SANDIA REPORT
SAND2025-10780
Printed September 2025

Sandia National Laboratories

Theoretical and methodological challenges in hierarchical Bayesian inference for model-form uncertainty

Teresa Portone & Rebekah D. White

Prepared by
Sandia National Laboratories
Albuquerque, New Mexico 87185
Livermore, California 94550

Issued by Sandia National Laboratories, operated for the United States Department of Energy by National Technology & Engineering Solutions of Sandia, LLC.

Available to DOE and DOE contractors from

U.S. Department of Energy
Office of Scientific and Technical Information
P.O. Box 62
Oak Ridge, TN 37831

Telephone: (865) 576-8401
Facsimile: (865) 576-5728
E-Mail: reports@osti.gov
Online ordering: http://www.osti.gov/scitech

Available to the public from

U.S. Department of Commerce
National Technical Information Service
5301 Shawnee Road
Alexandria, VA 22312

Telephone: (800) 553-6847
Facsimile: (703) 605-6900
E-Mail: orders@ntis.gov
Online order: https://classic.ntis.gov/help/order-methods

ABSTRACT

This report describes challenges associated with the hierarchical Bayesian approach to inform model-form uncertainty (MFU) representations, which are parameterized modifications to a mathematical models' governing equations to express uncertainty in form of the equations. To inform model-form uncertainties, hierarchical Bayesian inference is often employed. Here, the MFU parameters are distributed parametrically, and the hyperparameters of the parametric distribution are informed through Bayesian inference, with the aim of determining the MFU parameter distribution that best agrees with calibration data. In practice, however, we have found the hierarchical Bayesian approach falls short of this aim. We discuss theoretical and methodological challenges of the approach, and we present several numerical demonstrations of these challenges. To conclude, we suggest promising alternative approaches for future investigation.

This page intentionally left blank.

ACKNOWLEDGEMENT

This work was supported by the Laboratory Directed Research and Development program at Sandia National Laboratories, a multimission laboratory managed and operated by National Technology and Engineering Solutions of Sandia LLC, a wholly owned subsidiary of Honeywell International Inc. for the U.S. Department of Energy's National Nuclear Security Administration under contract DE-NA0003525.

This page intentionally left blank.

CONTENTS

This page intentionally left blank.

LIST OF FIGURES

1. INTRODUCTION

For many high-consequence design, engineering, and policy decisions, inability to gather observational data means computational models are the only sources of information to inform those decisions. However, such models are by definition simplifications of the complex realities they aim to represent, and there are often significant uncertainties in how to specify the model (e.g., model parameters, boundary and initial conditions, closure or constitutive models). For computational models to be useful for decision making, it is imperative to accurately characterize these uncertainties, endowing model predictions with a measure of confidence. In the practice of uncertainty quantification, these uncertainties are modeled probabilistically, e.g., as random variables or fields, and the probabilistic models are propagated through the model so that model predictions themselves have a probability distribution. By querying the probability distributions on model predictions it is possible to assess confidence in modeled outcomes, e.g., by computing moments or central tendencies or by computing probability of exceeding some threshold.

Bayesian inference provides a principled, probabilistic framework to update the probability distributions of uncertainty sources using data while coherently incorporating prior knowledge. Using Bayes' theorem, a prior distribution on uncertainty sources is combined with a likelihood distribution on observational data to produce a posterior distribution on uncertainty distributions. Unfortunately, when the simplifications (or model-form errors) in a computational model result in discrepancies between model outputs and the data employed in Bayesian inference, the result is posterior distributions which concentrate towards values that minimize these discrepancies [1], leading to overconfidence and biased distributions. This behavior is highly undesirable when the goal is to attain meaningful uncertainty characterizations to propagate to model predictions.

While there has been significant work to mitigate the impact of model-form errors and their resulting model discrepancies [2–6], we focus here on an emergent approach that aims not just to quantify uncertainties, but to enrich the model through embedded direct representations of model-form error, which we term model-form uncertainty (MFU) representations [7–13]. MFU representations are parameterized modifications to the governing equations of the model which represent uncertainty due to model-form error (MFE). By directly representing MFE, such representations have the promise of reducing or eliminating model discrepancies so that calibrated probability distributions using Bayesian inference are once again meaningful characterizations of uncertainty. Because they are embedded within a computational model and propagate to all model outputs, both observable and unobservable, MFU representations are attractive for problems where extrapolative predictions are required—the MFU representation can be informed with available data, then uncertainties can be propagated to unobservable prediction quantities of interest (QoIs).

MFU representations may reduce model-form errors, but in practice they cannot eliminate them, so they are commonly hyperparameterized in order to maintain random behavior no matter how much data is used to inform them (see Chapter 2 for more details). Hyperparameterization approaches

incorporate hyperparameters that define the MFU parameters' probability distributions into the inference problem [5, 8, 10–13]. The hyperparameters are themselves cast as random variables with prior probability distributions, and Bayesian inference is carried out with respect to the hyperparameters. The intuition behind this approach is to use data to inform the probability distribution of MFU parameters that is most consistent with the data. Ideally, it would lead to posterior distributions that are consistent with observational data without being over- or under-confident.

Unfortunately, in practice we have found the hierarchical Bayesian method for informing MFU representations to be unsatisfactory for attaining meaningful uncertainty characterizations. In this report, we detail the theoretical and methodological challenges we have encountered in the use of hierarchical Bayesian inference for MFU representation calibration. The report proceeds as follows: in Chapter 2 we introduce the mathematical framework for MFU representations and define the hierarchical Bayesian inverse problem; in Chapter 3 we discuss the theoretical and methodological challenges associated with the hierarchical Bayesian approach to MFU calibration; in Chapter 4 we provide a series of numerical demonstrations that illustrate the previously discussed challenges; in Chapter 5 we conclude and suggest future lines of inquiry.

2. BACKGROUND

Let a mathematical model governing the evolution of state variables v and model output mappings be denoted

$$0 = \mathcal{R}(v; \gamma), \tag{2.1}$$

$$d = \mathcal{O}(v) + \epsilon_m, \quad \epsilon_m \sim \pi_\epsilon(\epsilon_m), \tag{2.2}$$

$$Q = \mathcal{Q}(v), \tag{2.3}$$

where (2.1) represents a set of governing equations in their residual form with parameters γ; (2.2) is the mapping of v to observable data d through an observation operator $\mathcal{O}$, with measurement error ϵ_m typically assumed to be a random variable with distribution $\pi_\epsilon(\epsilon_m)$; and (2.3) is the mapping of v to a predicted quantity of interest (QoI) Q, which may not be observable (e.g., a prediction of the future). The state variables v depend on parameters γ through (2.1); to simplify notation we therefore denote the parameter-to-observable map as

$$m(\gamma) = \mathcal{O}(v(\gamma)). \tag{2.4}$$

Similarly we can define the parameter-to-QoI map:

$$q(\gamma) = \mathcal{Q}(v(\gamma)). \tag{2.5}$$

If γ are uncertain, we can pose probability distributions reflecting our prior knowledge about their plausible values, called the *prior distribution*: $\gamma \sim \pi_0(\gamma)$. Rewriting (2.2) using the parameter-to-observable map,

$$d = m(\gamma) + \epsilon_m, \quad \epsilon_m \sim \pi_\epsilon(\epsilon_m),$$

where this equation can be used to construct the probability of the data arising from the model for a given value of γ, called the *likelihood distribution*:

$$\pi(d|\gamma) = \pi_\epsilon(d - m(\gamma)). \tag{2.6}$$

Bayes' theorem defines the *posterior distribution* as

$$\pi(\gamma|d) = \frac{\pi(d|\gamma)\pi_0(\gamma)}{\int \pi(d|\gamma)\pi_0(\gamma)\mathrm{d}\gamma}. \tag{2.7}$$

In many practical applications, simplifications to the governing equations relative to the real phenomenon being modeled lead to discrepancies between observational data d, such that the

difference between data and model could not plausibly be drawn from the measurement error distribution for any γ:

$$d - m(\gamma) \nsim \pi_\epsilon(\epsilon_m) \quad \forall\ \gamma.$$

This phenomenon can also be described through the concept of *model misspecification* in the statistical literature. In this case, with increasing data the Bayesian posterior concentrates about the γ values which minimize misfit with the data, while never actually agreeing with it [1]. This is undesirable when the ultimate goal of Bayesian inference is to attain well-calibrated expressions of input uncertainty to attain meaningful characterization of uncertainties in prediction QoIs $q(\gamma)$. Propagating a posterior distribution from Bayesian inference with a misspecified model to prediction QoIs could result in biased and overconfident predictions with the potential to adversely affect decision making.

Model-form uncertainty (MFU) representations seek to characterize uncertainty in the mathematical form of the governing equations by introducing parameterized modifications embedded within the equations. Denoting an MFU representation at ξ and its parameters as θ, our system of governing equations and output mappings becomes

$$0 = \widetilde{\mathcal{R}}_{\xi(\theta)}(\tilde{v}; \gamma) \tag{2.8}$$

$$d = \mathcal{O}(\tilde{v}) + \epsilon_m, \quad \epsilon_m \sim \pi_\epsilon(\epsilon_m) \tag{2.9}$$

$$q' = \mathcal{Q}(\tilde{v}), \tag{2.10}$$

where the governing equations are denoted $\widetilde{\mathcal{R}}$ to signify that they have been modified by the MFU representation, and the state variables are denoted $\tilde{v}$ to indicate that the solution to the governing equations is subsequently modified as well. Similarly, we denote the parameter-to-observable and parameter-to-QoI maps with the incorporation of the MFU representation as

$$\tilde{m}(\gamma, \theta) = \mathcal{O}(\tilde{v}(\gamma, \theta)), \tag{2.11}$$

$$\tilde{q}(\gamma, \theta) = \mathcal{Q}(\tilde{v}(\gamma, \theta)). \tag{2.12}$$

MFU representations can mitigate but not fully eliminate model misspecification, so posterior concentration can still occur. Therefore, to prevent uncertainty collapse, MFU parameters are commonly hyperparameterized. This could be achieved multiple ways, but a common approach taken in [7, 8, 10–12] poses a hierarchical probabilistic model for the hyperparameters φ of the probability distributions of θ:

$$\pi(\theta, \varphi) = \pi(\theta|\varphi)\pi(\varphi). \tag{2.13}$$

The hierarchical Bayesian approach to informing MFU representations seeks the posterior distribution

$$\pi(\gamma, \varphi|d) \propto \int \pi(\gamma, \theta, \varphi|d)\mathrm{d}\theta \tag{2.14}$$

$$\propto \left(\int \pi_\epsilon(d|\gamma, \theta, \varphi)\pi(\theta|\varphi)\mathrm{d}\theta \right) \pi_0(\gamma)\pi_0(\varphi), \tag{2.15}$$

where the MFU parameters θ are marginalized out. By defining the probability distribution for θ hierarchically, even if hyperparameter uncertainty collapses in the limit of infinite data as per [1], we retain a probabilistic representation of θ. To guarantee this behavior, $\pi_0(\varphi)$ must be properly specified, e.g., if θ is modeled as a normal random variable, the hyperprior on standard deviation must not be allowed to collapse to zero. Additionally, by inferring the hyperparameters of the distribution for θ, the hope is that uncertainty can be "calibrated" such that the observational data is plausible according to the uncertainty characterization of the model, while at the same time the predictive uncertainties (those propagated through the model to observable or unobservable outputs) are not so large that they cannot inform decisions (see Figure 2-1).

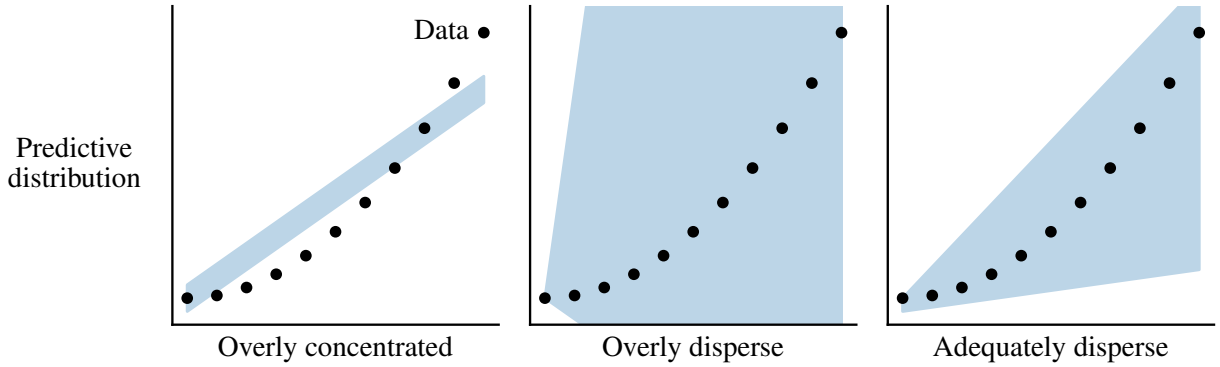

Figure 2-1. An illustration of different predictive uncertainty scenarios where data is nonlinear but the model is linear. On the left, MFU is not addressed, resulting in an overly-concentrated predictive distribution that assigns low probability to many of the data points. In the middle, the data is probable, but the uncertainty bounds are so great that they couldn't meaningfully inform a decision. On the right, the data is probable according to the predictive distribution, but the bounds may be tight enough to inform a decision. Note that an "adequately disperse" predictive distribution is necessary but not sufficient to determine whether an uncertainty characterization is meaningful for extrapolative prediction.

In practice, however, we have found that the hierarchical Bayesian approach to informing MFU representations does not achieve this hope. We have encountered a range of theoretical and methodological challenges with the approach which show that it does not align well with the goals of achieving meaningful uncertainty characterizations that can be propagated to predictive QoIs to represent confidence in those predictions. We present these challenges herein.

This page intentionally left blank.

3. THEORETICAL AND METHODOLOGICAL CHALLENGES

We begin by detailing the theoretical challenges associated with the hierarchical Bayesian approach to informing MFU representations presented in Chapter 2. Our first theoretical challenge is related to the implicit assumption associated with specifying a likelihood distribution, which when incorporating an MFU can be written as

$$d = \tilde{m}(\gamma, \theta) + \epsilon_m. \tag{3.1}$$

This mathematical statement implies an assumption that the model is capable of representing of the data-generating process, which by the need for an MFU representation has already been determined to be untrue. Even with the incorporation of the MFU representation, we do not expect the model to fully capture the modeled phenomenon. Nevertheless, the approach in Chapter 2 proceeds with this form of the likelihood, which is theoretically unsatisfying. Note that this challenge is not necessarily associated with the adoption of a hierarchical Bayesian approach, but rather with the form of (3.1).

A second theoretical challenge with the approach is in defining prior distributions for the hyperparameters, called *hyperpriors*. Prior distributions are an important aspect of the Bayesian inference problem where knowledge about the inferred quantity can be encoded. However, in general we lack information or even intuition about hyperparameters. For example, while we might have some sense of an appropriate prior on the MFU parameters directly, how that maps to the hyperpriors is not typically clear. If an MFU parameter is assumed to be a normal random variable and we have some intuition about the mode and confidence bounds of the distribution, there are many potential combinations of hyperparameters (mean and standard deviation) that could produce normal distributions consistent with that intuition. Without adequate information to pose hyperpriors, two different hyperprior specifications can results in very different posterior distributions, despite nearly identical induced prior distributions on the MFU parameters.

Scale parameters such as the standard deviation can be especially challenging. Without further intuition, the natural approach is to define an uninformative prior encoding positivity, however we have found that in practice, scale parameters are not well informed by data in the hierarchical Bayesian inference problem. Such lack of identifiability is problematic because the scale parameters play a critical role in how broad or concentrated our MFU parameters' uncertainty characterizations are, which in turn influences how broad or concentrated the predictive QoIs' uncertainty characterizations are. Another issue arising from the lack of identifiability of the hyperparameters is that the resulting uncertainty characterization is agnostic to the level of model-data discrepancy present. Thus, while the hierarchical approach may prevent complete collapse of the posterior predictive uncertainties, it is not inherently capable of capturing uncertainties arising from model misspecification. Loosely speaking, the hierarchical framework does not provide a sense of how much uncertainty is needed in the parameters to reasonably describe the observational data.

The hierarchical approach also poses the methodological challenge of increasing the dimension of the inference problem. A common choice for the hierarchical probabilistic model for the MFU parameters is a normal distribution; then by posing the Bayesian inverse problem with respect to its mean and standard deviation, we have effectively doubled the dimension of the inverse problem. This approach is therefore subject to the so-called "curse of dimensionality", which poses numerical challenges for many computational methods for Bayesian inverse problems and can limit the practical number of parameters and/or hyperparameters that can feasibly be included in a Bayesian inverse problem.

4. NUMERICAL DEMONSTRATIONS

Here we demonstrate some of the theoretical challenges described above, as well as challenges observed in practice when attempting to apply the hierarchical Bayesian approach to inform MFU representations. For simplicity, inference does not include other model parameters γ in these demonstrations; however, we observe the same trends illustrated here when other model parameters are included. As a result of considering only the MFU parameters, the hierarchical posterior we seek is simplified to the marginal on the MFU hyperparameters $\pi(\varphi|d)$. Since the hyperparameters do not explicitly appear in the parameter-to-observable map $\widetilde{m}(\theta)$, it is necessary to sample the joint distribution defined in Equation (2.13) when generating samples of θ and corresponding predictive samples using Equation (3.1). Prior samples are generated by taking $\pi(\varphi) = \pi_0(\varphi)$, while posterior samples consider $\pi(\varphi) = \pi(\varphi|d)$. We present these challenges in the context of two exemplar problems: a spatially varying linear model and a model for subsurface contaminant transport.

4.1. Spatially varying linear model

We first consider a simple problem where the data-generating function is nonlinear in x:

$$d_i = 3x_i^{\alpha} + \epsilon_m, \quad x_i = \frac{i}{3}, i = 0, \ldots, 9, \quad \alpha > 1, \quad \epsilon_m \sim \mathcal{N}(0, (0.1)^2),$$

where $\alpha = 2$ unless stated otherwise. However, the mathematical model assumes a linear dependence on x:

$$m(\theta) = \theta x.$$

For the sake of this demonstration, we assume that we cannot improve the linear model and thus aim to represent MFU by hyperparameterizing its slope. Therefore, we hierarchically model

$$\theta \sim \mathcal{N}(\mu, \sigma^2), \quad \mu \sim \pi(\mu), \quad \sigma \sim \pi(\sigma),$$

where μ is modeled as a normal random variable and σ is modeled as an exponential random variable defined by its rate parameter λ. Unless stated otherwise we take $\pi_0(\mu) = \mathcal{N}(6, 3^2)$ and $\pi_0(\sigma) \sim \exp(6)$.

As shown in Figure 4-1, the hierarchical Bayesian approach does not always yield posterior distributions that are in good agreement with data. The posterior predictive high-probability region has contracted significantly to the point where calibration data falls outside it. Such behavior is problematic because the aim of the predictive distribution is to assign meaningful probabilities to the model outputs. The fact that the predictive distribution assigns low probability to observational

data tells us our probabilistic representation is incorrect. This is exactly the situation we aimed to avoid by implementing the hierarchical approach: overconfidence in incorrect model predictions. In Figure 4-1, we observe that the posterior density for μ has concentrated in the region around

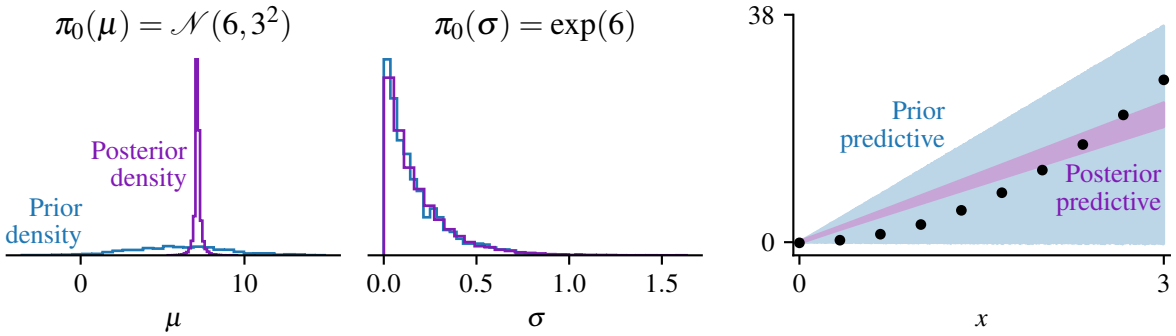

Figure 4-1. Left: Marginal prior and posterior densities for μ and σ. Right: 95% high-probability region for prior and posterior predictive distributions based on 0.025 and 0.975 quantiles, compared to calibration data (black).

the maximum likelihood estimate (MLE) for $\theta^* \approx 7$. However, the posterior marginal density for σ overlaps closely with its prior, indicating it was not informed by the calibration data. This lack of update in the distribution for a scale hyperparameter such as σ is problematic because it means that posterior predictive uncertainties are significantly influenced by our choice of hyperprior for σ. However, as previously discussed, we often do not have adequate information about σ to define informative hyperpriors. One means of mitigating this effect might be to define noninformative hyperpriors, however, since they are not informed by data, this can lead to overly dispersed predictive distributions, which lead to such high predictive uncertainties, they cannot be used to inform decision making.

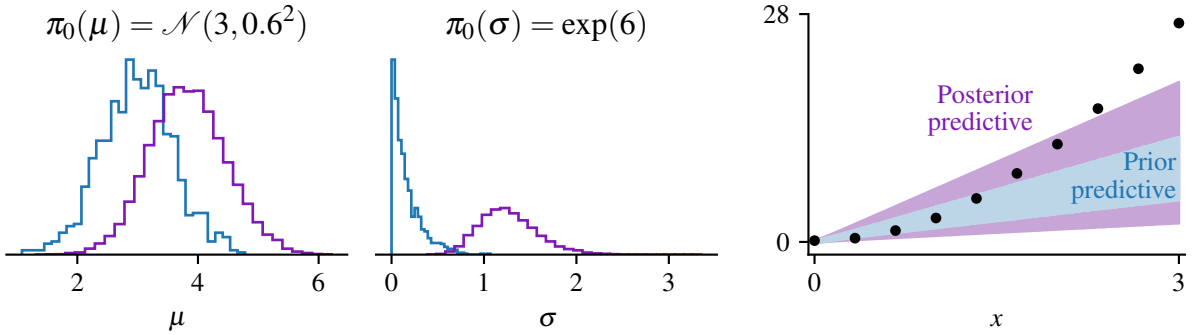

Figure 4-2. Left: Marginal prior and posterior densities for μ and σ. Right: 95% high-probability region for prior and posterior predictive distributions based on 0.025 and 0.975 quantiles, compared to calibration data (black).

A numerical exploration of the conditions under which the scale parameter is informed by data indicates it may require significantly degenerate cases, e.g., the case shown in Figure 4-2, where the overly restrictive hyperprior results in a prior predictive that assigns low probability to some of the observational data (i.e., the prior does not fully encompass the data). In comparison to the previous example shown in Figure 4-1, the hyperprior on μ has a smaller mean and standard deviation, such that the MLE for $\theta^* \approx 7$ is improbable. While the posterior marginal density on μ has shifted its

support towards this value, the high confidence of the hyperprior limits how much the posterior can shift. As a result, the posterior marginal density for σ has shifted its support to larger values so the MLE of θ is plausible under the hierarchically defined distribution for θ (taking the approximate mode of the hyperparameter posteriors, $\theta \sim \mathcal{N}(4, 1.5^2)$ such that the upper bound of the 95% CI is 7). However, while the uncertainty bounds have increased in the posterior predictive distribution, they did not increase enough to assign high probability to all the data. Furthermore, the fact that the scale parameter is identifiable primarily in such degenerate cases does not enable practical use of the hierarchical framework.

The challenge of hyperprior specification is further highlighted in Figure 4-3. Recall that often the only information available to inform hyperpriors is their impact on the distribution of the MFU parameter (in this case θ) and/or the predictive distribution. In this example we show two different hyperprior specifications that yield almost identical θ and predictive distributions. However, the resulting posterior predictive distributions differ significantly.

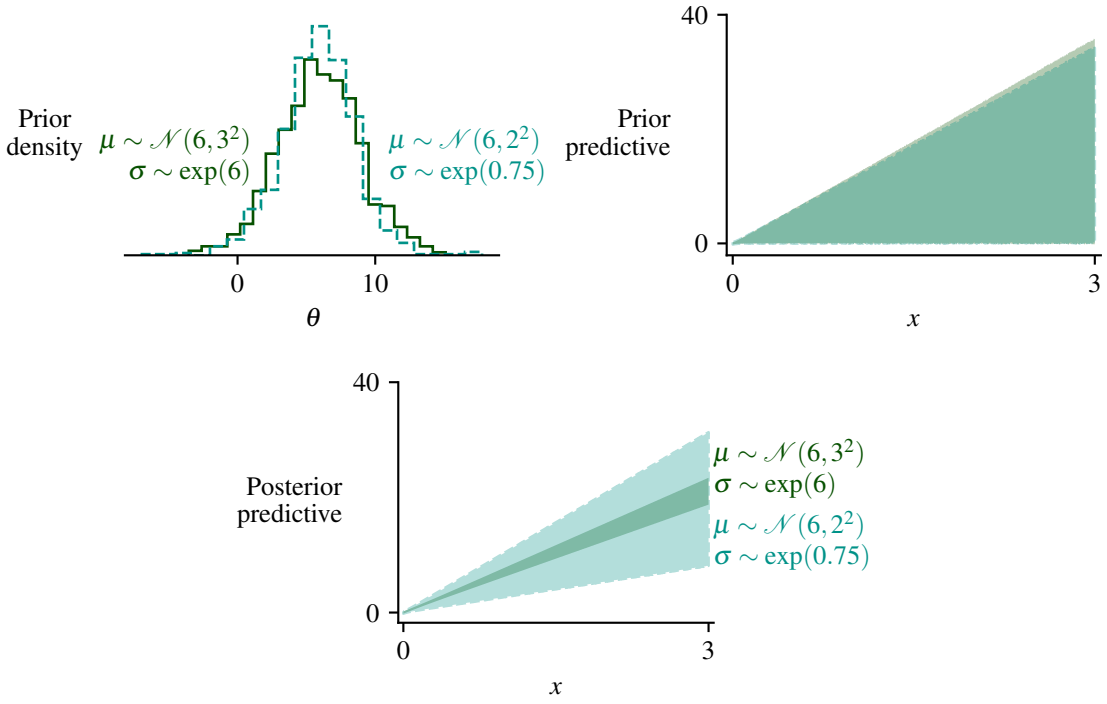

Figure 4-3. Top left: Marginal prior densities for θ under different hyperprior specifications, which are differentiated by shading and solid vs. dashed lines. Top right and bottom: 95% high-probability region for prior and posterior predictive distributions under different hyperprior specifications. The prior predictive distributions are nearly identical, making them difficult to distinguish.

Ideally, MFU would increase as the discrepancy between calibration data and model outputs increases; however, as shown in Figure 4-4, this is not necessarily the case. In this example, we fix our hyperpriors at $\mu \sim \mathcal{N}(6, 3^2)$ and $\sigma \sim \exp(6)$ and vary the exponent of the nonlinear data-generating function from weakly nonlinear ($\alpha = 1.3$) to almost quadratic ($\alpha = 1.9$) so that there is increasing discrepancy with the linear model. While the mode of the Gaussian describing μ

increases with the degree of misspecification, the variance associated with μ is not informed by the data for any considered degree of nonlinearity. This, in combination with the lack of identifiability of σ, results in a posterior predictive with similar point-wise variances, regardless of the degree of misspecification present. Hence, when using the hierarchical framework, the mean behavior of the predictive distribution may adjust with changing misspecification, but it does not lead to meaningful uncertainty characterizations capable of reflecting increased uncertainties.

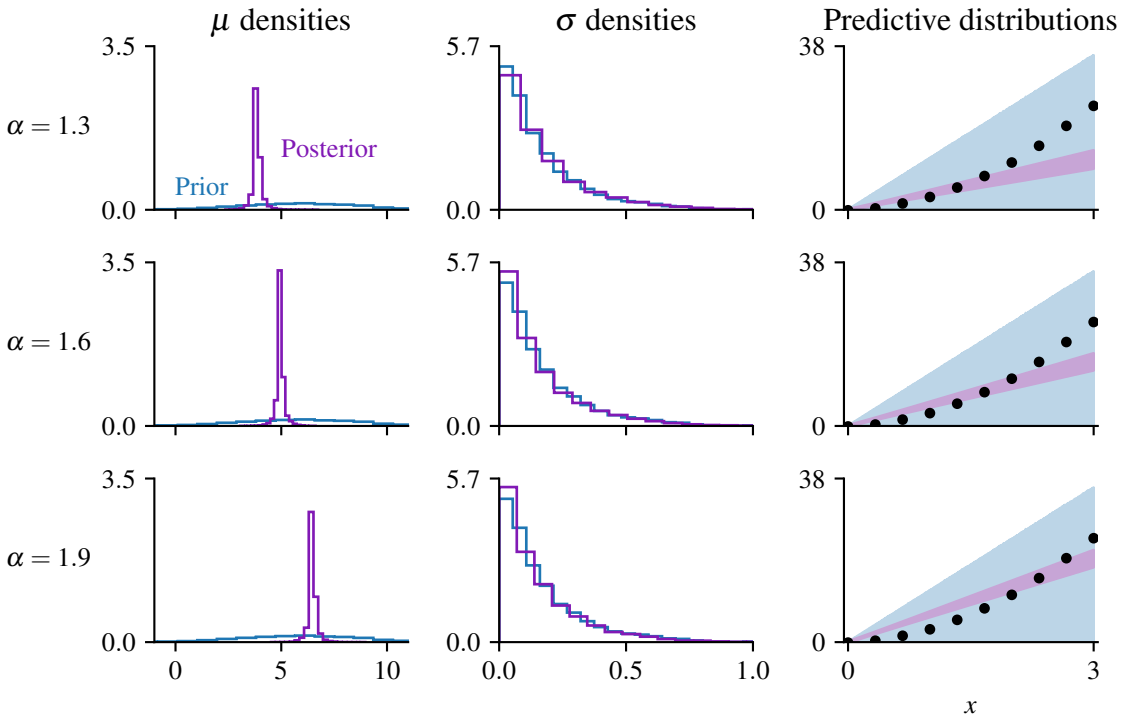

Figure 4-4. A comparison of prior and posterior marginal densities for μ and σ (left, center) and predictive densities (right).

4.2. Subsurface contaminant transport

As a model for subsurface contaminant transport through a heterogeneous porous medium, we pose the upscaled advection-diffusion equation:

$$\begin{aligned} \frac{\partial \langle c \rangle}{\partial t}(x,t) + \langle u \rangle \frac{\partial \langle c \rangle}{\partial x}(x,t) &= \nu_p \frac{\partial^2 \langle c \rangle}{\partial x^2}(x,t) - \frac{\partial \langle u'c' \rangle}{\partial x}(x,t), \\ \langle c \rangle (0,t) &= \langle c \rangle (L_x,t), \quad x \in [0, L_x], \\ \langle c \rangle (x,0) &= \exp\left(-\frac{(x-s)^2}{2\ell^2}\right), \end{aligned} \tag{4.1}$$

where $\langle c \rangle$ is the mean contaminant concentration, $\langle u \rangle$ is the mean velocity, $\nu_p \partial_{xx} \langle c \rangle$ is the pore-scale diffusion of the contaminant, and $\partial_x \langle u'c' \rangle$ is the *dispersion* resulting from unresolved fluctuations in the detailed velocity and concentration fields. The initial condition is a Gaussian pulse with

width $\ell = 0.1$ and mode s. We assume periodic boundary conditions, which is valid provided the velocity fluctuations are homogeneous with correlation lengths small compared to $L_x = 4$.

Some assumption must be made about the dispersion's mathematical form to close the equations, that is, to yield a set of governing equations that can be discretized and used to predict the contaminant's evolution in space and time. For heterogeneous porous media, the dispersion is known to have a nonlocal effect on the mean concentration, where the instantaneous concentration at a point x is influenced by the concentrations outside of a local region about x. However, the exact nature of this nonlocality is generally uncertain due to lack of detailed information about subsurface properties. This uncertainty motivates the introduction of a MFU representation for the dispersion, defined generally via a linear operator acting on $\langle c \rangle$:

$$\frac{\partial \langle u'c' \rangle}{\partial x} \approx -\mathcal{L} \langle c \rangle . \tag{4.2}$$

Under the mild assumption that $\mathcal{L}$ is shift invariant (does not depend on absolute location, just relative distances), the eigenfunctions of $\mathcal{L}$ are known to be the Fourier modes:

$$\mathcal{L} e^{ia_k} = \lambda_k e^{ia_k}, \quad a_k = \frac{2\pi k}{L_x}, \quad k \in \mathbb{Z},$$

where its eigenvalues λ_k may be complex valued, and Equation (4.1) admits a Fourier series solution. The action of $\mathcal{L}$ on $\langle c \rangle$ can thus be described entirely in terms of its eigenvalues, where real parts influence the magnitudes of the Fourier coefficients of the solution, while imaginary parts influence the phase [14]. Uncertainty in the dispersion can thus be represented through the eigenvalues of $\mathcal{L}$.

We consider two possible mathematical forms for $\mathcal{L}$. First, a general linear operator whose eigenvalues λ_k are informed from an ensemble of direct numerical simulations of the detailed advection-diffusion equation, as described in [9]:

$$\mathcal{L}_g[e^{ia_k}] = \lambda_k e^{ia_k}, \quad k = 1, \ldots, N_k. \tag{4.3}$$

Second, a Reisz fractional derivative operator, parameterized by its model coefficient ν and fractional power α:

$$\mathcal{L}_f[e^{ia_k}] = -\nu \frac{\partial^\alpha}{\partial x^\alpha}[e^{ia_k}] = -\nu (ia_k)^\alpha e^{ia_k}. \tag{4.4}$$

Fractional derivatives with non-integer powers are nonlocal in nature, which makes them attractive as a possible representation for dispersion. Note that the definition of eigenvalues for the fractional derivative implies the magnitude of its eigenvalues grows at a fixed rate α as a function of the index k. No such rate is imposed on the eigenvalues of the general linear operator, meaning it is more flexible than the fractional derivative in terms of the types of dynamics it can represent.

Our observable output for the Bayesian inverse problem is the concentration at $x = 1.4$ in the upstream half of the computational domain, collected for a short period of time, up to $t = 0.3$. The prediction QoI is the outflow concentration at $x = 4$ and $t = 1.5$. By selecting a prediction QoI that

is significantly downstream and in the future of when observations can be collected, this problem mimics the sorts of extrapolative prediction cases for which models are often needed to inform decisions.

To ensure concentration data is always positive and to represent uncertainty that increases as the observable increases in magnitude, we assume multiplicative measurement error, where the measurement error is assumed to be log-normally distributed with its median at 1:

$$d_i = \epsilon_m^{(i)} \langle c \rangle (1.4, t_i), \quad \log(\epsilon_m^{(i)}) \sim \mathcal{N}(0, (10^{-4})^2),$$

where t_i is a vector of 10 equally spaced time points between 0 and 0.3, excluding 0. The data-generating model sets $s = 0.9$, $\langle u \rangle = 1.05$, and $\nu_p = 9.5 \times 10^{-3}$, while the model used in inference sets the parameters to their nominal values, $s = 1$, $\langle u \rangle = 1$, and $\nu_p = 10^{-2}$.

We take as our MFU representation the fractional derivative and hyperparameterize ν and α. We have some prior knowledge about these parameters, which we can use to pose their distributions. First, we know $\nu > 0$ and that dispersion is influential; we are reasonably confident it falls above 0.1 (corresponding to 10x the pore-scale diffusion coefficient) and highly confident it falls below 0.5. Probabilistically this can be mapped to the statements

$$\begin{aligned} P(\nu \leq 0.1) &\approx 0.1, \\ P(\nu \leq 0.5) &\approx 0.99. \end{aligned} \tag{4.5}$$

This prior information can be encoded into the distribution for ν by representing it as a log-normal random variable whose hyperparameters are set by the probabilistic constraints in (4.5): $\pi(\nu) = \log \mathcal{N}(m_\nu, s_\nu)$. To define a distribution for α we know that $\alpha \in [1, 2]$ and want to allow for the hyperparameterized distribution to concentration about a mode. We therefore represent it as a truncated normal random variable defined on $[1, 2]$: $\pi(\alpha) = \text{trunc}\mathcal{N}(m_\alpha, s_\alpha; [1, 2])$.

We now demonstrate the challenges associated with defining prior distributions for the hyperparameters m_ν, s_ν, m_α, and s_α. Unfortunately, we have very limited information with which to constrain their prior distributions. For m_ν and s_ν we can use (4.5) to define "nominal" values m_ν^n and s_ν^n. However, the degree of uncertainty to impose about these nominal values is unclear. Since m_ν needn't be positive we define $m_\nu \sim \mathcal{N}(m_\nu^n, (0.5|m_\nu^n|)^2)$, where we choose $0.5|m_\nu^n|$ somewhat arbitrarily to indicate significant uncertainty about the nominal value. We know s_ν is positive, so we model it as log-normally distributed and parameterize it so that its mode is at s_ν^n and its value falls below $1.5 s_\nu$ with probability 0.99. Again the degree of uncertainty imposed on this hyperprior is somewhat arbitrary but aimed to express significant uncertainty.

For m_α, we expect the mean of α to be located away from the boundaries such that it yields nonlocal behavior, and we expect the mean to fall within $[1, 2]$; therefore, we assume $m_\alpha \sim \mathcal{N}(1.5, (0.25)^2)$ which yields a 95% confidence interval within the range. While we don't have any strong reason to believe that 1.5 is the most probable value for α, when we pose a less informative hyperprior on m_α, it results in a concentration of probability at the boundaries of α's distribution, which was nonphysical. Finally, lacking any information about s_α besides that it is positive, we represent it as an exponential random variable with a scale parameter of 1.

As this description of our procedure to define hyperpriors for this problem shows, it is nontrivial to incorporate prior information into the distributions, and often there is little or no information to incorporate at all. As a result, the hyperprior definitions can involve somewhat arbitrary choices. Since scale hyperparameters are often uninformed by the data in Bayesian inference, these arbitrary choices will likely persist and manifest in posterior predictive distributions, which significantly impacts the meaningfulness of the resulting uncertainty characterizations. This effect might be mitigated by performing robustness studies, where the stability of the solution to the Bayesian inverse problem is studied as the hyperpriors are varied. However, since even a single Bayesian inverse problem can be highly computationally intensive, such robustness studies are often infeasible in practice.

We now study the MFU characterizations achieved in the case where the model enriched with the MFU representation can reproduce the data-generating model (i.e., the data-generating model employs the fractional derivative to represent dispersion). In this case, there is no model-form error, and ideally this would be reflected in the MFU characterization by a significant contraction of posterior uncertainty. However, as is shown in Figure 4-5, only m_ν exhibits a substantial difference between prior and posterior. While m_α is weakly informed, neither scale parameter s_ν nor s_α are informed at all, compared to what one might intuitively expect in this context, which is that they would contract significantly. The predictive distributions for the outflow concentration QoI and for the calibration observable contracted somewhat, but considerable uncertainty remains, even when significantly more data is incorporated into the inference problem, as shown in Figure 4-6. Thus, the hierarchical approach may overestimate uncertainties when no model-form error is present.

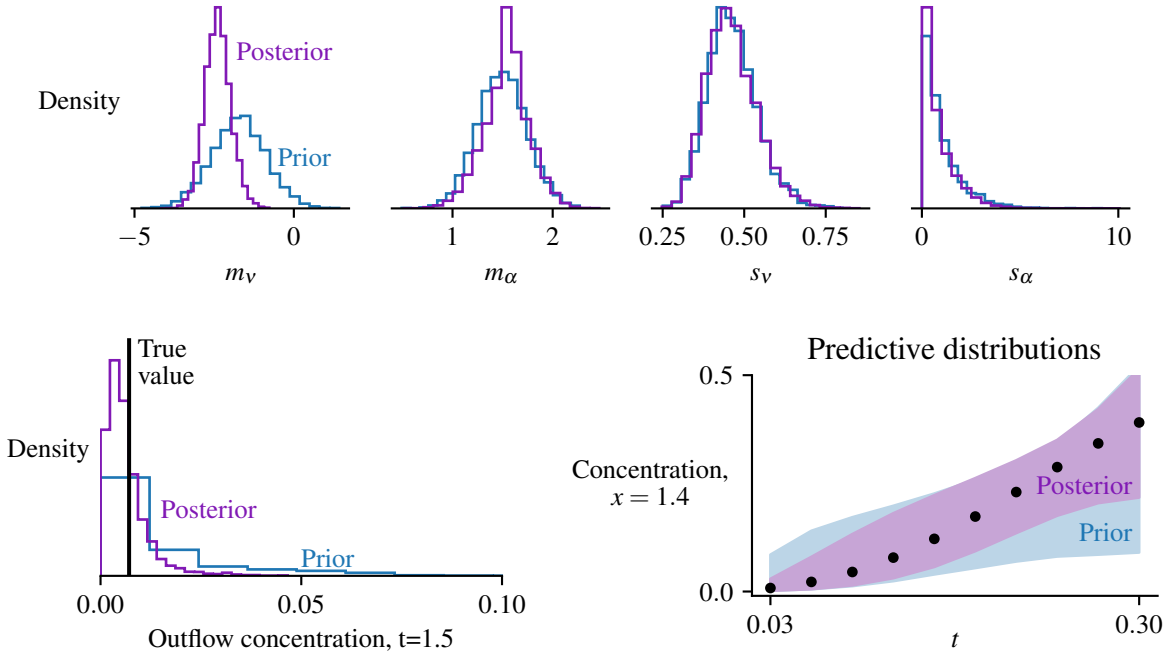

Figure 4-5. Marginal prior and posterior densities for (top) hyperparameters and (bottom left) outflow concentration QoI. Bottom right: 95% high-probability region for prior and posterior predictive distributions based on 0.025 and 0.975 quantiles compared to calibration data (black). Data was produced using the fractional derivative to represent dispersion.

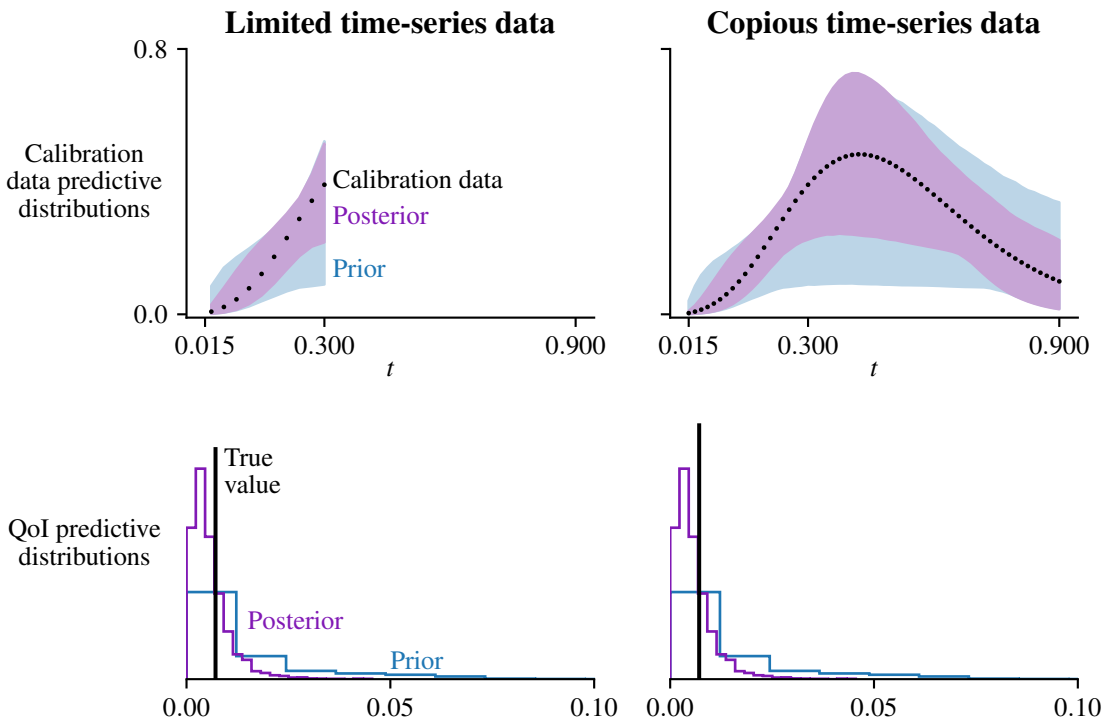

Figure 4-6. Calibration data predictive distributions (top) and outflow concentration QoI predictive distributions (bottom). Limited data (left) is gathered to time 0.3 with 10 data points. Copious data (right) is gathered to time 0.9 with 60 data points (a doubling in time resolution). Increased data did not substantively modify the predictive distributions.

We now consider the case where the model enriched with an MFU representation is unable to reproduce the data, which is generated using the general linear operator with eigenvalues from DNS to represent dispersion. Ideally, the MFU characterization in this case would assign high probability to the calibration data without being too disperse. As is shown in Figure 4-7, once again the scale parameters were not informed by the calibration data. Additionally, while the posterior predictive distribution for the outflow concentration QoI assigns high probability to the true value, its support has not significantly contracted relative to the prior, so a decision based on the range of plausible values for the concentration would not be better informed after calibration. Similarly, while the posterior predictive distribution for calibration observables has contracted somewhat relative to the prior, it still assigns high probability to a wide range of values for the concentration which fall far away from the calibration data, meaning it is overly dispersed relative to what might be desired in practice. While the hierarchical approach to MFU can update the MFU characterization, this example shows that it does not necessarily yield a *useful* characterization.

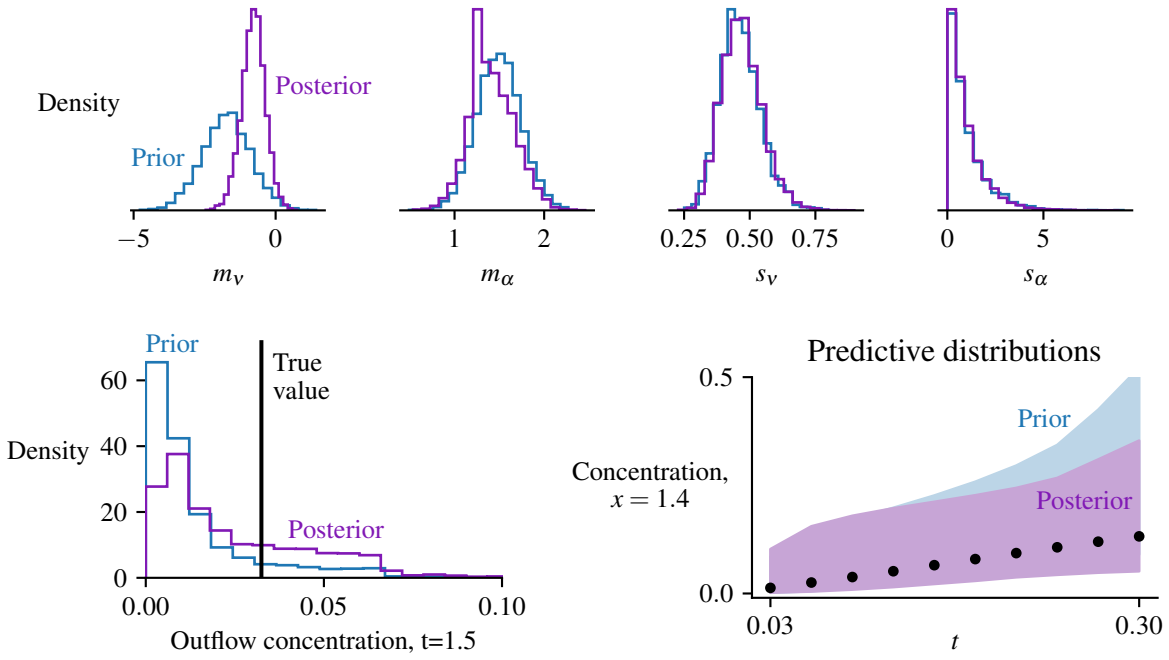

Figure 4-7. Marginal prior and posterior densities for (top) hyperparameters and (bottom left) outflow concentration QoI. Bottom right: 95% high-probability region for prior and posterior predictive distributions based on 0.025 and 0.975 quantiles compared to calibration data (black). Data was produced using the general linear operator to represent dispersion.

This page intentionally left blank.

5. CONCLUSIONS

This report focuses on the hierarchical Bayesian approach to informing model-form uncertainty (MFU) characterizations, which poses a hierarchical Bayesian inverse problem for the hyperparameters of MFU parameter distributions with the aim of identifying the probability distributions that best agree with calibration data. However, we have encountered several theoretical and methodological challenges while applying this approach, which are documented herein. Theoretical challenges include the lack of information with which to inform prior distributions for the hyperparameters, and the need to specify a likelihood, which implicitly assumes the model is capable of representing the data-generating process. Methodological challenges include an increase in the dimension of the inverse problem resulting from the fact that there are significantly more (often double) hyperparameters than model parameters. Furthermore, scale parameters are often uninformed by inference, except in undesirable degenerate cases. This latter challenge implies that in many cases there may be a limit to which the MFU representation can be "calibrated".

Overall, we show that the hierarchical Bayesian approach is not a turnkey solution to MFU characterization and can often yield nonintuitive results that are counter to the goals of informing MFU: namely, attaining uncertainty bounds that agree with data without being so broad as to be useless for decision making. Our findings indicate that alternative approaches to inform MFU characterizations are needed to meet the needs of practical engineering problems. The approach presented in [6] is a promising option. It augments MFU parameters with random perturbations (i.e., $\tilde{\theta} = \theta + \delta$ where δ is a random variable) and calibrates the parameters and perturbation distribution jointly. However, it maintains some of the challenges of the hierarchical approach. The inverse problem dimension is increased by including the random perturbations, which may introduce identifiability issues. Additionally, constraints on δ may be complicated by interactions with θ to determine feasible values for $\tilde{\theta}$.

Generalized Bayesian inference is a conceptually simple alternative that does not hyperparameterize the inputs but instead relaxes the assumption that the model can perfectly represent the data generating process. The resulting posteriors, often known as Gibbs posteriors [15], can be constructed through approaches such as tempering the likelihood or replacing it altogether with loss functions that are more robust to model misspecification. Note that related numerical approaches, such as approximate Bayesian computation [16], can also be employed to provide samples from a generalized posterior. Despite the promise of more generalized inference approaches capable of accounting for prediction goals, there does not exist an established way of accounting for uncertainty characterization goals—e.g., the predictive uncertainties should encompass the data without being too disperse. Methods that build off generalized Bayesian inverse problems and explicitly incorporate these goals could yield a simple yet effective method to inform MFU characterizations.

It is important to note that agreement of predictive distributions with calibration data is not sufficient to deem MFU characterization as meaningful for the purposes of extrapolative predictions.

Additional analyses that assess how well uncertainties extrapolate toward the target prediction are critical to build confidence in the characterization, e.g., the procedures for validation of extrapolative predictions under uncertainty as described in [7]. Hence, more work is needed to establish approaches capable of meaningfully characterizing uncertainties for the purpose of extrapolative predictions.

REFERENCES

[1] B.J.K. Kleijn and A.W. van der Vaart. The Bernstein-Von-Mises theorem under misspecification. *Electronic Journal of Statistics*, 6(none):354–381, January 2012. doi:10.1214/12-EJS675.

[2] Marc C. Kennedy and Anthony O'Hagan. Bayesian calibration of computer models. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 63(3):425–464, January 2001. doi:10.1111/1467-9868.00294.

[3] Jenný Brynjarsdóttir and Anthony O'Hagan. Learning about physical parameters: the importance of model discrepancy. *Inverse Problems*, 30(11):114007, October 2014. Publisher: IOP Publishing. URL: http://dx.doi.org/10.1088/0266-5611/30/11/114007, doi:10.1088/0266-5611/30/11/114007.

[4] Kathryn A. Maupin and Laura P. Swiler. Model discrepancy calibration across experimental settings. *Reliability Engineering & System Safety*, 200:106818, August 2020. URL: https://www.sciencedirect.com/science/article/pii/S0951832019301802, doi:10.1016/j.ress.2020.106818.

[5] K Sargsyan, HN Najm, and R Ghanem. On the statistical calibration of physical models. *International Journal of Chemical Kinetics*, 47(4):246–276, 2015. Publisher: Wiley Online Library. URL: https://doi.org/10.1002/kin.20906, doi:10.1002/kin.2090.

[6] Khachik Sargsyan, Xun Huan, and Habib N. Najm. Embedded model error representation for Bayesian model calibration. *International Journal for Uncertainty Quantification*, 9(4):365–394, 2019. URL: https://arxiv.org/pdf/1801.06768, doi:10.1615/int.j.uncertaintyquantification.2019027384.

[7] Todd A. Oliver, Gabriel Terejanu, Christopher S. Simmons, and Robert D. Moser. Validating predictions of unobserved quantities. *Computer Methods in Applied Mechanics and Engineering*, 283:1310–1335, January 2015. URL: https://www.sciencedirect.com/science/article/pii/S004578251400293X, doi:10.1016/j.cma.2014.08.023.

[8] Rebecca E. Morrison, Todd A. Oliver, and Robert D. Moser. Representing Model Inadequacy: A Stochastic Operator Approach. *SIAM/ASA Journal on Uncertainty Quantification*, 6(2):457–496, January 2018. Publisher: Society for Industrial and Applied Mathematics. doi:10.1137/16M1106419.

[9] Teresa Portone. *Representing model-form uncertainty from missing microstructural information*. PhD thesis, University of Texas at Austin, 2019. URL: http://dx.doi.org/10.26153/tsw/10112.

[10] Rileigh Bandy and Rebecca Morrison. Quantifying Model Form Uncertainty in Spring-Mass-Damper Systems. In Roland Platz, Garrison Flynn, Kyle Neal, and Scott Ouellette, editors, *Model Validation and Uncertainty Quantification, Volume 3*, pages 9–19, Cham, 2024. Springer Nature Switzerland.

[11] R. Bandy and R. Morrison. Stochastic model corrections for reduced Lotka–Volterra models exhibiting mutual, commensal, competitive, and predatory interactions. *Chaos: An Interdisciplinary Journal of Nonlinear Science*, 34(1):013116, January 2024. doi: 10.1063/5.0159043.

[12] Rileigh Bandy, Teresa Portone, and Rebecca Morrison. Stochastic Model Correction for the Adaptive Vibration Isolation Round-Robin Challenge. In Roland Platz, Garrison Flynn, Kyle Neal, and Scott Ouellette, editors, *Model Validation and Uncertainty Quantification, Vol. 3*, pages 53–62, Cham, 2025. Springer Nature Switzerland. doi: 10.1007/978-3-031-68893-5_8.

[13] Teresa Portone, Rebekah White, and Joseph L. Hart. Quantifying model prediction sensitivity to model-form uncertainty. *In preparation*.

[14] Teresa Portone and Robert D. Moser. Bayesian Inference of an Uncertain Generalized Diffusion Operator. *SIAM/ASA Journal on Uncertainty Quantification*, pages 151–178, February 2022. Publisher: Society for Industrial and Applied Mathematics. doi: 10.1137/21M141659X.

[15] P. G. Bissiri, C. C. Holmes, and S. G. Walker. A general framework for updating belief distributions. *Journal of the Royal Statistical Society. Series B (Statistical Methodology)*, 78(5):1103–1130, 2016. Publisher: [Royal Statistical Society, Wiley].

[16] Jean-Michel Marin, Pierre Pudlo, Christian P. Robert, and Robin J. Ryder. Approximate Bayesian computational methods. *Statistics and Computing*, 22:1167–1180, November 2012. doi:10.1007/s11222-011-9288-2.

DISTRIBUTION

Email—Internal

Name	Org.	Sandia Email Address
Technical Library	1911	sanddocs@sandia.gov

Hardcopy—Internal

Number of Copies	Name	Org.	Mailstop
1	L. Martin, LDRD Office	1910	0359

Sandia National Laboratories is a multimission laboratory managed and operated by National Technology & Engineering Solutions of Sandia LLC, a wholly owned subsidiary of Honeywell International Inc., for the U.S. Department of Energy's National Nuclear Security Administration under contract DE-NA0003525.