


A Bayesian multilevel approach to optimally estimate material properties

Conference Paper**Author(s):**

Nagel, Joseph B.; [Sudret, Bruno](#) 

Publication date:

2014

Permanent link:

<https://doi.org/10.3929/ethz-a-010247080>

Rights / license:

[In Copyright - Non-Commercial Use Permitted](#)

Originally published in:

<https://doi.org/10.1061/9780784413609.151>

A Bayesian Multilevel Approach to Optimally Estimate Material Properties

Joseph B. Nagel¹ and Bruno Sudret¹

¹ETH Zürich, Institute of Structural Engineering,
Chair of Risk, Safety & Uncertainty Quantification,
Stefano-Franscini-Platz 5, CH-8093 Zürich, Switzerland,
nagel@ibk.baug.ethz.ch, sudret@ibk.baug.ethz.ch

ABSTRACT

Bayesian multilevel modeling establishes a convenient framework to account for aleatory variability and epistemic uncertainty in inverse problems. In view of engineering applications we will place this framework into the context of classical and probabilistic inversion and demonstrate its additional potential to infer material properties as intermediate variables within a hierarchically defined Bayesian multilevel model. The objective of probabilistic inversion is to infer such probability distributions that describe the variability of model inputs across a number of experiments. Inference therefore focuses on so-called hyperparameters that parametrize population distributions. To that end intermediate or latent variables are considered nuisance and commonly marginalized out from the problem. But instead of integrating out intermediate quantities, one can just as well fade the hyperparameters from the inferential spotlight and marginalize them out. This allows for an “optimal” inference of the intermediate quantities with respect to the available prior information, structural assumptions and the acquired data. We will devise a simple academic example within the domain of mechanical engineering that will be used to study this optimal kind of inference. The system under consideration is an ensemble of simply supported beams with uncertain material properties. Over the sample of beams, Young’s moduli are subject to aleatory variability, while the knowledge about the true hyperparameter values suffers from epistemic uncertainty. We will show how individual Young’s moduli can be optimally estimated in a series of three-point bending experiments. It will be shown that in hierarchically defined models one can learn more about a specific specimen than what is encoded in dedicated data directly associated with that single specimen. One can indirectly utilize information from different specimens, within the same sample, in order to improve the assessment of a specific one. In this contribution we will demonstrate the potential and challenge of this learning mechanism and we will investigate the underlying flow of information.

PARAMETER ESTIMATION & PROBABILISTIC INVERSION

Inversion is the inference of model parameters from noisy and limited data and is often formulated as statistical estimation. This formulation encompasses a wide range of different problems with parameter estimation and probabilistic inversion being particular instances.

Classical *parameter estimation* aims at inferring unknown parameters \mathbf{x} of a physical forward model \mathcal{M} that predicts the outcome $\mathcal{M}(\mathbf{x}, \mathbf{d}_i)$ of an experiment under known experimental conditions \mathbf{d}_i . Discrepancy between model predictions and acquired data $\mathbf{y}_i = \mathcal{M}(\mathbf{x}, \mathbf{d}_i) + \varepsilon_i$ for $i = 1, \dots, n$ is accounted for by a residual term $\varepsilon_i \sim f_E(\varepsilon_i)$ that captures the variability in the data that is due to measurement errors, numerical approximations as well as model inadequacies. A widespread probabilistic residual model is a Gaussian distribution $f_E = \mathcal{N}(\mathbf{0}, \Sigma_i)$ with a covariance matrices Σ_i . Bayesian inversion of the data \mathbf{y}_i is to identify the fixed albeit unknown model parameter \mathbf{x} by updating a priorly elicited epistemic knowledge $\pi(\mathbf{x})$ about the true parameter value. With a likelihood function $\mathcal{L}(\mathbf{y}_1, \dots, \mathbf{y}_n | \mathbf{x}) = \prod_{i=1}^n f_E(\mathbf{y}_i - \mathcal{M}(\mathbf{x}_i, \mathbf{d}_i))$ the Bayesian rationale is based on the construction of a posterior probability density $\pi(\mathbf{x} | \mathbf{y}_1, \dots, \mathbf{y}_n) \propto \mathcal{L}(\mathbf{y}_1, \dots, \mathbf{y}_n | \mathbf{x}) \pi(\mathbf{x})$ that describes the knowledge about the unknown parameter after conditioning on the data. Apart from some exceptional cases this posterior density has no analytical closed-form solution, thus more often than not the challenge posed lies in sampling the posterior by means of Markov chain Monte Carlo (MCMC) methods (Robert 2004).

Another type of inverse problem is posed when variation in the data $\mathbf{y}_i = \mathcal{M}(\mathbf{x}_i, \mathbf{d}_i) + \varepsilon_i$ is not only attributed to different experimental conditions \mathbf{d}_i and a residual discrepancy ε_i , but to inherently varying model inputs $(\mathbf{x}_i | \boldsymbol{\theta}) \sim f_{\mathbf{X} | \boldsymbol{\theta}}(\mathbf{x}_i | \boldsymbol{\theta})$ throughout the experiments for $i = 1, \dots, n$. The random variables \mathbf{x}_i are conditionally independent and follow a density function $f_{\mathbf{X} | \boldsymbol{\theta}}$. The objective of *probabilistic inversion* focuses on the estimation of the hyperparameters $\boldsymbol{\theta}$. Application examples can be found in (Rocquigny 2009; Celeux 2010; Barbillion 2011). A likelihood for this class of problems can be obtained by the marginalization $\mathcal{L}(\mathbf{y}_1, \dots, \mathbf{y}_n | \boldsymbol{\theta}) = \prod_i \int_{\mathbf{x}_i} f_E(\mathbf{y}_i - \mathcal{M}(\mathbf{x}_i, \mathbf{d}_i)) f_{\mathbf{X} | \boldsymbol{\theta}}(\mathbf{x}_i | \boldsymbol{\theta}) d\mathbf{x}_i$. If one is able to quantify prior knowledge $\pi(\boldsymbol{\theta})$ about the unknown hyperparameters $\boldsymbol{\theta}$, the posterior $\pi(\boldsymbol{\theta} | \mathbf{y}_1, \dots, \mathbf{y}_n) \propto \mathcal{L}(\mathbf{y}_1, \dots, \mathbf{y}_n | \boldsymbol{\theta}) \pi(\boldsymbol{\theta})$ is the result of Bayesian data analysis.

MULTILEVEL MODELING

As there seems to be no universally valid definition of a *multilevel* or a *hierarchical model* we define it as “an assembly of submodels at different levels of a hierarchy”, where the hierarchical structure can be given by conditional dependencies and deterministic maps between the quantities involved. While multilevel models have been studied from a frequentist point of view (Davidian 2003; Banks 2012), the problem outlined as probabilistic inversion is indeed a Bayesian multilevel problem. Summing

it up we obtain the following hierarchically nested model specification

$$(\mathbf{y}_i | \mathbf{x}_i) \sim f_{\mathbf{E}}(\mathbf{y}_i - \mathcal{M}(\mathbf{x}_i, \mathbf{d}_i)), \quad (1a)$$

$$(\mathbf{x}_i | \boldsymbol{\theta}) \sim f_{\mathbf{X}|\Theta}(\mathbf{x}_i | \boldsymbol{\theta}), \quad (1b)$$

$$\boldsymbol{\theta} \sim \pi(\boldsymbol{\theta}). \quad (1c)$$

Models of such a kind can be intuitively represented by directed acyclic graphs (DAG) such as in Fig. 1. The Bayesian paradigm offers a uniquely elegant solution to inversion by formulating a joint posterior density of all unknowns $(\mathbf{x}_1, \dots, \mathbf{x}_n, \boldsymbol{\theta})$ conditioned on all knowns $(\mathbf{y}_1, \dots, \mathbf{y}_n)$. Up to a normalization constant it is given as

$$\pi(\mathbf{x}_1, \dots, \mathbf{x}_n, \boldsymbol{\theta} | \mathbf{y}_1, \dots, \mathbf{y}_n) \propto \left(\prod_{i=1}^n f_{\mathbf{E}}(\mathbf{y}_i - \mathcal{M}(\mathbf{x}_i, \mathbf{d}_i)) f_{\mathbf{X}|\Theta}(\mathbf{x}_i | \boldsymbol{\theta}) \right) \pi(\boldsymbol{\theta}). \quad (2)$$

Standard MCMC techniques for sampling this posterior are readily available and easy to implement. However, the possibly high-dimensional parameter space is a serious challenge that may necessitate more advanced MCMC sampling schemes. What is considered *nuisance* to the proclaimed inferential objective becomes marginalized out from the posterior. Classically the hyperparameters $\boldsymbol{\theta}$ are of interest and the intermediate parameters \mathbf{x}_i will be marginalized out (Nagel and Sudret 2013). Instead we will herein declare individual realizations \mathbf{x}_i as being the quantities of interest. The Bayesian approach to multilevel problems suggests an optimal strategy of estimating individual realizations \mathbf{x}_i that goes beyond separate parameter estimations. Indeed the crude alternative to inference within a hierarchical model would be to solve n different inverse problems of inferring \mathbf{x}_i with dedicated data \mathbf{y}_i , respectively.

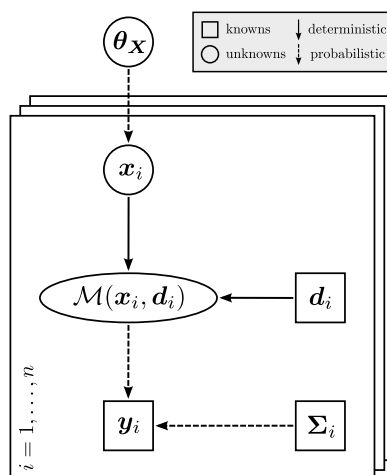


Figure 1. A DAG of the multilevel model is shown. Nodes represent known (\square) and unknown (\circ) quantities. Directed edges represent their probabilistic (\dashrightarrow) and deterministic (\longrightarrow) relations. Data \mathbf{y}_i , intermediate variables \mathbf{x}_i and the hyperparameters $\boldsymbol{\theta}$ constitute the hierarchical levels.

COMBINATION OF INFORMATION

Information-wise the estimation of the various unknown parameters in Eq. (1) can be based on the observed data \mathbf{y}_i , the Bayesian prior $\pi(\boldsymbol{\theta})$, the structural knowledge $f_{\mathbf{X}|\Theta}(\mathbf{x}_i|\boldsymbol{\theta})$ and the information being encoded in $f_{\mathbf{E}}(\mathbf{y}_i - \mathcal{M}(\mathbf{x}_i, \mathbf{d}_i))$. We will now focus on the optimal inference of individual parameters \mathbf{x}_i for $i = 1, \dots, n$. Instead of merely inverting each of the observations \mathbf{y}_i for the corresponding \mathbf{x}_i , we will solve one joint multilevel problem. As already mentioned, it will turn out that in doing so one can learn more about \mathbf{x}_i than what is encapsulated in \mathbf{y}_i . In order to demonstrate the effect and the underlying mechanism of this “optimal combination of information” or “borrowing strength” (Draper et al 1992), we will pursue the following three strategies for the inference of one specific parameter \mathbf{x}_i , say with $i = i_0$.

Bayesian Updating

In this first approach we base inference of \mathbf{x}_{i_0} solely on the data \mathbf{y}_{i_0} , the available structural knowledge $f_{\mathbf{X}|\Theta}(\mathbf{x}_{i_0}|\boldsymbol{\theta})$ and the prior belief $\pi(\boldsymbol{\theta})$. By marginalizing the joint prior density $f_{\mathbf{X}|\Theta}(\mathbf{x}_{i_0}|\boldsymbol{\theta})\pi(\boldsymbol{\theta})$ over the hyperparameter $\boldsymbol{\theta}$, the marginal prior information available about \mathbf{x}_{i_0} is given as

$$\pi(\mathbf{x}_{i_0}) = \int f_{\mathbf{X}|\Theta}(\mathbf{x}_{i_0}|\boldsymbol{\theta})\pi(\boldsymbol{\theta})\mathrm{d}\boldsymbol{\theta}. \quad (3)$$

This mixture or compound probability distribution represents the uncertainty about \mathbf{x}_{i_0} prior to any data analysis. The result of simple Bayesian updating of the prior $\pi(\mathbf{x}_{i_0})$ with the data \mathbf{y}_{i_0} is the posterior $\pi(\mathbf{x}_{i_0}|\mathbf{y}_{i_0})$. While the direct information \mathbf{y}_{i_0} about the parameter \mathbf{x}_{i_0} has entered the analysis, additional information $\mathbf{y}_{i \neq i_0}$ being available about $\boldsymbol{\theta}$ has been neglected. In other words, the hierarchical problem structure has been recognized but has not yet been utilized.

Multilevel Analysis

The joint multilevel analysis for estimating \mathbf{x}_{i_0} is accomplished by constructing the joint posterior Eq. (2) and subsequently integrating out nuisance. Hence when inferential attention is not directed to parameters $\mathbf{x}_{i \neq i_0}$ and hyperparameters $\boldsymbol{\theta}$, those will be marginalized over. The posterior knowledge about \mathbf{x}_{i_0} is then given as

$$\pi(\mathbf{x}_{i_0}|\mathbf{y}_1, \dots, \mathbf{y}_n) = \int \cdots \int \pi(\mathbf{x}_1, \dots, \mathbf{x}_n, \boldsymbol{\theta}|\mathbf{y}_1, \dots, \mathbf{y}_n)\mathrm{d}\boldsymbol{\theta}\mathrm{d}\mathbf{x}_{\sim i_0}, \quad (4)$$

where the simplifying notation $\mathbf{x}_{\sim i_0} = (\mathbf{x}_1, \dots, \mathbf{x}_{i_0-1}, \mathbf{x}_{i_0+1}, \dots, \mathbf{x}_n)$ has been introduced. Apart from the structural knowledge $f_{\mathbf{X}|\Theta}(\mathbf{x}_i|\boldsymbol{\theta})$ and the hyperprior $\pi(\boldsymbol{\theta})$, the total data \mathbf{y}_i for $i = 1, \dots, n$ has now contributed to the estimation of \mathbf{x}_{i_0} . Beyond just \mathbf{y}_{i_0} the estimation of \mathbf{x}_{i_0} has exploited information from $\mathbf{y}_{\sim i_0}$ in a joint learning process.

Bayesian Filtering

In order to give some insight into the aforementioned joint learning mechanism we propose a form of sequential Bayesian filtering. Initially we will infer the hyperparameters θ by probabilistic inversion of the data $\mathbf{y}_{\sim i_0}$. The posterior $\pi(\theta | \mathbf{y}_{\sim i_0})$ obtained in this first step can be translated into the prior distribution

$$\pi(\mathbf{x}_{i_0} | \mathbf{y}_{\sim i_0}) = \int f_{\mathbf{X}|\Theta}(\mathbf{x}_{i_0} | \theta) \pi(\theta | \mathbf{y}_{\sim i_0}) d\theta. \quad (5)$$

It represents the available knowledge about \mathbf{x}_{i_0} following the analysis of $\mathbf{y}_{\sim i_0}$ but prior to analyzing \mathbf{y}_{i_0} . With the prior $\pi(\mathbf{x}_{i_0} | \mathbf{y}_{\sim i_0})$ the data \mathbf{y}_{i_0} will be inverted for \mathbf{x}_{i_0} in a subsequent parameter estimation step. The result is a Bayesian posterior distribution $\pi(\mathbf{x}_{i_0} | \mathbf{y}_{\sim i_0}, \mathbf{y}_{i_0})$. In a sequential way the estimation of \mathbf{x}_{i_0} has been based on nearly the total amount of available information.

NUMERICAL EXPERIMENT

In order to demonstrate the optimal inference of individual parameters, we devise a simple example within the domain of structural engineering for which we will conduct a simulated computer experiment. It should be understood as a benchmark application for the optimal combination of information in data analysis of engineering systems. The system under consideration is chosen to be a set of beams $i = 1, \dots, n$ with well-known lengths L_i , widths b_i and heights h_i . Beams i are composed out of a material which is subject to aleatory uncertainty in its material properties, say the Young's modulus E_i . For each individual beam i the Young's modulus E_i is assumed to be constant along the main beam axis. Across the sample of beams Young's moduli E_i are assumed to be distributed according to a lognormal distribution $\mathcal{LN}(\lambda, \zeta)$ with mean value $\mu_E = \exp(\lambda + \zeta^2/2)$ and standard deviation $\sigma_E = (\exp(2\lambda + \zeta^2)(\exp(\zeta^2) - 1))^{1/2}$. At positions s_j with $0 \leq s_j \leq L_i/2$ and for $j = 1, \dots, n_i$ the deflections $v_i(s_j)$ of individual beams under a concentrated point load F_i at midspan are given as

$$v_i(s_j) = \frac{F_i s_j}{48 E_i I_i} (3L_i^2 - 4s_j^2). \quad (6)$$

The moment of inertia is given as $I_i = b_i h_i^3/12$. A symmetric expression holds for positions s_j with $L_i/2 \leq s_j \leq L_i$. In a series of experiments beam deflections can be measured and used to estimate individual Young's moduli $\mathbf{x}_i \equiv E_i$ or the hyperparameters $\theta \equiv (\mu_E, \sigma_E)$. We will herein consider the problem of inferring individual Young's moduli E_i that will be exemplified through the inference of a specific E_{i_0} .

To that end a simulated computer experiment is conducted as described below. We choose a set of $n = 100$ beams having well-known and constant dimensions $L_i = 1$ m and $b_i = h_i = 10$ cm and being subjected to loads $F_i = 30$ kN. "True" elastic moduli E_i are randomly sampled from a lognormal distribution with "true" mean $\mu_E = 15$ GPa and standard deviation $\sigma_E = 3$ GPa. These "true" values will be treated as

“unknowns” for the simulated experimental identification of the Young’s modulus E_{i_0} of a beam i_0 . For identifying E_{i_0} we simulate a synthetic set of pseudo-data $\{\mathbf{y}_i = (y_{i1}, y_{i2}, y_{i3})\}$ for each beam $i = 1, \dots, 100$. Pseudo-observations $y_{ij} = v_i(s_j) + \epsilon_{ij}$ with $j = 1, 2, 3$ are therefore generated for positions $\mathbf{d}_i = (s_1, s_2, s_3)$ with $s_1 = 25$ cm, $s_2 = 50$ cm and $s_3 = 75$ cm by perturbing the corresponding model predictions Eq. (6) with independently sampled residuals $\epsilon_{ij} \sim \mathcal{N}(0, \sigma_{ij}^2)$. We choose comparably small residual standard deviations $\sigma_{ij} = 0.01$ cm for $i \neq i_0$ and comparably large deviations $\sigma_{i_0j} = 0.1$ cm to clearly illustrate the inferential mechanisms of the three proposed estimation programs.

We assume that the “true” hyperparameter values (μ_E, σ_E) are not known, though, prior knowledge is available. The available inferential prior distribution of the hyperparameters is set priorly independent $\pi(\boldsymbol{\theta}) = \pi(\mu_E) \pi(\sigma_E)$ with only weakly informative but proper uniform distributions $\pi(\mu_E) = \mathcal{U}(0, 100)$ and $\pi(\sigma_E) \sim \mathcal{U}(0, 100)$ (in GPa) as marginals. The hyperparameters $\boldsymbol{\theta} = (\mu_E, \sigma_E)$ will be considered nuisance to the inference of $x_i = E_{i_0}$. Within this simulated experimental setup we will study the three proposed estimation strategies, i.e. simple Bayesian updating, joint multilevel analysis and sequential Bayesian filtering.

EXPERIMENTAL RESULTS

In the following we will conduct inference of the “unknown” Young’s modulus $E_{i_0} = 17.01$ GPa of a beam i_0 by means of simple Bayesian updating based on the prior Eq. (3), a full multilevel analysis based on the marginalization of a joint posterior Eq. (4) and a Bayesian filtering program based on the prior Eq. (5). Results from simple Bayesian inversion of \mathbf{y}_{i_0} to estimate E_{i_0} serves as a reference for assessing the results from the full hierarchical treatment of the whole set of data \mathbf{y}_i with $i = 1, \dots, 100$. These analyses will indicate the amount of information that can be additionally extracted by multilevel modeling. The sequential Bayesian filtering estimation will provide insight into the inferential mechanism involved.

Bayesian Updating

The compound prior Eq. (3) may not be available in analytical form, however, one can draw samples from the prior by the method of composition. To that end one draws K samples $\boldsymbol{\theta}^{(k)}$ from the hyperparameter distribution $\pi(\boldsymbol{\theta})$. Subsequently one draws one sample $E_{i_0}^{(k)}$ from each of the parameter distributions $f_{E_{i_0}|\boldsymbol{\theta}}(E_{i_0}|\boldsymbol{\theta}^{(k)})$ corresponding to $\boldsymbol{\theta}^{(k)}$. As desired the sample of parameter values $E_{i_0}^{(k)}$ is then distributed according to the compound distribution Eq. (3). We draw $N = 10^5$ samples of the mixture prior Eq. (3) which is shown in Fig. 2. A lognormal fit to the sample is shown as well as an approximation of the mixture prior by simple one-dimensional kernel smoothing with a Gaussian kernel. It is seen that the mixture is not well described by a lognormal shape and is more adequately approximated by the kernel density estimate. In the following MCMC analysis the prior knowledge about E_{i_0} is therefore represented by the abovementioned kernel density estimate.

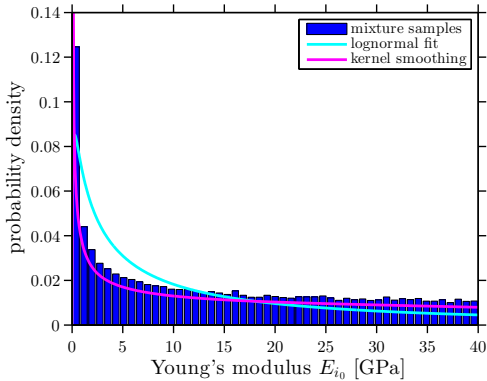


Figure 2. Bayesian updating: mixture prior

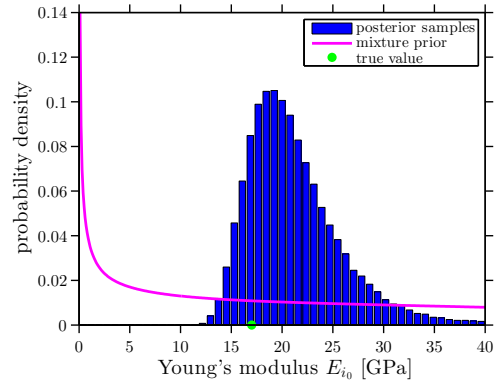


Figure 3. Bayesian updating: posterior

We accomplish simple Bayesian inversion by sampling the posterior $\pi(E_{i_0} | \mathbf{y}_{i_0})$ with a simple random walk Metropolis MCMC that uses a Gaussian proposal distribution. Although being a delicate issue we do not discuss convergence or its diagnostics at this place. With a chosen stepsize of $\varepsilon_{E_{i_0}} = 5$ GPa, i.e. the standard deviation of the Gaussian proposal distribution, the acceptance rate amounted to 63 %. The total program runtime for simulating $N = 10^5$ posterior samples amounted to $t = 3267$ s, which is comparably long due to evaluations of the mixture prior by kernel smoothing. The autocorrelation of MCMC posterior samples approaches zero after ca. 40 algorithm iterations. In Fig. 3 the resulting simulated posterior is shown.

Multilevel Analysis

While sampling the posterior $\pi(E_{i_0} | \mathbf{y}_{i_0})$ as above involved a 1-dim. parameter space, sampling the multilevel posterior $\pi(E_1, \dots, E_{100}, \mu_E, \sigma_E | \mathbf{y}_1, \dots, \mathbf{y}_n)$ involves a 102-dim. parameter space. In terms of MCMC sampling this generally demands dedicated schemes. We propose the Hamiltonian Monte Carlo (HMC) sampler (Neal 2011) in order to overcome this multilevel-typical challenge. An extensive presentation of the HMC and detailed description of our specific implementation is beyond the scope of this paper. In summary it can be said that the HMC is a highly efficient MCMC sampler for high-dimensional parameter spaces and highly correlated posteriors. It ensures superior mixing properties while maintaining high MCMC acceptance rates. However, it features additional algorithmic parameters that have to be tuned. With our final algorithm we draw $N = 10^5$ posterior samples within a total execution time of $t = 167$ s. With the HMC we achieve rapid mixing while maintaining an acceptance rate of 99.99 %. The autocorrelation of MCMC samples drops down to zero within ca. five MCMC iterations. Samples from the marginal posterior Eq. (4) can be easily extracted by sampling the joint posterior Eq. (2) and discarding samples of $E_{\sim i_0}$ and (μ_E, σ_E) . The simulated marginal density $\pi(E_{i_0} | \mathbf{y}_1, \dots, \mathbf{y}_{100})$ is shown in Fig. 4.

By comparison with the posterior $\pi(E_{i_0} | \mathbf{y}_{i_0})$ in Fig. 3 resulting from simple Bayesian updating, the additional amount of information that has been gained by solving the joint multilevel problem becomes manifest. The whole available information,

including the full set of observations \mathbf{y}_i for $i = 1, \dots, 100$, has been utilized for estimating E_{i_0} . Comparing Fig. 4 to Fig. 3 we observe the occurrence of a certain shrinkage effect. Note that while Fig. 4 shows only the single posterior marginal of E_{i_0} , MCMC sampling the joint posterior $\pi(E_1, \dots, E_{100}, \mu_E, \sigma_E | \mathbf{y}_1, \dots, \mathbf{y}_n)$ incidentally provides the marginal posterior information about (μ_E, σ_E) and $E_{\sim i_0}$.

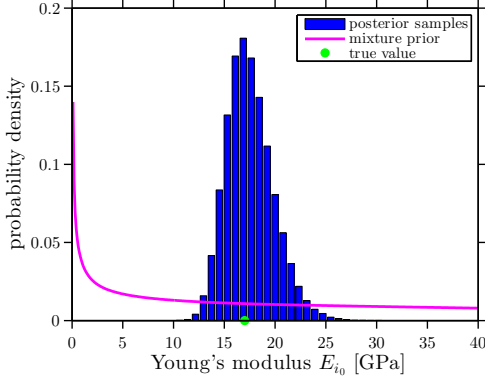


Figure 4. Multilevel analysis: posterior

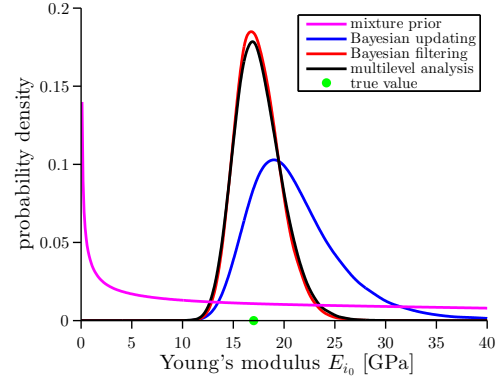


Figure 5. Summary

Bayesian Filtering

First we conduct probabilistic inversion to infer the hyperparameters (μ_E, σ_E) with data $\mathbf{y}_{\sim i_0}$. By means of the devised HMC sampler within a total runtime of $t = 167$ s we draw $N = 10^5$ samples from the posterior $\pi(\mu_E, \sigma_E | \mathbf{y}_{\sim i_0})$. We use this sample to draw samples from $\pi(E_{i_0} | \mathbf{y}_{\sim i_0})$ by the method of composition as described in the context of Eq. (5). The resulting sample is shown in Fig. 6 along with a lognormal fit to the sample and a kernel density estimate. Both the lognormal fit and the kernel density estimate reproduce the original population distribution $\mathcal{LN}(\lambda, \zeta)$ with $\mu_E = 15$ GPa and $\sigma_E = 3$ GPa quite well. With a lognormal fit of the 10^5 samples one obtains $\hat{\mu}_E = 15.04$ GPa and $\hat{\sigma}_E = 2.96$ GPa as hyperparameters. Fig. 7 show the results of a subsequent sampling from the posterior $\pi(E_{i_0} | \mathbf{y}_{\sim i_0}, \mathbf{y}_{i_0})$ with the abovementioned lognormal fit of the mixture $\pi(E_{i_0} | \mathbf{y}_{\sim i_0})$ being the prior knowledge. The total sample of size $N = 10^5$ is produced with a total runtime of $t = 25$ s with a stepsize 5 GPa and a resulting acceptance rate 44%. Execution time is comparably short because of evaluating the mixture prior by the fitted lognormal distribution.

Altogether the hyperprior $\pi(\mu_E, \sigma_E)$, the structural knowledge $f_{E|\Theta}(E_{i_0} | \mu_E, \sigma_E)$ and the observations $\mathbf{y}_{\sim i_0}$ have had influence on the estimation of $\pi(E_{i_0} | \mathbf{y}_{\sim i_0})$. In the subsequent updating step this knowledge has been updated by inversion of data \mathbf{y}_{i_0} . This shows the flow of information from data $\mathbf{y}_{\sim i_0}$ to the multilevel estimate of E_{i_0} . The information exchange takes place in an indirect way involving the hyperparameters (μ_E, σ_E) located at a “higher” level of the multilevel model. As a summary the simulated posteriors $\pi(E_{i_0} | \mathbf{y}_{i_0})$, $\pi(E_{i_0} | \mathbf{y}_{\sim i_0}, \mathbf{y}_{i_0})$ and $\pi(E_{i_0} | \mathbf{y}_1, \dots, \mathbf{y}_{100})$ that are relevant to the identification of E_{i_0} , are shown in Fig. 5. Table 1 provides a short statistical

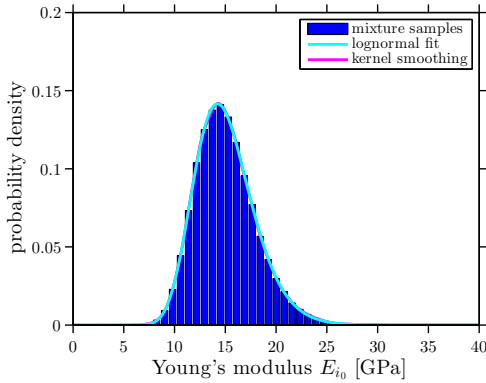


Figure 6. Bayesian filtering: updated prior

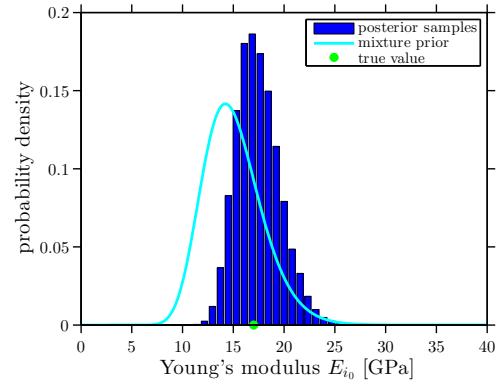


Figure 7. Bayesian filtering: posterior

summary of these posterior distributions that are obtained by simple updating, Bayesian filtering and multilevel analysis, respectively.

Table 1. Posterior summary of estimating $E_{i_0} = 17.01$ GPa

method	mean [GPa]	mode [GPa]	s.d. [GPa]
Bayesian updating	21.30	19.00	5.01
multilevel analysis	17.52	16.81	2.31
Bayesian filtering	17.45	17.78	2.17

CONCLUSION & OUTLOOK

We demonstrated the potential, the challenge and the mechanism of optimally estimating material properties as being intermediate quantities within a hierarchically defined Bayesian model. When the available information is a set of observations and prior knowledge about higher-level hyperparameters, multilevel modeling allows to obtain more information about the quantities of interest than classical Bayesian updating. The simple beam application has served as a benchmark example of Bayesian hierarchical modeling and optimal combination of information in engineering applications. Future research work will include attempts to apply the presented framework to more complex inverse problems, e.g. with additional forward model inputs that are subject to a prescribed aleatory uncertainty. Moreover a systematic investigation of how optimal combination of information influences epistemic uncertainty, e.g. with respect to the number of observations and their measurement error, is in progress.

REFERENCES

- Banks, H. T. and Zackary R. K. and Thompson, W. C. (2012). “A review of selected techniques in inverse problem nonparametric probability distribution estimation.” *J. Inv. Ill-Posed Problems*, 20(4), 429–460.
- Barbillon, P. and Celeux, G. and Grimaud, A. and Lefebvre, Y. and Rocquigny, E. (2010). “Nonlinear methods for inverse statistical problems.” *Comput. Stat. Data Anal.*, 55(1), 132–142.
- Celeux, G. and Grimaud, A. and Lefebvre, Y. and Rocquigny, E. (2010). “Identifying intrinsic variability in multivariate systems through linearized inverse methods.” *Inv. Prob. Sci. Eng.*, 18(3), 401–415
- Davidian, M. and Giltinan, D.M. (2003). “Nonlinear models for repeated measurement data: An overview and update.” *J. Agric. Biol. Environ. Stat.*, 8(4), 387–419.
- Draper, D. and Gaver, D. P. and Goel, P. K. and Greenhouse, J. B. and Hedges, L. V. and Morris, C. N. and Waternaux, C. M. (1992). “Combining Information: Statistical Issues and Opportunities for Research.” *Panel on Statistical Issues and Opportunities for Research in the Combination of Information*, National Research Council (US).
- Nagel, J. B., and Sudret, B. (2013). “Probabilistic Inversion for Estimating the Variability of Material Properties: A Bayesian Multilevel Approach.” *Proc. 11th Int. Probabilistic Workshop*, Brno (Czech Republic), November 6-8.
- Neal, R. M. (2011). “MCMC Using Hamiltonian Dynamics.” *Handbook of Markov Chain Monte Carlo*. Chapman & Hall/CRC, Handbooks of Modern Statistical Methods.
- Rocquigny, E. and Cambier, S. (2009). “Inverse probabilistic modelling of the sources of uncertainty: a non-parametric simulated-likelihood method with application to an industrial turbine vibration assessment.” *Inv. Prob. Sci. Eng.*, 17(7), 937–959.
- Robert, C. P. and Casella, G. (2004). “*Monte Carlo statistical methods (2nd)*.” Springer Verlag, Springer Series in Statistics.