to approximate $\pi\left(\mathbf{d}^{(j)}|\psi\right)$ can be done prior to any sample of the hyper-parameters in an "offline" step.

There are various approaches to approximate the integral $\mathcal{L}$, and which is optimal will be problem dependent. In the examples considered in this article, we choose (before sampling) to build a Gaussian process (GP) emulator to replace the expensive mapping $\psi \to \pi(\boldsymbol{d^j}|\psi)$. The approximation of likelihood using statistical emulators is described by (Rasmussen, 2003), see also (Fielding et al., 2011). As is done this contribution, the emulator of the likelihood is trained on a set of space-filling samples from the input space. Although not considered here, it is possible to consider an additional step, whereby a MCMC computation with a specific accept–reject step is carried out to improve the likelihood estimation, by ensuring samples come from the posterior distribution. Where a prior/initial parameter range is poorly specified, it is easy to see this could have a big effect of the quality of the emulator of likelihood, particularly in the case where very sparse evaluations of the model are possible due to computational constraints. This additional step in the article has not being consider, since via cross-validation, the quality of the approximations of the likelihood seem sufficient. However, this additional step could be readily added to the computational pipeline, with no methodological effect of the results presented in this contribution.

Since we have an emulator precomputed for $\pi(\boldsymbol{d^j}|\psi)$, it is natural to use Bayesian quadrature for the marginalization integrals in (10). Bayesian quadrature (Paleyes et al., 2019) gives a mean ($\mu_{\mathcal{L}}$) and standard deviation ($s_{\mathcal{L}}$) for integral $\mathcal{L}(\phi)$. For the MCMC step (in the hyperparameter space), a sample $\mathcal{L} = \mu_{\mathcal{L}} + s_{\mathcal{L}}\omega$ where $\omega \sim \mathcal{N}(0,1)$, can be drawn. By doing so, the MCMC traces will also account for the uncertainty in the approximation of (10). Details of Bayesian Quadrature implemented in this article can be provided in classic contribution (O'Hagan, 1991). Since also a Monte Carlo approximation can be used, without any changes to the computational framework presented, the Bayesian Quadrature step is not central to the core idea of this contribution, and we do not provided further details.

3. Demonstrative Examples

In this section, we demonstrate the hierarchical Bayesian framework set out with two computational examples. The first is a simple model for the calibration of an elastic spring, the second is an elastoplastic model for 3D-printed steel using real experimental data.

3.1. Example 1: spring model

In this first example, we consider a simple toy model to demonstrate the methodology presented before considering a much more computational intensive example in Section 3.2 based on a nonlinear finite element model. In this case, we wish to calibrate a simple linear Hookes law model for the load response of the springs under deformation. Due to the manufacturing process, material variability, and measurement

uncertainty, the springs display a distribution of responses. To understand the resulting distribution of model parameter, the behavior of N independent springs are tested by compressing each spring by an increasing displacement x_i and the force y_i for $i = 1, ..., M$ to achieve each displacement, which is recorded. From these experiments, we obtain the complete set of data:

$$\mathcal{D} := \left\{ \mathbf{d}^{(1)}, ..., \mathbf{d}^{(N)} \right\} \quad \text{where} \quad \mathbf{d}^{(j)} \in \mathbb{R}^M. \tag{11}$$

In this example, the data are generated synthetically, by the (nonlinear) spring model given by:

$$f_i^{(j)} = c_1 x_i + c_2 x_i^\alpha + \bar{\varepsilon}, \quad \text{so that} \quad \mathbf{d}^{(j)} = \left[f_1^{(j)}, ..., f_M^{(j)} \right]^T, \tag{12}$$

where $c_1 \sim \mathcal{N}^+(1.1, 0.1^2)$, $c_2 \sim \mathcal{N}^+(0.24, 0.2^2)$, $\alpha \sim \mathcal{N}^+(2.7, 0.3^2)$, $\bar{\varepsilon} \sim \mathcal{N}(0, 0.02^2)$, where $\mathcal{N}^+(.,.)$ denotes a truncated normal distribution. The control points x_i are taken to be $M = 5$ equally spaced displacements between $x_1 = 0$ and $x_5 = 0.5$, and synthetic data were generated for $J = 10$ experiments.

We assume the parameter-to-observable map is taken to be a linear mathematical relationship between force and displacement (Hooke's Law), and is the same for each experiment $j = 1, ..., J$:

$$f(x, k) = kx, \quad \text{where} \quad k \in \mathbb{R}^+. \tag{13}$$

For each experiment, the misfit between model and data points are assumed to be independent, isotropic, Gaussian with zero mean and unknown variance s^2, that is as defined by (3) with $\Sigma_\varepsilon = s^2 \mathbb{I} \in \mathbb{R}^{4 \times 4}$ with $\Sigma_{\bar{\varepsilon}} = s^2 \mathbb{I} \in \mathbb{R}^{5 \times 5}$. This captures the measurement uncertainty (generated by $\bar{\varepsilon}$ in (12)) and in this case, the model misspecification here is that the data are actually generated from a nonlinear model (12). The uncertain parameters for each experiment are therefore wrapped into a vector $\boldsymbol{\psi} = [k, s]^T$.

The problem can now be formulated into the hierarchical Bayesian setting, and therefore, hyper-parameters which describe the population statistics of k and s are introduced:

$$k \sim \mathcal{N}^+(\mu_k, v_k) \quad \text{and} \quad s \sim \mathcal{N}^+(\mu_s, v_s).$$

This gives four hyper-parameters, which are collected into a single vector $\boldsymbol{\phi} = [\mu_k, v_k, \mu_s, v_s]^T$. Finally, hyper-priors are defined for each component of $\boldsymbol{\phi}$, such that:

$$\begin{aligned} \mu_k \sim \mathcal{N}^+(1.0, 0.1), \quad v_k \sim \mathcal{N}^+(0.3, 0.1), \\ \mu_s \sim \mathcal{N}^+(0.3, 0.1) \quad \text{and} \quad v_s \sim \mathcal{N}^+(0.03, 0.01). \end{aligned} \tag{14}$$

The next step is to build $J = 10$ independent GPs, for the mapping between $\psi \to \pi(\psi | \mathbf{d}_j)$, for $j = 1, ..., J$. GPs were trained using the open source package `GPy` (GPy, 2012). After some initial testing, an ARD 5/2 Matérn kernel with a zero mean function was chosen to define each GP emulator. The kernel is defined by the two-point correlations function:

$$c(\boldsymbol{\psi}, \boldsymbol{\psi}') = \frac{\sigma_r^2 d(\boldsymbol{\psi}, \boldsymbol{\psi}')^{5/2}}{\Gamma(v_k) 2^{5/2}} K_{5/2}(d(\boldsymbol{\psi}, \boldsymbol{\psi}')) \quad \text{where} \quad d(\boldsymbol{\psi}, \boldsymbol{\psi}') := \sqrt{5 \sum_i^2 \left(\frac{\psi_i - \psi_i'}{\ell_i} \right)^2}, \tag{15}$$

where $K_{5/2}(\cdot)$ is a modified Bessel function and $\Gamma(\cdot)$ is the gamma function. The additional hyper-parameters for the Matérn covariance σ_r^2, the variance and $\ell = [\ell_1, \ell_2]$, the characteristic length scale, are free-parameters that are optimized during the training process. To train each model, we consider 100 samples drawn using a Latin Hypercube (SMT2019, 2019), a quasi-random (space-filling) sampler. Given initial starting points for the hyper-parameters of the Matérn kernel (15), the parameters are optimized to maximize the log-likelihood over the complete dataset using `GPy`'s default optimizer L-BFGS-B (GPy, 2012; Paleyes et al., 2019).

To test the resulting emulators, we perform a fivefold cross-validation test. The 100 training points are divided into five groups. The GP is trained with four groups and last group is used for the test, the process

is repeated five times assigning a different "test group" each time. In each case, we consider the individual standardized errors for $j = 1, \ldots, 20$ left out samples defined by:

$$e_j = \frac{f(x_j) - m^*(x_j)}{\sqrt{v^*(x_j, x_j)}},$$

where $f(x_j)$ is the value of the actual model, $m^*(x_j)$, $v^*(x_j, x_j)$ is the value of the mean and covariance function of our GP at x_j. Figure 2 (right) shows the standardized error for 20 points along with lines of one standard deviation. As it is assumed that the noise error follows a normal distribution $\mathcal{N}(0, 1)$ along with the natural uncertainty, we expect the standardized error to follow the same normal distribution and that no more than two points of each fold falls outside of the standard deviation line. This derives from the fact that we target in a 90% confidence interval (CI) which is a well-trained emulator. In every subfigure, each different fold is marked with a different color (and shape), so overall, we conclude that the emulators are well-trained as no more than two points of the same color (shape) are outside of the bounds in any figure.

Following the HBM setup in Section 2.2, a Metropolis-Hastings algorithm driven by precondition Crank–Nicolson (pCN) proposal distribution including a simple adaptive scaling was used (Cotter et al., 2013). A single MCMC of 25,000 samples was computed with a target acceptance rate of 25%. The first 5,000 samples were discarded for burn-in. The effective sample size for each of the hyper-parameters was computed such that $N_{\text{eff.}} = [1813, 424, 367, 349]^T$ giving sampling error $S_{err} = [0.090, 0.021, 0.018, 0.017]$, for μ_ψ, v_ψ, μ_s, v_s, respectively. The sampling errors are sufficiently small relative to the quantities estimated, and hence the number of effective samples in each case are deemed sufficient. Figure 3 shows the traces and the posterior distribution of all hyper-parameters considering 10 independent data. The effective sample sizes for each parameter could be improved using more sophisticated sampling methods, for example Hamiltonian Monte Carlo (Duane et al., 1987) or more refined adaptive Metropolis sampler (Haario et al., 2001).

We observe from Figure 3 that the posterior samples are strongly influenced by the data, demonstrated by the qualitative difference between prior and posterior distributions. This is expected since the synthetic data are generated from a nonlinear spring model, and the statistical model accounts for model misspecification and measurement uncertainty within a single parameter s. In particular, we observe an expected value $\mu_k = 1.17$, higher than the prior value specified ($\mu_k = 1$), but closer to the value used to generate the data $\mu_k = 1.1$ mean value of c_1 ($\mu_{c_1} = 1.1$). For real-world problems, additional work could be done to better specify a statistical model for the model missspecification, accounting for the fact there is greater model mismatch at larger strain values, and/or impose a more informed prior. However, this is outside the scope of this article for a simple demonstrative example.
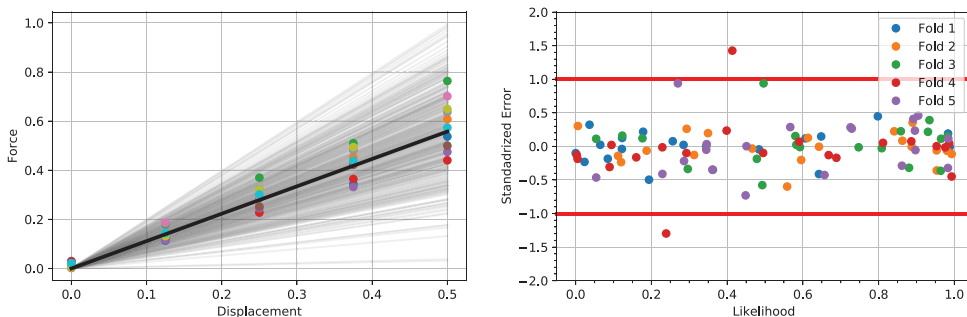

Figure 2. *(Left) Five hundred samples from the full posterior of $f = kx$ alongside experimental data points for $J = 10$. The mean of all posterior samples is plotted in as a solid black line. (Right) Shows the validation of a single experiment and associated Gaussian process using a fivefold analysis. Results demonstrate a well-trained model, since in the five folds only a single point has a standard error of magnitude greater 1. Similar good validation results are observed for all 10 experiments.*
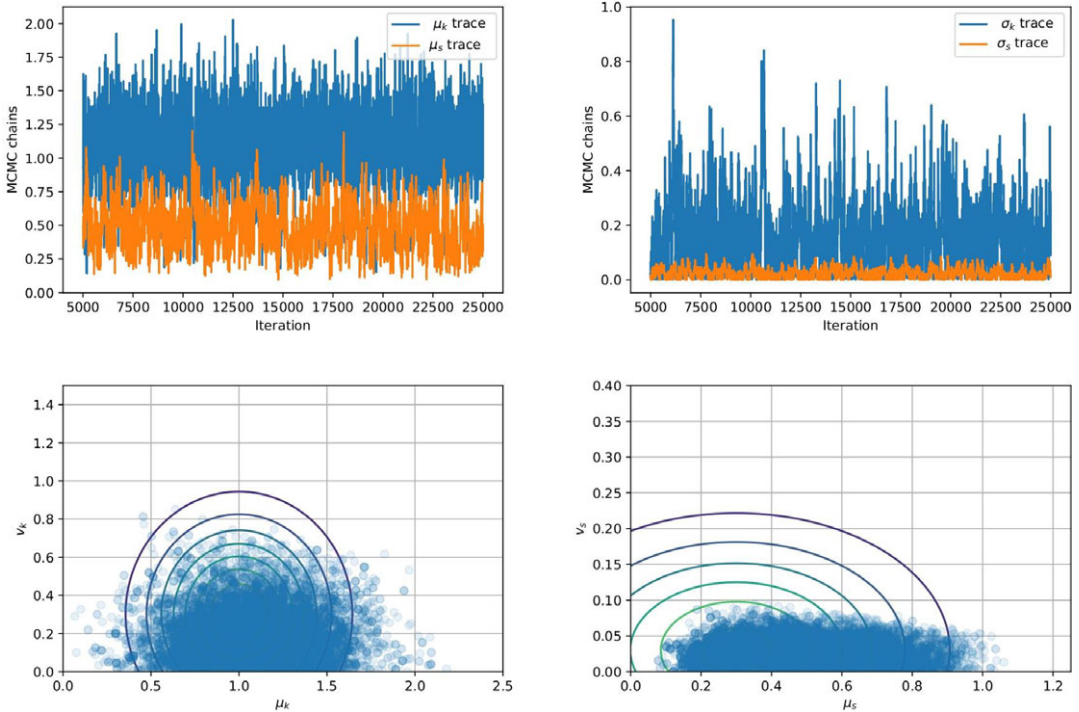
Figure 3. *(Top Row) Traces for the four hyper-parameters μ_k, v_k, μ_s, and v_s. (Bottom Row) Posterior distributions of the hyper-parameters (left) μ_k and v_k and (right) μ_s and v_s of the hyper-parameters.*

Table 1. *Descriptive statistics of posterior samples for an increasing number of experimental tests J*

J	$\mathbb{E}(\mu_k)$	$\mathbb{V}(\mu_k)$	$\mathbb{E}(v_k)$	$\mathbb{V}(v_k)$	$\mathbb{E}(\mu_s)$	$\mathbb{V}(\mu_s)$	$\mathbb{E}(v_s)$	$\mathbb{V}(v_s)$	Δ_{KL}
1	0.99	0.17	0.25	0.026	0.37	0.029	0.029	3.2e-4	–
5	1.05	0.09	0.22	0.018	0.40	0.017	0.025	3.1e-4	105
10	1.12	0.06	0.17	0.019	0.51	0.026	0.025	3.6e-4	15
15	1.19	0.05	0.17	0.015	0.53	0.021	0.024	1.8e-4	14
20	1.18	0.04	0.12	0.011	0.52	0.019	0.024	3.5e-4	6

Kullback–Leibler (KL) divergence Δ_{KL} is approximated between distribution as each set of data is added.

   In testing, MCMC simulations were repeated with increasing amounts of experimental data, so that in total, we have results for $J = 1, 5, 10, 15$, and 20. In each case, posterior chains were thinned using estimated auto-correlation lengths to achieve approximately independent samples for statistical analysis. For each group of experiments, descriptive statistics for each chain were computed and are summarized in Table 1. We analyze the convergence of posterior distribution under increasing number of experiments by estimating the Kullback–Leibler (KL) divergence (Kullback and Leibler, 1951), comparing the posterior distributions as new data. In the table, the KL divergence as the number of experiments are incremented are denoted Δ_{KL}. The results in the Table 1, in particular, the decreasing KL divergence between sequential distributions, and Figure 4 (left) shows the convergence of the distribution with increasing number of experiments as expected. With the exception of the case $J = 1$, we observe a reduction in uncertainty (decrease in estimated variance) as we increase the amount of data. This corresponds to the reduction in
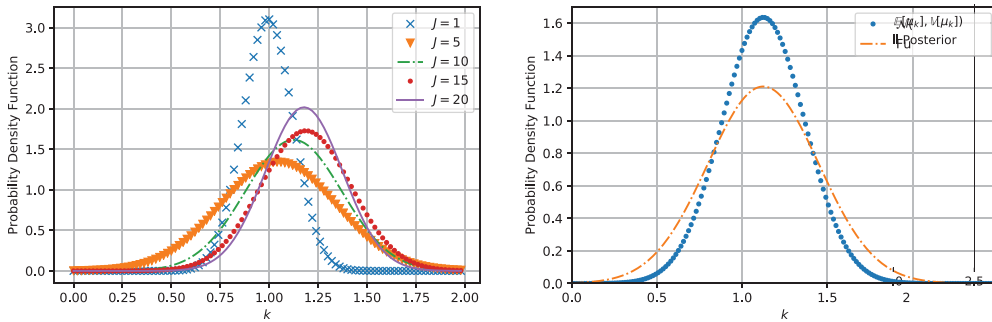
Figure 4. *(Left) Shows the distribution of $k \sim N^+\left(\mathbb{E}[\mu_k], \mathbb{E}[v_k]^2\right)$ for increasing amounts of data $J = 1, \ldots 20$. (Right) This figure shows the difference between the two distributions, the first which is denoted with "point marker" comes for $k \sim N^+\left(\mathbb{E}[\mu_k], \mathbb{E}[v_k]^2\right)$ while the other (denoted with "dash-dot line") is a maximum entropy approximated (with four moments matched) constructed from all posterior samples.*

the sampling error of the experimental data, allowing us to better represent the population statistics. A plot of a random 500 realization from the posterior for $J = 10$, alongside the original experimental data is shown in Fig.4 (right). All samples bound the experimental results, while the mean of the posterior samples provides a qualitatively sensible estimate for all the experiments.

Before considering an example with real experimental data, we apply the maximum entropy approximation of all posterior samples. A maximum entropy approximation for k was obtained with a four-moment representation of the distribution to all posterior samples. The results are shown in four (right), in which the maximum entropy approximation of the full posterior (dash-dot line) is plotted alongside the mean estimate $k \sim N\left(\mathbb{E}[\mu_k], \mathbb{E}[v_k]^2\right)$ (dot line). As expected, we see that full posterior distribution has greater uncertainty. This captures the effects of approximating population distributions with only a finite number of experiments, but also the uncertainty imposed by the assumption that the population distribution is Gaussian. The maximum entropy calculations allow the posterior distribution to be efficiently represented (here with just four numbers) rather than many thousands of samples, thereby simplifying the stochastic computational pipeline in a sequence of engineering calculations.

3.2. Example 2: stochastic elastoplastic model for 3D-printed steel

In this section, our approach is applied to a real experimental dataset provided by coupon tests of 3D-printed steel (Gardner et al., 2020). The study extends a previous Bayesian study for estimating uncertain material parameters using these data, as described in Dodwell et al. (2021). In the previous study, a standard Bayesian approach was presented, whereby all experiments were fitted with single values of parameters rather than population distribution as developed in the HBM. This section provides a brief overview of the engineering context, the mathematical and stastistical model setup and focuses on the results using our presented approach.

Additive manufacturing (AM) technologies are developing at a rapid pace, enabling the manufacture of advanced components with near arbitrary complexity. Various techniques have been developed for metallic AM including powder bed fusion and direct energy deposition. An interesting class of direct energy deposition are wire arc additive manufacturing (or WAAM), a process using a robotic arm and an

arc welding tool (Figure 5, left). This novel technology can build large scale structures in situ, at relatively high deposition rates (~ 4 kg/h). A well-documented example is the pedestrian bridge developed by Dutch startup MX3D (Figure 5, right). The complexity of possible parts along the uncertain thermal deformations of the material during welding, means that the as-manufactured material properties of WAAM are uncertain. In this section, we apply our hierarchical Bayesian approach to calibrate a stochastic elastoplastic model for WAAM.

The dataset reported by Gardner et al. (2020) are a set of six *smoothed/machined* tensile coupons, manufactured and tested according to the EN ISO 6892-1 standard (ISO, 2019), see Figure 6. The coupons were cut at angles $\psi = 0°$, 45° and 90° from a single larger WAAM steel plate. Here, θ is the angle of the layering relative to the build direction as shown in Figure 6. In total, two experiments are performed for each build direction, hence $J = 6$. For each test, the cross-sectional areas A_c of each coupon are measured, while the testing machine measures the applied tensile load and a four-camera LaVision digital image correlation system provides averaged surface strain measurements from both sides. Using these measurements, for each of the six experiments, the longitudinal stress σ_L at eight equally spaced longitudinal strain values ε_L between 0% and 0.8% is calculated. Each experimental test is modeled using a Ramberg–Osgood model (Ramberg and Osgood, 1943),

$$\varepsilon_L = \frac{\sigma_L}{E_L(\theta)} + K \left(\frac{\sigma_L}{\sigma_0(\theta)} \right)^n, \tag{16}$$

where $E_L(\theta)$ and $\sigma_0(\theta)$ are the unknown longitudinal elastic modulus and yield strengths, respectively, for a coupon with layers oriented at an angle θ, n is an unknown scalar hardening exponent and K is a constant. In this work, K is equal to 0.002 and, as a consequence, the yield strength σ_0 corresponds to the 0.2% proof stress, which is widely used to define the yield of metals in the literature (Gardner et al., 2020).

Collectively, by testing coupons over a range of angles, the Ramberg–Osgood model can be used to define a general anisotropic homogeneous, elastoplastic constitutive law under plane stress conditions over all angles θ. In this contribution, the elastic part of the deformation is consider to be transversely

Figure 5. (Left) The 3D-printing protocol developed by MX3D uses a weld head attached to a robotic arm (image by Joris Laarman, www.jorislaarman.com). (Middle) Pedestrian bridge manufactured using 3D-printed steel (7). (Right) Test set of up of individual coupons.
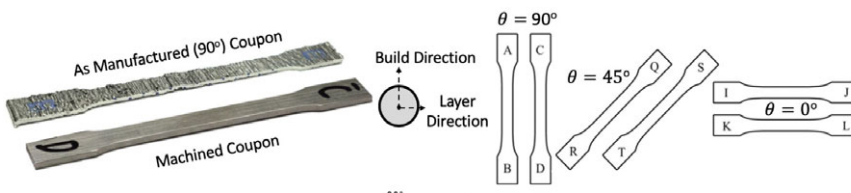

Figure 6. Test coupons were machined to remove influence of geometric surface features and cut at angles $\theta = 0°$, 45°, and 90° perpendicular to the printing direction (Buchanan et al., 2018).

isotropic, parameterized by E_1, E_2, G_{12}, and v_{12}, respectively the elastic moduli perpendicular and parallel to the build direction, a shear modulus and a Poisson ratio. For the plastic response, Hill's quadratic yield surface (Hill, 1948) is considered which, under plane stress assumptions, is uniquely parameterized by three parameters (F, G, N). Post-yield, a simple isotropic hardening rule is defined by the single exponent, n. The mathematical connection between the Ramberg–Osgood model and the two-dimensional elasto-plastic material parameters is provided in the Appendix. As a result, all experiments are described by a single vector of parameters, such that:

$$\boldsymbol{\psi} = \left[E_1, E_2, G_{12}, v_{12}, \sigma_y(0°), \sigma_y(90°), \sigma_y(45°), n, s_e\right].$$

The data for each experiment is represented in a vector $\mathbf{d}^{(j)} \in \mathbb{R}^8$, which contains the eight evenly spaced strain values. The mapping between parameter values and data are defined by the forward map:

$$\mathbf{d}^{(j)} = \mathbf{f}(\boldsymbol{\psi}) + \bar{\boldsymbol{\varepsilon}}.$$

Once again, $\bar{\boldsymbol{\varepsilon}}$ is a random vector that represents measurement error, whose components are independent and Gaussian with $e_i \sim \mathcal{N}(0, s_e^2 d_i^2)$, whereby the noise-to-signal parameter s_e is to be learned, providing a quantitative measurement of uncertainty in both the data and the model (i.e., as for the first example allowing for model misspecification). For all parameters except s_e, it is assumed that it comes from a normal distribution, that is $\psi_i \sim \mathcal{N}(\mu_i, v_i^2)$, while s_e is single value across all experiments. Gathering the mean and variance for each parameter and a single value for s_e, gives a vector of hyper-parameters of 17 parameters. Each of these is equipment with weakly informed prior distributions, defined by normal distributions over plausible ranges for the parameters given the least mean square fitting reported in (Gardner et al., 2020). All prior values are summarized and given in the Table 2.

MCMC simulations were carried out as described in Section 2. In this example, since the parameter space is larger than the previous example, a more sophisticated sampling approach is required. Hence all calculations are implemented in `PyMC3` (Salvatier et al., 2016), for which sampling was achieved using the inbuilt sampler `DEMetropolizZ`, a differential evolution Markov chain methods with embedding snooker updating (Braak and Vrugt, 2008). Once again, Gaussian emulators are built for each of the six experiments using `GPy`, for which a nine-dimensional ARD Matérn covariance (15) is used and trained with 300 samples per experiment and using a sixfold cross-validation analysis, very much as described in Example 1. In total, 5,000 samples were computed using MCMC, divided equally between four independent chains. The first 250 samples in each chain were neglected for burn in.

Summary statistics of prior and posterior samples are given in Table 2. For the first example, the distribution of possible observed individual samples can be constructed by sampling the hyper-distributions for hyper-parameters given by the posterior samples. Figure 7 (left) displays such posterior samples showing the possible distribution of stress–strain curves in relation to the experimental data (solid dots), with colors differentiating the build angles $\theta = 0°, 45°$, and $90°$. the two polar plots in Figure 7 (middle and right) shows posterior samples (green and red), the maximum a posteriori estimate of longitudinal elastic modulus and proof stress over the sweep of possible build angles θ. Here, isotropic behavior would be observed as a quarter circle (constant radius) for both properties. For both elastic and yield properties, the posterior samples clearly show an uncertain anisotropic material response, as discussed in more detail in Dodwell et al. (2021).

In our hierarchical Bayesian results, even with broad priors for the standard deviations of the population distributions for each parameter, for example s_{E_1}, s_{E_2}, we observe the posterior variances for the hyper-distributions are relatively small, while the variance of the means are large. This suggests the limited data show good agreement with the low predicted population variances. Qualitatively visualizing the evaluation of posterior samples of our model compares well with the data. However, some parameters in this approach are relatively insensitive to moderate variations in their values. For example, parameter E_1 effects only $0°$ coupons and the initial linear part of the curve (approximately $1/4$ of the curve), and is therefore informed by only $1/12$ of the data. Therefore, variations in fitting this value has a much smaller effect on the overall likelihood than one might expect.

Table 2. *Summary statistics of prior and posterior distributions, N—normal and HN—half normal.*

|	Units	Prior			Posterior			
		Dist.	Mean	Std. Dev.	Mean	Std. Dev.	Skew	Kurtosis
μ_{E_1}	GPa	N	150.00	4.00	153.97	5.00	2.82	55
μ_{E_2}	GPa	N	150.00	4.00	149.72	4.25	0.80	11
$\mu_{G_{12}}$	GPa	N	105.00	2.00	106.105	4.66	9.32	235
$\mu_{\sigma(0°)}$	GPa	N	0.35	0.04	0.36	0.06	10.56	288
$\mu_{\sigma(90°)}$	GPa	N	0.35	0.04	0.31	0.05	6.10	120
$\mu_{\sigma(45°)}$	GPa	N	0.40	0.04	0.42	0.07	12.39	270
n	–	N	13.00	0.5	12.31	1.89	0.15	2
s_{E_1}	GPa	HN	–	5.00	1.60	4.19	4.57	27
s_{E_2}	GPa	HN	–	5.00	1.20	3.55	6.55	65
$s_{G_{12}}$	GPa	HN	–	5.00	0.92	3.06	6.33	51
$s_{\sigma(0°)}$	GPa	HN	–	0.05	0.03	0.05	3.42	17
$s_{\sigma(0°)}$	GPa	HN	–	0.05	0.05	0.09	18.88	650
$s_{\sigma(0°)}$	GPa	HN	–	0.05	0.07	0.07	8.99	226
n	–	HN	–	0.05	0.30	0.69	3.37	13
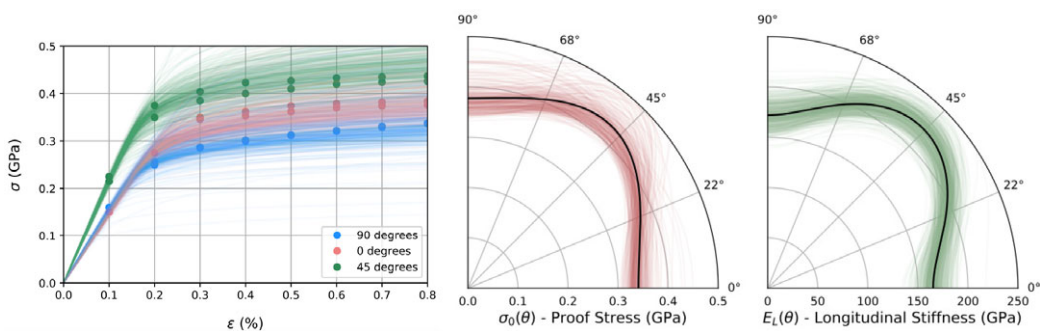
Figure 7. *(Left) Samples strain–stress curves via realization of the Ramberg–Osgood model (16). Data points are for two experiments, two for each build angle. (Right) Posterior samples for longitudinal stiffness and proof stress as a function of build angle. Bold lines show maximum a posteriori estimates.*

4. Conclusion

In this article, we develop a formal hierarchical Bayesian methodology for the calibration of parameters of a material model from a set of experiments. The article extends existing probabilistic mechanics literature which previously only considered the calibration specific to a single experiment. When only a single experimental result is used, this may lead to significant bias in the acquired posterior distributions. Clearly, this is because it is a single experimental test is drawn from a population distribution of possible tests, and a result inference from a single experiment will be over-optimistic. Using a formal HBM, we naturally account for the population variation in the experiments with resulting posterior distribution appropriately showing higher variance. This is supported in the test cases, both synthetic and real-world.

The computational challenges which arise for the hierarchical model are significant. Rather than inferring over a single experiment, inference is required over all experiments simultaneously. However, assuming that each experiment is independent, it is possible to marginalize all parameters for a given experiment and formulate the resulting Bayesian calculations in terms of the hyper (population) parameters alone. While ehim marginalization integral appears computationally expensive, we show that it can be readily approximated using Bayesian quadrature using a precomputed Gaussian emulator as a surrogate to good effect. There might be limitations where the number of material parameters for a model becomes particularly high, but this is not typical in a material model. In such cases, a more crude Monte Carlo approximation would work.

We compute full posterior distributions of hyper parameters using Metropolis Hasting MCMC. Clearly, when there is a large number of experiments and the model and computations are very expensive, then approximate approaches could be taken. For example, likelihoods could be approximated with only random batches of experiments, or clearly approximate Bayesian methods, for example, a Laplace approximation or variational methods could be readily deployed in the same framework.

References

Babapulle MN, **Joseph L**, **Bélisle P**, **Brophy J and Eisenberg MJ** (2004) A hierarchical Bayesian meta-analysis of randomised clinical trials of drug-eluting stents. *The Lancet 464*, 583–591.

Babuška I, **Sawlan Z**, **Scavino M**, **Szabó B and Tempone R** (2016) Bayesian inference and model comparison for metallic fatigue data. *Computer Methods in Applied Mechanics and Engineering 304*, 171–196.

Bayarri MJ, **Berger JO**, **Paulo R**, **Sacks J**, **Cafeo JCJ**, **Lin C-H and Tu J** (2007) A framework for validation of computer models. *Technometrics 49*(2), 138–154.

Bozorgzadeh N, **Harrison PJ and Escobar MD** (2019) Hierarchical Bayesian modelling of geotechnical data: Application to rock strength. *Géotechnique 69*(12), 1056–1070.

Braak CJF and Vrugt JA (2008) Differential evolution Markov chain with snooker updater and fewer chains. *Statistics and Computing 18*(4), 435–446.

Buchanan C, **Esther R and Leroy G** (2018) Testing, simulation and design of cold-formed stainless steel CHS columns. *Thin-Walled Structures 130*, 297–312.

Charnes A, **Frome EL and Yu P-L** (1976) The equivalence of generalized least squares and maximum likelihood estimates in the exponential family. *Journal of the American Statistical Association 71*(353), 169–171.

Chiachío J, **Chiachío M**, **Saxena A**, **Sankararaman S**, **Rus Gand Goebel K** (2015) Bayesian model selection and parameter estimation for fatigue damage progression models in composites. *International Journal of Fatigue*, *70*,361–373.

Chipman H, **George EIM**, **Robert E**, **Clyde MF**, **Dean P and Stine RA** (2001) The practical implementation of Bayesian model selection. *Lecture Notes-Monograph Series 38*, 65–134.

Cotter SL, **Roberts GOS**, **Andrew M and White D** (2013) MCMC methods for functions: modifying old algorithms to make them faster. *Statistical Science* 28(3), 424–446.

Dodwell TJ, **Fleming LR**, **Buchanan C**, **Kyvelou P**, **Detommaso G**, **Gosling PD**, **Scheichl R**, **Gardner WSK**, **Girolami MA and Oates, CJ** (2021) A data-centric approach to generative modelling for 3D-printed steel. *Proceedings of the Royal Society A* 477 (2255), 20210444.

Duane S, **Kennedy AD**, **Pendleton BJ and Roweth D** (1987) Hybrid Monte Carlo. *Physics Letters B* 195(2), 216–222.

Fielding M, **Nott DJ and Liong S-Y** (2011) Efficient MCMC schemes for computationally expensive posterior distributions. *Technometrics* 53, 16–28.

Gardner L, **Kyvelou P**, **Herbert G and Buchanan C** (2020) Testing and initial verification of the world's first metal 3D printed bridge. *Journal of Constructional Steel Research* 172.

Gelman A, **Carlin JB**, **Stern HS**, **Dunson DB**, **Vehtari A and Rubin DB** (2013) *Bayesian Data Analysis*. Boca Raton, FL: CRC press.

Gogu C, **Yin WH**, **Raphael I**, **Peter M**, **Le Riche J**, **Rodolphe and Vautrin A** (2013) Bayesian identification of elastic constants in multi-directional laminate from moiré interferometry displacement fields. *Experimental Mechanics* 53(4), 635–648.

GPy (2012) A Gaussian process framework in python. Available at http://github.com/SheffieldML/GPy

Haario H, **Saksman E**, **Tamminen J** (2001) An adaptive metropolis algorithm *Bernoulli* 7(2), 223–242.

Hill R (1948) A variational principle of maximum plastic work in classical plasticity. *The Quarterly Journal of Mechanics and Applied Mathematics* 1(1), 18–28.

Huang Y, **Liu D and Wu H** (2006) Hierarchical Bayesian methods for estimation of parameters in a longitudinal HIV dynamic system. *Biometrics* 62(2), 413–423.

Isenberg J (1979) *Progressing from least squares to Bayesian estimation. Proceedings of the 1979 ASME Design Engineering Technical Conference, New York*, 1–11.

ISO (2019) ISO 6892-12019 metallic materials — Tensile testing — Part 1: Method of test at room temperature. Available at https://www.iso.org/obp/ui/#iso:std:iso:6892:-1:ed-3:v1:en.

Johnson VE and Rossell D (2012) Bayesian model selection in high-dimensional settings. *Journal of the American Statistical Association* 107(498), 649–660.

Kennedy MC and O'Hagan A (2001) Bayesian calibration of computer models. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 63(3), 425–464.

Kong A (1992) A note on importance sampling using standardized weights.

Kong A, **Liu JS and Wong WH** (1994) Sequential imputations and Bayesian missing data problems. *Journal of the American Statistical Association* 89(425), 278–288.

Kullback S and Leibler R (1951) On information and sufficiency. *MathSciNet MATH* 22, 79–86.

Kwon H-H, **Brown C and Lall U** (2008) Climate informed flood frequency analysis and prediction in Montana using hierarchical Bayesian modeling. *Geophysical Research Letters* 5, 35.

Liu JS (2008). Monte Carlo strategies in scientific computing.

Loredo TJ (2013) Bayesian astrostatistics: a backward look to the future. *Astrostatistical Challenges for the New Astronomy*, 15–40.

Madireddy S, **Sista B and Vemaganti K** (2015) A Bayesian approach to selecting hyperelastic constitutive models of soft tissue. *Computer Methods in Applied Mechanics and Engineering* 291, 102–122.

O'Hagan A (1991) Bayes–hermite quadrature. *Journal of Statistical Planning and Inference* 29(3), 245–260.

Oden JT, **Prudencio EE and Hawkins-Daarud A** (2013) Selection and assessment of phenomenological models of tumor growth. *Mathematical Models and Methods in Applied Sciences* 23(07), 1309–1338.

Osborne M, **Garnett R**, **Ghahramani Z**, **Duvenaud DK**, **Roberts SJ and Rasmussen CE** (2012) Active learning of model evidence using Bayesian quadrature. *Advances in Neural Information Processing Systems*, 46–54.

Paleyes A, **Pullin M**, **Mahsereci M**, **Lawrence N and González J** (2019) *Emulation of physical processes with emukit. Second Workshop on Machine Learning and the Physical Sciences, NeurIPS*.

Ramberg W and Osgood WR (1943) Description of stress-strain curves by three parameters.

Rappel H, **Beex LAA and Bordas SPA** (2018) Bayesian inference to identify parameters in viscoelasticity. *Mechanics of Time-Dependent Materials* 22(2), 221–258.

Rasmussen CE (2003) Gaussian processes to speed up hybrid Monte Carlo for expensive Bayesian integrals. *Seventh Valencia international meeting, dedicated to Dennis V. Lindley*, 651–659.

Roberts GO and Rosenthal JS (2009) Examples of adaptive MCMC. *Journal of Computational and Graphical Statistics* 18(2), 349–367.

Roberts JS (2011) Optimal proposal distributions and adaptive MCMC. *Handbook of Markov Chain Monte Carlo* 4(10), 1201.

Salvatier J, **Wiecki TV and Fonnesbeck C** (2016) Probabilistic programming in python using pymc3. *PeerJ Computer Science* 2, e55.

Solonen A, **Ollinaho P**, **Laine M**, **Heikki H**, **Johanna T and Heikki J** (2012) Efficient MCMC for climate model parameter estimation: parallel adaptive chains and early rejection. *Bayesian Analysis*, 7(3) 715–736.

Thomas A, **Redd A**, **Khader K**, **Leecaster M**, **Greene T and Samore M** (2015) Efficient parameter estimation for models of healthcare-associated pathogen transmission in discrete and continuous time. *Mathematical Medicine and Biology: A Journal of the IMA* 32(1), 81–100.

Valderrama-Bahamóndez GI and Fröhlich H (2019) MCMC techniques for parameter estimation of ODE based models in systems biology. *Frontiers in Applied Mathematics and Statistics 55*.

Wolfgang A (2016) *Hierarchical Bayesian Modeling*. State College, PA: Academic Lecture, Pennsylvania State University.

Appendix

Plane-stress elastoplastic model

Consider a coupon which occupies the domain $\Omega \in \mathbb{R}^3$, with global coordinate system $(x, y, z) \in \mathbb{R}^3$, alongside a local coordinate system $(x_1, x_2, x_3) \in \mathbb{R}^3$. Here, the local coordinate x_1 is parallel to the layering at an angle θ to the global x axis, x_2 is the build direction and $x_3 \equiv z$ is normal to the plane of the coupon.

For each experiment with a layer orientation θ to the horizontal axis x, the experiments provide longitudinal stress values $\sigma_L \equiv \sigma_y$ at given values of longitudinal strains $\varepsilon_L \equiv \varepsilon_y$. The connection from this experimental data to global material parameters/properties is a two-stage process. Each experiment is modeled using a 1D Ramberg–Osgood model

$$\varepsilon_L = \frac{\sigma_L}{E_L(\theta)} + K \left(\frac{\sigma_L}{\sigma_0(\theta)} \right)^n. \tag{17}$$

This model is defined by three parameters $E_L(\theta)$, $\sigma_0(\theta)$, and n, where K is taken to be the fixed constant 0.02. In modeling the overall material, since the z dimension is relatively thin, plane-stress assumptions are imposed. In this section, the connection is set out between the independent Ramberg–Osgood parameters to eight global material parameters, which are made up of four elastic material parameters E_1, E_2, G_{12}, and v_{12}, three independent stress factors F, G, and N which define the onset of yield and a single (θ independent) hardening parameter, n.

The transversely anisotropic elastic properties of the material are defined uniquely by four parameters E_1, E_2, G_{12}, and v_{12}. The connection between local stresses and strains are defined by the compliance or stiffness matrices, given by

$$\mathbf{S} = \begin{bmatrix} \dfrac{1}{E_1} & -\dfrac{v_{12}}{E_2} & 0 \\ -\dfrac{v_{12}}{E_2} & \dfrac{1}{E_2} & 0 \\ 0 & 0 & \dfrac{1}{2G_{12}} \end{bmatrix} \quad \text{and} \quad \mathbf{C} = \mathbf{S}^{-1}, \tag{18}$$

respectively. Our controlled data from the experiments are measure values of stress in the longitudinal direction $\sigma_y \equiv \sigma_L$ at known longitudinal strain values $\varepsilon_y \equiv \varepsilon_L$. For the experimental setup, the applied values of σ_x and τ_{xy} are both zero. The resulting global strains, in global coordinates, can be calculated by rotating local coordinate properties by the layer orientation θ, so that

$$\begin{bmatrix} \varepsilon_x \\ \varepsilon_y \\ \gamma_{xy} \end{bmatrix} = \mathbf{T}^{-1} \mathbf{S} \mathbf{T} \begin{bmatrix} 0 \\ \sigma_y \\ 0 \end{bmatrix} \quad \text{where} \quad \mathbf{T} = \begin{bmatrix} c^2 & s^2 & 2sc \\ s^2 & c^2 & -2sc \\ -sc & sc & c \end{bmatrix}, \tag{19}$$

where $c = \cos(\theta)$ and $s = \sin(\theta)$. From this relationship, it follows that the elastic stiffness matrix in global coordinates is $\widehat{\mathbf{C}} = \widehat{\mathbf{S}}^{-1}$, whereby the longitudinal stiffness (the stiffness connecting longitudinal strains and stresses) is given by the entry $\widehat{C}_{22} = E_L$. This relationship provides a direct-mapping between material parameters E_1, E_2, G_{12}, and v_{12} and the observed/controlled parameters θ and E_L. In general, the inverse of this map is ill-posed. However, in our setting data from experiments at multiple angles are available, enabling all parameters to in principle be identified.