International Journal for Uncertainty Quantification, 5 (4): 341–359 (2015)

SURROGATE PREPOSTERIOR ANALYSES FOR PREDICTING AND ENHANCING IDENTIFIABILITY IN MODEL CALIBRATION

Zhen Jiang,[1] Daniel W. Apley,[2] & Wei Chen[1,]*

[1] *Department of Mechanical Engineering, Northwestern University, 2145 Sheridan Road, Evanston, Illinois 60208, USA*

[2] *Department of Industrial Engineering and Management Sciences, Northwestern University, 2145 Sheridan Road, Evanston, Illinois 60208, USA*

Original Manuscript Submitted: 12/07/2014; Final Draft Received: 05/05/2015

In physics-based engineering modeling and uncertainty quantification, distinguishing the effects of two main sources of uncertainty — calibration parameter uncertainty and model discrepancy — is challenging. Previous research has shown that identifiability, which is quantified by the posterior covariance of the calibration parameters, can sometimes be improved by experimentally measuring multiple responses of the system that share a mutual dependence on a common set of calibration parameters. In this paper, we address the issue of how to select the most appropriate subset of responses to measure experimentally, to best enhance identifiability. We use a preposterior analysis approach that, prior to conducting physical experiments but after conducting computer simulations, can predict the degree of identifiability that will result using different subsets of responses to measure experimentally. It predicts identifiability via the preposterior covariance from a modular Bayesian Monte Carlo analysis of a multi-response spatial random process (SRP) model. Furthermore, to handle the computational challenge in preposterior analysis, we propose a surrogate preposterior analysis based on Fisher information of the calibration parameters. The proposed methods are applied to a simply supported beam example to select two out of six responses to best improve identifiability. The estimated preposterior covariance is compared to the actual posterior covariance to demonstrate the effectiveness of the methods.

KEY WORDS: *model calibration, uncertainty quantification, parameter estimation, Bayesian inference*

1. INTRODUCTION

Computer simulations have been widely used for design and optimization in many fields of science and engineering, as they are often much less expensive than physical experiments in analyzing complex systems. However, computer simulations never agree completely with experiments because no computer model is perfect. Several sources of uncertainty accounting for the differences between computer simulations and physical experiments have been reported in the literature [1, 2]. *Parameter uncertainty* and *model discrepancy* are typically the two main sources; the former is due to the lack of knowledge of physical parameters (e.g., friction coefficient in a finite element analysis) that are naturally fixed but unknown and cannot be directly observed in physical experiments, while the latter is associated with the lack of knowledge of the underlying true physics. Other sources of uncertainty may include *numerical uncertainty* due to numerical errors in implementing computer models, *experimental uncertainty* due to random observational error when taking experimental measurements, and uncertainty due to randomly varying physical parameters. To analyze the differences between computer simulations and physical experiments and to adjust simulation models to better reflect the reality, several model uncertainty quantification methodologies [1–5] have been developed for learning these uncertainties via combining simulation data with physical experimental data. Adjusting predictive models based on

*Correspond to Wei Chen, E-mail: weichen@northwestern.edu, URL: http://ideal.mech.northwestern.edu/

2152–5080/15/$35.00

identifying the unknown physical parameters and the model discrepancy are referred to as *calibration* [1] and *bias correction* [6–8], respectively.

Recent studies [9–12] indicate that calibration is usually difficult and that existing methodologies often fail at distinguishing between the effects of parameter uncertainty and model discrepancy and even between the effects of different model parameters. We refer to this issue as (non)*identifiability*. Loosely speaking, identifiability problems occur when different but equally likely combinations of calibration parameters and discrepancy function result in equally good agreement with the observed data. In many engineering applications, good identifiability is virtually impossible if considering a single experimental response, although identifiability is often critically important for a number of reasons. For example, consider material science and mechanical engineering applications in which many material property parameters cannot be measured directly. If these parameters can be accurately identified via a model calibration study comparing experimental data with computer simulations, their estimated values can be used to predict the behaviors of materials in more complex cases. Even for similar cases, better knowledge of the parameters results in more accurate prediction over a broader range of input settings, because the model adjustment from learning unknown parameters is more global than the adjustment from learning the model discrepancy. In addition, learning the calibration parameters may in itself be a primary goal with broad-reaching implications (e.g., for scientific discovery purposes or if an unknown parameter reflects performance of interest but cannot be directly observed or calculated).

Our recent results [11] indicate that, in spite of the identifiability challenges, good identifiability may often be achieved in model uncertainty quantification by measuring multiple experimental responses that are automatically calculated in the simulation and that share a mutual dependence on a common set of calibration parameters. We observed an intriguing phenomenon that some combinations of responses may result in drastically different identifiability than others: We used a simply supported beam example in our previous papers [10, 11] (revisited in Section 4) to show that measuring certain responses will achieve substantial improvement in identifiability, while measuring other combinations of responses provides little improvement in identifiability beyond measuring only a single response. How to select the most appropriate subset of responses to measure experimentally to best enhance identifiability remains a research challenge.

Because of the cost and difficulty in developing and placing apparatus for measuring certain responses, it is generally not feasible to measure experimentally all of the great many responses that are automatically calculated in the simulations. Moreover, it is generally not necessary, because measuring only a subset of the responses may result in acceptable identifiability. One primary objective of this paper is to address the issue of how to select the most appropriate subset of responses to measure experimentally, to best enhance identifiability. We use a *preposterior analysis* approach built upon the approach introduced in [9, 13] that, prior to conducting the physical experiments but after conducting the computer simulations, can predict the relative improvement in identifiability that will result using different subsets of responses. Our preposterior analysis is based on Monte Carlo simulations within a modular Bayesian multi-response spatial random process (SRP) framework. For validation of the approach, the preposterior covariance predictions are compared with the actual posterior covariance calculated after the experiment is conducted. Furthermore, to handle the computational challenges in the preposterior analysis, we introduce a simpler, surrogate preposterior analysis based on the expected Fisher information of the calibration parameters. The expected Fisher information matrix is the frequentist counterpart to the Bayesian preposterior covariance matrix of the parameters. We demonstrate that, while being much faster to calculate than the preposterior covariance, it still provides a reasonable indication of the resulting identifiability. For real engineering applications with many system responses (and hence many different combinations of responses), we recommend using the surrogate preposterior analysis to eliminate the responses that are unlikely to lead to good identifiability and reduce the number of response combinations that are considered in preposterior analysis.

The remainder of the paper is organized as follows: In Section 2, we briefly review the standard approach to quantify identifiability after observing experimental data, via the posterior covariance of the calibration parameters using a multi-response modular Bayesian approach. Section 3 provides a detailed description of our proposed preposterior analysis and surrogate preposterior analysis, in conjunction with the multi-response modular Bayesian approach, for predicting identifiability prior to conducting physical experiments. In Section 4, the proposed methods are applied to a simply supported beam example to select the best two (out of six) responses to measure experimentally in order to best enhance identifiability. The results of the preposterior and surrogate preposterior analyses are compared to

the results of the posterior analysis (after observing the experimental data) to demonstrate the effectiveness of the methods. Conclusions are drawn in Section 5.

2. REVIEW OF MULTI-RESPONSE SRP MODELING AND CALIBRATION PARAMETER IDENTIFIABILITY

The most popular general model uncertainty quantification framework to assess parameter uncertainty and model discrepancy is a Bayesian one in which the simulation model responses and the physical experimental responses are all viewed as realizations of SRPs. A modular Bayesian approach [1, 11, 14, 15] is typically one of the most useful for calibration and bias correction. Consider the following model [1, 2, 4, 16] that relates the experimental and simulated values of q responses:

$$y_i^e(\mathbf{x}) = y_i^m(\mathbf{x}, \boldsymbol{\theta}^*) + \delta_i(\mathbf{x}) + \varepsilon_i \quad (i = 1, 2, ..., q), \tag{1}$$

where $\mathbf{x} = [x_1, ..., x_d]^T$ denotes a vector of d controllable input variables, $\boldsymbol{\theta} = [\theta_1, ..., \theta_r]^T$ denotes a vector of r calibration parameters (which, although unknown to modelers, can be specified like $\mathbf{x}$ in the simulation), $\boldsymbol{\theta}^*$ denotes their true values, $y_i^m(\mathbf{x}, \boldsymbol{\theta})(i = 1, ..., q)$ denotes the ith response from the simulation model as a function of $\mathbf{x}$ and $\boldsymbol{\theta}$, $y_i^e(\mathbf{x})(i = 1, ..., q)$ denotes the same response but as measured in the physical experiment, $\delta_i(\mathbf{x})(i = 1, ..., q)$ is a model discrepancy function that represents the difference between the ith model response (using the true $\boldsymbol{\theta}^*$) and ith experimental response, and $\varepsilon_i(i = 1, ..., q)$ is the random measurement error. Both $y_i^m(\mathbf{x}, \boldsymbol{\theta})$ and $\delta_i(\mathbf{x})$ are modeled as SRPs. After observing both simulation data and experimental data, the modular Bayesian approach calculates the posterior distributions of $\boldsymbol{\theta}$ and of the discrepancy functions, which subsequently can be combined with the simulation model to improve the response prediction.

In this section, we briefly review the multi-response modular Bayesian approach [5, 11, 14, 15, 17], in particular to calculate the (joint) posterior distribution of calibration parameters based on combining simulation and experimental data. The posterior distribution of calibration parameters provides a quantification of identifiability. The approach is an extension of the single response modular Bayesian approach of Kennedy and O'Hagan [1]. The following assumes the general relationship of Eq. (1), in which all of the responses share a mutual dependence on the same set of calibration parameters $\boldsymbol{\theta}$.

2.1 SRP-Based Metamodeling of the Computer Simulations and Discrepancy Functions

The Bayesian SRP-based metamodeling of simulation responses [18–22] provides an analytical basis for the modular Bayesian approach. Simulation runs, although generally much less expensive than experimental runs, still involve some expense and cannot densely cover the input space, especially with high-dimensional inputs. Hence, metamodels that replace the simulations are useful for inferring the response at input combinations for which no simulations have been conducted [23]. Moreover, metamodels for discrepancy functions are also needed to enable the comparisons between simulation data and experimental data, especially when simulations and experiments are conducted at two different sets of input settings. A Gaussian process model [24] is a widely used SRP model capable of capturing the trend and roughness of the simulation response surface by choosing and fitting a few key parameters, which are called *hyperparameters* of the Gaussian process. The well-known Kriging method [25] is also based on the Gaussian process approach. The SRP modeling technique is compatible with the Bayesian framework to quantify model uncertainty, particularly the type of model parameter uncertainty on which we focus in this paper. It also has the merit of providing a reasonable quantification of the uncertainty of the prediction via a prediction error variance (in addition to a prediction mean) of the experimental responses.

After collecting simulation and experimental data, two separate modules are used to fit the multi-response Gaussian process (MRGP) models for the computer simulations and for the model discrepancy functions, as follows.

Module 1. Multi-response Gaussian process modeling for the computer simulations

A MRGP model is fitted to the simulation data to replace the expensive computer simulations and to predict the value of the simulation response at any input site in the design domain. The MRGP model follows the form [14, 15] that is specified by its prior mean and prior covariance functions:

$$\mathbf{y}^m(\cdot,\cdot) \sim \mathcal{GP}\big(\mathbf{h}^m(\cdot,\cdot)\mathbf{B}^m, \mathbf{\Sigma}^m R^m((\cdot,\cdot),(\cdot,\cdot))\big), \tag{2}$$

where $\mathbf{y}^m(\mathbf{x},\boldsymbol{\theta}) = [y_1^m(\mathbf{x},\boldsymbol{\theta}),...,y_q^m(\mathbf{x},\boldsymbol{\theta})]$ denotes the multiple responses from the computer model. $\mathbf{h}^m(\mathbf{x},\boldsymbol{\theta}) = [h_1^m(\mathbf{x},\boldsymbol{\theta}),...,h_q^m(\mathbf{x},\boldsymbol{\theta})]$ denotes a set of pre-defined regression basis functions and $\mathbf{B}^m = [\boldsymbol{\beta}_1^m,...,\boldsymbol{\beta}_q^m]$ is a matrix of unknown regression coefficients associated with $\mathbf{h}^m(\mathbf{x},\boldsymbol{\theta})$; their product $\mathbf{h}^m(\mathbf{x},\boldsymbol{\theta})\mathbf{B}^m$ represents the prior mean function of the MRGP. The product of an unknown non-spatial $q \times q$ covariance matrix $\mathbf{\Sigma}^m$ and a pre-defined scalar spatial correlation function $R^m((\mathbf{x},\boldsymbol{\theta}),(\mathbf{x}',\boldsymbol{\theta}'))$ represents the prior covariance function of the MRGP, where $(\mathbf{x},\boldsymbol{\theta})$ and $(\mathbf{x}',\boldsymbol{\theta}')$ denote two sets of computer model inputs. A frequently used stationary spatial correlation function is the Gaussian correlation function

$$R^m((\mathbf{x},\boldsymbol{\theta}),(\mathbf{x}',\boldsymbol{\theta}')) = \exp\left\{-\sum_{k=1}^{d}\omega_k^m(x_k - x_k')^2 - \sum_{k=1}^{r}\omega_{d+k}^m(\theta_k - \theta_k')^2\right\}, \tag{3}$$

where $\boldsymbol{\omega}^m = [\omega_1^m,...,\omega_{d+r}^m]$ is the vector of roughness parameters that are used to capture the nonlinearity of the process. The unknowns in this MRGP model are its hyperparameters $\boldsymbol{\phi}^m = \{\mathbf{B}^m, \mathbf{\Sigma}^m, \boldsymbol{\omega}^m\}$.

Suppose that a set of simulation data $\mathbf{Y}^m = [\mathbf{y}_1^m,...,\mathbf{y}_q^m]$, where $\mathbf{y}_i^m = [y_i^m(\mathbf{x}_1^m,\boldsymbol{\theta}_1^m),...,y_i^m(\mathbf{x}_{N_m}^m,\boldsymbol{\theta}_{N_m}^m)]^T$, is collected at N_m input sites $\mathbf{X}^m = [\mathbf{x}_1^m,...,\mathbf{x}_{N_m}^m]^T$ and $\boldsymbol{\Theta}^m = [\boldsymbol{\theta}_1^m,...,\boldsymbol{\theta}_{N_m}^m]^T$. To obtain the estimates of the hyperparameters $\boldsymbol{\phi}^m$, the maximum likelihood estimation (MLE) method is used to maximize the multivariate normal *log-likelihood* function for the simulation data:

$$\begin{aligned}\ln p\big(\mathrm{vec}(\mathbf{Y}^m)|\boldsymbol{\phi}^m\big) = &-\frac{qN_m}{2}\ln(2\pi) - \frac{N_m}{2}\ln\left(|\mathbf{\Sigma}^m|\right) - \frac{q}{2}\ln\left(|\mathbf{R}^m|\right)\\ &-\frac{1}{2}\mathrm{vec}(\mathbf{Y}^m - \mathbf{H}^m\mathbf{B}^m)^T\left(\mathbf{\Sigma}^m \otimes \mathbf{R}^m\right)^{-1}\mathrm{vec}(\mathbf{Y}^m - \mathbf{H}^m\mathbf{B}^m),\end{aligned} \tag{4}$$

where $\mathrm{vec}(\cdot)$ is matrix vectorization (by stacking the columns of the matrix), $\otimes$ denotes the Kronecker product, $\mathbf{R}^m$ is an $N_m \times N_m$ correlation matrix whose ith-row, jth-column entry is $R^m((\mathbf{x}_i^m,\boldsymbol{\theta}_i^m),(\mathbf{x}_j^m,\boldsymbol{\theta}_j^m))$, and $\mathbf{H}^m = \big[\mathbf{h}^m(\mathbf{x}_1^m,\boldsymbol{\theta}_1^m)^T,...,\mathbf{h}^m(\mathbf{x}_{N_m}^m,\boldsymbol{\theta}_{N_m}^m)^T\big]^T$.

Module 2. Multi-response Gaussian process modeling for the discrepancy functions

Similarly to Module 1, a MRGP model

$$\boldsymbol{\delta}(\cdot) \sim \mathcal{GP}\big(\mathbf{h}^\delta(\cdot)\mathbf{B}^\delta, \mathbf{\Sigma}^\delta R^\delta(\cdot,\cdot)\big), \tag{5}$$

is created for the discrepancy functions, where $\boldsymbol{\delta}(\mathbf{x}) = [\delta_1(\mathbf{x}),...,\delta_q(\mathbf{x})]$ denotes the discrepancy functions for multiple responses. $\mathbf{h}^\delta(\mathbf{x}) = [h_1^\delta(\mathbf{x}),...,h_s^\delta(\mathbf{x})]$ denotes a set of pre-defined regression basis functions, $\mathbf{B}^\delta = [\boldsymbol{\beta}_1^\delta,...,\boldsymbol{\beta}_q^\delta]$ is a matrix of unknown regression coefficients, $\mathbf{\Sigma}^\delta$ is an unknown non-spatial $q \times q$ covariance matrix, and $R^\delta(\mathbf{x},\mathbf{x}')$ implemented in this paper is again the Gaussian spatial correlation function

$$R^\delta(\mathbf{x},\mathbf{x}') = \exp\left\{-\sum_{k=1}^{d}\omega_k^\delta(x_k - x_k')^2\right\}, \tag{6}$$

parameterized by a vector of roughness parameters $\boldsymbol{\omega}^\delta = [\omega_1^\delta,...,\omega_d^\delta]$. Assuming the experimental error ε_i $(i = 1,...,q)$ in Eq. (1) are independently, normally distributed with mean 0 and unknown variance λ_i, the prior for the experimental responses is also a MRGP that is the sum of the two MRGPs [Eqs. (2) and (5)] and the random error term, i.e.,

$$\begin{aligned}\mathbf{y}^e(\cdot)\big|(\boldsymbol{\theta} = \boldsymbol{\theta}^*) &\sim \mathcal{GP}\big(\mathbf{m}^e(\cdot,\boldsymbol{\theta}^*), \mathbf{V}^e((\cdot,\boldsymbol{\theta}^*),(\cdot,\boldsymbol{\theta}^*))\big),\\ \mathbf{m}^e(\mathbf{x},\boldsymbol{\theta}) &= \mathbf{h}^m(\mathbf{x},\boldsymbol{\theta})\mathbf{B}^m + \mathbf{h}^\delta(\mathbf{x})\mathbf{B}^\delta,\\ \mathbf{V}^e((\mathbf{x},\boldsymbol{\theta}),(\mathbf{x}',\boldsymbol{\theta})) &= \mathbf{\Sigma}^m R^m((\mathbf{x},\boldsymbol{\theta}),(\mathbf{x}',\boldsymbol{\theta})) + \mathbf{\Sigma}^\delta R^\delta(\mathbf{x},\mathbf{x}') + \mathrm{diag}(\boldsymbol{\lambda}),\end{aligned} \tag{7}$$

where $\mathbf{y}^e(\mathbf{x}) = [y_1^e(\mathbf{x}), ..., y_q^e(\mathbf{x})]$ denotes the multiple responses from experiments, $\boldsymbol{\lambda} = (\lambda_1, ..., \lambda_q)$ is a vector comprising the variance of measurement error, and $\mathrm{diag}(\boldsymbol{\lambda})$ is a diagonal matrix formed by the vector $\boldsymbol{\lambda}$.

The unknown hyperparameters of the MRGP model (together with the yet-to-be-estimated measurement error variance) are $\boldsymbol{\phi}^\delta = \{\mathbf{B}^\delta, \boldsymbol{\Sigma}^\delta, \boldsymbol{\omega}^\delta, \boldsymbol{\lambda}\}$. To obtain the MLEs of the hyperparameters $\boldsymbol{\phi}^\delta$ for this MRGP model, experimental data $\mathbf{Y}^e = [\mathbf{y}_1^e, ..., \mathbf{y}_q^e]$, where $\mathbf{y}_i^e = [y_i^e(\mathbf{x}_1^e), ..., y_i^e(\mathbf{x}_{N_e}^e)]^T$ $(i = 1, ..., q)$, are collected at N_e input sites $\mathbf{X}^e = [\mathbf{x}_1^e, ..., \mathbf{x}_{N_e}^e]^T$. A prior distribution $p(\boldsymbol{\theta})$ of the calibration parameters $\boldsymbol{\theta}$ is specified, with respect to which a likelihood function, constructed using the simulation data $\mathbf{Y}^m$ and experimental data $\mathbf{Y}^e$ and the MLEs of the computer model hyperparameters from Module 1, is marginalized. By maximizing this likelihood function, MLEs of the discrepancy function hyperparameters are obtained. Mathematical details of constructing the likelihood function and calculating the MLEs of hyperparameters $\boldsymbol{\phi}^\delta$ can be found in [9, 11].

Generally speaking, the higher dimensionality (the more input variables and/or calibration parameters) the problem has, the more training data are needed to estimate the hyperparameters of the MRGP models, and the higher the computational cost. The computational cost has at least two contributing factors. The first is the matrix inversion required in Eq. (4), and the second is the fact that there may be multiple local optima during the optimization when finding the MLEs of the hyperparameters. Regarding the latter, robust global optimization algorithms can often alleviate this.

2.2 Quantifying Identifiability of Calibration Parameters

After obtaining the MLEs of the hyperparameters of the two MRGP models in Modules 1 and 2, the posterior distribution of the calibration parameters $\boldsymbol{\theta}$ and of the simulation and experimental responses can be calculated as follows.

Module 3. Calculating posterior distribution of the calibration parameters

Based on Bayes theorem, the posterior of the calibration parameters is

$$p(\boldsymbol{\theta}\big|\mathbf{Y}^m, \mathbf{Y}^e, \hat{\boldsymbol{\varphi}}) = \frac{p(\mathbf{Y}^m, \mathbf{Y}^e\big|\boldsymbol{\theta}, \hat{\boldsymbol{\varphi}})p(\boldsymbol{\theta})}{p(\mathbf{Y}^m, \mathbf{Y}^e\big|\hat{\boldsymbol{\varphi}})}, \tag{8}$$

where $\hat{\boldsymbol{\varphi}}$ is the MLEs of $\boldsymbol{\varphi} = \{\boldsymbol{\Sigma}^m, \boldsymbol{\omega}^m, \boldsymbol{\Sigma}^\delta, \boldsymbol{\omega}^\delta, \boldsymbol{\lambda}\}$. $p(\mathbf{Y}^m, \mathbf{Y}^e\big|\boldsymbol{\theta}, \hat{\boldsymbol{\varphi}})$ is the likelihood function of the collected simulation and experimental data ($\mathbf{Y}^m$ and $\mathbf{Y}^e$) conditioned on $\boldsymbol{\theta}$ and $\hat{\boldsymbol{\varphi}}$; it is achieved by marginalizing $p(\mathbf{Y}^m, \mathbf{Y}^e\big|\boldsymbol{\theta}, \hat{\boldsymbol{\varphi}}, \mathbf{B}^m, \mathbf{B}^\delta)$ with respect to $p(\mathbf{B}^m, \mathbf{B}^\delta\big|\boldsymbol{\theta}, \hat{\boldsymbol{\varphi}}, \mathbf{Y}^m, \mathbf{Y}^e)$.

Due to the nature of the effects of the calibration parameters $\boldsymbol{\theta}$ and model discrepancy function $\boldsymbol{\delta}(\mathbf{x})$, different combinations of calibration parameters and discrepancy functions might result in equally good agreement with the physical experiments and equally high values for the likelihood function. Figure 1 illustrates this with an example having a single response ($q = 1$), a scalar input x ($d = 1$), and a scalar calibration parameter $\theta (r = 1)$. $\hat{\theta}^{(1)}$, $\hat{\theta}^{(2)}$ and $\hat{\theta}^{(3)}$ denote three possible estimates of the calibration parameter θ; $y^m(x, \hat{\theta}^{(1)})$, $y^m(x, \hat{\theta}^{(2)})$ and $y^m(x, \hat{\theta}^{(3)})$ are the computer simulation models corresponding to the three different values of θ. For each simulation realization $y^m(x, \hat{\theta}^{(i)})(i = 1, ..., 3)$, an estimated discrepancy function $\delta^{(i)}(x)(i = 1, ..., 3)$ can be found so that the resulting experimental response predictions $y^m(x, \hat{\theta}^{(1)}) + \delta^{(1)}(x)$, $y^m(x, \hat{\theta}^{(2)}) + \delta^{(2)}(x)$ and $y^m(x, \hat{\theta}^{(3)}) + \delta^{(3)}(x)$ are quite similar and in equally good agreement with the experimental data within the experimental region. Intuitively from the smoothness of the observed experimental data, the combination of $\hat{\theta}^{(3)}$ and $\delta^{(3)}(x)$ seems less likely than the other two combinations to be the true parameter value and true bias function as the simulation is highly nonlinear. Rigorous calculation can show that its posterior probability density $p(\hat{\theta}^{(3)})|y^m, y^e)$ is smaller than $p(\hat{\theta}^{(1)})|y^m, y^e)$ and $p(\hat{\theta}^{(2)})|y^m, y^e)$. However, it may be virtually and computationally impossible to identify which of $\hat{\theta}^{(1)}$ and $\hat{\theta}^{(2)}$ is a better estimate of θ, since in both cases [Figs. 1(a) and 1(b)] the computer simulations are consistent with the experiments to a similar degree, and the values of $p(\hat{\theta}^{(1)})|y^m, y^e)$ and $p(\hat{\theta}^{(2)})|y^m, y^e)$ are close.

Figure 2 depicts how the posterior distribution of $\boldsymbol{\theta}$ can be used to assess the level of identifiability for a hypothetical single-parameter example. A tight posterior distribution of θ with a clear mode indicates good identifiability. In sharp contrast, with a widely dispersed posterior distribution, the identifiability is poor. Throughout, to measure

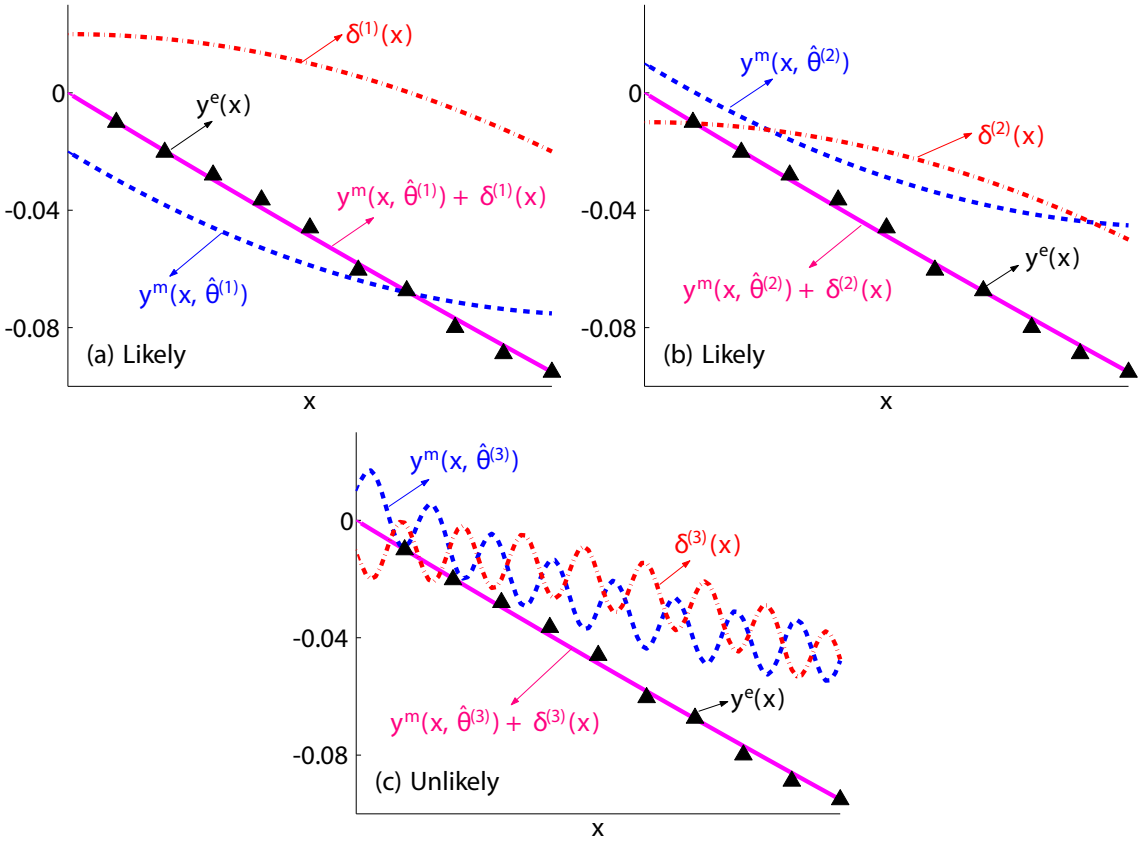

FIG. 1: An illustration of non-identifiability. Three different combinations of calibration parameter and discrepancy function result in equally good agreement with the experimental observations. While it is easy to tell that (c) is the least likely estimate of θ, it may be impossible to identify which of (a) and (b) is better.

the degree of identifiability, we will use the posterior variance (or posterior covariance matrix for cases with multiple calibration parameters) and its preposterior counterpart as its estimate prior to conducting the physical experiment.

While we do not need it in this paper, the multi-response modular Bayesian approach also includes an additional module to predict the experimental responses and the discrepancy function at untested sites $\mathbf{x}$ once the

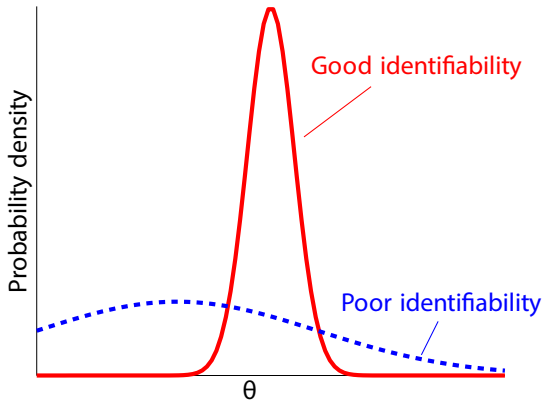

FIG. 2: Posterior distribution of calibration parameter as a demonstration of identifiability for a single-parameter case.

hyperparameters in Modules 1 and 2 are estimated. By combining the estimates of hyperparameters and the collected simulation data and experimental data, the conditional posterior distribution of the experimental responses (or the discrepancy functions if interested), given $\boldsymbol{\theta}$, can be calculated and subsequently marginalized with respect to the posterior distribution of the calibration parameters. Details can be found in [9, 11].

3. PREPOSTERIOR ANALYSIS TO SELECT THE RESPONSES TO MEASURE

In the previous section we discussed identifiability assessment using the posterior covariance calculated after observing both the simulation data and the experimental data. As discussed earlier, identifiability may possibly be enhanced by measuring multiple experimental responses. However, because each experimentally measured response incurs additional cost and effort, it is important to select an appropriate subset of responses that will reasonably enhance identifiability but that are economically feasible to measure. In this section we discuss our approach for using a preposterior analysis (Section 3.1) to *predict* (prior to conducting the experiments) the posterior covariance of the calibration parameters that will result after experimentally measuring each subset of responses from a collection of candidate subsets. The results of the preposterior analysis provides guidance on choosing which responses to measure experimentally in an economical yet effective manner. A surrogate preposterior analysis (Section 3.2) is also proposed based on the Fisher information of the calibration parameters in order to handle the computational challenge in preposterior analysis by eliminating combinations of responses that are unlikely to lead to good identifiability.

3.1 Multi-Response Preposterior Analysis

The preposterior analysis framework for selecting the experimentally measured responses is shown in Fig. 3. The procedure is based on an extension of our preposterior analysis for a single response [9, 13]. After conducting the simulations but before conducting the experiments, the user first defines a number of candidate subsets of responses for preposterior analysis. This could be accomplished based on some heuristics (e.g., one that will be discussed in Section 3.2) and/or on which responses are deemed inexpensive to measure. If there are N_c candidate subsets of responses, the analysis evaluates the degree of identifiability for each via the preposterior covariance. The subset of responses that yields the tightest preposterior distribution (cost/difficulty of measurement can also be taken into consideration) would be deemed the most likely to achieve good identifiability. Each step of the preposterior analysis is described in detail as follows.

Step 1. Preliminaries

In this step, a MRGP model is fit to the simulation data, and several quantities needed in the subsequent steps are defined. First, Module 1 of the multi-response modular Bayesian approach described in Section 2.1 is implemented to construct the MRGP model for the computer simulations based on the simulation data $\mathbf{Y}^m = [\mathbf{y}_1^m, ..., \mathbf{y}_q^m]$, where $\mathbf{y}_i^m = [y_i^m(\mathbf{x}_1^m, \boldsymbol{\theta}_1^m), ..., y_i^m(\mathbf{x}_{N_m}^m, \boldsymbol{\theta}_{N_m}^m)]^T, (i = 1, ..., q)$, collected at N_m input sites $\mathbf{X}^m = [\mathbf{x}_1^m, ..., \mathbf{x}_{N_m}^m]^T$ and $\boldsymbol{\Theta}^m = [\boldsymbol{\theta}_1^m, ..., \boldsymbol{\theta}_{N_m}^m]^T$. The maximum likelihood estimation (MLE) method estimates the hyperparameters $\boldsymbol{\phi}^m = \{\mathbf{B}^m, \boldsymbol{\Sigma}^m, \boldsymbol{\omega}^m\}$ of the computer simulations. In subsequent steps, the MRGP model is used to infer the simulation response at input settings within the design domain where no simulations are conducted.

Next, a total of N_e experimental input settings $\mathbf{X}^e = [\mathbf{x}_1^e, ..., \mathbf{x}_{N_e}^e]^T$ are defined, representing the input settings, i.e., the experimental design that one intends to use for the physical experiment (in the MC simulations of step 2, simulated experimental response values will be generated at these input settings). Step 1c then assigns the prior distribution parameters for the priors of MRGP hyperparameters in the discrepancy function, the prior for the experimental error, and the prior for the calibration parameters $\boldsymbol{\theta}$. The experimental error is assumed to be i.i.d. normal with mean 0 and a prior variance chosen based on prior knowledge of the level of measurement error. The prior for discrepancy function MRGP hyperparameters $\boldsymbol{\phi}^\delta \setminus \boldsymbol{\lambda} = \{\mathbf{B}^\delta, \boldsymbol{\Sigma}^\delta, \boldsymbol{\omega}^\delta\}$ captures the prior uncertainty and nonlinearity of the discrepancy functions; $\boldsymbol{\Sigma}^\delta$ in Eq. (5) controls the magnitude of uncertainty of the MRGP model, and $\boldsymbol{\omega}^\delta$ controls its roughness (rougher surfaces correspond to larger values of $\boldsymbol{\omega}^\delta$). For computational convenience, $\{\mathbf{B}^\delta, \boldsymbol{\Sigma}^\delta, \boldsymbol{\omega}^\delta\}$ can be assigned with point mass priors. However the prior for calibration parameters $\boldsymbol{\theta}$ should have support spanning the entire range

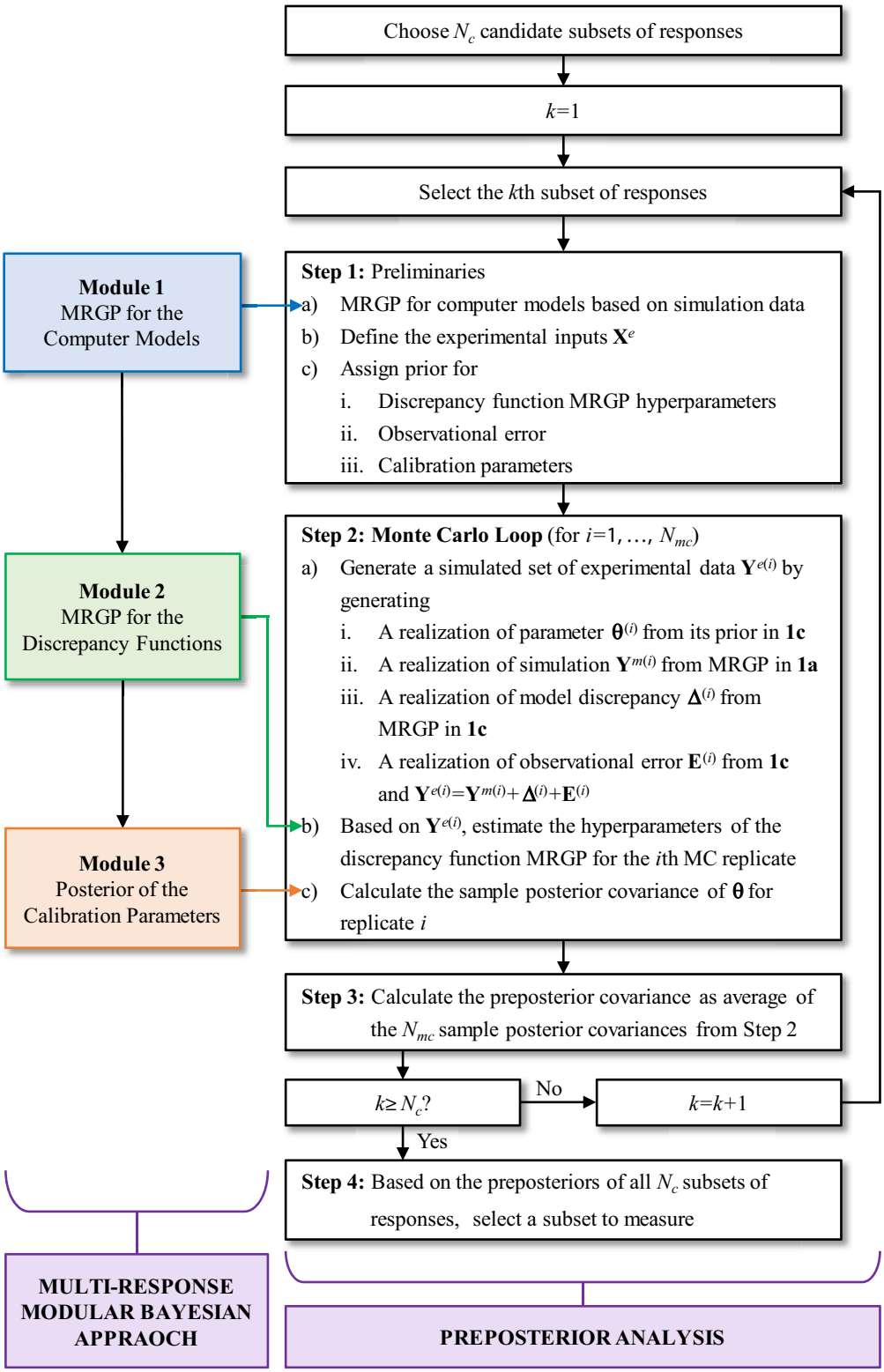

FIG. 3: Flowchart of the preposterior analysis for selecting experiment responses.

of possible values for $\boldsymbol{\theta}$; for a less informative prior, one can use a uniform distribution over a broad range, while a normal distribution with specified mean and variance can be used for a more informative prior.

Step 2. Monte Carlo (MC) loop (for $i = 1, ..., N_{mc}$)

In order to calculate the preposterior covariance, a Monte Carlo (MC) sampling strategy is applied to generate N_{mc} replicates of hypothetical experimental response data based on the information calculated or specified in step 1. For each replicate, after generating the hypothetical experimental data, the multi-response modular Bayesian approach is applied to estimate the hyperparameters of the discrepancy function MRGP models and to calculate the posterior covariance for this replicate. For the ith MC replicate, the following steps are involved. The superscript (i) added to a quantity indicates that it is for the ith MC replicate.

(a) Generate a simulated set of experimental data $\mathbf{Y}^{e(i)}$.

A realization $\boldsymbol{\theta}^{(i)}$ of the calibration parameters is first generated from its prior distribution specified in step 1. Next, using the MRGP model for the computer simulation obtained in step 1, we generate a realization of the computer simulation response values at parameter values $\boldsymbol{\theta}^{(i)}$ and input settings $\mathbf{X}^e$, denoted by $\mathbf{Y}^{m(i)} = \hat{\mathbf{y}}^m(\mathbf{X}^e, \boldsymbol{\theta}^{(i)}) = [\hat{\mathbf{y}}_1^{m(i)}, ..., \hat{\mathbf{y}}_q^{m(i)}]$, where $\hat{\mathbf{y}}_j^{m(i)} = \left[\hat{y}_j^{m(i)}(\mathbf{x}_1^e, \boldsymbol{\theta}^{(i)}), ..., \hat{y}_j^{m(i)}(\mathbf{x}_{N_e}^e, \boldsymbol{\theta}^{(i)})\right]^T (j = 1, ..., q)$. Here, $\mathbf{Y}^{m(i)}$ is generated from a multivariate normal distribution whose mean vector and covariance matrix are determined by the MRGP model. The hat notation '$\hat{}$' over $\mathbf{y}^m(\mathbf{X}^e, \boldsymbol{\theta}^{(i)})$ denotes that we are drawing interpolated data from the MRGP model (instead of the original model) of $\mathbf{y}^m$. Next, using the priors for $\boldsymbol{\phi}^\delta \setminus \boldsymbol{\lambda} = \{\mathbf{B}^\delta, \boldsymbol{\Sigma}^\delta, \boldsymbol{\omega}^\delta\}$ specified in step 1c and the MRGP model of Eq. (5), we generate a realization for the discrepancy functions at the design settings $\mathbf{X}^e$, denoted by $\boldsymbol{\Delta}^{(i)} = [\boldsymbol{\delta}_1^{(i)}, ..., \boldsymbol{\delta}_q^{(i)}]$, where $\boldsymbol{\delta}_j^{(i)} (j = 1, ..., q)$ represents the realization of discrepancy function for the ith response at $\mathbf{X}^e$. Similar to generating $\mathbf{Y}^{m(i)}$, a multivariate normal distribution is used to generate $\boldsymbol{\Delta}^{(i)}$, but with mean vector and covariance matrix determined by the MRGP model for the discrepancy functions.

Finally, we generate a realization $\mathbf{E}^{(i)}$ of the observation errors at the design settings $\mathbf{X}^e$, assuming that the experimental error $\varepsilon_j (j = 1, ..., q)$ follows i.i.d. normal distributions with mean 0 and variance λ_j specified in step 1c. $\mathbf{E}^{(i)}$ is an $N_e \times q$ matrix whose uth-row, vth-column entry is the realization of observation error for the vth response at the uth design setting $\mathbf{x}_u^e$. Based on Eq. (1), the realization $\mathbf{Y}^{e(i)}$ of the simulated experimental responses at the design settings $\mathbf{X}^e$ is calculated via

$$\mathbf{Y}^{e(i)} = \mathbf{Y}^{m(i)} + \boldsymbol{\Delta}^{(i)} + \mathbf{E}^{(i)}. \tag{9}$$

(b) Based on $\mathbf{Y}^{e(i)}$, estimate the hyperparameters of the discrepancy function MRGP $\boldsymbol{\delta}^{(i)}(\mathbf{x})$ for the ith MC replicate.

This is a direct application of Module 2 of the multi-response modular Bayesian approach described in Section 2.1, but with an actual set of experimental data $\mathbf{Y}^e$ replaced by the hypothetical set $\mathbf{Y}^{e(i)}$ generated on the ith MC replicate. Specifically, Module 2 estimates the hyperparameters of the MRGP model for the discrepancy functions for the ith MC replicate using MLE methods based on combining the simulation data $\mathbf{Y}^m$ obtained in step 1 and the hypothetical experimental data $\mathbf{Y}^{e(i)}$ generated in step 2a for the ith MC replicate.

(c) Calculate the sample posterior covariance of $\boldsymbol{\theta}$ for replicate i.

This is also a direct application of Module 3 of the multi-response modular Bayesian approach described in Section 2.2, but with an actual set of experimental data $\mathbf{Y}^e$ replaced by the hypothetical set $\mathbf{Y}^{e(i)}$ generated on the ith MC replicate. Specifically, Module 3 calculates the sample posterior covariance $\mathrm{Cov}^{(i)}[\boldsymbol{\theta}|\mathbf{Y}^m, \mathbf{Y}^{e(i)}]$ of the posterior distribution of theta in Eq. (8). For low-dimensional $\boldsymbol{\theta}$, Legendre-Gauss quadrature [26, 27] can be used to calculate the posterior covariance. For higher-dimensional $\boldsymbol{\theta}$, other numerical approaches such as Markov Chain Monte Carlo (MCMC) [28] or sampling-resampling [29] methods can be used.

Step 3. Calculate the preposterior covariance as average of the N_{mc} posterior covariances from step 2

Iterating through the N_{mc} MC replicates in step 2 produces $\{\mathrm{Cov}^{(i)}[\boldsymbol{\theta}|\mathbf{Y}^m, \mathbf{Y}^{e(i)}] : i = 1, 2, ..., N_{mc}\}$. The preposterior covariance matrix of the calibration parameters is defined as $\mathrm{E}_{\mathbf{Y}^e}[\mathrm{Cov}[\boldsymbol{\theta}|\mathbf{Y}^m, \mathbf{Y}^e]\big|\mathbf{Y}^m]$, where the outer expectation is with respect to the actual, yet-to-be-observed experimental response data $\mathbf{Y}^e$, conditioned on the already-observed simulation data $\mathbf{Y}^m$. As an estimate of the preposterior covariance matrix, we use

$$\tilde{\boldsymbol{\Sigma}}_{\boldsymbol{\theta}} \text{ (or } \tilde{\sigma}^2_{\theta} \text{ for a scalar } \theta) = \frac{1}{N_{mc}} \sum_{i=1}^{N_{mc}} \left\{ \mathrm{Cov}^{(i)}[\boldsymbol{\theta}|\mathbf{Y}^m, \mathbf{Y}^{e(i)}] \right\}. \tag{10}$$

Step 4. Based on the preposterior covariances of all subsets of responses, select a subset to measure experimentally

In Bayesian analyses, the posterior covariance constitutes a standard quantification of parameter identifiability [9–11]. The preposterior covariance serves as a prediction of the posterior covariance that we expect to result after conducting the physical experiment based on the knowledge we have obtained from the observed simulation response surface. Consequently, we use the preposterior covariance to guide the selection of the subset of responses to measure experimentally that has the most potential to enhance identifiability. For a simple case with a single calibration parameter, the subset of responses that leads to the smallest preposterior variance (tightest preposterior distribution) would be deemed as the most likely to achieve good identifiability. For cases with multiple calibration parameters, scalar metrics of the preposterior covariance matrix, such as its trace, determinant, maximum eigenvalue, etc., can be used to determine which subset yields the tightest preposterior distribution and, subsequently, which subset of responses to measure experimentally.

3.2 Fisher-Information-Based Surrogate Preposterior Analysis

Clearly, the proposed multi-response preposterior analysis is very computationally intensive. For a system with N responses, there are $N(N-1)/2$ combinations of two responses and 2^N total combinations of any subset of the N responses. Even for a single subset of responses, the computational cost can be substantial. The MC strategy requires a large number N_{mc} of replicates, and for each MC replicate, Modules 2 and 3 (which themselves involve a MC simulation or numerical integration) of the modular Bayesian approach must be implemented. Therefore, to make the preposterior analysis feasible for engineering applications with many system responses, we develop a more computationally efficient surrogate preposterior analysis that can be used to eliminate the responses that are unlikely to lead to good identifiability, thereby substantially reducing the number of response combinations that must be included in preposterior analysis.

For the r-dimensional calibration parameter vector $\boldsymbol{\theta}$, the *observed Fisher information* $\mathcal{I}(\boldsymbol{\theta})$ is a matrix whose uth-row, vth-column entry is the negative second-order derivative of the log-likelihood function:

$$[\mathcal{I}(\boldsymbol{\theta})]_{u,v} = -\frac{\partial^2}{\partial\theta_u \partial\theta_v} \log p(\mathbf{Y}^m, \mathbf{Y}^e\big|\boldsymbol{\theta}, \hat{\boldsymbol{\varphi}}) \quad (u, v = 1, 2, ..., r), \tag{11}$$

where $p(\mathbf{Y}^m, \mathbf{Y}^e\big|\boldsymbol{\theta}, \hat{\boldsymbol{\varphi}})$ is the likelihood function for the simulation and experimental data together, as in Eq. (8). It measures the amount of information that yet-to-be-collected experimental data $\mathbf{Y}^e$, together with the already observed $\mathbf{Y}^m$, carry about calibration parameter $\boldsymbol{\theta}$. We use Fisher information in our paper because most standard measures of the quality of a designed experiment are based on it (e.g., D-optimality, A-optimality, I-optimality, etc.), and we have a computationally reasonable way to calculate it. Note that $\boldsymbol{\theta}$ in Eq. (11) is meant to be the true values of the calibration parameters. Hence, in Eq. (11) the distribution of the computer response $\mathbf{Y}^m$ by itself does not depend on $\boldsymbol{\theta}$. Only the distribution of $\mathbf{Y}^e$ depends on $\boldsymbol{\theta}$.

The Fisher-like surrogate preposterior criterion that we propose in this section, which we use only for reducing the number of response combination that must be considered, is a modified version of Eq. (11). To handle the complication that we do not know the yet-to-be-collected experimental data $\mathbf{Y}^e$ in Eq. (11), we make the simple substitution $\hat{\mathbf{y}}^m(\mathbf{X}^e, \boldsymbol{\theta})$ for $\mathbf{Y}^e$. Here, $\hat{\mathbf{y}}^m(\mathbf{X}^e, \boldsymbol{\theta})$ represents the predicted value of $\mathbf{Y}^e$ via interpolating the data from the MRGP model of $\mathbf{y}^m$. After generating this fictitious realization of $\mathbf{Y}^e$, we use the modular Bayesian approach (step 2b of the flowchart in Fig. 3) to estimate the hyperparameters $\hat{\boldsymbol{\varphi}}$ for substitution into Eq. (11). Because we replace $\mathbf{Y}^e$ by

its prediction, this will tend to result in a small estimated variance parameters for the discrepancy function, which will naturally result in some level of underestimation of the identifiability in $\boldsymbol{\theta}$. However, we only use this surrogate procedure for *relative* ranking of the identifiability that results from the various combinations of responses, and our results (see Section 4) indicate that the surrogate analysis does a reasonable job of preserving the relative ranking. The advantage of this approach is computational — instead of generating multiple realizations of $\mathbf{Y}^e$, we only generate a single realization $\hat{\mathbf{y}}^m(\mathbf{X}^e, \boldsymbol{\theta})$.

There is an additional complication. The $\boldsymbol{\theta}$ in Eq. (11) represents the true parameters, and these are unknown. To handle this, we replace (11) by the "averaged" observed Fisher information matrix, averaging (11) with respect to the prior distribution $p(\boldsymbol{\theta})$ of $\boldsymbol{\theta}$. To calculate this, we use Monte Carlo simulation, as outlined in steps 2 and 3 of the flowchart in Fig. 4. Specifically, we draw N'_{mc} realizations $\boldsymbol{\theta}^{(i)} (i = 1, 2, ..., N'_{mc})$ from $p(\boldsymbol{\theta})$ and take the uth-row, vth-column entry $(u, v = 1, 2, ..., r)$ of our averaged observed Fisher information matrix to be

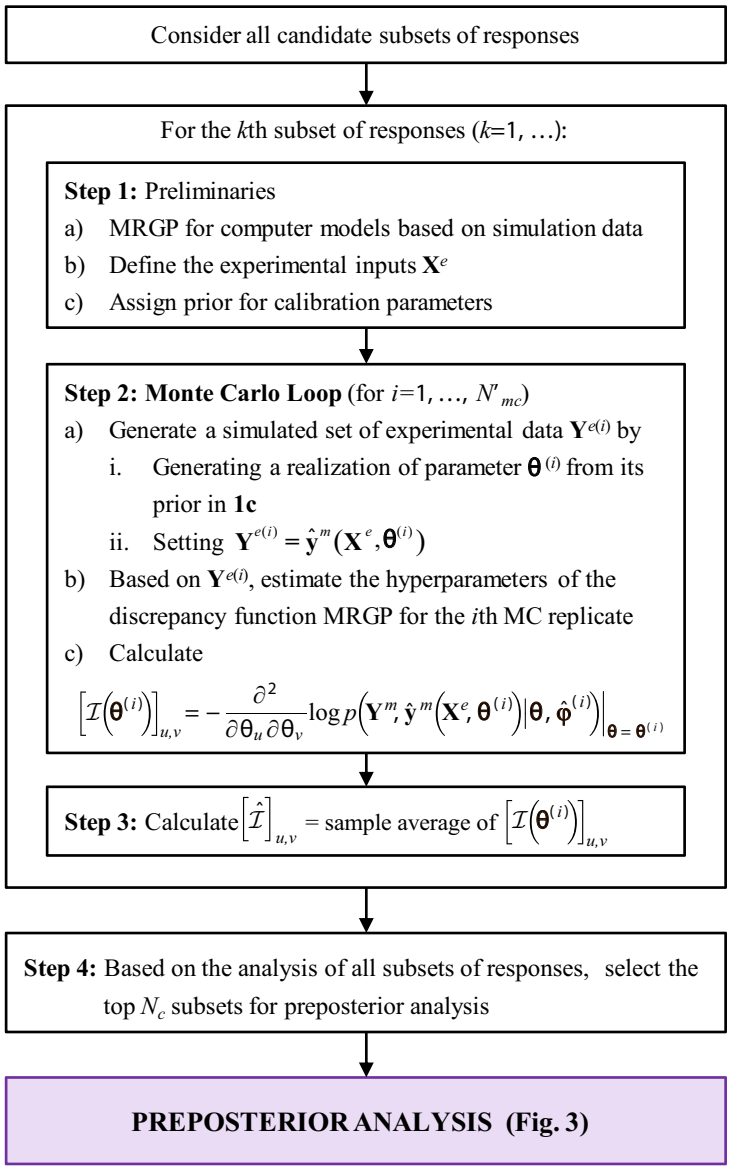

FIG. 4: Flowchart of the surrogate preposterior analysis.

$$\left[\hat{\mathcal{I}}\right]_{u,v} = \frac{1}{N'_{mc}} \sum_{i=1}^{N'_{mc}} \left\{ -\frac{\partial^2}{\partial\theta_u \partial\theta_v} \log p\left(\mathbf{Y}^m, \hat{\mathbf{y}}^m(\mathbf{X}^e, \boldsymbol{\theta}^{(i)}) \middle| \boldsymbol{\theta}, \hat{\boldsymbol{\varphi}}^{(i)}\right) \right|_{\boldsymbol{\theta}=\boldsymbol{\theta}^{(i)}} \right\}, \tag{12}$$

where $\hat{\boldsymbol{\varphi}}^{(i)}$ denotes the values of $\boldsymbol{\varphi}$ estimated in the modular Bayesian algorithm on the ith MC replicate. Various scalar metrics of $\hat{\mathcal{I}}$ (e.g., trace, determinant, maximum eigenvalue, etc.) can be used as the surrogate measure of identifiability. For a simple case with a single calibration parameter θ, $\hat{\mathcal{I}}$ is a scalar. The larger $\hat{\mathcal{I}}$ is, the more information we have about $\boldsymbol{\theta}$ from $\mathbf{Y}^m$ and $\mathbf{Y}^e$, and the more likely we are to achieve good identifiability.

After calculating $\hat{\mathcal{I}}$ from Eq. (12), we choose the top N_c subsets of responses (based on one of the scalar measures of identifiability extracted from $\hat{\mathcal{I}}$) to be further analyzed in the full preposterior analysis. Figure 4 is a flowchart of the entire Fisher-information-based surrogate preposterior procedure.

The surrogate analysis is clearly much more efficient than the full preposterior analysis. The number of MC replicates N'_{mc} can be significantly smaller than what is required in the preposterior analysis (N_{mc}), since the MC sampling now is only with respect to $\boldsymbol{\theta}$. In addition, and more importantly, the calculations in each MC replicate are much simpler. Within each MC replicate, neither an inner level MC simulation, nor numerical integration, is needed. The main computation within each MC replicate is to calculate the MLEs of the hyperparameters $\hat{\boldsymbol{\varphi}}^{(i)}$ and then to evaluate the likelihood at a few discrete values in the neighborhood of $\boldsymbol{\theta}^{(i)}$ to calculate the second-order derivative in Eq. (12) numerically.

4. CASE STUDY: A SIMPLY SUPPORTED BEAM EXAMPLE

In this section, a simply supported beam example (Fig. 5) used in [10, 11] is employed to demonstrate the effectiveness of the preposterior and surrogate preposterior analyses approach. The simply supported beam is fixed at one end and supported by a roller on the other end, with a length of 2 m and a rectangular cross section with a height of 52.5 mm and a width of 20 mm. The input variable x is the magnitude of the static force applied to the midpoint of the beam, and the calibration parameter θ is Young's modulus. The true value of Young's modulus is $\theta^* = 206.8$ GPa, which is assumed unknown to the modeler and the experimentalist. There are six output variables $y_i (i = 1, ..., 6)$ that are calculated in the computer simulation and that we consider as candidates to measure experimentally. The physical meanings of the six responses are listed in Table 1.

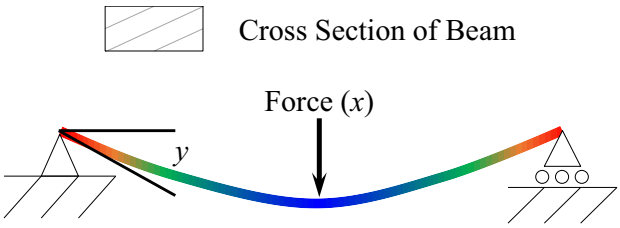

FIG. 5: Schematic of the simply supported beam.

TABLE 1: Physical meanings of the six responses that can be measured from the beam example

Response	Physical meaning
y_1	Strain at the midpoint of the beam
y_2	Plastic strain at the midpoint of the beam
y_3	Angle of deflection at the end of the beam (rad)
y_4	Internal energy of the beam (J)
y_5	Displacement at the middle of the beam (m)
y_6	Plastic dissipation energy (J)

As in [10, 11], what we treated as experimental data was from a higher-fidelity simulator for all the six responses, and the original objective is to calibrate the unknown parameter θ by integrating the experimental data and the simulation data. It turns out that the identifiability of this system is rather poor using any one of the six responses. Using two responses together may potentially enhance identifiability. Paper [11] provides detailed results of posterior analysis on each pair of the six responses and demonstrates the enhancement of identifiability. However, it did not provide a strategy for predicting which pair of responses will best enhance identifiability prior to conducting the physical experiment.

In this paper, we predict identifiability prior to collecting the physical experimental data, although the simulation data for all six responses are available on a 4×4 grid over the (x, θ) space with a range of $1300 \leqslant x \leqslant 2300$ N and $150 \leqslant \theta \leqslant 300$ GPa. Note that the design does not have to be factorial as in the example; any design of experiments (DOE) technique can be used to select the input settings. The objective is to select which two out of six responses to measure experimentally, in order to best enhance identifiability. The preposterior and surrogate preposterior analyses using the simulation data are used to predict the degree of identifiability. The candidate subsets of responses that we consider are all 15 $\{y_i, y_j : i, j = 1, ..., 6, i < j\}$. In the following discussion, we use the candidate subset $\{y_4, y_5\}$ to demonstrate the main procedures of the two analyses.

In the preliminary step of both preposterior and surrogate preposterior analyses, a MRGP model is built for $\{y_4, y_5\}$ from the simulation data. Figure 6 plots the predicted response surfaces from the MRGP model, which are quite close to the true simulation response surface, because the latter is relatively smooth in this example. The experimental settings $\mathbf{X}^e$ for the input variable are defined as 11 points uniformly spaced over the design region, i.e., $\mathbf{X}^e = [1300, 1400, ..., 2200, 2300]$ N. The prior for the calibration parameter θ is a uniform distribution over $[150, 300]$ GPa. Additionally for the preposterior analysis, we assign point mass priors for the hyperparameters of the discrepancy functions with masses at $\mathbf{B}^\delta = \mathbf{0}$, $\mathbf{\Sigma}^\delta = \text{diag}(1.11 \times 10^{-7}, 1.11 \times 10^{-7})$, and $\boldsymbol{\omega}^\delta = 2$. The experimental errors are assigned independent normal distributions: $\varepsilon_4 \sim \mathcal{N}(0, 0.0070)$ and $\varepsilon_5 \sim \mathcal{N}(0, 2.943 \times 10^{-9})$.

In the MC loop of the preposterior analysis, there are a total of $N_{mc} = 1600$ MC replicates for this specific subset of responses. Within the ith MC replicate, a realization of the simulation response $\mathbf{Y}^{m(i)}$ at the input settings $\mathbf{X}^e$ is generated from the MRGP model fitted in step 1a (the mean of which is shown in Fig. 6), a realization of the model discrepancy $\mathbf{\Delta}^{(i)}$ is generated based on the specified prior for the hyperparameters, and a realization of the observational error $\mathbf{E}^{(i)}$ is generated based on the distributions of ε_4 and ε_5. Based on this, a realization of experimental data $\mathbf{Y}^{e(i)}$ is calculated via step 2a-iv and can be used for the multi-response modular Bayesian approach to calculate the sample posterior variance for the ith MC replicate. A histogram of the 1600 sample posterior variances is plotted in Fig. 7. Assuming a uniform prior for θ over the range shown in Fig. 7, the preposterior variance is estimated as the average of all 1600 sample posterior variances, which is $\tilde{\sigma}^2_\theta = 2.0054$ GPa^2.

In the MC loop of the surrogate preposterior analysis, we consider $N'_{mc} = 16$ equally spaced values of θ. Within the ith MC replicate, a realization of the simulation $\mathbf{Y}^{m(i)}$ is generated, as in the preposterior analysis. However,

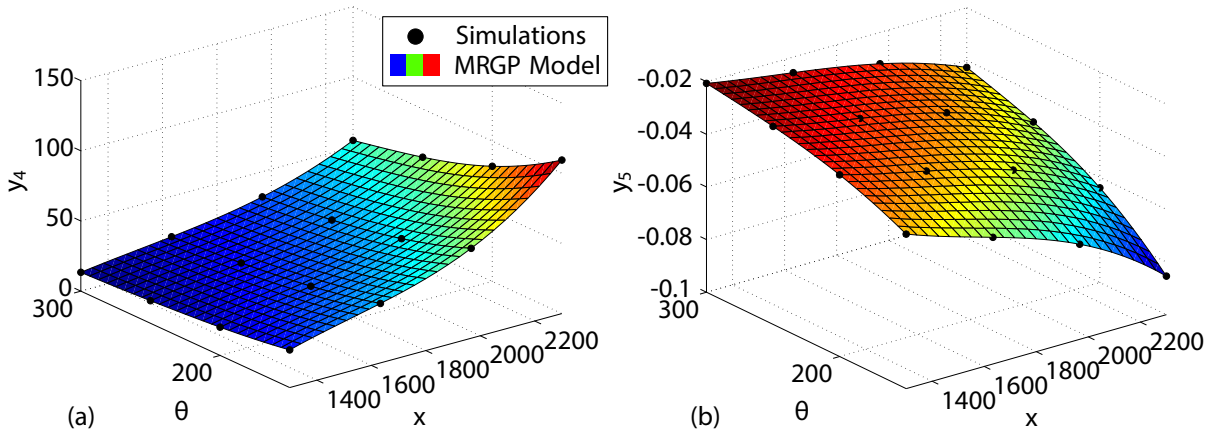

FIG. 6: The MRGP model for (a) y_4 and (b) y_5.

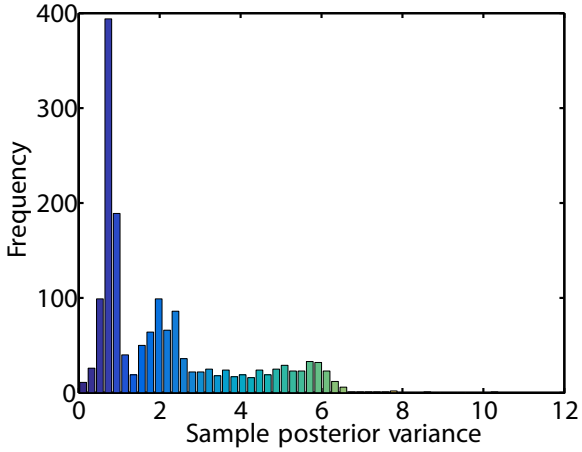

FIG. 7: Histogram of the 1600 sample posterior variances for $\{y_4, y_5\}$.

without generating $\boldsymbol{\Delta}^{(i)}$ and $\mathbf{E}^{(i)}$, we directly take $\mathbf{Y}^{e(i)} = \mathbf{Y}^{m(i)}$ and apply the multi-response modular Bayesian approach to calculate the negative second-order derivative of log-likelihood function [as the term inside the brackets in Eq. (12)]. The results for all 16 values of θ are plotted in Fig. 8(b). The Fisher-information-based scalar predictor $\hat{\mathcal{I}}$, taken to be the average of the 16 values, is 351.8. Because θ is scalar with a uniform prior for this example, taking the average over the 16 evenly spaced values of θ is more computationally efficient than generating random draws of θ from its prior and using Eq. (12).

To better illustrate the relationship between the preposterior and the surrogate preposterior analyses, we conduct the preposterior analysis in the same "fixed-θ" manner described in the preceding paragraph for the surrogate preposterior analysis. That is, we consider the 16 different values of θ equally spaced within $[150, 300]$ GPa, and for each specific θ value we use 100 MC replicates to calculate the sample posterior variance. A procedure identical to that describes in Section 3.1, but with θ fixed over the 100 MC replicates, was used to estimate the preposterior variance (as the average of the 100 sample posterior variances over the 16 values of θ). The results are shown in Fig. 8(a). Comparing Figs. 8(a) and 8(b), we see a clear negative correlation between the preposterior variance and the Fisher information criterion. This is expected, considering that the preposterior variance is an estimate of the actual posterior variance, which is closely related to the inverse of the Fisher information; a larger value of the former generally corresponds to a larger value of the latter.

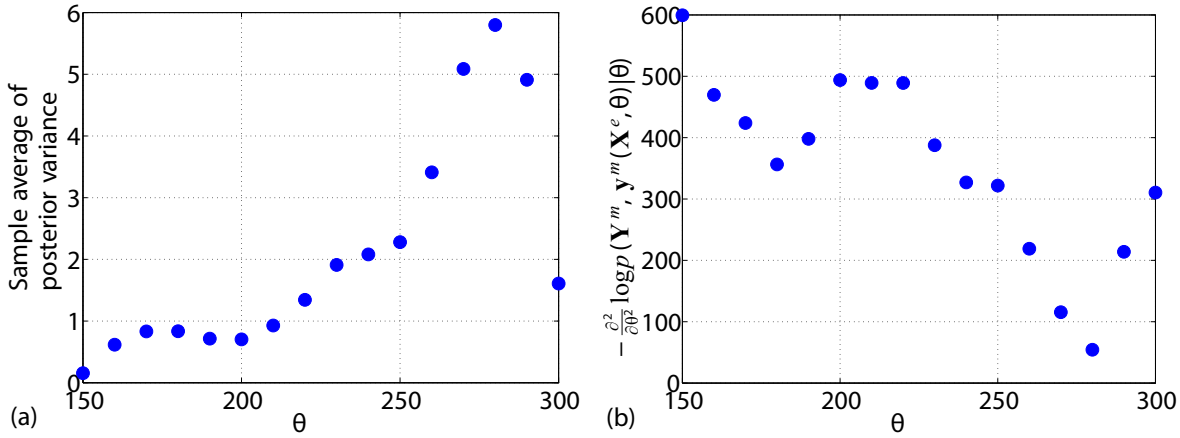

FIG. 8: 16 equally spaced values of θ, and the corresponding (a) sample average of posterior variance (unit: GPa^2) over 100 MC simulations per θ, and (b) negative second-order derivative of the log-likelihood function.

The same procedure was repeated for every other pair of responses. Table 2 shows the results of the preposterior and surrogate preposterior analyses. The ranks in columns 2 and 3 are based on the predicted identifiability; 1 corresponds to the smallest preposterior variance / largest $\hat{\mathcal{I}}$, and 15 to the largest preposterior variance / smallest $\hat{\mathcal{I}}$. The rankings provide us with predictions of which subsets are most likely to enhance identifiability (lower ranks) and which less likely (higher ranks). The "Posterior Analysis" columns in Table 2, which are taken from [11], show the actual posterior covariances that resulted from each combination of measured responses after the experiments were conducted. They serve as a basis for comparison, and ideally we would like the rankings from the posterior analyses to coincide with the predicted rankings from the preposterior and surrogate preposterior analyses. It can be seen that the preposterior analysis is in good relative agreement with the posterior. Although the values of preposterior and posterior standard deviations are off by roughly a factor of 2.5, both analyses indicate that $\{y_4, y_5\}$ together lead to the best identifiability, while $\{y_1, y_2\}$ together lead to the worst identifiability. Overall, the rankings are in very close agreement. The improvement on identifiability relative to the worst case (i.e., $\{y_1, y_2\}$ for both analyses) is also calculated and provided in the table. The relative improvements are also in very close agreement. The results provided by the surrogate preposterior analysis are also in close agreement with the posterior standard deviation results, although slightly less so than the preposterior standard deviations. The top seven pairs of responses coincide with the top seven pairs from the actual posterior analysis. Hence, the surrogate preposterior analysis would have effectively narrowed down the candidate pairs to consider in the preposterior analysis. Considering its extremely low computational cost, the surrogate preposterior analysis is a useful enhancement to the preposterior analysis for reducing the number of response pairs to consider.

The results from three analyses in Table 2 are in good accordance with the underlying physics of the system. For example, the strain y_1 and the plastic strain y_2 are perfectly correlated with each other; their values are off by a constant (equal to the value of elastic strain). Therefore, their combination adds no more information about θ and enhances identifiability little beyond using either single response. In contrast, the internal energy y_4 and the midpoint displacement y_5 follow a nonlinear relationship, and thus the degree of improvement in identifiability is substantial.

It is not surprising to observe the absolute differences between the posterior variance and preposterior variances. In the MC simulations of the preposterior analysis, the hypothetical experimental data are generated based on discrepancy functions generated from their assigned prior distribution. In contrast, the posterior variance calculation is based on the single realization that is the actual discrepancy function. Consequently, the Bayesian analysis modules inside the MC loops are hypothetical and are not expected to obtain same values of preposterior variance as the actual posterior variance. The surrogate preposterior analysis involves further approximations. However, it appears to accomplish

TABLE 2: Comparisons between posterior, preposterior, and surrogate preposterior analyses

Responses		**Posterior**			**Preposterior**			**Surrogate preposterior**	
y_i	y_j	σ_θ	Rank	Improvement	$\tilde{\sigma}_\theta$	Rank	Improvement	$\hat{\mathcal{I}}$	Rank
y_4	y_5	3.63	1	86.7%	1.4161	1	87.0%	351.8	1
y_4	y_6	3.90	2	85.8%	1.5224	3	86.0%	186.9	6
y_5	y_6	4.67	3	83.0%	1.6528	4	84.8%	218.2	3
y_3	y_6	4.86	4	82.2%	1.6618	5	84.7%	161.6	7
y_3	y_4	5.49	5	80.0%	1.4801	2	86.4%	273.0	2
y_3	y_5	6.27	6	77.1%	2.1117	6	80.6%	202.9	5
y_1	y_4	9.29	7	66.1%	3.0933	7	71.6%	209.6	4
y_1	y_3	10.13	8	63.0%	3.2583	8	70.1%	113.3	9
y_2	y_3	10.96	9	60.0%	3.6595	9	66.4%	93.34	10
y_2	y_5	15.42	10	43.4%	6.0765	11	44.2%	114.3	8
y_1	y_5	15.57	11	43.1%	5.5379	10	49.2%	90.92	11
y_2	y_4	18.35	12	33.9%	8.7816	12	19.4%	65.94	14
y_1	y_6	24.08	13	12.0%	9.1008	13	16.5%	81.19	12
y_2	y_6	26.46	14	3.3%	10.1850	14	6.5%	80.99	13
y_1	y_2	27.37	15	—	10.8931	15	—	47.09	15

its intended purpose, as the preposterior and surrogate preposterior analyses do a reasonable job of predicting the relative degree of identifiability and guide users in selecting the best set of responses to measure experimentally. Fig. 9 illustrates this by plotting the preposterior standard deviation and the Fisher information criterion $\hat{\mathcal{I}}$ versus the actual posterior standard deviation. The preposterior standard deviation is roughly in proportion to the actual posterior standard deviation, and $\hat{\mathcal{I}}$ is negatively correlated with the posterior standard deviation, which indicates that for this case study preposterior and surrogate preposterior analyses are sufficient in predicting identifiability.

We further illustrate the improvement on identifiability in Figs. 10 and 11. While neither analyzing y_4 nor analyzing y_5 alone can provide an informative posterior distribution of θ (Fig. 10), uncertainty quantification considering both y_4 and y_5 provides a much tighter posterior distribution of θ, and the mean of the posterior is close to the true value of θ [Fig. 11(a)]. In contrast, the subset of $\{y_1, y_2\}$ provides a dispersed posterior distribution of θ [Fig. 11(b)]. The level of uncertainty is well predicted by both the preposterior and surrogate preposterior analyses.

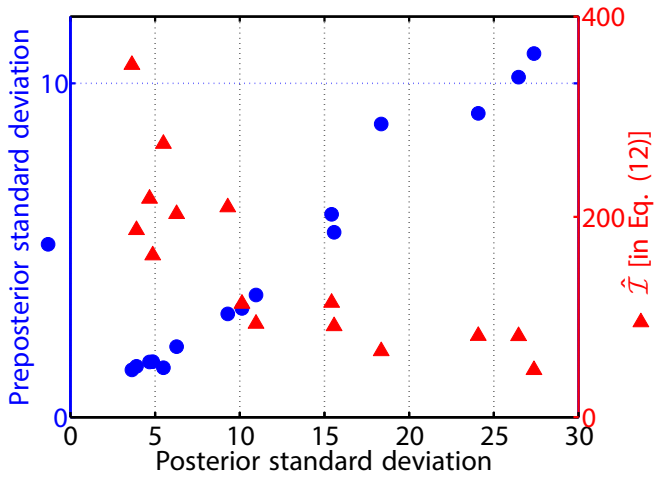

FIG. 9: Preposterior standard deviation and Fisher-information-based identifiability predictor, versus posterior standard deviation, demonstrating very high correlation.

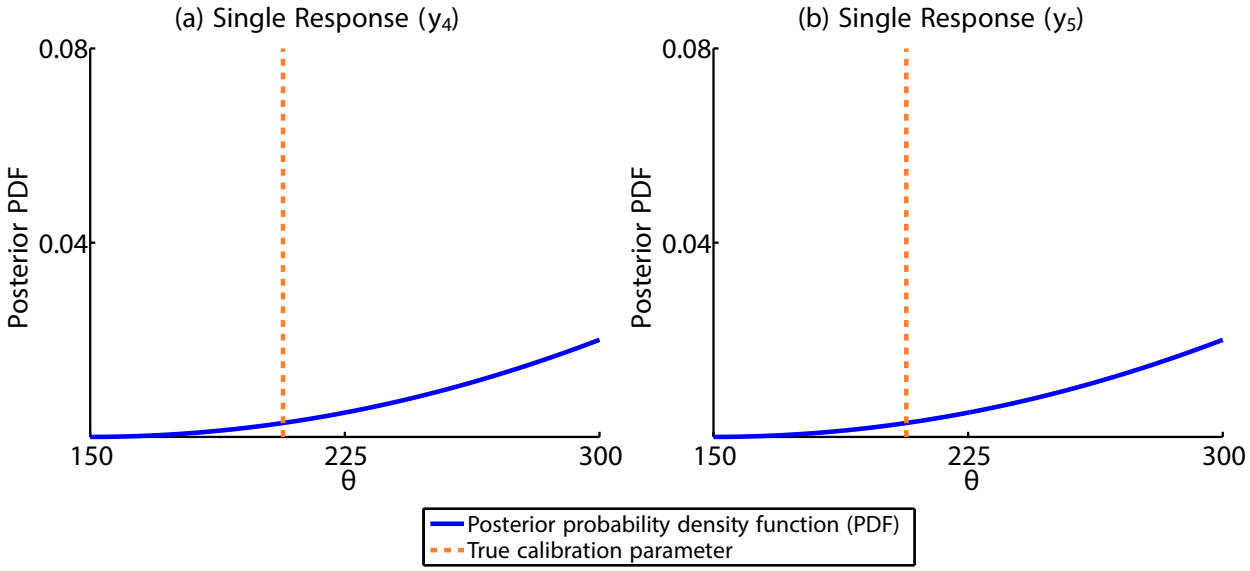

FIG. 10: Posterior distribution of the calibration parameter using a single response.

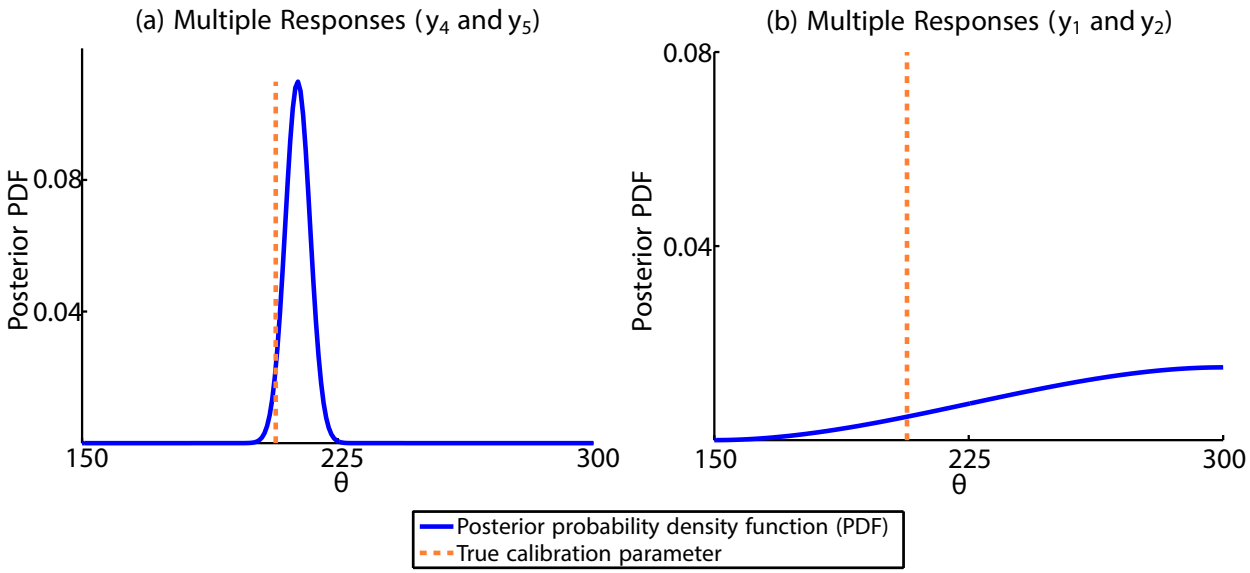

FIG. 11: Posterior distribution of the calibration parameter using multiple responses.

5. CONCLUSIONS

Identifiability is of major importance in model calibration and predictive modeling in all engineering disciplines. The degree of identifiability can be measured by the posterior covariance of the calibration parameters in a typical model uncertainty quantification framework. Earlier studies have demonstrated that identifiability can be enhanced by measuring multiple responses that share a mutual dependence on a common set of calibration parameters. However, to take advantage of this, a method is needed for predicting multi-response identifiability prior to conducting the physical experiments, to allow users to choose the most appropriate set of responses to measure experimentally. In this research, we propose a preposterior analysis that, prior to conducting the physical experiments but after conducting computer simulations, can predict the degree of identifiability that will result using different subsets of responses to measure experimentally. This is accomplished by calculating the preposterior covariance from a modular Bayesian Monte Carlo analysis of a MRGP model. To render the approach computationally feasible in engineering applications with a large number of responses, we also proposed a surrogate preposterior analysis based on the Fisher information of the calibration parameters, which is used to eliminate combinations of responses that are unlikely to provide substantial improvement in identifiability, thereby substantially reducing computational cost. The proposed methods were applied to a simply supported beam example to select which two out of six responses will best improve identifiability. Our study shows that the approach is effective in predicting which subset of responses will provide the largest improvement in identifiability. Even though there are absolute differences between the preposterior and actual posterior covariances, the relative differences and the rankings derived from them are quite consistent, indicating that the method can be used effectively to choose the best combination of responses to measure experimentally.

Future work in this direction includes examining the impact of using different priors for discrepancy function MRGP hyperparameters on the preposterior covariance. Also, the model input settings for physical experiments apparently affect identifiability; simultaneously optimizing the selected experimental responses and the design for the experimental input settings is another research direction.

ACKNOWLEDGMENTS

The grant support from the National Science Foundation (CMMI-1233403) and the Terminal Year Fellowship at Northwestern University are greatly acknowledged. Zhen Jiang would like to thank Dr. Paul D. Arendt for valuable discussions. This paper is revised based on DETC2013-12457 in the proceedings of the *ASME 2013 International*

Design Engineering Technical Conferences & Computers and Information in Engineering Conference. We are grateful for ASME to grant us the permission to publish the content of DETC2013-12457 in this paper with *International Journal for Uncertainty Quantification*.

REFERENCES

1. Kennedy, M. C. and O'Hagan, A., Bayesian calibration of computer models, *J. R. Stat. Soc. Ser. B (Statistical Methodology)*, 63(3):425–464, 2001.
2. Higdon, D., Kennedy, M., Cavendish, J. C., Cafeo, J. A., and Ryne, R. D., Combining field data and computer simulations for calibration and prediction, *SIAM J. Sci. Comput.*, 26(2):448–466, 2004.
3. Reese, C. S., Wilson, A. G., Hamada, M., Martz, H. F., and Ryan, K. J., Integrated analysis of computer and physical experiments, *Technometrics*, 46(2):153–164, 2004.
4. Bayarri, M. J., Berger, J. O., Paulo, R., Sacks, J., Cafeo, J. A., Cavendish, J., Lin, C.-H., and Tu, J., A framework for validation of computer models, *Technometrics*, 49(2):138–154, 2007.
5. Higdon, D., Gattiker, J., Williams, B., and Rightley, M., Computer model calibration using high-dimensional output, *J. Am. Stat. Assoc.*, 103(482):570–583, 2008.
6. Chen, W., Xiong, Y., Tsui, K.-L., and Wang, S., A design-driven validation approach using bayesian prediction models, *J. Mech. Design*, 130(2):021101, 2008.
7. Qian, P. Z. and Wu, C. J., Bayesian hierarchical modeling for integrating low-accuracy and high-accuracy experiments, *Technometrics*, 50(2):192–204, 2008.
8. Wang, S., Chen, W., and Tsui, K.-L., Bayesian validation of computer models, *Technometrics*, 51(4):439–451, 2009.
9. Arendt, P. D., *Quantification and Mitigation of Multiple Sources of Uncertainty in Simulation Based Design*, PhD Thesis, Northwestern University, Evanston, IL, 2012.
10. Arendt, P. D., Apley, D. W., and Chen, W., Quantification of model uncertainty: Calibration, model discrepancy, and identifiability, *J. Mech. Design*, 134(10):100908, 2012.
11. Arendt, P. D., Apley, D. W., Chen, W., Lamb, D., and Gorsich, D., Improving identifiability in model calibration using multiple responses, *J. Mech. Design*, 134(10):100909, 2012.
12. Loeppky, J., Bingham, D., and Welch, W., Computer model calibration or tuning in practice, Tech. Rep., University of British Columbia, 2006.
13. Arendt, P. D., Apley, D. W., and Chen, W., A preposterior analysis to predict identifiability in experimental calibration of computer models, *IIE Transactions*, accepted, DOI: 10.1080/0740817X.2015.1064554, 2015.
14. Conti, S., Gosling, J. P., Oakley, J., and O'Hagan, A., Gaussian process emulation of dynamic computer codes, *Biometrika*, 96(3):663–676, 2009.
15. Conti, S. and O'Hagan, A., Bayesian emulation of complex multi-output and dynamic computer models, *J. Stat. Planning Inference*, 140(3):640–651, 2010.
16. Liu, F., Bayarri, M., Berger, J., Paulo, R., and Sacks, J., A Bayesian analysis of the thermal challenge problem, *Comput. Methods Appl. Mech. Eng.*, 197(29):2457–2466, 2008.
17. McFarland, J., Mahadevan, S., Romero, V., and Swiler, L., Calibration and uncertainty analysis for computer simulations with multivariate output, *AIAA J.*, 46(5):1253–1265, 2008.
18. Cressie, N., *Statistics for Spatial Data: Wiley Series in Probability and Statistics*, Wiley-Interscience, New York, 1993.
19. Currin, C., Mitchell, T., Morris, M., and Ylvisaker, D., Bayesian prediction of deterministic functions, with applications to the design and analysis of computer experiments, *J. Am. Stat. Assoc.*, 86(416):953–963, 1991.
20. Handcock, M. S. and Stein, M. L., A bayesian analysis of Kriging, *Technometrics*, 35(4):403–410, 1993.
21. Santner, T. J., Williams, B. J., and Notz, W., *The Design and Analysis of Computer Experiments*, Springer, Berlin, 2003.
22. Johnson, R. T., Montgomery, D. C., Jones, B., and Parker, P. A., Comparing computer experiments for fitting high-order polynomial metamodels, *J. Quality Technol.*, 42(1):86–102, 2010.

23. Barton, R. R., Metamodeling: A state of the art review, in *Proc. of the 26th Conf. Winter Simulation*, Society for Computer Simulation International, pp. 237–244, 1994.
24. Rasmussen, C. E. and Williams, C. K. I., *Gaussian Processes for Machine Learning*, The MIT Press, Cambridge, MA, 2006.
25. Sacks, J., Welch, W. J., Mitchell, T. J., and Wynn, H. P., Design and Analysis of Computer Experiments, *Stat. Sci.*, 4(4):409–423, 1989.
26. Abramowitz, M. and Stegun, I. A., *Handbook of Mathematical Functions*, Vol. 1, Dover New York, 1972.
27. Beyer, W. H., *CRC Standard Mathematical Tables*, 29th ed., Vol. 1, CRC Press, Boca Raton, FL, 1991.
28. Robert, C. P. and Casella, G., *Monte Carlo Statistical Methods*, Springer-Verlag, New York, 1999.
29. Smith, A. F. and Gelfand, A. E., Bayesian statistics without tears: A sampling–resampling perspective, *The Am. Stat.*, 46(2):84–88, 1992.