Quantification of Model Uncertainty: Calibration, Model Discrepancy, and Identifiability

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

Abstract

To use predictive models in engineering design of physical systems, one should first quantify the model uncertainty via model updating techniques employing both simulation and experimental data. While calibration is often used to tune unknown calibration parameters of the computer model, the addition of a discrepancy function has been used to capture model discrepancy due to underlying missing physics, numerical approximations, and other inaccuracies of the computer model that would exist even if all calibration parameters are known. One of the main challenges in model updating is the difficulty in distinguishing between the effects of calibration parameters versus model discrepancy. We illustrate this identifiability problem with several examples, explain the mechanisms behind it, and attempt to shed light on when a system may or may not be identifiable. In some instances, identifiability is achievable under mild assumptions, whereas in other instances it is virtually impossible. In a companion paper, we demonstrate that using

multiple responses, each of which depends on a common set of calibration parameters, can substantially enhance identifiability.

Keywords: calibration, identifiability, model updating, uncertainty quantification, Kriging, Gaussian processes

1 Introduction

Uncertainty is ubiquitous in engineering design. Although recent years have seen a proliferation of research in design under uncertainty [1-5], the majority of the uncertainty analysis techniques were developed for uncertainty propagation, i.e., *forward* propagation of uncertainty, to study the effect of *given* random inputs on the response of a system. On the other hand, the *inverse* problem in uncertainty analysis, in which experimental data are used to learn about sources of modeling uncertainty such as calibration parameters of a computer model and computer model discrepancy (aka inadequacy or bias), is a much harder problem. The "inverse problem" is receiving increased attention in the design community, because quantifying the uncertainty of a model and the resulting system response predictions (e.g., in the form of probabilistic prediction intervals) is essential for robust and reliable design decision making. The objective of this paper is to closely examine existing mathematical frameworks for model uncertainty quantification and to offer insight into the associated challenges. We argue that model uncertainty quantification as it is typically implemented in a model updating process, using only a single response variable, is challenging due to a lack of identifiability of calibration parameters in conjunction with the model discrepancy. Identifiability refers to whether the single true value of a model's calibration parameters can theoretically be inferred [6] based on the available data. In a companion paper [7], we develop an approach to improve model identifiability using multiple observed responses that all depend on some common set of

calibration parameters.

To present this work in the context of predictive modeling in engineering design, Figure 1 provides a general flowchart for model updating, validation, and refinement. The goal of model updating is to create an updated model that integrates the computer model simulations and experimental data to better predict in regions where experimental data have not been collected. We denote the controllable inputs (aka design variables) by the vector $\mathbf{x} = [x_1, \ldots, x_d]^T$ and the calibration parameters, which are unknown model parameters, by the vector $\boldsymbol{\theta} = [\theta_1, \ldots, \theta_r]^T$. A calibration parameter is defined as any physical parameter that can be specified as an input to the computer model and that is unknown or not measureable when conducting the physical experiments [8]. The iterative model updating process begins by simulating the computer model response, $y^m(\mathbf{x},\boldsymbol{\theta})$, at a set of N_m input sites $\mathbf{X}^m = [\mathbf{x}_1^m, \ldots, \mathbf{x}_{N_m}^m]^T$ and corresponding calibration parameter values $\boldsymbol{\Theta}^m = [\boldsymbol{\theta}_1^m, \ldots, \boldsymbol{\theta}_{N_m}^m]^T$, where each $\{\mathbf{x}_i^m, \boldsymbol{\theta}_i^m\}$ is a combination of input and calibration parameter settings at which the simulation is run. One then observes the physical experimental response, $y^e(\mathbf{x})$, at a set of N_e input sites $\mathbf{X}^e = [\mathbf{x}_1^e, \ldots, \mathbf{x}_{N_e}^e]^T$, often using design of experiments methods [9-11] to select the input sites. After collecting the data, an updated model is created using various formulations and methods of inference (although it can be subject to the identifiability issues that we examine in this work). A by-product of the model updating process is quantification of uncertainty in the calibration parameters, in the model discrepancy (via some discrepancy function $\delta(\mathbf{x})$), in the experimental response predictions, and in the experimental uncertainty ε. To assess the predictive capability of a model, we use model validation as a follow-up process that determines whether the updated model accurately reflects reality based on additional experimental response data and various validation metrics [12]. If the validation indicates the model is inadequate, the designer can either collect more data to further update the

model or refine the computer model by changing its physics [13].

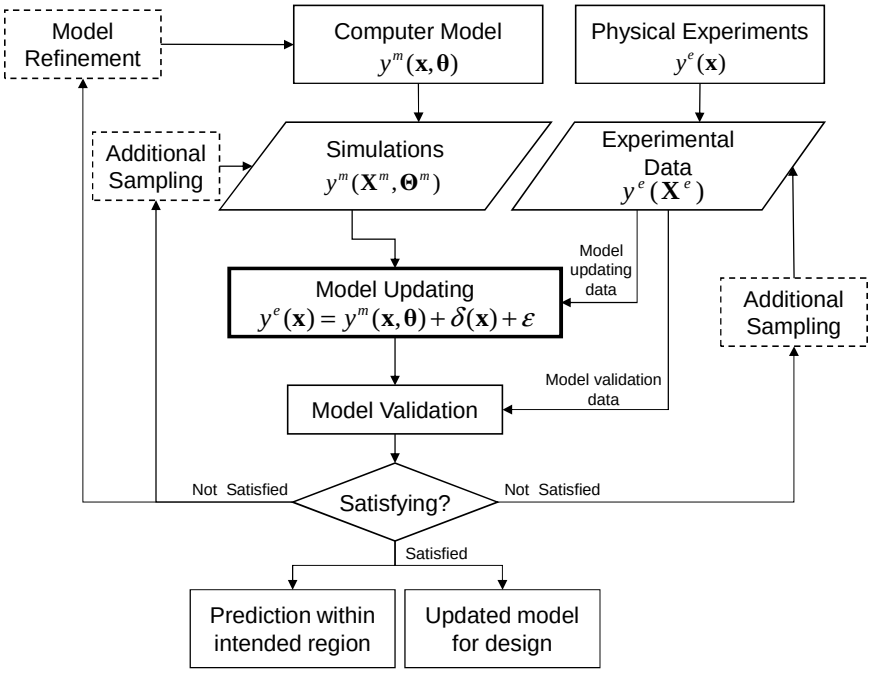

Fig. 1 Flowchart of model updating, model validation, and model refinement.

Within existing model updating techniques, the concept of calibration is frequently used to "tune" or "adjust" the calibration parameters $\boldsymbol{\theta}$, and/or an additive discrepancy function $\delta(\mathbf{x})$ is used to represent the model discrepancy. The majority of the approaches for model updating consider either calibration only [14, 15] or discrepancy function only [2, 16], and few consider both calibration and a discrepancy function [8, 13, 17-19] together, presumably because of the identifiability problem examined in this work.

The model updating formulation of Kennedy and O'Hagan [8] is one such approach that incorporates both calibration parameters and a discrepancy function, and we believe this to be the most applicable to design under uncertainty, although it is subject to the identifiability problems that we discuss in this paper. Most of the existing model updating techniques [8, 13, 17, 18] treat the process as a black-box and have the objective of improving the experimental response prediction but with little regard to identifying the true model calibration parameters and

model discrepancy. Loeppky et al. [19] explored the identifiability issue and concluded that accurate prediction of the experimental response is often possible, even if the individual calibration parameters and discrepancy function are not identified precisely. Furthermore, they concluded that the true values of the calibration parameters are only attainable when the model discrepancy between simulation and reality is neither present nor included in the model updating formulation. Higdon et al. [17] found that in some simple cases the calibration parameters and discrepancy function can be accurately identified, but in other cases their effects cannot be distinguished even though the response predictions may be reasonable.

In this paper we take a broader view that good identifiability is important in engineering applications for many reasons: (1) Learning the calibration parameters may in itself be a primary goal with broad-reaching implications for product/process improvement (e.g., if these calibration parameters are needed for a system-level simulation or if the calibration parameters themselves reflect the performance of interest but cannot be observed directly); (2) knowledge of the model discrepancy improves the understanding of the deficiencies of the computer model for improving future generations; and (3) it results in more accurate prediction over a broader set of input regions, because the model adjustment from learning the calibration parameters is more global than the adjustment from learning the model discrepancy. More specifically, adjusting the calibration parameters changes the prediction of the experimental response for a wide range of values for the design variables; whereas adjusting the discrepancy function tends to change the prediction of the experimental response predominantly within some neighborhood of the specific design variable settings that were used in the experiments.

The goal of this paper is to examine the model updating formulations and provide a better understanding of the issue of identifiability. In Section 2 we review and explore various sources

of uncertainty and various model updating formulations for combining computer models and physical experiments. Background on Gaussian process models to account for uncertainty due to lack of data (from either the computer model or physical experiments) is provided in Section 3. Details of the modular Bayesian procedure for calculating posterior distributions of the calibration parameters and the discrepancy function are provided in Section 4 [8]. Section 5 investigates the issue of identifiability using a simply supported beam example. Whereas the example and discussion in Section 5 convey the conclusion that identifiability is often very difficult, or impossible, in typical implementations, Section 6 discusses situations in which it is reasonable to expect that good identifiability can be achieved. Section 7 concludes the paper. In a companion paper [7], we further investigate the identifiability issue and demonstrate that incorporating multiple experimental responses that share a mutual dependence on a common set of calibration parameters can substantially improve identifiability.

2 Sources of Uncertainty and Model Updating Formulations

Following the seminal work of Kennedy and O'Hagan [8], we consider several different sources of uncertainty in model prediction: **Parameter uncertainty** derives from model calibration parameters that are inputs to the computer model but that are unknown and cannot be measured directly in physical experiments. An example is material damage parameters in finite element models. In this paper, we assume that the user has properly selected the set of calibration parameters. This should be based on prior knowledge of the computer model and perhaps parameter screening techniques, as discussed in ref. [18].

Parameter uncertainty in fact has two forms. A parameter may be constant but unknown and its uncertainty represented via a probability distribution. Alternatively, a parameter may vary randomly (from run to run over the physical experiments and/or over any future instances for

which the model is used to predict reality) according to some probability distribution. An example of the latter could be the blank thicknesses in a sheet metal stamping operation that vary randomly from blank to blank due to manufacturing variation. Calibration generally refers to the former form of parameter uncertainty, with the goal of identifying the true values of the constant parameters. By "constant" parameters, we mean those that do not change over the duration of the physical experiments. Accordingly, in this paper, we consider only this form of parameter uncertainty. Calibration in the case of parameters that vary randomly throughout the physical experiment would be far more challenging. One might assume some distribution for the randomly varying parameters and define the "calibration" goal as to identify the statistical parameters of the distribution (e.g., the mean and variance of a normal distribution) [13]. However, this would require far more experimental observations than are needed for the goal in this paper of identifying constant physical parameters.

Another source of uncertainty is computer model uncertainty (commonly termed **model discrepancy**, model bias, or model inadequacy), which results from underlying missing physics, numerical approximations, and/or other inaccuracies of the computer model that would exist even if all the parameters were known.

Another source of uncertainty introduced from using physical experiments is **experimental uncertainty** (aka observation error). This represents any variability in the experimental response that would occur if the experiment were repeated many times using the exact same settings for all of the design variables. This is often due to experimental measurement error.

Additional uncertainty results from having only a finite set of simulation and experimental data collected at discrete input settings, in which case one must interpolate or

extrapolate to predict the response at other input settings. This has been referred to as code uncertainty [12] and **interpolation uncertainty** [20]. For the computer model, this form of uncertainty would disappear if the simulations were inexpensive to run and one was able to simulate exhaustively over a densely spaced grid covering the entire input domain. However, in many instances computer models and experiments are expensive, and interpolation uncertainty is a major contribution to the uncertainty in prediction.

The three model updating formulations discussed below incorporate the various uncertainties outlined when combining a computer model and physical experiments. In all three model updating formulations, we assume that the design does not change so radically that the underlying variables change. Rather, we assume that the design space is represented by a space of changing variables **x** over a single design structure. For example, if the objective is to design a component using finite element analysis (FEA) by manipulating geometric design variables, where the calibration parameters dictate the material behavior of the component, then the different designs are associated with different settings of the design variables, while the underlying material behavior does not change. We note that there are other model updating formulations that do not consider a set of design variables **x**. However, we do not review them because they are not typically regarded as design under uncertainty.

The first model updating formulation considers a discrepancy function with no calibration parameters [21, 22]. A Bayesian approach is considered in references [2, 16, 23], which model the system as

$$y^e(\mathbf{x}) = y^m(\mathbf{x}) + \delta(\mathbf{x}) + \varepsilon \,, \tag{1}$$

where $y^e(\mathbf{x})$ is the experimental response, $y^m(\mathbf{x})$ denotes the computer model response, $\delta(\mathbf{x})$ is the additive discrepancy function, and ε accounts for the experimental uncertainty. Experimental

uncertainty is typically assumed to follow a normal distribution with mean 0 and variance λ, denoted $\varepsilon \sim N(0,\lambda)$. The advantage of considering only a discrepancy function is that a Bayesian closed form solution exists, since the discrepancy function is nearly directly observable from the collected simulations and experimental data. The obvious limitation of the formulation in Eq. (1) is its inability to account for the effects of unknown calibration parameters.

A second formulation incorporates calibration parameters with no discrepancy function. This formulation assumes the model

$$y^e(\mathbf{x}) = y^m(\mathbf{x}, \boldsymbol{\theta}^*) + \varepsilon \tag{2}$$

where $y^m(\mathbf{x},\boldsymbol{\theta})$ is the computer model response as a function of $\mathbf{x}$ and $\boldsymbol{\theta}$, and $\boldsymbol{\theta}^*$ (a $r \times 1$ vector of constants) denotes the true values of the unknown calibration parameters over the course of the physical experiments. Various methods, such as nonlinear least squares, can be used to estimate $\boldsymbol{\theta}^*$ [14], although this is not trivial if the input ($\mathbf{x}$) values are not the same over the physical experimental and simulation runs. For this reason, many instances of this approach [15, 24] only consider calibration parameters and omit $\mathbf{x}$ when fitting the model.

Fig. 2(a) shows a hypothetical example of $y^m(x,\theta)$ for different values of θ, together with $y^e(x)$ with no experimental uncertainty or model discrepancy (i.e., $y^m(x,\theta^*)$). With no model discrepancy, there exists a value of θ such that the computer model response is in perfect agreement with the physical response, and this value is precisely θ^*. In contrast, for the simple case of $\varepsilon = 0$, Fig. 2(b) illustrates the problem with not including a discrepancy function in the model when model discrepancy truly exists. In this case, there is no value of θ for which the computer model response will be in perfect agreement with the experimental response. However, inclusion of a discrepancy function will allow the model of Eq. (2) to agree with the experimental response.

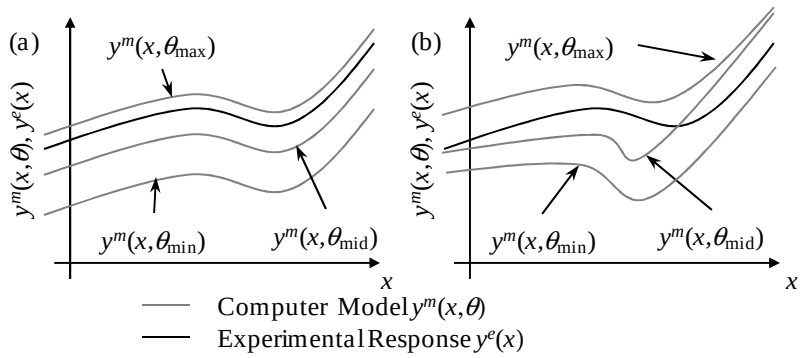

Fig. 2 Example of a response considering a calibration parameter (a) without and (b) with the need for a discrepancy function.

Previous work [8, 17, 18, 25] has combined the two approaches above into one model updating formulation represented by

$$y^e(\mathbf{x}) = y^m(\mathbf{x}, \boldsymbol{\theta}^*) + \delta(\mathbf{x}) + \varepsilon \,. \tag{3}$$

Notice that the discrepancy function $\delta(\mathbf{x})$ in Eq. (3) is written as a function of $\mathbf{x}$, which is only a consequence of the fundamental difference between $\mathbf{x}$ and $\boldsymbol{\theta}$. The design inputs $\mathbf{x}$ are variables that can be controlled and set by the user during the experiments; whereas, the calibration parameters $\boldsymbol{\theta}$ take on some constant but unknown true value $\boldsymbol{\theta}^*$ over the duration of all the experiments and cannot be controlled (or even directly observed) by the user. Hence, the discrepancy function is defined as the difference between the experimental response $y^e(\mathbf{x})$ and the computer model response $y^m(\mathbf{x},\boldsymbol{\theta}^*)$ at the true parameters $\boldsymbol{\theta}^*$. Although we could have written the discrepancy function as a function of $\boldsymbol{\theta}^*$, we suppress this notation because $\boldsymbol{\theta}^*$ is a vector of constants. We also note that even though $\mathbf{x}$ and $\boldsymbol{\theta}$ are fundamentally different in the physical experiments, in the computer simulations they are both treated as inputs that must be specified in order to run the simulations.

As illustrated in Fig. 3, the formulation in Eq. (3) addresses parameter uncertainty, model discrepancy, experimental uncertainty, and interpolation uncertainty. A quantification of

the latter is a by-product of the Gaussian process metamodeling approach discussed in the next section. To quantify uncertainty in the calibration parameters and discrepancy function, the modular Bayesian approach presented in Section 4 is used to calculate their posterior distributions. Notice that the discrepancy function in Eq. (3) is not directly observable from the collected data, since the true value of $\boldsymbol{\theta}$ is unknown. Because Eq. (3) accounts for several different forms of uncertainty and also includes $\mathbf{x}$, we view it as providing a comprehensive and widely applicable model updating formulation for design under uncertainty.

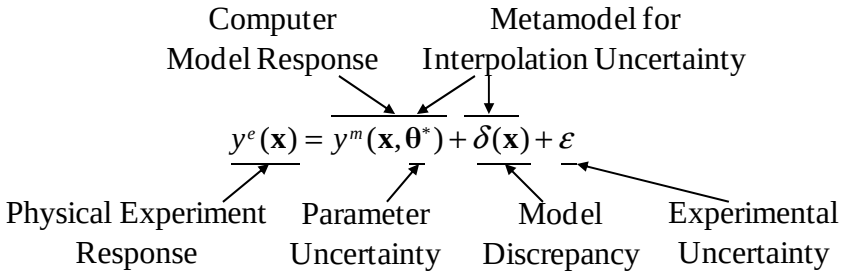

Fig. 3 Sources of uncertainty with calibration parameters and a discrepancy function.

3 Gaussian Processes Models for Interpolation

In this section we review Gaussian process (GP) modeling approaches for quantifying interpolation uncertainty [8, 11, 26-28], which is particularly important when data are limited. GP models provide a prediction that interpolates perfectly, in the sense that a fitted GP passes exactly through every observation, which is desirable for deterministic simulations; and it has a natural mechanism (via including a "nugget effect" or error term) for forgoing perfect interpolation if the response observations are noisy. The approach in this paper can be readily extended to this situation, although we restrict attention to the case of noiseless computer simulations. Additionally, inherent to GP modeling is a tractable and intuitively reasonable mechanism for quantifying the uncertainty of the metamodel response prediction at input sites where no data has been collected. Lastly, being a probabilistic model, GPs are well-suited for

combining data from different stages, with or without incorporating prior knowledge from the analyst.

In the GP approach, one assumes the response surface $y(\mathbf{x})$ (here we omit $\boldsymbol{\theta}$ for notational convenience, as it is treated no differently than the inputs $\mathbf{x}$ in the simulations) is a single realization of a spatial random process (assumed Gaussian for convenience) with prior mean function $\mathbf{h}(\mathbf{x})\boldsymbol{\beta}$, where $\mathbf{h}(\mathbf{x})$ is a row vector of regression functions (i.e., constant, linear, etc.) and $\boldsymbol{\beta}$ is a column vector of coefficients, and with covariance function $\sigma^2 R(\mathbf{x},\mathbf{x}')$, where σ^2 is a constant and $R(\mathbf{x},\mathbf{x}')$ is the correlation function between the responses at points $\mathbf{x}$ and $\mathbf{x}'$. These assumptions imply that for any set of input sites $\mathbf{X} = [\mathbf{x}_1, \ldots, \mathbf{x}_N]^T$, the vector $\mathbf{y} = [y(\mathbf{x}_1), \ldots, y(\mathbf{x}_N)]^T$ is multivariate normal with mean vector $\mathbf{H}\boldsymbol{\beta}$, where $\mathbf{H} = [\mathbf{h}^T(\mathbf{x}_1), \ldots, \mathbf{h}^T(\mathbf{x}_N)]^T$ (a $N \times 1$ vector of ones if the prior mean is a constant that does not depend on $\mathbf{x}$), and covariance matrix $\sigma^2\mathbf{R}$, where $\mathbf{R}$ is the $N \times N$ matrix whose ith row, jth column element is $R(\mathbf{x}_i,\mathbf{x}_j)$. One popular choice for $R(\bullet,\bullet)$ is the Gaussian correlation function

$$R(\mathbf{x},\mathbf{x}') = \exp\left\{-\sum_{k=1}^{d} \omega_k (x_k - x_k')^2\right\} \tag{4}$$

where $\boldsymbol{\omega} = [\omega_1, \ldots, \omega_d]^T$ is a vector of roughness parameters that represent the rate at which the correlation between $y(\mathbf{x})$ and $y(\mathbf{x}')$ decays to zero as $\mathbf{x}$ and $\mathbf{x}'$ diverge. The hyperparameters $\boldsymbol{\beta}$, σ^2, and $\boldsymbol{\omega}$ characterize the GP model. Fig. 4(a) depicts the prior distribution (prior to observing any response values) of the response function for a GP with $d = 1$, a constant prior mean of 0, and a Gaussian covariance function with $\sigma = 1.06$, and $\boldsymbol{\omega} = 0.67$. The grey shaded region is a 95% prediction interval (PI) for $y(\mathbf{x})$, defined as $E[y(x)] \pm 1.96\sqrt{Var(y(x))}$.

After observing $\mathbf{y}$, at any point $\mathbf{x}$ the posterior distribution for the response $y(\mathbf{x})$ given $\mathbf{y}$ (and given $\boldsymbol{\omega}$ and σ and assuming a noninformative prior for $\boldsymbol{\beta}$) is Gaussian with mean and covariance [29]

$$E[y(\mathbf{x}) \mid \mathbf{y}] = \mathbf{h}(\mathbf{x})\hat{\boldsymbol{\beta}} + \mathbf{r}^T(\mathbf{x})\mathbf{R}^{-1}(\mathbf{y} - \mathbf{H}\hat{\boldsymbol{\beta}}) \tag{5}$$

and

$$\begin{aligned} Cov[y(\mathbf{x}), y(\mathbf{x}') \mid \mathbf{y}] = \sigma^2 \{ & R(\mathbf{x}, \mathbf{x}') - \mathbf{r}^T(\mathbf{x})\mathbf{R}^{-1}\mathbf{r}(\mathbf{x}') + \\ & [\mathbf{h}^T(\mathbf{x}) - \mathbf{H}^T\mathbf{R}^{-1}\mathbf{r}(\mathbf{x})]^T [\mathbf{H}^T\mathbf{R}^{-1}\mathbf{H}]^{-1} [\mathbf{h}^T(\mathbf{x}') - \mathbf{H}^T\mathbf{R}^{-1}\mathbf{r}(\mathbf{x}')] \} \end{aligned} \tag{6}$$

where $\mathbf{r}(\mathbf{x})$ is a $N \times 1$ vector whose ith element is $R(\mathbf{x},\mathbf{x}_i)$, and $\hat{\boldsymbol{\beta}} = [\mathbf{H}^T\mathbf{R}^{-1}\mathbf{H}]^{-1}\mathbf{H}^T\mathbf{R}^{-1}\mathbf{y}$. From a non-Bayesian perspective, Eq. (5) is the best linear unbiased predictor for $y(\mathbf{x})$, and its variance given by Eq. (6) with $\mathbf{x} = \mathbf{x}'$. One can easily show that when $\mathbf{x}$ coincides with one $\mathbf{x}_i$, Eq. (5) reduces to $y(\mathbf{x}_i)$ and the variance of Eq. (6) is 0 (i.e., exact interpolation of each input site). The preceding treats the hyperparameters σ^2 and $\boldsymbol{\omega}$ as known.

If these hyperparameters are unknown, two methods are available to estimate them from the collected data. In practice, for computational reasons, one usually calculates the maximum likelihood estimates (MLEs) of σ^2 and $\boldsymbol{\omega}$, and then subsequently treats them as known in Eqs. (5) and (6) [30]. In contrast, a Bayesian approach calculates an entire posterior distribution for the hyperparameters, and is typically implemented by Markov Chain Monte Carlo (MCMC) [17]. Furthermore, in the Bayesian approach Eqs. (5) and (6) must be integrated with respect to the posterior distributions. This is computationally expensive and requires expert knowledge to tune the MCMC parameters to obtain accurate posterior distributions [31].

Fig. 4 depicts the posterior distributions of the response, with the assigned hyperparameters (σ = 1.06, and $\boldsymbol{\omega}$ = 0.67), after collecting two response observations (panel (b)) and then a third response observation (panel (c)). The uncertainty in predicting $y(x)$, in the form of a 95% PI, is depicted as the gray shaded regions, the width of which increases as x strays from the input sites (black dots). This simple example illustrates an attractive characteristic of the GP model, namely the ability to provide a reasonable predicted response surface that interpolates the

observed points, while simultaneously providing a reasonable quantification of prediction uncertainty that shrinks to zero at the observed points.

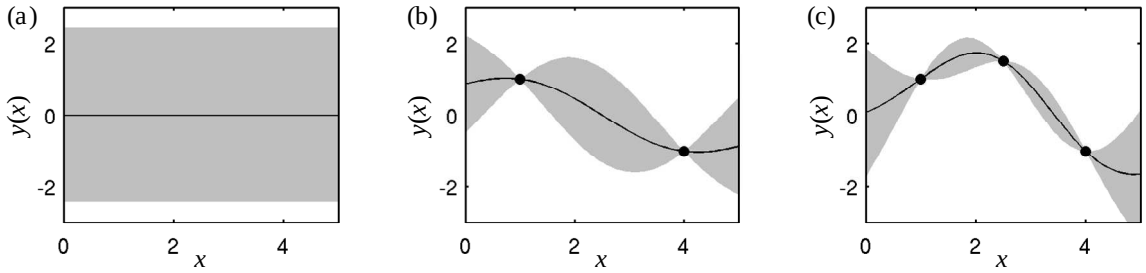

Fig. 4 Depiction of (a) prior distribution for a GP model with constant mean 0 and constant variance; and (b) and (c) posterior distributions for the same GP model after collecting response observations (bullets). The solid black lines are the predicted mean from Eq. (5), and the shaded regions are 95 % prediction intervals.

4 The Modular Bayesian Approach

In this section, we review the "modular" Bayesian approach to simultaneously provide posterior distributions for the calibration parameters, the discrepancy function, and the experimental response. This approach is part of the comprehensive model updating formulation based on the model in Eq. (3) [8, 18]. GP models are created for both the computer model and the discrepancy function to interpolate the limited data from the computer model and physical experiments.

While a fully Bayesian treatment of the computer model and the discrepancy function estimates all of the GP model hyperparameters simultaneously, a modular Bayesian approach, recommended in this paper, separates the estimation of the GP model hyperparameters for both the computer model and the discrepancy into two modules (Modules 1 and 2 of Fig. 5). The first module essentially replaces the computer model with a GP model and estimates the GP model hyperparameters based on only the simulation data. In the second module, the discrepancy

function is modeled by fitting another GP model based on the simulation data, the experimental data, and the prior for the calibration parameters. This separation of the two GP models is intuitive to most modelers, and thus the modular Bayesian approach has an advantage over a fully Bayesian approach [8, 31].

In both the fully Bayesian and the modular Bayesian approaches, estimates of the hyperparameters for both GP models can be obtained via MLE or via Bayesian posterior distributions, as described in Section 3. As stated previously, Bayesian posterior distributions can be computationally prohibitive. In this paper, we implement the modular Bayesian approach using MLEs for the hyperparameters. This approach decreases the computational cost by using only point estimates for the hyperparameters instead of full posterior distributions [8]. Moreover, ref. [18] found that both approaches provided similar results for the prediction of the experimental response, the prediction of the discrepancy function, and the posterior distribution of the calibration parameters.

In Modules 3 and 4 of Fig. 5, the posterior distributions of the calibration parameters, the discrepancy function, and the experimental response are calculated using the given data and the estimates of the hyperparameters from the modular Bayesian approach or the fully Bayesian approach. Details of each module are provided next.

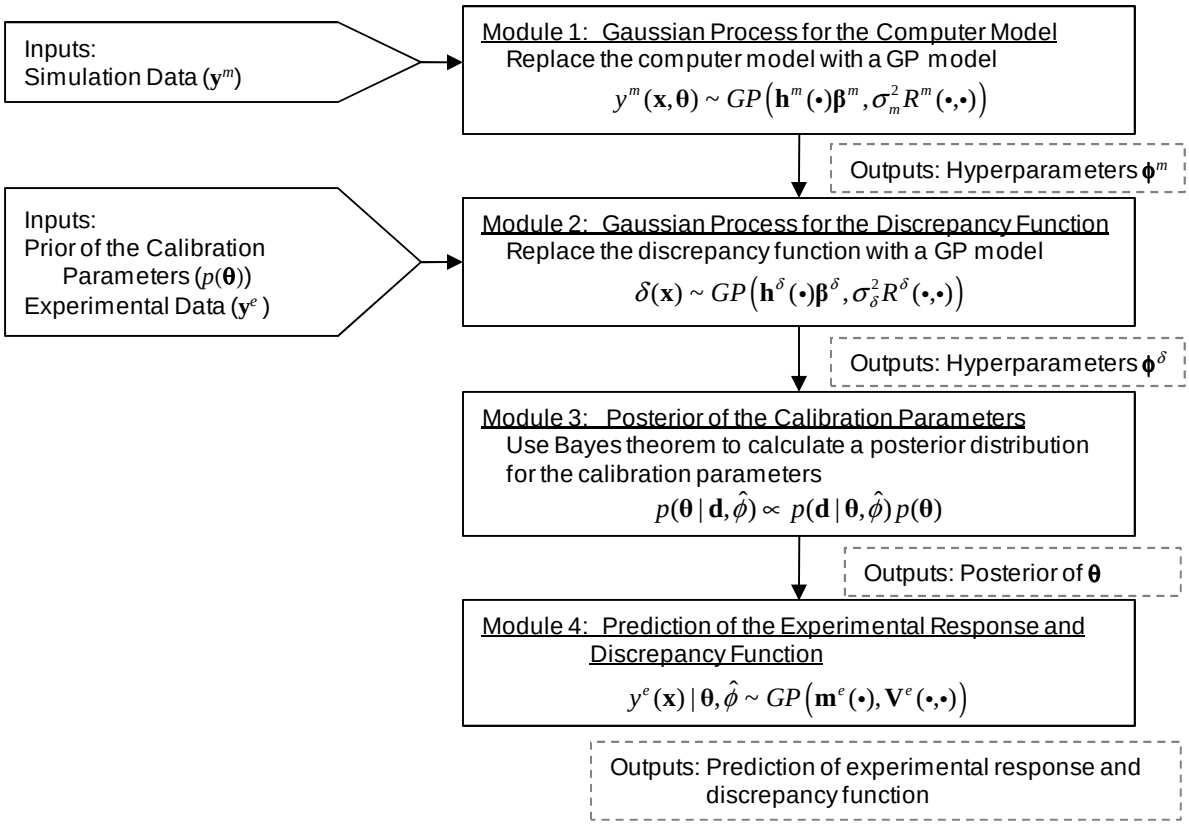

Fig. 5 Flowchart of the modular Bayesian approach.

4.1 Gaussian Process Model for the Computer Model (Module 1)

A GP model is used to infer the computer model response between the collected simulation data points. The prior of the computer model GP is defined as

$$y^m(\bullet,\bullet) \sim GP\left(\mathbf{h}^m(\bullet,\bullet)\boldsymbol{\beta}^m, \sigma_m^2 R^m((\bullet,\bullet),(\bullet,\bullet))\right). \tag{7}$$

The prior mean function is comprised of the unknown regression coefficients $\boldsymbol{\beta}^m$ and the known regression functions $\mathbf{h}^m(\mathbf{x},\boldsymbol{\theta})$ (for the examples in this paper, we use a constant $\mathbf{h}^m(\mathbf{x},\boldsymbol{\theta}) = 1$, which corresponds to a constant prior mean). The prior covariance function is the product of an unknown constant σ_m^2 and a 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. We use the Gaussian correlation function of Eq. (4), which, with the additional inputs $\boldsymbol{\theta}$, is parameterized by a $(d + r) \times 1$ vector $\boldsymbol{\omega}^m$.

To estimate the hyperparameters $\boldsymbol{\phi}^m = [\boldsymbol{\beta}^m, \sigma_m^2, \boldsymbol{\omega}^m]$ of the GP model, we maximize the multivariate normal log-likelihood function for the simulation data $\mathbf{y}^m = [y^m(\mathbf{x}_1^m, \boldsymbol{\theta}_1^m), \ldots, y^m(\mathbf{x}_{N_m}^m, \boldsymbol{\theta}_{N_m}^m)]^T$ collected at N_m input sites $\mathbf{X}^m$ and $\boldsymbol{\Theta}^m$. Previous literature [11, 32, 33] suggests a space-filling experimental design for the input settings $\mathbf{X}^m$ and $\boldsymbol{\Theta}^m$. To improve the numerical stability of the MLE algorithm, the inputs $\mathbf{X}^m$ and $\boldsymbol{\Theta}^m$ can be transformed to the range 0 to 1 and the simulation data $\mathbf{y}^m$ can be standardized to have a sample mean of 0 and a sample standard deviation of 1 [34].

To assess the predictive capability of the GP model, the MLEs can be plugged into Eq. (7), which in turn can be plugged into Eqs. (5) and (6), to calculate the posterior distribution of $y^m(\mathbf{x},\boldsymbol{\theta})$ at any $\mathbf{x}$ and $\boldsymbol{\theta}$. After obtaining estimates of the hyperparameters of the GP model for the computer model (Module 1), the next step is to estimate the hyperparameters of the GP model of the discrepancy function.

4.2 Gaussian Process Model for the Discrepancy Function (Module 2)

The GP model for the experimental response $y^e(\mathbf{x})$ is the sum of the GP models for the discrepancy function $\delta(\mathbf{x})$ and for the computer model $y^m(\mathbf{x},\boldsymbol{\theta})$. This GP model will then be used to infer the experimental response at any $\mathbf{x}$.

To define a GP model for the experimental response, we must first define the GP model for the discrepancy function. The prior for the discrepancy function GP model is

$$\delta(\bullet) \sim GP(\mathbf{h}^\delta(\bullet)\boldsymbol{\beta}^\delta, \sigma_\delta^2 R^\delta(\bullet,\bullet)). \tag{8}$$

The mean function is comprised of the unknown regression coefficients $\boldsymbol{\beta}^\delta$ and the known regression functions $\mathbf{h}^\delta(\mathbf{x})$ (we use the constant $\mathbf{h}^\delta(\mathbf{x}) = 1$). The covariance function is comprised of the unknown constant σ_δ^2 and the correlation function $R^\delta(\mathbf{x},\mathbf{x}')$, which for the Gaussian correlation of Eq. (4), is parameterized by the $d \times 1$ vector $\boldsymbol{\omega}^\delta$. Following the work of Kennedy

and O'Hagan [8], we assume the computer model, the discrepancy function, and the experimental uncertainty are a priori statistically independent (i.e. $Cov[y^m(\mathbf{x},\boldsymbol{\theta}), \delta(\mathbf{x}')] = 0$), $Cov[y^m(\mathbf{x},\boldsymbol{\theta}),\varepsilon] = 0$, and $Cov[\delta(\mathbf{x}), \varepsilon] = 0$ for all $\mathbf{x}$, $\mathbf{x}'$, and $\boldsymbol{\theta}$). These assumptions of a priori independence simplify many of the calculations required in the 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 GP [Eq. (7)] and the discrepancy function GP [Eq. (8)] are combined to form the prior for experimental response GP as

$$\begin{aligned} & y^e(\bullet) \mid \boldsymbol{\theta} \sim GP\left(m^e(\bullet,\bullet), V^e((\bullet,\bullet),(\bullet,\bullet))\right) \\ & m^e(\mathbf{x},\boldsymbol{\theta}) = \mathbf{h}^m(\mathbf{x},\boldsymbol{\theta})\boldsymbol{\beta}^m + \mathbf{h}^\delta(\mathbf{x})\boldsymbol{\beta}^\delta \\ & V^e((\mathbf{x},\boldsymbol{\theta}),(\mathbf{x}',\boldsymbol{\theta}')) = \sigma_m^2 R^m((\mathbf{x},\boldsymbol{\theta}),(\mathbf{x}',\boldsymbol{\theta}')) + \sigma_\delta^2 R^\delta(\mathbf{x},\mathbf{x}') + \lambda \end{aligned} \tag{9}$$

Since Eq. (9) depends on the calibration parameters $\boldsymbol{\theta}$, the problem now is to estimate the hyperparameters of the discrepancy function GP without knowing the true values of $\boldsymbol{\theta}$.

Kennedy and O'Hagan [8] developed a procedure to obtain MLEs of the hyperparameters $\boldsymbol{\phi}^\delta = [\boldsymbol{\beta}^\delta, \sigma_\delta^2, \lambda, \boldsymbol{\omega}^\delta]$ using only the prior distribution for the calibration parameters. Specifically, one considers the likelihood function for the simulation data and the experimental data, where the likelihood function is based on the estimates of the hyperparameters from Module 1 and the prior for the calibration parameters. Kennedy and O'Hagan [35] provide closed form solutions for the likelihood function under the conditions of Gaussian correlation functions, constant regression functions, and normal prior distributions for $\boldsymbol{\theta}$ (for details refer to Section 3 of [35] and Section 4.5 of [8]). In this work, we use a similar likelihood function with the exception that the prior distribution of $\boldsymbol{\theta}$ is uniform. For this likelihood function, the experimental data is obtained at N_e input sites, $\mathbf{X}^e$ (transformed to the range [0,1]) with response $\mathbf{y}^e = [y^e(\mathbf{x}_1), \ldots,$

$y^e(\mathbf{x}_{N_e})]^T$ (standardized). After estimating the hyperparameters, Bayes theorem is used to calculate the posterior distribution of $\boldsymbol{\theta}$.

If there is variability in the design variables (e.g., due to manufacturing variability), as long as one can measure $\mathbf{X}^e$ accurately during the physical experiments, then the modular Bayesian approach described in this section can still be applied directly to construct an updated model of the experimental response. One could then use the updated model to calculate predictions of the physical response in a manner that incorporates the variability of the design variables [8].

4.3 Posterior of the Calibration Parameters (Module 3)

This module calculates the posterior distribution of the calibration parameters based on the simulations, the experimental data, and the estimated hyperparameters from Modules 1 and 2. The posterior distribution is

$$p(\boldsymbol{\theta} \mid \mathbf{d}, \hat{\boldsymbol{\phi}}) \propto p(\mathbf{d} \mid \boldsymbol{\theta}, \hat{\boldsymbol{\phi}})\, p(\boldsymbol{\theta})\,. \tag{10}$$

where $\hat{\boldsymbol{\phi}}$ are the MLEs of $\boldsymbol{\phi} = [\boldsymbol{\phi}^m\ \boldsymbol{\phi}^\delta]$, $\mathbf{d}$ denotes all response observations $[\mathbf{y}^{mT}, \mathbf{y}^{eT}]^T$, and $p(\boldsymbol{\theta})$ is the prior of the calibration parameters. The likelihood function $p(\mathbf{d} \mid \boldsymbol{\theta}, \hat{\boldsymbol{\phi}})$ is multivariate normal. Its value is determined by the specified mean and covariance functions for both the computer model [Eq. (7)] and the experimental response [Eq. (9)] GP models. Section 4.4 of [8] provides detailed equations for the likelihood function. The posterior distribution of $\boldsymbol{\theta}$ and the hyperparameter estimates from Modules 1 and 2 influence the prediction of the experimental response in the next and final module.

4.4 Prediction of the Experimental Response and Discrepancy Function (Module 4)

After collecting the simulation and experimental data and estimating the hyperparameters in Modules 1 and 2, given a specific value of $\boldsymbol{\theta}$, the conditional posterior distribution of the

experimental response can be calculated at any point **x**, with mean and covariance similar to Eqs. (5) and (6), (see Section 4.6 of ref. [8] for detailed equations). The unconditional distribution of the experimental response is then obtained by marginalizing the conditional distribution with respect to the posterior distribution of the calibration parameters from Module 3. This unconditional distribution's mean and variance can be calculated by using the law of total expectation and the law of total variance. Thus, the marginalized posterior distribution of the experimental response accounts for parameter uncertainty, model discrepancy, interpolation uncertainty, and experimental uncertainty. Since the discrepancy function is also a GP model, its posterior distribution can be calculated in a similar manner.

5 Lack of Identifiability: An Illustrative Simply Supported Beam Example

The model updating formulation of Eq. (3) is comprehensive in that it accounts for many sources of uncertainty; however, one limitation is that for some systems it is very difficult to distinguish between the effects of parameter uncertainty and model discrepancy. More specifically, the same predicted experimental response can result from many different combinations of the calibration parameters and the discrepancy function. This constitutes a lack of identifiability [19, 36] of the calibration parameters and discrepancy function, even though the experimental response may be accurately predicted.

To further illustrate this identifiability problem, we present an example using a simply supported beam shown in Fig. 6(a). The beam has length (2 m) and a rectangular cross-section with a height of 52.5 mm and width of 20 mm. One end of the beam is fixed (no vertical or horizontal displacement permitted), while the other end of the beam is supported by a roller (no vertical displacement permitted). A static force is applied to the midpoint of the beam to induce various responses (e.g. stress and displacement). In this example, the magnitude of the applied

force is chosen as the design variable **x**. Refer to Table 1 for the notation of this example.

The computer model response $y^m(\mathbf{x},\boldsymbol{\theta})$ (angle of deflection at the end of the beam in radians) is obtained from a finite element analysis (FEA) model representing a simply supported beam (using Abaqus 6.9) with a simplified (perfectly plastic) material law. For illustrative purposes, we will take the "experimental response" $y^e(\mathbf{x})$ (angle of deflection at the end of the beam in radians) to be the response from a FEA model using a more elaborate material law (a power law for the plastic region with constant stress C = 2068 MPa and strain-hardening exponent n = 0.5) [37]. These material laws are in part governed by Young's modulus, E, which is treated as the unknown calibration parameter θ (Fig. 7). For the "physical experiments", the true value of the calibration parameter is set to θ^* = 206.8 GPa but is treated as unknown during the analysis and assigned a uniform prior over the range $150 \leq \theta \leq 300$ GPa. The prior distribution was chosen to be relatively wide in order to minimize the effect of the prior on the posterior distributions and to avoid choosing a prior that does not contain any support at the true value θ^* (in which case the posterior will not contain any support at θ^* either). In practice, we recommend choosing a conservatively wide prior in order to ensure that there is some support at the true value of the parameters.

The goal is to use the observed simulation and experimental data to infer the calibration parameter and the discrepancy function of the model updating formulation of Eq. (3). It should be noted that this example is intended only to illustrate the modular Bayesian approach and the identifiability challenges. There are many standard experimental methods to determine Young's modulus using specimens of the material.

Table 1 Notation for the simply supported beam example.

Variable	Description
x	Static force (N) applied to the midpoint of the beam (on the range [1300 2300] N)
θ	Young's Modulus (GPa), on which both the simulation model and physical reality depend (on the range [150 300] GPa), $\theta^* = 206.8$ GPa is the true value
y	The angle of deflection (radians) at the end of the beam
$y^m(x, \theta)$	y calculated via a finite element analysis (FEA) with a simplified material model
$y^e(x)$	"Experimentally measured" y (for illustrative purposes, calculated from a FEA model with a more elaborate material model assuming $\theta = \theta^*$)
$\delta(x)$	Discrepancy function
ε	Random error, assumed 0 in this example

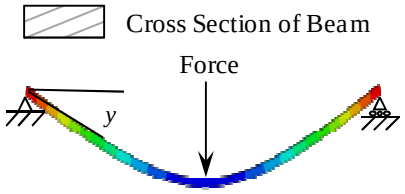

Fig. 6 Schematic of the simply supported beam.

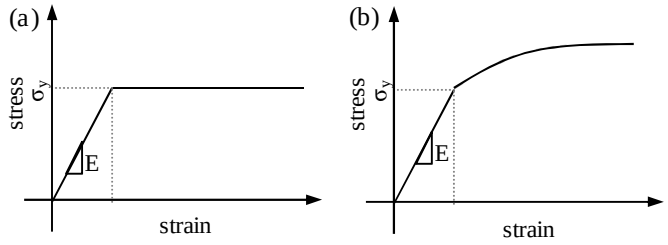

Fig. 7 Material stress-strain curves for (a) the computer model and (b) the "physical experiments". E is Young's modulus and the calibration parameter. σ_y is the yield stress (225 MPa).

To infer the calibration parameter and the discrepancy function, we use the modular Bayesian approach discussed in Section 4. We begin by fitting a GP model of the computer model (Module 1) 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 $\mathbf{y}^e$ is observed at $\mathbf{X}^e = [1300, 1400, 1500, \ldots, 2300]^T$ and indicated in Fig. 8(a) by the black dots. The experimental data, together with the uniform prior for θ and

the hyperparameter MLEs from Module 1, are used to estimate the hyperparameters of the GP model for the discrepancy function in Module 2. Finally, Modules 3 and 4 calculate, via Legendre-Gauss quadrature, the posterior distributions for the discrepancy function and the calibration parameter. The posterior distributions are shown in Fig. 8. From Fig. 8(a), the posterior distribution of the experimental response is very accurate and precise, because there was a relatively large amount of experimental data and no experimental error ε. In spite of this, the calibration parameter and the discrepancy function are not identifiable as illustrated below.

To assess identifiability, we use the posterior distributions of the calibration parameter (Fig. 8(c)) and the discrepancy function (Fig. 8(b)). The large posterior variance of the calibration parameter and the large width of the prediction intervals for the discrepancy function indicate the poor identifiability of the calibration parameter and discrepancy function. Indeed, several different combinations of these are approximately equally probable and give approximately the same (quite accurate and precise) predicted response in Figure 8(a). It is also worth noting neither posterior reflects the true values of the calibration parameter and the discrepancy function. Because of the high uncertainty reflected by the posterior distributions, the system is not identifiable, even with the relatively dense spacing of the experimental data. Moreover, collecting additional experimental observations would help little.

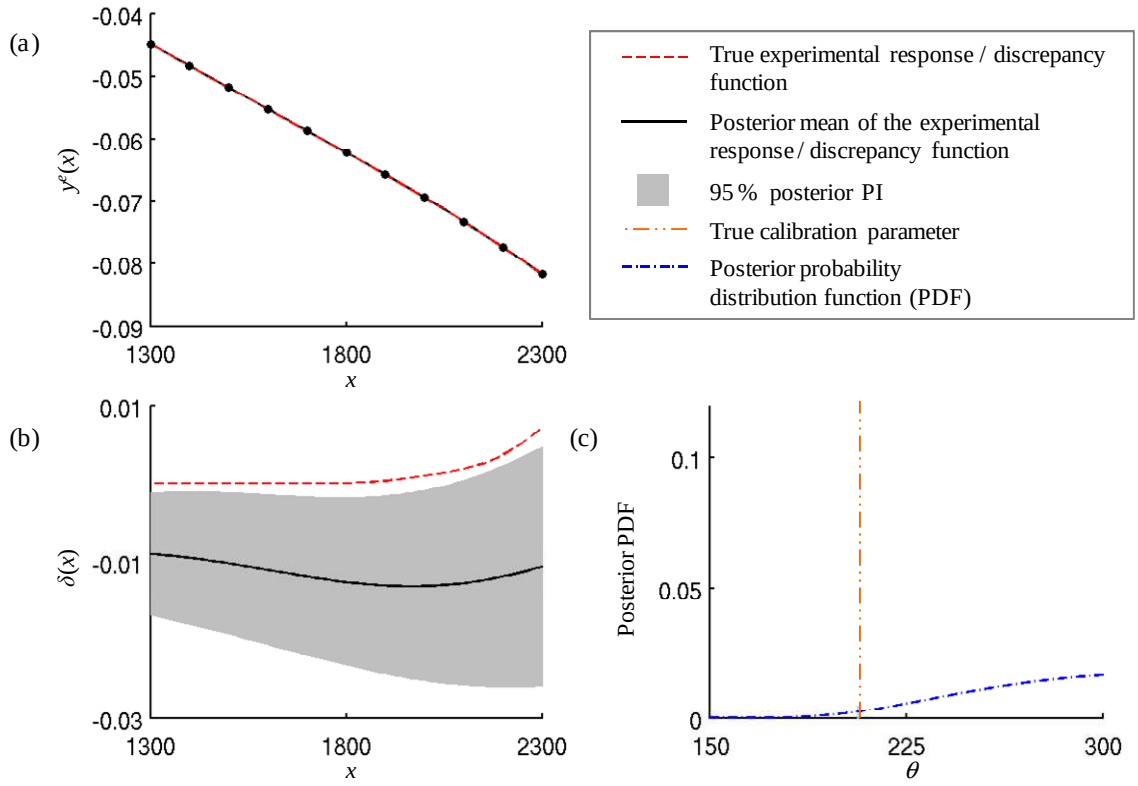

Fig. 8 The posterior distributions for (a) the experimental response, (b) the discrepancy function, and (c) the calibration parameter showing a lack of identifiability.

To understand how the large uncertainty and the inaccuracy of the posterior distributions lead to a lack of identifiability, consider the estimated discrepancy function $\hat{\delta}(x,\theta) = y^e(x) - y^m(x,\theta)$ shown in Fig. 9 for two different values of θ. For θ underestimated at 150 GPa in Fig. 9(a) (the true θ^* = 206.8 GPa) and for θ overestimated at 250 GPa in Fig. 9(b), neither θ nor the resulting estimated discrepancy function are inconsistent with their priors or in disagreement with the data. They both result in very similar, and quite accurate, prediction of the experimental response. Consequently, without additional information, it is virtually impossible to distinguish between the effects of θ and $\delta(x)$ and accurately identify each in this example.

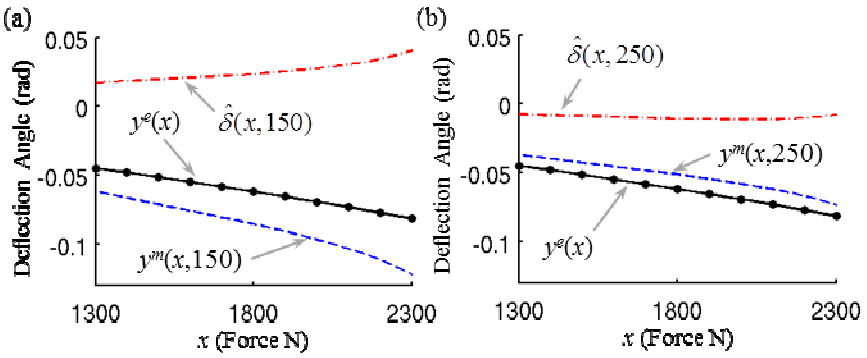

Fig. 9 The computer response $y^m(x,\theta)$, the experimental response $y^e(x)$ (bullets indicate experimental data), and the estimated discrepancy function $\hat{\delta}(x,\theta) = y^e(x) - y^m(x,\theta)$ for (a) θ = 150 GPa and (b) θ = 250 GPa.

To improve identifiability, previous literature [18, 31] recommended using informative (i.e, low variance or tightly dispersed) prior distributions for the calibration parameters, the discrepancy function, or both. Specifying informative priors for the discrepancy function involves assigning a specific functional form (e.g. linear or quadratic [13, 38]); however, assigning this form is often difficult because one rarely has significant prior knowledge.

Likewise, one does not have such prior information on the calibration parameters (otherwise, they would be viewed as known parameters and not considered calibration parameters). However, to illustrate how this would improve identifiability, we consider three versions of the preceding beam example using three different normal prior distributions for θ, each with different means and/or variances. The resulting posterior distributions for θ and the discrepancy function are shown in Fig. 10 and Table 2.

Case 1 assumes a less informative (larger standard deviation) prior distribution centered about the midpoint of the calibration parameter range (mean of 225 GPa). In this case, the posterior distribution for the calibration parameter and the discrepancy function were neither

accurate nor precise, which indicates a lack of identifiability. In Case 2, an informative prior (small standard deviation) with a mean equal to the true θ results in an identifiable system, which is evident by the accurate and precise posterior distributions. In contrast, Case 3 shows the inherent danger of using an informative prior that is inaccurate albeit precise. In this case, the posterior for the discrepancy function has a small prediction interval and the posterior of the calibration parameter has a small variance, which would lead one to believe that they are precisely estimated. However, the results of posteriors are inaccurate and do not reflect the true discrepancy function and θ.

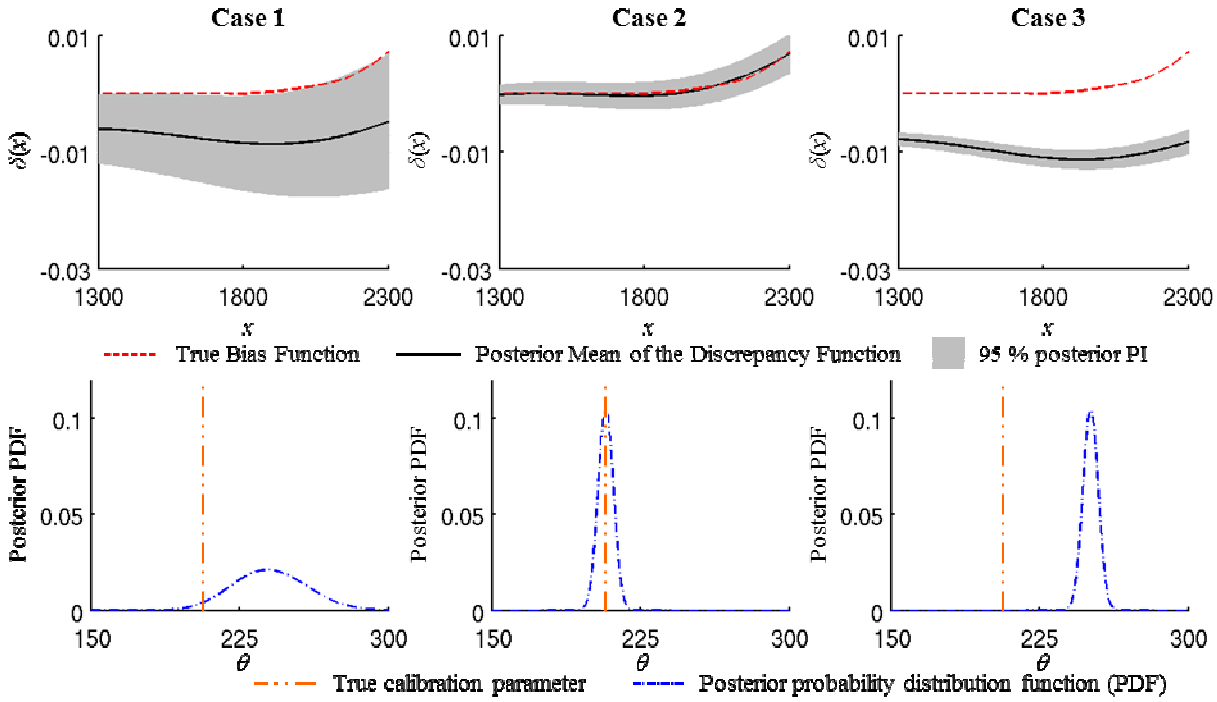

Fig. 10 Posterior distributions for the discrepancy function and the calibration parameter for the three prior distributions in Table 2.

Table 2 Prior distribution (normal) and posterior distribution means and standard deviations for the calibration parameter.

	Case 1	Case 2	Case 3
Prior Mean	225.00	206.80	250.00
Prior Std. Dev.	22.36	3.87	3.87
Post Mean	240.63	207.43	250.68
Post Std. Dev.	18.68	3.84	3.83

In conclusion, since one rarely has such accurate and informative prior knowledge about the sources of uncertainty present in the engineering system, assuming informative priors for the calibration parameters or the discrepancy function is not a satisfying solution to the identifiability problem.

6 When is Identifiability Possible?

The simply supported beam of Section 5 represents a system for which we cannot accurately and precisely identify both the calibration parameters and the discrepancy function (e.g., via their posterior distributions). This leads one to question whether identifiability is ever possible, given the nature of the model in Eq. (3). In this section we demonstrate that the answer is yes, but under certain assumptions. We illustrate this with a simple example in which the requisite assumptions are that the discrepancy function is reasonably smooth and can be represented by a smooth GP model (i.e., a GP with small roughness parameters, ω). Loosely speaking, "smooth" means that the function does not change abruptly for small changes in the design variables. In the context of our GP-based modeling, smooth means that the function is consistent with a GP model having relatively small roughness parameters ($\boldsymbol{\omega}$), as discussed later at the end of this section. Note that in our companion paper [7], we provide a solution to improving identifiability and discuss how incorporating experimental measurements of multiple

responses, with each depending on a common set of calibration parameters, can further enhance identifiability, sometimes substantially.

As evident from the posterior distributions of Fig. 8, the simply supported beam is not identifiable. The conceptual explanation (see the discussion surrounding Fig. 9) is that for many different values of θ, the estimated discrepancy function is smooth and consistent with a GP model. In contrast, we next consider an example in which the estimated discrepancy function has behavior that is inconsistent with a GP model when $\theta \neq \theta^*$, but consistent when $\theta = \theta^*$. We will show that this is the essential ingredient for good identifiability. Consider the following step function as the computer model

$$y^m(x,\theta) = \begin{cases} 1-\theta & x < 2 \\ 1+\theta & x \geq 2 \end{cases}, \tag{11}$$

with $1 \leq x \leq 3$ and $0.5 \leq \theta \leq 3$. Notice that the magnitude of the step is dictated by the calibration parameter θ. Plots of $y^m(x,\theta)$ for $\theta = 1.0$ and 3.0, respectively, are shown in Fig. 11(a) and Fig. 11(c). The physical experimental responses were generated according to

$$y^e(x) = y^m(x,\theta^*) + \delta(x) + \varepsilon \tag{12}$$

with the true calibration parameter $\theta^* = 2.5$, experimental error $\varepsilon = 0$, and the true discrepancy function $\delta(x)$ generated as one random realization of a smooth GP with a prior mean of 0 and a Gaussian covariance function with $\omega^\delta = \sigma_\delta = 1$.

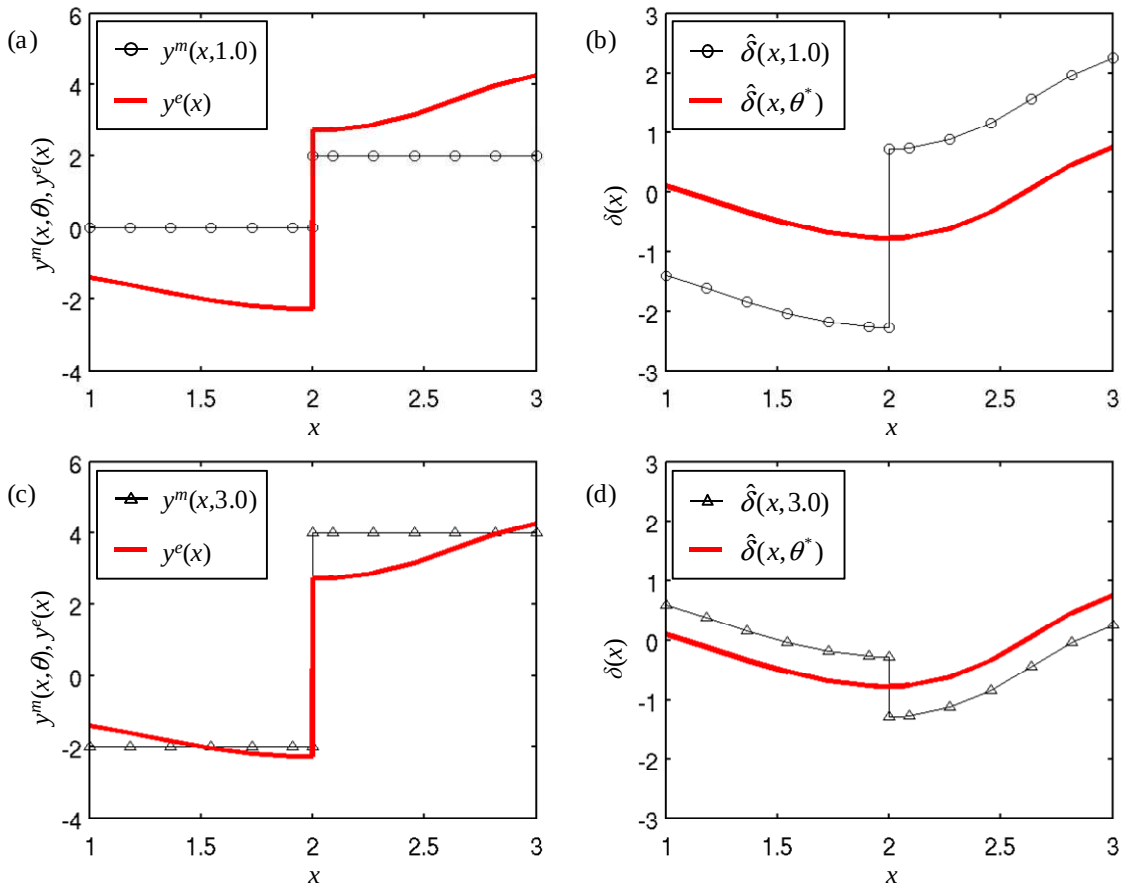

Fig. 11 Computer model and experimental response for Eq. (11) and Eq. (12) with (a) $\theta = 1.0$ and (c) $\theta = 3.0$. Corresponding estimated discrepancy function $\hat{\delta}(x,\theta)$ for (b) $\theta = 1.0$ and (d) $\theta = 3.0$.

To quantitatively assess identifiability, we again use the modular Bayesian approach outlined in Section 4. In Module 1, a GP model for the computer model (the step function of Eq. (11)) was created using a total of 168 simulation runs, as shown in Fig. 12. 144 runs were observed on a 12 by 12 grid. Additionally, to accurately capture the step behavior of the computer model, 24 additional runs were conducted at x = 1.99 and 2.01, for 12 evenly spaced values of θ. The GP model of the computer model was difficult to fit using the 168 simulations, due to the discontinuity at x = 2. Typically, a GP model with a Gaussian correlation function is

not a good choice to model a step function because of its inherent smoothness (see ref. [26] p. 123 and subsequent discussion). However, Fig. 12 shows that with the large amount of data and allowing large roughness parameters (the MLEs of the hyperparameters are $\boldsymbol{\omega}^m = [\omega^x \; \omega^\theta]^T$, where $\omega^x = 377.52$ and $\omega^\theta = 50$, and ω^θ fell at the upper boundary of what we allowed in the MLE algorithm for numerical purposes) a fairly accurate GP metamodel results. Although the fitted GP metamodel is somewhat rough and interpolates less accurately near the discontinuity, it serves to illustrate when identifiability is achievable.

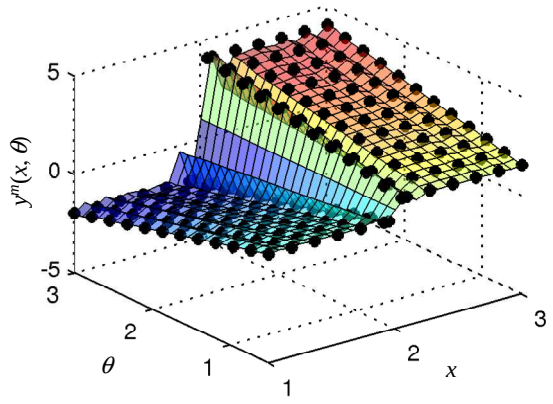

Fig. 12 Fitted GP model of the computer model of Eq. (11) based on 168 simulation runs, $\mathbf{y}^m$ (black dots).

After creating the GP model for the computer model, the modular Bayesian approach estimates the hyperparameters for the GP model representing the discrepancy function (Module 2). Within Module 2, 14 experimental data points were collected at $\mathbf{X}^e$ = [1 1.167 1.333, …, 3, 1.99, 2.01]^T (the black dots in Fig. 13(a)). The same x locations were chosen for the simulation and experimental runs in order to focus on identifiability rather than interpolation performance. Finally, Modules 3 and 4 calculate the posterior distributions for the calibration parameters, the experimental response, and the discrepancy function. Fig. 13 shows the resulting posterior distributions.

The accurate and precise (low posterior standard deviation) posterior distributions for the discrepancy function in Fig. 13(b) and the calibration parameter in Fig. 13(c) demonstrates that good identifiability is achievable for this example. Notice that the posterior mean of the experimental response in Fig. 13(a) is a less accurate and precise predictor near the step discontinuity. In spite of this, reasonable identifiability of the calibration parameter was achieved, which was the objective of this analysis.

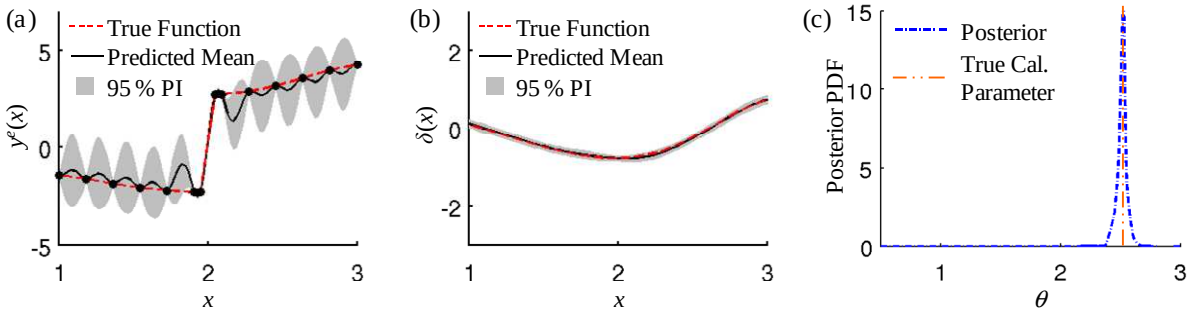

Fig. 13 Posterior distributions for (a) the experimental response (black dots indicate experimentally observed response values), (b) the discrepancy function, and (c) the calibration parameter θ.

To understand why the system is identifiable in this example, consider the estimated discrepancy functions $\hat{\delta}(x,\theta)$ for two incorrect values of θ shown in Fig. 11(b) (θ = 1.0) and Fig. 11(d) (θ = 3.0). For reference, the estimated discrepancy function $\hat{\delta}(x,\theta^*)$ for the true θ^* = 2.5 is shown in both figures. The salient feature is that $\hat{\delta}(x,\theta)$ for the incorrect values of θ have discontinuities that are inconsistent with a smooth GP model for the discrepancy function. The only value of θ that results in a $\hat{\delta}(x,\theta)$ consistent with model is the true value θ^* = 2.5. By a $\hat{\delta}(x,\theta)$ "inconsistent" with the model, we mean that when that value of θ and its corresponding $\hat{\delta}(x,\theta)$ are plugged into the likelihood function, the resulting likelihood is

exceptionally low. Because the estimated discrepancy function entails a high likelihood only when θ is close to θ^*, and the likelihood drops off abruptly as θ deviates from θ^*, the second derivative of the likelihood function (with respect to θ) is quite large. This translates to a large observed Fisher information (the expected value of the second derivative), which results in a tight posterior distribution for θ. This is, in essence, the definition of identifiability. The preceding example illustrates that identifiability is possible when small changes in the calibration parameters about their true values result in an estimated discrepancy function (which reflects the differences between the observed experimental data and the simulation model) that is inconsistent with the GP model of the discrepancy function.

We can also view the identifiability issues of the preceding discussion in a slightly different light. In essence, the approach evaluates identifiability by considering the simulation data; considering the experimental data; and then assessing the likelihood that the observed differences between the simulation and experimental data are explained by the discrepancy function, by an adjustment of the calibration parameters, or by various combinations thereof. Hypothetically, if there is roughly equal likelihood that the differences between the simulation and experimental data are explained by two (or more) different combinations of a discrepancy function and an adjustment in the calibration parameters, then identifiability would be poor in this case. On the other hand, if the likelihood is high that one particular combination of discrepancy function and adjustment in the calibration parameters accounts for the differences between the simulation and experimental data, then identifiability would be good in this case.

In either case, the likelihood must be evaluated with respect to the assumed prior distribution for the discrepancy function. And in this sense, identifiability is possible only by making certain assumptions on the prior distribution model for the discrepancy function.

Although identifiability is achieved in the step example via certain assumptions on the prior distributions, this is quite different than assuming a tight, informative prior distribution for the calibration parameters. The latter would also generally result in identifiability (i.e., a tight posterior), but it would be an artificial identifiability in that one would have to begin with precise knowledge of the parameters via a tight prior. In all of the examples of this paper, we have used only relatively non-informative prior distributions for the calibration parameters. We also note that in the preceding we use the term "likelihood", but it is technically a Bayesian posterior probability, which considers both the likelihood and the prior distributions for all parameters and hyperparameters.

7 Conclusions

To use a computer model for simulation based design, designers must build confidence in using the model for predicting physical systems, which is accomplished by quantifying uncertainty via model updating. In this work, we review various sources of uncertainty in predictive modeling and conclude that the model updating formulation proposed by Kennedy and O'Hagan is the most comprehensive and applicable to design under uncertainty. However, a limitation of this model updating formulation is that it can be difficult to distinguish between the effects of the calibration parameters (parameter uncertainty) and the discrepancy function (model discrepancy), which results in a lack of identifiability. It is important to identify, or differentiate, between these two sources of uncertainty in order to better understand the underlying sources of model uncertainty and ultimately further improve the model to better represent reality.

We have attempted to provide a better understanding of the identifiability problem in relation to model updating by using two very different illustrative examples. We first showed

how the calibration parameters and the discrepancy function were not identifiable in the simply supported beam example, even when significant amounts of simulation and experimental data are available. There are several different combinations of the computer model (with different values for the calibration parameter) and the estimated discrepancy function that combine to accurately predict the experimental response, which translates to poor identifiability. To improve identifiability, one can use informative prior distributions for the calibration parameters and/or the discrepancy function; however, this is a crude solution. Furthermore, in practice, informative priors typically do not exist. In the second example (a step function), we showed that identifiability is possible under the relatively mild assumption of a smooth discrepancy function. In this example, the estimated discrepancy function is only smooth and consistent with the GP model when the calibration parameter equals the true value, which resulted in good identifiability. As evident from the two examples, identifiability is a highly nuanced and difficult problem, but not impossible.

Another limitation of the modular Bayesian approach is the overall computational cost, which is mainly affected by the number of observed simulation and physical experimental runs. As the number of observations increases, the computational cost of the modular Bayesian approach also increases. For example, in the simply supported beam example with $N_m = 16$ and $N_e = 11$, the modular Bayesian approach took approximately 10 seconds to compute the posterior distributions (on a single Intel 2.66 GHz processer). In the step example of Section 6, with $N_m = 168$ and $N_e = 14$, the modular Bayesian approach required approximately 55 seconds to compute the posterior distributions. Although the computational cost was not prohibitive in either of these examples, it can become prohibitive as the number of observations increases.

It may sometimes be possible to enhance identifiability by incorporating additional

information. In our companion paper [7], the simply supported beam example is revisited to incorporate additional information in the form of multiple responses that mutually depend on a common calibration parameter, in order to improve identifiability of systems that are otherwise poorly identified. Together, this paper and the companion paper [7] shed light on the challenging problem of accurately identifying calibration parameters and a discrepancy function 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.

8 Acknowledgements

The grant support from the National Science Foundation (CMMI-0928320 and CMMI-0758557) is greatly acknowledged. The views expressed are those of the authors and do not necessarily reflect the views of the sponsors.

Nomenclature

C Constant stress of the material power law (MPa)

E Young's Modulus (GPa)

$\mathbf{H}$ Regression functions matrix for the mean of a GP model

$\mathbf{H}^m$ Regression functions matrix for the mean of the computer model GP

$\mathbf{H}^\delta$ Regression functions matrix for the mean of the discrepancy function GP

$GP(\bullet, \bullet)$ Gaussian process (GP) model

$N(\bullet, \bullet)$ Normal probability distribution

$\mathbf{R}$ Correlation matrix of a GP model

$R(\bullet, \bullet)$ Correlation function of a GP model for all inputs $\bullet$

$R^m((\mathbf{x}, \boldsymbol{\theta}),(\mathbf{x}', \boldsymbol{\theta}'))$ Correlation function, between points ($\mathbf{x}$, $\boldsymbol{\theta}$) and ($\mathbf{x}'$, $\boldsymbol{\theta}'$), for the computer model GP

$R^{\delta}(\mathbf{x}, \mathbf{x}')$ Correlation function, between points $\mathbf{x}$ and $\mathbf{x}'$, for the discrepancy function GP

$\mathbf{X}$ Inputs settings for the observed data ($N \times d$)

$\mathbf{X}^e$ Input settings of $\mathbf{x}$ for the experimental data ($N_e \times d$)

$\mathbf{X}^m$ Input settings of $\mathbf{x}$ for the simulations ($N_m \times d$)

$\mathbf{d}$ Vector of simulations and experimental data ($\mathbf{d} = [\mathbf{y}^{m\,T}\ \mathbf{y}^{e\,T}]^T$)

$\mathbf{h}(\mathbf{x})$ Vector of regression functions for the mean function of a GP model

$\mathbf{h}^m(\mathbf{x}, \boldsymbol{\theta})$ Vector of regression functions for the mean function of the computer model GP

$\mathbf{h}^{\delta}(\mathbf{x})$ Vector of regression functions for the mean function of the discrepancy function GP

n Strain-hardening exponent of the material power law

$p(\boldsymbol{\theta})$ Prior for the calibration parameters

$p(\mathbf{d} \mid \boldsymbol{\theta}, \hat{\boldsymbol{\phi}})$ Likelihood function given both the value of $\boldsymbol{\theta}$ and the estimates of the hyperparameters

$p(\boldsymbol{\theta} \mid \mathbf{d}, \hat{\boldsymbol{\phi}})$ Posterior distribution of the calibration parameters

x, $\mathbf{x}$ Controllable or design inputs

$\mathbf{y}$ Vector of responses collected at input settings $\mathbf{X}$ ($N \times 1$)

$y^e(\mathbf{x})$ Experimental response

$\mathbf{y}^e$ Experimental data collected at input settings $\mathbf{X}^e$ ($N_e \times 1$)

$y^m(\mathbf{x})$ and $y^m(\mathbf{x}, \boldsymbol{\theta})$ Computer model response

$\mathbf{y}^m$ Simulation data collected at input settings $\mathbf{X}^m$ ($N_m \times 1$)

$\boldsymbol{\beta}$ Regression coefficients for the mean function of a GP

$\boldsymbol{\beta}^m$ Regression coefficients for the mean function of the computer model GP

$\boldsymbol{\beta}^{\delta}$ Regression coefficients for the mean function of the discrepancy function GP

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

$\hat{\delta}(\mathbf{x},\boldsymbol{\theta})$ Estimate of the discrepancy function

ε Experimental uncertainty of the physical experiments

λ Variance of the experimental uncertainty

$\theta, \boldsymbol{\theta}$ Calibration parameters

$\boldsymbol{\Theta}^{m}$ Realizations of $\boldsymbol{\theta}$ for the simulation data ($N_m \times r$)

σ^2 Constant variance of the GP model

$\sigma_m^{\,2}$ Constant variance of the covariance function for the computer model GP

σ_δ^2 Constant variance of the covariance function for the discrepancy function GP

$\hat{\boldsymbol{\phi}}$ Estimated hyperparameters $\boldsymbol{\phi} = [\boldsymbol{\phi}^m \; \boldsymbol{\phi}^\delta]$

$\boldsymbol{\phi}^m$ Hyperparameters of the computer model GP $\boldsymbol{\phi}^m = [\boldsymbol{\beta}^m \; \sigma_m \; \boldsymbol{\omega}^m]$

$\boldsymbol{\phi}^\delta$ Hyperparameters of the discrepancy function GP $\boldsymbol{\phi}^\delta = [\boldsymbol{\beta}^\delta \; \sigma_\delta \; \boldsymbol{\omega}^\delta]$

$\boldsymbol{\omega}$ Roughness parameters for the correlation function of a GP model

$\boldsymbol{\omega}^m$ Roughness parameters for the correlation function of the computer model GP

$\boldsymbol{\omega}^\delta$ Roughness parameters for the correlation function of the discrepancy function GP

References

[1] Chen, W., Yin, X., Lee, S., and Liu, W., 2010, "A multiscale design methodology for hierarchical systems with random field uncertainty," Journal of Mechanical Design, 132(4), p. 041006.

[2] Chen, W., Xiong, Y., Tsui, K. L., and Wang, S., 2008, "A design-driven validation approach using Bayesian prediction models," Journal of Mechanical Design, 130(2), p. 021101.

[3] Wang, Z., Huang, H.-Z., and Liu, Y., 2010, "A unified framework for integrated optimization under uncertainty," Journal of Mechanical Design, 132(5), p. 051008.

[4] Singh, A., Mourelatos, Z. P., and Li, J., 2010, "Design for lifecycle cost using time-dependent reliability," Journal of Mechanical Design, 132(9), p. 091008.

[5] Apley, D. W., and Kim, J. B., "A cautious approach to robust parameter design with model uncertainty," IIE Transactions, to appear.

[6] Lancaster, T., 2004, An introduction to modern Bayesian econometrics, Blackwell Publishing, Malden, MA.

[7] Arendt, P., Apley, D., and Chen, W., 2011, "Improving identifiability in model calibration using multiple responses," *Submitted to the* Journal of Mechanical Design.

[8] Kennedy, M. C., and O'Hagan, A., 2001, "Bayesian calibration of computer models," Journal of the Royal Statistical Society: Series B, 63(3), p. 425-464.

[9] Box, G., Hunter, W., and Hunter, J., 1978, Statistics for experimenters: An introduction to design, data analysis, and model building, Wiley, New York.

[10] McKay, M. D., Beckman, R. J., and Conover, W. J., 1979, "A comparison of three methods for selecting values of input variables in the analysis of output from a computer code," Technometrics, 21(2), p. 239-245.

[11] Sacks, J., Welch, W. J., Mitchell, T. J., and Wynn, H. P., 1989, "Design and analysis of computer experiments," Statistical Science, 4(4), p. 409-423.

[12] Liu, Y., Chen, W., Arendt, P., and Huang, H.-Z., 2011, "Toward a better understanding of model validation metrics," Journal of Mechanical Design, 133(7), p. 071005.

[13] Xiong, Y., Chen, W., Tsui, K. L., and Apley, D. W., 2009, "A better understanding of model updating strategies in validating engineering models," Computer Methods in Applied Mechanics and Engineering, 198(15-16), p. 1327-1337.

[14] Lindgren, L.-E., Alberg, H., and Domkin, K., 2003, "Constitutive modelling and parameter optimization," 7th International Conference on Computational Plasticity, Barcelona, Spain.

[15] McFarland, J., Mahadevan, S., Romero, V., and Swiler, L., 2008, "Calibration and uncertainty analysis for computer simulations with multivariate output," AIAA Journal, 46(5), p. 1253-1265.

[16] Wang, S., Tsui, K. L., and Chen, W., 2009, "Bayesian validation of computer models," Technometrics, 51(4), p. 439-451.

[17] Higdon, D., Kennedy, M. C., Cavendish, J., Cafeo, J., and Ryne, R., 2004, "Combining field data and computer simulations for calibration and prediction," SIAM Journal on Scientific Computing, 26(2), p. 448-466.

[18] Bayarri, M. J., Berger, J. O., Paulo, R., Sacks, J., Cafeo, J. A., Cavendish, J., Lin, C. H., and Tu, J., 2007, "A framework for validation of computer models," Technometrics, 49(2), p. 138-154.

[19] Loeppky, J., Bingham, D., and Welch, W., 2006, "Computer model calibration or tuning in practice," University of British Columbia, Vancouver, BC, CA.

[20] 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), p. 945-958.

[21] Hasselman, T. K., Yap, K., Lin, C.-H., and Cafeo, J., 2005, "A case study in model improvement for vehicle crashworthiness simulation," Proceedings of the 23rd International Modal Analysis Conference, Orlando, FL.

[22] Easterling, R. G., and Berger, J. O., 2002, "Statistical foundations for the validation of computer models," presented at Computer Model Verification and Validation in the 21st Century Workshop, Johns Hopkins University.

[23] Qian, P. Z. G., and Wu, C. F. J., 2008, "Bayesian hierarchical modeling for integrating low-accuracy and high-accuracy experiments," Technometrics, 50(2), p. 192-204.

[24] Drignei, D., 2009, "A Kriging approach to the analysis of climate model experiments," Journal of Agricultural, Biological, and Environmental Statistics, 14(1), p. 99-114.

[25] Liu, F., Bayarri, M. J., Berger, J. O., Paulo, R., and Sacks, J., 2008, "A Bayesian analysis of the thermal challenge problem," Comput. Meth. Appl. Mech. Eng., 197(29-32), p. 2457-2466.

[26] Rasmussen, C. E., and Williams, C. K. I., 2006, Gaussian processes for machine learning, The MIT Press, Cambridge, Massachusetts.

[27] Jin, R., Chen, W., and Simpson, T. W., 2001, "Comparative studies of metamodelling techniques under multiple modelling criteria," Structural and Multidisciplinary Optimization, 23(1), p. 1-13.

[28] Santner, T. J., Williams, B. J., and Notz, W. I., 2003, The design and analysis of computer experiments, Springer Verlag, New York.

[29] Cressie, N., 1993, Statistics for spatial data, Wiley, New York.

[30] Schabenberger, O., and Gotway, C., 2005, Statistical methods for spatial data analysis Chapman & Hall/CRC, Boca Raton, Florida.

[31] Liu, F., Bayarri, M. J., and Berger, J. O., 2009, "Modularization in Bayesian analysis, with emphasis on analysis of computer models," Bayesian Analysis, 4(1), p. 119-150.

[32] 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), p. 1301-1309.

[33] Jin, R., 2004, "Enhancements of metamodeling techniques in engineering design," Ph.D., University of Illinois at Chicago, Chicago, Illinois.

[34] Rasmussen, C. E., 1996, "Evaluation of Gaussian processes and other methods for non-linear regression," Ph.D., University of Toronto, Toronto, ON, CA.

[35] Kennedy, M. C., and O'Hagan, A., 2000, "Supplementary details on Bayesian calibration of computer models," (last accessed on March 13, 2012), http://www.shef.ac.uk/~stlao/ps/calsup.ps.

[36] Arendt, P., Apley, D. W., and Chen, W., 2010, "Updating predictive models: Calibration, bias correction, and identifiability," ASME 2010 International Design Engineering Technical Conferences, Montreal, QC, CA.

[37] Chakrabarty, J., 2006, Theory of plasticity, Elsevier Butterworth-Heinemann.

[38] Joseph, V., and Melkote, S., 2009, "Statistical adjustments to engineering models," Journal of Quality Technology, 41(4), p. 362-375.

[39] Higdon, D., Gattiker, J., Williams, B., and Rightley, M., 2008, "Computer model calibration using high-dimensional output," Journal of the American Statistical Association, 103(482), p. 570-583.