A unified framework for multilevel uncertainty quantification in Bayesian inverse problems

Joseph B Nagel, Bruno B Sudret

HAL Id: hal-01432159
https://hal.science/hal-01432159

Submitted on 11 Jan 2017

A UNIFIED FRAMEWORK FOR MULTILEVEL UNCERTAINTY QUANTIFICATION IN BAYESIAN INVERSE PROBLEMS

Joseph B. Nagel, Bruno Sudret

Data Sheet

A Unified Framework for Multilevel Uncertainty Quantification in Bayesian Inverse Problems

Joseph B. Nagel[a,*], Bruno Sudret[a,**]

[a]*ETH Zürich, Institute of Structural Engineering*
Chair of Risk, Safety & Uncertainty Quantification
Stefano-Franscini-Platz 5, CH-8093 Zürich, Switzerland

Abstract

In this paper a unified probabilistic framework for solving inverse problems in the presence of epistemic and aleatory uncertainty is presented. The aim is to establish a flexible theory that facilitates Bayesian data analysis in experimental scenarios as they are commonly met in engineering practice. Problems are addressed where learning about unobservable inputs of a forward model, e.g. reducing the epistemic uncertainty of fixed yet unknown parameters and/or quantifying the aleatory uncertainty of variable inputs, is based on processing response measurements. Approaches to Bayesian inversion, hierarchical modeling and uncertainty quantification are combined into a generic framework that eventually allows to interpret and accomplish this task as multilevel model calibration. A joint problem formulation, where quantities that are not of particular interest are marginalized out from a joint posterior distribution, or an intrinsically marginal formulation, which is based on an integrated likelihood function, can be chosen according to the inferential objective and computational convenience. Fully Bayesian probabilistic inversion, i.e. the inference the variability of unobservable model inputs across a number of experiments, is derived as a special case of multilevel inversion. Borrowing strength, i.e. the optimal estimation of experiment-specific unknown forward model inputs, is introduced as a means for combining information in inverse problems. Two related statistical models for situations involving finite or zero model/measurement error are devised. Multilevel-specific obstacles to Bayesian posterior computation via Markov chain Monte Carlo are discussed. The inferential machinery of Bayesian multilevel model calibration and its underlying flow of information are studied on the basis of a system from the domain of civil engineering. A population of identically manufactured structural elements serves as an exemplary system for examining different experimental settings from the standpoint of uncertainty quantification and reduction. In a series of tests the material variability throughout the ensemble of specimens, the entirety of specimen-specific material properties and the measurement error level are inferred under various uncertainties in the problem setup.

Keywords: uncertainty quantification, inverse problems, Bayesian inference, hierarchical modeling, probabilistic inversion, borrowing strength, Markov chain Monte Carlo

1. Introduction

Main characteristics and challenges of inverse problems in engineering sciences subsume the following issues. Firstly, the ever-growing complexity of physical modeling increases the computational expense of deterministic forward simulations. Secondly, uncertainty is omnipresent and calls for an adequate mathematical formalism of representation and management. Thirdly, since data are commonly scarce or prohibitively expensive to acquire, the available information has to be carefully handled. An abstract inverse problem statement thus reads as follows. By analyzing a limited amount of data the endeavor is to optimally learn about unknown forward model inputs that are subject to epistemic uncertainty and aleatory variability. This includes deducing fixed albeit unknown forward model parameters as well as hyperparameters that determine the distribution of variable model inputs. Such a universal formulation describes a class of inverse problems that has hardly been satisfactorily solved yet. Our goal is therefore to develop a rigorous and extensive framework for formulating and solving such inverse problems in support of data analysis for engineering systems. The focus of this research is on experimental situations as they are typically encountered in this field. We emphasize aspects of uncertainty quantification and information accumulation. In order to establish a sound conceptional and computational basis for solving those problems one has to complement ideas and techniques that have been developed in different academic disciplines and scientific communities so far. This involves inverse modeling, Bayesian statistics and uncertainty quantification. In the following we will shortly survey relevant theories and practices.

In the first place we rely on the Bayesian approach to *classical inverse problems* [1, 2]. When a physical theory or a computational solver relates physical parameters to measurable quantities, i.e. the *forward model*, classical inversion is the process of reasoning or inferring unknown yet physically fixed model parameters from recorded data [3, 4]. Bayesian inference establishes a convenient probabilistic framework to accomplish this conventional type of parameter estimation and data assimilation. At least since the advent of the personal computer it is nowadays widely used in engineering applications [5, 6]. The stochastic paradigm provides a natural mechanism for the regularization of ill-posed problems, however, it requires the specification of a prior and a noise model. *Hierarchical inversion* is an extension of the classical framework that allows to set parameters of the prior and the noise model in a data-informed manner [7, 8]. While epistemic uncertainty is naturally incorporated, a shortcoming of these types of parameter estimation is that they do not account for aleatory variability.

In the second place *hierarchical statistical models* serve as the main tool for the analysis of complex systems. Those are systems that are hierarchically organized at multiple nested layers. Prominent instances include *random* and *mixed effects models* [9]. Historically those models were developed in social and biological

---

*Corresponding author
**Principal corresponding author
Email addresses: `nagel@ibk.baug.ethz.ch` (Joseph B. Nagel), `sudret@ibk.baug.ethz.ch` (Bruno Sudret)

sciences e.g. for purposes of educational research [10, 11] and pharmacokinetics/dynamics [12, 13]. Some recent reviews about the methods that were developed in these fields can be found in [14, 15]. Hierarchical modeling can be viewed from a more frequentist [16, 17] or a more Bayesian perspective [18, 19]. At the present day it is mature area of research that establishes sort of an overarching theme in modern multidisciplinary statistics. Dedicated chapters can be found in numerous standard references for Bayesian modeling and inference [20, 21]. A general observation is that hierarchical models may be complex in their probabilistic architecture whereas only little forward modeling takes place.

In the third place we respect the uncertainty taxonomy that is prevalent in risk assessment and decision making. According to this classification one distinguishes between epistemic and aleatory uncertainty [22, 23]. On one side, *epistemic uncertainty* refers to the ignorance or lack of knowledge of the observer and analyst. By taking further evidence this type of uncertainty is reducible in principle. On the contrary, *aleatory uncertainty* or *variability* refers to a trait of the system under consideration. It is a structural randomness of irreducible character. Uncertainties can be accounted for in distinct mathematical frameworks and especially the representation of ignorance is the subject matter of ongoing debates [24, 25]. Graphical statistical models such as *Bayesian probability networks* establish a powerful and widespread tool of uncertainty characterization [26, 27]. In risk-based decision making Bayesian belief networks have been adopted for their strength and flexibility in uncertainty modeling [28, 29] and their elegant mechanisms of information aggregation [30, 31].

In the fourth place *probabilistic inverse problems* constitute a challenging class of inverse problems that is of theoretical and practical relevance alike. While classical inversion is concerned with estimating uncertain yet physically fixed parameters in a series of experiments, i.e. identifying an epistemically uncertain quantity, probabilistic inversion deals with inferring the distribution of such forward model inputs that vary throughout the experiments, i.e. quantifying their aleatory variability. Previously established approaches to this interesting type of problems with *latent/hidden variable* structure subsume various approximate solutions. A frequentist technique that is premised on the simulation of an explicitly marginalized likelihood is proposed in [32]. There are also attempts to compute approximate solutions based on variants of the expectation-maximization algorithm within a linearized Gaussian frame [33] or with the aid of Kriging surrogates [34]. A methodological review of this school of probabilistic inversion is found in [35]. These methods are only partly Bayesian and suffer from the deficiency of providing mere point estimates.

The potential of hierarchical models as instruments of statistical modeling and uncertainty quantification have barely been acknowledged for the purposes of inversion in a classical sense. Hierarchical and probabilistic inversion are first steps towards preparing the Bayesian framework for the treatment of more realistic experimental scenarios. These approaches do not fully exhaust the inferential machinery of hierarchical models and the probability logic of Bayesian networks, though. In this contribution we thus aim at bridging that gap by developing a coherent Bayesian framework for managing uncertainties in such undertakings.

By drawing on the statistical theory of hierarchical models, we cast inversion under parameter uncertainty and variability as *Bayesian multilevel calibration*. This embeds a joint and a marginal problem formulation of Bayesian inference under uncertainty, both of which can be numerically solved with plain vanilla or specialized Markov chain Monte Carlo methods.

This new formulation of *multilevel inversion* is especially well-adapted to the challenges that engineers are frequently faced with. It naturally allows for sophisticated uncertainty modeling which comprises both epistemic and aleatory uncertainty. The inclusion of the former is straightforward whereas the introduction of the latter is an extension to classical parameter estimation. It also promotes a pervasive "blackbox" point of view on the forward model. While this is inevitable in many complex applications, it is not readily compliant with traditional hierarchical models. Previously established strategies of enhanced uncertainty quantification, e.g. hierarchical and probabilistic inversion, emerge as special cases of the proposed general problem formulation. This also offers the opportunity to cope with probabilistic inversion within a fully Bayesian setting. Beyond these extensions some fundamentally new possibilities are suggested. Based on the probabilistic calculus of multilevel models, we develop a novel formulation of multilevel inversion in the zero-noise and "perfect" data limit. The statistical effect of "borrowing strength" or "optimal combination of information" is transferred and applied to inverse problems.

The article is organized as follows. In Section 2 we will elaborate a general Bayesian framework for the treatment of uncertainty and variability in inverse problems. This is followed by a discussion about Bayesian inference in the context of multilevel inversion in Section 3. Thereafter Section 4 will provide an extension of the framework that will allow for handling "perfect" data. Probabilistic inversion and borrowing strength will be placed in context in Sections 5 and 6, respectively. Dedicated Bayesian computations based on Markov chain Monte Carlo are reviewed in Section 7. Lastly in Section 8 we will conduct a selection of numerical case studies, where by considering various experimental situations and uncertainty setups the very potential and the computational challenges of the devised modeling paradigm will become transparent.

2. Bayesian Multilevel Modeling

Due to the lack of a unified terminology, we define a *hierarchical* or *multilevel model* as "an overall system model that is hierarchically composed of deterministic and stochastic submodels". Important types of submodels comprise physical models of the deterministic system components (Section 2.1), prior descriptions of parameter uncertainty and variability (Section 2.2) and residual representations of forward model prediction errors (Section 2.3). From these submodels we will assemble a generic Bayesian multilevel model (Section 2.4). This will represent the overall system under consideration including its deterministic and probabilistic aspects.

2.1. Forward Model: Deterministic Subsystem

A so-called *forward model* is a mathematical representation of the physical system or phenomenon under investigation. More formally the forward model is a function

$$\mathcal{M} : \mathcal{D}_{\boldsymbol{m}} \times \mathcal{D}_{\boldsymbol{x}} \times \mathcal{D}_{\boldsymbol{\zeta}} \times \mathcal{D}_{\boldsymbol{d}} \to \mathcal{D}_{\tilde{\boldsymbol{y}}}$$
$$(\boldsymbol{m}, \boldsymbol{x}, \boldsymbol{\zeta}, \boldsymbol{d}) \mapsto \tilde{\boldsymbol{y}} = \mathcal{M}(\boldsymbol{m}, \boldsymbol{x}, \boldsymbol{\zeta}, \boldsymbol{d}), \tag{1}$$

that maps inputs $(\boldsymbol{m}, \boldsymbol{x}, \boldsymbol{\zeta}, \boldsymbol{d}) \in \mathcal{D}_{\boldsymbol{m}} \times \mathcal{D}_{\boldsymbol{x}} \times \mathcal{D}_{\boldsymbol{\zeta}} \times \mathcal{D}_{\boldsymbol{d}}$ from its domain to outputs $\tilde{\boldsymbol{y}} \in \mathcal{D}_{\tilde{\boldsymbol{y}}}$ from its codomain. Forward model arguments $(\boldsymbol{m}, \boldsymbol{x}, \boldsymbol{\zeta}, \boldsymbol{d})$ constitute physical parameters, while its responses $\tilde{\boldsymbol{y}}$ are predictions of observable quantities.

We distinguish between four different types of forward model inputs. They differ in their (un)certain nature when a number of experiments is carried out. There are fixed albeit unknown model parameters $\boldsymbol{m} \in \mathcal{D}_{\boldsymbol{m}}$ that are subject to epistemic uncertainty, two different types of inputs $\boldsymbol{x} \in \mathcal{D}_{\boldsymbol{x}}$ and $\boldsymbol{\zeta} \in \mathcal{D}_{\boldsymbol{\zeta}}$ that are subject to aleatory variability and well-known experimental conditions $\boldsymbol{d} \in \mathcal{D}_{\boldsymbol{d}}$.

2.2. Prior Model: Input Uncertainty

Forward model inputs $\boldsymbol{d}$ constitute perfectly known conditions that prevail during experimentation. In line with this they are deterministic arguments of the forward model. Experimental conditions may differ throughout the experiments, i.e. each of the experiments $i = 1, \ldots, n$ is conducted being subject to an experiment-specific condition $\boldsymbol{d}_i$.

Proper forward model parameters $\boldsymbol{m}$ are constant throughout the experiments $i = 1, \ldots, n$, yet they have unknown values. In Bayesian fashion the available prior or expert knowledge about the true parameter values is represented as a random variable or vector

$$\boldsymbol{M} \sim \pi_{\boldsymbol{M}}(\boldsymbol{m}). \tag{2}$$

The Bayesian prior distribution $\pi_{\boldsymbol{M}}(\boldsymbol{m})$ quantifies a subjective degree of plausibility or belief about the true parameter values $\boldsymbol{m}$. This is the Bayesian account for *epistemic uncertainty*. The uncertainty is reducible in the sense that Bayesian data analysis gives rise to a posterior probability model.

Forward model inputs $\boldsymbol{\zeta}$ are subject to a form of variability that is well-known, e.g. it could be ascertained in previous experiments or due to prior considerations. Rather than being constant throughout the experiments $i = 1, \ldots, n$, these variable inputs take on experiment-specific realizations $\boldsymbol{\zeta}_i$, all of which are unknown. The corresponding Bayesian prior representation is as mutually independent random variables

$$\boldsymbol{Z}_i \sim f_{\boldsymbol{Z}}(\boldsymbol{\zeta}_i; \boldsymbol{\theta}_{\boldsymbol{Z}_i}), \text{ for } i = 1, \ldots, n. \tag{3}$$

Distributions $f_{\boldsymbol{Z}}(\boldsymbol{\zeta}_i; \boldsymbol{\theta}_{\boldsymbol{Z}_i})$ specify prior knowledge about the experiment-specific unknowns that is of structural quality. They are prescribed by well-known hyperparameters $\boldsymbol{\theta}_{\boldsymbol{Z}_i} \in \mathcal{D}_{\boldsymbol{\theta}_{\boldsymbol{Z}}}$, e.g. shape, scale and dependency parameters, that possibly differ across the experiments. Due to stochastic independence, the

appropriate joint Bayesian prior model follows as

$$(\boldsymbol{Z}_1, \ldots, \boldsymbol{Z}_n) \sim \prod_{i=1}^{n} f_{\boldsymbol{Z}}(\boldsymbol{\zeta}_i; \boldsymbol{\theta}_{\boldsymbol{Z}_i}). \tag{4}$$

This is a Bayesian conception of *aleatory variability*, i.e. an uncertainty that is of structural nature. Hereinafter this probability model will also be referred to as *prescribed uncertainty*. It is irreducible in the sense that by Bayesian data analysis of the experiments $i = 1, \ldots, n$ "past" realizations $\boldsymbol{\zeta}_i$ can be inferred in principle, whereas the knowledge about "future" realizations $\boldsymbol{\zeta}_{i'}$ in further experiments $i' = n+1, \ldots, n+n'$ cannot be improved. "Future" realizations still feature a structural uncertainty $\boldsymbol{Z}_{i'} \sim f_{\boldsymbol{Z}}(\boldsymbol{\zeta}_{i'}; \boldsymbol{\theta}_{\boldsymbol{Z}_{i'}})$ that is prescribed by hyperparameters $\boldsymbol{\theta}_{\boldsymbol{Z}_{i'}} \in \mathcal{D}_{\boldsymbol{\theta}_{\boldsymbol{Z}}}$.

Another Bayesian notion of a similar type allows to account for forward model inputs $\boldsymbol{x}$ that are subject to a sort of variability which itself is unknown. For $1 = 1, \ldots, n$ these variables take on experiment-specific realizations $\boldsymbol{x}_i$, neither of which are known. Bayesian prior modeling is build upon conditionally independent random variables

$$(\boldsymbol{X}_i | \boldsymbol{\Theta}_{\boldsymbol{X}} = \boldsymbol{\theta}_{\boldsymbol{X}}) \sim f_{\boldsymbol{X} | \boldsymbol{\Theta}_{\boldsymbol{X}}}(\boldsymbol{x}_i | \boldsymbol{\theta}_{\boldsymbol{X}}), \ \text{for} \ i = 1, \ldots, n. \tag{5}$$

The conditional probability distribution $f_{\boldsymbol{X} | \boldsymbol{\Theta}_{\boldsymbol{X}}}(\boldsymbol{x}_i | \boldsymbol{\theta}_{\boldsymbol{X}})$ represents a structural kind of prior knowledge about the experiment-specific unknowns. Its determining hyperparameters $\boldsymbol{\theta}_{\boldsymbol{X}} \in \mathcal{D}_{\boldsymbol{\theta}_{\boldsymbol{X}}}$, e.g. location, dispersion and correlation parameters, themselves are fixed yet unknown. Hence these hyperparameters are priorly modeled as a random vector

$$\boldsymbol{\Theta}_{\boldsymbol{X}} \sim \pi_{\boldsymbol{\Theta}_{\boldsymbol{X}}}(\boldsymbol{\theta}_{\boldsymbol{X}}). \tag{6}$$

The Bayesian prior distribution $\pi_{\boldsymbol{\Theta}_{\boldsymbol{X}}}(\boldsymbol{\theta}_{\boldsymbol{X}})$ constitutes the subjective prior belief or available prior knowledge about the true hyperparameter values. In the statistical literature hyperprior elicitation is exhaustively discussed especially for variance hyperparameters [36–38]. Consequently the joint distribution of the unknowns of this prior model is given as

$$(\boldsymbol{X}_1, \ldots, \boldsymbol{X}_n, \boldsymbol{\Theta}_{\boldsymbol{X}}) \sim \left(\prod_{i=1}^{n} f_{\boldsymbol{X} | \boldsymbol{\Theta}_{\boldsymbol{X}}}(\boldsymbol{x}_i | \boldsymbol{\theta}_{\boldsymbol{X}}) \right) \pi_{\boldsymbol{\Theta}_{\boldsymbol{X}}}(\boldsymbol{\theta}_{\boldsymbol{X}}). \tag{7}$$

The joint prior distribution of experiment-specific realizations follows by marginalizing Eq. (7) over the hyperparameters $\boldsymbol{\theta}_{\boldsymbol{X}}$. Then one has

$$(\boldsymbol{X}_1, \ldots, \boldsymbol{X}_n) \sim \int_{\mathcal{D}_{\boldsymbol{\theta}_{\boldsymbol{X}}}} \left(\prod_{i=1}^{n} f_{\boldsymbol{X} | \boldsymbol{\Theta}_{\boldsymbol{X}}}(\boldsymbol{x}_i | \boldsymbol{\theta}_{\boldsymbol{X}}) \right) \pi_{\boldsymbol{\Theta}_{\boldsymbol{X}}}(\boldsymbol{\theta}_{\boldsymbol{X}}) \, \mathrm{d}\boldsymbol{\theta}_{\boldsymbol{X}}. \tag{8}$$

This is a form of *exchangeability* [39, 40] that realizes some "similarity" of the intermediate variables, i.e. the joint distribution of the sequence $(\boldsymbol{X}_1, \ldots, \boldsymbol{X}_n)$ equals the one of $(\boldsymbol{X}_{\tau(1)}, \ldots, \boldsymbol{X}_{\tau(n)})$ for any index permutation $\tau: \{1, \ldots, n\} \rightarrow \{1, \ldots, n\}$. In the present form Eq. (8), exchangeability establishes another Bayesian approach to aleatory variability. Unlike the prescribed uncertainty in Eq. (4), this form of uncertainty is

partially reducible in the sense that the "fuzziness" inherent in Eq. (8) can be reduced by learning about $\boldsymbol{\theta_X}$ in "past" experiments i. "Past" realizations $\boldsymbol{x}_i$ can also be inferred, however, even if the hyperparameters $\boldsymbol{\theta_X}$ would be known, the realizations $\boldsymbol{x}_{i'}$ of "future" experiments i' would still carry the structural prior uncertainty $\boldsymbol{X}_{i'} \sim f_{\boldsymbol{X}|\boldsymbol{\Theta_X}}(\boldsymbol{x}_{i'}|\boldsymbol{\theta_X})$.

In short, on the one hand we have *parametric priors* $\pi_{\boldsymbol{M}}(\boldsymbol{m})$ and $\pi_{\boldsymbol{\Theta_X}}(\boldsymbol{\theta_X})$ that in Eqs. (2) and (6) embody knowledge about global unknowns $\boldsymbol{m}$ and $\boldsymbol{\theta_X}$. On the other hand we have *structural priors* $f_{\boldsymbol{Z}}(\boldsymbol{\zeta}_i; \boldsymbol{\theta_{Z_i}})$ and $f_{\boldsymbol{X}|\boldsymbol{\Theta_X}}(\boldsymbol{x}_i|\boldsymbol{\theta_X})$ that encapsulate structural prior knowledge about the problem, and that for $i = 1, \ldots, n$ establish the prior model of experiment-specific unknowns $\boldsymbol{x}_i$ and $\boldsymbol{\zeta}_i$ through Eqs. (4) and (8).

2.3. Residual Model: Output Imperfection

Besides a representation of forward model input uncertainty and variability, an integral constituent of statistical approaches to inversion is a *residual representation* of forward model output discrepancy or imperfection. Due to measurement errors, numerical approximations and general inadequacies, even if all inputs $(\boldsymbol{m}, \boldsymbol{x}_i, \boldsymbol{\zeta}_i, \boldsymbol{d}_i)$ were perfectly known, predictions $\tilde{\boldsymbol{y}}_i = \mathcal{M}(\boldsymbol{m}, \boldsymbol{x}_i, \boldsymbol{\zeta}_i, \boldsymbol{d}_i)$ are expected to deviate from real observations $\boldsymbol{y}_i$. These imperfections can be accounted for by a *statistical data model*

$$\boldsymbol{y}_i = \tilde{\boldsymbol{y}}_i + \boldsymbol{\varepsilon}_i = \mathcal{M}(\boldsymbol{m}, \boldsymbol{x}_i, \boldsymbol{\zeta}_i, \boldsymbol{d}_i) + \boldsymbol{\varepsilon}_i, \text{ for } i = 1, \ldots, n, \tag{9}$$

where residual terms $\boldsymbol{\varepsilon}_i \in \mathcal{D}_{\boldsymbol{\varepsilon}}$ are assumed to be realizations of random variables $\boldsymbol{E}_i \sim f_{\boldsymbol{E}}(\boldsymbol{\varepsilon}_i; \boldsymbol{\Sigma}_i)$. Commonly one employs normal distributions $f_{\boldsymbol{E}}(\boldsymbol{\varepsilon}_i; \boldsymbol{\Sigma}_i) = \mathcal{N}(\boldsymbol{\varepsilon}_i; \boldsymbol{0}, \boldsymbol{\Sigma}_i)$ with mean $\boldsymbol{0}$ and possibly experiment-specific, symmetric and positive-semidefinite covariance matrices $\boldsymbol{\Sigma}_i$. Consequently, through a change of variables whose Jacobian determinant equals one, observations are viewed as outcomes $\boldsymbol{y}_i$ of random variables

$$(\boldsymbol{Y}_i | \boldsymbol{M} = \boldsymbol{m}, \boldsymbol{X}_i = \boldsymbol{x}_i, \boldsymbol{Z}_i = \boldsymbol{\zeta}_i) \sim f_{\boldsymbol{E}}(\boldsymbol{y}_i - \mathcal{M}(\boldsymbol{m}, \boldsymbol{x}_i, \boldsymbol{\zeta}_i, \boldsymbol{d}_i); \boldsymbol{\Sigma}_i), \text{ for } i = 1, \ldots, n. \tag{10}$$

For given values of the direct forward model inputs $(\boldsymbol{m}, \boldsymbol{x}_i, \boldsymbol{\zeta}_i, \boldsymbol{d}_i)$, data are viewed as random variables $(\boldsymbol{Y}_i | \boldsymbol{m}, \boldsymbol{x}_i, \boldsymbol{\zeta}_i)$ with conditional distributions $f(\boldsymbol{y}_i | \boldsymbol{m}, \boldsymbol{x}_i, \boldsymbol{\zeta}_i) = f_{\boldsymbol{E}}(\boldsymbol{y}_i - \mathcal{M}(\boldsymbol{m}, \boldsymbol{x}_i, \boldsymbol{\zeta}_i, \boldsymbol{d}_i); \boldsymbol{\Sigma}_i)$. Note that $f(\boldsymbol{y}_i | \boldsymbol{m}, \boldsymbol{x}_i, \boldsymbol{\zeta}_i) = f(\boldsymbol{y}_i | \boldsymbol{m}, \boldsymbol{x}_i, \boldsymbol{\zeta}_i, \boldsymbol{\theta_X})$ is independent of $\boldsymbol{\theta_X}$.

The specification of the residual model, i.e. quantifying the parameters of $\boldsymbol{\Sigma}_i$, is an essential part of calibrating the forward model and the experimental apparatus. In many experimental situations a model of the *prediction error* is not known a priori, though. Nevertheless, the structure of the prediction error model can be selected [41] and its parameters can be introduced as unknown hyperparameters that undergo calibration [42]. This also includes systematic forward model deviations [43, 44]. Moreover one could treat the form of the forward model $\mathcal{M}$ itself as uncertain/random [45, 46] and select the most plausible class via Bayesian model selection [47, 48]. By adding another layer of uncertainty on top of the outlined setup and at a higher associated cost, the aforementioned principles of assessing structural and parametric forward model uncertainty can be readily applied in multilevel models [49].

Based on random variable transformations, in Section 4 we will extend the framework by a model for analyzing "perfect" observations $\tilde{\boldsymbol{y}}_i = \mathcal{M}(\boldsymbol{m}, \boldsymbol{x}_i, \boldsymbol{\zeta}_i, \boldsymbol{d}_i)$ in the zero-noise limit $|\boldsymbol{\varepsilon}_i| \to 0$. This mathematical formulation will explain the variability in the data exclusively by a Bayesian prior model of input variability as outlined in the preceding Section 2.2.

2.4. Multilevel Model: Overall System

We start from the premise that if not denoted or stated otherwise, random vectors and variables are (conditionally) independent, e.g. the global forward model parameters $\boldsymbol{M}$ and the hyperparameters $\boldsymbol{\Theta_X}$ are understood to be priorly independent. Thus $\pi(\boldsymbol{m}, \boldsymbol{\theta_X}) = \pi_M(\boldsymbol{m})\,\pi_{\boldsymbol{\Theta_X}}(\boldsymbol{\theta_X})$ applies for their joint prior distribution. Note that this is not a necessity of the formulation, though. Moreover, we strictly reserve conditional notation for the stochastic dependency of random variables on outcomes of other random variables, e.g. the aleatory variables $(\boldsymbol{X}_i|\boldsymbol{\theta_X})$ are conditionally dependent on realizations $\boldsymbol{\Theta_X} = \boldsymbol{\theta_X}$. The stochastic variables $(\boldsymbol{Y}_i|\boldsymbol{m}, \boldsymbol{x}_i, \boldsymbol{\zeta}_i)$ are conditioned on random outcomes $\boldsymbol{M} = \boldsymbol{m}$, $\boldsymbol{X}_i = \boldsymbol{x}_i$ and $\boldsymbol{Z}_i = \boldsymbol{\zeta}_i$, nonetheless they depend on deterministic quantities $\boldsymbol{d}_i$ and $\boldsymbol{\Sigma}_i$, too. Similarly the aleatory variables $\boldsymbol{Z}_i$ are dependent on $\boldsymbol{\theta}_{\boldsymbol{Z}_i}$ in a way that is not explicitly indicated. In order to keep track of all stochastic and deterministic relations the index i serves as a bookkeeping mark.

Deterministic aspects of the system are covered by the forward model Eq. (1). Parametric priors in Eqs. (2) and (6) and structural priors in Eqs. (3) and (5) represent input uncertainty and variability. The model Eq. (10) condenses basic assumptions regarding the prediction error. Altogether those submodels are combined into a greater model of the whole system. The overall probability model is summarized as

$$(\boldsymbol{Y}_i|\boldsymbol{m}, \boldsymbol{x}_i, \boldsymbol{\zeta}_i) \sim f_{\boldsymbol{E}}\big(\boldsymbol{y}_i - \mathcal{M}(\boldsymbol{m}, \boldsymbol{x}_i, \boldsymbol{\zeta}_i, \boldsymbol{d}_i); \boldsymbol{\Sigma}_i\big), \tag{11a}$$

$$\boldsymbol{M} \sim \pi_M(\boldsymbol{m}), \tag{11b}$$

$$\boldsymbol{Z}_i \sim f_{\boldsymbol{Z}}(\boldsymbol{\zeta}_i; \boldsymbol{\theta}_{\boldsymbol{Z}_i}), \tag{11c}$$

$$(\boldsymbol{X}_i|\boldsymbol{\theta_X}) \sim f_{\boldsymbol{X}|\boldsymbol{\Theta_X}}(\boldsymbol{x}_i|\boldsymbol{\theta_X}), \tag{11d}$$

$$\boldsymbol{\Theta_X} \sim \pi_{\boldsymbol{\Theta_X}}(\boldsymbol{\theta_X}). \tag{11e}$$

Adopting a subjectivist viewpoint, this complex probability model Eq. (11) formalizes degrees of belief of how the data have been realized in the experiments $i = 1, \ldots, n$. According to our previous definition it is a generic Bayesian multilevel model. An intuitive representation of this multilevel model is provided by a directed acyclic graph (DAG) [26, 27] such as shown in Fig. 1.

3. Inference in Multilevel Models

We will now discuss statistical inference. In particular we will demonstrate how conditioning on the observables and marginalization out nuisance are elegant inferential tools of Bayesian multilevel inversion.
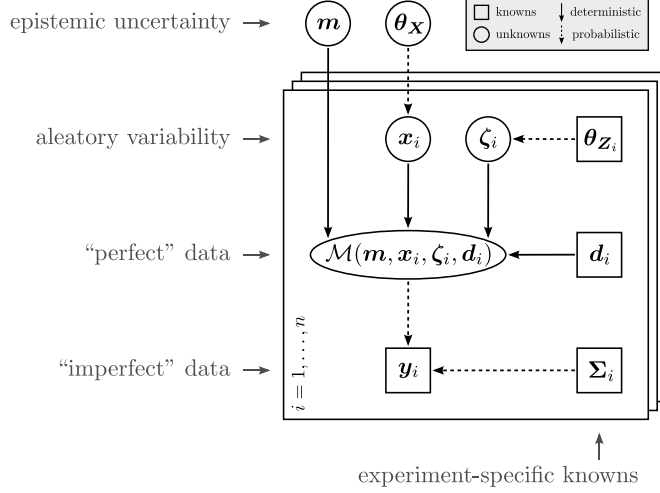
Figure 1: **DAG of the generic multilevel model.** Vertices symbolize known ($\square$) or unknown ($\bigcirc$) quantities, while directed edges represent their deterministic ($\longrightarrow$) or probabilistic ($\cdots\!\!\blacktriangleright$) relations. Global parameters $(\boldsymbol{m}, \boldsymbol{\theta_X})$ are subject to epistemic uncertainty, whereas experiment-specific realizations $(\langle \boldsymbol{x}_i \rangle, \langle \boldsymbol{\zeta}_i \rangle)$ are subject to aleatory variability. Known quantities comprise the data $\langle \boldsymbol{y}_i \rangle$ just as well as experiment-specific knowns $(\langle \boldsymbol{\theta_{Z_i}} \rangle, \langle \boldsymbol{d}_i \rangle, \langle \boldsymbol{\Sigma}_i \rangle)$ located at different levels of the hierarchy.

A pivotal joint problem formulation will be devised. Afterwards an intrinsically marginal problem variant will be presented in Section 3.1.

In the following $\langle \boldsymbol{q}_i \rangle$ denotes a sequence $\langle \boldsymbol{q}_i \rangle_{1 \leq i \leq n} = (\boldsymbol{q}_1, \boldsymbol{q}_2, \ldots, \boldsymbol{q}_n)$. Summarizing the available parametric and structural prior knowledge in Eqs. (11b) to (11e), the *joint prior* of the entirety of unknowns $(\boldsymbol{m}, \langle \boldsymbol{x}_i \rangle, \langle \boldsymbol{\zeta}_i \rangle, \boldsymbol{\theta_X})$ factorizes as

$$\pi\big(\boldsymbol{m}, \langle \boldsymbol{x}_i \rangle, \langle \boldsymbol{\zeta}_i \rangle, \boldsymbol{\theta_X}\big) = \left(\prod_{i=1}^{n} f_{\boldsymbol{X}|\boldsymbol{\Theta_X}}(\boldsymbol{x}_i|\boldsymbol{\theta_X}) \right) \left(\prod_{i=1}^{n} f_{\boldsymbol{Z}}(\boldsymbol{\zeta}_i; \boldsymbol{\theta_{Z_i}}) \right) \pi_{\boldsymbol{\Theta_X}}(\boldsymbol{\theta_X}) \, \pi_{\boldsymbol{M}}(\boldsymbol{m}). \tag{12}$$

This prior depends only on the collection of experiment-specific hyperparameters $\langle \boldsymbol{\theta_{Z_i}} \rangle$. With the model of single observations in Eq. (11a) one can formulate a conditional distribution for the total data $\langle \boldsymbol{y}_i \rangle$. For given values of the unknowns $(\boldsymbol{m}, \langle \boldsymbol{x}_i \rangle, \langle \boldsymbol{\zeta}_i \rangle)$ this yields the product $f(\langle \boldsymbol{y}_i \rangle | \boldsymbol{m}, \langle \boldsymbol{x}_i \rangle, \langle \boldsymbol{\zeta}_i \rangle) = \prod_{i=1}^{n} f_{\boldsymbol{E}}(\boldsymbol{y}_i - \mathcal{M}(\boldsymbol{m}, \boldsymbol{x}_i, \boldsymbol{\zeta}_i, \boldsymbol{d}_i); \boldsymbol{\Sigma}_i)$. It depends on experiment-specific knowns $(\langle \boldsymbol{d}_i \rangle, \langle \boldsymbol{\Sigma}_i \rangle)$.

With that said, one can derive the *joint posterior* of the totality of unknowns $(\boldsymbol{m}, \langle \boldsymbol{x}_i \rangle, \langle \boldsymbol{\zeta}_i \rangle, \boldsymbol{\theta_X})$ by conditioning on the acquired data $\langle \boldsymbol{y}_i \rangle$. By virtue of Bayes' theorem one obtains

$$\pi\big(\boldsymbol{m}, \langle \boldsymbol{x}_i \rangle, \langle \boldsymbol{\zeta}_i \rangle, \boldsymbol{\theta_X} | \langle \boldsymbol{y}_i \rangle\big) = \frac{1}{C} \left(\prod_{i=1}^{n} f_{\boldsymbol{E}}\big(\boldsymbol{y}_i - \mathcal{M}(\boldsymbol{m}, \boldsymbol{x}_i, \boldsymbol{\zeta}_i, \boldsymbol{d}_i); \boldsymbol{\Sigma}_i\big) \right) \pi\big(\boldsymbol{m}, \langle \boldsymbol{x}_i \rangle, \langle \boldsymbol{\zeta}_i \rangle, \boldsymbol{\theta_X}\big). \tag{13}$$

This posterior Eq. (13) is implicitly dependent on experiment-specific knowns $(\langle \boldsymbol{\theta_{Z_i}} \rangle, \langle \boldsymbol{d}_i \rangle, \langle \boldsymbol{\Sigma}_i \rangle)$. It is the central object in Bayesian multilevel model calibration.

The model evidence C is the total probability of the realized data $\langle \boldsymbol{y}_i \rangle$, given the underlying multilevel

9

model. When introducing the notation $\mathrm{d}\langle \boldsymbol{q}_i \rangle = \mathrm{d}\boldsymbol{q}_1 \, \mathrm{d}\boldsymbol{q}_2 \ldots \mathrm{d}\boldsymbol{q}_n$ one can write this as

$$C = \int_{\mathcal{D}_{\boldsymbol{m}}} \int_{\mathcal{D}_{\boldsymbol{x}}^n} \int_{\mathcal{D}_{\boldsymbol{\zeta}}^n} \int_{\mathcal{D}_{\boldsymbol{\theta}_{\boldsymbol{X}}}} \left(\prod_{i=1}^{n} f_{\boldsymbol{E}}\big(\boldsymbol{y}_i - \mathcal{M}(\boldsymbol{m}, \boldsymbol{x}_i, \boldsymbol{\zeta}_i, \boldsymbol{d}_i); \boldsymbol{\Sigma}_i \big) \right) \pi\big(\boldsymbol{m}, \langle \boldsymbol{x}_i \rangle, \langle \boldsymbol{\zeta}_i \rangle, \boldsymbol{\theta}_{\boldsymbol{X}}\big) \, \mathrm{d}\boldsymbol{m} \, \mathrm{d}\langle \boldsymbol{x}_i \rangle \, \mathrm{d}\langle \boldsymbol{\zeta}_i \rangle \, \mathrm{d}\boldsymbol{\theta}_{\boldsymbol{X}}. \quad (14)$$

For the Bayesian computations that will be reviewed in Section 7, the factor of proportionality C does not have to be computed explicitly. For that reason it will be occasionally omitted from now on.

One may define a likelihood in order to write the joint posterior Eq. (13) in the familiar textbook-form $\pi(\text{unknowns}|\text{data}) \propto \mathcal{L}(\text{unknowns}; \text{data}) \, \pi(\text{unknowns})$. Regarded as a function of the unknowns $(\boldsymbol{m}, \langle \boldsymbol{x}_i \rangle, \langle \boldsymbol{\zeta}_i \rangle)$, the *joint likelihood* evaluates the densities in Eq. (10) for the collected data $\langle \boldsymbol{y}_i \rangle$ by

$$\mathcal{L}\big(\boldsymbol{m}, \langle \boldsymbol{x}_i \rangle, \langle \boldsymbol{\zeta}_i \rangle; \langle \boldsymbol{y}_i \rangle \big) = \prod_{i=1}^{n} f_{\boldsymbol{E}}\big(\boldsymbol{y}_i - \mathcal{M}(\boldsymbol{m}, \boldsymbol{x}_i, \boldsymbol{\zeta}_i, \boldsymbol{d}_i); \boldsymbol{\Sigma}_i \big). \quad (15)$$

Apart from its functional arguments and the data it also depends on the total number of experiment-specific knowns $(\langle \boldsymbol{d}_i \rangle, \langle \boldsymbol{\Sigma}_i \rangle)$. It does not depend on $\boldsymbol{\theta}_{\boldsymbol{X}}$, though.

Subsequent to formulating the joint posterior Eq. (13) the marginal of the *quantities of interest* (QoI) is obtained by integrating out *nuisance* [50, 51]. For instance, given that $(\boldsymbol{m}, \boldsymbol{\theta}_{\boldsymbol{X}})$ are declared QoI and the latent variables $(\langle \boldsymbol{x}_i \rangle, \langle \boldsymbol{\zeta}_i \rangle)$ are considered nuisance, the correspondingly marginalized posterior becomes

$$\pi\big(\boldsymbol{m}, \boldsymbol{\theta}_{\boldsymbol{X}} | \langle \boldsymbol{y}_i \rangle\big) = \int_{\mathcal{D}_{\boldsymbol{x}}^n} \int_{\mathcal{D}_{\boldsymbol{\zeta}}^n} \pi\big(\boldsymbol{m}, \langle \boldsymbol{x}_i \rangle, \langle \boldsymbol{\zeta}_i \rangle, \boldsymbol{\theta}_{\boldsymbol{X}} | \langle \boldsymbol{y}_i \rangle\big) \, \mathrm{d}\langle \boldsymbol{x}_i \rangle \, \mathrm{d}\langle \boldsymbol{\zeta}_i \rangle. \quad (16)$$

Similarly, provided that hidden variables $(\langle \boldsymbol{x}_i \rangle, \langle \boldsymbol{\zeta}_i \rangle)$ are proclaimed QoI and $(\boldsymbol{m}, \boldsymbol{\theta}_{\boldsymbol{X}})$ are deemed nuisance parameters, appropriately marginalizing the posterior distribution gives

$$\pi\big(\langle \boldsymbol{x}_i \rangle, \langle \boldsymbol{\zeta}_i \rangle | \langle \boldsymbol{y}_i \rangle\big) = \int_{\mathcal{D}_{\boldsymbol{m}}} \int_{\mathcal{D}_{\boldsymbol{\theta}_{\boldsymbol{X}}}} \pi\big(\boldsymbol{m}, \langle \boldsymbol{x}_i \rangle, \langle \boldsymbol{\zeta}_i \rangle, \boldsymbol{\theta}_{\boldsymbol{X}} | \langle \boldsymbol{y}_i \rangle\big) \, \mathrm{d}\boldsymbol{m} \, \mathrm{d}\boldsymbol{\theta}_{\boldsymbol{X}}. \quad (17)$$

3.1. Marginalized Formulation

A common scenario is that inferential interest focuses on the global parameters $(\boldsymbol{m}, \boldsymbol{\theta}_{\boldsymbol{X}})$. In this particular case, instead of marginalizing the joint posterior distribution Eq. (13) as in Eq. (16), based on an integrated likelihood function one can formulate an inherently marginal problem [52–54]. One therefore constructs a marginalized observation model

$$(\boldsymbol{Y}_i | \boldsymbol{m}, \boldsymbol{\theta}_{\boldsymbol{X}}) \sim f(\boldsymbol{y}_i | \boldsymbol{m}, \boldsymbol{\theta}_{\boldsymbol{X}}), \ \text{ for } i = 1, \ldots, n, \quad (18a)$$

$$(\boldsymbol{M}, \boldsymbol{\Theta}_{\boldsymbol{X}}) \sim \pi(\boldsymbol{m}, \boldsymbol{\theta}_{\boldsymbol{X}}) = \pi_{\boldsymbol{M}}(\boldsymbol{m}) \, \pi_{\boldsymbol{\Theta}_{\boldsymbol{X}}}(\boldsymbol{\theta}_{\boldsymbol{X}}). \quad (18b)$$

The marginalized model consists of the prior distribution Eq. (18b) of the QoI $(\boldsymbol{m}, \boldsymbol{\theta}_{\boldsymbol{X}})$ and the probability model Eq. (18a) of the observations $\boldsymbol{y}_i$. By integrating out the aleatory variables $(\boldsymbol{x}_i, \boldsymbol{\zeta}_i)$ in the following way, one can obtain the marginal distributions of the observations

$$f(\boldsymbol{y}_i | \boldsymbol{m}, \boldsymbol{\theta}_{\boldsymbol{X}}) = \int_{\mathcal{D}_{\boldsymbol{x}}} \int_{\mathcal{D}_{\boldsymbol{\zeta}}} f_{\boldsymbol{E}}\big(\boldsymbol{y}_i - \mathcal{M}(\boldsymbol{m}, \boldsymbol{x}_i, \boldsymbol{\zeta}_i, \boldsymbol{d}_i); \boldsymbol{\Sigma}_i\big) \, f_{\boldsymbol{X}|\boldsymbol{\Theta}_{\boldsymbol{X}}}(\boldsymbol{x}_i | \boldsymbol{\theta}_{\boldsymbol{X}}) \, f_{\boldsymbol{Z}}(\boldsymbol{\zeta}_i; \boldsymbol{\theta}_{\boldsymbol{Z}_i}) \, \mathrm{d}\boldsymbol{x}_i \, \mathrm{d}\boldsymbol{\zeta}_i. \quad (19)$$

10

These distributions are conditional on $(\boldsymbol{m}, \boldsymbol{\theta_X})$ and dependent on $(\boldsymbol{\theta}_{\boldsymbol{Z}_i}, \boldsymbol{d}_i, \boldsymbol{\Sigma}_i)$. Following this, one can easily formulate an *integrated* or *marginalized likelihood*. Evaluated for the actual data $\langle \boldsymbol{y}_i \rangle$ and seen as a function of the QoI $(\boldsymbol{m}, \boldsymbol{\theta_X})$ this version of the likelihood reads as

$$\mathcal{L}\big(\boldsymbol{m}, \boldsymbol{\theta_X}; \langle \boldsymbol{y}_i \rangle\big) = f\big(\langle \boldsymbol{y}_i \rangle | \boldsymbol{m}, \boldsymbol{\theta_X}\big) = \prod_{i=1}^{n} f(\boldsymbol{y}_i | \boldsymbol{m}, \boldsymbol{\theta_X}). \tag{20}$$

It is the likelihood function corresponding to the case of eliminating all intermediate unobservables $(\langle \boldsymbol{x}_i \rangle, \langle \boldsymbol{\zeta}_i \rangle)$ with Eq. (19) on the likelihood rather than on the posterior level. Note that frequentist inference of $(\boldsymbol{m}, \boldsymbol{\theta_X})$ could be based on this integrated likelihood formulation. Fully Bayesian inference, however, proceeds by formulating the corresponding posterior distribution. With the prior Eq. (18b) and the likelihood Eq. (20), the posterior is obtained on grounds of Bayes' law

$$\pi\big(\boldsymbol{m}, \boldsymbol{\theta_X} | \langle \boldsymbol{y}_i \rangle\big) = \frac{1}{C} \, \mathcal{L}\big(\boldsymbol{m}, \boldsymbol{\theta_X}; \langle \boldsymbol{y}_i \rangle\big) \, \pi(\boldsymbol{m}, \boldsymbol{\theta_X}). \tag{21}$$

One can easily derive that the normalizing constant C equals Eq. (14) and show that the posteriors Eqs. (16) and (21) are identical. This means that, as far as the inference of $(\boldsymbol{m}, \boldsymbol{\theta_X})$ is concerned, the two problem formulations Eqs. (11) and (18) are equivalent. Those problem formulations pose different numerical obstacles, though. In Section 7 we will discuss Bayesian computations and their multilevel-related issues.

3.1.1. Monte Carlo Integration

In Eq. (20) the marginalized likelihood $\mathcal{L}(\boldsymbol{m}, \boldsymbol{\theta_X}; \langle \boldsymbol{y}_i \rangle)$ is a product of integrals $f(\boldsymbol{y}_i | \boldsymbol{m}, \boldsymbol{\theta_X})$. Most often it is not possible to perform the marginalization in Eq. (19) analytically. Still it can be approximately computed through deterministic or stochastic schemes of numerical integration.

The density $f(\boldsymbol{y}_i | \boldsymbol{m}, \boldsymbol{\theta_X})$ can be evaluated for arbitrary arguments $\boldsymbol{y}_i$ and for fixed values $(\boldsymbol{m}, \boldsymbol{\theta_X})$. A simple numerical means to that end rests upon stochastic integration via the Monte Carlo (MC) method

$$\hat{f}(\boldsymbol{y}_i | \boldsymbol{m}, \boldsymbol{\theta_X}) = \frac{1}{K} \sum_{k=1}^{K} f_{\boldsymbol{E}}\left(\boldsymbol{y}_i - \tilde{\boldsymbol{v}}_i^{(k)}; \boldsymbol{\Sigma}_i\right), \quad \text{with} \quad \begin{cases} \boldsymbol{x}_i^{(k)} \sim f_{\boldsymbol{X} | \boldsymbol{\Theta_X}}(\boldsymbol{x}_i^{(k)} | \boldsymbol{\theta_X}), \\ \boldsymbol{\zeta}_i^{(k)} \sim f_{\boldsymbol{Z}}(\boldsymbol{\zeta}_i^{(k)}; \boldsymbol{\theta}_{\boldsymbol{Z}_i}), \\ \tilde{\boldsymbol{v}}_i^{(k)} = \mathcal{M}(\boldsymbol{m}, \boldsymbol{x}_i^{(k)}, \boldsymbol{\zeta}_i^{(k)}, \boldsymbol{d}_i) \end{cases} \text{for } k = 1, \dots, K. \tag{22}$$

For $k = 1, \dots, K$ forward model inputs $\boldsymbol{x}_i^{(k)}$ and $\boldsymbol{\zeta}_i^{(k)}$ are independently sampled from their population distributions $f_{\boldsymbol{X} | \boldsymbol{\Theta_X}}(\boldsymbol{x}_i^{(k)} | \boldsymbol{\theta_X})$ and $f_{\boldsymbol{Z}}(\boldsymbol{\zeta}_i^{(k)}; \boldsymbol{\theta}_{\boldsymbol{Z}_i})$, respectively. In turn responses $\tilde{\boldsymbol{v}}_i^{(k)} = \mathcal{M}(\boldsymbol{m}, \boldsymbol{x}_i^{(k)}, \boldsymbol{\zeta}_i^{(k)}, \boldsymbol{d}_i)$ are computed accordingly. For evaluating $\mathcal{L}(\boldsymbol{m}, \boldsymbol{\theta_X}; \langle \boldsymbol{y}_i \rangle)$ as a function of the unknowns $(\boldsymbol{m}, \boldsymbol{\theta_X})$, one has to simulate Eq. (22) for the observations $\boldsymbol{y}_i$ that were taken in the experiments $i = 1, \dots, n$. Thus a simple MC-based estimator of the marginalized likelihood is given as

$$\hat{\mathcal{L}}\big(\boldsymbol{m}, \boldsymbol{\theta_X}; \langle \boldsymbol{y}_i \rangle\big) = \prod_{i=1}^{n} \hat{f}(\boldsymbol{y}_i | \boldsymbol{m}, \boldsymbol{\theta_X}). \tag{23}$$

11

The stochastic simulator Eq. (23) may be costly and numerically inefficient in terms of the number of runs K of the deterministic model. It should be understood as an instructive proof for the feasibility of computing the marginal posterior Eq. (21). In practice more advanced simulators, e.g. based on importance sampling, can be applied in similar fashion [55, 56]. More generally speaking, any method for computing the model evidence in classical Bayesian inference is applicable [57].

4. Zero-Noise & "Perfect" Data

In Section 2.3 the residual model was introduced as a representation of the discrepancy between model predictions and measurements. This conditional model had equipped the data space $\mathcal{D}_{\tilde{y}}$ with a probability measure. As a consequence, in Eq. (11a) observations were regarded as $\boldsymbol{y}_i = \tilde{\boldsymbol{y}}_i + \boldsymbol{\varepsilon}_i$ with a random outcome $\boldsymbol{\varepsilon}_i$. However, experimental situations may occur where direct access to

$$\tilde{\boldsymbol{y}}_i = \mathcal{M}(\boldsymbol{m}, \boldsymbol{x}_i, \boldsymbol{\zeta}_i, \boldsymbol{d}_i), \ \text{for} \ i = 1, \dots, n \tag{24}$$

is granted, e.g. due to noise-free measurements and a "sufficiently accurate" forward model [58]. The data $\langle \tilde{\boldsymbol{y}}_i \rangle$ is then only explained by uncertainty of the forward model inputs as described in Section 2.2, without being subject to prediction errors. Hereafter we will refer this scenario as to involve "perfect" data [59, 60]. A statistical model that is appropriate for "perfect" data can be formulated as

$$(\tilde{\boldsymbol{Y}}_i | \boldsymbol{m}, \boldsymbol{\theta_X}) \sim f(\tilde{\boldsymbol{y}}_i | \boldsymbol{m}, \boldsymbol{\theta_X}), \ \text{for} \ i = 1, \dots, n, \tag{25a}$$

$$(\boldsymbol{M}, \boldsymbol{\Theta_X}) \sim \pi(\boldsymbol{m}, \boldsymbol{\theta_X}) = \pi_M(\boldsymbol{m}) \, \pi_{\boldsymbol{\Theta_X}}(\boldsymbol{\theta_X}). \tag{25b}$$

As before, Eq. (25b) embodies the available prior knowledge about the unknowns $(\boldsymbol{m}, \boldsymbol{\theta_X})$. Conditional random variables in Eq. (25a) are constructed by forward uncertainty propagation as follows. The independent input uncertainties $(\boldsymbol{X}_i | \boldsymbol{\theta_X}) \sim f_{\boldsymbol{X}|\boldsymbol{\Theta_X}}(\boldsymbol{x}_i | \boldsymbol{\theta_X})$ and $\boldsymbol{Z}_i \sim f_{\boldsymbol{Z}}(\boldsymbol{\zeta}_i; \boldsymbol{\theta}_{\boldsymbol{Z}_i})$, that are defined for given $(\boldsymbol{\theta_X}, \boldsymbol{\theta}_{\boldsymbol{Z}_i})$, are propagated through the forward model $\mathcal{M}$, while the inputs $(\boldsymbol{m}, \boldsymbol{d}_i)$ are fixed. The density of the resulting random variables $(\tilde{\boldsymbol{Y}}_i | \boldsymbol{m}, \boldsymbol{\theta_X}) = \mathcal{M}(\boldsymbol{m}, (\boldsymbol{X}_i | \boldsymbol{\theta_X}), \boldsymbol{Z}_i, \boldsymbol{d}_i)$ at $\tilde{\boldsymbol{y}}_i \in \mathcal{D}_{\tilde{y}}$ is found as

$$f(\tilde{\boldsymbol{y}}_i | \boldsymbol{m}, \boldsymbol{\theta_X}) = \int\limits_{\mathcal{D}_{\boldsymbol{x}}} \int\limits_{\mathcal{D}_{\boldsymbol{\zeta}}} \delta\big(\tilde{\boldsymbol{y}}_i - \mathcal{M}(\boldsymbol{m}, \boldsymbol{x}_i, \boldsymbol{\zeta}_i, \boldsymbol{d}_i)\big) \, f_{\boldsymbol{X}|\boldsymbol{\Theta_X}}(\boldsymbol{x}_i | \boldsymbol{\theta_X}) \, f_{\boldsymbol{Z}}(\boldsymbol{\zeta}_i; \boldsymbol{\theta}_{\boldsymbol{Z}_i}) \, \mathrm{d}\boldsymbol{x}_i \, \mathrm{d}\boldsymbol{\zeta}_i, \tag{26}$$

where δ denotes the Dirac delta distribution. This endows the response space $\mathcal{D}_{\tilde{y}}$ with a proper probability model. Inspecting Eqs. (19) and (26) reveals that the marginal model Eq. (18) approaches the "perfect" data model Eq. (25) in the zero-noise limit $\|\boldsymbol{\Sigma}_i\| \to 0$. With the distributions $f(\tilde{\boldsymbol{y}}_i | \boldsymbol{m}, \boldsymbol{\theta_X})$, that are conditioned on $(\boldsymbol{m}, \boldsymbol{\theta_X})$ and dependent on experiment-specific knowns $(\boldsymbol{d}_i, \boldsymbol{\theta}_{\boldsymbol{Z}_i})$, one can formulate the corresponding likelihood function as

$$\mathcal{L}\big(\boldsymbol{m}, \boldsymbol{\theta_X}; \langle \tilde{\boldsymbol{y}}_i \rangle\big) = f\big(\langle \tilde{\boldsymbol{y}}_i \rangle | \boldsymbol{m}, \boldsymbol{\theta_X}\big) = \prod_{i=1}^{n} f(\tilde{\boldsymbol{y}}_i | \boldsymbol{m}, \boldsymbol{\theta_X}). \tag{27}$$

For given data $\langle \tilde{\boldsymbol{y}}_i \rangle$ it is viewed as a function of the unknowns $(\boldsymbol{m}, \boldsymbol{\theta_X})$ that also depends on $(\langle \boldsymbol{\theta}_{\boldsymbol{Z}_i} \rangle, \langle \boldsymbol{d}_i \rangle)$. As usual Bayesian data analysis proceeds by conditioning on the data $\langle \tilde{\boldsymbol{y}}_i \rangle$. With the prior Eq. (25b) and the likelihood Eq. (27) the posterior follows through Bayes' rule

$$\pi\big(\boldsymbol{m}, \boldsymbol{\theta_X} \,|\, \langle \tilde{\boldsymbol{y}}_i \rangle \big) = \frac{1}{\tilde{C}} \, \mathcal{L}\big(\boldsymbol{m}, \boldsymbol{\theta_X}; \langle \tilde{\boldsymbol{y}}_i \rangle \big) \, \pi(\boldsymbol{m}, \boldsymbol{\theta_X}). \tag{28}$$

The factor of proportionality $\tilde{C}$ in the posterior density Eq. (28) is given as the marginal probability density of the effectively acquired data $\langle \tilde{\boldsymbol{y}}_i \rangle$. It thus writes $\tilde{C} = \iint \mathcal{L}(\boldsymbol{m}, \boldsymbol{\theta_X}; \langle \tilde{\boldsymbol{y}}_i \rangle) \, \pi(\boldsymbol{m}, \boldsymbol{\theta_X}) \, \mathrm{d}\boldsymbol{m} \, \mathrm{d}\boldsymbol{\theta_X}$.

4.1. Kernel Density Estimation

The likelihood function $\mathcal{L}(\boldsymbol{m}, \boldsymbol{\theta_X}; \langle \tilde{\boldsymbol{y}}_i \rangle)$ in Eq. (27) is grounded on probability densities $f(\tilde{\boldsymbol{y}}_i | \boldsymbol{m}, \boldsymbol{\theta_X})$. Likelihood evaluations therefore require forward uncertainty propagation Eq. (26). In the majority of cases this complicated problem can only be approximately solved. A possible approach is to use MC uncertainty propagation in combination with kernel density estimation (KDE) [61].

Let $\mathcal{K}_{\boldsymbol{H}}(\tilde{\boldsymbol{y}}) = |\boldsymbol{H}|^{-1/2} \mathcal{K}(\boldsymbol{H}^{-1/2} \tilde{\boldsymbol{y}})$ be the scaled kernel that is defined by a kernel function $\mathcal{K}$ and the symmetric and positive-definite bandwidth matrix $\boldsymbol{H}$. A KDE of the density $f(\tilde{\boldsymbol{y}}_i | \boldsymbol{m}, \boldsymbol{\theta_X})$ in Eq. (26) as a function of $\tilde{\boldsymbol{y}}_i$ and for fixed values of $(\boldsymbol{m}, \boldsymbol{\theta_X})$ is given as

$$\hat{f}(\tilde{\boldsymbol{y}}_i | \boldsymbol{m}, \boldsymbol{\theta_X}) = \frac{1}{K} \sum_{k=1}^{K} \mathcal{K}_{\boldsymbol{H}} \left(\tilde{\boldsymbol{y}}_i - \tilde{\boldsymbol{v}}_i^{(k)} \right), \; \text{with} \; \left\{ \begin{array}{l} \boldsymbol{x}_i^{(k)} \sim f_{\boldsymbol{X}|\boldsymbol{\Theta_X}}(\boldsymbol{x}_i^{(k)} | \boldsymbol{\theta_X}), \\ \boldsymbol{\zeta}_i^{(k)} \sim f_{\boldsymbol{Z}}(\boldsymbol{\zeta}_i^{(k)}; \boldsymbol{\theta}_{\boldsymbol{Z}_i}), \\ \tilde{\boldsymbol{v}}_i^{(k)} = \mathcal{M}(\boldsymbol{m}, \boldsymbol{x}_i^{(k)}, \boldsymbol{\zeta}_i^{(k)}, \boldsymbol{d}_i) \end{array} \right\} \; \text{for} \; k = 1, \ldots, K. \tag{29}$$

Analogously to Eq. (22), for $k = 1, \ldots, K$ forward model inputs $\boldsymbol{x}_i^{(k)}$ and $\boldsymbol{\zeta}_i^{(k)}$ are randomly drawn from their parent distributions $f_{\boldsymbol{X}|\boldsymbol{\Theta_X}}(\boldsymbol{x}_i^{(k)} | \boldsymbol{\theta_X})$ and $f_{\boldsymbol{Z}}(\boldsymbol{\zeta}_i^{(k)}; \boldsymbol{\theta}_{\boldsymbol{Z}_i})$ and responses $\tilde{\boldsymbol{v}}_i^{(k)} = \mathcal{M}(\boldsymbol{m}, \boldsymbol{x}_i^{(k)}, \boldsymbol{\zeta}_i^{(k)}, \boldsymbol{d}_i)$ are computed. Subsequently the sample $(\tilde{\boldsymbol{v}}_i^{(1)}, \ldots, \tilde{\boldsymbol{v}}_i^{(K)})$ serves as a proxy for the distribution $f(\tilde{\boldsymbol{y}}_i | \boldsymbol{m}, \boldsymbol{\theta_X})$. Estimating $\mathcal{L}(\boldsymbol{m}, \boldsymbol{\theta_X}; \langle \tilde{\boldsymbol{y}}_i \rangle)$ is based on evaluating the KDE in Eq. (29) for arguments $(\boldsymbol{m}, \boldsymbol{\theta_X})$ and for the observations $\tilde{\boldsymbol{y}}_i$ corresponding to experiments $i = 1, \ldots, n$. On these grounds, the likelihood function $\mathcal{L}(\boldsymbol{m}, \boldsymbol{\theta_X}; \langle \tilde{\boldsymbol{y}}_i \rangle)$ is approximated as

$$\hat{\mathcal{L}}\big(\boldsymbol{m}, \boldsymbol{\theta_X}; \langle \tilde{\boldsymbol{y}}_i \rangle \big) = \prod_{i=1}^{n} \hat{f}(\tilde{\boldsymbol{y}}_i | \boldsymbol{m}, \boldsymbol{\theta_X}). \tag{30}$$

Similarly to Eq. (23) this is an expensive statistical estimation program that involves forward uncertainty quantification and tends to require a high number K of calls to the forward code. Further challenges intrinsically related to computing the posterior Eq. (28) of the "perfect" data model will be discussed in Section 7.

5. Probabilistic Inversion

The introduced Bayesian multilevel model Eq. (11) acts as a toolkit for statistical model building. It forms some kind of superstructure that embeds a variety of stochastic inverse problems as special cases. In this section we will show how different well-known types of inverse problems are obtained by omitting global parameters and/or experiment-specific variables accordingly.

Classical or *simple Bayesian inversion* is concerned with the estimation of fixed yet unknown parameters $\boldsymbol{m}$ of the physical simulator [3, 4]. The related DAG is pictured in Fig. 2a. In this context the term "simple" merely refers to the degree of sophistication of the input uncertainty model. As a matter of fact classical inversion may not be a simple problem at all. It typically calls for a high number of forward solves. The engineering community therefore relies on customized strategies in order to ameliorate the computational burden to Bayesian inference in real-case problems. This includes the employment of polynomial chaos expansions as forward model substitutes [62–64], advanced stochastic simulation techniques [65, 66] and forward model reduction methods [67, 68].

Probabilistic inversion features a more elaborate two-level representation of input uncertainty [69, 70]. Rather than aiming at an unknown constant $\boldsymbol{m}$, inference concentrates on the hyperparameters $\boldsymbol{\theta_X}$ that determine the variability of $\langle \boldsymbol{x}_i \rangle$ through $f_{\boldsymbol{X}|\boldsymbol{\Theta_X}}(\boldsymbol{x}_i|\boldsymbol{\theta_X})$. A DAG belonging to probabilistic inversion is depicted in Fig. 2b. Building upon probabilistic inversion one may have variable inputs $\langle \boldsymbol{\zeta}_i \rangle$, the distributions of which $f_{\boldsymbol{Z}}(\boldsymbol{\zeta}_i; \boldsymbol{\theta_{Z_i}})$ are prescribed by $\langle \boldsymbol{\theta_{Z_i}} \rangle$. Unless experiment-specific realizations of those variables are of inferential interest, they act as additional nuisance parameters impeding the inference of the QoI. The correspondingly extended DAG is provided in Fig. 2c. Of course, more complex modeling scenarios can be envisaged. An application example where inference targets both parameters of the type $\boldsymbol{m}$ and $\boldsymbol{\theta_X}$, in the presence of additional nuisance parameters $\langle \boldsymbol{\zeta}_i \rangle$, can be found in [59, 60].

The problem that we call probabilistic inversion shall not be confused with the identically named problem of finding an input distribution of a forward model given its output distribution [71, 72]. Commonly engineering applications do not allow to exercise this type of uncertainty backpropagation. The amount and structure of the data being available do not permit to fully specify a response distribution while expert knowledge refers to physical parameters instead.

At this point we have a closer look at probabilistic inversion. It results from removing the forward model inputs $\boldsymbol{m}$ and $\langle \boldsymbol{\zeta}_i \rangle$ from the overall system Eq. (11) and from declaring $\boldsymbol{\theta_X}$ as QoI and $\langle \boldsymbol{x}_i \rangle$ as nuisance variables. For the sake of completeness we summarize the associated multilevel model as

$$(\boldsymbol{Y}_i|\boldsymbol{x}_i) \sim f_{\boldsymbol{E}}\big(\boldsymbol{y}_i - \mathcal{M}(\boldsymbol{x}_i, \boldsymbol{d}_i); \boldsymbol{\Sigma}_i\big), \tag{31a}$$

$$(\boldsymbol{X}_i|\boldsymbol{\theta_X}) \sim f_{\boldsymbol{X}|\boldsymbol{\Theta_X}}(\boldsymbol{x}_i|\boldsymbol{\theta_X}), \tag{31b}$$

$$\boldsymbol{\Theta_X} \sim \pi_{\boldsymbol{\Theta_X}}(\boldsymbol{\theta_X}). \tag{31c}$$

(a) **Simple inversion.**   (b) **Probabilistic inversion.**   (c) **Additional nuisance.**
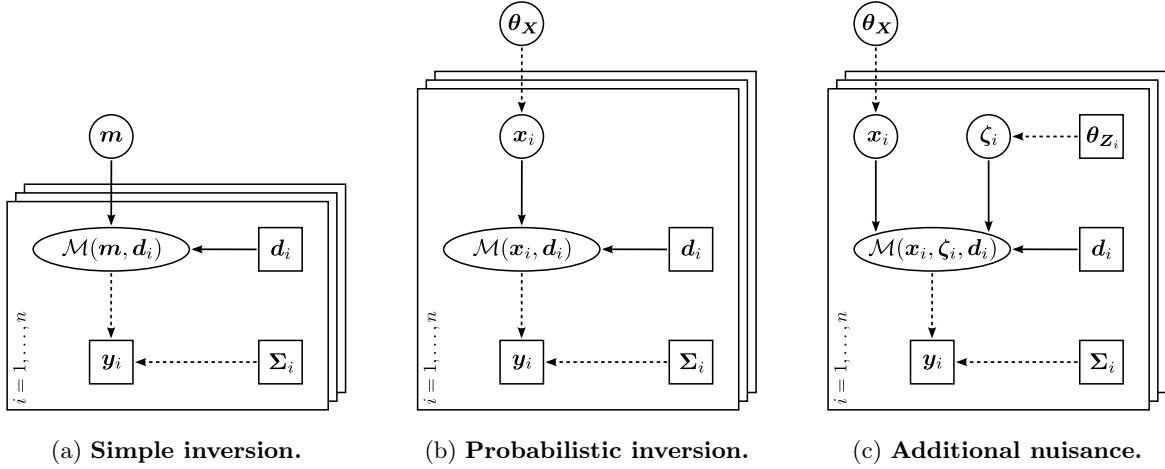
Figure 2: **Various DAGs.** Simple inversion, i.e. the estimation of an unknown $\boldsymbol{m}$, is visualized in (a), whereas (b) shows a DAG of probabilistic inversion, i.e. the inference of $\boldsymbol{\theta_X}$ that governs the variability of experiment-specific $\boldsymbol{x}_i$. An upgrade of probabilistic inversion, where a prescribed uncertainty has been introduced in nuisance variables $\boldsymbol{\zeta}_i$, is depicted in (c).

Joint Bayesian inference is accomplished by conditioning on the realized data $\langle \boldsymbol{y}_i \rangle$. Up to a normalization factor, according to Bayes' law the posterior density is given as

$$\pi\big(\langle \boldsymbol{x}_i \rangle, \boldsymbol{\theta_X} \,|\, \langle \boldsymbol{y}_i \rangle\big) \propto \left(\prod_{i=1}^{n} f_{\boldsymbol{E}}\big(\boldsymbol{y}_i - \mathcal{M}(\boldsymbol{x}_i, \boldsymbol{d}_i); \boldsymbol{\Sigma}_i\big) \right) \left(\prod_{i=1}^{n} f_{\boldsymbol{X}|\boldsymbol{\Theta_X}}(\boldsymbol{x}_i | \boldsymbol{\theta_X}) \right) \pi_{\boldsymbol{\Theta_X}}(\boldsymbol{\theta_X}). \tag{32}$$

Equivalent to integrating out nuisance $\langle \boldsymbol{x}_i \rangle$ from the joint posterior Eq. (32) as in Eq. (16), one can base inference of $\boldsymbol{\theta_X}$ on an inherently marginal problem formulation [32, 35]. Similar to Eqs. (19) and (20) the marginalized likelihood function for that case is derived as

$$\mathcal{L}\big(\boldsymbol{\theta_X}; \langle \boldsymbol{y}_i \rangle\big) = f\big(\langle \boldsymbol{y}_i \rangle | \boldsymbol{\theta_X}\big) = \prod_{i=1}^{n} \int_{\mathcal{D}_{\boldsymbol{x}}} f_{\boldsymbol{E}}\big(\boldsymbol{y}_i - \mathcal{M}(\boldsymbol{x}_i, \boldsymbol{d}_i); \boldsymbol{\Sigma}_i\big) \, f_{\boldsymbol{X}|\boldsymbol{\Theta_X}}(\boldsymbol{x}_i | \boldsymbol{\theta_X}) \, \mathrm{d}\boldsymbol{x}_i. \tag{33}$$

With the marginalized likelihood function Eq. (33) and the marginal prior distribution Eq. (31c), the unscaled version of the marginal posterior reduces to

$$\pi\big(\boldsymbol{\theta_X} \,|\, \langle \boldsymbol{y}_i \rangle\big) \propto \mathcal{L}\big(\boldsymbol{\theta_X}; \langle \boldsymbol{y}_i \rangle\big) \, \pi_{\boldsymbol{\Theta_X}}(\boldsymbol{\theta_X}). \tag{34}$$

Exemplary comparisons of the numerical efficiency for sampling joint posteriors of the form Eq. (32) and marginal posteriors of the form Eq. (34) are found in [69, 70].

Approximate *two-stage approaches* have been proposed for inferring aleatory parameter variability in inverse problems, e.g. the context of random fields [73–76]. In the first stage n separate inverse problems are solved, i.e. for each experiment $i = 1, \ldots, n$ an estimator $\hat{\boldsymbol{x}}_i$ of the realization $\boldsymbol{x}_i$ is computed. As a second step the hyperparameters $\boldsymbol{\theta_X}$ are identified by statistical analysis of the estimates $\langle \hat{\boldsymbol{x}}_i \rangle$. However, two-stage methods suffer from the dependence on a sufficient amount of data available for both of the stages and their

tendency to overestimate second-order central moments [14, 15]. Those issues are due to a fundamental inconsistency in treating epistemic and aleatory uncertainty.

Classical inverse problems are sometimes phrased within a hierarchical frame [7, 8]. Formally this is a special case of probabilistic inversion with $n = 1$. The intermediate unknowns $\boldsymbol{x}_1$ are commonly the QoI in this type of *hierarchical inversion*. Their prior $\pi(\boldsymbol{x}_1) = \int f_{\boldsymbol{X}|\boldsymbol{\Theta}_{\boldsymbol{X}}}(\boldsymbol{x}_1|\boldsymbol{\theta}_{\boldsymbol{X}}) \pi_{\boldsymbol{\Theta}_{\boldsymbol{X}}}(\boldsymbol{\theta}_{\boldsymbol{X}}) \, \mathrm{d}\boldsymbol{\theta}_{\boldsymbol{X}}$ decomposes into a conditional distribution $f_{\boldsymbol{X}|\boldsymbol{\Theta}_{\boldsymbol{X}}}(\boldsymbol{x}_1|\boldsymbol{\theta}_{\boldsymbol{X}})$ and a marginal one $\pi_{\boldsymbol{\Theta}_{\boldsymbol{X}}}(\boldsymbol{\theta}_{\boldsymbol{X}})$. However, other than in probabilistic inversion, Eq. (31b) is not interpreted as aleatory variability. Instead it can be viewed as leaving the prior for $\boldsymbol{x}_1$ incompletely specified [7], i.e. relaxing the assumption of a parametric prior $\pi(\boldsymbol{x}_1; \boldsymbol{\theta}_{\boldsymbol{X}}) = f_{\boldsymbol{X}|\boldsymbol{\Theta}_{\boldsymbol{X}}}(\boldsymbol{x}_1|\boldsymbol{\theta}_{\boldsymbol{X}})$ for a specific value $\boldsymbol{\theta}_{\boldsymbol{X}}$. Alternatively Eq. (32) suggests that the prior hyperparameters $\boldsymbol{\theta}_{\boldsymbol{X}}$ can be estimated along with $\boldsymbol{x}_1$. The prior in this case is given as $\pi(\boldsymbol{x}_1, \boldsymbol{\theta}_{\boldsymbol{X}}) = f_{\boldsymbol{X}|\boldsymbol{\Theta}_{\boldsymbol{X}}}(\boldsymbol{x}_1|\boldsymbol{\theta}_{\boldsymbol{X}}) \pi_{\boldsymbol{\Theta}_{\boldsymbol{X}}}(\boldsymbol{\theta}_{\boldsymbol{X}})$. For solving ill-posed problems this can be seen as an automatic determination of the regularization parameters [8].

6. Combination of Information

In the preceding Section 5 we declared the hyperparameters $\boldsymbol{\theta}_{\boldsymbol{X}}$ as QoI and latent quantities $\langle \boldsymbol{x}_i \rangle$ as nuisance. When this choice is reversed, i.e. proclaiming $\langle \boldsymbol{x}_i \rangle$ as the QoI and treating $\boldsymbol{\theta}_{\boldsymbol{X}}$ as nuisance, then the Bayesian multilevel model Eq. (31) allows for an optimal type of inference [77]. This effect is sometimes referred to as *optimal combination of information* or *borrowing strength*. To our best knowledge, it has been pointed out for the first time in [78]. As we will see, the term "optimal" has to be understood with respect to the total amount of information processed, e.g. the acquired data and the available parametric and structural prior knowledge. Optimal combination of information seems to be largely understudied in inverse problems with missing data structure. By taking the marginal viewpoint of Eq. (34), the additional advantages that the joint formulation Eq. (32) offers are often overlooked.

Based on the hierarchical model Eq. (31), in this section we will show how to "borrow strength" in inverse problems. The optimal inference of a specific $\boldsymbol{x}_{i_0}$ for some $i_0 \in \{1, \dots, n\}$ is demonstrated. We pursue three different estimation programs in order to investigate how inferring $\boldsymbol{x}_{i_0}$ can be accomplished by wholly or only partially utilizing the informational resources. In Section 6.1 we will present a simple Bayesian updating approach, in respect to which the principle and mechanism of borrowing strength is emphasized by means of multilevel inference in Section 6.3. Beforehand we will devise a sequential filtering approach in Section 6.2 that will serve as an illustration of the underlying flow of information.

6.1. Simple Updating

In this first approach, inference of $\boldsymbol{x}_{i_0}$ will be solely based on the single observation $\boldsymbol{y}_{i_0}$, the informational content of $f_{\boldsymbol{E}}(\boldsymbol{y}_{i_0} - \mathcal{M}(\boldsymbol{x}_{i_0}, \boldsymbol{d}_{i_0}); \boldsymbol{\Sigma}_{i_0})$, the structural prior $f_{\boldsymbol{X}|\boldsymbol{\Theta}_{\boldsymbol{X}}}(\boldsymbol{x}_{i_0}|\boldsymbol{\theta}_{\boldsymbol{X}})$ and the parametric prior $\pi_{\boldsymbol{\Theta}_{\boldsymbol{X}}}(\boldsymbol{\theta}_{\boldsymbol{X}})$. Utilizing the prior information one can formulate a Bayesian prior distribution for $\boldsymbol{x}_{i_0}$. By

marginalizing over the hyperparameters $\boldsymbol{\theta_X}$ this reads as

$$\pi(\boldsymbol{x}_{i_0}) = \int\limits_{\mathcal{D}_{\boldsymbol{\theta_X}}} f_{\boldsymbol{X}|\boldsymbol{\Theta_X}}(\boldsymbol{x}_{i_0}|\boldsymbol{\theta_X})\,\pi_{\boldsymbol{\Theta_X}}(\boldsymbol{\theta_X})\,\mathrm{d}\boldsymbol{\theta_X}. \tag{35}$$

This compound distribution represents the uncertainty that $\boldsymbol{x}_{i_0}$ priorly carries. Ensuing from the prior Eq. (35), analyzing the piece of data $\boldsymbol{y}_{i_0}$ is accomplished by constructing the corresponding posterior. It is proportional to $\pi(\boldsymbol{x}_{i_0}|\boldsymbol{y}_{i_0}) \propto f_{\boldsymbol{E}}(\boldsymbol{y}_{i_0} - \mathcal{M}(\boldsymbol{x}_{i_0}, \boldsymbol{d}_{i_0}); \boldsymbol{\Sigma}_{i_0})\,\pi(\boldsymbol{x}_{i_0})$. We remark that the approach is formally reminiscent of hierarchical inversion as discussed in Section 5.

While the observation $\boldsymbol{y}_{i_0}$ that is directly associated to $\boldsymbol{x}_{i_0}$ has been analyzed, the evidence that $\langle \boldsymbol{y}_{\neq i_0} \rangle$ carry about $\boldsymbol{\theta_X}$, and in turn about $\boldsymbol{x}_{i_0}$, has not yet been taken into consideration. Put another way, the hierarchical problem structure has been respected by formulating Eq. (35), however, it has only been partially exploited for learning about the QoI $\boldsymbol{x}_{i_0}$.

6.2. Sequential Filtering

For the second estimation scheme, which will be based on sequential updating, we introduce the simplifying notation $\langle \boldsymbol{q}_{\neq i_0} \rangle = (\boldsymbol{q}_1, \ldots, \boldsymbol{q}_{i_0-1}, \boldsymbol{q}_{i_0+1}, \ldots, \boldsymbol{q}_n)$. In a first step probabilistic inversion is accomplished by estimating $\boldsymbol{\theta_X}$ with the data $\langle \boldsymbol{y}_{\neq i_0} \rangle$. Similarly to Eq. (35), the resulting posterior $\pi(\boldsymbol{\theta_X}|\langle \boldsymbol{y}_{\neq i_0} \rangle)$ can be translated into a mixture distribution

$$\pi\big(\boldsymbol{x}_{i_0}|\langle \boldsymbol{y}_{\neq i_0} \rangle\big) = \int\limits_{\mathcal{D}_{\boldsymbol{\theta_X}}} f_{\boldsymbol{X}|\boldsymbol{\Theta_X}}(\boldsymbol{x}_{i_0}|\boldsymbol{\theta_X})\,\pi\big(\boldsymbol{\theta_X}|\langle \boldsymbol{y}_{\neq i_0} \rangle\big)\,\mathrm{d}\boldsymbol{\theta_X}. \tag{36}$$

It represents the uncertainty in $\boldsymbol{x}_{i_0}$ following the analysis of $\langle \boldsymbol{y}_{\neq i_0} \rangle$ but prior to analyzing $\boldsymbol{y}_{i_0}$. Thereupon the second stage of the filtering program consists in utilizing Eq. (36) as a Bayesian prior for inferring $\boldsymbol{x}_{i_0}$ by inverting $\boldsymbol{y}_{i_0}$. Bayesian updating yields the posterior distribution $\pi(\boldsymbol{x}_{i_0}|\langle \boldsymbol{y}_{\neq i_0} \rangle, \boldsymbol{y}_{i_0}) \propto f_{\boldsymbol{E}}(\boldsymbol{y}_{i_0} - \mathcal{M}(\boldsymbol{x}_{i_0}, \boldsymbol{d}_{i_0}); \boldsymbol{\Sigma}_{i_0})\,\pi(\boldsymbol{x}_{i_0}|\langle \boldsymbol{y}_{\neq i_0} \rangle)$.

Information-wise, the estimation of $\boldsymbol{\theta_X}$ has been initially based on the data $\langle \boldsymbol{y}_{\neq i_0} \rangle$, its conditional distributions $f_{\boldsymbol{E}}(\boldsymbol{y}_i - \mathcal{M}(\boldsymbol{x}_i, \boldsymbol{d}_i); \boldsymbol{\Sigma}_i)$ for $i \neq i_0$, the structural knowledge $f_{\boldsymbol{X}|\boldsymbol{\Theta_X}}(\boldsymbol{x}_i|\boldsymbol{\theta_X})$ and the parametric prior $\pi_{\boldsymbol{\Theta_X}}(\boldsymbol{\theta_X})$. While inheriting the obtained information about $\boldsymbol{\theta_X}$ by means of Eq. (36), the observation $\boldsymbol{y}_{i_0}$ has been eventually inverted for $\boldsymbol{x}_{i_0}$.

6.3. Multilevel Inversion

A full hierarchical analysis constitutes the third type of estimation. By formulating the joint posterior Eq. (32) of the collectivity of unknowns $(\langle \boldsymbol{x}_i \rangle, \boldsymbol{\theta_X})$ and marginalizing over nuisance $(\langle \boldsymbol{x}_{\neq i_0} \rangle, \boldsymbol{\theta_X})$, the posterior distribution of the QoI $\boldsymbol{x}_{i_0}$ can be written as

$$\pi\big(\boldsymbol{x}_{i_0}|\langle \boldsymbol{y}_i \rangle\big) = \int\limits_{\mathcal{D}_{\boldsymbol{x}}^{n-1}} \int\limits_{\mathcal{D}_{\boldsymbol{\theta_X}}} \pi\big(\langle \boldsymbol{x}_i \rangle, \boldsymbol{\theta_X}|\langle \boldsymbol{y}_i \rangle\big)\,\mathrm{d}\langle \boldsymbol{x}_{\neq i_0} \rangle\,\mathrm{d}\boldsymbol{\theta_X}, \tag{37}$$

17

where $\mathrm{d}\langle\boldsymbol{x}_{\neq i_0}\rangle = \mathrm{d}\boldsymbol{x}_1\ldots\mathrm{d}\boldsymbol{x}_{i_0-1}\,\mathrm{d}\boldsymbol{x}_{i_0+1}\ldots\mathrm{d}\boldsymbol{x}_n$. Note that when the joint posterior Eq. (32) is computed, other marginals than Eq. (37) can be extracted similarly.

In terms of estimating $\boldsymbol{x}_{i_0}$, the structure of the posterior Eq. (37) reveals that all the different pieces of information have been "optimally" combined during a joint learning process. From an informational point of view, the total data $\langle\boldsymbol{y}_i\rangle$, their conditional distributions $f_{\boldsymbol{E}}(\boldsymbol{y}_i - \mathcal{M}(\boldsymbol{x}_i, \boldsymbol{d}_i); \boldsymbol{\Sigma}_i)$, the structural knowledge $f_{\boldsymbol{X}|\boldsymbol{\Theta}_{\boldsymbol{X}}}(\boldsymbol{x}_i|\boldsymbol{\theta}_{\boldsymbol{X}})$ and the hyperprior $\pi_{\boldsymbol{\Theta}_{\boldsymbol{X}}}(\boldsymbol{\theta}_{\boldsymbol{X}})$ have been completely synthesized. This implies that inferring $\boldsymbol{x}_{i_0}$ "borrows" information encoded in the observations $\langle\boldsymbol{y}_{\neq i_0}\rangle$. A DAG-based visualization of the underlying flow of information is provided in Fig. 3. The deeper reason for borrowing strength to happen is the partial reducibility of the uncertainty model Eq. (8), i.e. the exchangeability of aleatory variables $\langle\boldsymbol{x}_i\rangle$.
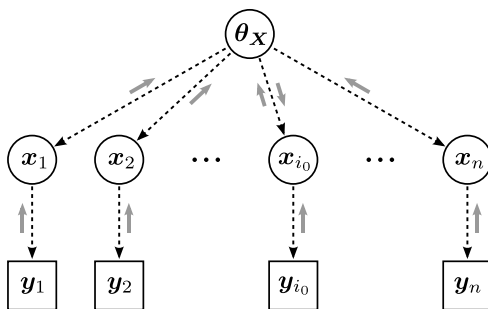

Figure 3: **Optimal combination of information.** A Bayesian network representation of probabilistic inversion is shown. Known (□) and unknown (○) quantities are related by probabilistic (⋯▸) relations. The "upstream" (◂─) and "downstream" (─▸) flow of information towards a specific $\boldsymbol{x}_{i_0}$ is indicated. This is a form of borrowing strength.

7. Bayesian Computations

Generally Bayesian posteriors feature an analytic closed-form expression only on a rare occasion. Specifically this applies to posteriors of the form Eqs. (13), (21) and (28). Notwithstanding the above, posteriors can be explored by means of Markov chain Monte Carlo (MCMC) [79, 80]. Principally this readily refers to posteriors stemming from multilevel inversion. The Metropolis-Hastings (MH) algorithm and the Gibbs sampler are prototypical MCMC techniques. In Section 7.1 we will review the MH algorithm and discuss classical MCMC key issues in Section 7.2. Additional computational key challenges posed by Bayesian multilevel model calibration will be discussed in Section 7.3. Some more sophisticated MCMC samplers that are suitable in a multilevel-context are surveyed in Section 7.4.

7.1. The Metropolis-Hastings Algorithm

MCMC is based on constructing a Markov chain such that its invariant distribution equals the posterior. Let $\pi(\boldsymbol{q})$ be the prior and $\pi(\boldsymbol{q}|\langle\boldsymbol{y}_i\rangle)$ the posterior density of some QoI $\boldsymbol{q}$. A Markov chain with equilibrium distribution $\pi(\boldsymbol{q}|\langle\boldsymbol{y}_i\rangle)$ is generated by initializing at $\boldsymbol{q}^{(0)}$ and repetitively proceeding as follows. Given a

state $\boldsymbol{q}^{(t)}$ that the Markov chain has taken on in some iteration, in the following iteration a candidate state $\boldsymbol{q}^{(\star)} \sim P(\boldsymbol{q}^{(\star)}|\boldsymbol{q}^{(t)})$ is randomly sampled from a proposal distribution $P(\boldsymbol{q}^{(\star)}|\boldsymbol{q}^{(t)})$. In the MH correction step the proposed state is approved as the new state $\boldsymbol{q}^{(t+1)} = \boldsymbol{q}^{(\star)}$ of the Markov chain with probability

$$\alpha\left(\boldsymbol{q}^{(\star)}, \boldsymbol{q}^{(t)}\right) = \min\left(1, \frac{\pi(\boldsymbol{q}^{(\star)}|\langle\boldsymbol{y}_i\rangle)\,P(\boldsymbol{q}^{(t)}|\boldsymbol{q}^{(\star)})}{\pi(\boldsymbol{q}^{(t)}|\langle\boldsymbol{y}_i\rangle)\,P(\boldsymbol{q}^{(\star)}|\boldsymbol{q}^{(t)})}\right). \tag{38}$$

Otherwise the proposal will be rejected, i.e. the Markov chain remains in its state $\boldsymbol{q}^{(t+1)} = \boldsymbol{q}^{(t)}$ of the preceding iteration. It is important to note that due to the MH acceptance probability Eq. (38), the algorithm calls for the computation of posterior ratios only. Thus for MCMC sampling the scale factors in Eqs. (13), (21) and (28) can be dropped and only unscaled posterior densities have to be evaluated.

Random walk Metropolis sampling rests upon local proposals, e.g. candidate states are sampled from a Gaussian distribution $\boldsymbol{q}^{(\star)} \sim \mathcal{N}(\boldsymbol{q}^{(\star)}; \boldsymbol{q}^{(t)}, \boldsymbol{\Sigma_q})$ that is centered around the current state $\boldsymbol{q}^{(t)}$. The covariance matrix $\boldsymbol{\Sigma_q}$ determines the "stepsizes" of the algorithm. Independence MH sampling is based on nonlocal proposals whose distribution $\boldsymbol{q}^{(\star)} \sim P(\boldsymbol{q}^{(\star)})$ is independent of $\boldsymbol{q}^{(t)}$, e.g. sampling candidate states from the prior $\boldsymbol{q}^{(\star)} \sim \pi(\boldsymbol{q}^{(\star)})$ or from some suitable approximation of the posterior $\boldsymbol{q}^{(\star)} \sim \hat{\pi}(\boldsymbol{q}^{(\star)}|\langle\boldsymbol{y}_i\rangle)$.

7.2. Classical Key Challenges

The performance of MCMC methods is governed by the mixing properties of the underlying Markov chain, i.e. the speed of convergence of the Markov chain towards the targeted posterior. As to which degree MCMC samples are autocorrelated has a determining influence on the convergence speed and their quality as posterior representatives. Hence MCMC algorithms are designed and tuned in pursuit of rapid mixing. Depending on the specific problem at hand, this may be a tricky business which requires to employ and combine sophisticated and highly specialized sampling schemes. Typically MCMC sampling calls for a high number of program iterations which in turn demands a high number of forward model runs for evaluating the likelihood function in the MH correction Eq. (38). Beyond that, careful convergence diagnostics are of particular importance for MCMC methods. One has to assess when the Markov chain has reached its stationary distribution, i.e. when it has lost any dependence on its initialization. Even though there are advanced convergence test [81, 82], e.g. Gelman-Rubin diagnostics for multiple over-dispersed chains [83, 84], we remark that from a pessimistic point of view any convergence diagnostic is heuristics [85]. Furthermore MCMC suffers from difficulties in exploring high-dimensional and multimodal posteriors.

7.3. Multilevel-Related Challenges

Multilevel posteriors can be readily sampled by means of classical MCMC techniques as they are commonly applied in "simple" Bayesian inversion. However, on top of the classical bottlenecks that were discussed above, one is faced with multilevel-specific MCMC challenges. The posteriors Eqs. (13) and (21), which are appertain to the joint and the marginal variant of multilevel calibration, are different in nature. Accordingly,

sampling these posteriors pose different computational burdens. The former requires a sampling scheme that performs efficiently in high-dimensional parameter spaces, whereas the latter suffers from computing the integrated likelihood Eq. (20). Similarly the posterior Eq. (28) of the "perfect" data model imposes forward uncertainty quantification for the computation of the likelihood Eq. (27).

Likelihood functions of the form Eqs. (23) and (30) suffer from another severe difficulty. It is well-known that statistical estimations of the likelihood ratio introduce an additional random component into the Markov chain transition kernel [86, 87]. Consequently the steady-state distribution of the chain may be modified. Therefore free parameters of the algorithm have to be chosen endeavoring high *posterior fidelity*, i.e. the degree as to which the induced long-run distribution conforms with the true posterior [59, 60].

7.4. Advanced MCMC Samplers

Summarized Bayesian multilevel model calibration requires an enormous number of forward model runs. Therefore in the statistical literature a wide range of advanced MCMC techniques, dedicated to posterior exploration in classical hierarchical models, have been devised. Some enhanced Gibbs sampling methods in this context are reviewed in [79] and references therein. However, in view of engineering problems they may not meet the challenges those applications usually pose. This is due to the inescapable "blackbox" character of the forward solver and nonconjugacy. Generally not all of the parameters will have full conditionals of a standard form that can be easily sampled. Despite that this paper does not focus on computational facets of uncertainty quantification, a short outlook on potentially efficient MCMC implementations is given.

Data augmentation is a powerful MCMC technique that aims at enhancing the numerical efficiency of posterior computation by introducing missing data as auxiliary variables [88, 89]. Note that the joint posterior Eq. (13) can be seen as an augmented form of the marginal one Eq. (21). Thus data augmentation naturally emerges in the context of Bayesian multilevel inversion. It has been beneficially applied for solving multilevel inverse problems within the domain of aerospace engineering [59, 60]. Vice versa, there are dedicated MCMC schemes for directly computing marginalized posteriors of the form Eq. (21), e.g. MC within Metropolis sampling [55, 86] or pseudo-marginalization [90]. The Hamiltonian Monte Carlo (HMC) algorithm is a sampler whose performance is remarkably efficient in high-dimensional parameter spaces and for highly correlated posteriors [91, 92]. Since multilevel models are higher-dimensional and correlated by definition, the HMC is a promising MCMC candidate in this context. Yet the HMC still occurs to be highly underacknowledged in Bayesian inference in general and for hierarchical models in particular.

8. Numerical Case Studies

In order to illustrate the power and versatility of the devised framework we conduct a selection of computer experiments. This shall be seen as a proof of concept and benchmark of the proposed methodology

in the context of engineering applications. A system of identically designed structural components functions as the basis for probing a range of experimental scenarios. Specifically we deal with an ensemble of simply supported beams that are tested in a series of three-point bending experiments. By multilevel analysis of measured beam deflections we highlight how different inferential goals, e.g. probabilistic inversion, residual calibration or optimal combination of information, can be achieved in the presence of material variability and uncertainties in the experimental setup. Keeping deterministic modeling simple and intuitive will allow us to focus on uncertainty quantification aspects that are the essential subject matter of this research. Incidentally we learn about the computational obstacles that must be overcome when aiming at "real-world" applications.

The forward problem will be shortly introduced in Section 8.1. Around this submodel, that covers the deterministic features of the system, Bayesian multilevel models will be built to capture uncertainty and variability. Probabilistic inversion, i.e. deducing the material variability throughout an ensemble of similar specimens, will be tackled in Section 8.2. The subsequent Section 8.3 will deal with residual model calibration. In Section 8.4 the impact of prescribed uncertainties in the test conditions will be investigated. In Section 8.5 borrowing strength will be utilized in order to ideally estimate the material characteristics of a single specimen by using information obtained from the other specimens.

8.1. Mechanical Model

The system under consideration is an ensemble of identically manufactured beams $i = 1, \ldots, n$ with well-known lengths L_i and rectangular cross sections with widths b_i and heights h_i. Yet the completed beams are only similar in the sense that we assume variability in the elastic moduli E_i across the ensemble, e.g. due to slight irregularities in the fabrication process. For each single beam i the Young's modulus E_i is assumed to be constant along the main beam axis. The deflections $\tilde{v}_i(s_{i,j})$ of a simply supported beam i under a concentrated point load F_i at midspan can be easily derived in Euler-Bernoulli beam theory. For positions $s_{i,j}$ along the beam axis with $0 \leq s_{i,j} \leq L_i/2$ and $j = 1, \ldots, n_i$ the deflections follow as

$$\tilde{v}_i(s_{i,j}) = \frac{F_i s_{i,j}}{48 E_i I_i} \left(3L_i^2 - 4s_{i,j}^2 \right), \text{ for } 0 \leq s_{i,j} \leq L_i/2, \tag{39}$$

where the moment of inertia is given as $I_i = b_i h_i^3/12$. Likewise a symmetric expression holds for positions $s_{i,j}$ along the main axis with $L_i/2 \leq s_{i,j} \leq L_i$. A single simply supported beam is visualized in Fig. 4.

Together with its symmetric counterpart, the algebraic formula Eq. (39) constitutes the deterministic submodel of the system under consideration. When a load F_i is applied to a beam i with physical dimensions $\boldsymbol{l}_i = (L_i, b_i, h_i)$ and an elastic modulus E_i, these relations predict the deflections $\tilde{\boldsymbol{v}}_i = (\tilde{v}_i(s_{i,1}), \ldots, \tilde{v}_i(s_{i,n_i}))$ at positions $\boldsymbol{s}_i = (s_{i,1}, \ldots, s_{i,n_i})$. We denote this as

$$\tilde{\boldsymbol{v}}_i = \mathcal{M}(E_i, F_i, \boldsymbol{l}_i, \boldsymbol{s}_i). \tag{40}$$
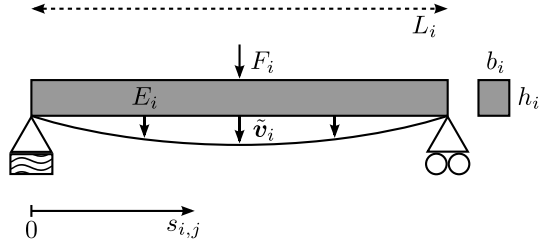
Figure 4: **A simply supported beam.**

When beam deflections are measured in three-point bending tests for each member $i = 1, \ldots, n$ in the population, multilevel inversion allows for optimal data analysis in experimental situations where the inputs of Eq. (40) are subject to uncertainty.

8.2. Probabilistic Inversion

We begin with Bayesian probabilistic inversion, on the basis of which we demonstrate how one can quantify the material variability within the ensemble of beams in a series of bending tests. A numerical experiment is therefore set up as follows. We consider a number of $n = 100$ beams with well-known dimensions $L_i = 1\,\mathrm{m}$ and $b_i = h_i = 10\,\mathrm{cm}$. Beams are subjected to concentrated loads $F_i = 30\,\mathrm{kN}$ that are applied at midspan. For $i = 1, \ldots, 100$ Young's moduli E_i are independently sampled from a lognormal distribution $\mathcal{LN}(E_i | \mu_E, \sigma_E)$ with mean $\mu_E = 15\,\mathrm{GPa}$ and standard deviation $\sigma_E = 3\,\mathrm{GPa}$. This corresponds to a coefficient of variation $c_E = 20\,\%$. After having set up the experiment, the hyperparameters $\boldsymbol{\theta}_E = (\mu_E, \sigma_E)$ as well as beam-specific moduli E_i will be treated as unknowns. At $n_i = 3$ positions $\boldsymbol{s}_i = (s_{i,1}, s_{i,2}, s_{i,3})$ with $s_{i,1} = 25\,\mathrm{cm}$, $s_{i,2} = 50\,\mathrm{cm}$ and $s_{i,3} = 75\,\mathrm{cm}$ beam deflections $\tilde{\boldsymbol{v}}_i = (\tilde{v}_i(s_{i,1}), \tilde{v}_i(s_{i,2}), \tilde{v}_i(s_{i,3}))$ are computed according to Eq. (39). In order to take measurement uncertainty and forward model imperfection into account, we perturb the predictions $\tilde{\boldsymbol{v}}_i$ with noise terms $\boldsymbol{\varepsilon}_i = (\varepsilon_{i,1}, \varepsilon_{i,2}, \varepsilon_{i,3})$. Those terms are independently sampled from Gaussian distributions $\mathcal{N}(\boldsymbol{\varepsilon}_i; \boldsymbol{0}, \boldsymbol{\Sigma}_i)$ with $\boldsymbol{\Sigma}_i = \sigma_i^2 \boldsymbol{I}_3$ and $\sigma_i = 0.1\,\mathrm{mm}$. Eventually $\boldsymbol{v}_i = \tilde{\boldsymbol{v}}_i + \boldsymbol{\varepsilon}_i$ represent the pseudo data that will become analyzed with respect to the QoI $\boldsymbol{\theta}_E = (\mu_E, \sigma_E)$.

In many circumstances expert knowledge about the QoI $\boldsymbol{\theta}_E$ is available prior to analyzing the data. This knowledge can be accounted for by eliciting a suitable prior distribution $\pi(\boldsymbol{\theta}_E)$. Herein we employ a proper Bayesian prior $\pi(\boldsymbol{\theta}_E) = \pi(\mu_E)\,\pi(\sigma_E)$ with independent marginals. As measured in units of GPa those marginals are given as uniform distributions $\pi(\mu_E) = \mathcal{U}(0, 100)$ and $\pi(\sigma_E) = \mathcal{U}(0, 30)$. This is supposed to represent an experimental situation where one cannot elicit informative priors, nonetheless one is confident enough to assign this weakly informative and flat prior with its upper and lower bounds.

Ultimately probabilistic inversion can be summarized as the estimation of the QoI $\boldsymbol{\theta}_X \equiv \boldsymbol{\theta}_E$ with the deflection measurements $\langle \boldsymbol{y}_i \rangle \equiv \langle \boldsymbol{v}_i \rangle$. Beam-specific Young's moduli $\langle \boldsymbol{x}_i \rangle \equiv \langle E_i \rangle$, that are not of immediate inferential interest, are considered nuisance to that end. Experimental conditions $\langle \boldsymbol{d}_i \rangle \equiv \langle (F_i, \boldsymbol{l}_i, \boldsymbol{s}_i) \rangle$, that the experiments where subject to, and prediction error models $\langle \boldsymbol{\Sigma}_i \rangle$ are assumed to be known. The distributions

22

$f_{\boldsymbol{X}|\boldsymbol{\Theta_X}}(\boldsymbol{x}_i|\boldsymbol{\theta_X}) \equiv \mathcal{LN}(E_i|\mu_E, \sigma_E)$ and $\pi_{\boldsymbol{\Theta_X}}(\boldsymbol{\theta_X}) \equiv \pi(\boldsymbol{\theta}_E)$ represent the available structural and parametric prior knowledge, respectively. The emerging posterior will be of the form $\pi(\boldsymbol{\theta_X}|\langle\boldsymbol{y}_i\rangle) \equiv \pi(\boldsymbol{\theta}_E|\langle\boldsymbol{v}_i\rangle)$. It can be directly sampled or accessed via the QoI-marginals of the joint posterior $\pi(\langle\boldsymbol{x}_i\rangle, \boldsymbol{\theta_X}|\langle\boldsymbol{y}_i\rangle) \equiv \pi(\langle E_i\rangle, \boldsymbol{\theta}_E|\langle\boldsymbol{v}_i\rangle)$. A DAG corresponding to probabilistic inversion is provided in Fig. 2b.

8.2.1. MCMC

Generally we employ a joint rather than a marginal problem formulation. For the fidelity reasons that were discussed in Section 7.3 this allows exact posterior computation where an approximation is only introduced in as much as MCMC sampling is concerned. Moreover a joint posterior features a richer structure which will provide new insights into multilevel inversion. All computations will be serially done on a contemporary Intel Xeon CPU.

The joint posterior $\pi(\langle E_i\rangle, \boldsymbol{\theta}_E|\langle\boldsymbol{v}_i\rangle)$ is sampled by means of a blockwise random walk Metropolis algorithm. A practical problem of random walk samplers in high dimension is to carefully tune the proposal distribution. For complex multivariate posterior distributions this is a cumbersome procedure that poses severe difficulties. However, in multilevel inversion one can advantageously exploit the "symmetry" of the problem in the latent variables. Assuming that separate inverse problems i with $1 \leq i \leq n$ are not severely ill-posed, latent variables of the same uncertainty type are expected to behave similarly in the sense that their marginal posteriors resemble one another. Moreover, due to the indirectness of borrowing strength, their mutual correlations are expected to be rather small. Along these lines the "effective dimensionality" is lower than the number of unknowns suggests. This discussion motivates that MCMC updates are done in blocks $\langle E_i\rangle$ and (μ_E, σ_E). We find that with Gaussian jumping distributions the algorithm can be easily tuned in such a way that blockwise acceptance rates range between 20 % and 40 %. Avoiding lengthy convergence times in high-dimensional problems requires smart initialization, too. Again we proceed by exploiting the structure of the multilevel system. The block $\langle E_i\rangle$ is initialized with solutions of separate inverse problems, while two-stage estimates are used in the hyperparameter block (μ_E, σ_E).

In order to assure duly completed posterior exploration we perform a number of convergence checks. The algorithm is initialized in regions of the parameter space that had not been visited before and the convergence behavior of the Markov chain is monitored. We detect that the chain eventually reaches the same posterior modes again. In Fig. 5 trace plots of a converging Markov chain are shown for its μ_E and σ_E components. They have been initialized at $\mu_E^{(0)} = 50\,\text{GPa}$ and $\sigma_E^{(0)} = 15\,\text{GPa}$, i.e. in the middle of their priorly admissible intervals. While the mean hyperparameter μ_E directly converges as shown in Fig. 5a, we observe a different behavior for the spread hyperparameter σ_E. From Fig. 5b it can be seen that the latter chain tends to higher values prior to attraction towards the posterior mean. For the given initialization this is a systematic effect that indicates a posterior correlation in the hyperparameters (μ_E, σ_E). Eventually the Markov chain converges within ca. 400 MCMC iterations. Apart from such visual inspections we generally

rely on Gelman-Rubin diagnostics for parallel chains [83, 84].

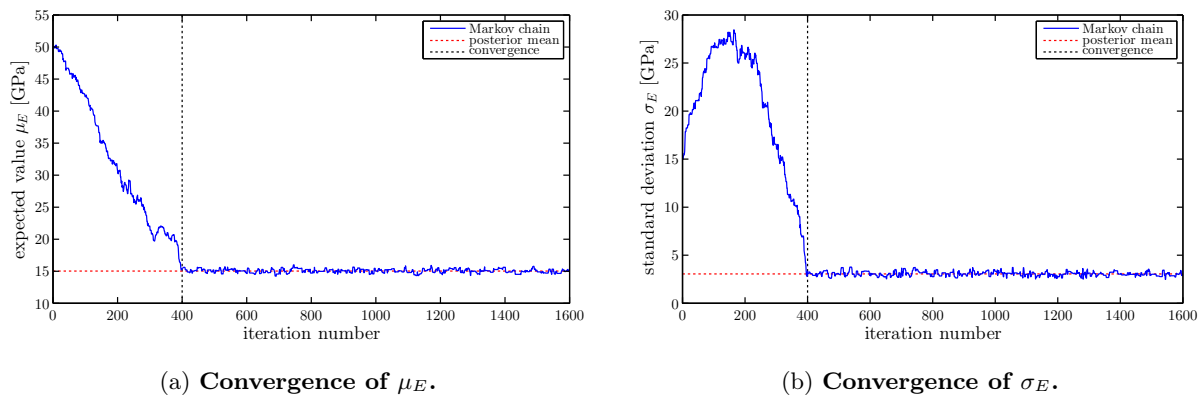(a) **Convergence of μ_E.**

(b) **Convergence of σ_E.**

Figure 5: **Trace plots of a converging Markov chain.** For $n = 100$ the converging Markov chain is shown for μ_E in (a) and for σ_E in (b). Being initialized at $\mu_E^{(0)} = 50\,\mathrm{GPa}$ and $\sigma_E^{(0)} = 15\,\mathrm{GPa}$ the Markov chain converges within ca. 400 MCMC iterations. In equilibrium the Markov chain samples the posterior around its mean.

In Fig. 6 the MCMC sample autocorrelations are plotted for the QoI (μ_E, σ_E) and for an intermediate variable E_i with $i = 1$. It can be seen how the autocorrelation function (ACF) drops until it becomes indistinguishable from zero. This behavior governs the quality of the sample as a posterior representative. Especially the ACF of E_i shown in Fig. 6c motivates more efficient updating schemes in future research.

(a) **Autocorrelation of μ_E.**

(b) **Autocorrelation of σ_E.**
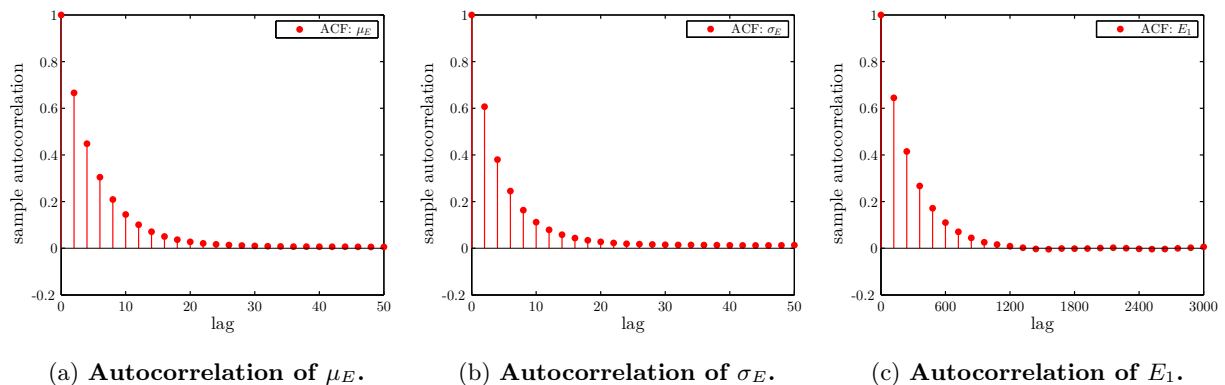
(c) **Autocorrelation of E_1.**

Figure 6: **Sample autocorrelation functions.** For a run with $n = 100$ the MCMC sample autocorrelation function is plotted for μ_E in (a), for σ_E in (b) and for E_1 in (c). The sample autocorrelation determines the effective MCMC sample size.

8.2.2. Results: Posterior Marginals

We analyze the data $\langle v_i \rangle_{1 \le i \le 100}$ as well as its subconfigurations $\langle v_i \rangle_{1 \le i \le 10}$, $\langle v_i \rangle_{1 \le i \le 20}$ and $\langle v_i \rangle_{1 \le i \le 50}$. This allows to assess how the number of experiments n influences the identification of the QoI. For each of the runs $N = 10^7$ MCMC iterations are performed. As a general rule we discard the initial 1% of the total number of iterations of each Markov chain as a burn-in period. The total algorithm runtime adds up

to $t = 3.85\,\mathrm{h}$ for $n = 10$ and to $t = 4.66\,\mathrm{h}$ for $n = 100$. The resulting posterior marginals of μ_E and σ_E are shown in Fig. 7. A statistical summary of these marginals can be found in Table 1, where the mean, mode, standard deviation (SD) and coefficient of variation (CV) are listed. With increasing number of processed experiments n, Bayesian point estimates (mean, mode) approach the true values $\mu_E = 15\,\mathrm{GPa}$ and $\sigma_E = 3\,\mathrm{GPa}$ while measures of estimation uncertainty (SD, CV) expectedly decrease.

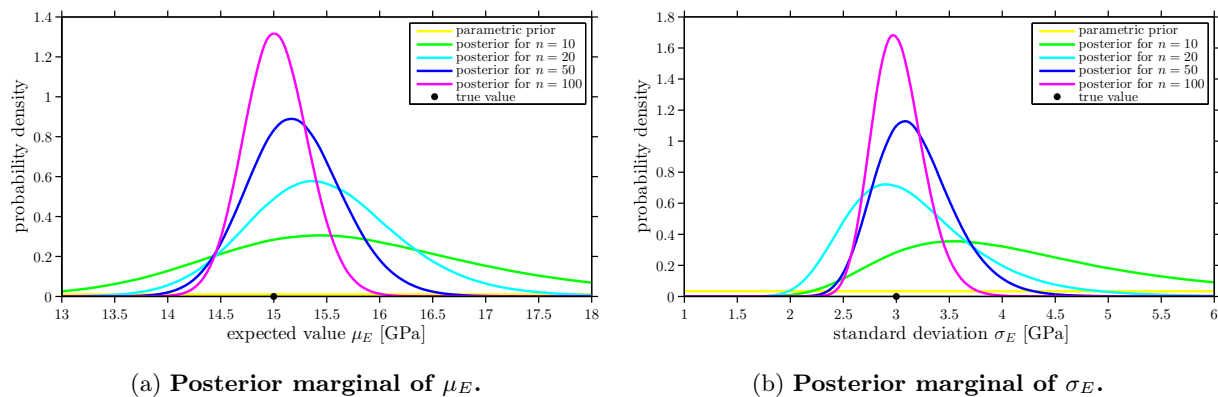(a) **Posterior marginal of μ_E.** (b) **Posterior marginal of σ_E.**

Figure 7: **Posterior marginals of the QoI.** Corresponding to various numbers of experiments n, the marginal posterior densities of μ_E and σ_E are shown in (a) and (b), respectively. For increasing n, the posterior uncertainty in estimating the QoI $\boldsymbol{\theta}_E = (\mu_E, \sigma_E)$ with $\mu_E = 15\,\mathrm{GPa}$ and $\sigma_E = 3\,\mathrm{GPa}$ steadily decreases.

Table 1: **Summary of the QoI posterior marginals.**

	μ_E [GPa]			[ ]	σ_E [GPa]			[ ]
	mean	mode	SD	CV	mean	mode	SD	CV
$n = 10$	15.98	15.43	2.06	0.13	4.73	3.54	3.55	0.75
$n = 20$	15.48	15.36	0.74	0.05	3.18	2.90	0.65	0.20
$n = 50$	15.20	15.17	0.46	0.03	3.17	3.08	0.37	0.12
$n = 100$	15.02	15.00	0.30	0.02	3.02	2.97	0.24	0.08

8.2.3. Results: Two-Dimensional Posteriors

Showing posterior marginals may hide possibly existing dependency structures or the lack thereof. Those constitute a substantial result of Bayesian data analysis, though. Hence Fig. 8 shows two-dimensional posteriors where interesting correlation properties were discovered. The two-dimensional posterior of (μ_E, σ_E) is plotted in Fig. 8a. According to the posterior probability model these two parameters are correlated with a linear Pearson coefficient of correlation $r_{\mu_E, \sigma_E} = 0.40$. Note that these parameters were assumed to be independent in accord with their prior model. The joint posterior Eq. (32) can also feature a correlation between

hyperparameters and experiment-specific parameters. In Figs. 8b and 8c the two-dimensional posteriors of (μ_E, E_i) and (E_j, E_i) with $i = 50$ and $j = 75$ are imaged.
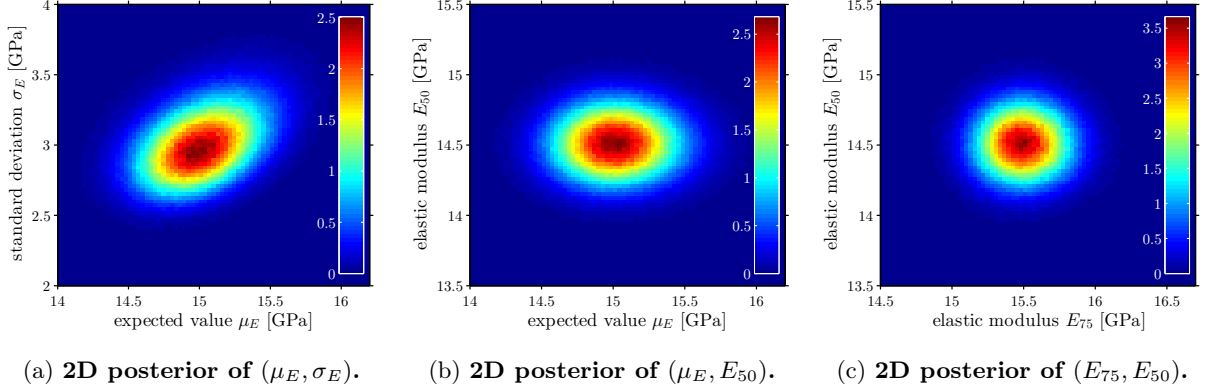

(a) **2D posterior of (μ_E, σ_E).**   (b) **2D posterior of (μ_E, E_{50}).**   (c) **2D posterior of (E_{75}, E_{50}).**

Figure 8: **2D posteriors of (μ_E, σ_E) and (μ_E, E_{50}).** The two-dimensional posteriors of (μ_E, σ_E), (μ_E, E_{50}) and (E_{75}, E_{50}) are shown. Being priorly independent the components μ_E and σ_E are seen to be correlated a posteriori. The linear Pearson coefficient of correlation amounts to $r_{\mu_E, \sigma_E} = 0.40$.

8.3. Residual Calibration

There are situations where the strong assumption of known residual variances $\boldsymbol{\Sigma}_i = \sigma_i^2 \boldsymbol{I}_3$ is somewhat restrictive. Thus we generalize multilevel inversion as in Section 8.2 by treating $\sigma_{\mathcal{E}} \equiv \sigma_i$ as a global unknown. In units of mm the corresponding parametric prior is set to a uniform distribution $\pi(\sigma_{\mathcal{E}}) = \mathcal{U}(0, 0.5)$. Otherwise the experimental setup of probabilistic inversion is used.

The standard deviation $\sigma_{\mathcal{E}}$ of the residual model $\mathcal{N}(\boldsymbol{\varepsilon}_i | \boldsymbol{0}, \sigma_{\mathcal{E}}^2 \boldsymbol{I}_3)$ is introduced as an extra unknown in the model Eq. (31) and in the posterior Eq. (32). Consequently the joint prior is given as $\pi(\langle E_i \rangle, \mu_E, \sigma_E, \sigma_{\mathcal{E}}) = \pi(\sigma_{\mathcal{E}}) \, \pi(\mu_E) \, \pi(\sigma_E) \prod_{i=1}^{n} \mathcal{LN}(E_i | \mu_E, \sigma_E)$. For the joint likelihood function one has $\mathcal{L}(\langle E_i \rangle, \sigma_{\mathcal{E}}; \langle \boldsymbol{v}_i \rangle) = \prod_{i=1}^{n} \mathcal{N}(\boldsymbol{v}_i | \mathcal{M}(E_i, F_i, \boldsymbol{l}_i, \boldsymbol{s}_i), \sigma_{\mathcal{E}}^2 \boldsymbol{I}_3)$. Brought together this leads to a joint posterior density that has the shape $\pi(\langle E_i \rangle, \mu_E, \sigma_E, \sigma_{\mathcal{E}} | \langle \boldsymbol{v}_i \rangle) \propto \mathcal{L}(\langle E_i \rangle, \sigma_{\mathcal{E}}; \langle \boldsymbol{v}_i \rangle) \, \pi(\langle E_i \rangle, \mu_E, \sigma_E, \sigma_{\mathcal{E}})$.

We sample from this posterior by appending a block for the additional unknown $\sigma_{\mathcal{E}}$ in the MCMC updating scheme. In order to assess the influence of the amount of data on the final results, independent runs are performed for $n = 10, 20, 50$ and 100. In Fig. 9 the relevant posterior marginals for the inference of the residual model $\sigma_{\mathcal{E}}$ are shown. A short summary of the these marginals is provided in Table 2. The higher the number of analyzed experiments n, the better the true value $\sigma_{\mathcal{E}} = 0.1 \, \text{mm}$ has been revealed. This proves that one can indeed estimate the parameters of the prediction error model in the context of multilevel calibration. If this is not of interest for its own sake, it still avoids the requirement of perfect knowledge of the error variance. In addition we observed that introducing an uncertainty in the residual model hardly affects the inference of the QoI in probabilistic inversion.
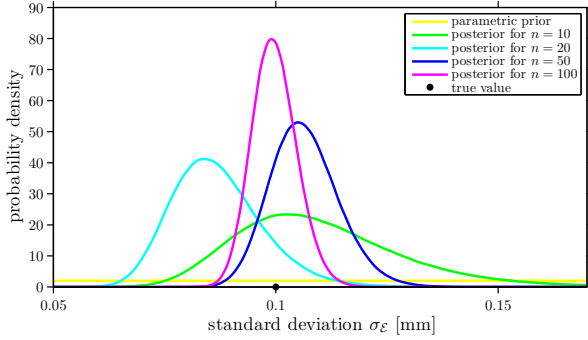
Figure 9: **Posterior marginals of $\sigma_{\mathcal{E}}$.**

Table 2: **Summary of the $\sigma_{\mathcal{E}}$-marginals.**

	$\sigma_{\mathcal{E}}$ [10^{-5} m]			[ ]
	mean	mode	SD	CV
$n = 10$	11.00	10.23	1.90	0.17
$n = 20$	8.68	8.38	1.01	0.12
$n = 50$	10.65	10.50	0.77	0.07
$n = 100$	9.97	9.90	0.50	0.05

8.4. Uncertain Conditions

In the following we describe an experimental situation where the inference of the QoI $\boldsymbol{\theta}_E$ is hampered by additional uncertainties in the experimental conditions. Experimental conditions are formally treated as nuisance parameters with prescribed uncertainties. More specifically, we do not assume that the loads F_i are perfectly known anymore. In contrast, we assume that they are $\boldsymbol{\zeta}_i$-type variables, i.e. they are uncertain yet they follow a known distribution. This represents a well-known situation where the loads F_i that the testing machine actually applies can only be imprecisely adjusted. In fact, while a targeted load in each experiment is chosen, the physically realized load F_i may be uncertain. This is accounted for by a prescribed distribution $\mathcal{N}(F_i; \mu_{F_i}, \sigma_{F_i}^2)$ where μ_{F_i} is the targeted load and σ_{F_i} represents the degree of uncertainty that is inherent to the test machinery.

The setup for conducting a numerical experiment is similar to the one specified in Section 8.2. For $n = 50$ beams we set the beam dimensions $\boldsymbol{l}_i$ and measurement positions $\boldsymbol{s}_i$ as before. Elastic moduli E_i are randomly drawn from $\mathcal{LN}(E_i | \mu_E, \sigma_E)$ as previously detailed. In contrast to plain probabilistic inversion, for $i = 1, \ldots, n$ experiment-specific loads F_i are independently sampled from normal distributions $\mathcal{N}(F_i; \mu_{F_i}, \sigma_{F_i}^2)$ with $\mu_{F_i} = 30\,\text{kN}$ and $\sigma_{F_i} = 3\,\text{kN}$. This equates to a coefficient of variation $c_{F_i} = 10\,\%$. Note that such a high degree of uncertainty is unlikely to be encountered in a real-case experiment. It is used here to accentuate the results presented below, though. The realized loads F_i will be treated as unknowns whereas the hyperparameters $\boldsymbol{\theta}_{F_i} = (\mu_{F_i}, \sigma_{F_i})$, i.e. the targeted load and its uncertainty, will be treated as knowns. In accordance with Eq. (39) synthetic measurements $\boldsymbol{v}_i = \tilde{\boldsymbol{v}}_i + \boldsymbol{\varepsilon}_i$ are generated again. The prior distribution $\pi(\boldsymbol{\theta}_E) = \pi(\mu_E)\,\pi(\sigma_E)$ is also chosen as previously stated.

The problem of probabilistic inversion under additional prescribed nuisance reads as follows. The hyperparameters $\boldsymbol{\theta}_X \equiv \boldsymbol{\theta}_E$ are the QoI whereas experiment-specific unknowns $\langle \boldsymbol{x}_i \rangle \equiv \langle E_i \rangle$ and $\langle \boldsymbol{\zeta}_i \rangle \equiv \langle F_i \rangle$ are considered nuisance. With measurements $\langle \boldsymbol{y}_i \rangle \equiv \langle \boldsymbol{v}_i \rangle$ the QoI can be inferred. Experimental-specific knowns consist of the hyperparameters $\langle \boldsymbol{\theta}_{Z_i} \rangle \equiv \langle \boldsymbol{\theta}_{F_i} \rangle$, the experimental conditions $\langle \boldsymbol{d}_i \rangle \equiv \langle (\boldsymbol{l}_i, \boldsymbol{s}_i) \rangle$ and the residual covariances $\langle \boldsymbol{\Sigma}_i \rangle$. Parametric Bayesian prior knowledge is given by $\pi_{\boldsymbol{\Theta}_X}(\boldsymbol{\theta}_X) \equiv \pi(\boldsymbol{\theta}_E)$ whereas

27

$f_{\boldsymbol{X}|\boldsymbol{\Theta_X}}(\boldsymbol{x}_i|\boldsymbol{\theta_X}) \equiv \mathcal{LN}(E_i|\mu_E, \sigma_E)$ and $f_{\boldsymbol{Z}}(\boldsymbol{\zeta}_i; \boldsymbol{\theta}_{\boldsymbol{Z}_i}) \equiv \mathcal{N}(F_i; \mu_{F_i}, \sigma_{F_i}^2)$ are structural prior distributions. Within a joint approach a posterior of the form $\pi(\langle \boldsymbol{x}_i \rangle, \langle \boldsymbol{\zeta}_i \rangle, \boldsymbol{\theta_X} | \langle \boldsymbol{y}_i \rangle) \equiv \pi(\langle E_i \rangle, \langle F_i \rangle, \boldsymbol{\theta}_E | \langle \boldsymbol{v}_i \rangle)$ arises. Eventually one is interested in the QoI-marginals $\pi(\boldsymbol{\theta_X} | \langle \boldsymbol{y}_i \rangle) \equiv \pi(\boldsymbol{\theta}_E | \langle \boldsymbol{v}_i \rangle)$ only. A DAG corresponding to this experimental situation is shown in Fig. 2c.

8.4.1. Results: Hyperparameters

We sample the joint posterior $\pi(\langle E_i \rangle, \langle F_i \rangle, \boldsymbol{\theta}_E | \langle \boldsymbol{v}_i \rangle)$ where nuisance variables $\langle F_i \rangle$ are explicitly accounted for. In a blockwise manner MCMC sweeps are accomplished for (μ_E, σ_E), $\langle E_i \rangle$ and $\langle F_i \rangle$ which constitute different blocks. Blockwise proposal distributions are again adjusted in order to obtain acceptance rates in between 20 % and 40 %. Each F_i in the block $\langle F_i \rangle$ is initialized at $F_i^{(0)} = \mu_{F_i}$, i.e. the structural prior mean. Other than that initialization, convergence checks and burn-in are accomplished as before. For $N = 10^7$ MCMC iterations the total computation time amounts to $t = 7.18\,\text{h}$. The resulting posterior marginals of μ_E and σ_E can be seen in Fig. 10. A statistical summary is provided in Table 3 where the mean, mode, SD and CV of the marginals are itemized.

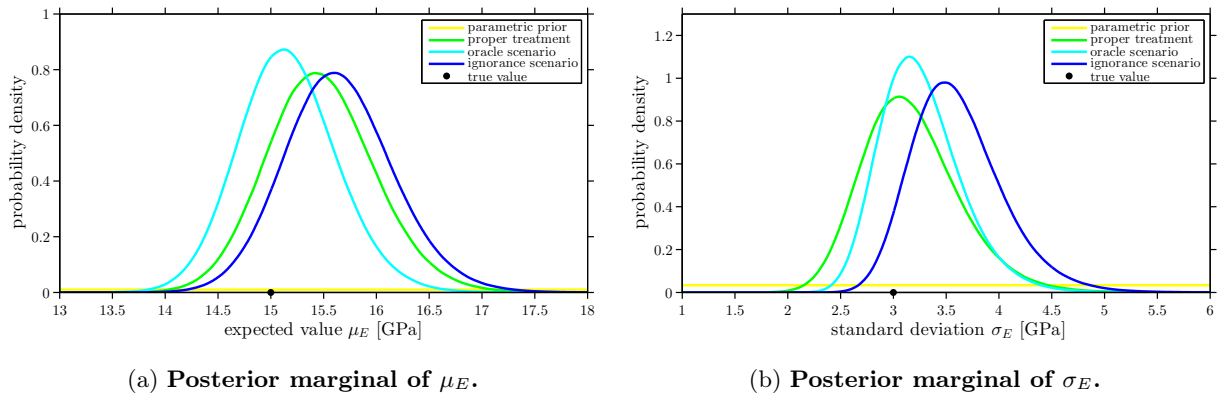(a) **Posterior marginal of μ_E.** (b) **Posterior marginal of σ_E.**

Figure 10: **Posterior marginals of the QoI.** The marginal posteriors of μ_E and σ_E are provided in (a) and (b), respectively. Three experimental scenarios are investigated: the proper treatment of the additional uncertainty, an idealized situation where one would precisely know the loads, and the case of a parsimonious model where the uncertainty remains unrecognized.

Table 3: **Summary of the QoI posterior marginals.**

	μ_E [GPa]			[ ]	σ_E [GPa]			[ ]
	mean	mode	SD	CV	mean	mode	SD	CV
proper treatment	15.47	15.41	0.51	0.03	3.17	3.05	0.46	0.14
oracle scenario	15.16	15.13	0.47	0.03	3.26	3.15	0.39	0.12
ignorance scenario	15.65	15.60	0.52	0.03	3.61	3.51	0.43	0.12

We try to assess the impact of the uncertainty that had been introduced in the loads F_i on the estimation of the QoI $\boldsymbol{\theta}_E = (\mu_E, \sigma_E)$. To that end we pursue the following two strategies. First of all we estimate the QoI while treating the realized loads F_i as if they were part of the experiment-specific knowns $\boldsymbol{d}_i$. This "what-if" or "oracle" scenario actually describes the hypothetical situation that we met in plain probabilistic inversion. It does not describe the realistic scenario of uncertain conditions $\boldsymbol{\zeta}_i$ that we are actually investigating. Yet this way of proceeding sheds light on how the prescribed uncertainty in the loads affects the inference of the QoI. For $N = 10^7$ and $t = 4.33\,\mathrm{h}$ the results to probabilistic inversion are added to Fig. 10. With respect to this idealized situation, one can reassess the previous results of properly treating the loads as uncertain. The introduction of the uncertainty in the loads had actually shifted the posterior modes and raised the level of estimation uncertainty accordingly.

Second of all we investigate the case that the uncertainty $\mathcal{N}(F_i; \mu_{F_i}, \sigma_{F_i}^2)$ in the applied loads F_i is simply disregarded. Either it has not been recognized by mistake or it has been intentionally dropped by making simplifying assumptions in favor of a parsimonious model. Rather than treating the loads as belonging to the unknowns $\boldsymbol{\zeta}_i$, we erroneously treat them as such experimental conditions $\boldsymbol{d}_i^{\widetilde{\approx}}$ that only approximately describe the prevailing conditions $\boldsymbol{d}_i$. While the data has been created under $\boldsymbol{d}_i$, data analysis is carried out under $\boldsymbol{d}_i^{\widetilde{\approx}}$. This describes a situation where the experimenter targets a load $F_i^{\widetilde{\approx}} = \mu_{F_i}$, but the testing machine actually realizes F_i. If this uncertainty $\mathcal{N}(F_i; \mu_{F_i}, \sigma_{F_i}^2)$ is not accounted for or not recognized at all, the analyst will accomplish inference under the spurious assumption that the loads had taken on their targeted values $F_i^{\widetilde{\approx}}$ during experiment execution. For $N = 10^7$ and $t = 3.75\,\mathrm{h}$ the resulting posteriors are added to Fig. 10. Our interpretation is that dropping the uncertainty of F_i corrupts the estimation of the QoI and results in misleading estimates of posterior uncertainty, whereas the proper treatment of all uncertainties yields results that are closer to the idealized "oracle" scenario.

8.4.2. Results: Intermediate Variables

Sampling the joint posterior $\pi(\langle E_i \rangle, \langle F_i \rangle, \boldsymbol{\theta}_E | \langle \boldsymbol{v}_i \rangle)$ of the entirety of unknowns provides further interesting insights. Apart from the QoI-marginals one can examine the posterior model of experiment-specific loads F_i, notwithstanding that they are considered nuisance. Fig. 11 contains two different posteriors involving some F_i. In Fig. 11a the posterior marginal of a pinpoint load F_i is shown for $i = 23$. The identification of specifically applied loads F_i is subject to rather high levels of posterior uncertainty. This is an issue of statistical identifiability. When both E_i and F_i are uncertain and various combinations of these can explain the observation $\boldsymbol{v}_i$ equally well, then those combinations (E_i, F_i) cannot be distinguished a posteriori. Of course, the reason is that only the ratio F_i/E_i in Eq. (39) can be identified. It is therefore interesting to investigate the posterior correlation between the load F_i and the modulus E_i of an experiment i. The two-dimensional posterior of (E_i, F_i) for $i = 20$ that is shown in Fig. 11b serves as an example. Posterior mass is assigned to suitable parameter constellations (E_i, F_i) that well-explain the measurement $\boldsymbol{v}_i$. As expected

the posterior is strongly correlated with a linear coefficient of correlation $r_{F_{20},E_{20}} = 0.99$.

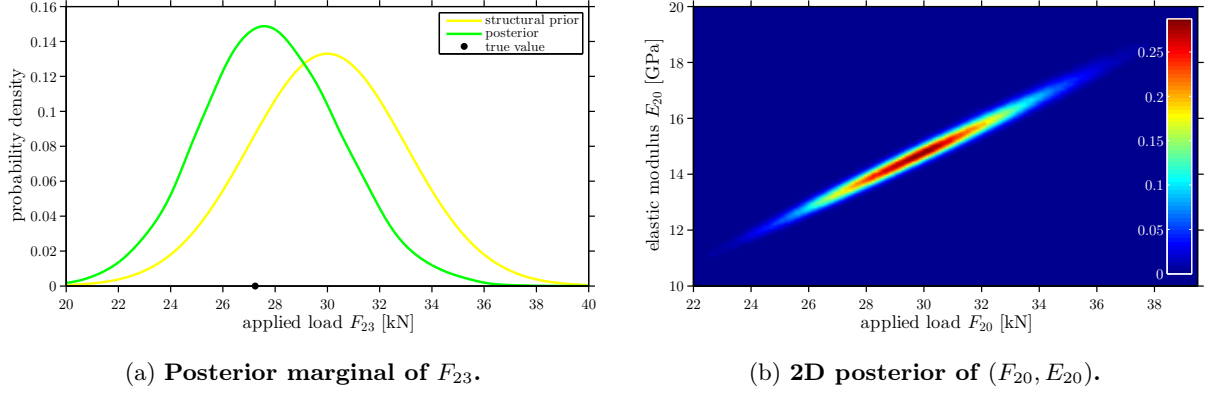(a) **Posterior marginal of F_{23}.** (b) **2D posterior of (F_{20}, E_{20}).**

Figure 11: **Posteriors of intermediate variables.** In (a) the posterior marginal of F_{23} and its structural prior $\mathcal{N}(F_{23}; \mu_{F_{23}}, \sigma_{F_{23}}^2)$ with $\mu_{F_{23}} = 30\,\text{kN}$ and $\sigma_{F_{23}} = 3\,\text{kN}$ are shown. The posterior is centered around the actual value $F_{23} = 27.24\,\text{kN}$. The two-dimensional posterior of (F_{20}, E_{20}) with $r_{F_{20},E_{20}} = 0.99$ is shown in (b).

8.5. Borrowing Strength

As pointed out in Section 6, Bayesian multilevel modeling allows for "optimal combination of information" or "borrowing strength". Here we demonstrate this inferential mechanism and investigate its underlying flow of information for the previous application example. The Bayesian model of probabilistic inversion Eq. (31) is considered. However, as opposed to probabilistic inversion we declare experiment-specific elastic moduli $\langle E_i \rangle$ as the QoI whereas the hyperparameters $\boldsymbol{\theta}_E$ are considered nuisance. Herein we highlight the optimal inference of a single E_{i_0} for some $i_0 \in \{1, \ldots, n\}$.

The experimental setup is similar to the one described in Section 8.2. For $n = 50$ beams, elastic moduli E_i are randomly sampled from $\mathcal{LN}(E_i | \mu_E, \sigma_E)$. Beam dimensions $\boldsymbol{l}_i$, measurement positions $\boldsymbol{s}_i$ and the applied loads F_i are chosen as before. With Eq. (39) beam deflections $\tilde{\boldsymbol{v}}_i$ are predicted. Synthetic data $\boldsymbol{v}_i = \tilde{\boldsymbol{v}}_i + \boldsymbol{\varepsilon}_i$ are generated by perturbing the predictions $\tilde{\boldsymbol{v}}_i$ with noise. For this purpose noise terms $\boldsymbol{\varepsilon}_i$ are independently sampled from Gaussian distributions $\mathcal{N}(\boldsymbol{\varepsilon}_i; \boldsymbol{0}, \boldsymbol{\Sigma}_i)$. We choose $\boldsymbol{\Sigma}_i = \sigma_i^2 \boldsymbol{I}_3$ with $\sigma_i = 0.1\,\text{mm}$ for $i \neq i_0$ and $\sigma_{i_0} = 0.1\,\text{cm}$. The latter describes a comparably large deviation that differs from the setup of Section 8.2. This choice serves the purpose of clearly illustrating the inferential mechanism of optimal combination of information.

Eventually optimal combination of information reads as the following problem. With noisy data $\langle \boldsymbol{y}_i \rangle \equiv \langle \boldsymbol{v}_i \rangle$ an experiment-specific $\boldsymbol{x}_{i_0} \equiv E_{i_0}$ has to be ideally estimated, i.e. taking all available sources of information into account. The hyperparameters $\boldsymbol{\theta_X} \equiv \boldsymbol{\theta}_E$ as well as $\langle \boldsymbol{x}_{\neq i_0} \rangle \equiv \langle E_{\neq i_0} \rangle$ are considered nuisance to that end. Experiment-specific knowns are $\langle \boldsymbol{d}_i \rangle \equiv \langle (F_i, \boldsymbol{l}_i, \boldsymbol{s}_i) \rangle$ and $\langle \boldsymbol{\Sigma}_i \rangle$. The resultant posterior will be of the form $\pi(\boldsymbol{x}_{i_0} | \langle \boldsymbol{y}_i \rangle) \equiv \pi(E_{i_0} | \langle \boldsymbol{v}_i \rangle)$. Subsequent to formulating the joint posterior $\pi(\langle \boldsymbol{x}_i \rangle, \boldsymbol{\theta_X} | \langle \boldsymbol{y}_i \rangle) \equiv$

$\pi(\langle E_i \rangle, \boldsymbol{\theta}_E | \langle \boldsymbol{v}_i \rangle)$, the QoI-marginals can be easily extracted. Other than that, the experimental setup of probabilistic inversion is adopted. Thus the experiment can be visualized by the DAG in Fig. 2b, too.

8.5.1. Results: Information Accumulation

We conduct simple updating, sequential filtering and multilevel inversion for estimating E_{i_0}, as introduced in Section 6. First of all we start with the simple Bayesian updating approach that was introduced in Section 6.1. By the method of composition we draw $K = 10^5$ samples $(E_{i_0}^{(1)}, \ldots, E_{i_0}^{(K)})$ from the mixture prior $\pi(E_{i_0})$ that corresponds to Eq. (35). With this sample the mixture prior can be evaluated as the corresponding one-dimensional KDE with Gaussian kernel functions. The posterior $\pi(E_{i_0} | \boldsymbol{v}_{i_0})$ results from conditioning on the piece of data $\boldsymbol{v}_{i_0}$. This univariate posterior is explored in $N = 10^5$ MCMC iterations for which the program execution time amounts to $t = 5.86\,\mathrm{h}$. The final result of this simple updating approach is shown in Fig. 12a.

Second of all we conduct the sequential Bayesian filtering program that was proposed in Section 6.2. In $N = 10^7$ MCMC iterations that take $t = 3.95\,\mathrm{h}$, probabilistic inversion for estimating $\boldsymbol{\theta}_E$ is executed with the data $\langle \boldsymbol{v}_{\neq i_0} \rangle$. MCMC samples from the resultant posterior $\pi(\boldsymbol{\theta}_E | \langle \boldsymbol{v}_{\neq i_0} \rangle)$ are used to sample the compound distribution $\pi(E_{i_0} | \langle \boldsymbol{v}_{\neq i_0} \rangle)$ in Eq. (36) via the composition method. Subsequently a lognormal fit to these samples acts as the prior for E_{i_0}. This prior and the arising posterior distribution $\pi(E_{i_0} | \langle \boldsymbol{v}_{\neq i_0} \rangle, \boldsymbol{v}_{i_0})$ are plotted in Fig. 12b. In $t = 0.01\,\mathrm{h}$ of execution time $N = 10^5$ MCMC samples of the univariate posterior were sampled. By comparison of the two posteriors in Fig. 12, the shrinkage of the posterior uncertainty from $\pi(E_{i_0} | \boldsymbol{v}_{i_0})$ to $\pi(E_{i_0} | \langle \boldsymbol{v}_{\neq i_0} \rangle, \boldsymbol{v}_{i_0})$ becomes apparent. Both posteriors follow from conditioning on the data $\boldsymbol{v}_{i_0}$, they update different priors $\pi(E_{i_0})$ and $\pi(E_{i_0} | \langle \boldsymbol{v}_{\neq i_0} \rangle)$, though. In the first place this proves that Bayesian priors are a valid source of information. Moreover, this principally shows how learning about E_{i_0} can be indirectly supported by the evidence that $\langle \boldsymbol{v}_{\neq i_0} \rangle$ contains with regard to $\boldsymbol{\theta}_E$.

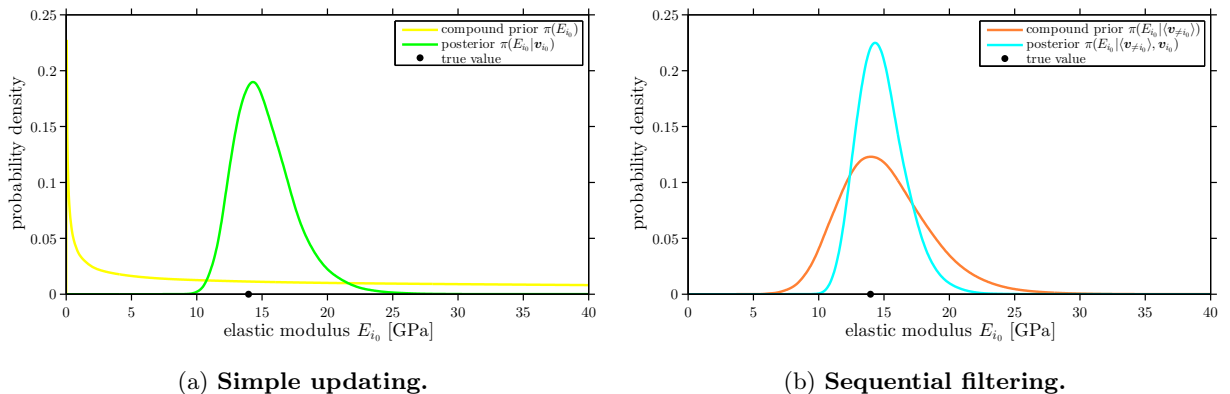(a) **Simple updating.**   (b) **Sequential filtering.**

Figure 12: **Bayesian updating and filtering.** The mixture prior $\pi(E_{i_0})$ and the posterior $\pi(E_{i_0} | \boldsymbol{v}_{i_0})$ of simple updating are shown in (a). Sequential filtering is based on the more informative mixture prior $\pi(E_{i_0} | \langle \boldsymbol{v}_{\neq i_0} \rangle)$ and the corresponding posterior $\pi(E_{i_0} | \langle \boldsymbol{v}_{\neq i_0} \rangle, \boldsymbol{v}_{i_0})$ that are given in (b).

Lastly we perform Bayesian multilevel analysis as described in Section 6.3. Sampling the joint posterior $\pi(\langle E_i \rangle, \boldsymbol{\theta}_E | \langle \boldsymbol{v}_i \rangle)$ allows to straightforwardly extract samples from its marginal $\pi(E_{i_0} | \langle \boldsymbol{v}_i \rangle)$ in Eq. (37). This is accomplished in $t = 4.57\,\mathrm{h}$ for $N = 10^7$ algorithm iterations. The posterior and the previous inferential distributions relevant for E_{i_0} are plotted in Fig. 13. In addition to that Table 4 recapitulates the different approaches. Results are also provided from a second series of runs that were independently carried out on top of the first one. The motivation is to show that borrowing strength is a not a random but a systematic effect. The accumulation of information concerning E_{i_0} manifests in the progressively decreasing uncertainty in the distributions. At every stage of the estimation plan, a certain proportion of the available information has entered the analysis and has been translated into a gain of knowledge related to E_{i_0}. Only the multilevel posterior $\pi(E_{i_0} | \langle \boldsymbol{v}_i \rangle)$ entirely aggregates the available information.

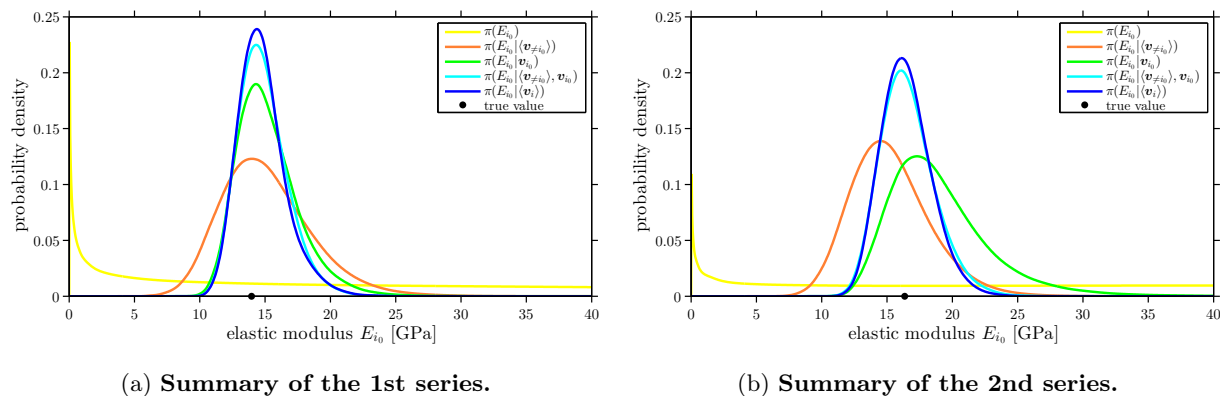(a) **Summary of the 1st series.**   (b) **Summary of the 2nd series.**

Figure 13: **Accumulation of information.** In (a) and (b) the estimations of E_{i_0} are summarized for two series of runs. The true values are $E_{i_0} = 13.96\,\mathrm{GPa}$ and $E_{i_0} = 16.35\,\mathrm{GPa}$ in the 1st and 2nd series, respectively. Uncertainties in identifying these values reflect the amount of information processed in simple updating, sequential filtering and multilevel inversion.

Table 4: **Posterior summaries of estimating E_{i_0}.**

	1st series: E_{i_0} [GPa]		[ ]		2nd series: E_{i_0} [GPa]		[ ]
	mean	mode	SD	CV	mean	mode	SD	CV
simple updating:	15.23	14.31	2.38	0.16	19.02	17.30	3.93	0.21
sequential filtering:	14.82	14.32	1.83	0.12	16.58	16.07	2.03	0.12
multilevel inversion:	14.75	14.37	1.79	0.12	16.47	16.12	1.85	0.11

The assumption of well-known loads F_i may be overly optimistic in experimental practice. As done in Section 8.4 one could attach an additional prescribed uncertainty to those model inputs. In doing so we expect similar results accompanied by a weakening of borrowing strength. Furthermore we expect an indirect form of borrowing strength also to occur for the inputs of a prescribed uncertainty type. Actually the

prescribed uncertainty model does not permit for learning about a specific F_{i_0} by borrowing strength directly from $\langle \boldsymbol{v}_{\neq i_0} \rangle$. However, by optimally estimating E_{i_0} also learning F_{i_0} would be indirectly strengthened.

9. Conclusion & Outlook

Bayesian multilevel model calibration has been developed as a consistent and comprehensive framework for managing uncertainties in inverse problems. At the core of the such problems a forward model relates physical parameters to observable quantities. This deterministic model has been surrounded by a probabilistic representation of uncertainty, variability and error. For this purpose classical Bayesian inversion, hierarchical statistical models and the predominant epistemic/aleatory conception of uncertainty have been utilized. The inferential rationale of multilevel inversion, based on the conditioning, marginalization and transformation of probability measures, has become transparent by laying the research focus on aspects of uncertainty quantification and information accumulation. Fully Bayesian probabilistic inversion and borrowing strength have been suggested. Furthermore we have originally elaborated on the "perfect" data limit. Our developments were driven by the challenges of engineering applications and they ultimately allow for optimal data analysis in intricate situations where evidence is scarce and uncertainty prevails.

An ensemble of structural elements of the same type, for all of which virtual tests are performed and pseudo data are gathered, served as the basis for investigating a variety of experimental scenarios. The amenities of Bayesian multilevel inversion were demonstrated by exercising inference in the chosen example applications under realistic uncertainty configurations. Probabilistic inversion, i.e. the identification of material variability throughout a population of specimens, was accomplished and it was investigated how the amount of data influences the estimation uncertainty . The constraints of perfectly known residual variances and experimental conditions were loosened. In this context we calibrated the forward model prediction error and we studied how the objective of probabilistic inversion is impeded by additional uncertainties in the experimental conditions. Optimal combination of information, i.e. the ideal inference of specimen-specific properties, has been introduced as a byproduct of the joint formulation of multilevel inversion. Especially in the engineering community this is an aspect that is often overlooked. We examined the underlying inferential mechanisms and we identified the computational obstacles, e.g. costly evaluations of the marginalized likelihood function or the curse of high-dimensionality.

In conclusion, innovative techniques must be developed in order to overcome these difficulties for solving "real-world" problems. Future research therefore includes the following items. For the marginal problem, numerically efficient and acceptably accurate approximations of the integrated likelihood have to be developed. Advanced MCMC techniques, that are custom-tailored for the specific structure of multilevel posteriors, have to be devised for the joint problem. In this connection a numerical study involving HMC is in progress. For both the marginal and the joint variant of multilevel inversion, the application of dedicated metamodeling

techniques promises drastic speedups. It will also be interesting to study the applicability and performance of optimal transportation approaches [93, 94] to classical Bayesian inference in the context of multilevel estimation. Another research question concerns the role of multimodality and severe ill-posedness of separate inverse problems in Bayesian multilevel inversion.

References

[1] A. M. Stuart. Inverse problems: A Bayesian perspective. *Acta Numer.*, 19:451–559, 2010.

[2] M. Allmaras, W. Bangerth, J. Linhart, J. Polanco, F. Wang, K. Wang, J. Webster, and S. Zedler. Estimating parameters in physical models through Bayesian inversion: A complete example. *SIAM Rev.*, 55(1):149–167, 2013.

[3] A. Tarantola. *Inverse Problem Theory and Methods for Model Parameter Estimation.* Society for Industrial and Applied Mathematics (SIAM), Philadelphia, Pennsylvania, USA, 2004.

[4] J. Kaipio and E. Somersalo. *Statistical and Computational Inverse Problems*, volume 160 of *Applied Mathematical Sciences.* Springer, New York, 2004.

[5] R. Hadidi and N. Gucunski. Probabilistic approach to the solution of inverse problems in civil engineering. *J. Comput. Civil Eng.*, 22(6):338–347, 2008.

[6] J. L. Beck. Bayesian system identification based on probability logic. *Struct. Control. Health Monit.*, 17(7):825–847, 2010.

[7] A. Malinverno and V. A. Briggs. Expanded uncertainty quantification in inverse problems: Hierarchical Bayes and empirical Bayes. *Geophysics*, 69(4):1005–1016, 2004.

[8] J. Wang and N. Zabaras. Hierarchical Bayesian models for inverse problems in heat conduction. *Inverse Prob.*, 21(1):183–206, 2005.

[9] L. Wu. *Mixed Effects Models for Complex Data*, volume 113 of *Monographs on Statistics & Applied Probability.* Chapman & Hall/CRC, Boca Raton, Florida, USA, 2010.

[10] S. W. Raudenbush. Educational applications of hierarchical linear models: A review. *J. Educ. Stat.*, 13(2):85–116, 1988.

[11] M. H. Seltzer, W. H. Wong, and A. S. Bryk. Bayesian analysis in applications of hierarchical models: Issues and methods. *J. Educ. Behav. Stat.*, 21(2):131–167, 1996.

[12] J. Wakefield. The Bayesian analysis of population pharmacokinetic models. *J. Am. Stat. Assoc.*, 91(433):62–75, 1996.

[13] H. T. Banks and L. K. Potter. Probabilistic methods for addressing uncertainty and variability in biological models: application to a toxicokinetic model. *Math. Biosci.*, 192(2):193–225, 2004.

[14] M. Davidian and D. M. Giltinan. Nonlinear models for repeated measurement data: An overview and update. *J. Agric. Biol. Environ. Stat.*, 8(4):387–419, 2003.

[15] H. T. Banks, Z. R. Kenz, and W. C. Thompson. A review of selected techniques in inverse problem nonparametric probability distribution estimation. *J. Inverse Ill-posed Prob.*, 20(4):429–460, 2012.

[16] M. Davidian and D. M. Giltinan. *Nonlinear Models for Repeated Measurement Data*, volume 62 of *Monographs on Statistics & Applied Probability.* Chapman & Hall/CRC, Boca Raton, Florida, USA, 1995.

[17] H. T. Banks, S. Hu, and W. C. Thompson. *Modeling and Inverse Problems in the Presence of Uncertainty.* Monographs and Research Notes in Mathematics. Chapman & Hall/CRC, Boca Raton, Florida, USA, 2014.

[18] A. Gelman and J. Hill. *Data Analysis Using Regression and Multilevel/Hierarchical Models.* Cambridge University Press, Cambridge, UK, 2006.

[19] P. D. Congdon. *Applied Bayesian Hierarchical Methods.* Chapman & Hall/CRC, Boca Raton, Florida, USA, 2010.

[20] S. Jackman. *Bayesian Analysis for the Social Sciences.* Wiley Series in Probability and Statistics. John Wiley & Sons, Ltd., Chichester, West Sussex, UK, 2009.

[21] A. Gelman, J. B. Carlin, H. S. Stern, D. B. Dunson, A. Vehtari, and D. B. Rubin. *Bayesian Data Analysis*. Texts in Statistical Science. CRC Press, Boca Raton, Florida, USA, 3rd edition, 2014.

[22] M. H. Faber. On the treatment of uncertainties and probabilities in engineering decision analysis. *J. Offshore Mech. Arct. Eng.*, 127(3):243–248, 2005.

[23] A. Der Kiureghian and O. Ditlevsen. Aleatory or epistemic? Does it matter? *Struct. Saf.*, 31(2):105–112, 2009.

[24] J. C. Helton and W. L. Oberkampf. Alternative representations of epistemic uncertainty. *Reliab. Eng. Syst. Saf.*, 85(1–3):1–10, 2004.

[25] J. C. Helton and J. D. Johnson. Quantification of margins and uncertainties: Alternative representations of epistemic uncertainty. *Reliab. Eng. Syst. Saf.*, 96(9):1034–1052, 2011.

[26] T. Koski and J. M. Noble. *Bayesian Networks: An Introduction*. Wiley Series in Probability and Statistics. John Wiley & Sons, Ltd., Chichester, West Sussex, UK, 2009.

[27] U. B. Kjærulff and A. L. Madsen. *Bayesian Networks and Influence Diagrams: A Guide to Construction and Analysis*, volume 22 of *Information Science and Statistics*. Springer, New York, 2nd edition, 2013.

[28] Y. Y. Bayraktarli, J. W. Baker, and M. H. Faber. Uncertainty treatment in earthquake modelling using Bayesian probabilistic networks. *Georisk*, 5(1):44–58, 2011.

[29] M. Deublein, M. Schubert, B. T. Adey, J. Köhler, and M. H. Faber. Prediction of road accidents: A Bayesian hierarchical approach. *Accid. Anal. Prev.*, 51:274–291, 2013.

[30] D. L. Kelly and C. L. Smith. Bayesian inference in probabilistic risk assessment – The current state of the art. *Reliab. Eng. Syst. Saf.*, 94(2):628–643, 2009.

[31] A. Urbina, S. Mahadevan, and T. L. Paez. A Bayes network approach to uncertainty quantification in hierarchically developed computational models. *Int. J. Uncertainty Quantification*, 2(2):173–193, 2012.

[32] E. de Rocquigny and S. Cambier. Inverse probabilistic modelling of the sources of uncertainty: A non-parametric simulated-likelihood method with application to an industrial turbine vibration assessment. *Inverse Prob. Sci. Eng.*, 17(7):937–959, 2009.

[33] G. Celeux, A. Grimaud, Y. Lefèbvre, and E. de Rocquigny. Identifying intrinsic variability in multivariate systems through linearized inverse methods. *Inverse Prob. Sci. Eng.*, 18(3):401–415, 2010.

[34] P. Barbillon, G. Celeux, A. Grimaud, Y. Lefèbvre, and E. de Rocquigny. Nonlinear methods for inverse statistical problems. *Comput. Stat. Data Anal.*, 55(1):132–142, 2011.

[35] E. de Rocquigny. *Modelling Under Risk and Uncertainty: An Introduction to Statistical, Phenomenological and Computational Methods*. Wiley Series in Probability and Statistics. John Wiley & Sons, Ltd., Chichester, West Sussex, UK, 2012.

[36] J. O. Berger and W. E. Strawderman. Choice of hierarchical priors: admissibility in estimation of normal means. *Ann. Stat.*, 24(3):931–951, 1996.

[37] J. O. Berger, W. Strawderman, and D. Tang. Posterior propriety and admissibility of hyperpriors in normal hierarchical models. *Ann. Stat.*, 33(2):606–646, 2005.

[38] A. Gelman. Prior distributions for variance parameters in hierarchical models (Comment on article by Browne and Draper). *Bayesian Anal.*, 1(3):515–534, 2006.

[39] D. Draper, J. S. Hodges, C. L. Mallows, and D. Pregibon. Exchangeability and data analysis. *J. R. Stat. Soc. Ser. A*, 156(1):9–37, 1993.

[40] J. M. Bernardo. The concept of exchangeability and its applications. *Far East J. Math. Sci.*, 4:111–121, 1996.

[41] E. Simoen, C. Papadimitriou, and G. Lombaert. On prediction error correlation in Bayesian model updating. *J. Sound Vib.*, 332(18):4136–4152, 2013.

[42] E. L. Zhang, P. Feissel, and J. Antoni. A comprehensive Bayesian approach for model updating and quantification of

modeling errors. *Probab. Eng. Mech.*, 26(4):550–560, 2011.

[43] M. C. Kennedy and A. O'Hagan. Bayesian calibration of computer models. *J. R. Stat. Soc. Ser. B*, 63(3):425–464, 2001.

[44] P. D. Arendt, D. W. Apley, and W. Chen. Quantification of model uncertainty: Calibration, model discrepancy, and identifiability. *J. Mech. Des.*, 134(10):100908:1–12, 2012.

[45] E. L. Droguett and A. Mosleh. Bayesian methodology for model uncertainty using model performance data. *Risk Anal.*, 28(5):1457–1476, 2008.

[46] I. Park and R. V. Grandhi. A Bayesian statistical method for quantifying model form uncertainty and two model combination methods. *Reliab. Eng. Syst. Saf.*, 129:46–56, 2014.

[47] J. Beck and K. Yuen. Model selection using response measurements: Bayesian probabilistic approach. *J. Eng. Mech.*, 130(2):192–203, 2004.

[48] K.-V. Yuen. Recent developments of Bayesian model class selection and applications in civil engineering. *Struct. Saf.*, 32(5):338–346, 2010.

[49] D. Draper. Assessment and propagation of model uncertainty. *J. R. Stat. Soc. Ser. B*, 57(1):45–97, 1995.

[50] D. Basu. On the elimination of nuisance parameters. *J. Am. Stat. Assoc.*, 72(358):355–366, 1977.

[51] A. P. Dawid. A Bayesian look at nuisance parameters. *Trabajos de Estadística*, 31(1):167–203, 1980.

[52] J. O. Berger, B. Liseo, and R. L. Wolpert. Integrated likelihood methods for eliminating nuisance parameters. *Stat. Sci.*, 14(1):1–28, 1999.

[53] T. A. Severini. On the relationship between Bayesian and non-Bayesian elimination of nuisance parameters. *Stat. Sinica*, 9(3):713–724, 1999.

[54] T. A. Severini. Integrated likelihood functions for non-Bayesian inference. *Biometrika*, 94(3):529–542, 2007.

[55] M. A. Beaumont. Estimation of population growth or decline in genetically monitored populations. *Genetics*, 164(3):1139–1160, 2003.

[56] Y. J. Sung and C. J. Geyer. Monte Carlo likelihood inference for missing data models. *Ann. Stat.*, 35(3):990–1011, 2007.

[57] C. S. Bos. A comparison of marginal likelihood computation methods. In W. Härdle and B. Rönz, editors, *Compstat: Proceedings in Computational Statistics*, pages 111–116. Physica-Verlag Heidelberg New York, 2002.

[58] L. G. Crespo, S. P. Kenny, and D. P. Giesy. The NASA Langley multidisciplinary uncertainty quantification challenge. In *Proc. 16th AIAA Non-Deterministic Approaches Conference (SciTech 2014)*, National Harbor, Maryland, USA, January 13-17 2014.

[59] J. B. Nagel and B. Sudret. A Bayesian multilevel framework for uncertainty characterization and the NASA Langley multidisciplinary UQ challenge. In *Proc. 16th AIAA Non-Deterministic Approaches Conference (SciTech 2014)*. American Institute of Aeronautics and Astronautics (AIAA), Reston, Virginia, USA, 2014.

[60] J. B. Nagel and B. Sudret. Bayesian multilevel model calibration for inverse problems under uncertainty with perfect data. *J. Aerosp. Inf. Syst.*, 12(1):97–113, 2015.

[61] B. Sudret, F. Perrin, and M. Pendola. Use of polynomial chaos expansions in stochastic inverse problems. In *Proc. 4th Int. ASRANet Colloquium*. ASRANet Ltd, Glasgow, Scotland, UK, 2008.

[62] Y. M. Marzouk, H. N. Najm, and L. A. Rahn. Stochastic spectral methods for efficient Bayesian solution of inverse problems. *J. Comput. Phys.*, 224(2):560–586, 2007.

[63] Y. M. Marzouk and H. N. Najm. Dimensionality reduction and polynomial chaos acceleration of Bayesian inference in inverse problems. *J. Comput. Phys.*, 228(6):1862–1902, 2009.

[64] P. M. Tagade and H.-L. Choi. A generalized polynomial chaos-based method for efficient Bayesian calibration of uncertain computational models. *Inverse Prob. Sci. Eng.*, 22(4):602–624, 2014.

[65] J. Beck and S. Au. Bayesian updating of structural models and reliability using Markov chain Monte Carlo simulation. *J. Eng. Mech.*, 128(4):380–391, 2002.

[66] J. Ching and Y. Chen. Transitional Markov chain Monte Carlo method for Bayesian model updating, model class selection, and model averaging. *J. Eng. Mech.*, 133(7):816–832, 2007.

[67] C. Papadimitriou and D.-C. Papadioti. Component mode synthesis techniques for finite element model updating. *Comput. Struct.*, 126:15–28, 2013.

[68] H. A. Jensen, E. Millas, D. Kusanovic, and C. Papadimitriou. Model-reduction techniques for Bayesian finite element model updating using dynamic response data. *Comput. Meth. Appl. Mech. Eng.*, 279:301–324, 2014.

[69] J. B. Nagel and B. Sudret. Probabilistic inversion for estimating the variability of material properties: A Bayesian multilevel approach. In D. Novák and M. Vořechovský, editors, *Proc. 11th Int. Probabilistic Workshop (IPW11)*, pages 293–303. Litera, Brno, Czech Republic, 2013.

[70] G. C. Ballesteros, P. Angelikopoulos, C. Papadimitriou, and P. Koumoutsakos. Bayesian hierarchical models for uncertainty quantification in structural dynamics. In M. Beer, S.-K. Au, and J. W. Hall, editors, *Vulnerability, Uncertainty, and Risk: Quantification, Mitigation, and Management*, chapter 162, pages 1615–1624. American Society of Civil Engineers (ASCE), Reston, Virginia, USA, 2014.

[71] B. Kraan and T. Bedford. Probabilistic inversion of expert judgments in the quantification of model uncertainty. *Manage. Sci.*, 51(6):995–1006, 2005.

[72] C. Du, D. Kurowicka, and R. M. Cooke. Techniques for generic probabilistic inversion. *Comput. Stat. Data Anal.*, 50(5):1164–1187, 2006.

[73] C. Desceliers, R. Ghanem, and C. Soize. Maximum likelihood estimation of stochastic chaos representations from experimental data. *Int. J. Numer. Methods Eng.*, 66(6):978–1001, 2006.

[74] C. Desceliers, C. Soize, and R. Ghanem. Identification of chaos representations of elastic properties of random media using experimental vibration tests. *Comput. Mech.*, 39(6):831–838, 2007.

[75] L. Mehrez, A. Doostan, D. Moens, and D. Vandepitte. Stochastic identification of composite material properties from limited experimental databases, part II: Uncertainty modelling. *Mech. Syst. Sig. Process.*, 27:484–498, 2012.

[76] S. Debruyne, D. Vandepitte, and D. Moens. Identification of design parameter variability of honeycomb sandwich beams from a study of limited available experimental dynamic structural response data. *Comput. Struct.*, 146:197–213, 2015.

[77] J. B. Nagel and B. Sudret. A Bayesian multilevel approach to optimally estimate material properties. In M. Beer, S.-K. Au, and J. W. Hall, editors, *Vulnerability, Uncertainty, and Risk: Quantification, Mitigation, and Management*, chapter 151, pages 1504–1513. American Society of Civil Engineers (ASCE), Reston, Virginia, USA, 2014.

[78] D. Draper, D. P. Gaver, P. K. Goel, J. B. Greenhouse, L. V. Hedges, C. N. Morris, and C. M. Waternaux. *Combining Information: Statistical Issues and Opportunities for Research.* Panel on Statistical Issues and Opportunities for Research in the Combination of Information, Committee on Applied and Theoretical Statistics, Board on Mathematical Sciences, Commission on Physical Sciences, Mathematics, and Applications, National Research Council, 1992.

[79] W. R. Gilks, S. Richardson, and D. J. Spiegelhalter, editors. *Markov Chain Monte Carlo in Practice.* Interdisciplinary Statistics. Chapman & Hall/CRC, Boca Raton, Florida, USA, 1996.

[80] C. P. Robert and G. Casella. *Monte Carlo Statistical Methods.* Springer Series in Statistics. Springer, New York, 2nd edition, 2004.

[81] M. K. Cowles and B. P. Carlin. Markov chain Monte Carlo convergence diagnostics: A comparative review. *J. Am. Stat. Assoc.*, 91(434):883–904, 1996.

[82] S. P. Brooks and G. O. Roberts. Convergence assessment techniques for Markov chain Monte Carlo. *Stat. Comput.*, 8(4):319–335, 1998.

[83] A. Gelman and D. B. Rubin. Inference from iterative simulation using multiple sequences. *Stat. Sci.*, 7(4):457–472, 1992.

[84] S. P. Brooks and A. Gelman. General methods for monitoring convergence of iterative simulations. *J. Comp. Graph. Stat.*, 7(4):434–455, 1998.

[85] C. J. Geyer. Introduction to Markov chain Monte Carlo. In S. Brooks, A. Gelman, G. L. Jones, and X.-L. Meng, editors, *Handbook of Markov Chain Monte Carlo*, Handbooks of Modern Statistical Methods, pages 3–48. Chapman & Hall/CRC, Boca Raton, Florida, USA, 2011.

[86] P. D. O'Neill, D. J. Balding, N. G. Becker, M. Eerola, and D. Mollison. Analyses of infectious disease data from household outbreaks by Markov chain Monte Carlo methods. *J. R. Stat. Soc. Ser. C*, 49(4):517–542, 2000.

[87] G. Bal, I. Langmore, and Y. Marzouk. Bayesian inverse problems with Monte Carlo forward models. *Inverse Prob. Imaging*, 7(1):81–105, 2013.

[88] D. A. van Dyk and X.-L. Meng. The art of data augmentation. *J. Comp. Graph. Stat.*, 10(1):1–50, 2001.

[89] D. A. van Dyk. Hierarchical models, data augmentation, and Markov chain Monte Carlo. In *Statistical Challenges in Astronomy*, pages 41–55. Springer, New York, 2003.

[90] C. Andrieu and G. O. Roberts. The pseudo-marginal approach for efficient Monte Carlo computations. *Ann. Stat.*, 37(2):697–725, 2009.

[91] S. Duane, A. D. Kennedy, B. J. Pendleton, and D. Roweth. Hybrid Monte Carlo. *Phys. Lett. B*, 195(2):216–222, 1987.

[92] R. M. Neal. MCMC using Hamiltonian dynamics. In S. Brooks, A. Gelman, G. L. Jones, and X.-L. Meng, editors, *Handbook of Markov Chain Monte Carlo*, Handbooks of Modern Statistical Methods, pages 113–162. Chapman & Hall/CRC, Boca Raton, Florida, USA, 2011.

[93] S. Reich. A dynamical systems framework for intermittent data assimilation. *BIT Numer. Math.*, 51(1):235–249, 2011.

[94] T. A. El Moselhy and Y. M. Marzouk. Bayesian inference with optimal maps. *J. Comput. Phys.*, 231(23):7815–7850, 2012.