Improving Identifiability in Model Calibration Using Multiple Responses

Paul D. Arendt
Northwestern University, Department of Mechanical Engineering
2145 Sheridan Road Room B214
Evanston, IL, 60208
Phone: 847-491-5066, Fax: 847-491-3915, Email: paularendt2012@u.northwestern.edu

Daniel W. Apley
Northwestern University, Department of Industrial Engineering and Management Sciences
2145 Sheridan Road Room C150
Evanston, IL, 60208
Phone: 847-491-2397, Fax: 847-491-8005, Email: apley@northwestern.edu

Wei Chen*
***Corresponding Author
Northwestern University, Department of Mechanical Engineering
2145 Sheridan Road Room A216
Evanston, IL, 60208
Phone: 847-491-7019, Fax: 847-491-3915, Email: weichen@northwestern.edu

David Lamb
U.S. Army Tank-Automotive Research Development & Engineering Center
6501 E. Eleven Mile Road
Warren, MI 48397

David Gorsich
U.S. Army Tank-Automotive Research Development & Engineering Center
6501 E. Eleven Mile Road
Warren, MI 48397

Abstract

In physics-based engineering modeling, the two primary sources of model uncertainty, which account for the differences between computer models and physical experiments, are parameter uncertainty and model discrepancy. Distinguishing the effects of the two sources of uncertainty can be challenging. For situations in which identifiability cannot be achieved using only a single response, we propose to improve identifiability by using multiple responses that share a mutual dependence on a common set of calibration parameters. To that end, we extend the single response modular Bayesian approach for calculating posterior distributions of the calibration parameters and the discrepancy function to multiple responses. Using an engineering example, we demonstrate that including multiple responses can improve identifiability (as measured by posterior standard deviations) by an amount that ranges from minimal to substantial, depending on the characteristics of the specific responses that are combined.

Keywords: multiple responses, Gaussian process, model updating, calibration, identifiability, uncertainty quantification, multi-output emulator

1 Introduction

The quantification of model uncertainty is important to better understand how a computer model represents physical reality. Two primary sources of uncertainty that account for differences between a computer model and physical reality are parameter uncertainty and model discrepancy [1]. Parameter uncertainty results from unknown calibration parameters of the computer model, whereas model discrepancy results from missing physics and other inaccuracies of the computer model. Identifiability of the calibration parameters and the discrepancy function (which represents the effects of model discrepancy) is important because it allows one to (1) learn the underlying values of the calibration parameters, (2) better understand the deficiencies of

the computer model, and (3) more accurately predict the experimental response. However, as demonstrated in our companion paper [2], distinguishing these two sources of uncertainty based on using only a single response is usually difficult (albeit possible in some situations).

In engineering systems where there is a lack of identifiability with a single response, we propose in this paper the use of multiple responses as additional information to improve identifiability. In this work, multiple responses are considered to be a collection of single responses of different quantities (e.g., load force, displacement, stress, etc.), each of which can also be measured over time and/or space. When the multiple responses are mutually dependent on the same set of calibration parameters, the information obtained from them can be combined to better infer the true value of the calibration parameters. We note that multiple responses usually can be "measured" for free in computer simulations, as they are automatically calculated in the simulation. Hence, the added cost of observing multiple responses is that associated with their measurement in the physical experiments.

To quantify the uncertainty that results from having to interpolate the responses between discrete input sites at which they are observed in expensive computer simulations or physical experiments, multiple response surrogate models are necessary. One existing body of work uses a multiple response Gaussian process (MRGP) model, which is defined by its mean and covariance functions, as a surrogate model representing only the computer model. Most existing work uses the same mathematical structure for the mean function. However, for the covariance function two different mathematical structures are common. References [3-6] assign an indexing variable to each response, which is treated as an additional input over which the covariance function is defined. This is only applicable if the multiple responses are inherently ordered in a manner that permits such an indexing.

More recently, Conti et al. [7-8] defined a covariance function that is the combination of a spatial correlation function and a discrete covariance matrix, the latter representing the covariance between responses. This covariance parameterization is relatively simple yet flexible, and the resulting MRGP model is a straightforward extension of the well established single response Gaussian process model [8]. Therefore, we adopt this covariance function to construct the MRGP model for both the computer model and the discrepancy functions.

Although using multiple responses to quantify computer model uncertainty based on experimental data has been considered in previous literature, the “multiple responses” were restricted to a single response measured at multiple spatial and/or temporal locations [9-12]. Furthermore, the prior work focuses solely on predicting the experimental response and not on identifiability of the calibration parameters and the discrepancy functions. As demonstrated in our companion paper [2], different combinations of the calibration parameters and the discrepancy functions can often result in the same experimental response prediction, which is the root of the identifiability problem.

In our companion paper [2], we shed light on the challenging problem of identifying the calibration parameters and the discrepancy function when combining single-response data from a computer model and physical experiments. The main objective of this paper is to show that identifiability can be improved by including multiple responses in the model updating formulation. We first extend the single response modular Bayesian approach [1] to multiple responses in Section 2, which we use to calculate the posterior distributions of the calibration parameters and the discrepancy functions. In Section 3, we apply the multiple response modular Bayesian approach to the same simply supported beam example introduced in [2]. Whereas identifiability was not achievable using only a single response in [2], we show here that

identifiability can be substantially enhanced using certain combinations of multiple responses. Section 4 concludes the paper.

2 Multiple Response Modular Bayesian Approach

In [2], we reviewed the single response modular Bayesian approach to calculate the posterior distributions, which quantify the posterior uncertainty of the calibration parameters and the discrepancy function, by integrating data from both simulations and experiments. In this section, we extend the modular Bayesian approach to incorporate multiple responses. To do this, we reformulate the model updating formulation of Kennedy and O'Hagan [1] as

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

for $i = 1, \ldots, q$, where q denotes the number of response variables. The vector $\mathbf{x} = [x_1, \ldots, x_d]^T$ denotes the design variables (aka controllable inputs) and $\boldsymbol{\theta} = [\theta_1, \ldots, \theta_r]^T$ denotes a vector of calibration parameters with $\boldsymbol{\theta}^*$ denoting their true physical values, which are unknown to the user. $y_i^m(\mathbf{x},\boldsymbol{\theta})$ ($i = 1, \ldots, q$) denotes the ith response from the computer model as a function of $\mathbf{x}$ and $\boldsymbol{\theta}$, and $y_i^e(\mathbf{x})$ denotes the corresponding experimentally observed response.

Note that all of the responses from the computer model mutually depend on the same set of calibration parameters $\boldsymbol{\theta}$, and the experimental responses depend on the same true calibration parameter value $\boldsymbol{\theta}^*$. $\delta_i(\mathbf{x})$ denotes the ith discrepancy function, which represents the difference between the computer model (with input $\boldsymbol{\theta}^*$) and the physical experiments. Finally, ε_i is independent random experimental error (e.g., observation error), which is assumed normal with mean 0 and variance λ_i. A detailed discussion of the terms in Eq. (1) is provided in our companion paper [2].

Similar to the procedure for implementing the single response modular Bayesian approach (see Fig. 5 of [2]), the multiple response modular Bayesian approach fits a MRGP

model to the computer model and the discrepancy functions in two separate modules (Modules 1 and 2). These two MRGP models account for the interpolation uncertainty, which is a result of having to interpolate between the observations of the finite set of simulation and experimental data [1-2, 13]. After fitting each MRGP model, one calculates the posterior distributions of the calibration parameters, the discrepancy functions, and the experimental responses (Modules 3 and 4). Details of each module in the multiple response modular Bayesian approach are provided in the following subsections.

2.1 Multiple Response Gaussian Process Model for the Computer Model (Module 1)

In Module 1, a MRGP model is fitted to the computer responses and used to infer their values at input sites other than those that were simulated. Following Conti et al. [7-8], the prior for the MRGP model is

$$\mathbf{y}^m(\bullet,\bullet) \sim GP\left(\mathbf{h}^m(\bullet,\bullet)\mathbf{B}^m, \boldsymbol{\Sigma}^m R^m\left((\bullet,\bullet),(\bullet,\bullet)\right)\right), \tag{2}$$

where $\mathbf{y}^m(\mathrm{x},\boldsymbol{\theta}) = [y_1{}^m(\mathbf{x},\boldsymbol{\theta}), \ldots, y_q{}^m(\mathbf{x},\boldsymbol{\theta})]$ denotes the multiple responses from the computer model. The prior mean function is comprised of an arbitrary vector of specified regression functions $\mathbf{h}^m(\mathbf{x},\boldsymbol{\theta}) = [h_1^m(\mathbf{x},\boldsymbol{\theta}),\ldots,h_p^m(\mathbf{x},\boldsymbol{\theta})]$ and a matrix of unknown regression coefficients $\mathbf{B}^m = [\boldsymbol{\beta}_1^m, \ldots, \boldsymbol{\beta}_q^m]$, where $\boldsymbol{\beta}_i^m = [\beta_{1,i}^m, \ldots, \beta_{p,i}^m]^T$. For the examples in this paper, we use $\mathbf{h}^m(\mathbf{x},\boldsymbol{\theta}) = 1$, which corresponds to a constant prior mean. The prior covariance function is the product of an unknown non-spatial $q \times q$ covariance matrix $\boldsymbol{\Sigma}^m$ and a spatial correlation function $R^m((\mathbf{x},\boldsymbol{\theta}),(\mathbf{x}',\boldsymbol{\theta}'))$, where $(\mathbf{x},\boldsymbol{\theta})$ and $(\mathbf{x}',\boldsymbol{\theta}')$ denote two sets of computer model inputs. This prior covariance function can be rewritten as $Cov[y_i^m(\mathbf{x},\boldsymbol{\theta}), y_j^m(\mathbf{x}',\boldsymbol{\theta}')] = \boldsymbol{\Sigma}_{ij}^m R^m((\mathbf{x},\boldsymbol{\theta}),(\mathbf{x}',\boldsymbol{\theta}'))$ where $\boldsymbol{\Sigma}_{ij}^m$ is the covariance between the *i*th and *j*th computer model responses at the same input

values. Further, this covariance structure reduces to the covariance function for a single response Gaussian process model when $q = 1$. We use a Gaussian correlation function

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

parameterized by the $(d + r) \times 1$ vector of roughness parameters $\boldsymbol{\omega}^m$, which represent the rate at which the correlation between $y_i^m(\mathbf{x},\boldsymbol{\theta})$ and $y_i^m(\mathbf{x}',\boldsymbol{\theta}')$ decays to zero as $(\mathbf{x},\boldsymbol{\theta})$ and $(\mathbf{x}',\boldsymbol{\theta}')$ diverge. Lower values of $\boldsymbol{\omega}^m$ indicate a smoother MRGP model for the response $y_i^m(\mathbf{x},\boldsymbol{\theta})$.

Reference [7] provides some justification, beyond the relatively simple parameterization that results in tractable computations, for using the covariance structure in Eq. (2). In particular, if multiple responses mutually depend on the same set of inputs (both design variables and calibration parameters), then one might expect the responses to be correlated and to share a similar spatial correlation structure or smoothness. The formulation of Eq. (2) assumes such a shared spatial correlation structure. On the other hand, if the responses are weakly correlated with different spatial correlations, then one can model each response independently, allowing each to have a different spatial correlation function [7].

Fig. 1 illustrates the separation of the covariance function into spatial and non-spatial portions for the case of two responses. The gray arrow below the x-axis of Fig. 1 signifies the spatial correlation portion of the covariance function, which is represented by $R^m((x,\theta),(x',\theta))$ for computer model inputs x and x' with fixed θ. Alternatively, the non-spatial covariance matrix portion of the covariance function, $\boldsymbol{\Sigma}^m$, represents the covariance across response functions, as signified in Fig. 1 by the gray arrow between the responses. Whereas the spatial correlation function dictates the smoothness of both responses, the non-spatial covariance matrix dictates the similarity between any trends or fluctuations (not represented by the prior mean function) in the two responses.

To obtain maximum likelihood estimates (MLEs) of the hyperparameters $\boldsymbol{\phi}^m = \{\mathbf{B}^m, \boldsymbol{\Sigma}^m, \boldsymbol{\omega}^m\}$ for the computer model MRGP model, we maximize the multivariate normal log-likelihood function based on the simulation data $\mathbf{Y}^m = [\mathbf{y}_1^m, \ldots, \mathbf{y}_q^m]$, where $\mathbf{y}_i^m = [y_i^m(\mathbf{x}_1^m, \boldsymbol{\theta}_1^m), \ldots, y_i^m(\mathbf{x}_{N_m}^m, \boldsymbol{\theta}_{N_m}^m)]^T$, collected at N_m input sites $\mathbf{X}^m = [\mathbf{x}_1^m, \ldots, \mathbf{x}_{N_m}^m]^T$ and $\boldsymbol{\Theta}^m = [\boldsymbol{\theta}_1^m, \ldots, \boldsymbol{\theta}_{N_m}^m]^T$. Previous literature [14-16] suggest a space-filling experimental design for the input settings of $\mathbf{X}^m$ and $\boldsymbol{\Theta}^m$. Furthermore, to improve the numerical stability of the MLE algorithm, the inputs $\mathbf{X}^m$ and $\boldsymbol{\Theta}^m$ can be transformed to the range of 0 to 1, and the simulation data $\mathbf{y}_i^m$ can be standardized to have a sample mean of 0 and a sample standard deviation of 1 [17]. Details of the MLE algorithm can be found in Appendix A.

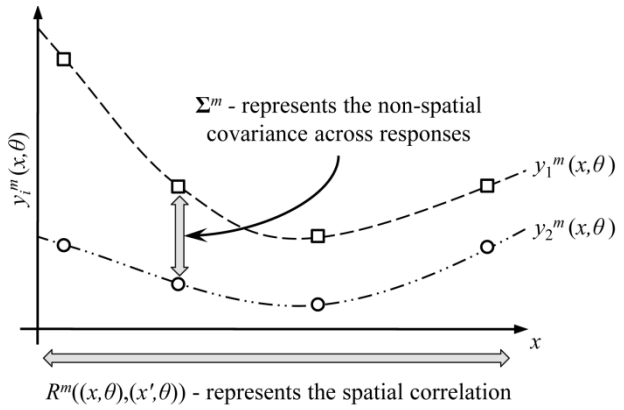

Fig. 1 Schematic showing the MRGP model's covariance function represented by spatial correlation and non-spatial covariance for a fixed θ.

Inference of the computer model response $y_i^m(\mathbf{x},\boldsymbol{\theta})$ at any $\mathbf{x}$ and $\boldsymbol{\theta}$, can be determined by inserting the MLEs of the hyperparameters into the MRGP model posterior mean and covariance equations, Eqs. (B1) and (B2) of Appendix B. After obtaining the MLEs of the hyperparameters, the next step is to estimate the hyperparameters of the MRGP model representing the discrepancy functions, as follows.

2.2 Multiple Response Gaussian Process Model for the Discrepancy Functions (Module 2)

The MRGP model of the experimental response is the sum of two MRGP models representing both the computer model $y_i^m(\mathbf{x},\boldsymbol{\theta})$ and the discrepancy functions $\delta_i(\mathbf{x})$. This MRGP model will then infer the experimental response $y_i^e(\mathbf{x})$ at any point $\mathbf{x}$.

To define the MRGP model for the experimental response, we first need to define the MRGP model for the discrepancy functions, which is the focus of Module 2. The prior of the MRGP model for the discrepancy functions is

$$\boldsymbol{\delta}(\bullet) \sim GP\left(\mathbf{h}^{\delta}(\bullet)\mathbf{B}^{\delta}, \boldsymbol{\Sigma}^{\delta} R^{\delta}(\bullet,\bullet)\right), \tag{4}$$

where $\boldsymbol{\delta}(\mathbf{x}) = [\delta_1(\mathbf{x}), \ldots, \delta_q(\mathbf{x})]$ denotes the multiple discrepancy functions. The mean function is comprised of a vector of specified regression functions $\mathbf{h}^{\delta}(\mathbf{x}) = [h_1^{\delta}(\mathbf{x}), \ldots, h_s^{\delta}(\mathbf{x})]^T$ (we use the constant $\mathbf{h}_i^{\delta}(\mathbf{x}) = 1$) and a matrix of unknown regression coefficients $\mathbf{B}^{\delta} = [\boldsymbol{\beta}_1^{\delta}, \ldots, \boldsymbol{\beta}_q^{\delta}]$, where $\boldsymbol{\beta}_i^{\delta} = [\beta_{1,i}^{\delta}, \ldots, \beta_{s,i}^{\delta}]^T$. The prior covariance function of this MRGP model is the product of an unknown non-spatial covariance matrix $\boldsymbol{\Sigma}^{\delta}$ and a spatial correlation function $R^{\delta}(\mathbf{x},\mathbf{x}')$, which is a Gaussian correlation function (similar to Eq. (3) but without the $\boldsymbol{\theta}$ inputs) parameterized by the $d \times 1$ vector of roughness parameters $\boldsymbol{\omega}^{\delta}$. This prior covariance function can be rewritten as $Cov[\delta_i(\mathbf{x}), \delta_j(\mathbf{x}')] = \boldsymbol{\Sigma}_{ij}^{\delta} R^{\delta}(\mathbf{x}, \mathbf{x}')$, where $\boldsymbol{\Sigma}_{ij}^{\delta}$ is the covariance between the ith and jth discrepancy functions at the same input values. Following the work of Kennedy and O'Hagan [1], we assume the computer model, the discrepancy functions, and the experimental uncertainty are a priori statistically independent (i.e. $Cov[y_i^m(\mathbf{x},\boldsymbol{\theta}), \delta_j(\mathbf{x}')] = 0$, $Cov[y_i^m(\mathbf{x},\boldsymbol{\theta}), \varepsilon_j] = 0$, and $Cov[\delta_i(\mathbf{x}), \varepsilon_j] = 0$ for $i,j \in \{1, \ldots, q\}$ and all $\mathbf{x}$, $\mathbf{x}'$, and $\boldsymbol{\theta}$). These assumptions of a priori independence simplify many of the calculations required in the multiple response modular

Bayesian approach. We note that the posterior distributions for these quantities are generally highly correlated, even though their prior distributions are independent.

The priors for the computer model MRGP model [Eq. (2)] and the discrepancy functions MRGP model [Eq. (4)] are combined via Eq. (1) to form the prior for the experimental responses MRGP model:

$$\begin{aligned} &\mathbf{y}^e(\bullet)\,|\,\boldsymbol{\theta}=\boldsymbol{\theta}^* \sim GP(m^e(\bullet,\boldsymbol{\theta}^*), V^e((\bullet,\boldsymbol{\theta}^*),(\bullet,\boldsymbol{\theta}^*)) \\ &m^e(\mathbf{x},\boldsymbol{\theta}) = \mathbf{h}^m(\mathbf{x},\boldsymbol{\theta})\mathbf{B}^m + \mathbf{h}^\delta(\mathbf{x})\mathbf{B}^\delta \\ &V^e((\mathbf{x},\boldsymbol{\theta}),(\mathbf{x}',\boldsymbol{\theta})) = \boldsymbol{\Sigma}^m R^m((\mathbf{x},\boldsymbol{\theta}),(\mathbf{x}',\boldsymbol{\theta})) + \boldsymbol{\Sigma}^\delta R^\delta(\mathbf{x},\mathbf{x}') + \boldsymbol{\lambda} \end{aligned} \quad , \qquad (5)$$

where $\mathbf{y}^e(\mathbf{x}) = [y_1^e(\mathbf{x}), \ldots, y_q^e(\mathbf{x})]$ denote the multiple experimental responses. The $q \times q$ diagonal covariance matrix $\boldsymbol{\lambda}$ with diagonal entries $\lambda_1, \ldots, \lambda_q$ represents the experimental uncertainty, i.e. $\varepsilon_i \sim N(0, \lambda_i)$. Since Eq. (5) depends on the true calibration parameters $\boldsymbol{\theta}^*$, the objective is now to estimate the hyperparameters of the MRGP model for the discrepancy functions without knowing the true value of $\boldsymbol{\theta}^*$.

Kennedy and O'Hagan [1] developed a procedure to obtain estimates of the hyperparameters of a single discrepancy function using the prior distribution of the calibration parameters. Specifically, this procedure constructs a likelihood function, marginalized with respect to the prior distribution of the calibration parameters, from the simulation and experimental data and using the MLEs of the computer model hyperparameters from Module 1. The experimental data consists of experimental response observations at N_e input sites, $\mathbf{X}^e = [\mathbf{x}_1^e, \ldots, \mathbf{x}_{N_e}^e]^T$ (transformed to the range [0,1]) with responses $\mathbf{Y}^e = [\mathbf{y}_1^e, \ldots, \mathbf{y}_q^e]$, where $\mathbf{y}_i^e = [y_i^e(\mathbf{x}_1^e), \ldots, y_i^e(\mathbf{x}_{N_e}^e)]^T$ (standardized). Then the MLE estimates of the hyperparameters $\boldsymbol{\phi}^\delta = \{\mathbf{B}^\delta, \boldsymbol{\Sigma}^\delta, \boldsymbol{\omega}^\delta, \boldsymbol{\lambda}\}$ occur at the location of the maximum of the likelihood function (marginalized with respect to the priors of the calibration parameters). Appendix C contains a detailed

description of this procedure. After estimating the hyperparameters of the two MRGP models in Modules 1 and 2, we use Bayes theorem to calculate the posterior distribution of $\boldsymbol{\theta}$, as follows.

2.3 Posterior of the Calibration Parameters (Module 3)

Module 3 calculates the posterior of the calibration parameters using the MLEs of the hyperparameters from Modules 1 and 2, the simulation data, and the experimental data. The posterior of the calibration parameters is

$$p(\boldsymbol{\theta} \mid \mathbf{d}, \hat{\boldsymbol{\phi}}) = \frac{p(\mathbf{d} \mid \boldsymbol{\theta}, \hat{\boldsymbol{\phi}})\, p(\boldsymbol{\theta})}{p(\mathbf{d} \mid \hat{\boldsymbol{\phi}})}, \tag{6}$$

where $\hat{\boldsymbol{\phi}}$ are the MLEs of $\boldsymbol{\phi} = \{\boldsymbol{\phi}^m, \boldsymbol{\phi}^\delta\}$, $\mathbf{d} = [vec(\mathbf{Y}^m)^T\ vec(\mathbf{Y}^e)^T]^T$ is the complete response data, and $p(\boldsymbol{\theta})$ is the prior of the calibration parameters (e.g., a uniform distribution defined *a priori*). The likelihood function $p(\mathbf{d} \mid \boldsymbol{\theta}, \hat{\boldsymbol{\phi}})$ is a multivariate normal distribution whose mean and covariance are determined by the two MRGP models for the computer model [Eq. (2)] and the experimental responses [Eq. (5)]. The denominator $p(\mathbf{d} \mid \hat{\boldsymbol{\phi}})$ of Eq. (6) is the marginal distribution of the data [18], which does not depend upon $\boldsymbol{\theta}$ (i.e. the denominator is a normalizing constant). Appendix D details the equations for calculating the posterior distribution in Eq. (6). The posterior distribution of the calibration parameters and the MLEs of the hyperparameters from Modules 1 and 2 influence the prediction of the experimental responses in the next and final module.

2.4 Prediction of the Experimental Responses and Discrepancy Function (Module 4)

After estimating the hyperparameters in Modules 1 and 2 and collecting the simulation and experimental data, the conditional (given a specific value of $\boldsymbol{\theta}$) posterior distribution for the experimental responses can be calculated at any point $\mathbf{x}$. The unconditional posterior distribution is then obtained by marginalizing the conditional posterior distribution with respect to the

posterior of the calibration parameters calculated in Module 3, as discussed in Appendix E. Therefore, the marginalized posterior distribution of the multiple experimental responses accounts for parameter uncertainty, model discrepancy, interpolation uncertainty, and experimental uncertainty (see our companion paper [2] for more details on the sources of uncertainty). Since the discrepancy functions are also represented by a MRGP model, their posterior distributions can be calculated in a similar manner. One should note that the MRGP model prediction of the discrepancy functions will depend on the value of the calibration parameters. Therefore, similar to the prediction of the physical experiments, we marginalize the prediction of the discrepancy functions with respect to the posterior distribution of the calibration parameters.

3 Identifiability: A Simply Supported Beam

In this section, we revisit the simply supported beam example, introduced in our companion paper [2], to show how using multiple responses can improve the identifiability of the calibration parameters and the discrepancy functions. Briefly (refer to [2] for details), the simply supported beam is fixed at one end and supported by a roller on the other end. A static force is applied to the midpoint of the beam to induce various responses, e.g., deflection and stress. The magnitude of this force was chosen as the design variable x, while Young's modulus was treated as the calibration parameter θ. For generating the "physical experiment" (which we take to be the same computer simulations but with a more sophisticated material law) data, the true value of the calibration parameter is set to $\theta^* = 206.8$ GPa. However, we treat θ^* as unknown during the analysis and assign a uniform prior distribution over the range $150 \leq \theta \leq 300$ GPa. The experimental uncertainty was set to zero (i.e., $\lambda_i = 0$).

3.1 Identifiability with Single Responses

Before using multiple responses for the simply supported beam, we first examine the resulting identifiability from using only single responses. Table 1 shows the six responses measured from the beam (for both the computer model y_i^m and the physical experiments y_i^e). The simulation data for each response $\mathbf{y}_i^m$ was collected using a 4 by 4 grid over the input space ($1300 \le x \le 2300$ N and $150 \le \theta \le 300$ GPa). The experimental data $\mathbf{y}_i^e$ was collected at $\mathbf{X}^e = [1300, 1400, 1500, \ldots, 2300]^T$ ($N_e = 11$). For each single response, the modular Bayesian approach, reviewed in ref. [2], was used to calculate the posterior distributions of the calibration parameter, the discrepancy function, and the experimental response.

Table 1 Responses measured in the simply supported beam.

Response	Description
y_1	Strain at the midpoint of the beam (mm)
y_2	Plastic strain at the midpoint of the beam (mm)
y_3	Angle of deflection at the end of the beam (radians) (y of ref. [2])
y_4	Internal energy of the beam (Joules)
y_5	Displacement at the middle of the beam (m)
y_6	Plastic dissipation energy (Joules)

Table 2 Single response (SR) posterior mean ($\mu_{i,post}^{SR}$ in GPa) and standard deviation ($\sigma_{i,post}^{SR}$ in GPa) of the calibration parameter θ.

Response	SR posterior distribution of θ	
y_i	$\mu_{i,post}^{SR}$	$\sigma_{i,post}^{SR}$
y_1	256.74	31.68
y_2	264.70	27.15
y_3	262.18	26.67
y_4	261.79	26.92
y_5	261.15	27.84
y_6	254.48	33.30

Because of the relatively large amount of experimental data and no experimental uncertainty, the posterior distribution of the experimental response for each single response was accurate and precise (see Fig. 8(a) of ref. [2] for y3 of this paper). Even so, the calibration parameter and the discrepancy function are not identifiable as evident from the high uncertainty reflected by their posterior distributions, as shown in Fig. 2 for responses 1 through 4 (similar results were obtained for y5 and y6 but are not shown due to space). Table 2 shows the resulting posterior mean and standard deviation of the calibration parameter for each single response. The large posterior standard deviations in Table 2 again indicate a lack of identifiability. Since the experimental data are already relatively dense, additional data will not improve identifiability when using only a single response. Instead, in the remainder of this section we explore the use of multiple responses to enhance identifiability.

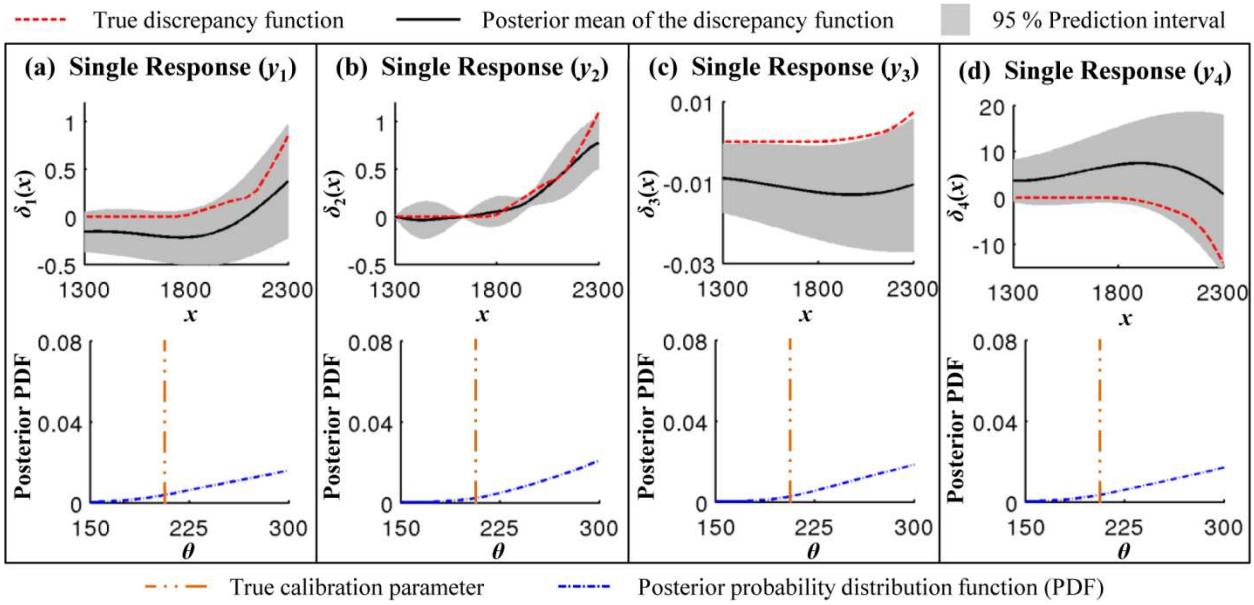

Fig. 2 Posterior distributions of the discrepancy function and the calibration parameter using a single response for (a) y_1, (b) y_2, (c) y_3, and (d) y_4.

3.2 Identifiability with Multiple Responses

With the goal of enhancing identifiability, the procedure described in Section 2 combines information from multiple responses. In general, there will often be many responses that are automatically calculated in the computer model and that could potentially be measured experimentally. Ideally, one might like to include all of the responses in the model updating formulation, but this is not practical for a number of reasons. First, it can become more computationally expensive to implement the modular Bayesian approach with additional responses and including too many responses may result in numerical instabilities. Furthermore, some responses may be largely redundant, containing nearly the same information as other responses, which will not improve identifiability. As seen below, different combinations of responses result in drastically different identifiability. Third, although multiple responses are available for free in the computer simulation, their experimental measurement may involve prohibitive costs.

For ease of illustration and computational reasons, the multiple responses that we consider are various pairs of responses. To explore how different pairs of responses affect identifiability in the simply supported beam example, we calculate the relevant posterior distributions for the 15 pairs of responses from Table 1. Specifically, for each pair of responses, we use the multiple response modular Bayesian approach discussed in Section 2 to calculate the posterior distributions of the calibration parameter, the discrepancy functions, and the experimental responses. The modular Bayesian approach begins with Module 1, which creates a MRGP model of the computer model using the simulation data $\mathbf{y}_i^m$ (observed at the same input settings $\mathbf{X}^m$ as in Section 3.1). In Module 2, the hyperparameters of the MRGP model for the discrepancy functions are estimated from the experimental data $\mathbf{y}_i^e$ (observed at the same input

settings $\mathbf{X}^e$ as in Section 3.1), the prior distribution of the calibration parameter, and the hyperparameter MLEs from Module 1. Finally, Modules 3 and 4 calculate the posterior distributions for the calibration parameter, the discrepancy functions, and the experimental responses. As with the single responses in Section 3.1, the posterior of the experimental responses were extremely accurate and precise, although the identifiability of the calibration parameters varied widely from pair to pair of responses.

Table 3 Multiple response posterior mean $\mu_{ij,post}^{MR}$ (in GPa) and standard deviation $\sigma_{ij,post}^{MR}$ (in GPa) of the calibration parameter θ and identifiability improvement compared to the single response.

Responses			MR posterior distribution of θ		$\sigma_{ij,min}^{SR}$	Std. dev. improvement
y_i	y_j		$\mu_{ij,post}^{MR}$	$\sigma_{ij,post}^{MR}$		
y_1	y_2	(1)	259.81	27.37	31.68	14 %
y_1	y_3		213.70	10.13	26.67	62 %
y_1	y_4		215.00	9.29	26.92	65 %
y_1	y_5		219.83	15.57	27.84	44 %
y_1	y_6		247.33	24.08	31.68	24 %
y_2	y_3	(2)	218.82	10.96	26.67	59 %
y_2	y_4		228.28	18.35	26.92	32 %
y_2	y_5		224.50	15.42	27.84	45 %
y_2	y_6		265.96	26.46	33.30	21 %
y_3	y_4	(3)	209.23	5.49	26.67	79 %
y_3	y_5		210.39	6.27	26.67	76 %
y_3	y_6		210.30	4.86	26.67	82 %
y_4	y_5		213.23	3.63	26.92	87 %
y_4	y_6		211.91	3.90	26.92	86 %
y_5	y_6		210.35	4.67	27.84	83 %

Table 3 shows the resulting posterior mean and standard deviation of the calibration parameter for each set of responses. In cases where the posterior distribution is not normal, as in Fig. 2, the standard deviation can be an oversimplification of the dispersion of the posterior

distribution, but it is still perhaps the most relevant single measure of identifiability. To quantify identifiability improvement, we compare the multiple response posterior standard deviation $\sigma_{ij,post}^{MR}$ to the best single response posterior standard deviation of the pair (i.e. the smaller of the two single response posterior standard deviations $\sigma_{ij,min}^{SR} = \min(\sigma_{i,post}^{SR}, \sigma_{j,post}^{SR})$). The posterior standard deviation improvement, "std. dev. improvement" in Table 3, is defined as $\left(\sigma_{ij,min}^{SR} - \sigma_{ij,post}^{MR}\right)/\sigma_{ij,min}^{SR}$. We focus on this as the key metric, as it represents the improvement in identifiability that is achieved by considering multiple responses in this example. The posterior standard deviations in absolute terms are also important, but these depend on a host of other complex factors and considerations.

To better visualize improved identifiability, Fig. 3 shows the posterior distributions of the calibration parameter and the discrepancy functions for three sets of multiple responses with different degrees of identifiability improvement. The sets of responses shown in Fig. 3 (a), (b), and (c) will from hereon be referred to as Case 1, 2, and 3, respectively, and are labeled in Table 3. These three cases were chosen as a representative set of the 15 combinations of responses show in Table 3.

As can be seen in Table 3 and Fig. 3, the extent of the identifiability improvement varies widely from pair to pair. In Case 1, the combination of the strain y_1 and the plastic strain y_2 enhances identifiability little beyond that of a single response, and the resulting posterior distributions in Fig. 2 (a) and (b) for the single responses are similar to those in Fig. 3(a) for the pair of responses (a 14 % improvement, from Table 3). In contrast, in Case 2, the combination of the plastic strain y_2 and the angle of deflection y_3 resulted in a much more substantial

improvement in identifiability. Specifically, the posterior distributions shown in Fig. 3(b) represent a 59% improvement in posterior standard deviation (see Table 3).

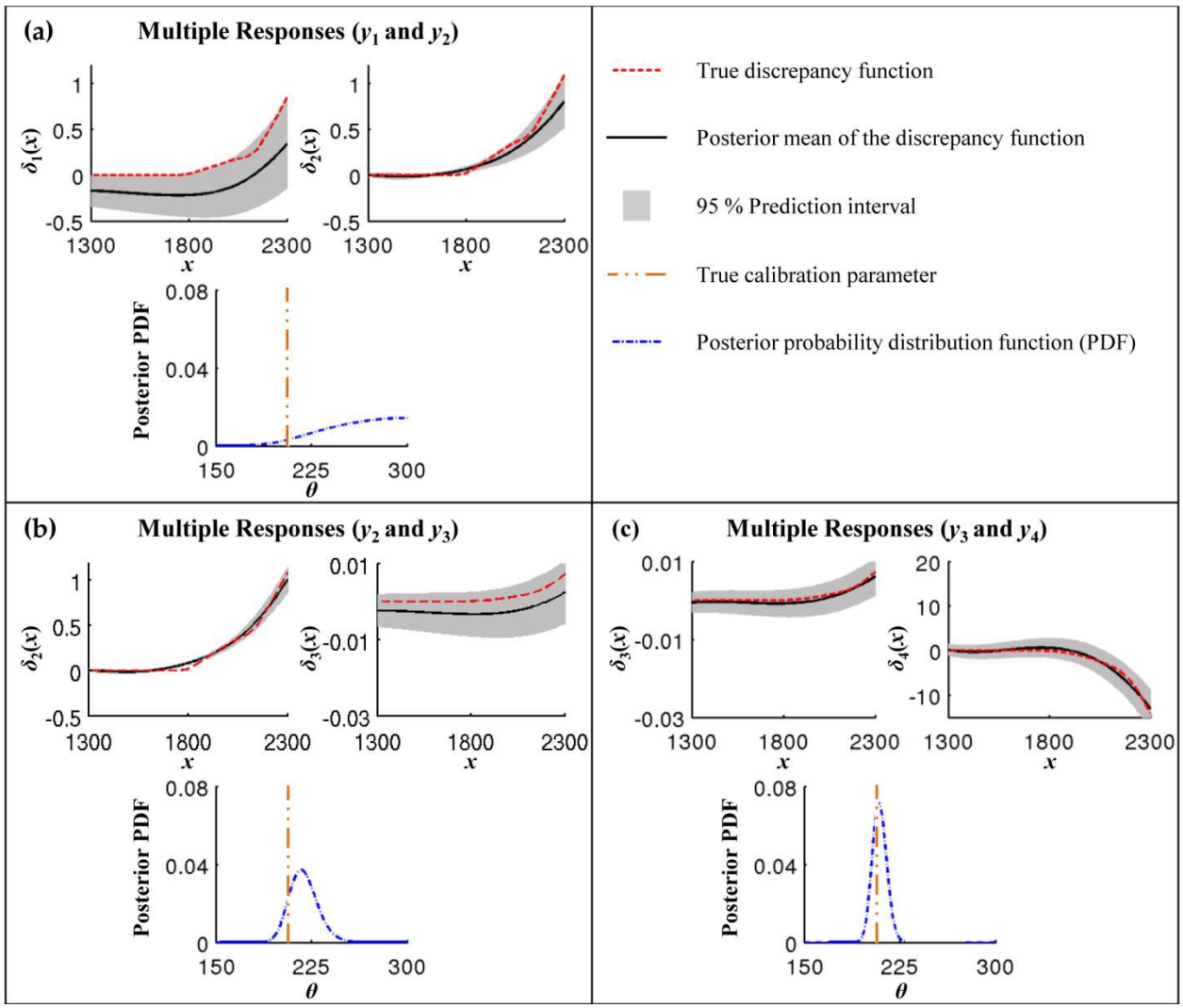

Fig. 3 Posterior distributions of the discrepancy functions and the calibration parameter using multiple responses for (a) y_1 and y_2, (b) y_2 and y_3, ad (c) y_3 and y_4.

In Case 3, the combination of the angle of deflection y_3 and the internal energy y_4 resulted in even more substantially enhanced identifiability. The corresponding posterior distribution in Fig. 3(c) is much more narrowly dispersed (a 79% improvement, from Table 3) than for the single response cases. Additionally, the posterior mean of the calibration parameter is quite close to the true value of $\theta^* = 206.8$ (marked by the vertical line in Fig. 3(c)). Moreover, the

posteriors of the discrepancy functions $\delta_3(\mathbf{x})$ and $\delta_4(\mathbf{x})$ capture the true functions with relatively small uncertainty. These three cases are representative of the varied improvements in identifiability for all pairs of responses listed in Table 3.

In the simply supported beam example, 12 out of the 15 pairs of responses resulted in an identifiability improvement of 25% or more. The degree of improvement varied from little (e.g., 14%) to moderate (e.g., 24% and 32%) to substantial (e.g., 87%). This example illustrates that if identifiability cannot be obtained using a single response, multiple responses may be used to achieve improved identifiability.

4 Conclusions

Computer models in engineering applications are not perfect representations of physical reality, and distinguishing between the sources of model uncertainty is often difficult. In this work, we have shown that identifiability can be enhanced by using multiple responses in model updating. The approach is most beneficial when data from a single response results in a lack of identifiability.

We first extended the existing single response modular Bayesian approach to multiple responses. This extension required the use of multiple response Gaussian process (MRGP) models to represent the multiple computer simulation responses and the multiple discrepancy functions. The multiple response modular Bayesian approach allows one to calculate the posterior distributions of the calibration parameters upon which the multiple responses share a mutual dependence, the posterior distributions of the discrepancy functions of each response, and the predictions of multiple experimental responses.

By applying the multiple response modular Bayesian approach developed in this work to the simply supported beam example, we showed that including multiple responses can improve

identifiability. In this example, the majority of paired responses resulted in an identifiability improvement of over 50%, and some pairs resulted in an improvement of over 85%. Across all pairs, the resulting improvement in identifiability varied from 14% to 87%. Overall, the simply supported beam example illustrates the effectiveness of using additional information from multiple responses to improve identifiability, when identifiability cannot be achieved using a single response.

As with the single response modular Bayesian approach, one limitation of the proposed multiple response modular Bayesian approach is the computational cost, which is predominantly affected by the number of observations (see our companion paper [2]). Using additional responses also increases the computational cost, but not as significantly as adding observed data. For example, when using two responses, 17 seconds (on a single Intel 2.66 GHz processer) of computation time was required to calculate the posterior distributions of the simply supported beam example; whereas, when using a single response, only 10 seconds of computation time was required. The increase in computation time was a result of calculating the posterior distributions for two experimental responses and two discrepancy functions in Module 4. Also, one should notice that the increase in computational cost was due to the Bayesian computations, and not the computer simulations, because no additional computer simulations were run.

Another limitation of the proposed approach is that considering additional responses can result in numerical instabilities. Theoretically, using additional responses should potentially improve, and never worsen, identifiability. However, additional responses can also introduce numerical conditioning problems. For example, when using more than two responses in the simply supported beam example, estimation of $\mathbf{\Sigma}^{\delta}$ for the discrepancy functions in Module 2 became problematic. Due to the high correlations between the responses, the estimate of $\mathbf{\Sigma}^{\delta}$ was

close to singular, which caused numerical instability in the subsequent calculations. A potentially fruitful avenue for further research is to develop more numerically stable methods of handling a larger number of responses.

Although in the examples the number of calibration parameters was less than or equal to the number of responses, this is not a requirement for identifiability (much like how regression modeling with a single response variable can be used to fit a model with multiple parameters). Of course, using more responses will generally improve identifiability, but this improvement will depend on the specifics of the system (including the form of the computer model, the physical experiments, and the discrepancy functions), the multiple responses being measured (as seen in Fig. 2 and Fig. 3), and the location of the measurements in $\mathbf{x}$. In light of this, it is difficult to generalize how many responses are necessary to achieve identifiability based on the number of calibration parameters.

Future research is needed to better understand why certain combinations of multiple responses improve identifiability much more than others. This understanding could be used to develop a systematic methodology for choosing the additional responses to measure experimentally that result in optimal improvement in identifiability. In the simply supported beam example, it appears that combining the two most similar responses (y_1 and y_2, which are both strains) results in the least improvement. But no other clear patterns emerge from the example. Further study of this important issue is the subject of ongoing research.

Together, this paper and the companion paper [2] shed light on the challenging problem of identifying calibration parameters and discrepancy functions when combining data from a computer model and physical experiments. They demonstrate that identifiability is often possible and can be reasonably achieved with proper analyses in certain engineering systems.

5 Acknowledgements

The grant support from the National Science Foundation (CMMI-0928320 and CMMI-0758557) and the U.S. Army Tank-Automotive Research Development & Engineering Center (TARDEC) (contract number W911NF11D0001-0037) is greatly acknowledged. The views expressed are those of the authors and do not necessarily reflect the views of the sponsors.

Nomenclature

$\mathbf{B}^m$ Regression coefficients for the computer model MRGP

$\mathbf{B}^\delta$ Regression coefficients for the discrepancy functions MRGP

$\otimes$ Kronecker product operator

$\mathbf{d}$ Combined vector of data ($\mathbf{d} = [vec(\mathbf{Y}^m)^T \quad \text{vec}(\mathbf{Y}^e)^T]^T$)

$\mathbf{h}^m(\mathbf{x}, \boldsymbol{\theta})$, $\mathbf{H}^m$ Regression functions for the computer model MRGP

$\mathbf{h}^\delta(\mathbf{x})$, $\mathbf{H}^\delta$ Regression functions for the discrepancy function MRGP

$vec(\bullet)$ Vectorization operator

$y_i^e(\mathbf{x})$, $\mathbf{y}^e(\mathbf{x})$ Experimental response for $i = 1, \ldots, q$

$\mathbf{y}_i^e$, $\mathbf{Y}^e$ Experimental data (vector ($N_e \times 1$) and matrix ($N_e \times q$)) collected at input settings $\mathbf{X}^e$

$y_i^m(\mathbf{x}, \boldsymbol{\theta})$, $\mathbf{y}^m(\mathbf{x}, \boldsymbol{\theta})$ Computer model for $i = 1, \ldots, q$

$\mathbf{y}_i^m$, $\mathbf{Y}^m$ Simulation data (vector ($N_m \times 1$) and matrix ($N_m \times q$)) collected at input settings $\mathbf{X}^m$ and $\boldsymbol{\Theta}^m$

$\delta(\mathbf{x})$, $\boldsymbol{\delta}(\mathbf{x})$ Discrepancy function

λ, $\boldsymbol{\lambda}$ Variance of the experimental uncertainty

$\mu_{ij,post}^{MR}$ Posterior mean of the calibration parameter using multiple responses i and j

$\mu_{i,post}^{SR}$ Posterior mean of the calibration parameter using a single response i

$\boldsymbol{\Sigma}^m$ Covariance between the responses of the computer model MRGP

$\boldsymbol{\Sigma}^\delta$ Covariance between the discrepancy functions of the discrepancy functions MRGP

$\sigma_{ij,post}^{MR}$ Posterior standard deviation of the calibration parameter using multiple responses *i* and *j*

$\sigma_{i,post}^{SR}$ Posterior standard deviation of the calibration parameter using a single response *i*

$\sigma_{ij,min}^{SR}$ Minimum of the posterior standard deviations resulting from single responses *i* and *j*

Appendix A: Estimates of the Hyperparameters for the Computer Model MRGP

To obtain the MLEs of the hyperparameters for the computer model MRGP model, we first construct the multivariate normal likelihood function as

$$p(vec(\mathbf{Y}^m)\,|\,\mathbf{B}^m,\boldsymbol{\Sigma}^m,\boldsymbol{\omega}^m) = (2\pi)^{-qN_m/2}\left|\Sigma^m\right|^{-N_m/2}\left|\mathbf{R}^m\right|^{-q/2} \times\exp\left\{-\frac{1}{2}vec(\mathbf{Y}^m-\mathbf{H}^m\mathbf{B}^m)^T\left(\Sigma^m\otimes\mathbf{R}^m\right)^{-1}vec(\mathbf{Y}^m-\mathbf{H}^m\mathbf{B}^m)\right\}, \quad \text{(A1)}$$

where *vec*(•) is the vectorization of the matrix (stacking of the columns), $\otimes$ denotes the Kronecker product, $\mathbf{R}^m$ is a $N_m \times N_m$ correlation matrix whose *i*th-row, *j*th-column entry is $R^m((\mathbf{x}_i^m,\boldsymbol{\theta}_i^m),(\mathbf{x}_j^m,\boldsymbol{\theta}_j^m))$, and $\mathbf{H}^m=[\mathbf{h}^m(\mathbf{x}_1^m,\boldsymbol{\theta}_1^m)^T,\ldots,\mathbf{h}^m(\mathbf{x}_{N_m}^m,\boldsymbol{\theta}_{N_m}^m)^T]^T$. Taking the log of Eq. (A1) yields

$$\ln(p(vec(\mathbf{Y}^m)\,|\,\mathbf{B}^m,\boldsymbol{\Sigma}^m,\boldsymbol{\omega}^m)) = -\frac{qN_m}{2}\ln(2\pi)-\frac{N_m}{2}\ln\left(\left|\Sigma^m\right|\right)-\frac{q}{2}\ln\left(\left|\mathbf{R}^m\right|\right) -\frac{1}{2}vec(\mathbf{Y}^m-\mathbf{H}^m\mathbf{B}^m)^T\left(\Sigma^m\otimes\mathbf{R}^m\right)^{-1}vec(\mathbf{Y}^m-\mathbf{H}^m\mathbf{B}^m) \quad \text{(A2)}$$

The MLE of $\mathbf{B}^m$ is found by setting the derivative of Eq. (A2) with respect to $\mathbf{B}^m$ equal to zero, which gives

$$\hat{\mathbf{B}}^m=[(\mathbf{H}^m)^T(\mathbf{R}^m)^{-1}\mathbf{H}^m]^{-1}(\mathbf{H}^m)^T(\mathbf{R}^m)^{-1}\mathbf{Y}^m. \quad \text{(A3)}$$

The MLE of $\boldsymbol{\Sigma}^m$ is found using result 4.10 of ref. [19], which yields

$$\hat{\boldsymbol{\Sigma}}^m = \frac{1}{N^m}(\mathbf{Y}^m - \mathbf{H}^m\hat{\mathbf{B}}^m)^T(\mathbf{R}^m)^{-1}(\mathbf{Y}^m - \mathbf{H}^m\hat{\mathbf{B}}^m). \quad \text{(A4)}$$

Finally, the MLE of $\boldsymbol{\omega}^m$, denoted by $\hat{\boldsymbol{\omega}}^m$, is found by numerically maximizing Eq. (A2) after plugging in the MLEs of $\mathbf{B}^m$ and $\boldsymbol{\Sigma}^m$.

Appendix B: Posterior Distributions of the Computer Responses

After observing $\mathbf{Y}^m$, the posterior of the computer response $y_i^m(\mathbf{x},\boldsymbol{\theta})$ given $\mathbf{Y}^m$ (and given $\boldsymbol{\omega}^m$ and $\boldsymbol{\Sigma}^m$ and assuming a noninformative prior for $\mathbf{B}^m$) is Gaussian with mean and covariance

$$E[\mathbf{y}^m(\mathbf{x},\boldsymbol{\theta}) \mid \mathbf{Y}^m, \boldsymbol{\phi}^m] = \mathbf{h}^m(\mathbf{x},\boldsymbol{\theta})\hat{\mathbf{B}}^m + \mathbf{r}^m(\mathbf{x},\boldsymbol{\theta})^T(\mathbf{R}^m)^{-1}(\mathbf{Y}^m - \mathbf{H}^m\hat{\mathbf{B}}^m) \quad \text{(B1)}$$

$$\begin{aligned} Cov[\mathbf{y}^m(\mathbf{x},\boldsymbol{\theta})), \mathbf{y}^m(\mathbf{x}',\boldsymbol{\theta}') \mid \mathbf{Y}^m, \boldsymbol{\phi}^m] = \boldsymbol{\Sigma}^m \Big\{ & R^m((\mathbf{x},\boldsymbol{\theta}),(\mathbf{x}',\boldsymbol{\theta}')) \\ & - \mathbf{r}^m(\mathbf{x},\boldsymbol{\theta})^T(\mathbf{R}^m)^{-1}\mathbf{r}^m(\mathbf{x}',\boldsymbol{\theta}') + [\mathbf{h}^m(\mathbf{x},\boldsymbol{\theta})^T - (\mathbf{H}^m)^T(\mathbf{R}^m)^{-1}\mathbf{r}^m(\mathbf{x},\boldsymbol{\theta})]^T, \\ & \times [(\mathbf{H}^m)^T(\mathbf{R}^m)^{-1}\mathbf{H}^m]^{-1}[\mathbf{h}^m(\mathbf{x},\boldsymbol{\theta})^T - (\mathbf{H}^m)^T(\mathbf{R}^m)^{-1}\mathbf{r}^m(\mathbf{x},\boldsymbol{\theta})] \Big\} \end{aligned} \quad \text{(B2)}$$

where $\mathbf{r}^m(\mathbf{x},\boldsymbol{\theta})$ is a $N_m \times 1$ vector whose ith element is $R^m((\mathbf{x}_i^m,\boldsymbol{\theta}_i^m),(\mathbf{x},\boldsymbol{\theta}))$. Using an empirical Bayes approach, the MLEs of the hyperparameters from Appendix A are plugged into Eqs. (B1) and (B2) to calculate the posterior distribution of the computer responses. Notice that Eqs. (B1) and (B2) are analogous to the single response GP model results (reviewed in Section 3 of our companion paper [2]).

Appendix C: Estimates of the Hyperparameters for the Discrepancy Functions MRGP

To estimate the hyperparameters $\boldsymbol{\phi}^\delta = \{\mathbf{B}^\delta, \boldsymbol{\Sigma}^\delta, \boldsymbol{\omega}^\delta, \boldsymbol{\lambda}\}$ of the MRGP model representing the discrepancy functions, we use the procedure outlined by Kennedy and O'Hagan [1], which is modified in this paper to handle multiple responses. This procedure begins by obtaining a posterior of the experimental responses given the simulation data and the hyperparameters from Module 1, which has prior mean and covariance

$$E[\mathbf{y}^e(\mathbf{x}) \mid \mathbf{Y}^m, \hat{\boldsymbol{\phi}}^m, \boldsymbol{\theta} = \boldsymbol{\theta}^*] = E[\mathbf{y}^m(\mathbf{x},\boldsymbol{\theta}^*) \mid \mathbf{Y}^m, \hat{\boldsymbol{\phi}}^m] + \mathbf{h}^\delta(\mathbf{x})\mathbf{B}^\delta \quad \text{(C1)}$$

$$Cov[\mathbf{y}^e(\mathbf{x}), \mathbf{y}^e(\mathbf{x}') \mid \mathbf{Y}^m, \hat{\boldsymbol{\phi}}^m, \boldsymbol{\theta} = \boldsymbol{\theta}^*] = \boldsymbol{\Sigma}^\delta R^\delta(\mathbf{x}, \mathbf{x}) + \boldsymbol{\lambda} + Cov[\mathbf{y}^m(\mathbf{x}, \boldsymbol{\theta}^*)), \mathbf{y}^m(\mathbf{x}', \boldsymbol{\theta}^*)) \mid \mathbf{Y}^m, \hat{\boldsymbol{\phi}}^m] \quad \text{(C2)}$$

where $\hat{\boldsymbol{\phi}}^m$ are the MLEs of the hyperparameters for the computer model MRGP model. Since Eqs. (C1) and (C2) depend on the unknown true value of $\boldsymbol{\theta}^*$, we integrate these two equations with respect to the prior distribution of $\boldsymbol{\theta}$ ($p(\boldsymbol{\theta})$) via

$$\begin{aligned} E[\mathbf{y}^e(\mathbf{x}) \mid \mathbf{Y}^m, \hat{\boldsymbol{\phi}}^m] &= \int E[\mathbf{y}^e(\mathbf{x}) \mid \mathbf{Y}^m, \hat{\boldsymbol{\phi}}^m, \boldsymbol{\theta}] p(\boldsymbol{\theta}) d\boldsymbol{\theta} \\ Cov[\mathbf{y}^e(\mathbf{x}), \mathbf{y}^e(\mathbf{x}') \mid \mathbf{Y}^m, \hat{\boldsymbol{\phi}}^m] &= \int Cov[\mathbf{y}^e(\mathbf{x}), \mathbf{y}^e(\mathbf{x}') \mid \mathbf{Y}^m, \hat{\boldsymbol{\phi}}^m, \boldsymbol{\theta}] p(\boldsymbol{\theta}) d\boldsymbol{\theta} \end{aligned} \quad \text{(C3)}$$

Kennedy and O'Hagan [20] provide closed form solutions for Eq. (C3) under the conditions of Gaussian correlation functions, constant regression functions, and normal prior distributions for $\boldsymbol{\theta}$ (for details refer to Section 3 of [20] and Section 4.5 of [1]). In this work, we use similar closed form solutions except that we assume a uniform prior distribution for $\boldsymbol{\theta}$.

After observing the experimental data, $\mathbf{Y}^e$, one can construct a multivariate normal likelihood function with mean and variance from Eq. (C3). The MLEs of $\boldsymbol{\phi}^\delta$ maximize this likelihood function. The MLE of $\mathbf{B}^\delta$ can found by setting the analytical derivative of this likelihood function with respect to $\mathbf{B}^\delta$ equal to zero (see Section 2 of ref. [20]). However, there are no analytical derivatives with respect to the hyperparameters $\boldsymbol{\Sigma}^\delta$, $\boldsymbol{\omega}^\delta$, and $\boldsymbol{\lambda}$. Therefore, numerical optimization techniques are needed to find these MLEs.

Appendix D: Posterior Distribution of the Calibration Parameters

The posterior for the calibration parameters in Eq. (6) involves the likelihood function $p(\mathbf{d} \mid \boldsymbol{\theta}, \hat{\boldsymbol{\phi}})$ and the marginal posterior distribution for the data $p(\mathbf{d} \mid \hat{\boldsymbol{\phi}})$. The likelihood function is multivariate normal with mean vector and covariance matrix defined as

$$\mathbf{m}(\boldsymbol{\theta}) = \mathbf{H}(\boldsymbol{\theta})\hat{\mathbf{B}} = \begin{bmatrix} \mathbf{I}_q \otimes \mathbf{H}^m & \mathbf{0} \\ \mathbf{I}_q \otimes \mathbf{H}^m(\mathbf{X}^e, \boldsymbol{\theta}) & \mathbf{I}_q \otimes \mathbf{H}^\delta \end{bmatrix} \begin{bmatrix} vec(\hat{\mathbf{B}}^m) \\ vec(\hat{\mathbf{B}}^\delta) \end{bmatrix} \quad \text{(D1)}$$

$$\mathbf{V}(\boldsymbol{\theta}) = \begin{bmatrix} \hat{\boldsymbol{\Sigma}}^m \otimes \mathbf{R}^m & \hat{\boldsymbol{\Sigma}}^m \otimes \mathbf{C}^m \\ \hat{\boldsymbol{\Sigma}}^m \otimes \mathbf{C}^{mT} & \hat{\boldsymbol{\Sigma}}^m \otimes \mathbf{R}^m(\mathbf{X}^e, \boldsymbol{\theta}) + \hat{\boldsymbol{\Sigma}}^\delta \otimes \mathbf{R}^\delta + \hat{\boldsymbol{\lambda}} \otimes \mathbf{I}_{N_e} \end{bmatrix}, \quad \text{(D2)}$$

where $\hat{\mathbf{B}} = \left(\mathbf{H}(\boldsymbol{\theta})^T \mathbf{V}(\boldsymbol{\theta})^{-1} \mathbf{H}(\boldsymbol{\theta})\right)^{-1} \mathbf{H}(\boldsymbol{\theta})^T \mathbf{V}(\boldsymbol{\theta})^{-1} \mathbf{d}$, which is calculated based on the entire data set (instead of using the estimates from Modules 1 and 2 for $\mathbf{B}^m$ and $\mathbf{B}^\delta$) as detailed in Section 4 of [20]. $\mathbf{H}^m(\mathbf{X}^e, \boldsymbol{\theta}) = [\mathbf{h}^m(\mathbf{x}_1^e, \boldsymbol{\theta})^T, \ldots, \mathbf{h}^m(\mathbf{x}_{N_e}^m, \boldsymbol{\theta})^T]^T$ and $\mathbf{H}^\delta = [\mathbf{h}^\delta(\mathbf{x}_1^e)^T, \ldots, \mathbf{h}^\delta(\mathbf{x}_{N_e}^\delta)^T]^T$ denote the specified regression functions for the computer model and the discrepancy functions at the input settings $\mathbf{X}^e$. $\mathbf{C}^m$ denotes the $N_m \times N_e$ matrix with ith-row, jth-column entries $R^m((\mathbf{x}_i^m, \boldsymbol{\theta}_i^m), (\mathbf{x}_j^e, \boldsymbol{\theta}))$. $\mathbf{R}^m(\mathbf{X}^e, \boldsymbol{\theta})$ denotes the $N_e \times N_e$ matrix with ith-row, jth-column entries $R^m((\mathbf{x}_i^e, \boldsymbol{\theta}), (\mathbf{x}_j^e, \boldsymbol{\theta}))$. $\mathbf{R}^\delta$ denotes the $N_e \times N_e$ matrix with ith-row, jth-column entries $R^\delta(\mathbf{x}_i^e, \mathbf{x}_j^e)$. Finally, $\mathbf{I}_q$ and $\mathbf{I}_{N_e}$ denote the $q \times q$ and $N_e \times N_e$ identity matrices.

The marginal posterior distribution for the data $p(\mathbf{d} \mid \hat{\boldsymbol{\phi}})$ is

$$p(\mathbf{d} \mid \hat{\boldsymbol{\phi}}) = \int p(\mathbf{d} \mid \boldsymbol{\theta}, \hat{\boldsymbol{\phi}}) \, p(\boldsymbol{\theta}) d\boldsymbol{\theta}, \quad \text{(D3)}$$

which can be calculated using any numerical integration technique. In this paper, we use Legendre-Gauss quadrature for our low dimensional examples. Alternatively, Markov Chain Monte Carlo (MCMC) could be used to sample complex posterior distributions such as those in Eq. (6).

Appendix E: Posterior Distribution of the Experimental Responses

Since a MRGP model represents the experimental responses, the conditional (given $\boldsymbol{\theta}$) posterior distribution at any point $\mathbf{x}$ is Gaussian with mean and covariance defined as (assuming

a noninformative prior on $\mathbf{B}^m$ and $\mathbf{B}^\delta$ and using the empirical Bayes approach that treats $\boldsymbol{\phi} = \hat{\boldsymbol{\phi}}$ as fixed)

$$E[\mathbf{y}^e(\mathbf{x})^T \mid \boldsymbol{\theta}, \mathbf{d}, \hat{\boldsymbol{\phi}}] = \mathbf{h}(\mathbf{x}, \boldsymbol{\theta})\hat{\mathbf{B}} + \mathbf{t}(\mathbf{x}, \boldsymbol{\theta})^T \mathbf{V}(\boldsymbol{\theta})^{-1}(\mathbf{d} - \mathbf{H}(\boldsymbol{\theta})\hat{\mathbf{B}}) \tag{E1}$$

$$\begin{aligned} Cov[\mathbf{y}^e(\mathbf{x})^T, \mathbf{y}^e(\mathbf{x}')^T \mid \boldsymbol{\theta}, \mathbf{d}, \hat{\boldsymbol{\phi}}] &= \hat{\boldsymbol{\Sigma}}^m R^m((\mathbf{x}, \boldsymbol{\theta}), (\mathbf{x}', \boldsymbol{\theta})) + \hat{\boldsymbol{\Sigma}}^\delta R^\delta(\mathbf{x}, \mathbf{x}') \\ &+ \hat{\boldsymbol{\lambda}} - \mathbf{t}(\mathbf{x}, \boldsymbol{\theta})^T \mathbf{V}(\boldsymbol{\theta})^{-1}\mathbf{t}(\mathbf{x}', \boldsymbol{\theta}) + \left(\mathbf{h}(\mathbf{x}, \boldsymbol{\theta})^T - \mathbf{H}(\boldsymbol{\theta})^T \mathbf{V}(\boldsymbol{\theta})^{-1}\mathbf{t}(\mathbf{x}, \boldsymbol{\theta})\right)^T \\ &\times \left(\mathbf{H}(\boldsymbol{\theta})^T \mathbf{V}(\boldsymbol{\theta})^{-1}\mathbf{H}(\boldsymbol{\theta})\right)^{-1} \left(\mathbf{h}(\mathbf{x}', \boldsymbol{\theta})^T - \mathbf{H}(\boldsymbol{\theta})^T \mathbf{V}(\boldsymbol{\theta})^{-1}\mathbf{t}(\mathbf{x}', \boldsymbol{\theta})\right) \end{aligned} \tag{E2}$$

where

$$\mathbf{t}(\mathbf{x}, \boldsymbol{\theta}) = \begin{bmatrix} \hat{\boldsymbol{\Sigma}}^m \otimes \mathbf{R}^m((\mathbf{X}^m, \boldsymbol{\Theta}^m), (\mathbf{x}, \boldsymbol{\theta})) \\ \hat{\boldsymbol{\Sigma}}^m \otimes \mathbf{R}^m((\mathbf{X}^e, \boldsymbol{\theta}), (\mathbf{x}, \boldsymbol{\theta})) + \hat{\boldsymbol{\Sigma}}^\delta \otimes \mathbf{R}^\delta(\mathbf{X}^e, \mathbf{x}) \end{bmatrix} \tag{E3}$$

$$\mathbf{h}(\mathbf{x}, \boldsymbol{\theta}) = \begin{bmatrix} \mathbf{I}_q \otimes \mathbf{h}^m(\mathbf{x}, \boldsymbol{\theta}) & \mathbf{I}_q \otimes \mathbf{h}^\delta(\mathbf{x}) \end{bmatrix}. \tag{E4}$$

$\mathbf{R}^m((\mathbf{X}^m, \boldsymbol{\Theta}^m), (\mathbf{x}, \boldsymbol{\theta}))$ is a $N_m \times 1$ vector whose ith entry is $R^m((\mathbf{x}_i^m, \boldsymbol{\theta}_i^m), (\mathbf{x}, \boldsymbol{\theta}))$, $\mathbf{R}^m((\mathbf{X}^e, \boldsymbol{\theta}), (\mathbf{x}, \boldsymbol{\theta}))$ is a $N_e \times 1$ vector whose ith entry is $R^m((\mathbf{x}_i^e, \boldsymbol{\theta}), (\mathbf{x}, \boldsymbol{\theta}))$, and $\mathbf{R}^\delta(\mathbf{X}^e, \mathbf{x})$ is a $N_e \times 1$ vector whose ith entry is $R^\delta(\mathbf{x}_i^e, \mathbf{x})$.

To calculate the unconditional posterior distributions (marginalized with respect to $\boldsymbol{\theta}$) of the experimental responses, we marginalize the conditional posterior distributions with respect to the posterior distribution of the calibration parameters from Module 3. The mean and covariance of the unconditional posterior distributions can be written as

$$E[\mathbf{y}^e(\mathbf{x})^T \mid \mathbf{d}, \hat{\boldsymbol{\phi}}] = E[E[\mathbf{y}^e(\mathbf{x})^T \mid \boldsymbol{\theta}, \mathbf{d}, \hat{\boldsymbol{\phi}}]] \tag{E5}$$

$$\begin{aligned} Cov[\mathbf{y}^e(\mathbf{x})^T, \mathbf{y}^e(\mathbf{x}')^T \mid \mathbf{d}, \hat{\boldsymbol{\phi}}] &= E[Cov[\mathbf{y}^e(\mathbf{x})^T, \mathbf{y}^e(\mathbf{x}')^T \mid \boldsymbol{\theta}, \mathbf{d}, \hat{\boldsymbol{\phi}}]] \\ &+ Cov[E[\mathbf{y}^e(\mathbf{x})^T \mid \boldsymbol{\theta}, \mathbf{d}, \hat{\boldsymbol{\phi}}], E[\mathbf{y}^e(\mathbf{x}')^T \mid \boldsymbol{\theta}, \mathbf{d}, \hat{\boldsymbol{\phi}}]] \end{aligned}, \tag{E6}$$

where the outer expectation and covariance are with respect to the posterior distribution of the calibration parameters. Eqs. (E5) and (E6) are derived using the law of total expectation and the

law of total covariance [21]. Due to the complexity of the posterior distribution of the calibration parameters, the marginalization requires numerical integration methods. For the examples in this paper, we use Legendre-Gauss quadrature.

References

[1] Kennedy, M. C., and O'Hagan, A., 2001, "Bayesian Calibration of Computer Models," Journal of the Royal Statistical Society: Series B, 63(3), pp. 425-464.

[2] Arendt, P., Apley, D., and Chen, W., 2012, "Quantification of Model Uncertainty: Calibration, Model Discrepancy, and Identifiability," Journal of Mechanical Design, accepted.

[3] Qian, P., Wu, H., and Wu, C., 2008, "Gaussian Process Models for Computer Experiments with Qualitative and Quantitative Factors," Technometrics, 50(3), pp. 383-396.

[4] McMillan, N., Sacks, J., Welch, W., and Gao, F., 1999, "Analysis of Protein Activity Data by Gaussian Stochastic Process Models," Journal of Biopharmaceutical Statistics, 9(1), pp. 145-160.

[5] Cressie, N., 1993, Statistics for Spatial Data, Wiley, New York.

[6] Ver Hoef, J., and Cressie, N., 1993, "Multivariable Spatial Prediction," Mathematical Geology, 25(2), pp. 219-240.

[7] Conti, S., Gosling, J. P., Oakley, J. E., and O'Hagan, A., 2009, "Gaussian Process Emulation of Dynamic Computer Codes," Biometrika, 96(3), pp. 663-676.

[8] Conti, S., and O'Hagan, A., 2010, "Bayesian Emulation of Complex Multi-Output and Dynamic Computer Models," Journal of Statistical Planning and Inference, 140(3), pp. 640-651.

[9] McFarland, J., Mahadevan, S., Romero, V., and Swiler, L., 2008, "Calibration and Uncertainty Analysis for Computer Simulations with Multivariate Output," AIAA Journal, 46(5), pp. 1253-1265.

[10] Bayarri, M. J., Berger, J. O., Cafeo, J., Garcia-Donato, G., Liu, F., Palomo, J., Parthasarathy, R. J., Paulo, R., Sacks, J., and Walsh, D., 2007, "Computer Model Validation with Functional Output," Annals of Statistics, 35(5), pp. 1874-1906.

[11] Williams, B., Higdon, D., Gattiker, J., Moore, L. M., McKay, M. D., and Keller-McNulty, S., 2006, "Combining Experimental Data and Computer Simulations, with an Application to Flyer Plate Experiments," Bayesian Analysis, 1(4), pp. 765-792.

[12] Drignei, D., 2009, "A Kriging Approach to the Analysis of Climate Model Experiments," Journal of Agricultural, Biological, and Environmental Statistics, 14(1), pp. 99-114.

[13] Apley, D., Liu, J., and Chen, W., 2006, "Understanding the Effects of Model Uncertainty in Robust Design with Computer Experiments," Journal of Mechanical Design, 128(4), pp. 945-958.

[14] Sacks, J., Welch, W. J., Mitchell, T. J., and Wynn, H. P., 1989, "Design and Analysis of Computer Experiments," Statistical Science, 4(4), pp. 409-423.

[15] Jin, R., 2004, "Enhancements of Metamodeling Techniques in Engineering Design," Ph.D., University of Illinois at Chicago, Chicago, Illinois.

[16] Kennedy, M. C., Anderson, C. W., Conti, S., and O'Hagan, A., 2006, "Case Studies in Gaussian Process Modelling of Computer Codes," Reliability Engineering and System Safety, 91(10-11), pp. 1301-1309.

[17] Rasmussen, C. E., 1996, "Evaluation of Gaussian Processes and Other Methods for Non-Linear Regression," Ph.D., University of Toronto, Toronto, ON, CA.

[18] Lancaster, T., 2004, An Introduction to Modern Bayesian Econometrics, Blackwell Publishing, Malden, MA.

[19] Johnson, R., and Wichern, D., 2007, Applied Multivariate Statistical Analysis, Prentice Hall, Upper Saddle River, N.J.

[20] Kennedy, M. C., and O'Hagan, A., 2000, "Supplementary Details on Bayesian Calibration of Computer Models," University of Sheffield, Sheffield, UK, http://www.isds.duke.edu/~fei/samsi/Oct_09/01Sup-KenOHa.pdf, (last accessed on June 27, 2012).

[21] Billingsley, P., 1995, Probability and Measure, John Wiley & Sons, Inc., New York, NY.