International Journal for Uncertainty Quantification, 15(2):65–90 (2025)

VARIANCE-BASED SENSITIVITY OF BAYESIAN INVERSE PROBLEMS TO THE PRIOR DISTRIBUTION

John E. Darges,[1,*] Alen Alexanderian,[1] & Pierre A. Gremaud[1,2]

[1]*Department of Mathematics, North Carolina State University, Raleigh, North Carolina, 27607, USA*

[2]*The Graduate School, North Carolina State University, Raleigh, North Carolina, 27607, USA*

*Address all correspondence to: John E. Darges, North Carolina State University, Raleigh, North Carolina, 27607, USA, E-mail: jedarges@ncsu.edu

The formulation of Bayesian inverse problems involves choosing prior distributions; choices that seem equally reasonable may lead to significantly different conclusions. We develop a computational approach to understand the impact of the hyperparameters defining the prior on the posterior statistics of the quantities of interest. Our approach relies on global sensitivity analysis (GSA) of Bayesian inverse problems with respect to the prior hyperparameters. This, however, is a challenging problem—a naive double loop sampling approach would require running a prohibitive number of Markov chain Monte Carlo (MCMC) sampling procedures. The present work takes a foundational step in making such a sensitivity analysis practical by combining efficient surrogate models and a tailored importance sampling approach. In particular, we can perform accurate GSA of posterior statistics of quantities of interest with respect to prior hyperparameters without the need to repeat MCMC runs. We demonstrate the effectiveness of the approach on a simple Bayesian linear inverse problem and a nonlinear inverse problem governed by an epidemiological model.

KEY WORDS: *prior hyperparameters, global sensitivity analysis, Sobol' indices, Bayesian inverse problems, importance sampling, surrogate modeling*

1. INTRODUCTION

Consider a Bayesian inverse problem governed by a system of differential equations. The inverse problem uses a vector $\boldsymbol{d}$ of measurement data to estimate the uncertain model parameters, θ. The solution of the Bayesian inverse problem is the posterior distribution $\pi_{\text{post}}(\theta|\boldsymbol{d})$. After solving the inverse problem, typically we seek to make some predictions based on the posterior. For example, for a prediction quantity $q(\theta)$ we may consider

$$\mathbb{E}_{\text{post}}(q) := \int q(\theta)\pi_{\text{post}}(\theta|\boldsymbol{d})\, d\theta.$$

A crucial component of this analysis is to know how the choice of prior hyperparameters affects such predictions. We present a practical variance-based global sensitivity analysis (GSA) approach to study how the statistics (e.g., mean or variance) of q vary with respect to prior hyperparameters. The global sensitivity indices associated to each prior hyperparameter enable us to identify which hyperparameters carry the most influence over the prediction.

Bayesian inference is pervasive; this perspective makes inferences not just from data, but also by incorporating prior beliefs and assumptions. In practice, these prior assumptions are often subjective choices made by the researcher. Different prior assumptions can result in differing outcomes from inference in Bayesian inverse problems [1]. This well-known issue motivated statisticians in the 1980s and 1990s to develop a methodology, known as robust Bayesian analysis or Bayesian sensitivity analysis [2–5], for ensuring the robustness of Bayesian inference across a class of priors. These ideas have continued to receive attention over the past two decades [6–11]. The goals of robust Bayesian analysis are related to, but distinct from those of our approach. Rather than answering whether our inference is sensitive to prior assumptions, we aim to characterize the uncertainty distribution resulting from uncertainty in the prior hyperparameters and identify which hyperparameters have the greatest impact on uncertainty.

Submitted: 10/29/23; Accepted: 6/12/24; Online: 8/14/24 https://doi.org/10.1615/Int.J.UncertaintyQuantification.2024051475
2152–5080/25/$35.00

Related work. Sensitivity analysis of Bayesian inverse problems has been subject to several recent research efforts. The articles [12–14] consider hyperdifferential sensitivity analysis (HDSA) of Bayesian inverse problems. HDSA is a technique used originally for (deterministic) PDE-constrained optimization problems. HDSA, as a practical framework for sensitivity analysis of optimal control problems governed by PDEs, was considered in [15]. In [16], HDSA was used for sensitivity analysis of deterministic inverse problems to auxiliary model parameters and parameters specifying the experimental setup (experimental parameters). In [12], use of HDSA is extended to nonlinear Bayesian inverse problems. Specifically, the authors consider the Bayes risk and the maximum a posterior probability (MAP) point as quantities of interest for sensitivity analysis.

In [13], the HDSA framework is used to study Bayesian inverse problems governed by ice sheet models. The sensitivity of information gain, measured by the Kullback–Leibler (KL) divergence between the prior and posterior, to uncertain model parameters in linear Bayesian inverse problems is studied in [14]. HDSA provides valuable insight for experimenters on where to focus resources during experimental design and when measuring auxiliary parameters. The previous works on HDSA of Bayesian inverse problems have focused primarily on sensitivity analysis with respect to auxiliary or experimental rather than prior hyperparameters. More importantly, HDSA is local, relying on derivative information evaluated at a set of nominal parameters. Variance-based GSA, see Section 3, accounts for the uncertainty in the hyperparameters globally.

The work of [17], which is closely related to our work, examines statistical models using Bayesian inference. In that paper, the authors perform variance-based GSA on posterior statistics with respect to both prior and likelihood hyperparameters. Their method uses Gaussian process (GP) surrogates to emulate the mapping from the hyperparameters to the posterior distribution. This requires many Markov chain Monte Carlo (MCMC) runs to build the GP surrogate. For the Bayesian inverse problems we target, it is impractical to repeat the many MCMC runs needed to build a surrogate. This is due to the high cost of evaluating the forward model and the difficulty of tuning each MCMC run. Our interest in the influence of prior assumptions means we focus only on the prior hyperparameters. Restricting to prior hyperparameters avoids the challenges associated with multiple MCMC runs by enabling an importance sampling approach to integrating the QoIs under study with respect to multiple posterior distributions. Strategies for importance sampling on multiple distributions have been subject to several previous works; see e.g., [18–22]. We exploit structure in the Bayesian inverse problem to derive a tailored importance sampling approach. Another related work that has partly inspired the approach in the present work is [23]. That article outlines a method for GSA of rare event probabilities and combines surrogate-assisted GSA with subset simulation.

Our approach and contributions. We present a viable computational approach to analyze the sensitivity of Bayesian inverse problems to prior hyperparameters. The proposed approach is goal oriented—the focus is on the posterior statistics of prediction/goal QoIs that are functions of the inversion parameters. We first frame the problem in a manner conducive to variance-based GSA in Section 2. We detail the computational strategy for sensitivity analysis in Section 4. Our method combines two key techniques. Importance sampling eliminates the need for repeated MCMC runs for different choices of the prior. Then, sparse polynomial chaos expansion (PCE) and extreme learning machine (ELM) surrogate models emulate the mapping from prior hyperparameters to statistics of q. Use of surrogate models further eases the computational burden. The combined approach enables prior hyperparameter sensitivity analysis for many Bayesian inverse problems. If one has access to a single MCMC run, then one can ascertain prior hyperparameter importance. To demonstrate the effectiveness of the proposed approach, we present computational experiments in the context of two examples: a simple linear inverse problem in Section 5.1 and a nonlinear inverse problem governed by an epidemiological model in Section 5.2.

2. HYPERPARAMETER-TO-STATISTIC MAPPING OF BAYESIAN INVERSE PROBLEMS

In an inverse problem [24], we use a model and observed data to estimate unknown model parameters of interest. We consider the inverse problem of estimating a parameter vector $\boldsymbol{\theta}$ in models of the form

$$\begin{cases} \boldsymbol{y}' = f(\boldsymbol{y}; \boldsymbol{\theta}), \\ \boldsymbol{y}(t_0) = \boldsymbol{y}_0. \end{cases} \tag{1}$$

Here, $\boldsymbol{y} \in \mathbb{R}^{n_{\text{state}}}$ is the state vector. In a deterministic formulation of the inverse problem, we typically seek a $\boldsymbol{\theta} \in \mathbb{R}^n$ that minimizes the cost functional,

$$J(\boldsymbol{\theta}) := \|\mathbf{B}\boldsymbol{y}(\boldsymbol{\theta}) - \boldsymbol{d}\|^2. \tag{2}$$

Here, $\boldsymbol{d} \in \mathbb{R}^{n_{\text{data}}}$ is a vector of data measurements, $\mathbf{B}$ is a linear operator that selects the corresponding model responses, and $\boldsymbol{y}$ is obtained by solving Eq. (1).

Real-world measurements are always subject to some observation error. The Bayesian approach to inverse problems [24] describes the inversion parameters through a statistical distribution that takes uncertainty into account. This distribution for $\boldsymbol{\theta}$, known as the posterior distribution, is conditioned on the observed data and consistent with the prior distribution. In this context, the prior distribution encodes our prior knowledge regarding the parameters. The Bayes formula shows how the model, data, and the prior are combined to obtain the posterior distribution:

$$\pi_{\text{post}}(\boldsymbol{\theta}|\boldsymbol{d}) \propto \pi_{\text{like}}(\boldsymbol{d}|\boldsymbol{\theta}) \times \pi_{\text{pr}}(\boldsymbol{\theta}), \tag{3}$$

where π_{like} is the data likelihood and π_{pr} is the prior probability density function (PDF). The likelihood of the data is assessed according to the cost (2) and the noise model that describes the error distribution for each measurement. For example, suppose the observation error follows a Gaussian noise model so that $\mathbf{B}\boldsymbol{y}(\boldsymbol{\theta}) + \boldsymbol{\eta} = \boldsymbol{d}, \boldsymbol{\eta} \sim \mathcal{N}(0, \Gamma_{\text{noise}})$. In this case, the Bayes formula reads

$$\pi_{\text{post}}(\boldsymbol{\theta}|\boldsymbol{d}) \propto \exp\left(-\frac{1}{2}(\mathbf{B}\boldsymbol{y}(\boldsymbol{\theta}) - \boldsymbol{d})^\top \Gamma_{\text{noise}}^{-1}(\mathbf{B}\boldsymbol{y}(\boldsymbol{\theta}) - \boldsymbol{d})\right) \times \pi_{\text{pr}}(\boldsymbol{\theta}). \tag{4}$$

In practice, we are often interested in scalar prediction quantities of interest (QoIs) that depend on $\boldsymbol{\theta}$. Let $q(\boldsymbol{\theta})$ be such a QoI. Solving the Bayesian inverse problem enables reducing the uncertainty in $\boldsymbol{\theta}$ and consequently in $q(\boldsymbol{\theta})$. In this case, the statistical properties of q depend on π_{post}. Let $\Psi(q)$ denote a generic statistic of q. Examples include $\Psi(q) = \text{var}(q)$ or $\Psi(q) = \mathbb{E}(q)$, where the expectation and variance are with respect to the posterior distribution. Another example is $\Psi(q) = q(\boldsymbol{\theta}_{\text{MAP}})$; i.e., QoI evaluated at the maximum a posteriori (MAP) point estimate of $\boldsymbol{\theta}$. Recall that the MAP point, $\boldsymbol{\theta}_{\text{MAP}}$, is a point where the posterior PDF attains its maximum value. Using the Bayes formula (4), we note that the MAP point is the solution to the nonlinear least squares problem,

$$\boldsymbol{\theta}_{\text{MAP}} = \underset{\boldsymbol{\theta}}{\text{argmin}}\ J(\boldsymbol{\theta}) := (\mathbf{B}\boldsymbol{y}(\boldsymbol{\theta}) - \boldsymbol{d})^\top \Gamma_{\text{noise}}^{-1}(\mathbf{B}\boldsymbol{y}(\boldsymbol{\theta}) - \boldsymbol{d}) - 2\log(\pi_{\text{pr}}(\boldsymbol{\theta})). \tag{5}$$

We consider how the choice of prior affects $\Psi(q)$. Narrowing this question, we take a parameterized family of prior distributions $\pi_{\text{pr}}^{\boldsymbol{\xi}}(\boldsymbol{\theta})$ determined by a vector $\boldsymbol{\xi} \in \mathbb{R}^d$ of scalar hyperparameters. For a Gaussian prior, the hyperparameters can be taken as the prior means and variances. With this setup, the choice of $\boldsymbol{\xi}$ will determine our statistic of interest so that $\Psi(q) = \Psi^{\boldsymbol{\xi}}(q)$. In what follows, the *hyperparameter-to-statistic (HS) mapping* $F : \mathbb{R}^d \to \mathbb{R}$ is given by

$$F(\boldsymbol{\xi}) := \Psi^{\boldsymbol{\xi}}(q). \tag{6}$$

To model the uncertainty in the hyperparameters, we consider them as random variables and then analyze how the uncertainty in the entries of $\boldsymbol{\xi}$ contributes to the uncertainty in $F(\boldsymbol{\xi})$. To this end, we follow a variance-based sensitivity analysis framework, and compute the Sobol' indices [25,26] of the HS mapping F with respect to $\boldsymbol{\xi}$.

For the purposes of this study, we let the prior hyperparameters $\boldsymbol{\xi}$ follow uniform distributions, $\xi_j \sim \mathcal{U}[a_j, b_j]$, for $j = 1, \ldots, d$. We focus on three choices for the statistic of interest Ψ in Eq. (6):

- the mean: $F_{\text{mean}}(\boldsymbol{\xi}) = \mathbb{E}_{\text{post}}^{\boldsymbol{\xi}}(q)$;
- the variance: $F_{\text{var}}(\boldsymbol{\xi}) = \mathbb{E}_{\text{post}}^{\boldsymbol{\xi}}(q^2) - (\mathbb{E}_{\text{post}}^{\boldsymbol{\xi}}(q))^2$; and
- the QoI evaluated at the MAP point: $F_{\text{atMAP}}(\boldsymbol{\xi}) := q(\boldsymbol{\theta}_{\text{MAP}}(\boldsymbol{\xi}))$, with $\boldsymbol{\theta}_{\text{MAP}}(\boldsymbol{\xi})$ from Eq. (5).

The mean and variance are computed from moments of the posterior PDF. These two quantities can be estimated at each $\boldsymbol{\xi}$ by Monte Carlo integration. Estimating F_{atMAP} instead requires solving the nonlinear least-squares problem (5) for each $\boldsymbol{\xi}$.

3. GLOBAL SENSITIVITY ANALYSIS AND SURROGATE-ASSISTED APPROACHES

We focus on variance-based GSA using Sobol' indices [25–30]. Consider a (scalar-valued) model

$$y = F(\boldsymbol{x}), \quad \boldsymbol{x} \in \mathbb{R}^d.$$

We assume that the components of $\boldsymbol{x}$ are independent random variables. In variance-based GSA, the most important inputs are those that contribute the most to the output variance $\text{var}(F(\boldsymbol{x}))$. Sobol' indices are quantitative measures of this contribution. Specifically, the first-order Sobol' indices S_k, and the total Sobol' indices S_k^{tot}, are defined by

$$S_k = \frac{\text{var}(F_k)}{\text{var}(F)}, \quad S_k^{\text{tot}} = 1 - \frac{\text{var}(\mathbb{E}(F|x_l, l \neq k))}{\text{var}(F)}, \tag{7}$$

where $F_k(x_k) := \mathbb{E}(f|x_k) - \mathbb{E}(f)$. In practice, the Sobol' indices are approximated by pick-freeze estimators [31]. The strategy for computing these estimators uses Monte Carlo sampling and requires evaluating the model at many sample points. This can be too costly, especially when the model F is expensive to evaluate. In such cases, it is common practice to construct a surrogate model $\widehat{F} \approx F$ whose Sobol' indices can be efficiently computed [32,33]. In the best case scenario, the Sobol' indices of the surrogate model can be computed analytically. We detail two such surrogate models below.

Polynomial chaos surrogates. Polynomial chaos expansions (PCEs) take advantage of orthogonal polynomials to approximate expensive-to-evaluate models; see [34,35]. The standard approach is to truncate the PCE based on the total polynomial degree. PCE surrogates are advantageous because they admit analytic formulas for Sobol' indices that depend only on the PC coefficients [34]. In practice, the PC coefficients are typically computed using nonintrusive approaches that involve sampling the model F. These include nonintrusive spectral projection or regression-based methods [36]. In the present work, we build PCE surrogates using sparse regression [37,38]. As noted in [23], this approach is particularly useful in the case where function evaluations are noisy. Solving the sparse regression problem can be formulated as a linear least-squares problem regularized by an ℓ^1-penalty [39,40]. In our numerical computations, we use the SPGL1 solver [41,42] to solve such problems. Note that an ℓ^1-penalty approach also involves choosing a penalty parameter. In our experiments, we perform a tenfold cross validation over training sets to choose the ℓ^1-penalty parameters.

Sparse weight-ELM surrogates. Sparse weight extreme learning machines (SW-ELMs), introduced in [43], are a class of neural network surrogates that build on the standard extreme learning machines (ELMs). These surrogates are closely related to random feature expansions, which have been proposed for use in uncertainty quantification [44]. SW-ELM surrogates are specialized for applications in global sensitivity analysis. They are single-layer neural networks of the form

$$\widehat{F}(\boldsymbol{x}) = \boldsymbol{\beta}^\top \phi(\mathbf{W}\boldsymbol{x} + \boldsymbol{b}), \quad \boldsymbol{x} \in \mathbb{R}^d. \tag{8}$$

Here, $\boldsymbol{\beta}$ denotes the output weight vector, $\mathbf{W}$ the hidden layer weight matrix, $\boldsymbol{b}$ the hidden layer bias vector, and ϕ the activation function. Whereas with traditional neural networks, all weights and biases are trained at once by solving a nonlinear least-squares problem, ELMs instead randomly choose the hidden layer weights and biases. Training an ELM then only involves determining the output layer weights by solving a linear least-squares problem; see [45,46] for details. This means training ELMs is very simple and fast. SW-ELM modifies the weight sampling step of standard ELM to improve performance for GSA. The method introduces a validation step to choose a sparsification parameter p. Similar to PCE, analytical formulas for Sobol' indices of SW-ELM, as shown in [43], can be derived for specific choice of activation function. Unlike PCE surrogates, which are expansions over an orthogonal basis, SW-ELM surrogates are expansions over random nonorthogonal bases. SW-ELM surrogates are quick and easy to train, but are also structurally different from PCE surrogates. In our numerical results, we compare the Sobol' index estimates from both surrogates to gain further confidence in the results. For the SW-ELM surrogates used in our experiments, the number of neurons used is half the number of training points. A fraction of the training points are used as a validation set to choose the sparsification parameter. See [43] for further details.

4. METHOD

In this section, we outline our proposed approach for GSA of hyperparameter-to-statistic (HS) mappings of the form Eq. (6). Our focus will be mainly on HS mappings that involve integrating over the posterior. Examples are the posterior mean or variance. For simplicity, we focus on

$$F(\boldsymbol{\xi}) = \mathbb{E}^{\boldsymbol{\xi}}_{\text{post}}(q) = \int_{\mathbb{R}^d} q(\boldsymbol{\theta})\pi^{\boldsymbol{\xi}}_{\text{post}}(\boldsymbol{\theta})\, d\boldsymbol{\theta}. \tag{9}$$

We generalize the strategies described below to the cases of variance and higher-order moments later in this section.

Estimating the Sobol' indices of Eq. (9) is generally challenging. Computing $F(\boldsymbol{\xi})$ using samples from the posterior distribution of $\boldsymbol{\theta}$ often requires a MCMC method. With MCMC, the mean ergodic theorem guarantees convergence to $F(\boldsymbol{\xi})$ as we take more MCMC samples as long as the Markov kernel is ergodic [47]. A naive approach for computing the Sobol' indices of $F(\boldsymbol{\xi})$ would be to follow a sampling procedure where an MCMC simulation is carried out for each realization of $\boldsymbol{\xi}$. This is typically infeasible. For one thing, the computational cost of this naive approach will be prohibitive for most practical problems. In addition, performing multiple runs of an MCMC algorithm can be problematic, because such methods typically have algorithm-specific parameters that might need tuning for different realizations of $\boldsymbol{\xi}$.

In Section 4.1, we outline an approach that combines MCMC and importance sampling for fast computation of moment-based HS mappings under study. Then, in Section 4.2, we present an algorithm that combines the approach in Section 4.1 and surrogate models to facilitate GSA of moment-based HS maps. In that section, we also discuss the computational cost of the proposed approach, in terms of the number of required forward model evaluations. We also briefly discuss GSA of F_{atMAP} in Section 4.3.

4.1 Importance Sampling for Fast Evaluation of Moment-Based HS Maps

Importance sampling [22,48] aims at accelerating the computation of integrals such as Eq. (9), where the target distribution $\pi^{\boldsymbol{\xi}}_{\text{post}}$ is difficult to sample from. This is done by introducing an importance sampling distribution $\pi_{\text{IS}}(\boldsymbol{\theta})$, which is tractable to work with, and from which we are likely to sample points where the target posterior distribution takes high density.

Let π_{IS} be an importance sampling distribution. The integral (9) can be written as

$$\int_{\mathbb{R}^n} q(\boldsymbol{\theta})\pi^{\boldsymbol{\xi}}_{\text{post}}(\boldsymbol{\theta}|\boldsymbol{d})\, d\boldsymbol{\theta} = \int_{\mathbb{R}^n} w^{\boldsymbol{\xi}}(\boldsymbol{\theta}) q(\boldsymbol{\theta})\pi_{\text{IS}}(\boldsymbol{\theta})\, d\boldsymbol{\theta}, \quad \text{with } w^{\boldsymbol{\xi}}(\boldsymbol{\theta}) = \frac{\pi^{\boldsymbol{\xi}}_{\text{post}}(\boldsymbol{\theta}|\boldsymbol{d})}{\pi_{\text{IS}}(\boldsymbol{\theta})}, \tag{10}$$

provided that $\pi_{\text{IS}}(\boldsymbol{\theta}) > 0$ whenever $q(\boldsymbol{\theta})\pi_{\text{post}}(\boldsymbol{\theta}) \neq 0$ [48]. When this holds, we can create a Monte Carlo estimate of Eq. (10),

$$\int_{\mathbb{R}^n} q(\boldsymbol{\theta})\pi^{\boldsymbol{\xi}}_{\text{post}}(\boldsymbol{\theta}|\boldsymbol{d})\, d\boldsymbol{\theta} \approx \sum_{i=1}^{M} w_i\, q(\boldsymbol{\theta}_i), \quad \boldsymbol{\theta}_i \sim \pi_{\text{IS}}, \tag{11}$$

where $w_i = w^{\boldsymbol{\xi}}(\boldsymbol{\theta}_i), i = 1, \ldots, M$, define the importance sampling weights. For our purposes, we desire weights that are much greater than zero and have little variation over different samples. Our motivation for using importance sampling is to compute Eq. (11) for different realizations of $\boldsymbol{\xi}$ without the need for multiple MCMC runs. For this to be successful, there must be an importance sampling distribution suitable for all the target posteriors. This is a fundamental assumption behind the proposed method. We propose an importance sampling approach tailored to the Bayesian inverse problem of interest that enables computing Eq. (11) for different choices of $\boldsymbol{\xi}$ using the same importance sampling distribution.

Because the choices of prior distribution belong to a parameterized family, the target posterior distributions belong to a parameterized family (parameterized by the same prior hyperparameters) as well. We let the importance sampling distribution be the posterior $\pi_{\text{IS}} = \pi^{\text{IS}}_{\text{post}}$ constructed using a specific choice of prior, $\pi^{\text{IS}}_{\text{pr}}$. This $\pi^{\text{IS}}_{\text{pr}}$ is chosen from the same family as the priors in such a way that its high-probability region covers that of the family of target priors. See Fig. 1 for an illustration, for the case of Gaussian priors. We then consider

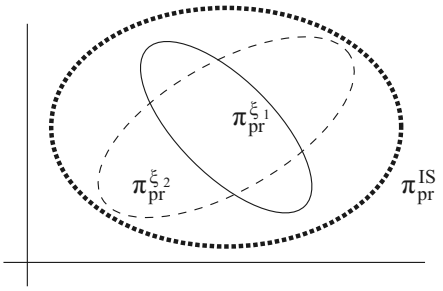

FIG. 1: The interiors of the solid-line and dashed-line ellipses represent the high density regions of two priors $\pi_{\text{pr}}^{\xi_1}$ and $\pi_{\text{pr}}^{\xi_2}$, respectively. They are both enclosed by the high density region of $\pi_{\text{pr}}^{\text{IS}}$, represented by the interior of the dotted-line ellipse.

$$\pi_{\text{post}}^{\text{IS}}(\theta|\boldsymbol{d}) \propto \pi_{\text{like}}(\boldsymbol{d}|\theta) \times \pi_{\text{pr}}^{\text{IS}}(\theta). \tag{12}$$

Importance sampling often breaks down if the importance sampling distribution fails to cover the density of the target, especially when the target distribution has a heavy tail. As noted in our computational results, choosing a prior that "covers" all the target priors typically results in a suitable importance sampling posterior $\pi_{\text{post}}^{\text{IS}}$. With the present strategy, it is possible to sample from $\pi_{\text{post}}^{\text{IS}}$ with one run of MCMC and gather information for all the target posteriors. Choosing the importance sampling prior from the same family as the target priors is expedient but not essential to this strategy. However, doing so makes it easy to compare the importance sampling distribution against the target distributions. The importance sampling prior can be from different family or mixture distribution, though the cases where these choices are more advantageous require future study.

We derive an expression for the estimator (11) when $\pi_{\text{IS}} = \pi_{\text{post}}^{\text{IS}}$. We let θ_{IS} and Γ_{IS} denote the mean and covariance of $\pi_{\text{pr}}^{\text{IS}}$ while θ_ξ and Γ_ξ will denote the mean and covariance of π_{pr}^{ξ}. Let P^ξ and P^{IS} be the normalization constants that correspond to π_{post}^{ξ} and $\pi_{\text{post}}^{\text{IS}}$, respectively:

$$P^\xi := \int_{\mathbb{R}^d} \pi_{\text{like}}(\boldsymbol{d}|\theta)\pi_{\text{pr}}^{\xi}(\theta)\, d\theta, \quad P^{\text{IS}} := \int_{\mathbb{R}^d} \pi_{\text{like}}(\boldsymbol{d}|\theta)\pi_{\text{pr}}^{\text{IS}}(\theta)\, d\theta. \tag{13}$$

We can write the importance sampling weights in Eq. (10) as

$$w^\xi(\theta) = \frac{\pi_{\text{post}}^{\xi}(\theta)}{\pi_{\text{post}}^{\text{IS}}\theta)} = \frac{\pi_{\text{pr}}^{\xi}(\theta)\pi_{\text{like}}(\theta)/P^\xi}{\pi_{\text{pr}}^{\text{IS}}(\theta)\pi_{\text{like}}(\theta)/P^{\text{IS}}} = \frac{1}{P^\xi/P^{\text{IS}}}\frac{\pi_{\text{pr}}^{\xi}(\theta)}{\pi_{\text{pr}}^{\text{IS}}(\theta)}. \tag{14}$$

Letting the importance sampling weight in Eq. (11) be given by Eq. (14), we obtain

$$\int_{\mathbb{R}^n} q(\theta)\pi_{\text{post}}^{\xi}(\theta|\boldsymbol{d})\, d\theta = \frac{1}{P^\xi/P^{\text{IS}}} \int_{\mathbb{R}^n} q(\theta)\frac{\pi_{\text{pr}}^{\xi}(\theta)}{\pi_{\text{pr}}^{\text{IS}}(\theta)}\pi_{\text{post}}^{\text{IS}}(\theta|\boldsymbol{d})\, d\theta. \tag{15}$$

We can use the importance sampling distribution to rewrite the ratio of normalization constants P^ξ/P^{IS} as

$$\begin{aligned}
\frac{P^\xi}{P^{\text{IS}}} &= \frac{1}{P^{\text{IS}}} \int_{\mathbb{R}^n} \pi_{\text{like}}(\boldsymbol{d}|\theta)\pi_{\text{pr}}^{\xi}(\theta)\, d\theta \\
&= \frac{1}{P^{\text{IS}}} \int_{\mathbb{R}^n} \pi_{\text{like}}(\boldsymbol{d}|\theta)\pi_{\text{pr}}^{\xi}(\theta)\frac{P^{\text{IS}}}{\pi_{\text{like}}(\boldsymbol{d}|\theta)\pi_{\text{pr}}^{\text{IS}}(\theta)}\pi_{\text{post}}^{\text{IS}}(\theta|\boldsymbol{d})\, d\theta \\
&= \frac{1}{P^{\text{IS}}} \int_{\mathbb{R}^n} P^{\text{IS}}\frac{\pi_{\text{pr}}^{\xi}(\theta)}{\pi_{\text{pr}}^{\text{IS}}(\theta)}\pi_{\text{post}}^{\text{IS}}(\theta|\boldsymbol{d})\, d\theta \\
&= \int_{\mathbb{R}^n} \frac{\pi_{\text{pr}}^{\xi}(\theta)}{\pi_{\text{pr}}^{\text{IS}}(\theta)}\pi_{\text{post}}^{\text{IS}}(\theta|\boldsymbol{d})\, d\theta.
\end{aligned} \tag{16}$$

Combining the expressions (15) and (16) yields the estimator

$$F(\boldsymbol{\xi}) = \int_{\mathbb{R}^n} q(\boldsymbol{\theta})\pi^{\boldsymbol{\xi}}_{\text{post}}(\boldsymbol{\theta}|\boldsymbol{d})\,d\boldsymbol{\theta} \approx \frac{1}{C(\boldsymbol{\theta}_1,\dots,\boldsymbol{\theta}_M)}\sum_{i=1}^{M} q(\boldsymbol{\theta}_i)\frac{\pi^{\boldsymbol{\xi}}_{\text{pr}}(\boldsymbol{\theta}_i)}{\pi^{\text{IS}}_{\text{pr}}(\boldsymbol{\theta}_i)}, \quad \boldsymbol{\theta}_i \sim \pi^{\text{IS}}_{\text{post}}, \tag{17}$$

where $C(\boldsymbol{\theta}_1,\dots,\boldsymbol{\theta}_M) = \sum_{i=1}^{M}(\pi^{\boldsymbol{\xi}}_{\text{pr}}(\boldsymbol{\theta}_i)/\pi^{\text{IS}}_{\text{pr}}(\boldsymbol{\theta}_i))$ is from the estimator of Eq. (16). For the case of Gaussian priors,

$$\frac{\pi^{\boldsymbol{\xi}}_{\text{pr}}(\boldsymbol{\theta})}{\pi^{\text{IS}}_{\text{pr}}(\boldsymbol{\theta})} = \exp\left[\frac{1}{2}\big((\boldsymbol{\theta}_{\text{IS}}-\boldsymbol{\theta})^\top\Gamma_{\text{IS}}^{-1}(\boldsymbol{\theta}_{\text{IS}}-\boldsymbol{\theta}) - (\boldsymbol{\theta}_{\boldsymbol{\xi}}-\boldsymbol{\theta})^\top\Gamma_{\boldsymbol{\xi}}^{-1}(\boldsymbol{\theta}_{\boldsymbol{\xi}}-\boldsymbol{\theta})\big)\right]. \tag{18}$$

There are some diagnostics for evaluating the effectiveness of a sample set from the importance sampling distribution—see Chapter 9 in [22]. We use effective sample size in our experiments. A large effective sample size is desirable as it indicates small variation in the estimator (15). For a given $\boldsymbol{\xi}$, the effective sample size is

$$n^{\boldsymbol{\xi}}_{\text{E}} := \frac{\left(\sum_{i=1}^{M} w^{\boldsymbol{\xi}}(\boldsymbol{\theta}_i)\right)^2}{\sum_{i=1}^{M} w^{\boldsymbol{\xi}}(\boldsymbol{\theta}_i)^2}, \quad \boldsymbol{\theta}_i \sim \pi^{\text{IS}}_{\text{post}}. \tag{19}$$

Recall from Eq. (14) that we can rewrite $w^{\boldsymbol{\xi}} = \pi^{\boldsymbol{\xi}}_{\text{post}}/\pi^{\text{IS}}_{\text{post}} = (1/(P^{\boldsymbol{\xi}}/P^{\text{IS}}))(\pi^{\boldsymbol{\xi}}_{\text{pr}}/\pi^{\text{IS}}_{\text{pr}})$. We can write Eq. (19) as

$$\begin{aligned} n^{\boldsymbol{\xi}}_{\text{E}} &= \left(\sum_{i=1}^{M}\frac{\pi^{\boldsymbol{\xi}}_{\text{post}}(\boldsymbol{\theta}_i)}{\pi^{\text{IS}}_{\text{post}}(\boldsymbol{\theta}_i)}\right)^2 \Big/ \sum_{i=1}^{M}\left(\frac{\pi^{\boldsymbol{\xi}}_{\text{post}}(\boldsymbol{\theta}_i)}{\pi^{\text{IS}}_{\text{post}}(\boldsymbol{\theta}_i)}\right)^2 = \left(\sum_{i=1}^{M}\frac{1}{P^{\boldsymbol{\xi}}/P^{\text{IS}}}\frac{\pi^{\boldsymbol{\xi}}_{\text{pr}}(\boldsymbol{\theta}_i)}{\pi^{\text{IS}}_{\text{pr}}(\boldsymbol{\theta}_i)}\right)^2 \Big/ \left(\sum_{i=1}^{M}\left(\frac{1}{P^{\boldsymbol{\xi}}/P^{\text{IS}}}\frac{\pi^{\boldsymbol{\xi}}_{\text{pr}}(\boldsymbol{\theta}_i)}{\pi^{\text{IS}}_{\text{pr}}(\boldsymbol{\theta}_i)}\right)^2\right) \\ &= \left(\sum_{i=1}^{M}\frac{\pi^{\boldsymbol{\xi}}_{\text{pr}}(\boldsymbol{\theta}_i)}{\pi^{\text{IS}}_{\text{pr}}(\boldsymbol{\theta}_i)}\right)^2 \Big/ \left(\sum_{i=1}^{M}\left(\frac{\pi^{\boldsymbol{\xi}}_{\text{pr}}(\boldsymbol{\theta}_i)}{\pi^{\text{IS}}_{\text{pr}}(\boldsymbol{\theta}_i)}\right)^2\right). \end{aligned} \tag{20}$$

In practice, we assess the suitability of $\pi^{\text{IS}}_{\text{post}}$ as an importance sampling distribution by examining the distribution of $n^{\boldsymbol{\xi}}_{\text{E}}$ for an ensemble of realizations of $\boldsymbol{\xi}$. This is illustrated in our computational results in Section 5.

Such diagnostics highlight potential avenues for choosing the importance sampling procedure in a more systematic manner. For example, Chapter 5 in [49] explores finding the importance sampling distribution that minimizes the variance of all importance sampling weight functions. Alternatively, one could can minimize the Hellinger distance between the importance sampling distribution and the target distributions [50].

Higher-order moments. We examine the mth order moment of q,

$$\mathbb{E}^{\boldsymbol{\xi}}_{\text{post}}(q^m) = \int_{\mathbb{R}^d}\big(q(\boldsymbol{\theta})\big)^m\pi^{\boldsymbol{\xi}}_{\text{post}}(\boldsymbol{\theta})\,d\boldsymbol{\theta}. \tag{21}$$

The estimator (17) approximates the expectation of q with respect to the posterior. The formula can be applied to estimate the mth-order moment,

$$\mathbb{E}^{\boldsymbol{\xi}}_{\text{post}}(q^m) \approx \frac{1}{C(\boldsymbol{\theta}_1,\dots,\boldsymbol{\theta}_M)}\sum_{i=1}^{M}\big(q(\boldsymbol{\theta}_i)\big)^m\frac{\pi^{\boldsymbol{\xi}}_{\text{pr}}(\boldsymbol{\theta}_i)}{\pi^{\text{IS}}_{\text{pr}}(\boldsymbol{\theta}_i)}, \quad \boldsymbol{\theta}_i \sim \pi^{\text{IS}}_{\text{post}}. \tag{22}$$

In this way, statistics that depend on higher-order moments can be estimated. For example, the variance, which depends on the second moment and expectation, has the estimator

$$\mathbb{E}^{\boldsymbol{\xi}}_{\text{post}}(q^2) - \big(\mathbb{E}^{\boldsymbol{\xi}}_{\text{post}}(q)\big)^2 \approx \frac{1}{C(\boldsymbol{\theta}_1,\dots,\boldsymbol{\theta}_M)}\sum_{i=1}^{M}\big(q(\boldsymbol{\theta}_i)\big)^2\frac{\pi^{\boldsymbol{\xi}}_{\text{pr}}(\boldsymbol{\theta}_i)}{\pi^{\text{IS}}_{\text{pr}}(\boldsymbol{\theta}_i)} - \left(\frac{1}{C(\boldsymbol{\theta}_1,\dots,\boldsymbol{\theta}_M)}\sum_{i=1}^{M}q(\boldsymbol{\theta}_i)\frac{\pi^{\boldsymbol{\xi}}_{\text{pr}}(\boldsymbol{\theta}_i)}{\pi^{\text{IS}}_{\text{pr}}(\boldsymbol{\theta}_i)}\right)^2, \quad \boldsymbol{\theta}_i \sim \pi^{\text{IS}}_{\text{post}}. \tag{23}$$

Accurate estimation of $F_{\text{var}}(\boldsymbol{\xi})$ using Eq. (23) then requires accurate estimation of $F_{\text{mean}}(\boldsymbol{\xi})$.

4.2 Algorithm for GSA of Moment Based HS Maps

We show in Algorithm 1 how GSA can be done on the HS mappings, by using M samples from one MCMC run to yield approximations to the HS mappings $\tilde{F}^M_{\text{mean}} \approx F_{\text{mean}}$ and $\tilde{F}^M_{\text{var}} \approx F_{\text{var}}$. The approximate $\tilde{F}^M_{\text{mean}}$ mappings and $\tilde{F}^M_{\text{var}}$ are defined with respect to the estimators (17) and (23), respectively. For example,

$$\tilde{F}^M_{\text{mean}} = \frac{1}{C(\boldsymbol{\theta}_1, \ldots, \boldsymbol{\theta}_M)} \sum_{i=1}^{M} q(\boldsymbol{\theta}_i) \frac{\pi^{\boldsymbol{\xi}}_{\text{pr}}(\boldsymbol{\theta}_i)}{\pi^{\text{IS}}_{\text{pr}}(\boldsymbol{\theta}_i)}, \quad \boldsymbol{\theta}_i \sim \pi^{\text{IS}}_{\text{post}}. \tag{24}$$

Variance-based sensitivity analysis can be done through pick-freeze estimators for the Sobol' indices of the approximate HS mappings. If many function evaluations and MCMC samples are used, this process can be expensive. A faster alternative is to used surrogate models. We employ sparse regression PCE and sparse weight ELM, discussed in Section 3. Sample realizations of $\boldsymbol{\xi}$ are generated by HS [51,52]. These samples serve as a training set for building surrogate models for GSA, as discussed in Section 3. The purpose of using two different surrogate methods is to help gain further confidence in the computed results.

MCMC generates correlated samples. One strategy for reducing the correlation between samples is thinning [53]. Thinning an MCMC chain means to keep every kth sample in the chain while discarding the rest. This reduces the set to M/k samples. Because the remaining samples are spaced farther apart in the original MCMC chain, the thinned set overall has less correlation than the original set. It is also cheaper to compute the statistical moments of the smaller thinned set than of those of the original set. An extra step can be added to Algorithm 1 where the MCMC chain is thinned. This can be helpful if M and N are large so that computing all HS mapping approximations requires many, many operations.

Algorithm 1 incurs much of its cost during the MCMC sampling stage, assuming that the model and QoI q are expensive to evaluate. In this work, we use the delayed-rejection adaptive Metropolis (DRAM) [28,54,55] algorithm to perform MCMC. Our method does not specifically depend on DRAM and can be used in conjunction with any MCMC algorithm. DRAM, however, is state-of-the-art and has broad applicability. With delayed rejection, each MCMC stage can include up to a fixed number of extra delayed-rejection steps. Each of these steps requires us to evaluate the model

Algorithm 1: Variance-based prior sensitivity analysis through importance sampling

Input: (i) Likelihood PDF $\pi_{\text{like}}(\boldsymbol{d}|\boldsymbol{\theta})$. (ii) Hyperparameter-dependent prior PDF $\pi^{\boldsymbol{\xi}}_{\text{pr}}(\boldsymbol{\theta})$. (iii) Importance sampling prior PDF $\pi^{\text{IS}}_{\text{pr}}(\boldsymbol{\theta})$. (iv) Collection of hyperparameter samples $\{\boldsymbol{\xi}_k\}_{k=1}^N$. (v) QoI function $q(\boldsymbol{\theta})$. (vi) Monte Carlo sample size M.

Output: (i) First-order Sobol' indices. (ii) Total Sobol' indices.

Choose importance sampling prior $\pi^{\text{IS}}_{\text{pr}}$
Perform MCMC to generate samples, $\{\boldsymbol{\theta}_i\}_{i=1}^M$, from $\pi^{\text{IS}}_{\text{post}}$, of which $\widehat{M}$ are distinct
for $i = 1, \ldots, \widehat{M}$
    Compute and store $q(\boldsymbol{\theta}_i)$
    Compute and store $\pi^{\text{IS}}_{\text{pr}}(\boldsymbol{\theta}_i)$
    for $k = 1, \ldots, N$
        Compute and store $\pi^{\boldsymbol{\xi}_k}_{\text{pr}}(\boldsymbol{\theta}_i)$
    end for
end for
for $k = 1, \ldots, N$
    Approximate $F_{\text{mean}}(\boldsymbol{\xi}_k)$ and $F_{\text{var}}(\boldsymbol{\xi}_k)$, using the estimators (17) and (23), with $\left\{q(\boldsymbol{\theta}_i), \pi^{\boldsymbol{\xi}_k}_{\text{pr}}(\boldsymbol{\theta}_i), \pi^{\text{IS}}_{\text{pr}}(\boldsymbol{\theta}_i)\right\}_{i=1}^M$
end for
Estimate first-order and total Sobol' indices of F_{mean} and F_{var}

an additional time. Typically, one initially runs MCMC for M_{burn} burn-in stages. These burn-in samples are discarded and not included in the set of posterior draws. The cost of running the MCMC stage in Algorithm 1 with DRAM is $\mathcal{O}(M + M_{\text{burn}})$ model evaluations. In the second stage, we evaluate q at the distinct MCMC samples. Because the MCMC samples usually include repeated draws, the number of these QoI evaluations is less than M.

4.3 GSA of the MAP Point

The MAP point is an important point estimator and studying its sensitivity to prior parameters complements the study of other moment-based HS maps such as the posterior mean or variance. The approach described in Algorithm 1 can be used in cases where $F(\boldsymbol{\xi})$ involves moments of the posterior, as in the case of the mean and variance. On the other hand, evaluating F_{atMAP} requires solving the regularized nonlinear least-squares problem (5) for each realization of $\boldsymbol{\xi}$. No numerical integration is needed. One does not even need to know the normalization constant of the posterior to find its MAP point. While we do not use Algorithm 1 to study F_{atMAP}, we evaluate it at the same set of realizations $\{\boldsymbol{\xi}_k\}_{k=1}^N$ used in Algorithm 1. These evaluations are used to build surrogate models for F_{atMAP}. The computed surrogate is then used for fast GSA of F_{atMAP}.

5. COMPUTATIONAL RESULTS

In this section, we consider two model inverse problems as test beds for our proposed approach. Specifically, we use Algorithm 1 for global sensitivity analysis (GSA) of hyperparameter-to-statistic (HS) mappings from the inverse problems under study. These examples are used to examine various aspects of the proposed method. In Section 5.1, we consider a simple linear inverse problem. Specifically, we formulate fitting a line to noisy data as a linear Bayesian inverse problem. In this case, the posterior distribution is known analytically. This means that the HS mappings admit analytical forms, and we can perform GSA without Algorithm 1. This problem serves as a benchmark where we gauge the accuracy of GSA with Algorithm 1 against reference values. The QoI in this example is a quadratic function. For this QoI, we study the HS mappings for the mean and variance. The Sobol' indices, approximated using Algorithm 1, of these HS mappings are compared to the true Sobol' indices. Overall, we note close agreement between the results produced by our method and the analytic results.

Next, we apply our method to a nonlinear Bayesian inverse problem in Section 5.2. The inverse problem is governed by the SEIR epidemiological model [56,57]. It exemplifies the type of problem that Algorithm 1 is designed and intended for. Our numerical results provide a unique perspective on the impact of uncertainty in prior hyperparameters. The QoI is the basic reproductive number. We quantify the uncertainty in the mean, variance, and MAP point that is caused by uncertainty in the prior hyperparameters. The Sobol' indices of the mean, variance, and MAP point HS mappings are computed using Algorithm 1 and highlight the most influential hyperparameters in each case. We use two different surrogate modeling approaches in these computations: one based on sparse polynomial chaos expansions (PCEs) and the other based on sparse weight extreme learning machines (SW-ELMs). The two approaches provide results that match closely.

5.1 Linear Bayesian Inverse Problem

We consider the problem of fitting a line $y = mt + b$ to noisy measurements $\{(t_i, y_i)\}_{i=1}^4$ at times $t = 0, 0.5, 1.5, 2.5$. The slope m and intercept b are treated as unknown parameters, which we seek to estimate. We cast this problem in a Bayesian framework. This serves to illustrate various properties of our proposed framework.

5.1.1 Bayesian Inverse Problem Setup

Let the inversion parameter vector be denoted by $\boldsymbol{\theta} = \begin{bmatrix} b & m \end{bmatrix}^\top$. We consider estimation of $\boldsymbol{\theta}$ from

$$\mathbf{A}\boldsymbol{\theta} + \boldsymbol{\eta} = \boldsymbol{y}, \tag{25}$$

where $\mathbf{A} = \begin{bmatrix} 1 & 1 & 1 & 1 \\ 0 & 0.5 & 1.5 & 2.5 \end{bmatrix}^\top$ is the forward operator, $\boldsymbol{\eta}$ models measurement noise, and $\boldsymbol{y}$ is the data.

We assume noise at each measurement independently follows the standard normal distribution, i.e., $\eta_i \sim \mathcal{N}(0,1)$. The noise covariance is $\mathbf{\Gamma}_{\text{noise}} = \mathbf{I}_{4\times 4}$. We assume a "ground-truth" parameter vector $\boldsymbol{\theta}_{\text{true}} = [1 \ \ -2]^\top$ and generate measurements by adding sampled noise η_i to $y_i = -2t_i + 1$ for $i = 1, \ldots, 4$; see Fig. 2. We assume a Gaussian prior distribution $\mathcal{N}(\boldsymbol{\mu}_{\text{pr}}, \boldsymbol{\Sigma}_{\text{pr}})$ for the inversion parameters $\boldsymbol{\theta}$ with

$$\boldsymbol{\mu}_{\text{pr}} = \begin{bmatrix} \mu_b \\ \mu_m \end{bmatrix}, \quad \boldsymbol{\Sigma}_{\text{pr}} = \begin{bmatrix} \sigma_b^2 & 0 \\ 0 & \sigma_m^2 \end{bmatrix}. \tag{26}$$

Due to linearity of the parameter-to-observable map and Gaussian prior and noise models, the posterior distribution for $\boldsymbol{\theta}$ is also Gaussian and explicitly known. It is the Gaussian distribution $\mathcal{N}(\boldsymbol{\mu}_{\text{post}}, \boldsymbol{\Sigma}_{\text{post}})$, where

$$\boldsymbol{\Sigma}_{\text{post}} = (\mathbf{A}^\top \mathbf{\Gamma}_{\text{noise}}^{-1} \mathbf{A} + \boldsymbol{\Sigma}_{\text{pr}}^{-1})^{-1}, \quad \boldsymbol{\mu}_{\text{post}} = \mathbf{\Gamma}_{\text{post}}(\mathbf{A}^\top \mathbf{\Gamma}_{\text{noise}}^{-1} \boldsymbol{y} + \boldsymbol{\Sigma}_{\text{pr}}^{-1} \boldsymbol{\mu}_{\text{pr}}). \tag{27}$$

Quantity of interest. We introduce the QoI which depends on the inversion parameters $\boldsymbol{\theta}$. The QoI is the quadratic form

$$q(\boldsymbol{\theta}) = \boldsymbol{\theta}^\top \boldsymbol{\theta} = m^2 + b^2, \quad \boldsymbol{\theta} \sim \mathcal{N}(\boldsymbol{\mu}_{\text{post}}, \boldsymbol{\Sigma}_{\text{post}}). \tag{28}$$

As $\boldsymbol{\theta}$ is a Gaussian random variable, we have access to expressions for the first and second moments [58–61]. of the QoI. We can therefore express the mean and variance of the QoI analytically.

$$\mathbb{E}_{\text{post}}(q) = \text{tr}(\boldsymbol{\Sigma}_{\text{post}}) + \boldsymbol{\mu}_{\text{post}}^\top \boldsymbol{\mu}_{\text{post}}, \quad \text{var}(q) = 2\,\text{tr}(\boldsymbol{\Sigma}_{\text{post}}^2) + 4\boldsymbol{\mu}_{\text{post}}^\top \boldsymbol{\Sigma}_{\text{post}} \boldsymbol{\mu}_{\text{post}}. \tag{29}$$

Uncertainty in prior hyperparameters. Before building the posterior distribution, we must choose values for the prior hyperparameters $\boldsymbol{\xi} = \begin{bmatrix} \mu_b & \mu_m & \sigma_b^2 & \sigma_m^2 \end{bmatrix}^\top$ that appear in Eq. (26). We assume these parameters are specified within some interval around their nominal values and are modeled as independent uniformly distributed random variables. We use a nominal value of 1 for each of the parameters and let the upper and lower bounds of the distributions be $\pm$ 50% perturbations of the nominal value.

5.1.2 Parameter Estimation and Importance Sampling

To understand how the uncertainty in the prior hyperparameters affects the QoI, we employ Algorithm 1 from Section 4. The first step is to choose a prior $\pi_{\text{pr}}^{\text{IS}}$ to build the importance sampling distribution $\pi_{\text{post}}^{\text{IS}}$. We take $\mathcal{N}(\boldsymbol{\mu}_{\text{pr}}^{\text{IS}}, \boldsymbol{\Sigma}_{\text{pr}}^{\text{IS}})$ with

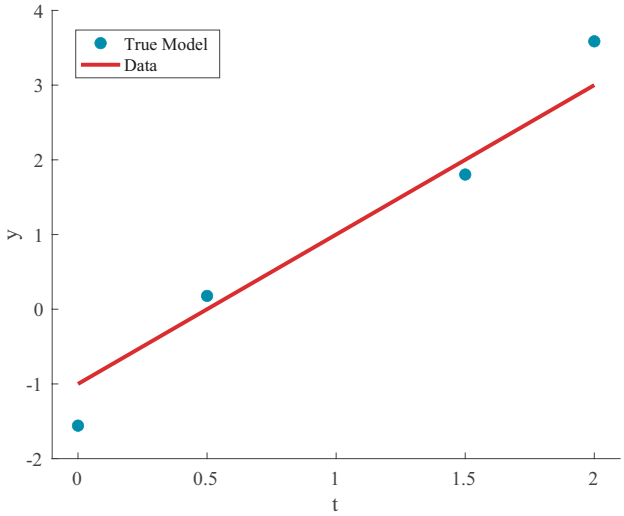

FIG. 2: The true trajectory of the linear model plotted with the noisy measurements at times $t = 0, 0.5, 1.5, 2$

$$\boldsymbol{\mu}_{\text{pr}} = \begin{bmatrix} 1 \\ 1 \end{bmatrix}, \quad \boldsymbol{\Sigma}_{\text{pr}} = \begin{bmatrix} 1.5^2 & 0 \\ 0 & 1.5^2 \end{bmatrix}. \tag{30}$$

We use the DRAM algorithm, discussed in Section 4, to draw 10^5 samples from $\pi^{\text{IS}}_{\text{post}}$. In Fig. 3, we compare the prior, analytic posterior, and MCMC-constructed posterior marginal distributions of b and m. Before we implement Algorithm 1, we evaluate whether $\pi^{\text{IS}}_{\text{post}}$ is an acceptable importance sampling distribution. As discussed in Section 4, we use Eq. (20) to compute the effective sample size over the distribution of prior hyperparameters $\boldsymbol{\xi}$. The distribution of effective sample sizes, given in Fig. 4(left), shows that $\pi^{\text{IS}}_{\text{post}}$ is an effective importance sampling distribution over many realizations of $\boldsymbol{\xi}$. In Fig. 4, we give a further visual of how $\pi^{\text{IS}}_{\text{post}}$ serves as an effective importance sampling distribution. In the right panel, the distribution of q, when $\boldsymbol{\theta} \sim \pi^{\text{IS}}_{\text{post}}$, is compared to the distributions of $q(\boldsymbol{\theta})$ when $\boldsymbol{\theta} \sim \pi^{\boldsymbol{\xi}}_{\text{post}}$, for three realizations of $\boldsymbol{\xi}$. The realizations are

- $\boldsymbol{\xi}_1 = \begin{bmatrix} 1.5 & 0.5 & 1.5 & 1 \end{bmatrix}^\top$,
- $\boldsymbol{\xi}_2 = \begin{bmatrix} 1.5 & 1.5 & 1.5 & 1.5 \end{bmatrix}^\top$,
- and $\boldsymbol{\xi}_3 = \begin{bmatrix} 1 & 1.5 & 1 & 0.5 \end{bmatrix}^\top$.

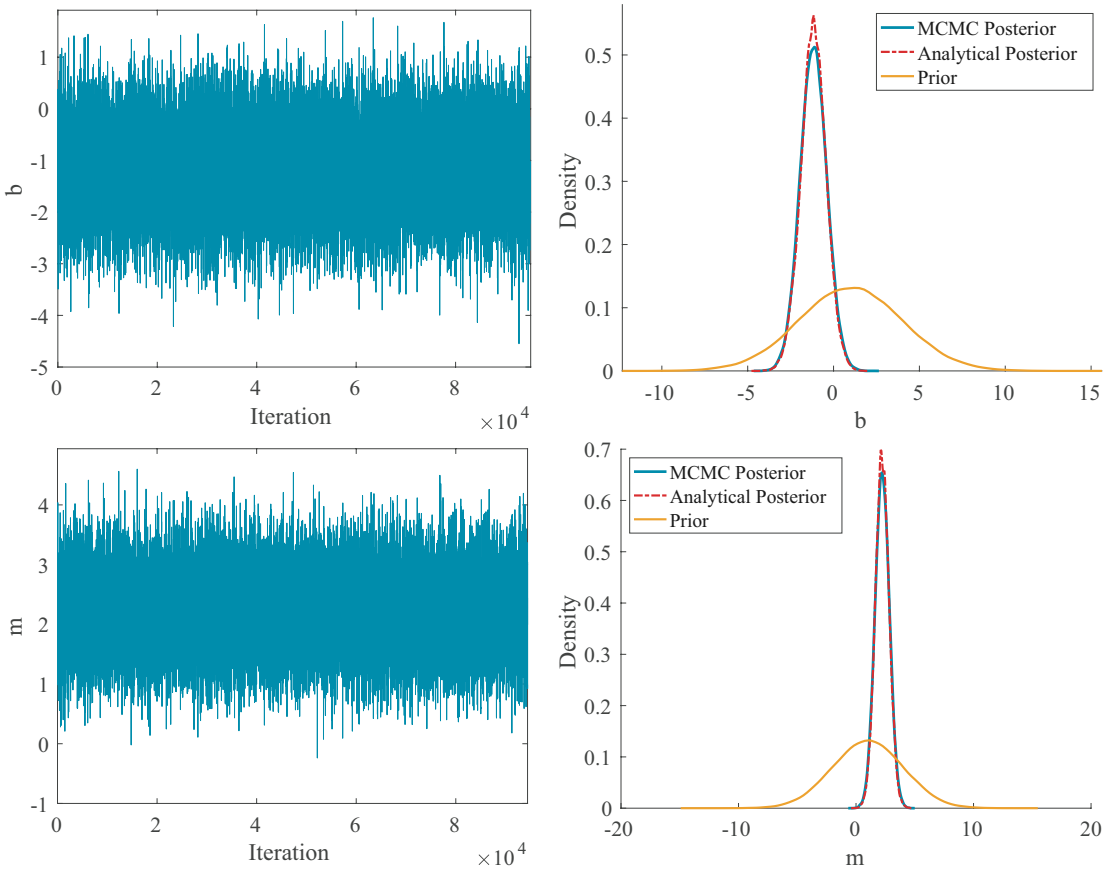

FIG. 3: MCMC chains of the inversion parameters m and b along with corresponding marginal posterior distributions compared to prior distributions

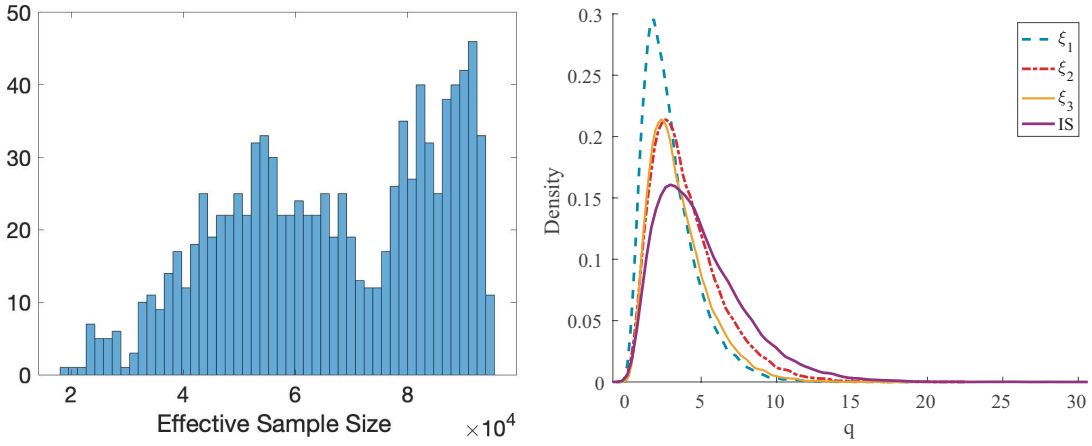

FIG. 4: Left: Histogram, over 1000 realizations of $\boldsymbol{\xi}$, of effective sample size for importance sampling of $M = 10^5$ samples from $\pi^{\text{IS}}_{\text{post}}$. Right: Distribution of q when $\boldsymbol{\theta} \sim \pi^{\text{IS}}_{\text{post}}$ compared to when $\boldsymbol{\theta} \sim \pi^{\boldsymbol{\xi}}_{\text{post}}$ for three realizations of $\boldsymbol{\xi}$.

5.1.3 Sensitivity Analysis

We now study q given in Eq. (28). We are interested in the variance and mean HS mappings (6) $F_{\text{mean}}(\boldsymbol{\xi}) = \mathbb{E}^{\boldsymbol{\xi}}_{\text{post}}(q)$ and $F_{\text{var}}(\boldsymbol{\xi}) = \text{var}^{\boldsymbol{\xi}}(q)$. As shown in Eq. (29), these HS mappings take analytically known forms.

We use Algorithm 1 to yield estimations of the HS mappings for use in uncertainty quantification and computing Sobol' indices. The importance sampling distribution is given by $\pi^{\text{IS}}_{\text{post}}$, as described in Section 4, and with $\pi^{\text{IS}}_{\text{pr}}$ as specified in Eq. (30). We study how the Sobol' indices, computed via Algorithm 1, converge as we increase MCMC sample size M.

The Sobol' indices estimated by Algorithm 1 are compared against benchmark indices. We compute the benchmark indices by applying the standard sampling approach from [31] to the HS mappings. This yields accurate indices because we have access to the analytic expressions of F_{var} and F_{mean}.

In Fig. 5, the study for F_{mean} is given. The tests of convergence for the first-order and total Sobol' indices compare the pick-freeze estimators of the approximate HS mapping $\tilde{F}_{\text{mean}}$ to the pick-freeze estimators of the exact HS mapping F_{mean}, given by Eq. (29), which we treat as reference values. Each set of estimators uses 6×10^6 evaluations to compute. We see how closely the Sobol' indices of the approximate HS mapping converge to the reference values as M increases.

As the pick-freeze estimators are expensive to evaluate, we demonstrate in Fig. 5 how surrogates can speed up GSA and give accurate results. In our computations, we build sparse PCE and sparse weight ELM [43] surrogate models, discussed in Section 3 using 10^3 realizations of $\boldsymbol{\xi}$, drawn using LHS. SW-ELM surrogates also use another 200 realizations for validation during the weight sparsification step. For PCE surrogates, the PC basis is truncated at total degree 5. With only a modest MCMC sample size of $M = 10^4$, we can ascertain the correct importance ranking of the total Sobol' indices of F_{mean}. The surrogate models provide us with accurate GSA, but require few model evaluations.

In Fig. 6, we have results of repeating the above study for F_{var}. A slower rate of convergence may be expected, but is not observed. The numerical studies for the present model linear inverse problem provide a proof-of-concept study of Algorithm 1. In particular, availability of analytic expressions for the HS mappings enables testing the accuracy of the computed results. In all the numerical tests, we note that a modest MCMC sample size is sufficient to obtain the correct parameter rankings.

We show in Fig. 7 the uncertainty in the posterior mean and variance of the QoI caused by uncertainty in the prior hyperparameters. Kernel density estimates of the distribution of the approximate HS mapping are compared to those of the true HS mapping. We estimate the approximation errors that result from using MCMC in Algorithm 1 and from building the surrogate models. A verification set of 1000 realizations of $\boldsymbol{\xi}$ is drawn from the distribution on the prior hyperparameters. We evaluate relative error over this set, so that

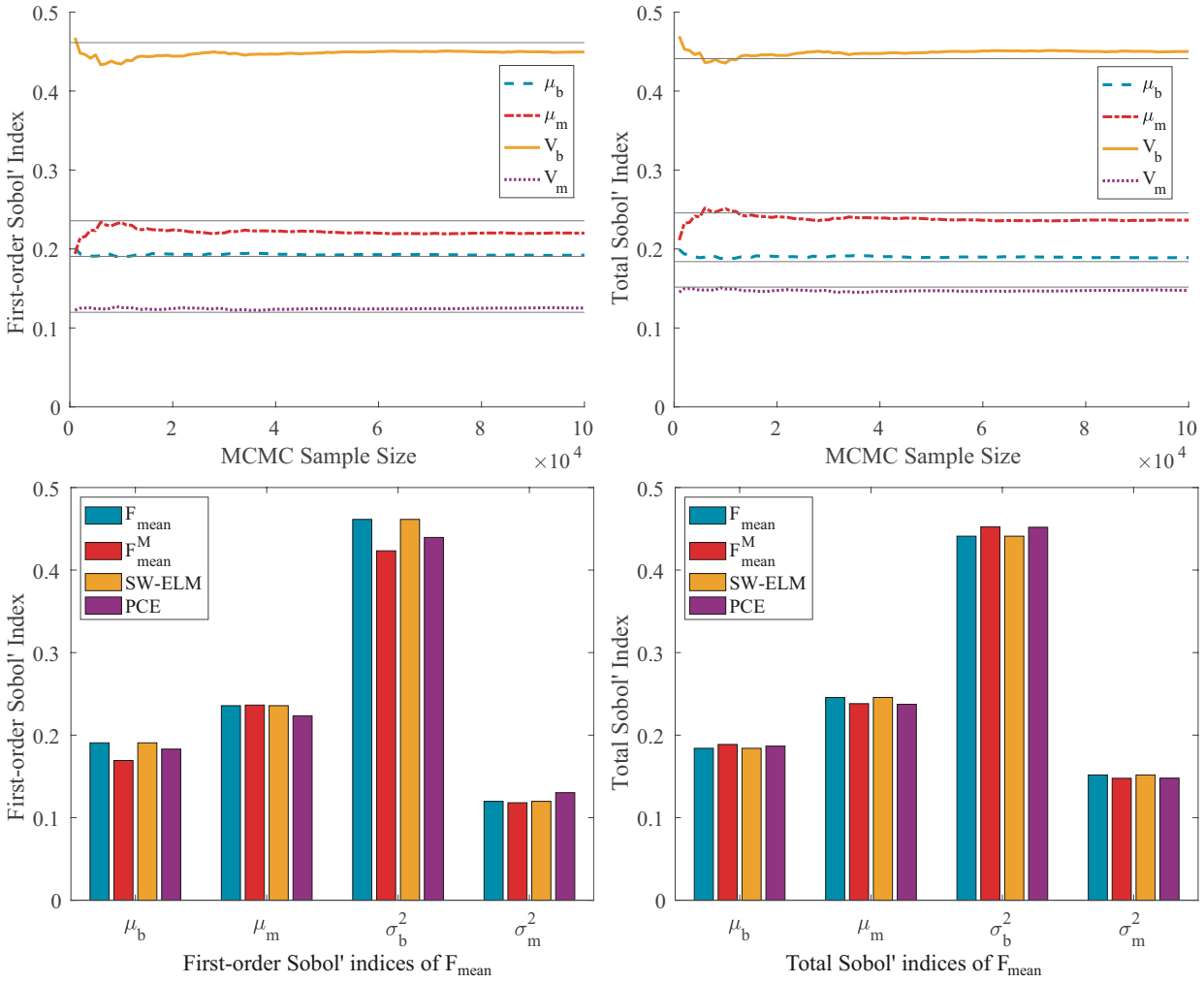

FIG. 5: Top left: Convergence experiment showing how the first-order Sobol' indices of $\tilde{F}^M_{\text{mean}}$ converge to those of F_{mean} with increasing MCMC sample size M. Sobol' indices are approximated using pick-freeze estimators with 6×10^6 evaluations. Pick-freeze estimators for Sobol' indices of F_{mean} (horizontal lines) are treated as values. Top right: Convergence experiment showing how the total Sobol' indices of $\tilde{F}^M_{\text{mean}}$ converge to those of F_{mean} with increasing MCMC sample size M. Bottom left: Comparison of reference first-order Sobol' indices to surrogate-estimated indices where $\tilde{F}^M_{\text{mean}}$ evaluations are approximated with $M = 10^5$ MCMC samples. Bottom right: Comparison of reference total Sobol' indices to pick-freeze estimators and surrogate-estimated indices.

$$\epsilon_{\text{rel}}(F, G) := \frac{\sum_{k=1}^{N} (F(\boldsymbol{\xi}_k) - G(\boldsymbol{\xi}_k))^2}{\sum_{k=1}^{N} (F(\boldsymbol{\xi}_k))^2}. \tag{31}$$

We also assess the relative error in total Sobol' indices using

$$S_{\text{rel}} := \frac{\sum_{i=1}^{d} \left(S_i^{\text{tot}}(F) - S_i^{\text{tot}}(G)\right)^2}{\sum_{i=1}^{d} \left(S_i^{\text{tot}}(F)\right)^2}. \tag{32}$$

The relative approximation errors are given in Table 1 along with the maximum relative errors in approximation of the Sobol' indices.

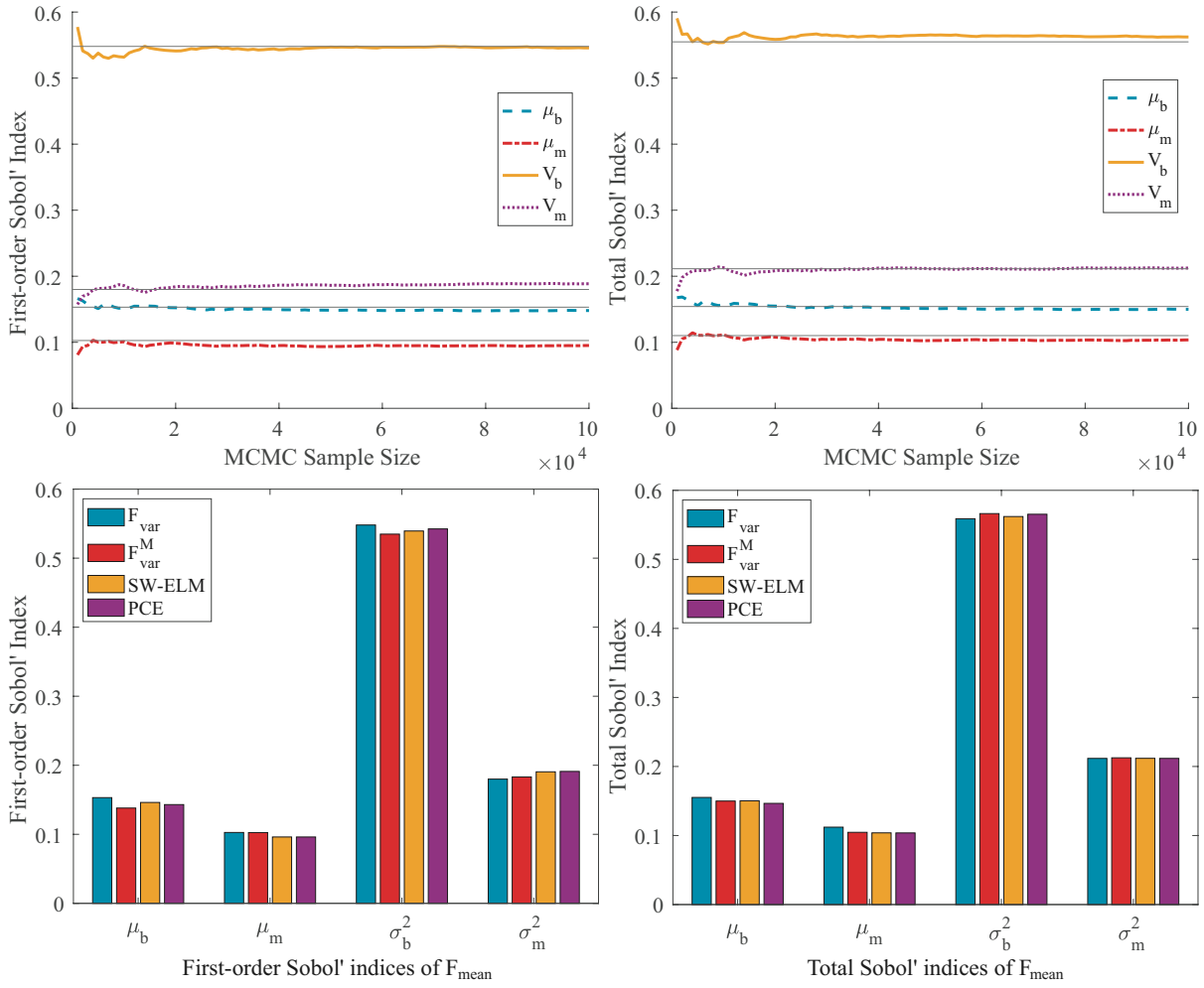

FIG. 6: Top left: Convergence experiment showing how the first-order Sobol' indices of $\tilde{F}_{\text{var}}^{M}$ converge to those of F_{var} with increasing MCMC sample size M. Sobol' indices are approximated using pick-freeze estimators with 6×10^6 evaluations. Pick-freeze estimators for Sobol' indices of F_{var} (horizontal lines) are treated as values. Top right: Convergence experiment showing how the total Sobol' indices of $\tilde{F}_{\text{var}}^{M}$ converge to those of F_{var} with increasing MCMC sample size M. Bottom left: Comparison of reference first-order Sobol' indices to surrogate-estimated indices where $\tilde{F}_{\text{var}}^{M}$ evaluations are approximated with $M = 10^5$ MCMC samples. Bottom right: Comparison of reference total Sobol' indices to pick-freeze estimators and surrogate-estimated indices.

5.2 Nonlinear Bayesian Inverse Problem Based on SEIR Model

In this section, we consider a Bayesian inverse problem governed by the susceptible-exposed-infected-recovered (SEIR) model [56,57] epidemic model. In Section 5.2.1 we discuss the governing SEIR model and the Bayesian inverse problem under study. In Section 5.2.2, we study the proposed importance sampling procedure for computing the HS mappings under study. Finally, in Section 5.2.3, we present our computational results for GSA of the present Bayesian inverse problem with respect to prior hyperparameters.

5.2.1 The Inverse Problem

The SEIR model simulates the time dynamics of an epidemic outbreak in a population. The model has four compartments, S, E, I, and R, corresponding to the susceptible, exposed, infected, and recovered populations. The individuals

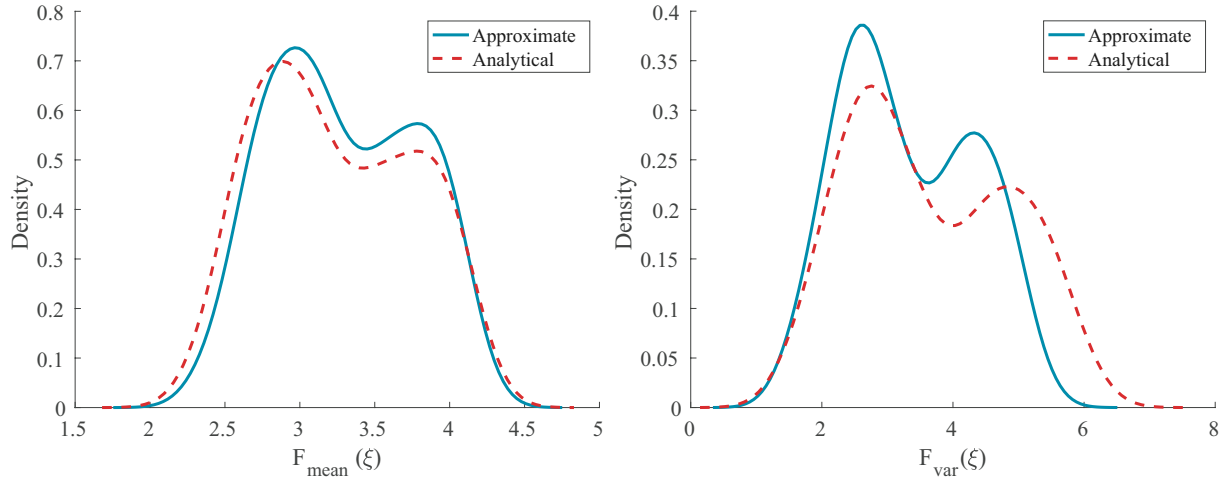

FIG. 7: Left: True distribution of F_{mean} compared to its approximation, estimated using $\tilde{F}^M_{\text{mean}}$ with $M = 10^5$ MCMC samples. Right: True distribution of F_{var} compared to its approximation, estimated using $\tilde{F}^M_{\text{var}}$ with $M = 10^5$ MCMC samples.

TABLE 1: Left: Relative approximation errors, defined by Eq. (31), capturing error from MCMC and from surrogate approximation. Right: Relative error in total Sobol' index estimates, defined by Eq. (32), resulting from error from MCMC and from surrogate approximation. Surrogate approximation captures deviation of the surrogate from the approximate HS mapping. MCMC error captures deviation of the approximate HS mapping from the true HS mapping

Error type	F_{mean}	F_{var}
$\epsilon_{\text{rel}}(F, \tilde{F}^M)$	1.78×10^{-2}	0.100
$\epsilon_{\text{rel}}(F, \tilde{F}^M_{\text{ELM}})$	1.78×10^{-2}	0.100
$\epsilon_{\text{rel}}(F, \tilde{F}^M_{\text{PCE}})$	1.78×10^{-2}	0.100
$\epsilon_{\text{rel}}(\tilde{F}^M, \tilde{F}^M_{\text{ELM}})$	4.56×10^{-6}	6.73×10^{-6}
$\epsilon_{\text{rel}}(\tilde{F}^M, \tilde{F}^M_{\text{PCE}})$	2.95×10^{-5}	1.53×10^{-4}

Error type	F_{mean}	F_{var}
$S_{\text{rel}}(F, \tilde{F}^M)$	2.85×10^{-2}	1.98×10^{-2}
$S_{\text{rel}}(F, \tilde{F}^M_{\text{ELM}})$	1.31×10^{-2}	6.24×10^{-3}
$S_{\text{rel}}(F, \tilde{F}^M_{\text{PCE}})$	1.69×10^{-2}	8.43×10^{-3}
$S_{\text{rel}}(\tilde{F}^M, \tilde{F}^M_{\text{ELM}})$	2.52×10^{-2}	3.89×10^{-3}
$S_{\text{rel}}(\tilde{F}^M, \tilde{F}^M_{\text{PCE}})$	1.55×10^{-2}	1.24×10^{-2}

in the exposed compartment are those who have been exposed to the disease but are not yet displaying signs of infection. The individuals in the I compartment are infected and infectious. We consider a standard SEIR model where we assume recovered individuals cannot be reinfected. Additionally, we assume that the natural birth and death rates are equal and neglect disease related mortality. This ensures that the total population $N = S + E + I + R$ remains constant over time. The present model is described by the following system of nonlinear ordinary differential equations (ODEs):

$$\begin{aligned} \dot{S} &= \mu N - \beta SI/N - \mu S, \\ \dot{E} &= \beta SI/N - (\sigma + \mu)E, \\ \dot{I} &= \sigma E - (\gamma + \mu)I, \\ \dot{R} &= \gamma I - \mu R. \end{aligned} \tag{33}$$

There are four model parameters in the above system which we seek to estimate. The infection rate β, in units of days^{-1}, represents how quickly an infected individual infects a susceptible individual. The recovery rate γ, in units of days^{-1}, represents how fast an infected individual recovers from infection. The latency rate σ, in units of days^{-1}, represents how long it takes for an exposed individual to display symptoms. Lastly, there is also a parameter μ, with units of individuals per day, which represents both the natural birth rate and the natural death rate. In the model,

individuals are only born susceptible while individuals in any compartment can die a natural death. As noted before, since the birth and death rates are the same, the total population size remains constant.

Setup. Bayesian analysis of an epidemic using the SEIR model requires constructing a posterior of the model parameters from available infection data [62]. For the purposes of this example, we simulate an epidemic governed by the SEIR model for a population of $N = 1000$ individuals. The nominal parameters and initial conditions are detailed in Table 2. The nominal parameter values will be used as "ground-truth" in the computational studies that follow. The dynamics of the epidemic under these conditions are shown in Fig. 8(left). Next, we formulate a Bayesian inverse problem. In what follows, we formulate the inverse problem as that of estimating the log of the uncertain model parameters. Hence, we consider the inversion parameter vector, $\theta = \begin{bmatrix} \log\mu & \log\beta & \log\sigma & \log\gamma \end{bmatrix}^\top$. The data measurements, used to solve the inverse problem, consist of simulated data $\{(t_k, I_k)\}$ at times $t_k = 3k + 30$, where $k = 1, \ldots, 15$. These simulated data measurements are obtained by solving the SEIR model with ground-truth parameter values and adding random noise. The noise at each measurement is identically independently distributed from a normal distribution $\mathcal{N}(0, 30^2)$. The simulated data compared to the true model are shown in Fig. 8(right).

We use a Gaussian prior $\mathcal{N}(\boldsymbol{m}_{\text{pr}}, \boldsymbol{\Sigma}_{\text{pr}})$ on the inversion parameter vector θ with

$$\boldsymbol{m}_{\text{pr}} = \begin{bmatrix} m_{\log\mu} \\ m_{\log\beta} \\ m_{\log\sigma} \\ m_{\log\gamma} \end{bmatrix}, \quad \boldsymbol{\Sigma}_{\text{pr}} = \begin{bmatrix} s^2_{\log\mu} & 0 & 0 & 0 \\ 0 & s^2_{\log\beta} & 0 & 0 \\ 0 & 0 & s^2_{\log\sigma} & 0 \\ 0 & 0 & 0 & s^2_{\log\gamma} \end{bmatrix}. \tag{34}$$

TABLE 2: Model parameters and initial conditions used to simulate the SEIR model (33) in Fig. 8

Model parameter	Value	Initial condition	Value
μ	5.48×10^{-5}	S_0	999
β	1/2.5	$E(0)$	0
σ	1/3	$I(0)$	1
γ	1/7	$R(0)$	0

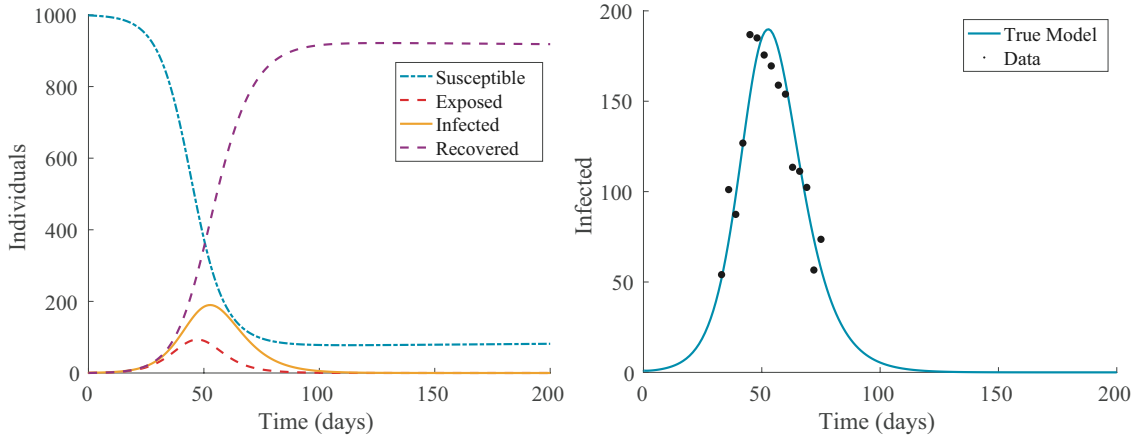

FIG. 8: Left: Simulated dynamics of an epidemic outbreak following the SEIR model. The total population is $N = 1000$ individuals and the model parameters are $\mu = 5.48 \times 10^{-5}$ individuals/day, $\beta = 1/2.5$ days^{-1}, $\sigma = 1/3$ days^{-1}, and $\gamma = 1/7$ days^{-1}. Initially, there is one infected individual and there are no exposed individuals. Right: Simulated data of the epidemic outbreak compared to the true infected dynamics. The simulated infected data are taken by adding noise to values from the true model. The noise is sampled from the normal distribution $\mathcal{N}(0, 30^2)$.

Unlike the inverse problem in Section 5.1, this Bayesian inverse problem is nonlinear. In this case, we do not have access to an analytically known posterior distribution. This means MCMC is needed to sample from the posterior distribution.

Uncertainty in prior hyperparameters. We assume there is uncertainty in the hyperparameters that appear in Eq. (34). Specifically, we consider the vector

$$\xi = \begin{bmatrix} m_{\log\mu} & m_{\log\beta} & m_{\log\sigma} & m_{\log\gamma} & s^2_{\log\mu} & s^2_{\log\beta} & s^2_{\log\sigma} & s^2_{\log\gamma} \end{bmatrix}^\top$$

of parameters that define the prior as uncertain. In the present study, we assume that the entries of ξ are independent uniformly distributed random variables, as specified in Table 3.

Quantity of interest. An important quantity of interest in epidemiology is the basic reproduction number, denoted R_0. It can be interpreted as the number of secondary infections caused, on average, by a single individual [56]. Determining R_0 of an epidemic is key to understanding how severe the outbreak could be. For the SEIR model (33), R_0 takes the form

$$R_0 = \frac{\beta}{\gamma+\mu}\frac{\sigma}{\sigma+\mu}. \tag{35}$$

For the simulated epidemic in Fig. 8, $R_0 = 2.7985$. The importance of R_0 makes it a prime area to apply uncertainty quantification and robustness analysis. In [63], the robustness of R_0 estimates to model parameters is considered through local derivative-based methods. Hence, we focus on R_0 as the QoI:

$$q(\theta) = \frac{e^{\theta_2}}{e^{\theta_4}+e^{\theta_1}}\frac{e^{\theta_3}}{e^{\theta_3}+e^{\theta_1}}.$$

5.2.2 Parameter Estimation and Importance Sampling

Before we can implement Algorithm 1, we have to choose the importance sampling distribution. In accordance with the discussion in Section 4, we choose the importance sampling distribution $\pi^{\text{IS}}_{\text{pr}}$ as $\mathcal{N}(\boldsymbol{m}^{\text{IS}}_{\text{pr}}, \boldsymbol{\Sigma}^{\text{IS}}_{\text{pr}})$ with

$$\boldsymbol{m}^{\text{IS}}_{\text{pr}} = \begin{bmatrix} -10 \\ -1.5 \\ -1.5 \\ -1.5 \end{bmatrix}, \quad \boldsymbol{\Sigma}^{\text{IS}}_{\text{pr}} = \begin{bmatrix} 2^2 & 0 & 0 & 0 \\ 0 & 2^2 & 0 & 0 \\ 0 & 0 & 2^2 & 0 \\ 0 & 0 & 0 & 2^2 \end{bmatrix}. \tag{36}$$

Because $m_{\log\mu}$ takes a wider range of values compared to the other means, we impose a large variance on $\log\mu$ in $\pi^{\text{IS}}_{\text{pr}}$. We construct the corresponding posterior $\pi^{\text{IS}}_{\text{post}}$ using the DRAM algorithm. The first 10^4 samples are removed for burn-in. We adapt the proposal distribution for 10^6 iterations. After this we take 10^6 draws from the MCMC chain. Because the posterior is non-Gaussian, the resulting MCMC chain has high levels of correlation. We present the MCMC chains of log parameters and their respective marginal posterior distributions in Fig. 9. In Fig. 10(left), we evaluate the effectiveness of our importance sampling distribution by examining the distribution of effective sample sizes. We also compare the distribution of R_0 values, with respect to $\pi^{\text{IS}}_{\text{post}}$, compared to the posterior distributions for three realizations of the prior hyperparameters in Fig. 10(right). The realizations are

TABLE 3: Intervals for admissible hyperparameter values of the prior $\theta \sim \mathcal{N}(\boldsymbol{m}_{\text{pr}}, \boldsymbol{\Sigma}_{\text{pr}})$. Each hyperparameter is uniformly distributed on an interval perturbed $\pm$ 50% of the respective nominal value

Mean hyperparameter	Distribution	Variance hyperparameter	Distribution
$m_{\log\mu}$	$\mathcal{U}([-15,-5])$	$s^2_{\log\mu}$	$\mathcal{U}([0.5,1.5])$
$m_{\log\beta}$	$\mathcal{U}([-2.25,-0.75])$	$s^2_{\log\beta}$	$\mathcal{U}([0.5,1.5])$
$m_{\log\sigma}$	$\mathcal{U}([-2.25,-0.75])$	$s^2_{\log\sigma}$	$\mathcal{U}([0.5,1.5])$
$m_{\log\gamma}$	$\mathcal{U}([-2.25,-0.75])$	$s^2_{\log\gamma}$	$\mathcal{U}([0.5,1.5])$

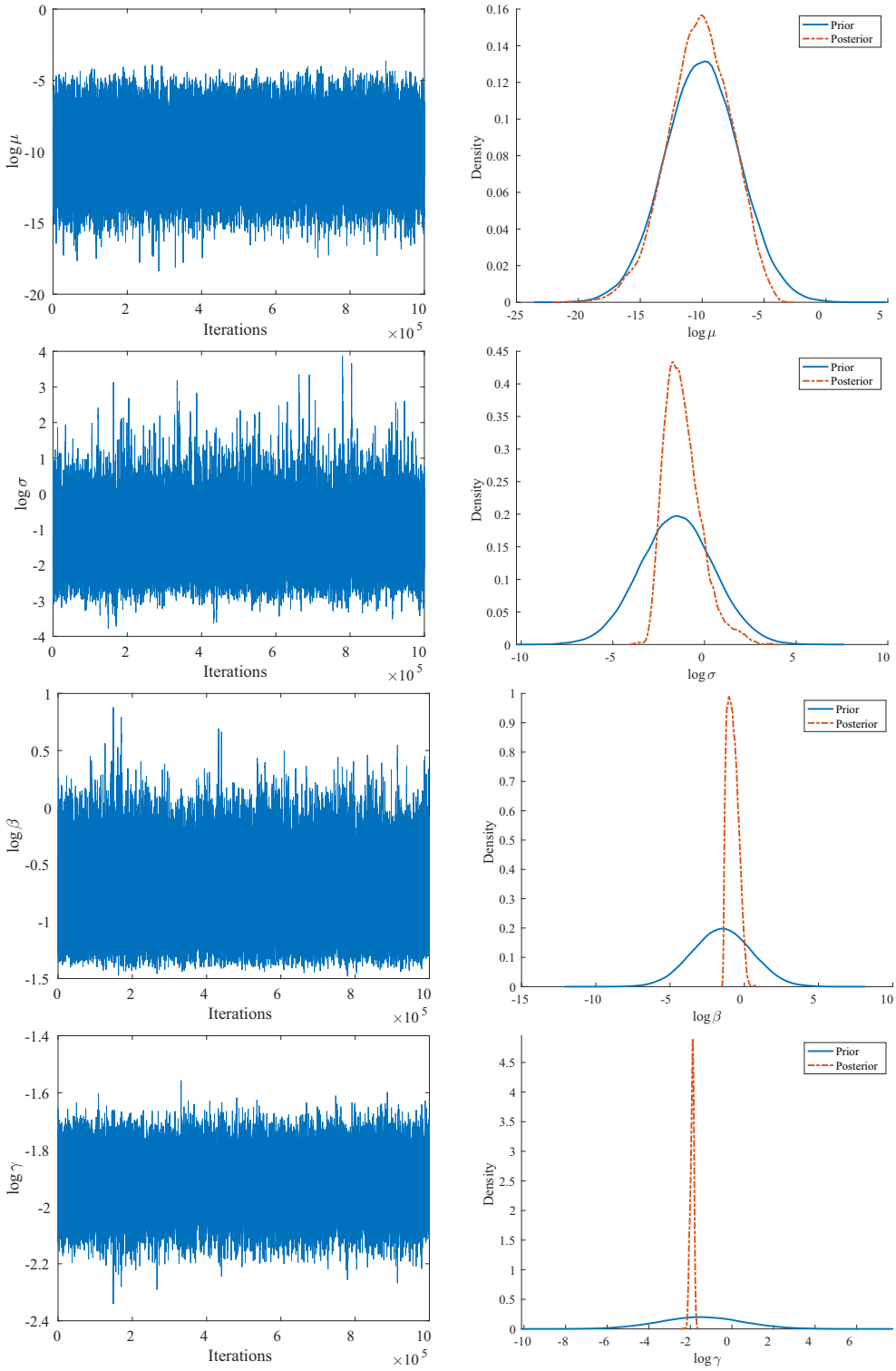

FIG. 9: Results of Bayesian parameter estimation for SEIR example. Chains are constructed by taking 10^6 posterior samples after 10^4 iterations of burn-in and 10^6 iterations for adapting the proposal distribution. Marginal distributions are constructed by kernel density estimation on the respective MCMC chains.

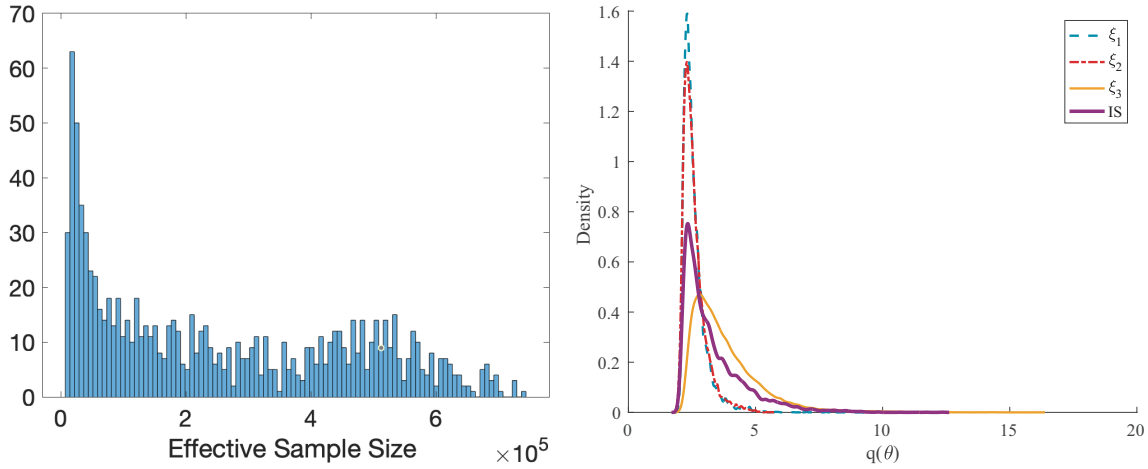

FIG. 10: Left: Distribution of effective sample sizes when using 10^6 samples from $\pi^{\text{IS}}_{\text{post}}$ to perform importance sampling for 1000 target distributions $\pi^{\boldsymbol{\xi}}_{\text{post}}$. Right: Kernel density estimates of the distribution for $\pi^{\text{IS}}_{\text{post}}$ compared to those for selected target posterior distributions with prior hyperparameters.

- $\boldsymbol{\xi}_1 = \begin{bmatrix} -15 & -2.25 & -0.75 & -0.75 & 0.5 & 1.5 & 0.5 & 1.5 \end{bmatrix}^\top$,
- $\boldsymbol{\xi}_2 = \begin{bmatrix} -5 & -0.75 & -2.25 & -2.25 & 1.5 & 0.5 & 1.5 & 0.5 \end{bmatrix}^\top$,
- and $\boldsymbol{\xi}_3 = \begin{bmatrix} -15 & -0.75 & -0.75 & -0.75 & 0.5 & 1.5 & 0.5 & 1.5 \end{bmatrix}^\top$.

Compared to the effective sample sizes observed for the linear example in Fig. 4, there is a larger proportion of target distributions $\pi^{\boldsymbol{\xi}}_{\text{post}}$ for which the effective sample size is small. Still, for the majority of the target distributions, the effective sample size is sufficient.

5.2.3 Sensitivity Analysis

Here, we study the sensitivity of the HS mappings F_{mean}, F_{var}, and F_{atMAP} to prior hyperparameters, relative to the QoI $q(\boldsymbol{\theta}) = R_0$. As discussed in Section 4, F_{atMAP} is not evaluated the same way as the other two HS mappings—it is evaluated by solving an optimization problem. Therefore, we only include convergence studies for F_{mean} and F_{var}.

We start by studying the pick-freeze estimators of total Sobol' indices for F_{mean} and F_{var}. To reduce the computational burden, we thin the MCMC chain from 10^6 iterates to 2×10^4 iterates. We track the convergence of these indices as we increase the number of MCMC samples from $\pi^{\text{IS}}_{\text{post}}$ up to 2×10^4 samples. The pick-freeze estimators use 10^6 design points, requiring 10^7 total evaluations of the HS mapping. The results are reported in Fig. 11 for F_{mean} and in Fig. 12 for F_{var}.

For each HS mapping, we also construct surrogate models using 10^3 realizations of $\boldsymbol{\xi}$, drawn using LHS. For polynomial chaos expansion surrogates, we use expansions of total degree 5. SW-ELM surrogates use 1000 realizations for training and an additional 500 realizations for validation during the weight sparsification step. We use these surrogates to approximate the first-order and total Sobol' indices of the approximate HS mappings. The surrogate estimates are compared to the pick-freeze estimators in Fig. 11 for F_{mean} and in Fig. 12 for F_{var}.

We must use a large proportion of the MCMC chain before the pick-freeze estimators fully stabilize. Despite this, the hyperparameter sensitivity rankings stabilize much faster, after one-fourth of the chain is used to approximate the HS mappings in Fig. 12.

GSA of F_{atMAP}, defined in Eq. (5), does not make use of Algorithm 1. Instead, as discussed in Section 4.3, evaluating the HS mapping for each realization requires solving an optimization problem. This costly process makes estimating the pick-freeze estimators of the HS mapping too expensive. We instead perform GSA using only the

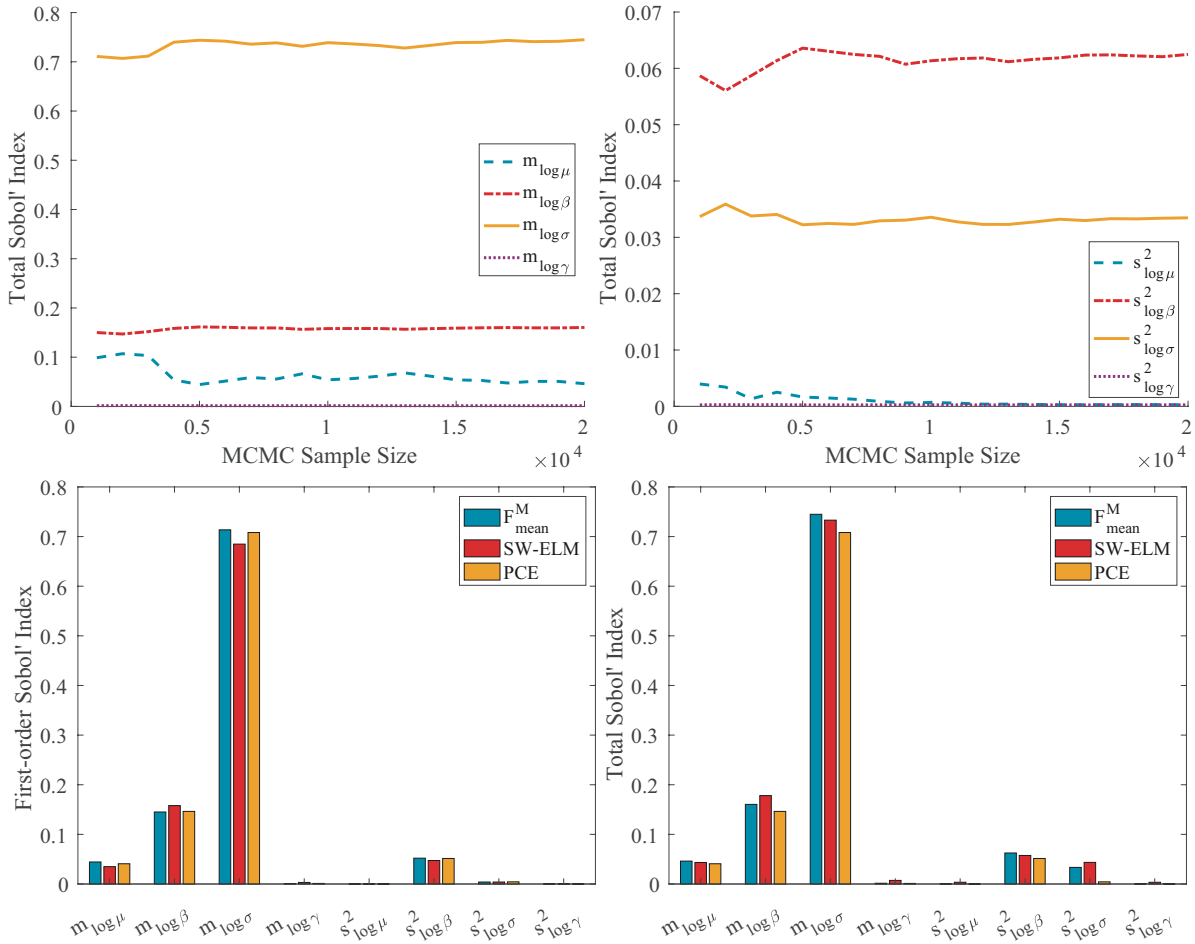

FIG. 11: Top row: Convergence of pick-freeze estimators of the total indices of $\tilde{F}^M_{\text{mean}}(\xi)$ with increasing MCMC sample size M. Total indices of the mean hyperparameters and variance hyperparameters are displayed separately. Bottom row: First-order and total Sobol indices estimated by SW-ELM and PCE surrogates compared to pick-freeze estimators, labeled $F^M_{\text{mean}}(\xi)$. Pick-freeze estimators use 10^6 design points. HS mapping approximations use $M = 2 \times 10^4$ samples.

surrogates. SW-ELM and PCE surrogates are constructed using the same sets of realizations used to build surrogates for $\tilde{F}^M_{\text{mean}}$ and $\tilde{F}^M_{\text{var}}$. The Sobol' index estimates are summarized in Fig. 13.

The results from the SW-ELM and sparse regression PCE surrogates provide comparable summaries. This suggests that the present computations are stable with respect to the choice of the surrogate model. The global sensitivity analysis of F_{mean}, F_{var}, and F_{atMAP} allow us to infer much information about which hyperparameters in the prior matter and which do not. The Sobol' indices suggest that the uncertainty in the prior mean of $\log\gamma$ and prior variances of $\log\mu, \log\gamma$ can be ignored because their respective Sobol' indices are close to zero. It is encouraging that GSA of the different HS mappings yields similar conclusions because the methods for GSA of F_{mean} and F_{var} are different from the methods used for F_{atMAP}. To illustrate this is the case, we compare the distributions of F_{mean}, F_{var}, and F_{atMAP} to the distributions of their respective reduced mappings, for which the unimportant prior hyperparameter inputs are fixed at nominal values in Fig. 14. The density estimates in Fig. 14 confirm that those three prior hyperparameters have little influence over the posterior mean, variance, and MAP point. Thus, the experimental resources should be put towards finding more information about the other hyperparameters.

We lastly quantify the approximation errors resulting from building the surrogate models. A verification set of 1000 realizations of ξ is drawn from the distribution on the prior hyperparameters. For F_{mean} and F_{var}, we can only

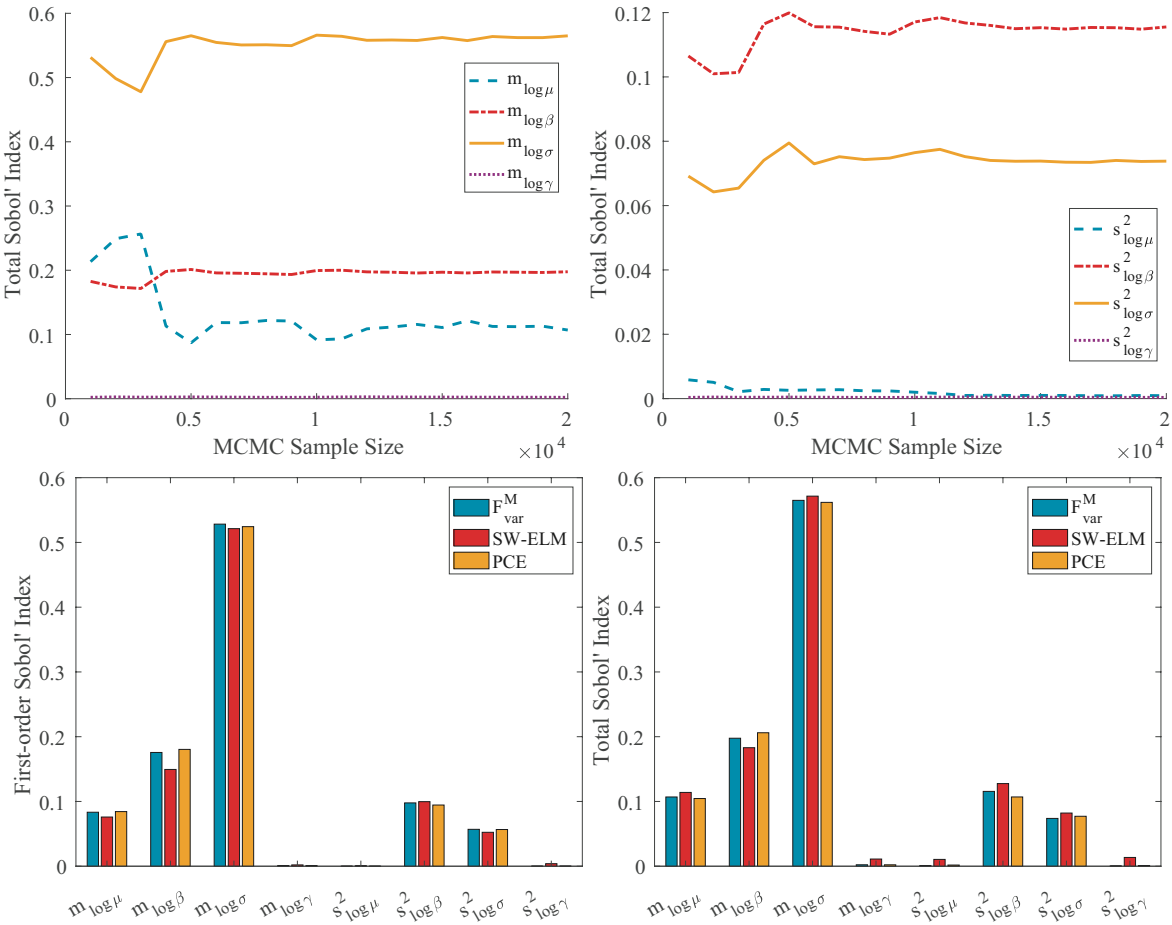

FIG. 12: Top row: Convergence of pick-freeze estimators of the total indices of $\tilde{F}^M_{\text{var}}(\boldsymbol{\xi})$ with increasing MCMC sample size M. Total indices of the mean hyperparameters and variance hyperparameters are displayed separately. Bottom row: First-order and total Sobol indices estimated by SW-ELM and PCE surrogates compared to pick-freeze estimators, labeled $F^M_{\text{var}}(\boldsymbol{\xi})$. Pick-freeze estimators use 10^6 design points. HS mapping approximations use $M = 2 \times 10^4$ samples.

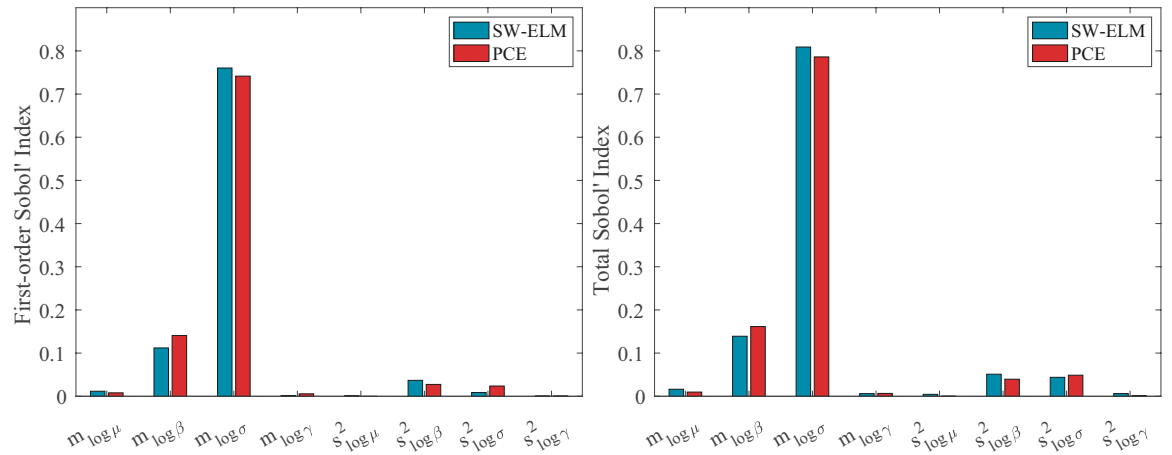

FIG. 13: First-order and total Sobol' indices of F_{atMAP} estimated by SW-ELM and PCE surrogates. Each evaluation of F_{atMAP} (5) is computed by solving an optimization problem.

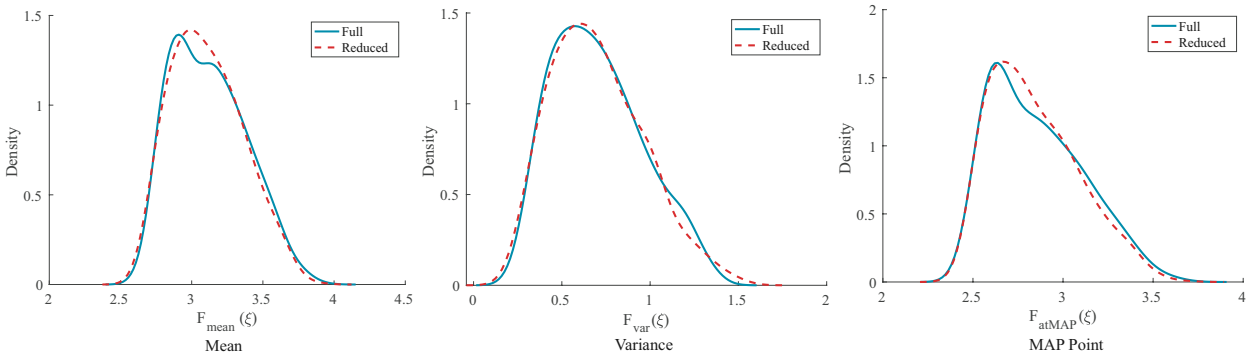

FIG. 14: Kernel density estimates of $F(\boldsymbol{\xi})$ when sampled over all prior hyperparameters compared to when the prior mean of $\log\gamma$ and prior variances of $\log\mu$, $\log\gamma$ are fixed at the nominal values of –1, 1, and 1, respectively. These estimates on superimposed on one another for $F_{\text{mean}}(\boldsymbol{\xi})$ (left), $F_{\text{var}}(\boldsymbol{\xi})$ (right), and $F_{\text{atMAP}}(\boldsymbol{\xi})$ (bottom).

examine the error between the surrogates and the approximate HS mappings, rather than the true HS mappings. For F_{atMAP} we can assess the error between the surrogate and the true HS mapping, but do not have pick-freeze estimators to compare the surrogate estimated Sobol' indices against. The error estimates are summarized in Table 4.

6. CONCLUSION

We have developed a computational approach quantifying sources of uncertainty unique to Bayesian inverse problems. The approach enables in-depth analysis of a problem's dependence on uncertainties in prior hyperparameters. Global sensitivity analysis indicates that the posterior distribution can exhibit complex dependence on such hyperparameters. Consequently, the uncertainty in the prior hyperparameters leads to uncertainty in posterior statistics of the prediction/goal quantities of interest which needs to be accounted for. The results of GSA provide valuable insight in this context. Such an analysis reveals the prior hyperparameters that are most influential on the posterior statistics of prediction quantities of interest and whose specification requires care. Our computational studies provide a proof-of-concept of the proposed approach and indicate its viability. In particular, at the cost of one MCMC run, we can obtain reliable estimates of the sensitivity of moment-based hyperparameter-to-statistic mappings with respect to prior hyperparameters.

The steps taken in this work show that there are numerous challenges and promising future directions to build towards. An important aspect of our approach is the proposed importance sampling procedure. A limitation of the

TABLE 4: Left: Relative approximation errors, defined by Eq. (31), capturing error in approximating $\tilde{F}^M_{\text{mean}}$ and $\tilde{F}^M_{\text{mean}}$ using SW-ELM and PCE surrogates. Right: Relative approximation errors, defined by Eq. (31), capturing error in approximating F_{atMAP} using SW-ELM and PCE surrogates. Bottom: Relative error in total Sobol' index estimates, defined by Eq. (32), comparing surrogate estimates to pick-freeze estimators. Surrogate approximation captures deviation of the surrogate from the approximate HS mappings $\tilde{F}^M_{\text{mean}}$ and $\tilde{F}^M_{\text{mean}}$

Error type	$\boldsymbol{F_{\text{mean}}}$	$\boldsymbol{F_{\text{var}}}$
$\epsilon_{\text{rel}}(\tilde{F}^M, \tilde{F}^M_{\text{ELM}})$	1.93×10^{-3}	1.49×10^{-2}
$\epsilon_{\text{rel}}(\tilde{F}^M, \tilde{F}^M_{\text{PCE}})$	1.40×10^{-3}	1.46×10^{-2}

Error type	$\boldsymbol{F_{\text{atMAP}}}$
$\epsilon_{\text{rel}}(F, \tilde{F}^M_{\text{ELM}})$	8.77×10^{-4}
$\epsilon_{\text{rel}}(F, \tilde{F}^M_{\text{PCE}})$	5.41×10^{-4}

Error type	$\boldsymbol{F_{\text{mean}}}$	$\boldsymbol{F_{\text{var}}}$
$S_{\text{rel}}(\tilde{F}^M, \tilde{F}^M_{\text{ELM}})$	2.52×10^{-2}	3.83×10^{-2}
$S_{\text{rel}}(\tilde{F}^M, \tilde{F}^M_{\text{PCE}})$	3.92×10^{-3}	2.11×10^{-2}

present study is that the importance sampling prior in Eq. (12) was chosen in an empirical manner. While this can be practical in many cases, developing a systematic approach for picking this distribution is an interesting and important avenue of future investigations. This can be facilitated, e.g., by considering an appropriate optimization problem for finding $\pi_{\text{pr}}^{\text{IS}}$. This requires definition of suitable performance objectives for $\pi_{\text{pr}}^{\text{IS}}$ that are tractable to optimize. Progress has been made in this area in Chapter 5 of [49]. There are still cases where there is no single importance sampling posterior that can cover all high-density regions of the target posteriors. This can occur, for example, if the high-density regions of the target priors have little or no overlap. Future work should explore possible remedies, including choosing the importance sampling prior from a different family than that of the target priors, building an importance sampling prior out of a suitable mixture of target priors, using different importance sampling distributions for different groupings of target distributions, and multiple importance sampling—see Chapter 9 in [22].

The computational studies in this article are limited to inverse problems with low-dimensional inversion parameters. Also, our framework assumes a prior that belongs to a parameterized family. There are many opportunities for expanding the present framework. MCMC for high-dimensional Bayesian inverse problems is computationally expensive but the cost can be eased by replacing the forward model with a surrogate or using multifidelity methods [64]. Infinite-dimensional Bayesian inverse problems can be solved through scalable MCMC methods [65–67]. Exploration of hyperparameters that affect the likelihood, as done in [17] for statistical models, is particularly intriguing. Our importance sampling approach promises computational efficiency as long as excess evaluations of the forward model are avoided. Uncertainty in certain likelihood parameters, such as hyperparameters affecting the noise model, can be studied through importance sampling. The practicality of using importance sampling to study uncertainty in model parameters depends on the model structure.

Future work should consider an alternate framework of approximating the posterior using, for example, the Laplace approximation as in [68,69] to propagate uncertainty from model hyperparameters. Approximating posteriors in this way is appropriate for examining the effects of hyperparameter uncertainty on other types of analysis, such as optimal experimental design [70,71].

Yet another direction for future work is the development of hyperparameter screening steps. A tried-and-true approach is to screen via derivative-based global sensitivity measures [72,73], after which a variance-based analysis may be conducted. Such derivative-based approaches may be accelerated using tools from hyperdifferential sensitivity analysis [12,16]. This would be important for inverse problems with a large number of prior hyperparameters.

ACKNOWLEDGMENTS

The work of J.E. Darges, A. Alexanderian, and P.A. Gremaud was supported in part by the U.S. National Science Foundation under Grant Nos. DMS-1745654 and DMS-1953271. Additionally, the work of A. Alexanderian was also supported in part by the U.S. National Science Foundation under Grant No. DMS-2111044.

REFERENCES

1. Scales, J.A. and Tenorio, L., Prior Information and Uncertainty in Inverse Problems, *Geophysics*, **66**(2):389–397, 2001.
2. Berger, J.O., Robust Bayesian Analysis: Sensitivity to the Prior, *J. Stat. Plann. Inference*, **25**(3):303–328, 1990.
3. Berger, J.O., Moreno, E., Pericchi, L., Bayarri, M., Bernardo, J., Cano, J., Horra, J., Martín, J., Rios, D., Betrò, B., Dasgupta, A., Gustafson, P., Wasserman, L., Kadane, J., Srinivasan, C., Lavine, M., O'Hagan, A., Polasek, W., Robert, C., and Sivaganesan, S., An Overview of Robust Bayesian Analysis, *Test*, **3**:5–124, 1994.
4. Hill, S.D. and Spall, J.C., Sensitivity of a Bayesian Analysis to the Prior Distribution, *IEEE Trans. Sys. Man Cybern.*, **24**:216–221, 1994.
5. Berger, J.O., Insua, D.R., and Ruggeri, F., Bayesian Robustness, in *Robust Bayesian Analysis*, D.R. Insua and F. Ruggeri, Eds., Berlin: Springer, pp. 1–32, 2000.
6. Lopes, H. and Tobias, J., Confronting Prior Convictions: On Issues of Prior Sensitivity and Likelihood Robustness in Bayesian Analysis, *Ann. Rev. Econ.*, **3**:107–131, 2011.
7. Owhadi, H., Scovel, C., and Sullivan, T., On the Brittleness of Bayesian Inference, *SIAM Rev.*, **57**(4):566–582, 2015.
8. Watson, J. and Holmes, C., Approximate Models and Robust Decisions, *Stat. Sci.*, **31**(4):465–489, 2016.

9. Giordano, R., Broderick, T., and Jordan, M.I., Covariances, Robustness, and Variational Bayes, *J. Mach. Learn. Res.*, **19**:1–58, 2018.
10. Gelman, A., Carlin, J.B., Stern, H.S., Dunson, D.B., Vehtari, A., and Rubin, D.B., *Bayesian Data Analysis*, 3rd ed., Boca Raton, FL: Chapman and Hall/CRC, 2014.
11. van de Schoot, R., Depaoli, S., King, R., Kramer, B., Märtens, K., Tadesse, M.G., Vannucci, M., Gelman, A., Veen, D., Willemsen, J., and Yau, C., Bayesian Statistics and Modelling, *Nat. Rev. Methods Primers*, **1**(1):1–26, 2021.
12. Sunseri, I., Alexanderian, A., Hart, J., and van Bloemen Waanders, B., Hyper-Differential Sensitivity Analysis for Nonlinear Bayesian Inverse Problems, *Int. J. Uncertainty Quantif.*, **14**(2):1–20, 2024.
13. Reese, W.M., Hart, J., van Bloemen Waanders, B., Perego, M., Jakeman, J., and Saibaba, A., Hyper-Differential Sensitivity Analysis in the Context of Bayesian Inference Applied to Ice-Sheet Problems, *Int. J. Uncertainty Quantif.*, **14**(3):1–20, 2024.
14. Chowdhary, A., Tong, S., Stadler, G., and Alexanderian, A., Sensitivity Analysis of the Information Gain in Infinite-Dimensional Bayesian Linear Inverse Problems, *Int. J. Uncertainty Quantif.*, **14**(6):17–35, 2024.
15. Hart, J., van Bloemen Waanders, B., and Herzog, R., Hyper-Differential Sensitivity Analysis of Uncertain Parameters in PDE-Constrained Optimization, *Int. J. Uncertainty Quantif.*, **10**(3):225–248, 2020.
16. Sunseri, I., Hart, J., van Bloemen Waanders, B., and Alexanderian, A., Hyper-Differential Sensitivity Analysis for Inverse Problems Constrained by Partial Differential Equations, *Inv. Probl.*, **36**(12):125001, 2020.
17. Vernon, I. and Gosling, J.P., A Bayesian Computer Model Analysis of Robust Bayesian Analyses, *Bayesian Anal.*, 18(4):1367–1399, 2022.
18. Geyer, C.J. and Thompson, E.A., Constrained Monte Carlo Maximum Likelihood for Dependent Data, *J. R. Stat. Soc.*, **54**(3):657–699, 1992.
19. Madras, N. and Piccioni, M., Importance Sampling for Families of Distributions, *Ann. Appl. Prob.*, **9**(4):1202–1225, 1999.
20. Geyer, C.J. and Thompson, E.A., Annealing Markov Chain Monte Carlo with Applications to Ancestral Inference, *J. Am. Stat. Assoc.*, **90**(431):909–920, 1995.
21. Geyer, C.J., Importance Sampling, Simulated Tempering, and Umbrella Sampling, in *Handbook of Markov Chain Monte Carlo*, S. Brooks, A. Gelman, G. Jones, and X.L. Meng, Eds., Boca Raton, FL: Chapman and Hall/CRC Press, pp. 295–311, 2011.
22. Owen, A.B., Monte Carlo Theory, Methods and Examples, from https://artowen.su.domains/mc/, 2013.
23. Merritt, M., Alexanderian, A., and Gremaud, P.A., Global Sensitivity Analysis of Rare Event Probabilities Using Sublet Simulation and Polynomial Chaos Expansions, *Int. J. Uncertainty Quantif.*, **13**(1):53–67, 2023.
24. Tarantola, A., *Inverse Problem Theory and Methods for Model Parameter Estimation*, 1st ed., Philadelphia, PA, USA: Society for Industrial and Applied Mathematics, 2005.
25. Saltelli, A. and Sobol', I., Sensitivity Analysis for Nonlinear Mathematical Models: Numerical Experience, *Mat. Model.*, **7**(11):16–28, 1995.
26. Sobol', I., Global Sensitivity Indices for Nonlinear Mathematical Models and Their Monte Carlo Estimates, *Math. Comput. Simul.*, **55**(1-3):271–280, 2001.
27. Saltelli, A., Ratto, M., Andres, T., Campolongo, F., Cariboni, J., Gatelli, D., Saisana, M., and Tarantola, S., *Global Sensitivity Analysis: The Primer*, 1st ed., Chichester, UK: John Wiley & Sons, 2008.
28. Smith, R.C., *Uncertainty Quantification: Theory, Implementation, and Applications*, 1st ed., Philadelphia, PA, USA: Society for Industrial and Applied Mathematics, 2014.
29. Iooss, B. and Lemaître, P., A Review on Global Sensitivity Analysis Methods, *Uncertainty Management in Simulation-Optimization of Complex Systems: Algorithms and Applications*, G. Dellino and C. Meloni, Eds., Berlin: Springer, pp. 101–122, 2015.
30. Prieur, C. and Tarantola, S., Variance-Based Sensitivity Analysis: Theory and Estimation Algorithms, in *Handbook of Uncertainty Quantification*, R. Ghanem, D. Higdon, and H. Owhadi, Eds., Berlin: Springer, pp. 1217–1239, 2017.
31. Saltelli, A., Annoni, P., Azzini, I., Campolongo, F., Ratto, M., and Tarantola, S., Variance Based Sensitivity Analysis of Model Output. Design and Estimator for the Total Sensitivity Index, *Comput. Phys. Commun.*, **181**(2):259–270, 2010.
32. Sargsyan, K., Surrogate Models for Uncertainty Propagation and Sensitivity Analysis, in *Handbook of Uncertainty Quantification*, R. Ghanem, D. Higdon, and H. Owhadi, Eds., Berlin: Springer, pp. 673–698, 2017.

33. Le Gratiet, L., Marelli, S., and Sudret, B., Metamodel-Based Sensitivity Analysis: Polynomial Chaos Expansions and Gaussian Processes, in *Handbook of Uncertainty Quantification*, R. Ghanem, D. Higdon, and H. Owhadi, Eds., Berlin: Springer, pp. 1289–1326, 2017.
34. Sudret, B., Global Sensitivity Analysis Using Polynomial Chaos Expansions, *Reliab. Eng. Syst. Saf.*, **93**(7):964–979, 2008.
35. Crestaux, T., Le Maître, O., and Martinez, J.M., Polynomial Chaos Expansion for Sensitivity Analysis, *Reliab. Eng. Syst. Saf.*, **94**(7):1161–1172, 2009.
36. Le Maître, O.P. and Knio, O.M., *Spectral Methods for Uncertainty Quantification*, 1st ed., Dordrecht, Netherlands: Springer, 2010.
37. Blatman, G. and Sudret, B., Efficient Computation of Global Sensitivity Indices Using Sparse Polynomial Chaos Expansions, *Reliab. Eng. Syst. Saf.*, **95**(11):1216–1229, 2010.
38. Blatman, G. and Sudret, B., Adaptive Sparse Polynomial Chaos Expansion Based on Least Angle Regression, *J. Comput. Phys.*, **230**(6):2345–2367, 2011.
39. Hampton, J. and Doostan, A., Compressive Sampling Methods for Sparse Polynomial Chaos Expansions, in *Handbook of Uncertainty Quantification*, R. Ghanem, D. Higdon, and H. Owhadi, Eds., Berlin: Springer, pp. 827–855, 2017.
40. Fajraoui, N., Marelli, S., and Sudret, B., Sequential Design of Experiment for Sparse Polynomial Chaos Expansions, *SIAM/ASA J. Uncertainty Quantif.*, **5**(1):1061–1085, 2017.
41. van den Berg, E. and Friedlander, M.P., Probing the Pareto Frontier for Basis Pursuit Solutions, *SIAM J. Sci. Comput.*, **31**(2):890–912, 2008.
42. van den Berg, E. and Friedlander, M.P., SPGL1: A Solver for Large-Scale Sparse Reconstruction, from https://friedlander.io/spgl1, 2019.
43. Darges, J.E., Alexanderian, A., and Gremaud, P.A., Extreme Learning Machines for Variance-Based Global Sensitivity Analysis, *Int. J. Uncertainty Quantif.*, **14**(4):83–103, 2024.
44. Hashemi, A., Schaeffer, H., Shi, R., Topcu, U., Tran, G., and Ward, R., Generalization Bounds for Sparse Random Feature Expansions, *Appl. Comput. Harmonic Anal.*, **62**:310–330, 2023.
45. Huang, G.B., Zhu, Q.Y., and Siew, C.K., Extreme Learning Machine: Theory and Applications, *Neurocomputing*, **70**(1):489–501, 2006.
46. Huang, G.B., Wang, D., and Lan, Y., Extreme Learning Machines: A Survey, *Int. J. Mach. Learn. Cybernet.*, **2**(2):107–122, 2011.
47. Besag, J., Green, P., Higdon, D., and Mengersen, K., Bayesian Computation and Stochastic Systems, *Stat. Sci.*, **10**(1):3–41, 1995.
48. Tokdar, S. and Kass, R., Importance Sampling: A Review, *Wiley Interdisc. Rev.: Comput. Stat.*, **2**:54–60, 2010.
49. Darges, J.E., Sensitivity Analysis in Forward and Inverse Problems, PhD, North Carolina State University, 2024.
50. Zhang, J. and Shields, M.D., On the Quantification and Efficient Propagation of Imprecise Probabilities Resulting from Small Datasets, *Mech. Syst. Signal Process.*, **98**:465–483, 2018.
51. McKay, M.D., Beckman, R.J., and Conover, W.J., A Comparison of Three Methods for Selecting Values of Input Variables in the Analysis of Output from a Computer Code, *Technometrics*, **21**(2):239–245, 1979.
52. Viana, F.A., A Tutorial on Latin Hypercube Design of Experiments, *Qual. Reliab. Eng. Int.*, **32**(5):1975–1985, 2016.
53. Geyer, C.J., Practical Markov Chain Monte Carlo, *Stat. Sci.*, **7**(4):473–483, 1992.
54. Haario, H., Laine, M., Mira, A., and Saksman, E., DRAM: Efficient Adaptive MCMC, *Stat. Comput.*, **16**(4):339–354, 2006.
55. Laine, M., Adaptive MCMC Methods with Applications in Environmental and Geophysical Model, PhD, Finnish Meteorological Institute, 2008.
56. Hethcote, H.W., The Mathematics of Infectious Diseases, *SIAM Rev.*, **42**(4):599–653, 2000.
57. Fox, M.P. and Gower, E.W., Infectious Disease Epidemiology, in *Modern Epidemiology*, T.L. Lash, T.J. VanderWeele, S. Haneuese, and K.J. Rothman, Eds., Philadelphia, PA: Lippincott Williams & Wilkin, pp. 1175–1232, 2021.
58. Withers, C., The Moments of the Multivariate Normal, *Bull. Austral. Math. Soc.*, **32**:103–107, 1985.
59. Holmquist, B., Moments and Cumulants of the Multivariate Normal Distribution, *Stoch. Anal. Appl.*, **6**(3):273–278, 1985.
60. Triantafyllopoulos, K., On the Central Moments of the Multidimensional Gaussian Distribution, *Math. Sci.*, **28**:125–128, 2003.

61. Alexanderian, A., On the Mean and Variance of Quadratic Functionals of Gaussian Random Vectors, from https://aalexan3.math.ncsu.edu/articles/moments_quad.pdf, 2021.
62. Ndanguza Rusatsi, D., Bayesian Analysis of SEIR Epidemic Models, PhD, Lappeenranta University of Technology, 2015.
63. Lloyd, A., Sensitivity of Model-Based Epidemiological Parameter Estimation to Model Assumptions, in *Mathematical and Statistical Estimation Approaches in Epidemiology*, G. Chowell, J.M. Hyman, L.M.A. Bettencourt, and C. Castillo-Chavez, Eds., Berlin: Springer, pp. 123–141, 2009.
64. Peherstorfer, B., Cui, T., Marzouk, Y., and Willcox, K., Multifidelity Importance Sampling, *Comput. Methods Appl. Mech. Eng.*, **300**:490–509, 2016.
65. Petra, N., Martin, J., Stadler, G., and Ghattas, O., A Computational Framework for Infinite-Dimensional Bayesian Inverse Problems, Part II: Stochastic Newton MCMC with Application to Ice Sheet Flow Inverse Problems, *SIAM J. Sci. Comput.*, **36**(4):A1525–A1555, 2014.
66. Rudolf, D. and Sprungk, B., On a Generalization of the Preconditioned Crank-Nicolson Metropolis Algorithm, *Found. Comput. Math.*, **18**:309–343, 2018.
67. Saibaba, A.K., Bardsley, J., Brown, D.A., and Alexanderian, A., Efficient Marginalization-Based MCMC Methods for Hierarchical Bayesian Inverse Problems, *SIAM/ASA J. Uncertainty Quantif.*, **7**(3):1105–1131, 2019.
68. Isaac, T., Petra, N., Stadler, G., and Ghattas, O., Scalable and Efficient Algorithms for the Propagation of Uncertainty from Data through Inference to Prediction for Large-Scale Problems, with Application to Flow of the Antarctic Ice Sheet, *J. Comput. Phys.*, **296**:348–368, 2015.
69. Villa, U., Petra, N., and Ghattas, O., HIPPYlib: An Extensible Software Framework for Large-Scale Inverse Problems Governed by PDEs: Part I: Deterministic Inversion and Linearized Bayesian Inference, *ACM Trans. Math. Software (TOMS)*, **47**(2):1–34, 2021.
70. Attia, A., Alexanderian, A., and Saibaba, A.K., Goal-Oriented Optimal Design of Experiments for Large-Scale Bayesian Linear Inverse Problems, *Inv. Probl.*, **34**(9):095009, 2018.
71. Alexanderian, A., Optimal Experimental Design for Infinite-Dimensional Bayesian Inverse Problems Governed by PDEs: A Review, *Inv. Probl.*, **37**(4):043001, 2021.
72. Sobol', I. and Kucherenko, S., Derivative Based Global Sensitivity Measures and Their Link with Global Sensitivity Indices, *Math. Comput. Simul.*, **79**(10):3009–3017, 2009.
73. Kucherenko, S. and Iooss, B., Derivative Based Global Sensitivity Measures, in *Handbook of Uncertainty Quantification*, R. Ghanem, D. Higdon, and H. Owhadi, Eds., Berlin: Springer, pp. 1241–1263, 2017.