JOURNAL OF AEROSPACE INFORMATION SYSTEMS
Vol. 12, No. 1, January 2015

Bayesian Multilevel Model Calibration for Inverse Problems Under Uncertainty with Perfect Data

Joseph B. Nagel* and Bruno Sudret†
Swiss Federal Institute of Technology, CH-8093 Zürich, Switzerland

DOI: 10.2514/1.I010264

A probabilistic framework for Bayesian inference and uncertainty analysis is developed. It allows inverse problems to be addressed in experimental situations where data are scarce and uncertainty is ubiquitous. The uncertainty characterization subproblem of the NASA Langley Multidisciplinary Uncertainty Quantification Challenge serves as the motivating application example. From the responses of a computational model, the goal is to learn about unknown model inputs that are subject to multiple types of uncertainty. This objective is interpreted and solved as Bayesian multilevel model calibration. The zero-noise or "perfect" data limit is investigated. Thereby, the likelihood function is defined as a solution to forward uncertainty propagation. Posterior explorations are based on suitable Markov chain Monte Carlo algorithms and stochastic likelihood simulations. An unforeseen finding in this context is that the posterior distribution can only be sampled with a certain degree of fidelity. Partial data augmentation is introduced as a means to improve the error statistics of likelihood estimations and the fidelity of posterior computations.

I. Introduction

THE NASA Langley Multidisciplinary Uncertainty Quantification Challenge has raised contemporary open questions to uncertainty quantification (UQ) [1,2]. Altogether, it consists of five subproblems that deal with uncertainty characterization, sensitivity analysis, uncertainty propagation, extreme-case analysis, and robust design. These problems originate from a specific aerospace application that is part of greater efforts to reduce the rate of fatal loss-of-control accidents [3–5]. An abstract and widely discipline-independent problem formulation prompts researchers and practitioners from various fields in academia and industry to devise generic solutions to the problems. A dynamically scaled free-flight model of a remotely piloted twin-turbine-powered transport aircraft is the physical system under consideration. It serves as a prototyping and experimentation testbed for flight control in adverse situations, e.g., under structural damage or component failure. Parameter uncertainties of this subscaled model reflect the uncertainties in aerodynamic conditions and losses in control effectiveness.

In this contribution, we address the uncertainty characterization subproblem of the challenge posed. With given responses of a computational model, the challenge is to learn about the unknown inputs that parametrize the flying conditions. Throughout the experiments, data are collected while model inputs are subject to epistemic uncertainty and aleatory variability [6,7]. Inference therefore focuses on physically fixed model parameters, as well as on so-called hyperparameters. The latter determine the distribution of such model inputs that are variable during experimentation. We approach the problem from a Bayesian perspective to statistical inversion and uncertainty quantification [8–10]. Although classical Bayesian inversion allows the estimation of the constant model parameters, the additional identification of hyperparameters requires hierarchical modeling approaches. Hierarchical models were mainly developed in biological statistics [11,12], and they are only slowly being adopted within the engineering community [13–15].

The goal of this paper is the development of a framework, along with computational tools for attacking inverse problems under aleatory and epistemic parameter uncertainties. We combine classical inversion and hierarchical modeling into a Bayesian multilevel framework that allows us to tackle the general class of problems that the uncertainty characterization subproblem typifies. Within a probabilistic setting, this eventually allows for an elegant formulation and efficient numerical solution. The foundations of inverse modeling in conjunction with perfect data (i.e., in the zero-noise limit) and parameter uncertainty are laid. Randomness in the data is then solely attributed to a probability model of the input arguments of a computational "black-box" solver. The likelihood is formulated as a solution to uncertainty propagation. Since this renders its evaluation analytically intractable, statistical estimators based on the Monte Carlo method and kernel density estimation are proposed. In this context, the induced type of posterior approximation is investigated. Heuristic ways of tuning free algorithmic parameters (e.g., the kernel bandwidth) are presented. Partial data augmentation is proposed in order to improve likelihood estimations through automatic kernel bandwidth selection and to enhance the fidelity of the posterior.

The paper is organized as follows. In Sec. II, a generic Bayesian multilevel framework for inversion under uncertainty will be initially formulated. For the solution of the NASA Langley Multidisciplinary UQ Challenge, we will devise a statistical model involving perfect data in Sec. III. Computational key challenges posed by Bayesian inference in the present context will be discussed in Sec. IV. The challenge problem will be cast as multilevel inversion in Sec. V and, in the subsequent Sec. VI, our results will be presented. In Sec. VII, data augmentation will be used in order to ensure a sufficient degree of algorithmic efficiency and posterior fidelity. We will conclude in Sec. VIII where the gathered experience from solving the NASA UQ Challenge Problem is summarized.

II. Bayesian Multilevel Modeling

Due to the lack of a universally accepted terminology, we define a multilevel or hierarchical model as "an assembly of submodels at different levels of a hierarchy." The hierarchical structure can be constituted by stochastic dependencies and deterministic maps between the quantities

Presented as Paper 2014-1502 at the 16th AIAA Non-Deterministic Approaches Conference (SciTech 2014), National Harbor, MD, 13–17 January 2014; received 3 April 2014; revision received 15 September 2014; accepted for publication 21 November 2014; published online 28 January 2015.

*Ph.D. Candidate, D-BAUG, Institute of Structural Engineering, Chair of Risk, Safety and Uncertainty Quantification, Stefano-Franscini-Platz 5; nagel@ibk.baug.ethz.ch.

†Professor, D-BAUG, Institute of Structural Engineering, Chair of Risk, Safety and Uncertainty Quantification, Stefano-Franscini-Platz 5; sudret@ibk.baug.ethz.ch.

involved. According to that definition, multilevel modeling forms sort of an overarching theme in modern cross-disciplinary statistics. In the last two decades, it has been extensively studied from a frequentist [11,12] and a Bayesian [16,17] point of view. Adopting the latter paradigm, prior elicitation [18,19] and posterior computation [20,21] are delicate issues that have been discussed in the statistical literature. Applications of multilevel modeling encompass probabilistic inversion [22,23] and optimal combination of information [24,25]. Based on a probabilistic representation of both epistemic uncertainty and aleatory variability, Bayesian multilevel modeling establishes a natural framework for solving complex inverse problems under uncertainty. Inference can be accomplished by transforming, conditioning, and marginalizing probability distributions.

A. Uncertainty and Variability

A forward model $\mathcal{M}$: $(\boldsymbol{m}, \boldsymbol{x}, \boldsymbol{\zeta}, \boldsymbol{d}) \mapsto \tilde{\boldsymbol{y}}$ represents the system or phenomenon under consideration. It formally maps model inputs $(\boldsymbol{m}, \boldsymbol{x}, \boldsymbol{\zeta}, \boldsymbol{d}) \in \mathcal{D}_{\boldsymbol{m}} \times \mathcal{D}_{\boldsymbol{x}} \times \mathcal{D}_{\boldsymbol{\zeta}} \times \mathcal{D}_{\boldsymbol{d}}$ to outputs $\tilde{\boldsymbol{y}} = \mathcal{M}(\boldsymbol{m}, \boldsymbol{x}, \boldsymbol{\zeta}, \boldsymbol{d}) \in \mathcal{D}_{\tilde{\boldsymbol{y}}} \subset \mathbb{R}^d$. When carrying out a number of experiments, the variability of measured forward model responses can be attributed to models of input uncertainty. There are fixed yet unknown model parameters $\boldsymbol{m} \in \mathcal{D}_{\boldsymbol{m}} \subset \mathbb{R}^p$, model inputs $\boldsymbol{\zeta} \in \mathcal{D}_{\boldsymbol{\zeta}} \subset \mathbb{R}^r$ with perfectly known aleatory variability, input variables $\boldsymbol{x} \in \mathcal{D}_{\boldsymbol{x}} \subset \mathbb{R}^q$ with imperfectly known aleatory variability, and experimental conditions $\boldsymbol{d} \in \mathcal{D}_{\boldsymbol{d}} \subset \mathbb{R}^s$ that are entirely known.

With respect to a number of $i = 1, \ldots, n$ experiments, forward model inputs are represented as deterministic or stochastic objects within the Bayesian multilevel framework. Throughout the experiments, data are acquired under known but possibly different experimental conditions $\boldsymbol{d}_i$. These model inputs $\boldsymbol{d}_i$ are therefore deterministically represented. Fixed albeit unknown model parameters $\boldsymbol{m}$ are assumed to be constant over the experiments. In Bayesian fashion, they are represented as random variables $\boldsymbol{M} \sim \pi_{\boldsymbol{M}}(\boldsymbol{m})$, where the Bayesian prior distribution $\pi_{\boldsymbol{M}}(\boldsymbol{m})$ accounts for a subjective degree of belief or prior knowledge about their true values. This is the Bayesian conception of epistemic uncertainty.

Over the number of experiments, varying model inputs $\boldsymbol{\zeta}$ take on unknown experiment-specific realizations $\boldsymbol{\zeta}_i$ of conditionally independent random variables $(\boldsymbol{Z}_i|\boldsymbol{\theta}_{\boldsymbol{Z}}) \sim f_{\boldsymbol{Z}|\boldsymbol{\Theta}_{\boldsymbol{Z}}}(\boldsymbol{\zeta}_i|\boldsymbol{\theta}_{\boldsymbol{Z}})$. The conditional distribution $f_{\boldsymbol{Z}|\boldsymbol{\Theta}_{\boldsymbol{Z}}}(\boldsymbol{\zeta}_i|\boldsymbol{\theta}_{\boldsymbol{Z}})$ with known hyperparameters $\boldsymbol{\theta}_{\boldsymbol{Z}}$ states a subjective degree of belief or prior knowledge about the individual realizations $\boldsymbol{\zeta}_i$. This is a Bayesian notion of aleatory variability. Similarly model inputs $\boldsymbol{x}$ are subject to variability and take on unknown experiment-specific realizations $\boldsymbol{x}_i$ of conditionally independent random variables $(\boldsymbol{X}_i|\boldsymbol{\theta}_{\boldsymbol{X}}) \sim f_{\boldsymbol{X}|\boldsymbol{\Theta}_{\boldsymbol{X}}}(\boldsymbol{x}_i|\boldsymbol{\theta}_{\boldsymbol{X}})$. The hyperparameters $\boldsymbol{\theta}_{\boldsymbol{X}}$ determine this variability throughout the experiments and are fixed but unknown. In turn, they are modeled as random variables $\boldsymbol{\Theta}_{\boldsymbol{X}} \sim \pi_{\boldsymbol{\Theta}_{\boldsymbol{X}}}(\boldsymbol{\theta}_{\boldsymbol{X}})$, where the distribution $\pi_{\boldsymbol{\Theta}_{\boldsymbol{X}}}(\boldsymbol{\theta}_{\boldsymbol{X}})$ quantifies an a priori degree of plausibility. Random variables

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

embody the prior knowledge about the experiment-specific realizations $\boldsymbol{x}_i$.

Summarizing, marginal distributions $\pi_{\boldsymbol{M}}(\boldsymbol{m})$ and $\pi_{\boldsymbol{\Theta}_{\boldsymbol{X}}}(\boldsymbol{\theta}_{\boldsymbol{X}})$ represent parametric prior knowledge about the true values of the model parameters $\boldsymbol{m}$ and the hyperparameters $\boldsymbol{\theta}_{\boldsymbol{X}}$, whereas conditional distributions $f_{\boldsymbol{X}|\boldsymbol{\Theta}_{\boldsymbol{X}}}(\boldsymbol{x}_i|\boldsymbol{\theta}_{\boldsymbol{X}})$ and $f_{\boldsymbol{Z}|\boldsymbol{\Theta}_{\boldsymbol{Z}}}(\boldsymbol{\zeta}_i|\boldsymbol{\theta}_{\boldsymbol{Z}})$ encapsulate structural prior knowledge about the problem, i.e., information about experiment-specific $\boldsymbol{x}_i$ and $\boldsymbol{\zeta}_i$.

B. Statistical Data Model

An integral constituent of many statistical approaches to inverse problems is a residual model. Real observations $\boldsymbol{y}_i$ often deviate from model predictions $\tilde{\boldsymbol{y}}_i = \mathcal{M}(\boldsymbol{m}, \boldsymbol{x}_i, \boldsymbol{\zeta}_i, \boldsymbol{d}_i)$, even if forward model inputs were known with certainty. This discrepancy, which is due to measurement errors, numerical approximations, and model inadequacies, is often accounted for by a statistical data model $\boldsymbol{y}_i = \tilde{\boldsymbol{y}}_i + \boldsymbol{\varepsilon}_i$. Prediction errors $\boldsymbol{\varepsilon}_i$ are assumed to be realizations of random variables $\boldsymbol{E}_i \sim f_{\boldsymbol{E}_i}(\boldsymbol{\varepsilon}_i)$, e.g., with normal distributions $f_{\boldsymbol{E}_i}(\boldsymbol{\varepsilon}_i) = \mathcal{N}(\boldsymbol{\varepsilon}_i|\boldsymbol{0}, \boldsymbol{\Sigma}_i)$ and experiment-specific symmetric and positive-semidefinite covariance matrices $\boldsymbol{\Sigma}_i$. It quantifies a degree of imperfection of the forward model and experimental apparatus. Hence, observations are viewed as realizations $\boldsymbol{y}_i$ of random variables $(\boldsymbol{Y}_i|\boldsymbol{m}, \boldsymbol{x}_i, \boldsymbol{\zeta}_i)$ with distributions $f(\boldsymbol{y}_i|\boldsymbol{m}, \boldsymbol{x}_i, \boldsymbol{\zeta}_i) = f_{\boldsymbol{E}_i}(\boldsymbol{y}_i - \mathcal{M}(\boldsymbol{m}, \boldsymbol{x}_i, \boldsymbol{\zeta}_i, \boldsymbol{d}_i))$. The overall model formulated thus far can be summarized as

$$(\boldsymbol{Y}_i|\boldsymbol{m}, \boldsymbol{x}_i, \boldsymbol{\zeta}_i) \sim f_{\boldsymbol{E}_i}(\boldsymbol{y}_i - \mathcal{M}(\boldsymbol{m}, \boldsymbol{x}_i, \boldsymbol{\zeta}_i, \boldsymbol{d}_i)), \tag{2a}$$

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

$$(\boldsymbol{X}_i|\boldsymbol{\theta}_{\boldsymbol{X}}) \sim f_{\boldsymbol{X}|\boldsymbol{\Theta}_{\boldsymbol{X}}}(\boldsymbol{x}_i|\boldsymbol{\theta}_{\boldsymbol{X}}), \tag{2c}$$

$$\boldsymbol{\Theta}_{\boldsymbol{X}} \sim \pi_{\boldsymbol{\Theta}_{\boldsymbol{X}}}(\boldsymbol{\theta}_{\boldsymbol{X}}), \tag{2d}$$

$$(\boldsymbol{Z}_i|\boldsymbol{\theta}_{\boldsymbol{Z}}) \sim f_{\boldsymbol{Z}|\boldsymbol{\Theta}_{\boldsymbol{Z}}}(\boldsymbol{\zeta}_i|\boldsymbol{\theta}_{\boldsymbol{Z}}). \tag{2e}$$

This model is composed of conditional probabilistic and deterministic relations between the quantities involved. As per our previous definition, it is a generic Bayesian multilevel model. An intuitive model representation is provided by a directed acyclic graph (DAG) [26,27], such as shown in Fig. 1. Vertices symbolize unknown (○) or known (□) quantities, and directed edges represent their deterministic (→) or probabilistic (⇢) relations. In Sec. III, we will devise a model for analyzing perfect observations $\tilde{\boldsymbol{y}}_i$ instead of the "imperfect" ones $\boldsymbol{y}_i = \tilde{\boldsymbol{y}}_i + \boldsymbol{\varepsilon}_i$. Unlike the latter, the randomness of the former is exclusively attributed to forward model input variability, as discussed in Sec. II.A. Unless stated or denoted otherwise, random variables in Eq. (2) are assumed to be (conditionally) independent. This defines a joint overall probability density of all probabilistic quantities. By conditioning and marginalizing this overall density at one's convenience, one can derive meaningful probability densities. For inferential purposes, these are certain prior and posterior distributions that we will explain in the following.

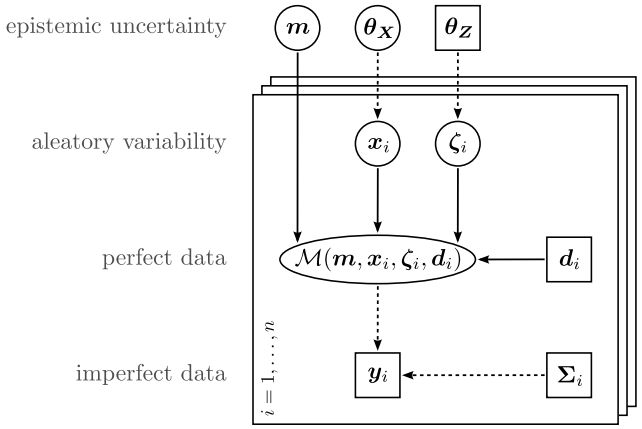

Fig. 1 DAG of the generic multilevel model.

C. Inference in Multilevel Models

In what follows, $\langle \boldsymbol{q}_i \rangle$ denotes a tuple $\langle \boldsymbol{q}_i \rangle_{1 \le i \le n} = (\boldsymbol{q}_1, \boldsymbol{q}_2, \ldots, \boldsymbol{q}_n)$. Conditioned on the priorly known hyperparameters $\boldsymbol{\theta}_Z$, the joint prior distribution of the unknowns $(\boldsymbol{m}, \langle \boldsymbol{x}_i \rangle, \langle \boldsymbol{\zeta}_i \rangle, \boldsymbol{\theta}_X)$ is given as

$$\pi(\boldsymbol{m}, \langle \boldsymbol{x}_i \rangle, \langle \boldsymbol{\zeta}_i \rangle, \boldsymbol{\theta}_X | \boldsymbol{\theta}_Z) = \left(\prod_{i=1}^{n} f_{X|\boldsymbol{\Theta}_X}(\boldsymbol{x}_i | \boldsymbol{\theta}_X) \right) \left(\prod_{i=1}^{n} f_{Z|\boldsymbol{\Theta}_Z}(\boldsymbol{\zeta}_i | \boldsymbol{\theta}_Z) \right) \pi_{\boldsymbol{\Theta}_X}(\boldsymbol{\theta}_X) \pi_M(\boldsymbol{m}). \tag{3}$$

It summarizes the available parametric and structural prior knowledge. The joint posterior distribution of the unknowns $(\boldsymbol{m}, \langle \boldsymbol{x}_i \rangle, \langle \boldsymbol{\zeta}_i \rangle, \boldsymbol{\theta}_X)$ is obtained by further conditioning the prior Eq. (3) on the data $\langle \boldsymbol{y}_i \rangle$. By virtue of Bayes' law, this posterior is, up to a scale factor, found as

$$\pi(\boldsymbol{m}, \langle \boldsymbol{x}_i \rangle, \langle \boldsymbol{\zeta}_i \rangle, \boldsymbol{\theta}_X | \langle \boldsymbol{y}_i \rangle, \boldsymbol{\theta}_Z) \propto \left(\prod_{i=1}^{n} f_{\boldsymbol{E}_i}(\boldsymbol{y}_i - \mathcal{M}(\boldsymbol{m}, \boldsymbol{x}_i, \boldsymbol{\zeta}_i, \boldsymbol{d}_i)) \right) \pi(\boldsymbol{m}, \langle \boldsymbol{x}_i \rangle, \langle \boldsymbol{\zeta}_i \rangle, \boldsymbol{\theta}_X | \boldsymbol{\theta}_Z). \tag{4}$$

The posterior degree of plausibility about the quantities of interest (QOI) can be extracted by marginalizing the posterior Eq. (4) over parameters considered nuisance [28,29]. Provided $(\boldsymbol{m}, \boldsymbol{\theta}_X)$ are QOI and $(\langle \boldsymbol{x}_i \rangle, \langle \boldsymbol{\zeta}_i \rangle)$ are nuisance parameters, the correspondingly marginalized posterior is

$$\pi(\boldsymbol{m}, \boldsymbol{\theta}_X | \langle \boldsymbol{y}_i \rangle, \boldsymbol{\theta}_Z) = \int_{\mathcal{D}_x^n} \int_{\mathcal{D}_\zeta^n} \pi(\boldsymbol{m}, \langle \boldsymbol{x}_i \rangle, \langle \boldsymbol{\zeta}_i \rangle, \boldsymbol{\theta}_X | \langle \boldsymbol{y}_i \rangle, \boldsymbol{\theta}_Z) \, \mathrm{d}\langle \boldsymbol{x}_i \rangle \, \mathrm{d}\langle \boldsymbol{\zeta}_i \rangle, \tag{5}$$

where $\mathrm{d}\langle \boldsymbol{x}_i \rangle = \mathrm{d}\boldsymbol{x}_1 \ldots \mathrm{d}\boldsymbol{x}_n$, and $\mathrm{d}\langle \boldsymbol{\zeta}_i \rangle = \mathrm{d}\boldsymbol{\zeta}_1 \ldots \mathrm{d}\boldsymbol{\zeta}_n$. To summarize the genuinely unique approach to Bayesian inference in multilevel models is to construct the posterior of the QOI $(\boldsymbol{m}, \boldsymbol{\theta}_X)$ by conditioning on the knowns $(\langle \boldsymbol{y}_i \rangle, \boldsymbol{\theta}_Z)$ and, subsequently, marginalizing out nuisance $(\langle \boldsymbol{x}_i \rangle, \langle \boldsymbol{\zeta}_i \rangle)$.

D. Marginalized Formulation

Equivalently, one could solve a marginal formulation of the multilevel calibration problem, with a marginal prior $\pi(\boldsymbol{m}, \boldsymbol{\theta}_X) = \pi_M(\boldsymbol{m}) \pi_{\boldsymbol{\Theta}_X}(\boldsymbol{\theta}_X)$ and a marginalized version of the likelihood $\mathcal{L}(\langle \boldsymbol{y}_i \rangle | \boldsymbol{m}, \boldsymbol{\theta}_X, \boldsymbol{\theta}_Z)$ [30,31]. One therefore factorizes the marginalized posterior Eq. (5) as

$$\pi(\boldsymbol{m}, \boldsymbol{\theta}_X | \langle \boldsymbol{y}_i \rangle, \boldsymbol{\theta}_Z) \propto \left(\prod_{i=1}^{n} f(\boldsymbol{y}_i | \boldsymbol{m}, \boldsymbol{\theta}_X, \boldsymbol{\theta}_Z) \right) \pi_M(\boldsymbol{m}) \pi_{\boldsymbol{\Theta}_X}(\boldsymbol{\theta}_X), \tag{6}$$

where $f(\boldsymbol{y}_i | \boldsymbol{m}, \boldsymbol{\theta}_X, \boldsymbol{\theta}_Z)$ is the marginalized density of the observation $\boldsymbol{y}_i$. The aleatory variables $(\boldsymbol{x}_i, \boldsymbol{\zeta}_i)$ have been eliminated based on the integration

$$f(\boldsymbol{y}_i | \boldsymbol{m}, \boldsymbol{\theta}_X, \boldsymbol{\theta}_Z) = \int_{\mathcal{D}_x} \int_{\mathcal{D}_\zeta} f_{\boldsymbol{E}_i}(\boldsymbol{y}_i - \mathcal{M}(\boldsymbol{m}, \boldsymbol{x}_i, \boldsymbol{\zeta}_i, \boldsymbol{d}_i)) f_{X|\boldsymbol{\Theta}_X}(\boldsymbol{x}_i | \boldsymbol{\theta}_X) f_{Z|\boldsymbol{\Theta}_Z}(\boldsymbol{\zeta}_i | \boldsymbol{\theta}_Z) \, \mathrm{d}\boldsymbol{x}_i \, \mathrm{d}\boldsymbol{\zeta}_i. \tag{7}$$

When defining the marginalized or integrated likelihood as $\mathcal{L}(\langle \boldsymbol{y}_i \rangle | \boldsymbol{m}, \boldsymbol{\theta}_X, \boldsymbol{\theta}_Z) = \prod_{i=1}^{n} f(\boldsymbol{y}_i | \boldsymbol{m}, \boldsymbol{\theta}_X, \boldsymbol{\theta}_Z)$, the marginal posterior Eq. (6) simply writes as $\pi(\boldsymbol{m}, \boldsymbol{\theta}_X | \langle \boldsymbol{y}_i \rangle, \boldsymbol{\theta}_Z) \propto \mathcal{L}(\langle \boldsymbol{y}_i \rangle | \boldsymbol{m}, \boldsymbol{\theta}_X, \boldsymbol{\theta}_Z) \pi(\boldsymbol{m}, \boldsymbol{\theta}_X)$. Directly computing the marginal posterior Eq. (5) involves a lower-dimensional parameter space compared to computing the joint posterior Eq. (4). However, it requires the computation of the integrals in Eq. (7).

III. Perfect Data Model

In Sec. II.B, the residual model was introduced as a representation of the imperfection of the forward solver and measurement device. As a consequence, in Eq. (2a), observations were regarded as realizations $\boldsymbol{y}_i = \tilde{\boldsymbol{y}}_i + \boldsymbol{\varepsilon}_i$ of random variables $(\boldsymbol{Y}_i | \boldsymbol{m}, \boldsymbol{x}_i, \boldsymbol{\zeta}_i)$ conditioned on direct forward model inputs. The residual assumptions that led to Eq. (2a) had equipped the data space $\mathcal{D}_{\tilde{\boldsymbol{y}}}$ with a probability model. We introduce the term imperfect for this statistical data model in order to distinguish it from the following.

In this paper, we are interested in the experimental situation where $\tilde{\boldsymbol{y}}_i = \mathcal{M}(\boldsymbol{m}, \boldsymbol{x}_i, \boldsymbol{\zeta}_i, \boldsymbol{d}_i)$ can be directly observed, e.g., due to noise-free measurements and a "sufficiently accurate" forward simulator. Hereinafter, we will refer to this limiting case as to involve perfect data. Not being premised on a residual model, the data are viewed as realizations $\tilde{\boldsymbol{y}}_i$ of random variables $(\tilde{\boldsymbol{Y}}_i | \boldsymbol{m}, \boldsymbol{\theta}_X, \boldsymbol{\theta}_Z)$ conditioned on the "highest-level"

quantities. When provided an appropriate probability model $f(\tilde{\boldsymbol{y}}_i|\boldsymbol{m},\boldsymbol{\theta}_{\boldsymbol{X}},\boldsymbol{\theta}_{\boldsymbol{Z}})$ of those random variables, a Bayesian multilevel model for perfect data can be written as

$$(\tilde{\boldsymbol{Y}}_i|\boldsymbol{m},\boldsymbol{\theta}_{\boldsymbol{X}},\boldsymbol{\theta}_{\boldsymbol{Z}}) \sim f(\tilde{\boldsymbol{y}}_i|\boldsymbol{m},\boldsymbol{\theta}_{\boldsymbol{X}},\boldsymbol{\theta}_{\boldsymbol{Z}}), \tag{8a}$$

$$(\boldsymbol{M},\boldsymbol{\Theta}_{\boldsymbol{X}}) \sim \pi(\boldsymbol{m},\boldsymbol{\theta}_{\boldsymbol{X}}) = \pi_{\boldsymbol{M}}(\boldsymbol{m})\pi_{\boldsymbol{\Theta}_{\boldsymbol{X}}}(\boldsymbol{\theta}_{\boldsymbol{X}}). \tag{8b}$$

As before, prior knowledge about the unknowns $(\boldsymbol{m},\boldsymbol{\theta}_{\boldsymbol{X}})$ is embodied in Eq. (8b). As a function of the unknowns $(\boldsymbol{m},\boldsymbol{\theta}_{\boldsymbol{X}})$, with the density Eq. (8a), one can formulate a residual-free version of the likelihood

$$\mathcal{L}(\langle\tilde{\boldsymbol{y}}_i\rangle|\boldsymbol{m},\boldsymbol{\theta}_{\boldsymbol{X}},\boldsymbol{\theta}_{\boldsymbol{Z}}) = \prod_{i=1}^{n} f(\tilde{\boldsymbol{y}}_i|\boldsymbol{m},\boldsymbol{\theta}_{\boldsymbol{X}},\boldsymbol{\theta}_{\boldsymbol{Z}}). \tag{9}$$

For reasons that will be discussed next, we call Eq. (9) the transformed likelihood. As usual, Bayesian analysis proceeds by conditioning the prior on the acquired data $\langle\tilde{\boldsymbol{y}}_i\rangle$. The posterior follows as

$$\pi(\boldsymbol{m},\boldsymbol{\theta}_{\boldsymbol{X}}|\langle\tilde{\boldsymbol{y}}_i\rangle,\boldsymbol{\theta}_{\boldsymbol{Z}}) \propto \mathcal{L}(\langle\tilde{\boldsymbol{y}}_i\rangle|\boldsymbol{m},\boldsymbol{\theta}_{\boldsymbol{X}},\boldsymbol{\theta}_{\boldsymbol{Z}})\pi(\boldsymbol{m},\boldsymbol{\theta}_{\boldsymbol{X}}). \tag{10}$$

Note that the notation of Eq. (10) is reminiscent of classical Bayesian inversion. Indeed, the multilevel character of the problem manifests in the likelihood function Eq. (9), which we will now specify in greater detail.

A. Forward Uncertainty Propagation

Let $\boldsymbol{X}_i \sim f_{\boldsymbol{X}|\boldsymbol{\Theta}_{\boldsymbol{X}}}(\boldsymbol{x}_i|\boldsymbol{\theta}_{\boldsymbol{X}})$ and $\boldsymbol{Z}_i \sim f_{\boldsymbol{Z}|\boldsymbol{\Theta}_{\boldsymbol{Z}}}(\boldsymbol{\zeta}_i|\boldsymbol{\theta}_{\boldsymbol{Z}})$ be the aleatory variables in Eqs. (2c) and (2e) that have independent marginal densities for given hyperparameters $(\boldsymbol{\theta}_{\boldsymbol{X}},\boldsymbol{\theta}_{\boldsymbol{Z}})$. Now, consider the map $\mathcal{M}_{\boldsymbol{m},\boldsymbol{d}_i}\colon (\boldsymbol{x}_i,\boldsymbol{\zeta}_i) \mapsto \tilde{\boldsymbol{y}}_i = \mathcal{M}(\boldsymbol{m},\boldsymbol{x}_i,\boldsymbol{\zeta}_i,\boldsymbol{d}_i)$ that the forward model defines for fixed inputs $(\boldsymbol{m},\boldsymbol{d}_i)$. For given $(\boldsymbol{m},\boldsymbol{d}_i,\boldsymbol{\theta}_{\boldsymbol{X}},\boldsymbol{\theta}_{\boldsymbol{Z}})$, the random variables in Eq. (8a) are constructed by forward propagation of the aleatory input uncertainties through the function $\mathcal{M}_{\boldsymbol{m},\boldsymbol{d}_i}(\boldsymbol{x}_i,\boldsymbol{\zeta}_i)$ into an output uncertainty. Provided the existence of a corresponding density, the transformed random variables $(\tilde{\boldsymbol{Y}}_i|\boldsymbol{m},\boldsymbol{\theta}_{\boldsymbol{X}},\boldsymbol{\theta}_{\boldsymbol{Z}}) = \mathcal{M}_{\boldsymbol{m},\boldsymbol{d}_i}(\boldsymbol{X}_i,\boldsymbol{Z}_i)$ follow

$$f(\tilde{\boldsymbol{y}}_i|\boldsymbol{m},\boldsymbol{\theta}_{\boldsymbol{X}},\boldsymbol{\theta}_{\boldsymbol{Z}}) = \int_{\mathcal{D}_{\boldsymbol{x}}}\int_{\mathcal{D}_{\boldsymbol{\zeta}}} \delta(\tilde{\boldsymbol{y}}_i - \mathcal{M}_{\boldsymbol{m},\boldsymbol{d}_i}(\boldsymbol{x}_i,\boldsymbol{\zeta}_i)) f_{\boldsymbol{X}|\boldsymbol{\Theta}_{\boldsymbol{X}}}(\boldsymbol{x}_i|\boldsymbol{\theta}_{\boldsymbol{X}}) f_{\boldsymbol{Z}|\boldsymbol{\Theta}_{\boldsymbol{Z}}}(\boldsymbol{\zeta}_i|\boldsymbol{\theta}_{\boldsymbol{Z}})\,\mathrm{d}\boldsymbol{x}_i\,\mathrm{d}\boldsymbol{\zeta}_i. \tag{11}$$

Here, δ denotes the Dirac delta distribution. At this point, it is noted that epistemic uncertainties are not propagated through the forward model. The transformed density Eq. (11) establishes the proper probability model in the space of data $\mathcal{D}_{\tilde{\boldsymbol{y}}}$, and thereby defines the transformed likelihood function Eq. (9). More formally, one can derive Eq. (11) as the density function of a push-forward measure [32].

B. Relation Between Imperfect and Perfect Data

It is interesting to examine the relation between the perfect and the imperfect data model. The full multilevel model Eq. (2) was defined by a cascade of random variables that were conditioned on one another. Constructing the marginal model $(\boldsymbol{Y}_i|\boldsymbol{m},\boldsymbol{\theta}_{\boldsymbol{X}},\boldsymbol{\theta}_{\boldsymbol{Z}})$ in Eq. (7) was then based on the marginalization of aleatory variables, whereas the mechanism to formulate the perfect data model $(\tilde{\boldsymbol{Y}}_i|\boldsymbol{m},\boldsymbol{\theta}_{\boldsymbol{X}},\boldsymbol{\theta}_{\boldsymbol{Z}})$ in Eq. (11) was their propagation. Those two operations are related by writing $(\boldsymbol{Y}_i|\boldsymbol{m},\boldsymbol{\theta}_{\boldsymbol{X}},\boldsymbol{\theta}_{\boldsymbol{Z}}) = (\tilde{\boldsymbol{Y}}_i|\boldsymbol{m},\boldsymbol{\theta}_{\boldsymbol{X}},\boldsymbol{\theta}_{\boldsymbol{Z}}) + \boldsymbol{E}_i$ as the sum of independent random variables. The convolution integral

$$f(\boldsymbol{y}_i|\boldsymbol{m},\boldsymbol{\theta}_{\boldsymbol{X}},\boldsymbol{\theta}_{\boldsymbol{Z}}) = \int_{\mathcal{D}_{\tilde{\boldsymbol{y}}}} f(\tilde{\boldsymbol{y}}_i|\boldsymbol{m},\boldsymbol{\theta}_{\boldsymbol{X}},\boldsymbol{\theta}_{\boldsymbol{Z}}) f_{\boldsymbol{E}_i}(\boldsymbol{y}_i - \tilde{\boldsymbol{y}}_i)\,\mathrm{d}\tilde{\boldsymbol{y}}_i \tag{12}$$

then establishes the relation between the distributions $f(\boldsymbol{y}_i|\boldsymbol{m},\boldsymbol{\theta}_{\boldsymbol{X}},\boldsymbol{\theta}_{\boldsymbol{Z}})$ and $f(\tilde{\boldsymbol{y}}_i|\boldsymbol{m},\boldsymbol{\theta}_{\boldsymbol{X}},\boldsymbol{\theta}_{\boldsymbol{Z}})$ of imperfect and perfect data, respectively. Of course, when plugging Eq. (11) into Eq. (12), one easily rederives Eq. (7).

For finite measurement uncertainty $\|\boldsymbol{\Sigma}_i\| > 0$ the two models involving imperfect and perfect data describe distinct experimental situations. However, in the limiting case $\|\boldsymbol{\Sigma}_i\| \to 0$ with $\mathbb{E}[(\boldsymbol{Y}_i|\boldsymbol{m},\boldsymbol{x}_i,\boldsymbol{\zeta}_i) - \tilde{\boldsymbol{y}}_i)^2] \to 0$ and $f_{\boldsymbol{E}_i} \to \delta$, the marginalized likelihood approaches the transformed likelihood. Naturally, this meets one's expectations. In classical Bayesian inversion, the small-noise limit implies that the posterior of the unobservables shrinks to exact solutions of the inverse problem, i.e., posterior consistency [33,34]. In multilevel inversion, however, the zero-noise limit leads to convergence of the posterior Eq. (6) to Eq. (10).

C. Kernel Density Estimation

Since the transformed likelihood Eq. (9) will be rarely available in analytical form, commonly one has to rely on numerical approximations. A possible approach is to simulate the response density Eq. (11) by Monte Carlo (MC) sampling and kernel density estimation (KDE) [35,36] and evaluate the likelihood accordingly [37,38].

In the d-variate case, given a sample $(\tilde{\boldsymbol{y}}^{(1)},\ \ldots,\ \tilde{\boldsymbol{y}}^{(K)})$ from some distribution with an unknown density $f(\tilde{\boldsymbol{y}})$, a kernel smoothing (KS) estimate of this density is given as $\hat{f}(\tilde{\boldsymbol{y}}) = K^{-1}\sum_{k=1}^{K}\mathcal{K}_{\boldsymbol{H}}(\tilde{\boldsymbol{y}} - \tilde{\boldsymbol{y}}^{(k)})$. The scaled kernel $\mathcal{K}_{\boldsymbol{H}}(\tilde{\boldsymbol{y}}) = |\boldsymbol{H}|^{-1/2}\mathcal{K}(\boldsymbol{H}^{-1/2}\tilde{\boldsymbol{y}})$ is defined by a kernel function $\mathcal{K}$ and a symmetric and positive-definite bandwidth matrix $\boldsymbol{H}$. Common types of bandwidth matrices are multiples of the identity matrix $\boldsymbol{H} = h^2 1_d$ for $h > 0$ or diagonal matrices $\boldsymbol{H} = \mathrm{diag}(h_1^2,\ \ldots,\ h_d^2)$ with $h_1,\ \ldots,\ h_d > 0$. According to certain criteria and assumptions, "optimal" bandwidths are commonly selected to prevent over- and undersmoothing. This amounts to a classical tradeoff between the bias and the variance of the estimation. The mean squared error (MSE) of a KDE $\hat{f}(\tilde{\boldsymbol{y}})$ at a point $\tilde{\boldsymbol{y}}$ can be decomposed into its variance and the square of its bias:

$$\mathrm{MSE}[\hat{f}(\tilde{\boldsymbol{y}})] = \mathbb{E}[(\hat{f}(\tilde{\boldsymbol{y}}) - f(\tilde{\boldsymbol{y}}))^2] = \mathrm{Var}[\hat{f}(\tilde{\boldsymbol{y}})] + \mathrm{Bias}^2[\hat{f}(\tilde{\boldsymbol{y}})]. \tag{13}$$

An optimal bandwidth normally strives to minimize the mean integrated squared error (i.e., when Eq. (13) is integrated over $\tilde{\boldsymbol{y}}$) by balancing the variance against the bias. The KDE of a univariate density similar to a Gaussian with kernels of the same type can be based on Silverman's normal reference rule [35], i.e., $h = (4/(3\mathrm{K}))^{1/5}\hat{\sigma}$. Here, $\hat{\sigma}$ is the standard deviation of the sample $(\tilde{y}^{(1)}, \ldots, \tilde{y}^{(K)})$.

Based on MC and KDE techniques, one can estimate the target density Eq. (11). On top of that, we propose to estimate the transformed likelihood Eq. (9) through

$$\hat{\mathcal{L}}_{\mathrm{KS}}(\langle\tilde{\boldsymbol{y}}_i\rangle|\boldsymbol{m},\boldsymbol{\theta}_X,\boldsymbol{\theta}_Z) = \prod_{i=1}^{n}\left(\frac{1}{K}\sum_{k=1}^{K}\mathcal{K}_{\boldsymbol{H}}(\tilde{\boldsymbol{y}}_i - \tilde{\boldsymbol{y}}_i^{(k)})\right), \quad \text{with} \quad \begin{cases} \boldsymbol{x}^{(k)} \sim f_{X|\boldsymbol{\Theta}_X}(\boldsymbol{x}^{(k)}|\boldsymbol{\theta}_X), \\ \boldsymbol{\zeta}^{(k)} \sim f_{\boldsymbol{Z}|\boldsymbol{\Theta}_Z}(\boldsymbol{\zeta}^{(k)}|\boldsymbol{\theta}_Z), \\ \tilde{\boldsymbol{y}}_i^{(k)} = \mathcal{M}(\boldsymbol{m},\boldsymbol{x}^{(k)},\boldsymbol{\zeta}^{(k)},\boldsymbol{d}_i). \end{cases} \tag{14}$$

For $k = 1, \ldots, K$, forward model responses $\tilde{\boldsymbol{y}}_i^{(k)} = \mathcal{M}(\boldsymbol{m},\boldsymbol{x}^{(k)},\boldsymbol{\zeta}^{(k)},\boldsymbol{d}_i)$ are computed for given arguments $(\boldsymbol{m},\boldsymbol{d}_i)$ and for further inputs $\boldsymbol{x}^{(k)}$ and $\boldsymbol{\zeta}^{(k)}$ that are randomly sampled from their parent distributions. These distributions, $f_{X|\boldsymbol{\Theta}_X}(\boldsymbol{x}^{(k)}|\boldsymbol{\theta}_X)$ and $f_{\boldsymbol{Z}|\boldsymbol{\Theta}_Z}(\boldsymbol{\zeta}^{(k)}|\boldsymbol{\theta}_Z)$, are defined by values of the hyperparameters $(\boldsymbol{\theta}_X,\boldsymbol{\theta}_Z)$, respectively.

With Eq. (14) we have a statistical estimator of the likelihood Eq. (9) at hand. However, this estimator is accompanied by additional free parameters, i.e., the type of kernel function $\mathcal{K}$, the number of samples K, and the bandwidth $\boldsymbol{H}$. As will be discussed in the following, the application of classical criteria, which usually assist in adjusting free parameters, is questionable in the context of posterior computation via Markov chain Monte Carlo.

IV. Bayesian Computations

More often than not, Bayesian posterior distributions do not have analytic closed-form solutions. Nevertheless, one can explore posteriors through Markov chain Monte Carlo (MCMC) sampling techniques [39]. The principle of MCMC is to realize an ergodic Markov chain over the prior support for which the invariant distribution equals the posterior. Let $\pi_0(\boldsymbol{q})$ be the prior and $\pi_1(\boldsymbol{q}) \propto \mathcal{L}(\boldsymbol{q})\pi_0(\boldsymbol{q})$ the posterior of an unknown QOI $\boldsymbol{q}$. The Markov kernel $\mathcal{K}$ defines the density $\mathcal{K}(\boldsymbol{q}^{(t)},\boldsymbol{q}^{(t+1)})$ of the transition probability from $\boldsymbol{q}^{(t)}$ to $\boldsymbol{q}^{(t+1)}$, i.e., the state of the chain at times t and $t+1$. The posterior $\pi_1(\boldsymbol{q})$ is said to be an invariant distribution of the Markov chain if $\pi_1(\boldsymbol{q}^{(t+1)}) = \int\pi_1(\boldsymbol{q}^{(t)})\mathcal{K}(\boldsymbol{q}^{(t)},\boldsymbol{q}^{(t+1)})\,\mathrm{d}\boldsymbol{q}^{(t)}$. This is abbreviated as $\pi_1 = \pi_1\mathcal{K}$. Detailed balance [i.e., time reversibility $\pi_1(\boldsymbol{q}^{(t)})\mathcal{K}(\boldsymbol{q}^{(t)},\boldsymbol{q}^{(t+1)}) = \pi_1(\boldsymbol{q}^{(t+1)})\mathcal{K}(\boldsymbol{q}^{(t+1)},\boldsymbol{q}^{(t)})$] is a sufficient condition for the Markov chain to leave the posterior $\pi_1(\boldsymbol{q})$ invariant. It normally serves as the guiding principle for the construction of a Markov chain appropriate for posterior exploration. The Metropolis–Hastings (MH) algorithm establishes a prototypical class of MCMC techniques that relies on this principle [40,41].

A. The Metropolis–Hastings Algorithm

Initialized at $\boldsymbol{q}^{(0)}$, the MH algorithm generates a Markov chain with steady-state distribution $\pi_1(\boldsymbol{q})$ by iteratively applying the Markov chain transition kernel as follows. For a current state $\boldsymbol{q}^{(t)}$ of the Markov chain, a candidate state $\boldsymbol{q}^{(\star)} \sim P(\boldsymbol{q}^{(\star)}|\boldsymbol{q}^{(t)})$ is sampled from a proposal distribution $P(\boldsymbol{q}^{(\star)}|\boldsymbol{q}^{(t)})$. The proposal state becomes accepted (i.e., $\boldsymbol{q}^{(t+1)} = \boldsymbol{q}^{(\star)}$) with probability

$$\alpha(\boldsymbol{q}^{(\star)}|\boldsymbol{q}^{(t)}) = \min\left(1, \frac{\pi_1(\boldsymbol{q}^{(\star)})P(\boldsymbol{q}^{(t)}|\boldsymbol{q}^{(\star)})}{\pi_1(\boldsymbol{q}^{(t)})P(\boldsymbol{q}^{(\star)}|\boldsymbol{q}^{(t)})}\right). \tag{15}$$

Otherwise, it is rejected, i.e., $\boldsymbol{q}^{(t+1)} = \boldsymbol{q}^{(t)}$. The MH transition kernel defined this way satisfies detailed balance. Note that the MH correction Eq. (15) requires the computation of posterior ratios; hence, only unscaled posterior densities have to be evaluated. Classical random walk Metropolis sampling is based on local proposals, e.g., sampling candidate states from a Gaussian $\boldsymbol{q}^{(\star)} \sim \mathcal{N}(\boldsymbol{q}^{(\star)}|\boldsymbol{q}^{(t)},\boldsymbol{\Sigma}_{\boldsymbol{q}})$ with mean $\boldsymbol{q}^{(t)}$ and covariance matrix $\boldsymbol{\Sigma}_{\boldsymbol{q}}$. Independence MH samplers are based on nonlocal proposals, e.g., sampling candidate states from the prior $\boldsymbol{q}^{(\star)} \sim \pi_0(\boldsymbol{q}^{(\star)})$ or from some suitable approximation of the posterior $\boldsymbol{q}^{(\star)} \sim \hat{\pi}_1(\boldsymbol{q}^{(\star)})$.

B. Key Challenges

Typically, MCMC sampling calls for a high number of forward model runs for likelihood evaluations in Eq. (15). Besides that, the degree as to which MCMC samples are autocorrelated governs their quality as posterior representatives. The design and efficient tuning of MCMC algorithms therefore aims at optimizing the mixing properties, i.e., the speed of convergence of the Markov chain toward its equilibrium distribution. This is a challenging and highly problem-dependent task. MCMC methods demand careful convergence diagnostics, i.e., the assessment of when the Markov chain has reached its target distribution and has lost the dependency on its initialization [42,43]. Moreover, MCMC suffers from difficulties in exploring high-dimensional parameter spaces and multimodal posteriors. Multilevel model calibration poses further multilevel-specific MCMC burdens. Sampling the posterior Eq. (10) imposes estimations of the likelihood Eq. (9), which is analytically intractable [44].

C. Posterior Fidelity

Due to Bayes' law, deterministic closed-form approximations $\bar{\mathcal{L}}$ of the likelihood $\mathcal{L}$ directly induce approximations on the level of the posterior $\bar{\pi}_1(\boldsymbol{q}) \propto \bar{\mathcal{L}}(\boldsymbol{q})\pi_0(\boldsymbol{q})$. However, if the posterior is explored by means of MCMC and calls to the likelihood function $\mathcal{L}$ are replaced by calls to a statistical estimator $\hat{\mathcal{L}}$, then a modification is introduced on the level of the Markov chain transition kernel [45,46]. There is no reason to expect that the modified MH transition kernel with the "randomized" version of Eq. (15) leaves the posterior invariant, i.e., in general, $\pi_1 \neq \pi_1\hat{\mathcal{K}}$. Consequently, there arises the question of whether an equilibrium distribution $\hat{\pi}_1 = \hat{\pi}_1\hat{\mathcal{K}}$ actually exists and, in the event that it does, as to what extent the induced distribution $\hat{\pi}_1$ is in congruence with the true posterior π_1. To pay tribute to these issues, we introduce the term posterior fidelity as a qualitative measure of the similarity between $\hat{\pi}_1$ and π_1.

Moreover, there is the practical question of how free algorithmic parameters (e.g., the number of response samples K and the kernel bandwidth $\boldsymbol{H}$) can be set in order to provide a convenient tradeoff between posterior fidelity and computational feasibility, i.e., an optimal parameter tuning. We suppose that it is indeed possible to define certain criteria that parameter tuning can be based on. Even though this is beyond the scope of this paper, we have some preliminary comments. As done in the following, when postulating the existence and uniqueness of an equilibrium

distribution, one can further argue that "small" changes in the transition kernel only cause small changes in the distribution. Given the current state $\boldsymbol{q}^{(t)}$ of the Markov chain, which is assumed to be ergodic, in the MH correction Eq. (15), a certain random decision is made whether to approve or to refuse a candidate state $\boldsymbol{q}^{(\star)}$. This binary decision follows the computation of the posterior ratio $\pi_1(\boldsymbol{q}^{(\star)})/\pi_1(\boldsymbol{q}^{(t)})$. Thus, provided that the ratio of estimated likelihoods approximates the true ratio "reasonably well," i.e., in some sense

$$\frac{\hat{\mathcal{L}}(\boldsymbol{q}^{(\star)})}{\hat{\mathcal{L}}(\boldsymbol{q}^{(t)})} \approx \frac{\mathcal{L}(\boldsymbol{q}^{(\star)})}{\mathcal{L}(\boldsymbol{q}^{(t)})}, \tag{16}$$

an "appropriate" decision is being made. High posterior fidelity is ensured on the condition that appropriate decisions are being frequently made over the course of the Markov process, i.e., detailed balance is maintained in some average sense. That this is indeed the case depends on a complex interplay between the quality of the estimation $\hat{\mathcal{L}}$ of $\mathcal{L}$, the true posterior π_1, and the proposal distribution P.

Similar arguments have been invoked in connection with Bayesian inversion of MC forward models, where algorithms have been designed that make use of forward model evaluations at multiple MC resolutions [47]. Moreover, approximate MH corrections have been proposed in the context of big data, where evaluating the likelihood for partial chunks of observations from a large dataset trades off the bias and the variance of MCMC posterior sampling [48,49]. To provide a likelihood simulator that does not exhibit variations between consecutive evaluations for the same arguments, it was proposed to employ a common random number generator seed [50]. Once the seed is chosen, likelihood evaluations became effectively deterministic. The pseudo-marginal approach to MCMC [51] provides a strong theoretical result regarding posterior fidelity. It is shown that, for unbiased likelihood estimators, the exact posterior can be sampled. Although, this is accompanied by a slowdown in the mixing properties of the Markov chain. With that said, one may suppose that it is preferable to minimize the bias in Eq. (13) instead of the total mean squared error.

V. The NASA Langley Multidisciplinary UQ Challenge

Now, we will interpret the uncertainty characterization subproblem A of the NASA Langley Multidisciplinary UQ Challenge [1,2] in Bayesian terms. For that purpose, we will compose an appropriate Bayesian multilevel model and formulate the main objective (i.e., the reduction of epistemic uncertainties) as Bayesian calibration of this multilevel model. Inputs $(p_1, p_2, p_3, p_4, p_5)$ of the forward model $\mathcal{M} \equiv h_1$ are subject to uncertainty. Model inputs $\boldsymbol{\zeta} \equiv p_3$, which are subject to aleatory variability, constitute the category 1 parameters. There is epistemic uncertainty about the true value of the category 2 model parameter $\boldsymbol{m} \equiv p_2$. Category 3 subsumes those parameters $\boldsymbol{x} \equiv (p_1, p_4, p_5)$ that are subject to a mixed-type uncertainty. Generally, population distributions of experiment-specific variables are provided by the organizers of the challenge problem, whereas prior marginals of the QOI are uninformative interpretations of the epistemic intervals given.

A. Category 1: Aleatory Uncertainty

For experiments $i = 1, \ldots, n$, category 1 model inputs $\boldsymbol{\zeta} \equiv p_3 \in [0, 1]$ take on experiment-specific realizations $p_{3,i}$. The population distribution is a uniform distribution $\mathcal{U}(p_{3,i}|a_3, b_3)$ determined by perfectly known hyperparameters $\boldsymbol{\theta}_Z \equiv \boldsymbol{\theta}_3 = (a_3, b_3)$ with $(a_3, b_3) = (0, 1)$. We write this as follows:

$$(P_{3,i}|\boldsymbol{\theta}_3) \sim f_3(p_{3,i}|\boldsymbol{\theta}_3) = \mathcal{U}(p_{3,i}|0, 1). \tag{17}$$

It corresponds to a prescribed aleatory variability or structural uncertainty that is irreducible in the sense that, by analyzing available data $\tilde{y}_i$ for $i = 1, \ldots, n$, "past" realizations $p_{3,i}$ could be inferred in principle; whereas the knowledge about "future" realizations $p_{3,i'}$ with $i' > n$ cannot be improved.

B. Category 2: Epistemic Uncertainty

Category 2 model inputs are physically fixed yet unknown model parameters $\boldsymbol{m} \equiv p_2 \in [0, 1]$. A given epistemic interval $\Delta = [0, 1]$ is known to contain the true value of p_2 before any data analysis. We translate this available information into a flat and uniform Bayesian prior probability density

$$P_2 \sim \pi_2(p_2) = \mathcal{U}(p_2|0, 1). \tag{18}$$

It represents an a priori degree of plausibility of the true value p_2, and it is reducible in the sense that Bayesian updating provides an a posteriori degree of evidence. The quantification of parametric Bayesian priors is a controversial business. Priors go beyond bare interval-like statements by assigning a relative probability structure over the set of admissible values. Thus, more generally, any prior distribution with nonzero support over the priorly admissible set Δ that vanishes elsewhere could be considered appropriate.

C. Category 3: Mixed Uncertainty

Category 3 comprises those model inputs $\boldsymbol{x} \equiv (p_1, p_4, p_5)$ that are subject to aleatory variability across experiments $i = 1, \ldots, n$. The natural variability is parametrized by hyperparameters $\boldsymbol{\theta}_X \equiv (\boldsymbol{\theta}_1, \boldsymbol{\theta}_{45})$ that are epistemically uncertain themselves. This is a mixed-type uncertainty model that, in less Bayesian contexts, is sometimes referred to as imprecise probability or distributional p-box [52,53].

1. Unimodal Beta

Model inputs $p_1 \in [0, 1]$ are distributed according to a unimodal beta distribution. Beta distributions $\mathrm{Beta}(p_{1,i}|\alpha_1, \beta_1)$ are commonly parametrized by shape hyperparameters $\alpha_1, \beta_1 > 0$. Instead, we herein parametrize the beta distribution $\mathrm{Beta}(p_{1,i}|\mu_1, \sigma_1^2)$ by its mean μ_1 and variance σ_1^2. Thus, with unknown hyperparameters $\boldsymbol{\theta}_1 \equiv (\mu_1, \sigma_1^2)$ experiment-specific realizations $p_{1,i}$ are drawn from the population distribution

$$(P_{1,i}|\boldsymbol{\theta}_1) \sim f_1(p_{1,i}|\boldsymbol{\theta}_1) = \mathrm{Beta}(p_{1,i}|\mu_1, \sigma_1^2). \tag{19}$$

To begin with, given the shape parameters (α_1, β_1), the expected value $\mu_1 = \mathbb{E}[p_1]$ and the variance $\sigma_1^2 = \mathrm{Var}[p_1]$ of the density function $\mathrm{Beta}(p_{1,i}|\mu_1, \sigma_1^2)$ are given as

$$\mu_1 = \frac{\alpha_1}{\alpha_1 + \beta_1}, \qquad \sigma_1^2 = \frac{\alpha_1 \beta_1}{(\alpha_1 + \beta_1)^2(\alpha_1 + \beta_1 + 1)}. \tag{20}$$

Conversely, given the statistical moments (μ_1, σ_1^2), the shape parameters (α_1, β_1) of the density function $\text{Beta}(p_{1,i}|\alpha_1, \beta_1)$ can be obtained by

$$\alpha_1 = \left(\frac{\sigma_1^2 + \mu_1^2 - \mu_1}{\sigma_1^2}\right)(-\mu_1), \qquad \beta_1 = \left(\frac{\sigma_1^2 + \mu_1^2 - \mu_1}{\sigma_1^2}\right)(\mu_1 - 1). \tag{21}$$

The required unimodality (i.e., the fact that the distribution features a single mode within its support) translates into $\alpha_1, \beta_1 > 1$. Moreover the problem setup requires $3/5 \le \mu_1 \le 4/5$ and $1/50 \le \sigma_1^2 \le 1/25$. To adopt this epistemic uncertainty model for the hyperparameters $\boldsymbol{\theta}_1$, we state the uniform hyperprior

$$\boldsymbol{\Theta}_1 \sim \pi_1(\boldsymbol{\theta}_1) = \mathcal{U}(\boldsymbol{\theta}_1|\mathcal{D}_{\boldsymbol{\theta}_1}), \quad \text{with} \quad \mathcal{D}_{\boldsymbol{\theta}_1} = \{(\mu_1, \sigma_1^2) \in \mathbb{R}^2 | 3/5 \le \mu_1 \le 4/5, 1/50 \le \sigma_1^2 \le 1/25, \alpha_1 > 1, \beta_1 > 1\}. \tag{22}$$

If $\lambda(\mathcal{D}_{\boldsymbol{\theta}_1})$ is the volume of the set $\mathcal{D}_{\boldsymbol{\theta}_1} \subset \mathbb{R}^2$, then the uniform density Eq. (22) is $1/\lambda(\mathcal{D}_{\boldsymbol{\theta}_1})$ on $\mathcal{D}_{\boldsymbol{\theta}_1}$ and zero elsewhere. In practice, the normalization constant $\lambda(\mathcal{D}_{\boldsymbol{\theta}_1})$ is unknown, but since priors are flat and only ratios are compared in the MH correction Eq. (15), only the set membership of MCMC proposals has to be determined.

Consequently, we can treat the prior Eq. (22) as $\pi_1(\boldsymbol{\theta}_1) = \pi(\mu_1)\pi(\sigma_1^2)$, with independent marginals $\pi(\mu_1) = \mathcal{U}(\mu_1|3/5, 4/5)$ and $\pi(\sigma_1^2) = \mathcal{U}(\sigma_1^2|1/50, 1/25)$, and reject MCMC proposals that do not respect $\alpha_1, \beta_1 > 1$ with the aid of Eq. (21). This practical prior choice is ambiguous in the sense that priors could be assumed for shape parameters (α_1, β_1), too. However, this could yield improper prior distributions. From an engineering point of view, we consider (μ_1, σ_1^2) statistically more "natural" than the shape parameters. In addition, they underlie strong prior constraints, which is advantageous to exploring the posterior by means of MCMC.

2. Correlated Gaussian

The model inputs $p_4, p_5 \in \mathbb{R}$ are modeled as possibly correlated Gaussian random variables. Across the experiments $i = 1, \ldots, n$, these model inputs take on different unknown realizations $(p_{4,i}, p_{5,i})$. This inherently aleatory variability is represented by the population distribution

$$((P_{4,i}, P_{5,i})|\boldsymbol{\theta}_{45}) \sim f_{45}((p_{4,i}, p_{5,i})|\boldsymbol{\theta}_{45}) = \mathcal{N}((p_{4,i}, p_{5,i})|\boldsymbol{\mu}_{45}, \boldsymbol{\Sigma}_{45}). \tag{23}$$

For $j = 4, 5$, the means $\mu_j = \mathbb{E}[p_j]$, variances $\sigma_j^2 = \text{Var}[p_j]$, and the coefficient of correlation $\rho_{45} = \mathbb{E}[(p_4 - \mu_4)(p_5 - \mu_5)]$ constitute the hyperparameters $\boldsymbol{\theta}_{45} \equiv (\mu_4, \sigma_4^2, \mu_5, \sigma_5^2, \rho_{45})$. Those hyperparameters are unknown constants that determine the mean $\boldsymbol{\mu}_{45}$ and the covariance matrix $\boldsymbol{\Sigma}_{45}$ of the bivariate normal density by

$$\boldsymbol{\mu}_{45} = \begin{pmatrix} \mu_4 \\ \mu_5 \end{pmatrix}, \qquad \boldsymbol{\Sigma}_{45} = \begin{pmatrix} \sigma_4^2 & \rho_{45}\sigma_4\sigma_5 \\ \rho_{45}\sigma_4\sigma_5 & \sigma_5^2 \end{pmatrix}. \tag{24}$$

Besides the natural bounds $|\rho_{45}| \le 1$, it is requested that $-5 \le \mu_j \le 5$ and $1/400 \le \sigma_j^2 \le 4$. We translate these intervals into flat and independent marginals $\pi(\mu_j)$, $\pi(\sigma_j^2)$, and $\pi(\rho_{45})$ of the common hyperprior $\pi_{45}(\boldsymbol{\theta}_{45})$ by

$$\left.\begin{aligned} \pi(\mu_j) &= \mathcal{U}(\mu_j| -5, 5), \\ \pi(\sigma_j^2) &= \mathcal{U}(\sigma_j^2|1/400, 4), \\ \pi(\rho_{45}) &= \mathcal{U}(\rho_{45}| -1, 1), \end{aligned}\right\} \boldsymbol{\Theta}_{45} \sim \pi_{45}(\boldsymbol{\theta}_{45}) = \left(\prod_{j=4}^{5} \pi(\mu_j)\pi(\sigma_j^2)\right)\pi(\rho_{45}). \tag{25}$$

Insofar as priors for spread hyperparameters could refer to standard deviations or variances alike, the ambiguity in quantifying parametric Bayesian priors becomes especially obvious for these types of hyperparameters.

D. Bayesian Problem Statement

The primary objective of the NASA UQ challenge subproblem A is the reduction of epistemic uncertainties of the true values of the forward model parameter p_2 and the hyperparameters $(\boldsymbol{\theta}_1, \boldsymbol{\theta}_{45})$ [1]. To accomplish that goal, the forward model, a set of data, and prior knowledge are available. In order to prevent from reverse-engineering of the mathematical character and numerical implementation, the forward model h_1 is distributed as a protected MATLAB p-code file, i.e., a black-box model. Available data $\langle \tilde{y}_i \rangle_{1 \le i \le 50}$ comprise $n = 50$ scalar observations $\tilde{y}_i = h_1(p_{1,i}, p_2, p_{3,i}, p_{4,i}, p_{5,i})$, which have been realized as forward model responses complying with the true uncertainty model of forward model inputs, i.e., the model parameter p_2 takes on its true value and $(\langle p_{1,i} \rangle, \langle p_{3,i} \rangle, \langle p_{4,i} \rangle, \langle p_{5,i} \rangle)$ have been randomly sampled from their true population distributions. Notwithstanding that the observations provided are perfect, in general, they might very well be subject to an additional model-measurement discrepancy, i.e., imperfect [2]. Data have been arranged into two distinct configurations of observations, $\langle \tilde{y}_i \rangle_{1 \le i \le 25}$ and $\langle \tilde{y}_i \rangle_{26 \le i \le 50}$, for which separate and joint analyses are envisaged to indicate how the number n of processed samples impacts the significance of the final results.

The available prior knowledge has been translated into parametric and structural Bayesian prior distributions. We have pointed out that this formulation endows the problem with a subjectivist interpretation of probability and suffers from the ambiguity in the chosen parametric prior and its influence on the resulting posterior. The problem statement, as well as the framework and the algorithms introduced so far, grant an ample scope of formulating and solving the problem as Bayesian inference of the QOI $(p_2, \boldsymbol{\theta}_1, \boldsymbol{\theta}_{45})$ within a multilevel context. In the first place, the Bayesian multilevel model Eq. (8), defined by parametric prior Eqs. (18), (22), and (25) and structural priors Eqs. (17), (19), and (23), establishes the natural framework for solving the original problem posed. For the sake of completeness, the devised multilevel model is summarized as

$$
\begin{aligned}
(\tilde{Y}_i|p_2,\boldsymbol{\theta}_1,\boldsymbol{\theta}_{45},\boldsymbol{\theta}_3) &\sim f(\tilde{y}_i|p_2,\boldsymbol{\theta}_1,\boldsymbol{\theta}_{45},\boldsymbol{\theta}_3),\\
P_2 &\sim \pi_2(p_2) = \mathcal{U}(p_2|0,1),\\
(P_{1,i}|\boldsymbol{\theta}_1) &\sim f_1(p_{1,i}|\boldsymbol{\theta}_1) = \mathrm{Beta}(p_{1,i}|\mu_1,\sigma_1^2),\\
((P_{4,i},P_{5,i})|\boldsymbol{\theta}_{45}) &\sim f_{45}((p_{4,i},p_{5,i})|\boldsymbol{\theta}_{45}) = \mathcal{N}((p_{4,i},p_{5,i})|\boldsymbol{\mu}_{45},\boldsymbol{\Sigma}_{45}),\\
\boldsymbol{\Theta}_1 &\sim \pi_1(\boldsymbol{\theta}_1) = \mathcal{U}(\boldsymbol{\theta}_1|\mathcal{D}_{\boldsymbol{\theta}_1}),\\
\boldsymbol{\Theta}_{45} &\sim \pi_{45}(\boldsymbol{\theta}_{45}) = \pi(\mu_4)\pi(\sigma_4^2)\pi(\mu_5)\pi(\sigma_5^2)\pi(\rho_{45}),\\
(P_{3,i}|\boldsymbol{\theta}_3) &\sim f_3(p_{3,i}|\boldsymbol{\theta}_3) = \mathcal{U}(p_{3,i}|0,1).
\end{aligned}
\tag{26}
$$

The corresponding posterior distribution Eq. (10) of the QOI $(p_2,\boldsymbol{\theta}_1,\boldsymbol{\theta}_{45})$ follows the Bayesian data analysis of the given forward model responses $\langle\tilde{y}_i\rangle$, i.e., realizations of random variables $(\tilde{Y}_i|p_2,\boldsymbol{\theta}_1,\boldsymbol{\theta}_{45},\boldsymbol{\theta}_3)$.

In the second place, one could solve the inverse problem posed in the presence of additional measurement noise. To that end, the Bayesian multilevel model Eq. (2) establishes the proper framework. Synthetic and noisy observations $y_i = \tilde{y}_i + \varepsilon_i$ could be obtained by perturbing the given model responses $\tilde{y}_i$ with residuals ε_i that are randomly sampled from prescribed distributions $f_{E_i}(\varepsilon_i) = \mathcal{N}(\varepsilon_i|0,\sigma_i^2)$. Parameters of the residual model (i.e., the residual variances σ_i^2) could either be treated as knowns or as further unknowns. By analyzing imperfect data $\langle y_i\rangle$ [i.e., realizations of random variables $(Y_i|p_{1,i},p_2,p_{3,i},p_{4,i},p_{5,i})$] and treating latent variables $(\langle p_{1,i}\rangle,\langle p_{3,i}\rangle,\langle p_{4,i}\rangle,\langle p_{5,i}\rangle)$ as nuisance, inference of the QOI would be based on the posterior Eq. (5). A DAG of the Bayesian multilevel model corresponding to our challenge problem interpretation with perfect and imperfect data, respectively, is depicted in Fig. 2.

VI. Bayesian Data Analysis

We will now apply the inferential machinery of multilevel calibration for solving the Bayesian interpretation of the uncertainty characterization subproblem A of the NASA Langley Multidisciplinary UQ Challenge. The problem will be solved in its original formulation involving perfect data. Motivated by findings from the first preliminary problem analyses, posterior densities of the QOI will be computed by a suitable MH independence sampler. This sampler will be implemented in MATLAB and serially run on a modern CPU. Nevertheless, we will discuss possible parallelization strategies. The total data $\langle\tilde{y}_i\rangle_{1\le i\le 50}$ and their subconfigurations $\langle\tilde{y}_i\rangle_{1\le i\le 25}$ and $\langle\tilde{y}_i\rangle_{26\le i\le 50}$ will be analyzed with the devised algorithm. Based on heuristic parameter tuning and plausibility checks, we will assess the fidelity of the posterior. Promising a boost of posterior fidelity, we will lastly devise a hybrid MCMC scheme that is based on data augmentation and both independence and random walk sampling.

A. Preliminary Analyses

A basic understanding of an inverse problem under consideration allows us to judge the performance of various potential MCMC schemes. This allows us to design efficient algorithms, and it is indispensable because it prevents us from obtaining misleading results that are due to inappropriate samplers. To gain first insights into the present multilevel calibration problem, we perform a number of initial MCMC runs that are based on a crude random walk Metropolis sampling. Thereby, we could provisionally assess the principal nature of the posterior Eqs. (4) and (10). The main findings from sampling the posterior Eq. (10) indicate that posterior marginals of the QOI $(p_2,\boldsymbol{\theta}_1,\boldsymbol{\theta}_{45})$ may very well be multimodal or broad distributions that significantly overlap with the marginal priors.

Solving a joint problem in the presence of additional measurement noise provides further insight. Notwithstanding that, this is actually a different problem, it will eventually prove valuable. Sampling the joint posterior Eq. (4) of the entirety of unknowns $(\langle p_{1,i}\rangle, p_2, \langle p_{3,i}\rangle, \langle p_{4,i}\rangle, \langle p_{5,i}\rangle,\boldsymbol{\theta}_1,\boldsymbol{\theta}_{45})$ reveals further information about the unknowns, e.g., we find that experiment-specific unknowns $\langle p_{1,i}\rangle$ are identifiable and their posterior marginals feature single modes. This does not refer to the posteriors of $\langle p_{3,i}\rangle$ and $\langle p_{4,i}\rangle$ that are rather flat and the ones of $\langle p_{5,i}\rangle$ that are bimodal. Altogether, those preliminary analyses have provided useful information that will eventually motivate the final MCMC samplers.

B. Perfect Data Analysis

For the calibration of the Bayesian multilevel model Eq. (8), we devise a blockwise independence MCMC sampler. Since the algorithm is based on MCMC, MC, and KS techniques, hereinafter, it will be referred to as MC^{3}KS. QOI are grouped in blocks (p_2), (μ_1,σ_1^2), $(\mu_4,\sigma_4^2,\rho_{45})$, and (μ_5,σ_5^2) that are consecutively updated by sampling blockwise candidates from the corresponding prior distributions. In many cases, independence sampling from the priors is inefficient due to a negligible overlap between the prior and the posterior distributions and the resulting low acceptance rates. However, on account of the multimodality of the posteriors and their overlap with the priors, which were indicated by first analyses, independence sampling promises rapid mixing for the problem at hand.

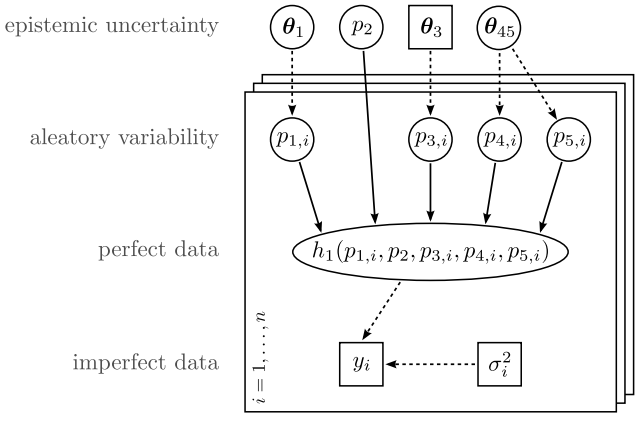

Fig. 2 DAG of the NASA UQ Challenge subproblem A.

Moreover, in the context of Eq. (16), we suppose that wide jumps in the parameter space, which are induced by independence sampling on average, are beneficial in terms of posterior fidelity. For wide jumps, the difference of the likelihood at the current and the candidate state of the Markov chain tends to be larger than for small jumps. Following previous discussions, to some extent, this alleviates the error statistics of repeated likelihood estimations for the same state. Another advantage of the devised MCMC scheme over random walk sampling is that, apart from its blockwise updating structure, it does not require extensive fine tuning of the proposal distribution. Updating in blocks intends to minimize the number of calls to the likelihood Eq. (14) that are necessary for each block in each MCMC iteration while maintaining high acceptance rates that are favorable for independence sampling. With the help of Eq. (21), the constraints $\alpha_1, \beta_1 > 1$ are enforced by rejecting nonconforming proposals in the block (μ_1, σ_1^2). The MC^{3}KS sampler is initialized by setting parameters in the middle of their admissible intervals. Due to rapid mixing, the initialization is not of crucial importance for the employed sampling scheme. Generally speaking, we expect that forward model parameters and mean hyperparameters are easier to identify than spread or even correlation hyperparameters.

C. Likelihood Estimation and Posterior Fidelity

For the estimation Eq. (14) of the transformed likelihood Eq. (9), we choose kernel functions $\mathcal{K}$ of Gaussian type. To achieve a convenient tradeoff between the conflicting endeavors fidelity of the posterior and the ease of its computation, the number of samples K and the bandwidth h have to be set. In practice, computational resource limitations restrict the total number of affordable forward model runs; hence, we approach parameter tuning from the situation of a given K.

Owing to the absence of a rigorous means to define a corresponding and optimal bandwidth h, we study the posteriors obtained for fixed $K = 10^4$ and decreasing h in a cascade of runs. We observe an initial shrinkage of the posterior (i.e., evolving from the flat prior, it takes on definite shape) and an eventual collapse (i.e., the posterior flattens out again and loses its structure). The initial shrinkage is associated with significant changes of the posterior shape, the eventual breakdown is QOI dependent, and, in between, the posterior is relatively stable with respect to h. We remark that this behavior is consistent with Eqs. (15) and (16). Significant oversmoothing of the target density Eq. (11) [i.e., strongly biased estimator Eq. (14)] can falsely assign posterior mass to QOI values that do not well explain or even contradict the data. Considerable undersmoothing of the target density [i.e., a high variance of the estimator Eq. (14)] can cause "arbitrary" acceptances in the MH correction. We speculate that, in between those extremes, the more stable the posterior is with respect to small changes in h, the more confident we can be to have revealed the true posterior. Beyond that, we presume that a high degree of distinctiveness of the posterior with respect to the prior indicates high posterior fidelity. The converse statement does not hold though.

In addition to those heuristics, we perform a plausibility check as follows. During preliminary analysis, we have solved the UQ Challenge Problem in the presence of additional measurement noise ε_i with $E_i \sim \mathcal{N}(\varepsilon_i|0, \sigma_i^2)$, i.e., we sampled a joint posterior of the form Eq. (4). If the corresponding noise level σ_i^2 tends to zero, the results of analyzing imperfect data should approach the ones of analyzing perfect data. Indeed, we find that, for low levels of noise $\sigma_i^2 \gtrsim 0$, the joint problem solution resembles our final results for analyzing perfect data. Although the posterior Eq. (10) can only be approximately explored with the dubious aid of statistical likelihood estimations Eq. (14), the joint posterior Eq. (4) can be sampled exactly. Thus, we have found that our approximate solution to the actual problem reminds us of an exact solution to an only slightly different problem. For "well-behaved" problems, we take this observation as an indication of an acceptable degree of posterior fidelity.

Following this discussion, $K = 10^4$ and $h = 0.002$ constitute our final parameter setup. The principle of estimating the density Eq. (11) and the transformed likelihood Eq. (9) is visualized in Fig. 3. Samples of $K = 10^4$ and $K = 10^7$ forward model responses are simulated for two different (hyper)parameter values $(p_2, \boldsymbol{\theta}_1, \boldsymbol{\theta}_{45})_{\text{high}}$ and $(p_2, \boldsymbol{\theta}_1, \boldsymbol{\theta}_{45})_{\text{low}}$. As judged from our final results, these are (hyper)parameter values of high and low degrees of posterior evidence, respectively. For the smaller sample with $K = 10^4$, estimates of the sought densities $f(\tilde{y}_i|p_2, \boldsymbol{\theta}_1, \boldsymbol{\theta}_{45}, \boldsymbol{\theta}_3)$ are shown. For reference purposes, a histogram of the larger sample with $K = 10^7$ is shown. It can be seen that response densities $f(\tilde{y}_i|p_2, \boldsymbol{\theta}_1, \boldsymbol{\theta}_{45}, \boldsymbol{\theta}_3)$ for $(p_2, \boldsymbol{\theta}_1, \boldsymbol{\theta}_{45})_{\text{high}}$ and $(p_2, \boldsymbol{\theta}_1, \boldsymbol{\theta}_{45})_{\text{low}}$ significantly overlap. This is a problem characteristic that complicates the statistical identification of the QOI $(p_2, \boldsymbol{\theta}_1, \boldsymbol{\theta}_{45})$.

It can also be seen that the employed bandwidth $h = 0.002$ amounts to a slight undersmoothing of the target density, i.e., a bias-variance tradeoff favoring lower bias yet acceptable variance. This is advantageous because it allows us to capture local small-scale features of the target density (e.g., sharp peaks and edges) in the posterior. We remark that, in the context of the pseudo-marginal MCMC [51], this observation supports the speculation that it is preferable to minimize the bias in Eq. (13). Since the target density significantly differs from a normal distribution, automatic bandwidth selection cannot be based on the normal reference rule. The resulting oversmoothing of the target density (i.e., a significantly biased KDE) would veil its important characteristics. Finally, the (hyper)parameter values $(p_2, \boldsymbol{\theta}_1, \boldsymbol{\theta}_{45})_{\text{high}}$ can be seen to lead to a response density that explains the data sample $\langle \tilde{y}_i \rangle$ reasonably well.

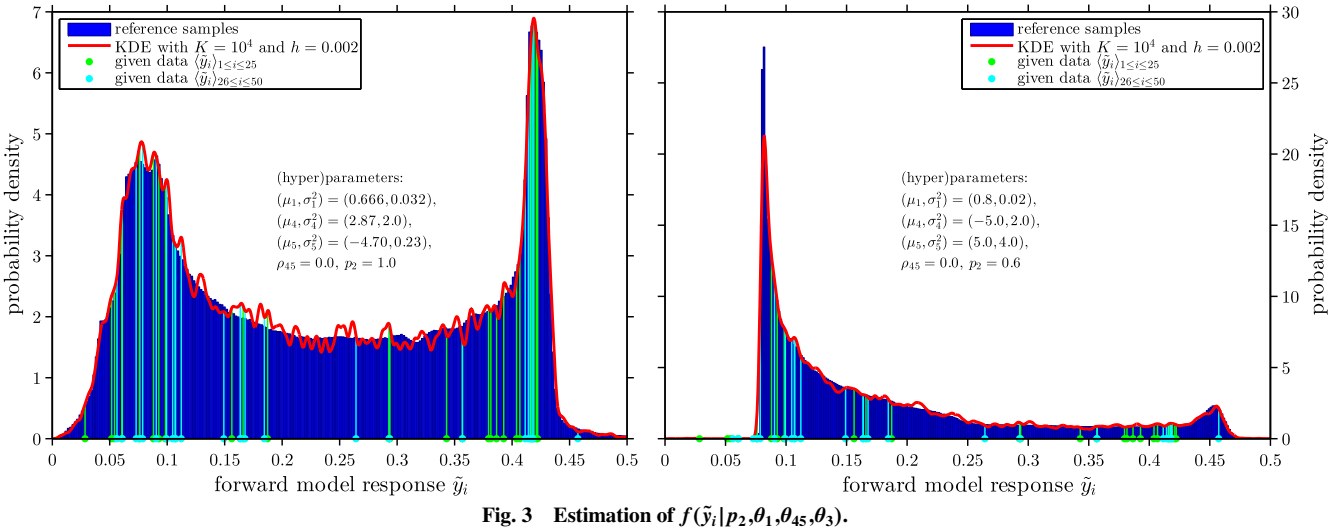

Fig. 3 Estimation of $f(\tilde{y}_i|p_2, \boldsymbol{\theta}_1, \boldsymbol{\theta}_{45}, \boldsymbol{\theta}_3)$.

D. Final Results

First of all, we jointly analyze the total data $\langle \tilde{y}_i \rangle_{1 \le i \le 50}$. For $N = 10^5$ iterations of the MC^{3}KS algorithm, the total program runtime amounts to $t \approx 30$h on a single core. Blockwise acceptance rates were found to be approximately 20% for (p_2), 40% for (μ_1, σ_1^2), 60% for $(\mu_4, \sigma_4^2, \rho_{45})$, and 10% for (μ_5, σ_5^2). With Eq. (20), a number of 10,327 blockwise proposals (μ_1, σ_1^2) had been rejected because of violation of the prior requirement $\alpha_1, \beta_1 > 1$. Marginal posterior densities of the QOI are shown in Figs. 4–7. Based on appropriate boundary correction methods, the densities shown have been obtained by kernel smoothing of the MCMC posterior samples. Following precursory discussions, we attribute an acceptable

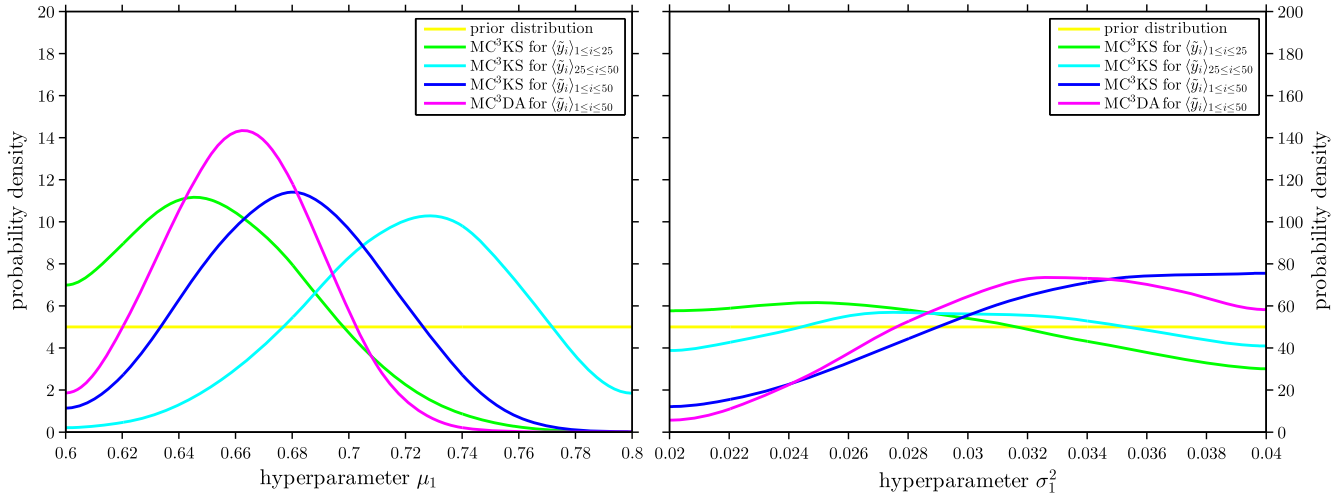

Fig. 4 Posterior marginals of μ_1 and σ_1^2.

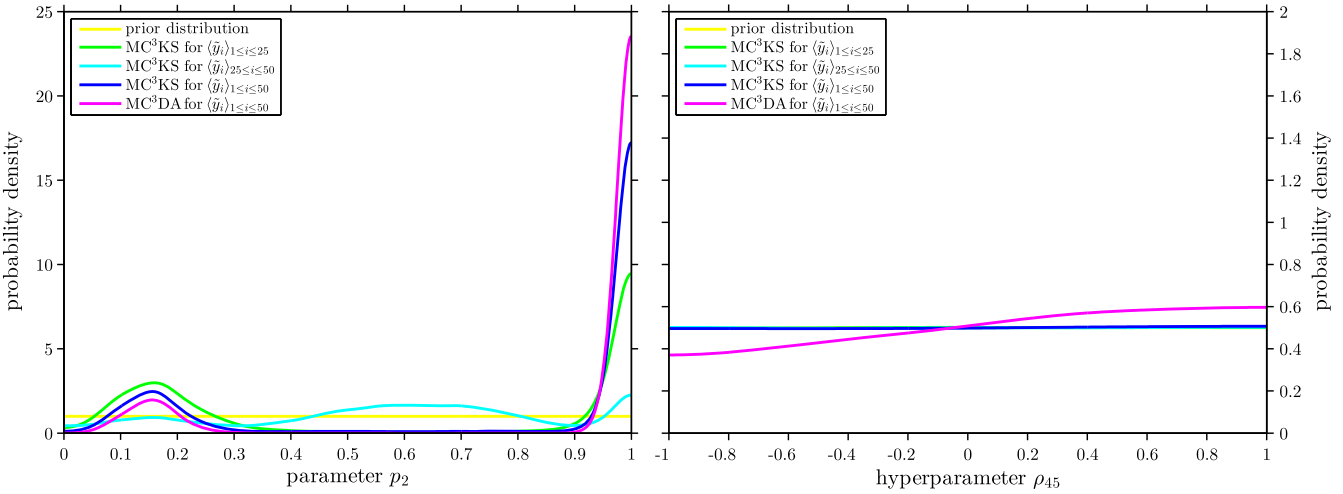

Fig. 5 Posterior marginals of p_2 and ρ_{45}.

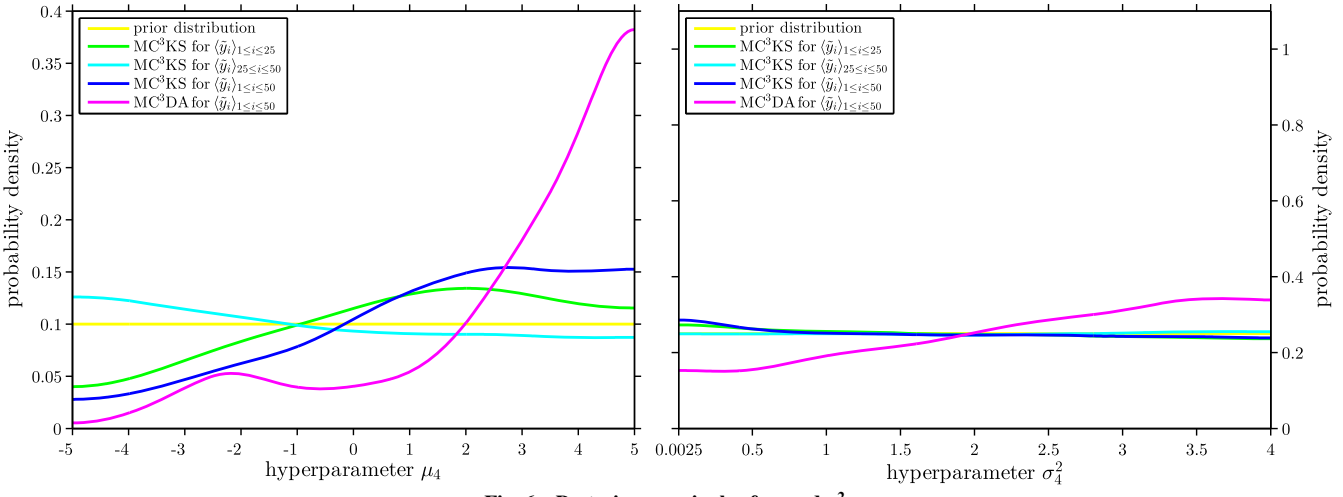

Fig. 6 Posterior marginals of μ_4 and σ_4^2.

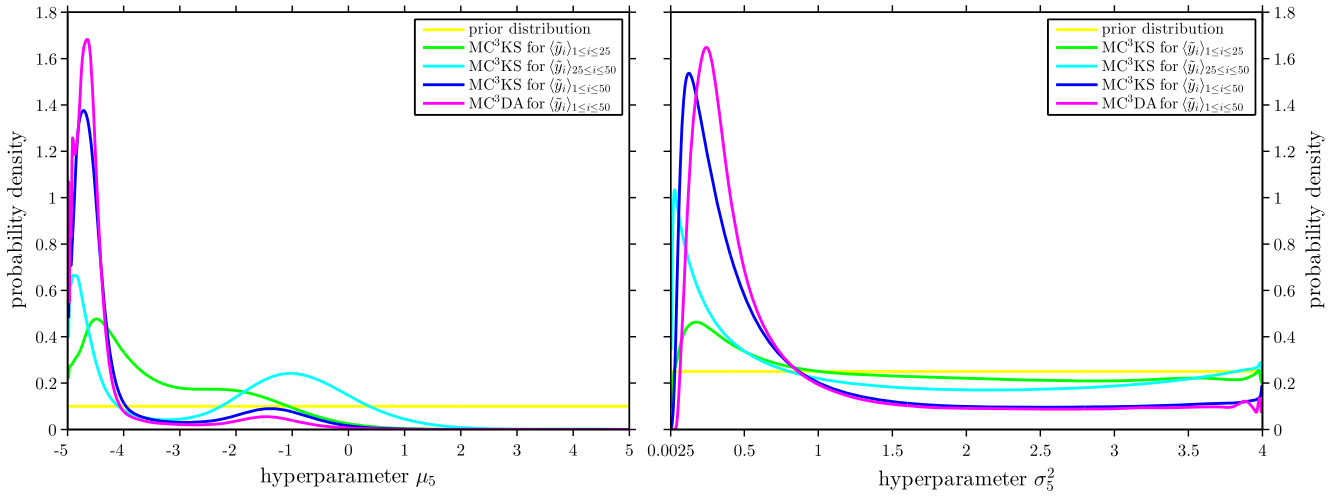

Fig. 7 Posterior marginals of μ_5 and σ_5^2.

degree of fidelity to the posteriors obtained. We are confident that we have revealed a "good" approximation of the true posteriors, regardless of whether some of them are flat and only weakly informative.

We also analyze the data subconfigurations $\langle\tilde{y}_i\rangle_{1\le i\le 25}$ and $\langle\tilde{y}_i\rangle_{26\le i\le 50}$ separately. The posterior densities produced by separate analyses may differ considerably. With respect to the posteriors yielded by analyzing $\langle\tilde{y}_i\rangle_{1\le i\le 50}$, the two data subconfigurations are representative to a different degree. Those findings indicate that $n = 25$ is a comparably low number of observations, whereas $n = 50$ is moderately satisfying for the Bayesian calibration of mean hyperparameters and the forward model parameter. Properly identifying the variance and correlation hyperparameters would require a higher number of observations. This is hardly surprising regarding the complex uncertainty setup, the number of unknowns, the unknown character of the forward model, and the inverse nature of the calibration problem.

At this point, it is important to mention that multilevel model calibration shares and combines aspects of classical inverse problems (i.e., the inference of an unknown constant forward model parameter) and direct sample statistics (i.e., fitting a parametric distribution to a random data sample). Thereby, Bayesian multilevel model calibration also inherits the usual difficulties inherent in inversion and distribution fitting.

The marginal posteriors of μ_1 and σ_1^2 are shown in Fig. 4. In comparison to the prior, the posterior of μ_1 shows a pronounced structure. Separately analyzing $\langle\tilde{y}_i\rangle_{1\le i\le 25}$ and $\langle\tilde{y}_i\rangle_{26\le i\le 50}$ gives rise to two different posterior modes. Those are suggested by the corresponding experiment-specific realizations $\langle p_{1,i}\rangle_{1\le i\le 25}$ and $\langle p_{1,i}\rangle_{26\le i\le 50}$. A joint analysis of $\langle\tilde{y}_i\rangle_{1\le i\le 50}$ leads to a mode that lies in between the two aforementioned ones. The posterior marginal of the spread hyperparameter σ_1^2 is comparably structureless, and therefore less informative. In Fig. 5, the posterior marginals of p_2 and ρ_{45} are depicted. Due to the fact that data subconfigurations $\langle\tilde{y}_i\rangle_{1\le i\le 25}$ and $\langle\tilde{y}_i\rangle_{26\le i\le 50}$ are, to a different degree, informative about further unknowns (e.g., about (μ_1, σ_2^2), which was discussed previously), the posteriors obtained for the constant model parameter p_2 may deviate as well. The analysis of $\langle\tilde{y}_i\rangle_{1\le i\le 50}$ by MC³KS reveals two clear and separated posterior modes in the posterior of the forward model parameter p_2, whereas the posterior of the correlation hyperparameter ρ_{45} is flat. This is according to our previous expectations. Posteriors of μ_4 and σ_4^2 are given in Fig. 6. That they are comparably flat and uninformative prevents us from drawing clear inferential conclusions. This statement does not hold for the posteriors of μ_5 and σ_5^2 that can be seen in Fig. 7. Although the former features a bimodal structure and drops to zero for higher values of μ_5, the latter is unimodal and reaches a nonzero "plateau" for higher values of σ_5^2. Since this region accumulates considerable posterior probability mass, one cannot exclude those values of σ_5^2. Apart from the results of MC³KS, Figs. 4–7 also contain the results of analyzing $\langle\tilde{y}_i\rangle_{1\le i\le 50}$ by MC³DA, i.e., an alternative algorithm that is based on data augmentation. This technique will be further detailed in Sec. VII.

Note that Bayesian probabilities feature a richer structure than mere epistemic intervals. Conforming with a subjective Bayesian paradigm, probabilities are identified as relative degrees of belief or plausibility. Thus, multivariate probability distributions, which may contain complex dependency structures that are not entirely defined by their marginals only, have to be interpreted accordingly. The marginal densities shown hide these possibly existing posterior correlations. We provide a selection of two-dimensional posterior projections for analyzing $\langle\tilde{y}_i\rangle_{1\le i\le 50}$ by MC³KS in Figs. 8 and 9.

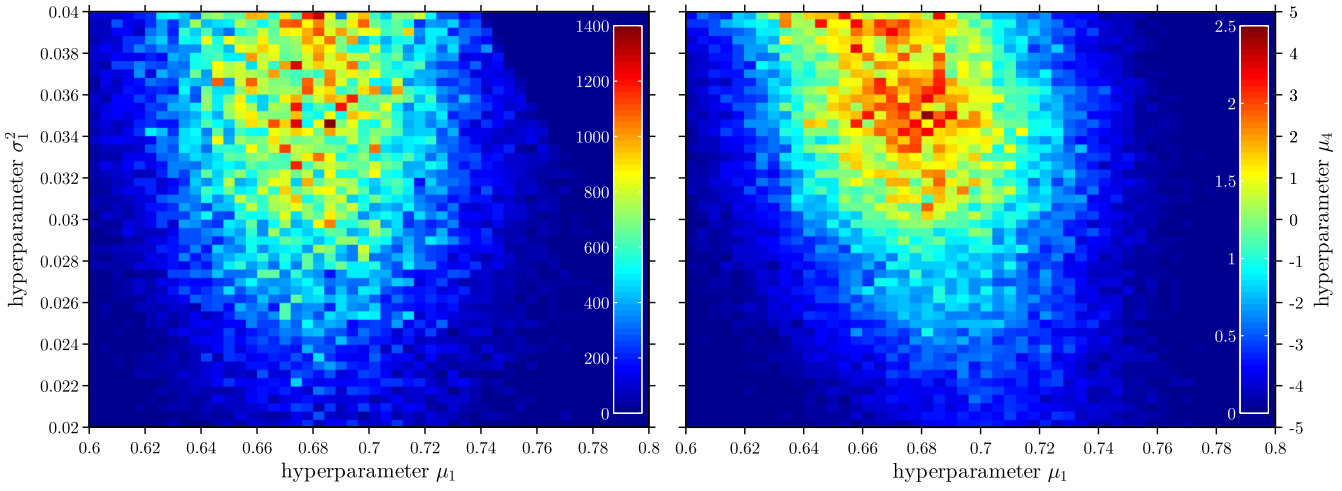

Fig. 8 Two-dimensional posteriors of (μ_1, σ_1^2) and (μ_1, μ_4).

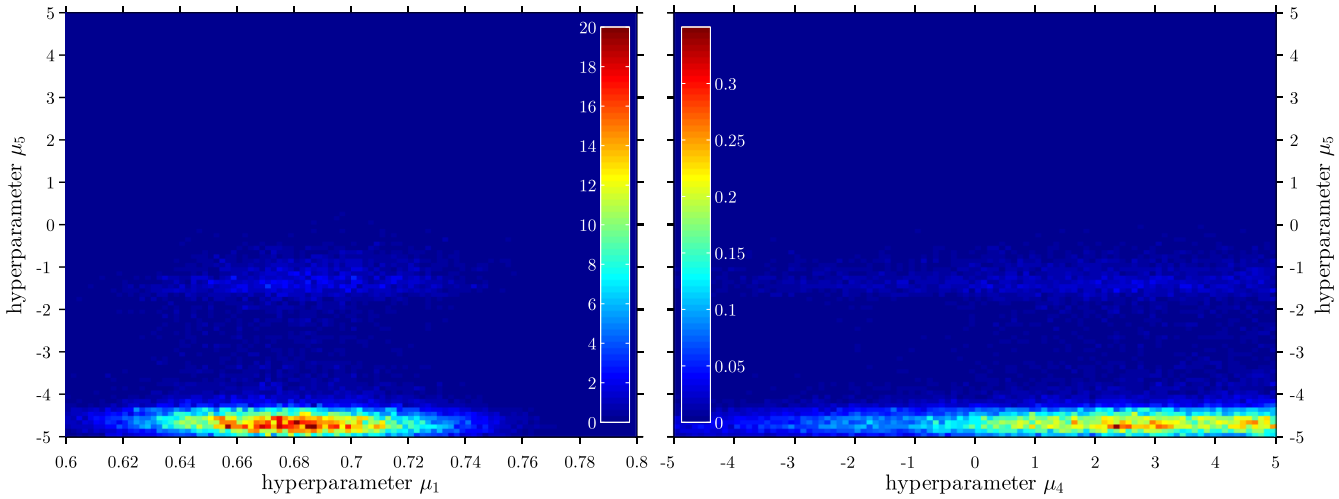

Fig. 9 Two-dimensional posteriors of (μ_1, μ_5) and (μ_4, μ_5).

Parameters that were assumed to be statistically independent a priori (e.g., the parameter p_2, the hyperparameters $\boldsymbol{\theta}_1$, and the hyperparameters $\boldsymbol{\theta}_{45}$) can be statistically dependent a posteriori. Small negative correlations in the posteriors of (μ_1, σ_1^2) and (μ_1, μ_4), shown in Fig. 8, with linear Pearson coefficients of correlation $r_{\mu_1,\sigma_1^2} = -0.08$ and $r_{\mu_1,\mu_4} = -0.22$, were discovered, respectively. In Fig. 9 the two-dimensional posteriors of (μ_1, μ_5) and (μ_4, μ_5) are shown. Those posteriors do not feature any marked dependency structure. To provide final results of interval-like character, one could define suitable Bayesian credible intervals or sets that accumulate a certain proportion (e.g., 95%) of the total posterior mass. However, the definition of such intervals is ambiguous and would still bear the probabilistic interpretation; therefore, we refrain from defining Bayesian credible intervals.

E. First Conclusion

With the proposed MC^{3}KS algorithm, the Bayesian formulation of the challenge problem could be solved. The numerical efficiency of the MCMC sampling scheme, which was based on independently sampling from the priors, could be easily increased. Obtained posteriors could be approximated by suitable distributions that are easy to sample. Using these as proposal distributions would lead to higher acceptance rates and better mixing properties. Most Bayesian computations can only be parallelized by running several Markov chains simultaneously. An obvious parallelization strategy for the devised algorithm is to parallelize the estimation of the transformed likelihood on the level of forward model runs. This also suggests the possibility of studying the posterior for significantly larger K and smaller h. Moreover, different classes of kernel functions $\mathcal{K}$ (e.g., with bounded nonzero support) or more advanced KDE techniques (e.g., locally adaptive schemes or other bias reduction and correction methods) could be employed. The major shortcoming of the approach was the dependency of the final results on free algorithmic tuning parameters. Parameter tuning had to be based on heuristic criteria and plausibility checks, and the fidelity of the final posteriors could only be provisionally assessed. In the following section, we will therefore propose a complementary multilevel approach that aims at enhancing the level of posterior fidelity.

VII. Partial Data Augmentation

As a potential improvement over the employed MC^{3}KS sampler, we will devise a new hybrid MCMC sampling scheme. Since the scheme will be based on data augmentation (DA), it will henceforth be referred to as MC^{3}DA. Traditionally, DA can be a powerful tool for enhancing the computational efficiency of MCMC posterior sampling [54–56]. Instead, we will herein use DA as a means to reformulate the multilevel model calibration problem in such a complementary way that it allows for more adequate likelihood estimations in view of Eq. (13). In turn, this promises an enhancement of the posterior fidelity through Eq. (15). The approach will also allow for automatic kernel bandwidth selection based on a classical yet well-approved criterion, namely, the normal reference rule [35]. This is appealing, since it avoids the cumbersome procedure of tuning free algorithmic parameters of the KDE that was described in Sec. VI.C.

Rather than directly sampling the posterior of the QOI $(p_2, \boldsymbol{\theta}_1, \boldsymbol{\theta}_{45})$, one can sample the posterior of an augmented number of unknowns $(\langle p_{1,i}\rangle, p_2, \boldsymbol{\theta}_1, \boldsymbol{\theta}_{45})$ and obtain the posterior of the QOI by subsequently marginalizing over nuisance $\langle p_{1,i}\rangle$. Presuming that sampling from $\pi(\langle p_{1,i}\rangle, p_2, \boldsymbol{\theta}_1, \boldsymbol{\theta}_{45}|\langle\tilde{y}_i\rangle, \boldsymbol{\theta}_3)$ is "easier" to accomplish than straightforwardly sampling from $\pi(p_2, \boldsymbol{\theta}_1, \boldsymbol{\theta}_{45}|\langle\tilde{y}_i\rangle, \boldsymbol{\theta}_3)$, a de facto improvement is achieved. The introduction of $\langle p_{1,i}\rangle$ as auxiliary variables is a partial form of data augmentation. As indicated by preliminary problem analyses, the forward model h_1 seems to be dependent on its input p_1 in such a strong way that the data $\langle\tilde{y}_i\rangle$ can be inverted for the unknown $\langle p_{1,i}\rangle$, under uncertainty of the remaining unknowns. Even though $\langle p_{1,i}\rangle$ are not QOI, this provides additional insight into the inverse problem posed. Moreover, the likelihood function corresponding to partial data augmentation can be estimated more adequately. Presumably, within a feasible computation time, the aforementioned facts will allow us to sample $\pi(\langle p_{1,i}\rangle, p_2, \boldsymbol{\theta}_1, \boldsymbol{\theta}_{45}|\langle\tilde{y}_i\rangle, \boldsymbol{\theta}_3)$ with higher fidelity than sampling $\pi(p_2, \boldsymbol{\theta}_1, \boldsymbol{\theta}_{45}|\langle\tilde{y}_i\rangle, \boldsymbol{\theta}_3)$. We will introduce the formalism of partial data augmentation next.

A. Augmented Multilevel Model

If the unknowns $(p_2, \boldsymbol{\theta}_1, \boldsymbol{\theta}_{45})$ of the Bayesian multilevel model Eq. (8) are augmented by experiment-specific realizations $\langle p_{1,i}\rangle$, then the collective of unknowns $(\langle p_{1,i}\rangle, p_2, \boldsymbol{\theta}_1, \boldsymbol{\theta}_{45})$ has to be explicitly taken into account. The associated Bayesian prior is given as

$$\pi(\langle p_{1,i}\rangle, p_2, \boldsymbol{\theta}_1, \boldsymbol{\theta}_{45}) = \left(\prod_{i=1}^{n} f_1(p_{1,i}|\boldsymbol{\theta}_1)\right)\pi_2(p_2)\pi_1(\boldsymbol{\theta}_1)\pi_{45}(\boldsymbol{\theta}_{45}). \tag{27}$$

This distribution comprises both parametric and structural prior knowledge. Given an appropriate probability model $f(\tilde{y}_i|p_{1,i}, p_2, \boldsymbol{\theta}_{45}, \boldsymbol{\theta}_3)$ of random variables $(\tilde{Y}_i|p_{1,i}, p_2, \boldsymbol{\theta}_{45}, \boldsymbol{\theta}_3)$, the corresponding augmented likelihood follows as

$$\mathcal{L}(\langle\tilde{y}_i\rangle|\langle p_{1,i}\rangle, p_2, \boldsymbol{\theta}_{45}, \boldsymbol{\theta}_3) = \prod_{i=1}^{n} f(\tilde{y}_i|p_{1,i}, p_2, \boldsymbol{\theta}_{45}, \boldsymbol{\theta}_3). \tag{28}$$

The definition of the augmented likelihood Eq. (28) rests upon a probability model $(\tilde{Y}_i|p_{1,i}, p_2, \boldsymbol{\theta}_{45}, \boldsymbol{\theta}_3) \sim f(\tilde{y}_i|p_{1,i}, p_2, \boldsymbol{\theta}_{45}, \boldsymbol{\theta}_3)$. Analogous to the propagated uncertainty in Eq. (11), such a model is established through a transformed random variable $\mathcal{M}_{p_{1,i},p_2}(P_3, P_4, P_5)$ with a density function

$$f(\tilde{y}_i|p_{1,i}, p_2, \boldsymbol{\theta}_{45}, \boldsymbol{\theta}_3) = \int_0^1 \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} \delta(\tilde{y}_i - \mathcal{M}_{p_{1,i},p_2}(p_3, p_4, p_5)) f_3(p_3|\boldsymbol{\theta}_3) f_{45}((p_4, p_5)|\boldsymbol{\theta}_{45})\, \mathrm{d}p_3\, \mathrm{d}p_4\, \mathrm{d}p_5. \tag{29}$$

Here, δ denotes the Dirac delta function and $\mathcal{M}_{p_{1,i},p_2}\colon (p_3, p_4, p_5) \mapsto \mathcal{M}(p_{1,i}, p_2, p_3, p_4, p_5)$ formalizes the map that the forward model $\mathcal{M} \equiv h_1$ defines for fixed inputs $(p_{1,i}, p_2)$ and functional arguments (p_3, p_4, p_5). With the combined parametric and the structural prior Eq. (27) and augmented likelihood Eq. (28), the augmented posterior of the unknowns $(\langle p_{1,i}\rangle, p_2, \boldsymbol{\theta}_1, \boldsymbol{\theta}_{45})$ is, according to Bayes' law, proportional to

$$\pi(\langle p_{1,i}\rangle, p_2, \boldsymbol{\theta}_1, \boldsymbol{\theta}_{45}|\langle\tilde{y}_i\rangle, \boldsymbol{\theta}_3) \propto \mathcal{L}(\langle\tilde{y}_i\rangle|\langle p_{1,i}\rangle, p_2, \boldsymbol{\theta}_{45}, \boldsymbol{\theta}_3)\pi(\langle p_{1,i}\rangle, p_2, \boldsymbol{\theta}_1, \boldsymbol{\theta}_{45}). \tag{30}$$

Since we are not interested in inferring experiment-specific realizations $\langle p_{1,i}\rangle$ per se, they are treated as nuisance. Similar to the marginalization Eq. (5), the posterior of the QOI $(p_2, \boldsymbol{\theta}_1, \boldsymbol{\theta}_{45})$ is thus found by integrating the posterior Eq. (30) as follows:

$$\pi(p_2, \boldsymbol{\theta}_1, \boldsymbol{\theta}_{45}|\langle\tilde{y}_i\rangle, \boldsymbol{\theta}_3) = \int_0^1 \dots \int_0^1 \pi(\langle p_{1,i}\rangle, p_2, \boldsymbol{\theta}_1, \boldsymbol{\theta}_{45}|\langle\tilde{y}_i\rangle, \boldsymbol{\theta}_3)\, \mathrm{d}\langle p_{1,i}\rangle, \tag{31}$$

where $\mathrm{d}\langle p_{1,i}\rangle = \mathrm{d}p_{1,1} \dots \mathrm{d}p_{1,n}$ is as before. Both of the distributions Eqs. (10) and (31) equivalently define the desired posterior $\pi(\boldsymbol{m}, \boldsymbol{\theta}_X|\langle\tilde{\boldsymbol{y}}_i\rangle, \boldsymbol{\theta}_Z) \equiv \pi(p_2, \boldsymbol{\theta}_1, \boldsymbol{\theta}_{45}|\langle\tilde{y}_i\rangle, \boldsymbol{\theta}_3)$. Although Eq. (10) straightforwardly conditions via $\pi(p_2, \boldsymbol{\theta}_1, \boldsymbol{\theta}_{45}|\langle\tilde{y}_i\rangle, \boldsymbol{\theta}_3) \propto \mathcal{L}(\langle\tilde{y}_i\rangle|p_2, \boldsymbol{\theta}_1, \boldsymbol{\theta}_{45}, \boldsymbol{\theta}_3)\pi(p_2, \boldsymbol{\theta}_1, \boldsymbol{\theta}_{45})$, Eqs. (30) and (31) provide a rearrangement where nuisance variables $\langle p_{1,i}\rangle$ are first factored out before they are eventually eliminated.

Even though this scheme of data augmentation leads to a higher-dimensional estimation problem, it may be computationally advantageous. In practice, the marginal Eq. (31) can be computed by sampling the joint distribution Eq. (30) and simply discarding samples of $\langle p_{1,i}\rangle$, i.e., the multidimensional integral does not have to be calculated explicitly. Writing the compound distribution

$$f(\tilde{y}_i|p_2, \boldsymbol{\theta}_1, \boldsymbol{\theta}_{45}, \boldsymbol{\theta}_3) = \int_0^1 f(\tilde{y}_i|p_{1,i}, p_2, \boldsymbol{\theta}_{45}, \boldsymbol{\theta}_3) f(p_{1,i}|\boldsymbol{\theta}_1)\, \mathrm{d}p_{1,i} \tag{32}$$

establishes the connection between the transformed likelihood $\mathcal{L}(\langle\tilde{y}_i\rangle|p_2, \boldsymbol{\theta}_1, \boldsymbol{\theta}_{45}, \boldsymbol{\theta}_3)$ of the form Eq. (9) and the augmented likelihood $\mathcal{L}(\langle\tilde{y}_i\rangle|\langle p_{1,i}\rangle, p_2, \boldsymbol{\theta}_{45}, \boldsymbol{\theta}_3)$ in Eq. (28). The relation Eq. (32) suggests that $f(\tilde{y}_i|p_{1,i}, p_2, \boldsymbol{\theta}_{45}, \boldsymbol{\theta}_3)$ could be a "simplified" version of the "complex" distribution $f(\tilde{y}_i|p_2, \boldsymbol{\theta}_1, \boldsymbol{\theta}_{45}, \boldsymbol{\theta}_3)$ that was exemplarily shown in Fig. 3. Ideally, it would be a unimodal distribution that resembles a Gaussian. Indeed, in this case, the augmented likelihood could be estimated more adequately than the transformed one. Consequently, the QOI marginal of the induced limiting distribution of the MC^{3}DA scheme can be expected to be closer to the true posterior than the long-run distribution of the MC^{3}KS approach.

An augmented model can be analogously defined when latent variables other than $\langle p_{1,i}\rangle$ are introduced as auxiliary variables. The more variables that are introduced, the smaller the variance of the target density similar to Eq. (29) is expected to be. In the case that all variables $(\langle p_{1,i}\rangle, \langle p_{3,i}\rangle, \langle p_{4,i}\rangle, \langle p_{5,i}\rangle)$ are jointly introduced, the conditional distribution of the data would even shrink to a Dirac delta $f(\tilde{y}_i|p_{1,i}, p_2, p_{3,i}, p_{4,i}, p_{5,i}) = \delta(\tilde{y}_i - \mathcal{M}_{p_{1,i},p_2,p_{3,i},p_{4,i},p_{5,i}})$. Note that, if imperfect data $y_i = \tilde{y}_i + \varepsilon_i$ were involved, a proper distribution would still be defined as $f(y_i|p_{1,i}, p_2, p_{3,i}, p_{4,i}, p_{5,i}) = f_{E_i}(y_i - \mathcal{M}_{p_{1,i},p_2,p_{3,i},p_{4,i},p_{5,i}})$. In fact, this exactly defines the distribution Eq. (2a) of the joint problem in Eq. (2). The motivation for augmenting with $\langle p_{1,i}\rangle$ is the expectation that this provides a convenient tradeoff between fidelity and feasibility, i.e., likelihood evaluations are optimally facilitated with respect to the implied increase in dimensionality.

B. Augmented Likelihood Estimation

In practical terms, the augmented likelihood Eq. (28) can be estimated analogously to Eq. (14). To that end, the response density $f(\tilde{y}_i|p_{1,i}, p_2, \boldsymbol{\theta}_{45}, \boldsymbol{\theta}_3)$ is estimated for each $p_{1,i}$ and evaluated for the given responses $\tilde{y}_i$. Hence, a KDE-based estimate of the augmented likelihood as a function of $(\langle p_{1,i}\rangle, p_2, \boldsymbol{\theta}_{45})$ is given as

$$\hat{\mathcal{L}}_{\mathrm{DA}}(\langle\tilde{y}_i\rangle|\langle p_{1,i}\rangle, p_2, \boldsymbol{\theta}_{45}, \boldsymbol{\theta}_3) = \prod_{i=1}^{n} \left(\frac{1}{K} \sum_{k=1}^{K} \mathcal{K}_{h_i}(\tilde{y}_i - \tilde{y}_i^{(k)}) \right), \quad \text{with} \quad \begin{cases} p_{3,i}^{(k)} \sim f_3(p_{3,i}^{(k)}|\boldsymbol{\theta}_3), \\ (p_{4,i}^{(k)}, p_{5,i}^{(k)}) \sim f_{45}((p_{4,i}^{(k)}, p_{5,i}^{(k)})|\boldsymbol{\theta}_{45}), \\ \tilde{y}_i^{(k)} = \mathcal{M}_{p_{1,i},p_2}(p_{3,i}^{(k)}, p_{4,i}^{(k)}, p_{5,i}^{(k)}), \\ h_i = (4/(3\mathrm{K}))^{1/5}\hat{\sigma}_i. \end{cases} \tag{33}$$

For $k = 1, \dots, K$, inputs $p_{3,i}^{(k)} \sim f_3(p_{3,i}^{(k)}|\boldsymbol{\theta}_3)$ and $(p_{4,i}^{(k)}, p_{5,i}^{(k)}) \sim f_{45}((p_{4,i}^{(k)}, p_{5,i}^{(k)})|\boldsymbol{\theta}_{45})$ are sampled from the corresponding population distributions, and responses $\tilde{y}_i^{(k)} = \mathcal{M}_{p_{1,i},p_2}(p_{3,i}^{(k)}, p_{4,i}^{(k)}, p_{5,i}^{(k)})$ are computed accordingly. Furthermore, $\hat{\sigma}_i$ denotes the standard deviation of the response samples $(\tilde{y}_i^{(1)}, \dots, \tilde{y}_i^{(K)})$. Note that, in Eq. (33), the density $f(\tilde{y}_i|p_{1,i}, p_2, \boldsymbol{\theta}_{45}, \boldsymbol{\theta}_3)$ is individually estimated for each $p_{1,i}$ with $i = 1, \dots, n$. The number of samples for each of these estimations is set to $K = 10^3$ and selection of the bandwidths follows the normal reference rule $h_i = (4/(3\mathrm{K}))^{1/5}\hat{\sigma}_i$.

Let us compare the transformed densities Eqs. (11) and (29). The random variable $\mathcal{M}_{p_2}(P_1, P_3, P_4, P_5) \sim f(\tilde{y}_i|p_2, \boldsymbol{\theta}_1, \boldsymbol{\theta}_{45}, \boldsymbol{\theta}_3)$ is conditioned on $\boldsymbol{\theta}_1$ and involves the uncertainty of $P_1 \sim f_1(p_1|\boldsymbol{\theta}_1)$. In contrast, $\mathcal{M}_{p_{1,i},p_2}(P_3, P_4, P_5) \sim f(\tilde{y}_i|p_{1,i}, p_2, \boldsymbol{\theta}_{45}, \boldsymbol{\theta}_3)$ is conditioned on the realization $p_{1,i}$ and does not bear reference to $\boldsymbol{\theta}_1$. Hence, $f(\tilde{y}_i|p_2, \boldsymbol{\theta}_1, \boldsymbol{\theta}_{45}, \boldsymbol{\theta}_3)$ is the broader and more complex density, whereas $f(\tilde{y}_i|p_{1,i}, p_2, \boldsymbol{\theta}_{45}, \boldsymbol{\theta}_3)$ is

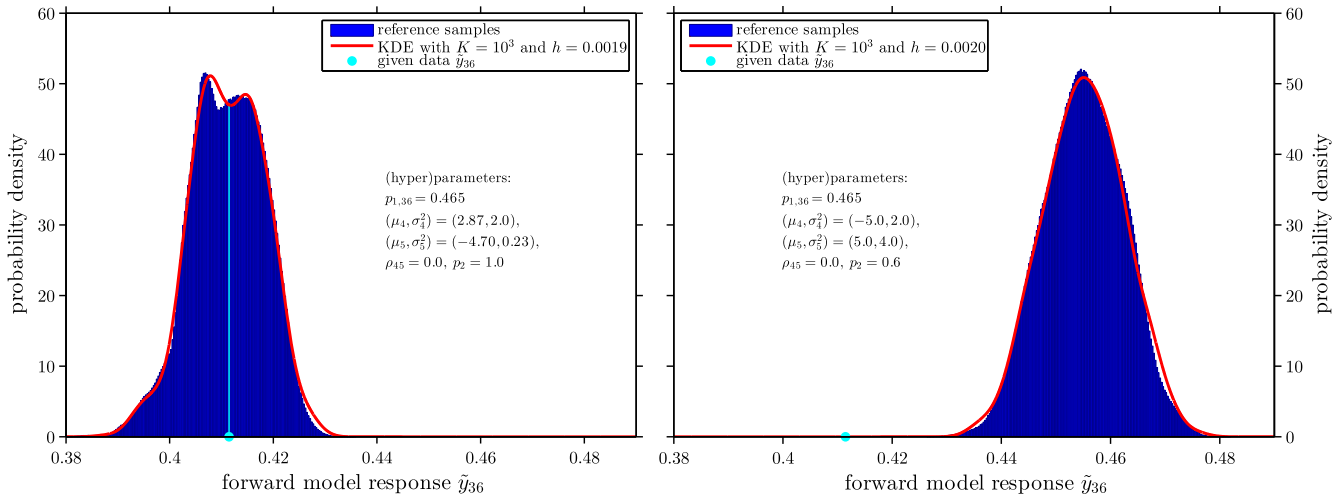

Fig. 10 Estimation of $f(\tilde{y}_{36}|p_{1,36}, p_2, \boldsymbol{\theta}_{45}, \boldsymbol{\theta}_3)$.

simpler and easier to estimate. In turn, likelihood estimations for MC^{3}DA are less biased and have a smaller variance than for MC^{3}KS. Thus, the approach promises a higher degree of posterior fidelity. In Fig. 10, the density $f(\tilde{y}_i|p_{1,i}, p_2, \boldsymbol{\theta}_{45}, \boldsymbol{\theta}_3)$ is shown for values $(p_2, \boldsymbol{\theta}_{45})_{\text{high}}$ and $(p_2, \boldsymbol{\theta}_{45})_{\text{low}}$ that have been chosen as the same values already used in Fig. 3. Following our final results, the value $p_{1,i} = 0.465$ has been exemplarily chosen as the posterior mean of $p_{1,i}$ with $i = 36$. KDEs of $f(\tilde{y}_{36}|p_{1,36}, p_2, \boldsymbol{\theta}_{45}, \boldsymbol{\theta}_3)$ with $K = 10^3$ are shown. The bandwidths $h = 0.0019$ and $h = 0.0020$ were automatically selected according to the normal reference rule. Histograms with a larger number $K = 10^7$ are shown for reference purposes. When comparing Figs. 3 and 10, one can clearly see the essential difference between the densities $f(\tilde{y}_i|p_2, \boldsymbol{\theta}_1, \boldsymbol{\theta}_{45}, \boldsymbol{\theta}_3)$ and $f(\tilde{y}_i|p_{1,i}, p_2, \boldsymbol{\theta}_{45}, \boldsymbol{\theta}_3)$. The latter is distinctly simpler and clearly better resembles a Gaussian density. Moreover, the densities for (hyper)parameter values of high and low posterior evidence only negligibly overlap. It can also be seen that the chosen values $p_{1,36}$ and $(p_2, \boldsymbol{\theta}_{45})_{\text{high}}$ lead to a response density, which is consistent with the observation $\tilde{y}_{36}$.

C. MCMC

The augmented posterior Eq. (30) is explored by means of a suitable MC^{3}DA sampler. Updating is done in blocks $\langle p_{1,i}\rangle$, (μ_1, σ_1^2), (p_2), (μ_4), (μ_5), and $(\sigma_4^2, \sigma_5^2, \rho_{45})$. Each $p_{1,i}$ in the block $\langle p_{1,i}\rangle$ is concurrently updated with a random walk Metropolis sampler based on independent Gaussian proposals with standard deviation $\sigma_{p_{1,i}} = 0.01$. As before, the remaining blocks are initialized in the middle of the corresponding epistemic intervals and updated with independent prior proposals. Acceptance rates amounted to approximately 10% for $\langle p_{1,i}\rangle$, (p_2), and (μ_5); 15% for (μ_4) and $(\sigma_4^2, \sigma_5^2, \rho_{45})$; and 30% for (μ_1, σ_1^2). A number of 10,219 proposals in the aforementioned block were rejected due to violating α_1, $\beta_1 > 1$. We start with preliminary MCMC runs with $N = 10^5$, $K = 10^3$ and constant bandwidths $h_i = 0.02$ (i.e., without automatic bandwidth selection) in order to identify the posterior modes of $\langle p_{1,i}\rangle$. Experiment-specific realizations $\langle p_{1,i}\rangle$ are initialized in the middle of their epistemic intervals and converge within approximately 1000 MCMC iterations. The initial convergence and final posterior of an experiment-specific realization $p_{1,i}$ with $i = 10$ are shown in Fig. 11. This shows that individual experiment-specific realizations $\langle p_{1,i}\rangle$ can indeed be inferred. The danger of the approach is that missing further posterior modes of $\langle p_{1,i}\rangle$ would alter the sampled posteriors of the remaining unknowns, above all, the one of $\boldsymbol{\theta}_1$. Convergence checks have therefore been accomplished by initializing $\langle p_{1,i}\rangle$ within admissible regions of the parameter space that have not been visited in previous runs. Ultimately, the chains converged to the same posterior modes that were found before. We conclude that the parameter space has been properly explored.

We point out that, although partial data augmentation has been motivated by considerations of posterior fidelity, it also gives additional insight into the inverse problems posed. Incidentally, the posterior of experiment-specific realizations $\langle p_{1,i}\rangle$ is explored and its modes are identified. Thus, we have gained knowledge about unknown problem quantities that are not primary QOI. Eventually, we initialize the final sampler within the detected posterior modes of $\langle p_{1,i}\rangle$. With $K = 10^3$ and automatic selection of the bandwidths h_i, we draw $N = 10^5$ posterior samples. Total

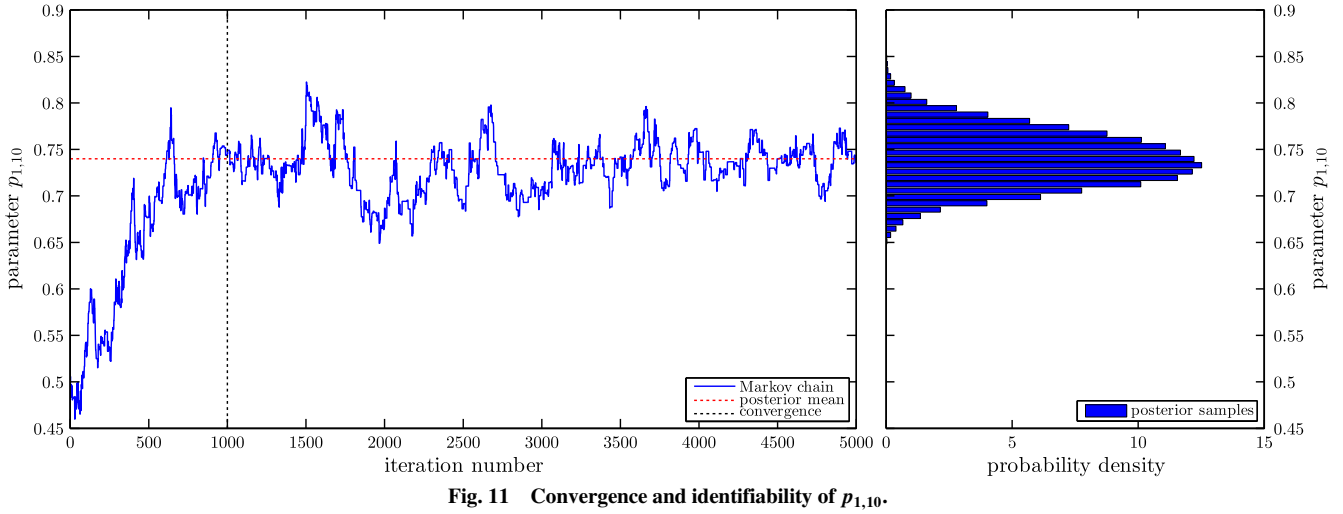

Fig. 11 Convergence and identifiability of $p_{1,10}$.

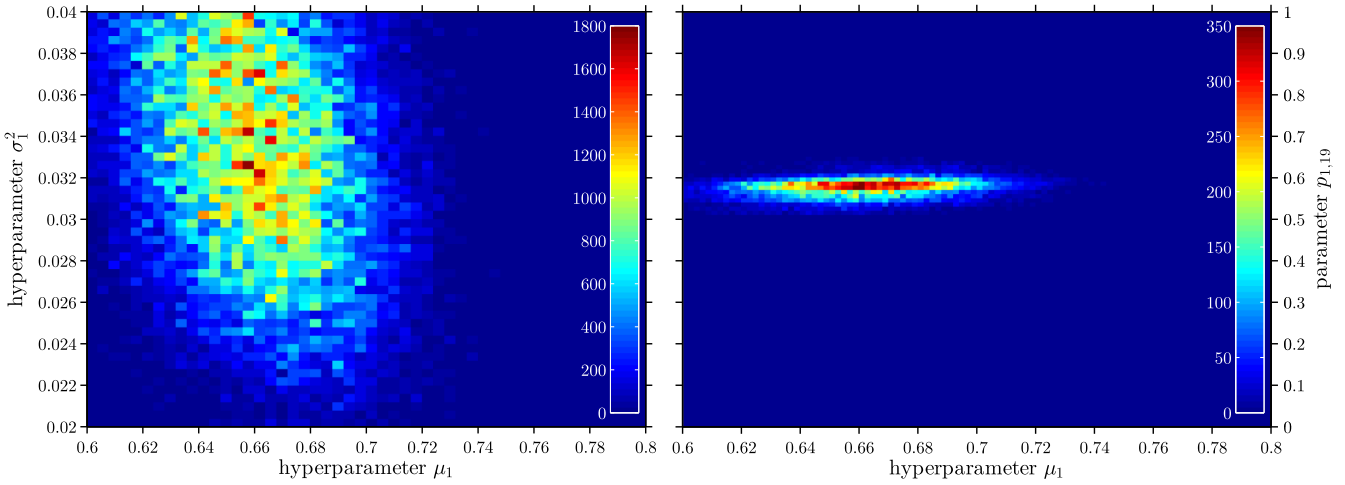

Fig. 12 Two-dimensional posteriors of (μ_1, σ_1^2) and $(\mu_1, p_{1,19})$.

execution time amounts to $t \approx 90h$ on a single core. The resulting posterior marginals of the QOI are added to Figs. 4–7. As compared to the results obtained by MC^{3}KS, the posteriors found by MC^{3}DA have been slightly shrunk and evolved in structure. Resting upon the assumption that the posterior modes of $\langle p_{1,i} \rangle$ have been correctly identified, we take this as an indication of a gain in posterior fidelity. In Fig. 12, two-dimensional posteriors are shown for (μ_1, σ_1^2) and $(\mu_1, p_{1,i})$ with $i = 19$. The corresponding linear coefficients of correlation are found to be $r_{\mu_1,\sigma_1^2} = -0.25$ and $r_{\mu_1,p_{1,19}} = 0.19$. Generally, we find small linear correlations $r_{\mu_1,p_{1,i}} \gtrsim 0$ between the mean hyperparameter μ_1 and experiment-specific realizations $p_{1,i}$ for nearly all $i = 1, \ldots, n$. This is plausible, since higher values of μ_1 increase the plausibility of higher values of each $p_{1,i}$, and vice versa.

VIII. Conclusions

Addressing the uncertainty characterization subproblem of the NASA Langley Multidisciplinary UQ Challenge has turned out to be a challenging yet rewarding task. It was began with the formulation of a generic Bayesian multilevel framework for managing different types of forward model input uncertainties in complex inverse problems. Incidentally, this showed how the problem could be solved for imperfect data (e.g., in the presence of additional measurement noise) and how the entirety of problem unknowns, including those that are not of declared inferential interest, could be deduced. Although these were not the guiding questions, this is a future research direction on its own. Bayesian structural prior modeling served as a foundation for devising a multilevel model in the zero-noise or perfect data limit, i.e., the data space has been endowed with a probability model that was based on uncertainty propagation. Ensuing from those general considerations, the challenge problem was interpreted and solved as Bayesian calibration of a suitably defined multilevel model. It was thoroughly commented on the assumptions that the adopted approach rests upon, as well as the interpretations it entails.

In turn, the problem solution has given rise to new questions of theoretical and practical relevance alike. Posterior fidelity was discussed in the context of MCMC posterior exploration and online uncertainty propagation. First, related thoughts were given and an in-depth consideration was initiated. The starting point of the latter could be Eqs. (13) to (16). With the objective of improving the fidelity of the final results, it was demonstrated how one can exploit partial data augmentation. In addition to improving the estimation of the QOI, in principle, this approach allowed inference of such problem unknowns that inferential interest was not particularly focused on. That way, partial data augmentation provided further insight into the calibration problem posed. In summary, it is hoped that these efforts prove to be a solid contribution to the NASA challenge problem in particular and to the theory and practice of Bayesian data analysis and uncertainty quantification in general. Future research work encompasses the design of more sophisticated methods to simulate the likelihood function and the rigorous assessment of the posterior fidelity.

References

[1] Crespo, L. G., Kenny, S. P., and Giesy, D. P., "The NASA Langley Multidisciplinary Uncertainty Quantification Challenge," *16th AIAA Non-Deterministic Approaches Conference (SciTech 2014)*, AIAA Paper 2014-1347, Jan. 2014.

[2] Crespo, L. G., Kenny, S. P., and Giesy, D. P., "NASA LaRC UQ Challenge 2014: Challenge Problem FAQ," NASA, May 2013, http://uqtools.larc.nasa.gov/nda-uq-challenge-problem-2014/faq/, [retrieved 2014].

[3] Foster, J. V., Cunningham, K., Fremaux, C. M., Shah, G. H., Stewart, E. C., Rivers, R. A., Wilborn, J. E., and Gato, W., "Dynamics Modeling and Simulation of Large Transport Airplanes in Upset Conditions," *AIAA Guidance, Navigation, and Control Conference and Exhibit*, AIAA Paper 2005-5933, Aug. 2005.

[4] Jordan, T. L., and Bailey, R. M., "NASA Langley's AirSTAR Testbed: A Subscale Flight Test Capability for Flight Dynamics and Control System Experiments," *AIAA Guidance, Navigation and Control Conference and Exhibit*, AIAA Paper 2008-6660, Aug. 2008.

[5] Crespo, L. G., Matsutani, M., and Annaswamy, A. M., "Design of an Adaptive Controller for a Remotely Operated Air Vehicle," *Journal of Guidance, Control, and Dynamics*, Vol. 35, No. 2, 2012, pp. 406–422. doi:10.2514/1.54779

[6] Faber, M. H., "On the Treatment of Uncertainties and Probabilities in Engineering Decision Analysis," *Journal of Offshore Mechanics and Arctic Engineering*, Vol. 127, No. 3, 2005, pp. 243–248. doi:10.1115/1.1951776

[7] Der Kiureghian, A., and Ditlevsen, O., "Aleatory or Epistemic? Does it Matter?" *Structural Safety*, Vol. 31, No. 2, 2009, pp. 105–112. doi:10.1016/j.strusafe.2008.06.020

[8] Hadidi, R., and Gucunski, N., "Probabilistic Approach to the Solution of Inverse Problems in Civil Engineering," *Journal of Computing in Civil Engineering*, Vol. 22, No. 6, 2008, pp. 338–347. doi:10.1061/(ASCE)0887-3801(2008)22:6(338)

[9] Beck, J. L., "Bayesian System Identification Based on Probability Logic," *Structural Control and Health Monitoring*, Vol. 17, No. 7, 2010, pp. 825–847. doi:10.1002/stc.v17:7

[10] Stuart, A. M., "Inverse Problems: A Bayesian Perspective," *Acta Numerica*, Vol. 19, No. 19, 2010, pp. 451–559. doi:10.1017/S0962492910000061
[11] Davidian, M., and Giltinan, D. M., "Nonlinear Models for Repeated Measurement Data: An Overview and Update," *Journal of Agricultural, Biological, and Environmental Statistics*, Vol. 8, No. 4, 2003, pp. 387–419. doi:10.1198/1085711032697
[12] Banks, H. T., Kenz, Z. R., and Thompson, W. C., "A Review of Selected Techniques in Inverse Problem Nonparametric Probability Distribution Estimation," *Journal of Inverse and Ill-Posed Problems*, Vol. 20, No. 4, 2012, pp. 429–460. doi:10.1515/jip-2012-0037
[13] de Rocquigny, E., and Cambier, S., "Inverse Probabilistic Modelling of the Sources of Uncertainty: A Non-Parametric Simulated-Likelihood Method with Application to an Industrial Turbine Vibration Assessment," *Inverse Problems in Science and Engineering*, Vol. 17, No. 7, 2009, pp. 937–959. doi:10.1080/17415970902916987
[14] Celeux, G., Grimaud, A., Lefèbvre, Y., and de Rocquigny, E., "Identifying Intrinsic Variability in Multivariate Systems Through Linearized Inverse Methods," *Inverse Problems in Science and Engineering*, Vol. 18, No. 3, 2010, pp. 401–415. doi:10.1080/17415971003624330
[15] Barbillon, P., Celeux, G., Grimaud, A., Lefèbvre, Y., and de Rocquigny, E., "Nonlinear Methods for Inverse Statistical Problems," *Computational Statistics and Data Analysis*, Vol. 55, No. 1, 2011, pp. 132–142. doi:10.1016/j.csda.2010.05.030
[16] Wakefield, J., "The Bayesian Analysis of Population Pharmacokinetic Models," *Journal of the American Statistical Association*, Vol. 91, No. 433, 1996, pp. 62–75. doi:10.1080/01621459.1996.10476664
[17] Lunn, D. J., "Bayesian Analysis of Population Pharmacokinetic/Pharmacodynamic Models," *Probabilistic Modeling in Bioinformatics and Medical Informatics*, edited by Husmeier, D., Dybowski, R., and Roberts, S., Advanced Information and Knowledge Processing, Springer, London, 2005, pp. 351–370.
[18] Daniels, M. J., "A Prior for the Variance in Hierarchical Models," *Canadian Journal of Statistics*, Vol. 27, No. 3, 1999, pp. 567–578. doi:10.2307/3316112
[19] Gelman, A., "Prior Distributions for Variance Parameters in Hierarchical Models (Comment on Article by Browne and Draper)," *Bayesian Analysis*, Vol. 1, No. 3, 2006, pp. 515–534. doi:10.1214/06-BA117A
[20] Wakefield, J. C., Smith, A. F. M., Racine-Poon, A., and Gelfand, A. E., "Bayesian Analysis of Linear and Non-Linear Population Models by Using the Gibbs Sampler," *Journal of the Royal Statistical Society. Series C (Applied Statistics)*, Vol. 43, No. 1, 1994, pp. 201–221. doi:10.2307/2986121
[21] Bennett, J. E., Racine-Poon, A., and Wakefield, J. C., "MCMC for Nonlinear Hierarchical Models," *Markov Chain Monte Carlo in Practice*, edited by Gilks, W. R., Richardson, S., and Spiegelhalter, D. J., Interdisciplinary Statistics, Chapman and Hall/CRC Press, Boca Raton, FL, 1995, pp. 339–357.
[22] Nagel, J. B., and Sudret, B., "Probabilistic Inversion for Estimating the Variability of Material Properties: A Bayesian Multilevel Approach," *Proceedings of the 11th International Probabilistic Workshop (IPW11)*, edited by Novák, D., and Vořechovský, M., Litera, Brno, Czech Republic, Nov. 2013, pp. 293–303.
[23] Ballesteros, G. C., Angelikopoulos, P., Papadimitriou, C., and Koumoutsakos, P., "Bayesian Hierarchical Models for Uncertainty Quantification in Structural Dynamics," *Vulnerability, Uncertainty, and Risk: Quantification, Mitigation, and Management*, edited by Beer, M., Au, S.-K., and Hall, J. W., American Soc. of Civil Engineers, Reston, VA, 2014, pp. 1615–1624, Chap. 162. doi:10.1061/9780784413609.162
[24] Draper, D., Gaver, D. P., Goel, P. K., Greenhouse, J. B., Hedges, L. V., Morris, C. N., and Waternaux, C. M., *Combining Information: Statistical Issues and Opportunities for Research*, Panel on Statistical Issues and Opportunities for Research in the Combination of Information, National Research Council, Washington, D.C., 1992.
[25] Nagel, J. B., and Sudret, B., "A Bayesian Multilevel Approach to Optimally Estimate Material Properties," *Vulnerability, Uncertainty, and Risk: Quantification, Mitigation, and Management*, edited by Beer, M., Au, S.-K., and Hall, J. W., American Soc. of Civil Engineers, Reston, VA, 2014, pp. 1504–1513, Chap. 151. doi:10.1061/9780784413609.151
[26] Koski, T., and Noble, J. M., *Bayesian Networks: An Introduction*, Wiley Series in Probability and Statistics, John Wiley and Sons, Chichester, West Sussex, U.K., 2009.
[27] Kjærulff, U. B., and Madsen, A. L., *Bayesian Networks and Influence Diagrams: A Guide to Construction and Analysis*, Vol. 22, Information Science and Statistics, 2nd ed., Springer, New York, 2013.
[28] Basu, D., "On the Elimination of Nuisance Parameters," *Journal of the American Statistical Association*, Vol. 72, No. 358, 1977, pp. 355–366. doi:10.1080/01621459.1977.10481002
[29] Dawid, A. P., "A Bayesian Look at Nuisance Parameters," *Trabajos de Estadística Y de Investigación Operativa*, Vol. 31, No. 1, 1980, pp. 167–203. doi:10.1007/BF02888351
[30] Berger, J. O., Liseo, B., and Wolpert, R. L., "Integrated Likelihood Methods for Eliminating Nuisance Parameters," *Statistical Science*, Vol. 14, No. 1, 1999, pp. 1–28.
[31] Severini, T. A., "Integrated Likelihood Functions for Non-Bayesian Inference," *Biometrika*, Vol. 94, No. 3, 2007, pp. 529–542. doi:10.1093/biomet/asm040
[32] Nagel, J. B., and Sudret, B., "A Bayesian Multilevel Framework for Uncertainty Characterization and the NASA Langley Multidisciplinary UQ Challenge," *16th AIAA Non-Deterministic Approaches Conference (SciTech 2014)*, AIAA Paper 2014-1502, Jan. 2014.
[33] Diaconis, P., and Freedman, D., "On the Consistency of Bayes Estimates," *Annals of Statistics*, Vol. 14, No. 1, 1986, pp. 1–26. doi:10.1214/aos/1176349830
[34] Vollmer, S. J., "Posterior Consistency for Bayesian Inverse Problems Through Stability and Regression Results," *Inverse Problems*, Vol. 29, No. 12, 2013, Paper 125011. doi:10.1088/0266-5611/29/12/125011
[35] Silverman, B. W., *Density Estimation for Statistics and Data Analysis*, Vol. 26, Monographs on Statistics and Applied Probability, Chapman and Hall/CRC Press, Boca Raton, FL, 1986.
[36] Wand, M. P., and Jones, M. C., *Kernel Smoothing*, Vol. 60, Monographs on Statistics and Applied Probability, Chapman and Hall/CRC Press, Boca Raton, FL, 1994.
[37] Sudret, B., *Uncertainty Propagation and Sensitivity Analysis in Mechanical Models: Contributions to Structural Reliability and Stochastic Spectral Methods*, Habilitation à Diriger des Recherches, Univ. Blaise Pascal, Clermont-Ferrand, France, 2007, p. 252.
[38] Sudret, B., Perrin, F., and Pendola, M., "Use of Polynomial Chaos Expansions in Stochastic Inverse Problems," *Proceedings of the 4th International ASRANet Colloquium*, ASRANet, Glasgow, U.K., 2008.
[39] Robert, C. P., and Casella, G., *Monte Carlo Statistical Methods*, Springer Series in Statistics, 2nd ed., Springer, New York, 2004.
[40] Metropolis, N., Rosenbluth, A. W., Rosenbluth, M. N., Teller, A. H., and Teller, E., "Equation of State Calculations by Fast Computing Machines," *Journal of Chemical Physics*, Vol. 21, No. 6, 1953, pp. 1087–1092. doi:10.1063/1.1699114

[41] Hastings, W. K., "Monte Carlo Sampling Methods Using Markov Chains and Their Applications," *Biometrika*, Vol. 57, No. 1, 1970, pp. 97–109. doi:10.1093/biomet/57.1.97
[42] Cowles, M. K., and Carlin, B. P., "Markov Chain Monte Carlo Convergence Diagnostics: A Comparative Review," *Journal of the American Statistical Association*, Vol. 91, No. 434, 1996, pp. 883–904. doi:10.1080/01621459.1996.10476956
[43] Brooks, S. P., and Roberts, G. O., "Convergence Assessment Techniques for Markov Chain Monte Carlo," *Statistics and Computing*, Vol. 8, No. 4, 1998, pp. 319–335. doi:10.1023/A:1008820505350
[44] Diggle, P. J., and Gratton, R. J., "Monte Carlo Methods of Inference for Implicit Statistical Models," *Journal of the Royal Statistical Society. Series B (Methodological)*, Vol. 46, No. 2, 1984, pp. 193–227.
[45] O'Neill, P. D., Balding, D. J., Becker, N. G., Eerola, M., and Mollison, D., "Analyses of Infectious Disease Data from Household Outbreaks by Markov Chain Monte Carlo Methods," *Journal of the Royal Statistical Society. Series C (Applied Statistics)*, Vol. 49, No. 4, 2000, pp. 517–542. doi:10.1111/1467-9876.00210
[46] Beaumont, M. A., "Estimation of Population Growth or Decline in Genetically Monitored Populations," *Genetics*, Vol. 164, No. 3, 2003, pp. 1139–1160.
[47] Bal, G., Langmore, I., and Marzouk, Y., "Bayesian Inverse Problems with Monte Carlo Forward Models," *Inverse Problems and Imaging*, Vol. 7, No. 1, 2013, pp. 81–105. doi:10.3934/ipi.2013.7.81
[48] Korattikara, A., Chen, Y., and Welling, M., "Austerity in MCMC Land: Cutting the Metropolis–Hastings Budget," *JMLR Workshop and Conference Proceedings: 31st International Conference on Machine Learning (ICML 2014)*, Vol. 32, No. 1, 2014, pp. 181–189.
[49] Bardenet, R., Doucet, A., and Holmes, C., "Towards Scaling up Markov Chain Monte Carlo: An Adaptive Subsampling Approach," *JMLR Workshop and Conference Proceedings: 31st International Conference on Machine Learning (ICML 2014)*, Vol. 32, No. 1, 2014, pp. 405–413.
[50] McFarland, J. M., Bichon, B. J., and Riha, D. S., "A Probabilistic Treatment of Multiple Uncertainty Types: NASA UQ Challenge," *16th AIAA Non-Deterministic Approaches Conference (SciTech 2014)*, AIAA Paper 2014-1500, Jan. 2014.
[51] Andrieu, C., and Roberts, G. O., "The Pseudo-Marginal Approach for Efficient Monte Carlo Computations," *Annals of Statistics*, Vol. 37, No. 2, 2009, pp. 697–725. doi:10.1214/07-AOS574
[52] Helton, J. C., and Oberkampf, W. L., "Alternative Representations of Epistemic Uncertainty," *Reliability Engineering and System Safety*, Vol. 85, Nos. 1–3, 2004, pp. 1–10. doi:10.1016/j.ress.2004.03.001
[53] Helton, J. C., and Johnson, J. D., "Quantification of Margins and Uncertainties: Alternative Representations of Epistemic Uncertainty," *Reliability Engineering and System Safety*, Vol. 96, No. 9, 2011, pp. 1034–1052. doi:10.1016/j.ress.2011.02.013
[54] Tanner, M. A., and Wong, W. H., "The Calculation of Posterior Distributions by Data Augmentation," *Journal of the American Statistical Association*, Vol. 82, No. 398, 1987, pp. 528–540. doi:10.1080/01621459.1987.10478458
[55] van Dyk, D. A., and Meng, X.-L., "The Art of Data Augmentation," *Journal of Computational and Graphical Statistics*, Vol. 10, No. 1, 2001, pp. 1–50. doi:10.1198/10618600152418584
[56] van Dyk, D. A., "Hierarchical Models, Data Augmentation, and Markov Chain Monte Carlo," *Statistical Challenges in Astronomy*, Springer, New York, 2003, pp. 41–55.

L. Crespo
Associate Editor