Comput. Methods Appl. Mech. Engrg. 198 (2009) 1327–1337

Contents lists available at ScienceDirect

Comput. Methods Appl. Mech. Engrg.

journal homepage: www.elsevier.com/locate/cma

A better understanding of model updating strategies in validating engineering models

Ying Xiong [a], Wei Chen [a,*], Kwok-Leung Tsui [b], Daniel W. Apley [c]

[a] Department of Mechanical Engineering, Northwestern University, Evanston, IL 60208, USA
[b] School of Industrial and Systems Engineering, Georgia Institute of Technology, Atlanta, GA 30332, USA
[c] Department of Industrial Engineering and Management Sciences, Northwestern University, 2145 Sheridan Road, Evanston, IL 60208, USA

ARTICLE INFO

Article history:
Received 17 July 2008
Received in revised form 20 November 2008
Accepted 27 November 2008
Available online 24 December 2008

Keywords:
Model updating
Model validation
Bayesian approach
Maximum likelihood estimation
Thermal challenge problem

ABSTRACT

Our objective in this work is to provide a better understanding of the various model updating strategies that utilize mathematical means to update a computer model based on both physical and computer observations. We examine different model updating formulations, e.g. calibration and bias-correction, as well as different solution methods. Traditional approaches to calibration treat certain computer model parameters as fixed over the physical experiment, but unknown, and the objective is to infer values for the so-called calibration parameters that provide a better match between the physical and computer data. In many practical applications, however, certain computer model parameters vary from trial to trial over the physical experiment, in which case there is no single calibrated value for a parameter. We pay particular attention to this situation and develop a maximum likelihood estimation (MLE) approach for estimating the distributional properties of the randomly varying parameters which, in a sense, calibrates them to provide the best agreement between physical and computer observations. Furthermore, we employ the newly developed u-pooling method (by Ferson et al.) as a validation metric to assess the accuracy of an updated model over a region of interest. Using the benchmark thermal challenge problem as an example, we study several possible model updating formulations using the proposed methodology. The effectiveness of the various formulations is examined. The benefits and limitations of using the MLE method versus a Bayesian approach are presented. Our study also provides insights into the potential benefits and limitations of using model updating for improving the predictive capability of a model.

1. Introduction

Computer models have been widely used in engineering design and analysis to simulate complex physical phenomena. The accuracy or adequacy of a computer model can be assessed by means of model validation, which refers to the process of determining the degree to which a computational simulation is an accurate representation of the real world from the perspective of the intended uses of the model [1]. While there exists no unified approach to model validation, it is increasingly recognized that model validation is not merely a process of assessing the accuracy of a computer model, but should also help improve the model based on the validation results.

Strategies for model improvement roughly fall into two categories: *model refinement* and *model updating*. *Model refinement* involves changing the physical principles in modeling or using other means to build a more sophisticated model that better represents the physics of the problem by, for example, using a non-linear finite element method to replace a linear method, correcting and refining boundary conditions, or introducing microscale modeling in addition to macroscale modeling, etc. *Model updating*, on the other hand, utilizes mathematical means (e.g. calibrating model parameters and bias-correction) to match model predictions with the physical observations. While model refinement is desirable for fundamentally improving the predictive capability, the practical feasibility of refinement is often restricted by available knowledge and computing resources. In contrast, model updating is a cheaper means that can be practical and useful if done correctly. Here, *predictive capability* refers to the capability of making accurate predictions in domains (or locations) where no physical data are available.

While various model updating strategies (formulations and solution methods) exist, there is a lack of understanding of the effectiveness and efficiency of these methods. It is our interest in this work to examine various model updating strategies to achieve a better understanding of their merits. We are particularly interested in the role that model updating plays versus model validation and prediction. A detailed review is provided in Section 3. In summary, conventional calibration approaches [2] assume

* Corresponding author. Tel.: +1 847 491 7019; fax: +1 847 491 3915.
E-mail address: weichen@northwestern.edu (W. Chen).

0045-7825/$ - see front matter © 2008 Elsevier B.V. All rights reserved.
doi:10.1016/j.cma.2008.11.023

Nomenclature

$\mathbf{x} = \{x_1, x_2, \ldots, x_n\}$	n controllable input variables
$\boldsymbol{\theta} = \{\theta_1, \theta_2, \ldots, \theta_m\}$	m uncontrollable input variables
$y^e(\mathbf{x})$	physical experiments
$y^m(\mathbf{x})$ or $y^m(\mathbf{x}, \boldsymbol{\theta})$	computer model
$\delta(\mathbf{x})$	bias function
ε	experimental error
$y^{m'}(\mathbf{x}, \Theta)$	updated model
$y^{pred}(\mathbf{x})$	predictive model
$\boldsymbol{\Theta}$	model updating parameters
$L(\boldsymbol{\Theta})$	likelihood function
$F_{\mathbf{x}_i^e}(y_i^e)$	cumulative distribution function (CDF) at y_i^e

calibration parameters are fixed and estimated, typically using least squares to match the model with the physical observations. This type of approach for model updating is inconsistent with the primary concerns of model validation and prediction in which various uncertainties should be accounted for either explicitly or implicitly. Examples of such uncertainties include experimental error, lack of data, uncertain model parameters, and model uncertainty (systematic model inadequacy). The more recent Bayesian style calibration, also named calibration under uncertainty (CUU) or stochastic calibration, treats calibration parameters as unknown entities that are fixed over the course of the physical experiment. Initial lack of knowledge of the parameters is represented by assigning prior distributions to them, and, given the experimental data, this lack of knowledge is revised by updating their distributions (from priors to posteriors) based on the observed data through Bayesian analysis [3,4]. However, as we discuss in a more thorough examination in Section 3.2, several limitations of applying the Bayesian calibration approaches are identified in this work.

One limitation of the aforementioned calibration approaches is that the calibration parameters are assumed to remain fixed over the entire course of the physical experiment and beyond. In contrast, it is frequently the case that some parameters vary randomly over the physical experiment, perhaps due to manufacturing variation, variation in raw materials, variation in environmental or usage conditions, etc. This violates the assumptions under which the Bayesian or regression-based calibration analyses are derived. In this situation, rather than assuming fixed parameters and updating their posterior distributions to represent our lack of knowledge of them, it is more reasonable to treat them a randomly varying and estimate their distributional properties by integrating the physical data with the model. In essence, the distributional properties (e.g. the mean and variance of the randomly varying parameters) become the calibration parameters for the model, and the objective is to identify values for them that provide the best agreement with the observed distributional properties (e.g. the dispersion [5]) of the physical experimental data. In this paper, we present a maximum likelihood estimation (MLE) [6] approach for accomplishing this. The MLE method is used to estimate a set of unknown parameters (heretofore called model updating parameters) associated with several modeling updating formulations, which include the distributional properties of parameters that vary randomly over the experiment, as well as well as more traditional fixed calibration parameters and quantities associated with bias-correction and random experimental error.

The remainder of the paper is organized as follows. In Section 2, we discuss the role that model updating plays versus model validation and prediction. In Section 3, we examine the existing model updating formulations under two categories, namely, model bias-correction and calibration. The popular Bayesian approach is described and its limitations are highlighted. In Section 4, we describe our proposed MLE based model updating approach, together with the introduction of the u-pooling validation metric. In Section 5, a benchmark thermal challenge problem adopted by the Sandia Validation workshop [7,8] is used as an example to illustrate the proposed approach, draw important conclusions, and portray these conclusions in relation to conclusions from prior studies. Section 6 is the closure with a summary of the features of the proposed method, the relative merits of different approaches, the insights gained, and future research directions.

2. Role of model updating vs. model validation

In this work, model updating is viewed as a process that continuously improves a computer model through mathematical means based on the results from newly added physical experiments, until the updated model satisfies the validation requirement or the resource is exhausted. Therefore, even though model updating is interrelated with model validation, it is viewed as a separate activity that occurs before "validation". As shown in Fig. 1, the model updating procedure integrates the computer simulation model y^m with the physical experiment data $\mathbf{y}^e$ to yield an updated model $y^{m'}(\cdot)$. This updated model is then subject to a validation procedure that utilizes additional physical experiments $\mathbf{y}^e$ in the intended region of interest for validation. As noted from this diagram, unlike many contemporary model validation works, model validation in this work is used to evaluate an evolved, updated model $y^{m'}(\cdot)$, rather than the original computer model $y^m(\cdot)$. Besides, the updated model $y^{m'}(\cdot)$ is the one used for making future prediction $y^{pred}(\cdot)$ with the consideration of various sources of uncertainties. For implementing model updating and validation in a computationally efficient manner, it is indicated in Fig. 1 that a metamodel (surrounded by a dashed box) may be used to substitute the original computer model if it is expensive to compute.

As more details are provided in the remaining sections,*model updating* utilizes mathematical means (e.g. calibrating model parameters, bias-correction) to match model predictions with the physical observations. Model updating provides not only the formulation of an *updated model*, but also the characterization of model updating parameters $\boldsymbol{\Theta}$, together with the associated assumptions. As noted, the model updating procedure, during which $y^m(\cdot)$ is treated as a black-box, is largely driven by the ob-

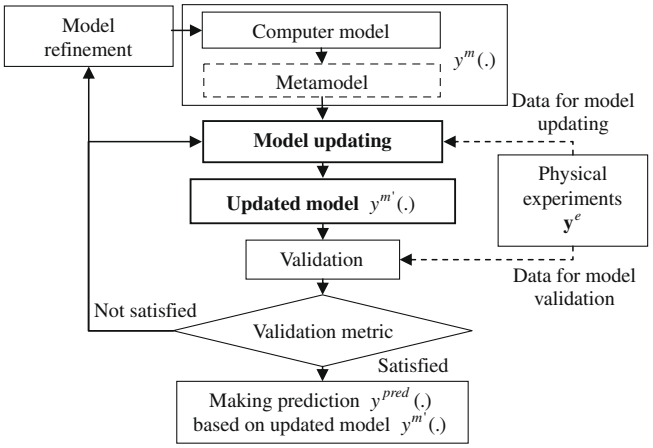

Fig. 1. Relationship of model updating, model refinement, and model validation.

served experimental data. It is our interest in this work to examine whether such a data-driven approach can improve the predictive capability of a computer model. As argued by Ferson et al. [9], the extrapolation capability in using a model to make prediction should be carefully validated.

3. Examination of existing model updating methods

The existing model updating strategies differ in their formulations, the solution method used, and the physical interpretations. In the following overview, we first review two categories of formulations, namely, bias-correction and calibration. We then discuss the limitations of using the Bayesian approach as the solution method. The physical interpretations are provided throughout the review and will be further expanded in Section 4.

3.1. Model bias-correction approaches

Bias-correction is useful when accuracy improvement cannot be accomplished by calibrating model parameters [10,11]. One widely accepted interpretation of the bias-correction approach [3] is that it captures the potential model error due to the use of incorrect modeling method (e.g. modeling a non-linear behavior with a linear model), which often cannot be compensated by other means. There are various formulations of bias-correction in the literature. In the Bayesian bias-correction model proposed by Chen et al. [12] and Wang et al. [13], a plain additive bias-correction model is formulated as

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

where the bias function $\delta(\mathbf{x})$ is a direct measure of the difference between the computer model $y^m(\mathbf{x})$ and the physical process $y^e(\mathbf{x})$. The bias function $\delta(\mathbf{x})$ is assumed to be a Gaussian Process model, the uncertainty of which reflects the uncertainty involved in a model updating process such as the experimental error, lack of data, etc. One advantage of using the above formulation is that the closed form Bayesian posterior of the Gaussian process model $\delta(\mathbf{x})$ can be derived. In addition, the bias function $\delta(\mathbf{x})$ provides a direct measure of the assessed accuracy (or validity) of a computer model within a particular design region or at an application site [13].

In addition to using the additive bias shown in Eq. (1), a bias-correction approach may employ a combination of multiplicative bias and additive bias, as shown in the following formulation [14],

$$y^e(\mathbf{x}) = \rho(\mathbf{x}) y^m(\mathbf{x}) + \delta(\mathbf{x}) + \varepsilon, \tag{2}$$

where $\rho(\mathbf{x})$ is modeled as a simple linear regression model w.r.t. $\mathbf{x}$, ε is assumed to be a zero-mean Gaussian random variable. The scaling function $\rho(\mathbf{x})$ in Eq. (2) brings more flexibility to the constant adjustment parameter ρ used in Kennedy and O'Hagan [3]. The Maximum Likelihood Estimation (MLE) method is utilized in their work [14] to estimate the regression coefficients of $\rho(\mathbf{x})$, while the closed forms Bayesian posteriors of the hyperparameters in the Gaussian Process $\delta(\mathbf{x})$ are derived for given prior distributions.

One inherent limitation of the bias-correction method is that it assumes all inputs ($\mathbf{x}$) of both the computer model ($y^m(\mathbf{x})$) and the physical process ($y^e(\mathbf{x})$) are observable and controllable. In practice, it often occurs that some of the model input parameters cannot be directly observed and measured in the physical experiments. This limitation can be addressed using the model calibration approach introduced next.

3.2. Model calibration approaches

With a typical model calibration approach, the inputs of a computer model are divided into controllable inputs ($\mathbf{x}$) and uncontrollable parameters ($\boldsymbol{\theta}$) that are assumed to be fixed over the experiment. Note that it is $\boldsymbol{\theta}$ that are to be calibrated. A computer model for the given input vector $(\mathbf{x},\boldsymbol{\theta})$ is denoted as $y^m(\mathbf{x},\boldsymbol{\theta})$, while the physical process is denoted to be $y^e(\mathbf{x})$ as a function of controllable inputs $\mathbf{x}$ only.

3.2.1. Deterministic calibration approach

A conventional way to carry out a deterministic parameter calibration is to formulate the problem in a fashion similar to that of the non-linear regression analysis [5,15,16] through the following equation

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

where e is the residual between the prediction from the calibrated computer model $y^m(\mathbf{x},\boldsymbol{\theta})$ and the experimental observation $y^e(\mathbf{x})$. The optimal values of the calibration parameters $\boldsymbol{\theta}$ are found by minimizing the (weighted) sum of the squared error (SSE) between the model predictions and the physical experiments [17], i.e.,

$$\text{Find } \boldsymbol{\theta} \text{ Minimizing SSE} = \sum_{i=1}^{N} w_i e_i^2 = \sum_{i=1}^{N} w_i [y^m(\mathbf{x}_i,\boldsymbol{\theta}) - y^e(\mathbf{x}_i)]^2, \tag{4}$$

where $\mathbf{x}_i = [x_{i1}, x_{i2}, \ldots, x_{ik}]^T$ $(i = 1,2,\ldots,N)$ are sample points, w_i $(i = 1,2,\ldots,N)$ are the weights for different experimental observations reflecting the quality of experimental data, $\boldsymbol{\theta} = [\theta_1, \theta_2, \ldots, \theta_m]^T$ are unknown physical constants, and k is the number of input variables. Although deterministic calibration approaches are generally plausible and easy to apply, the limitation is that they cannot account for uncertainties in both computer simulation and physical experimentation.

3.2.2. Non-deterministic Bayesian calibration approach

Non-deterministic parameter calibration is also called calibration under uncertainty (CUU) [16]. Kennedy and O'Hagan [3] first developed a Bayesian approach to simultaneously calibrate a computer model and characterize the potential bias (discrepancy) between the model output and the physical experiments. Their method is based on the following relation,

$$y^e(\mathbf{x}) = \rho \cdot y^m(\mathbf{x},\boldsymbol{\theta}) + \delta(\mathbf{x}) + \varepsilon, \tag{5}$$

where ρ is an unknown regression parameter (an adjustment parameter), $\delta(\mathbf{x})$ is a bias (discrepancy) function assumed to be the realization of a Gaussian Process, ε is the experimental error assumed to be a zero-mean Gaussian random variable. In essence, the formulation shown in Eq. (5) is a combination of both parameter calibration and bias-correction. In Kennedy and O'Hagan's work, the Bayesian analysis is performed to update the prior distributions of the calibration parameters $\boldsymbol{\theta}$and the hyperparameters underlying two separate Gaussian Process models: one for the bias function $\delta(\mathbf{x})$, and the other for replacing the original expensive computer model $y^m(\mathbf{x},\boldsymbol{\theta})$. To manage the computational complexity, priors are often only specified for a very few calibration parameters $\boldsymbol{\theta}$ and a small set of Gaussian Process model parameters, while the rest are assumed unknown but fixed.

Several variants and applications of Kennedy and O'Hagan's approach [3] exist in the literature. In the Simulator Assessment and Validation Engine (SAVE) framework developed by Bayarri et al. [18], followed by Higdon et al. [19], and Liu et al. [20], a formulation that is similar to the one used by Kennedy and O'Hagan is shown as follows with the regression parameter ρ omitted.

$$y^e(\mathbf{x}) = y^m(\mathbf{x},\boldsymbol{\theta}) + \delta(\mathbf{x}) + \varepsilon. \tag{6}$$

McFarland et al. [21] developed a simplified Bayesian calibration approach in the form of

$$y^e(\mathbf{x}^*) = y^m(\mathbf{x}^*,\boldsymbol{\theta}) + \varepsilon. \tag{7}$$

Their method does not consider the bias-correction, and poses the prior belief of the calibration parameters $\boldsymbol{\theta}$ as uniformly distributed. Unlike others, their calibration is only performed at one particular site $\mathbf{x}^*$, based on the assumption that the results of calibration are identical at different input sites. Such an assumption is questionable if the computer model is so wrong that the calibration at one single site could be heavily biased and is hard to be extrapolated to other sites.

3.3. Limitations of Bayesian approaches

While the Bayesian approach is useful when limited data are available, there are several common issues. First, as indicated in Trucano et al. [16], the prior distributions of calibration parameters are often difficult to specify due to the lack of prior knowledge. Subjectively assigned prior distributions may yield unstable posterior distributions [22], which undermines the advantage of Bayesian updating. Second, the Markov Chain Monte Carlo (MCMC) method used in most Bayesian calibration practices for obtaining the posterior distributions requires a significant amount of iterations, while the criterion for ceasing the Markov Chain growth is not established [16].

Loeppky et al. [23] examined a non-Bayesian version of the approach from Kennedy and O'Hagan [3], but using the MLE to estimate the calibration parameters which are assumed deterministic. The issue of identifiability of model bias was addressed by examining the likelihood ratio of two model versions: one with the bias term, the other without. It was demonstrated that the MLE estimates of calibration parameters will asymptotically attain an unbiased computer model if such a model exists. However, their method provides deterministic MLE estimates without acknowledging the uncertainty of model input.

4. A Maximum Likelihood Estimation (MLE) based model updating methodology

We present an alternative model updating approach that differs from the existing Bayesian approach in both uncertainty treatment and the computation of model updating parameters. The basic principle of this approach is to determine the model updating parameters with the MLE, so that the best agreement between the distribution of model outputs and the dispersion of the observed physical observations $\mathbf{y}^e$ can be achieved, while the experiment-to-experiment variation of calibration parameters is captured by parameter distributions.

4.1. Model updating formulations and parameters

As reviewed in Section 3, various formulations are available for constructing an updated model based on the original computer model $y^m(\mathbf{x},\boldsymbol{\theta})$. In our view, the choice of the updated formulation (denoted as $y^{m'}(\mathbf{x},\Theta)$) and the model updating parameters $\boldsymbol{\Theta}$ are problem dependent and will require insight into the error sources. In Section 5, we will investigate three possible model updating formulations for the specific thermal challenge problem. One typical formulation that combines both bias-correction and parameter calibration is illustrated here in Eq (8):

$$y^{m'}(\mathbf{x},\Theta) = y^m(\mathbf{x},\boldsymbol{\theta}) + \delta(\mathbf{x}) + \varepsilon. \tag{8}$$

In Eq. (8), $\mathbf{x} = \{x_1, x_2, \ldots, x_n\}$ are n controllable input variables, which are always deterministic. $\boldsymbol{\theta} = \{\theta_1, \theta_2, \ldots, \theta_m\}$ are m uncontrollable input variables, assumed random to capture the uncertainty associated with model input. The uncertainties of $\boldsymbol{\theta}$ are parameterized by distribution parameters $\{\mu_{\theta 1}, \sigma_{\theta 1}, \ldots, \mu_{\theta m}, \sigma_{\theta m}\}$, independent from model input $\mathbf{x}$. ε is an unobservable output variable, also assumed random, to capture the experimental uncertainty associated with a model output. Similar to $\boldsymbol{\theta}$, the distribution of ε is parameterized by $\{\mu_\varepsilon, \sigma_\varepsilon\}$. The bias function $\delta(\mathbf{x})$ is used to capture the model systematic bias, but not intended to account for the experimental uncertainty. $\delta(\mathbf{x})$could be parameterized in various ways, for example, with a regression model $\delta(\mathbf{x}) = \beta_{\delta 0} + \beta_{\delta 1} x_1 + \cdots + \beta_{\delta n} x_n$ parameterized by $\{\beta_{\delta 0}, \beta_{\delta 1}, \ldots, \beta_{\delta n}\}$. Here the bias function $\delta(\mathbf{x})$ is treated to be a deterministic function that does not contribute to the model output uncertainty. Other possible choices could be using a constantδ, which is less flexible, or a Gaussian Process (GP) model, which is more complex. Collectively, the model updating parameters for the above formulation are denoted as $\boldsymbol{\Theta} = \{\mu_{\theta 1}, \sigma_{\theta 1}, \ldots, \mu_{\theta m}, \sigma_{\theta m}; \beta_{\delta 0}, \beta_{\delta 1}, \ldots, \beta_{\delta n};\ \mu_\varepsilon, \sigma_\varepsilon\}$. Notice that in contrast to traditional calibration approaches, our model updating parameters $\boldsymbol{\Theta}$ do not directly include the parameters $\boldsymbol{\theta}$, because they are not assumed to be fixed. Rather, $\boldsymbol{\Theta}$ includes the means and variances of the parameters $\boldsymbol{\theta}$. Fig. 2 shows the collection of model updating parameters in a formulation with two calibration parameters and two model inputs, i.e., $\mathbf{x} = \{x_1, x_2\}$ and $\boldsymbol{\theta} = \{\theta_1, \theta_2\}$. Note that in the right hand side of Eq. (8), only $\boldsymbol{\theta}$ and ε are random quantities, as illustrated by the shaded vertical PDF profiles in Fig. 2. Also note that we assume $\boldsymbol{\theta}$ and ε follow normal distributions, thus only two parameters are needed to determine the distribution of each calibration parameter. With the statistical moments defined for the calibration parameters, the various sources of uncertainties in a model updating process can be propagated to form the uncertainty of the updated model $y^{m'}(\mathbf{x},\Theta)$, as illustrated by the shaded horizontal PDF profile in Fig. 2.

4.2. Determining model updating parameters via MLE

To determine the values of all model updating parameters $\boldsymbol{\Theta}$, the MLE method is adopted towards matching the output distribution of the updated model $y^{m'}(\mathbf{x},\Theta)$ with the dispersion observed in physical experiments $y^e(\mathbf{x})$.

To construct a likelihood function, the following equation relating data $y^e(\mathbf{x})$ with the probabilistic output from $y^{m'}(\mathbf{x},\Theta)$ is established,

$$y^e(\mathbf{x}) = y^{m'}(\mathbf{x},\Theta), \tag{9}$$

based on the assumption that the experimental data $y^e(\mathbf{x})$ can be hypothetically regenerated through the updated model $y^{m'}(\mathbf{x},\Theta)$. Therefore, the likelihood $L(\boldsymbol{\Theta}|\mathbf{y}^e)$ as a function of $\boldsymbol{\Theta}$ conditioned on all observations $\mathbf{y}^e$ is equal to the joint PDF of a N^e dimensional multivariate distribution of $y^{m'}(\mathbf{x},\Theta)$ evaluated at $\mathbf{y}^e$. In this work, we assume the N^e observations are independent, then the likelihood function is the multiplication of N^e separate PDFs, i.e.,

$$L(\Theta) = L(\Theta|\mathbf{y}^e) = p(\mathbf{y}^e|\Theta) = \prod_{i=1}^{N^e} p(y_i^e|\Theta), \tag{10}$$

where $\boldsymbol{\Theta}$ are all model updating parameters to be estimated, $p(y_i^e|\Theta)$ is the value of PDF yielded from $y^{m'}(\mathbf{x},\Theta)$ at (x_i^e, y_i^e). Fig. 3 depicts the plots of output distributions (the PDFs of which are represented by shaded PDF profiles) of two $y^{m'}(\mathbf{x},\Theta)$ models. With the

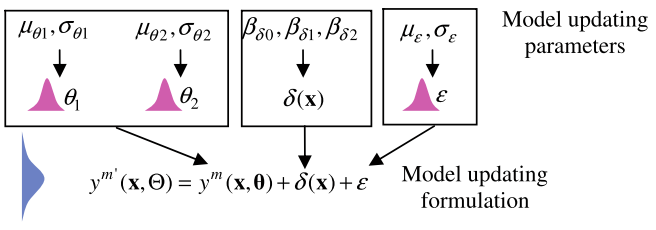

Fig. 2. Model updating parameters $\boldsymbol{\Theta}$ in formulation $y^{m'}(\mathbf{x},\Theta)$.

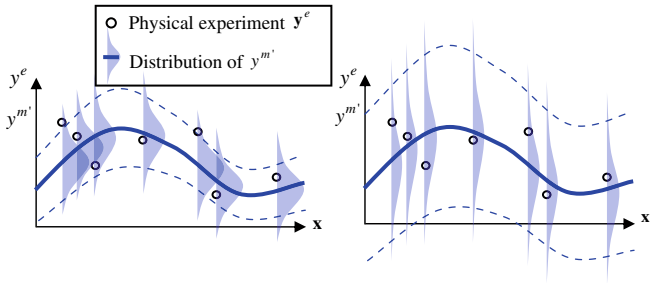

Fig. 3. Likelihood value indicates the agreement between the output distribution of the updated model and the dispersion of physical experiments.

same experimental data (empty circles), the model of the left side figure, which corresponds to a larger likelihood function value, shows a better match between the two distributions.

To alleviate the computational burden, we approximate the output distribution of $y^{m'}(\mathbf{x},\Theta)$ with a Gaussian distribution. As the result, the PDF $p(y_i^e|\Theta)$ in Eq. (10) for experiment y_i^e can be expressed through the Gaussian distribution,

$$y_i^e|\Theta \sim N(E_\Theta\{y^{m'}(\mathbf{x}_i,\Theta)\}, Var_\Theta\{y^{m'}(\mathbf{x}_i,\Theta)\}). \tag{11}$$

To efficiently compute the mean $E_\Theta\{y^{m'}(\mathbf{x}_i,\Theta)\}$ and the variance $Var_\Theta\{y^{m'}(\mathbf{x}_i,\Theta)$ in the above equation, the numerical integration approach based on the tensor-product quadrature rule [24,25] is adopted in this work.

4.3. Comparison of the MLE based model updating with traditional Bayesian approach

With a traditional Bayesian calibration approach [18], for the model updating formulation shown in Eq. (6), the likelihood function $L(\sigma_m,\mathbf{r}_m,\mu_m;\sigma_\delta,\mathbf{r}_\delta;\sigma_\varepsilon;\boldsymbol{\theta}|\mathbf{y}^{em})$ can be obtained from the PDF of a multivariate Gaussian distribution as follows [23]

$$\mathbf{y}^{em}|\sigma_m,\mathbf{r}_m,\mu_m;\sigma_\delta,\mathbf{r}_\delta;\sigma_\varepsilon;\boldsymbol{\theta} \sim N(\mathbf{y}^{em}-\mu_m\mathbf{1},\Sigma_{em}), \tag{12}$$

with covariance matrix Σ_{em} expressed by

$$\Sigma_{em} = \sigma_m^2\mathbf{R}_m + \begin{pmatrix} 0 & 0 \\ 0 & \sigma_\delta^2\mathbf{R}_\delta + \sigma_\varepsilon^2\mathbf{I} \end{pmatrix}, \tag{13}$$

where $\mathbf{y}^{em}$ denotes the joint data $(\mathbf{y}^e,\mathbf{y}^m)$ of computer and physical experiments, $\sigma_m,\mathbf{r}_m,\mu_m$ are the parameters of the GP model that replaces the expensive computer model $y^m(\mathbf{x},\boldsymbol{\theta})$ with $\mathbf{R}_m$ being the correlation matrix; $\sigma_\delta,\mathbf{r}_\delta$ are the parameters of the GP model for $\delta(\mathbf{x})$ with $\mathbf{R}_\delta$ being the correlation matrix; σ_ε is the standard deviation of the experimental error ε; and $\mathbf{I}$ is the identity matrix of the same size of $\mathbf{R}_\delta$. It is noted from Eqs. (12) and (13) that the observed experimental uncertainty of $y^e(\mathbf{x})$ using the Bayesian approach is essentially attributed to three sources, namely, random variable ε, GP model of $\delta(\mathbf{x})$, and GP model of $y^m(\mathbf{x},\boldsymbol{\theta})$.

Based on the above introduction of Bayesian approach, we generalize several major differences between the Bayesian approach and the MLE based model updating approach. First, different types of experimental uncertainty are accounted for in different approaches. With the MLE based model updating approach, the experimental uncertainty is explicitly accounted for by both the random parameters $\boldsymbol{\theta}$ (for experiment-to-experiment variation) and the error term ε. In Bayesian approach, $\boldsymbol{\theta}$ is assumed fixed but unknown due to the lack of knowledge, while only one random parameter ε accounts for the experimental uncertainty, which is caused by random measurement error.

The second difference is associated with the handling of expensive original computer models. In the traditional Bayesian calibration approaches, if an original computer model is expensive, a Gaussian Process model is used to replace it and the GP parameters are estimated using the Bayesian analysis together with other unknown model updating parameters. This adds much computational complexity to the Bayesian approach. With the MLE based model updating approach, a metamodel is first constructed to replace the expensive computer model even before the model updating procedure is initiated. It is then the metamodel, but not the original computer model, that is updated and used for prediction.

4.4. Prediction using the updated model

Once an updated model $y^{m'}(\mathbf{x},\Theta)$ is determined, it is used to form a predictive model $y^{pred}(\mathbf{x})$ for prediction, i.e.,

$$y^{pred}(\mathbf{x}) = y^{m'}(\mathbf{x},\Theta). \tag{14}$$

The prediction uncertainty of $y^{pred}(\mathbf{x})$ is evaluated by propagating the uncertainties defined by the model updating parameters $\boldsymbol{\Theta}$ through the updated model $y^{m'}(\mathbf{x},\Theta)$, which by itself is deterministic. With our approach, since the original expensive computer model $y^m(\mathbf{x},\boldsymbol{\theta})$ is replaced by a metamodel, uncertainty propagation can be done in a rather efficient manner using a combination of Monte Carlo simulations and numerical methods, given the mean and standard deviation of $\boldsymbol{\theta}$ estimated from MLE. As a result, the prediction incorporates the uncertainties involved in a model updating and validation process.

5. Comparative studies using the thermal challenge problem

5.1. Problem description

The thermal challenge problem was developed by the Sandia National Laboratory (http://www.esc.sandia.gov/VCWwebsite/vcwhome.html) as a testbed for demonstrating various methods of prediction and validation [7,8]. The same problem is adopted in this work to demonstrate the features of our model updating approach. A schematic figure of the thermal challenge problem is provided in Fig. 4, in which the device is a material layer that is exposed to a constant heat flux. To predict the temperature T^m of a spot in the device at a specific location and time, an analytical computer model $y^m(\mathbf{x},\boldsymbol{\theta})$ given in Eq. (15) was used as the original model.

$$\begin{aligned} y^m(\mathbf{x},\boldsymbol{\theta}) &= T^m(q,L,x_l,t,\kappa,\rho) \\ &= T_0 + \frac{qL}{\kappa}\left[\frac{(\kappa/\rho)t}{L^2} + \frac{1}{3} - \frac{x_l}{L} + \frac{1}{2}\left(\frac{x_l}{L}\right)^2 \right. \\ &\quad \left. - \frac{2}{\pi^2}\sum_{n=1}^{6}\frac{1}{n^2}e^{-n^2\pi^2\frac{(\kappa/\rho)t}{L^2}}\cos\left(n\pi\frac{x_l}{L}\right)\right], \end{aligned} \tag{15}$$

where $\mathbf{x} = (q,L,x_l,t)$ are controllable input variables, and $\boldsymbol{\theta} = (\kappa,\rho)$ are uncontrollable input parameters for calibration. Among the controllable input variables, x_l is the distance from the surface to the point being measured, q is the applied heat flux, L is the thickness, T_0 is the initial temperature of the device at time zero, and t is time.

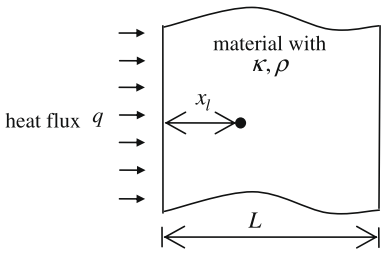

Fig. 4. Schematic of thermal challenge problem.

Since T_0 is fixed at 25 °C for all data and analyses, it is considered as a static model parameter instead of an input variable. Among the uncontrollable input parameters, κ and ρ stand for the thermal conductivity and the volumetric heat capacity, respectively; both κ and ρ are material properties, the values of which may vary from unit to unit due to the variability in the manufacturing process. The goal of this challenge problem is to assess if a regulatory requirement is satisfied for a specified setting of model input, i.e., at an application site $\mathbf{x}^* = (q^*, L^*, x_l^*, t^*)$. To satisfy the regulatory requirement, the probability that the predicted temperature at a particular time not exceeding a threshold value 900 °C, should be less than the target probability limit (0.01), that is,

$$P\{T^{pred}(q^* = 3500, L^* = 0.019, x_l^* = 0, t^* = 1000) > 900\} < 0.01. \tag{16}$$

As described in the original problem statement, the prior knowledge about κ, ρ are given in the form of material property characterization (MPC) data. Note that the measurements of MPC data are collected under different temperatures T at the material subcomponent level rather than at the device level. Therefore the information of MPC cannot be directly used for device level prediction. The first two moments of the distributions of the above two parameters are evaluated based on the MPC data and summarized in Table 1. The linear dependency of κ versus temperature T can be easily observed (Fig. 5). Since temperature T is also an output response of the device-level model (Eq. (15)), this creates a closed-loop situation where T is both an input and output of the model.

The full set of physical experiments consist of a subset of 'ensemble' (EN) data of 176, and a subset of 'accreditation' (AC) data of 126. The primary difference between the EN data and the AC data is that the former is gathered at model input sites far from the application site, while the latter is closer. Both EN and AC data are collected at several configurations (different settings of $\mathbf{x}$), and each configuration is a combination of model inputs (q, L, x_l), while t varies at discrete time interval spots (11 spots for EN, 21 spots for AC) within the range 0–1000. For each configuration, data is collected respectively for 1–4 times. Table 2 lists the values of these configurations. The EN data and the AC data are selectively used at three levels of data sufficiency, namely, low, medium, and high. The sizes of EN data and AC data are considered at three different levels of sufficiency: 44 + 63 for low level, 88 + 63 for medium level, and 176 + 126 for high level. In this work, we use EN and AC data at the high level of sufficiency by default.

Table 1
Statistics of the given material property characterization (MPC) data.

κ		ρ	
μ_κ	σ_κ	μ_ρ	σ_ρ
0.0628	0.0099	393900	36252

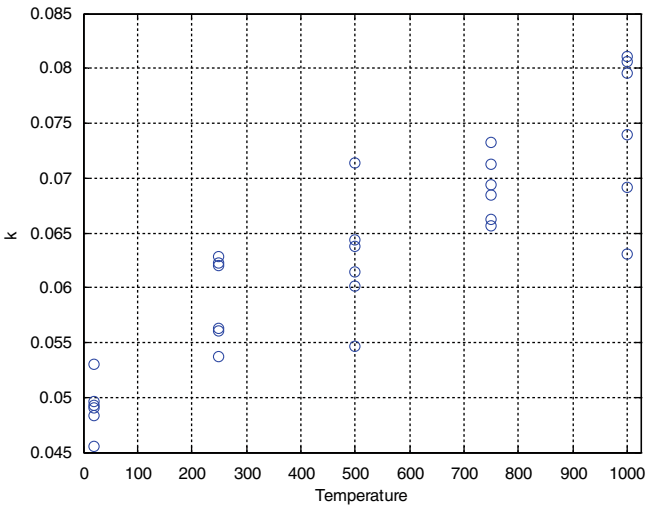

Fig. 5. The dependency of thermal conductivity κ to temperature T.

Table 2
Statistics of the given characterization data.

Data set	Config #	(q, L, x_l)	
EN	Config 1	$x_l = 0$	$q = 1000$, $L = 1.27$
	Config 2		$q = 1000$, $L = 2.54$
	Config 3		$q = 2000$, $L = 1.27$
	Config 4		$q = 2000$, $L = 2.54$
AC	Config 5.1	$x_l = 0$	$q = 3000$, $L = 1.9$
	Config 5.2	$x_l = L/2$	
	Config 5.3	$x_l = L$	

5.2. Bayesian approaches to the thermal challenge problem

Several different approaches [20,26–31] have been developed and presented in the literature on the thermal challenge problem as a result from the Sandia Validation workshop. We find these methods differ in how they utilize three different data sources (MPC, EN and AC data), the model updating formulations (e.g. including bias or not), and the solution method (e.g. Bayesian or non-Bayesian). Additionally, some of these works focus on prediction (whether the regulatory requirement will be met), while some others also study the model validity (accuracy). The readers should refer to Hills et al. [27] for a complete summary of the existing approaches.

In terms of the solution method, the Bayesian calibration methodology of Kennedy and O'Hagan [3] was followed by Higdon et al. [26], Liu et al.[32], McFarland and Mahadevan [28] to calibrate κ andρ with the bias function included. In Refs. [32,26], no formal model validation is considered, while in Ref. [28], model validation metrics based on the significance test are employed. These works use different prior specifications of Gaussian Process hyperparameters and assume different prior distributions for calibration parameters κ and ρ. Refs. [32,26] assign prior distributions for parameters κ and ρ based on either full or partial MPC data. Ref. [28] specifies vague priors for κ and ρ without using any MPC data. By utilizing both the EN and AC data, the predicted failure probabilities are determined to be 0.03 for using all levels of data sufficiency in Ref. [26], and are determined to be 0.02 and 0.04, respectively for using the medium and high levels of data sufficiency in Ref. [32]. In Ref. [28], the failure probability is computed as 0.166 using the high level of data sufficiency. All studies indicate that the thermal device cannot meet the regulatory requirement (<0.01) as specified in Eq. (16).

5.3. Three model updating formulations for testing the MLE method

In this study, we test the MLE based model updating approach with three different model updating formulations. The formulations of the updated model $y^{m'}(\mathbf{x}, \Theta)$ and the corresponding model updating parameters $\mathbf{\Theta}$ are listed in Table 3.

In all formulations, we assume that the uncontrollable output variable ε is a zero-mean random variable ($\mu_\varepsilon = 0$). Formulation (2) is exactly the one used in Eq. (8) to explain the MLE method first described in Section 4. While Formulation (1) is the simplest updating formulation, the bias function $\delta(\mathbf{x})$ is introduced in both Formulations (2) and (3), where a first-order polynomial regression model, governed by parameters $\beta_{\delta 0}, \beta_{\delta 1}, \ldots, \beta_{\delta 4}$, is used to represent $\delta(\mathbf{x})$. To capture the linear dependency of κ versus the temperature

Table 3
Model updating formulations and model updating parameters.

Model form. #	$y^{m'}(\mathbf{x},\Theta)$	Model updating parameters $\boldsymbol{\Theta}$		
		For $\boldsymbol{\theta}$ (i.e.,κ and ρ)	For ε	For $\delta(\mathbf{x})$
(1)	$y^m(\mathbf{x},\boldsymbol{\theta})+\varepsilon$	μ_κ, σ_κ,μ_ρ,σ_ρ	μ_ε(=0), σ_ε	
(2)	$y^m(\mathbf{x},\boldsymbol{\theta})+\delta(\mathbf{x})+\varepsilon$	μ_κ, σ_κ, μ_ρ, σ_ρ	μ_ε(=0), σ_ε	$\beta_{\delta 0}$, $\beta_{\delta 1},\ldots,\beta_{\delta 4}$
(3)	$y^m(\mathbf{x},\boldsymbol{\theta}(\mathbf{x}))+\delta(\mathbf{x})+\varepsilon$	$\beta_{\kappa 0},\beta_{\kappa 1}$, σ_κ, μ_ρ, σ_ρ	μ_ε(=0), σ_ε	$\beta_{\delta 0},\beta_{\delta 1},\ldots,\beta_{\delta 4}$

T observed from the MPC data (Fig. 5), in Formulation (2), we assume that $\boldsymbol{\theta}$ is a function of $\mathbf{x}$, i.e., $\boldsymbol{\theta}=\boldsymbol{\theta}(\mathbf{x})$. Given that the temperature field of the thermal device is primarily influenced by the distance (x_l) from the surface to the measured point, we further simplify the function of $\boldsymbol{\theta}(\mathbf{x})$, by modeling the mean of κ, namely μ_κ, as linearly dependent on x_l (rather than all $\mathbf{x}$) through a linear model $\mu_\kappa=\beta_0+\beta_1 x_l$.

5.4. Results of model updating parameters for different formulations

Using the MLE method described in Section 4.2, the model updating parameters for each formulation are obtained based on the selected data from the given data sets EN and AC. To study the extrapolation capability of the updated model, three scenarios are considered. For Scenario 1, only EN data set is included; for Scenario 2, only AC data set is included; For Scenario 3, both EN and AC data sets are included. In searching for model updating parameters $\boldsymbol{\Theta}$ via MLE optimiation, the mean and variance values from the MPC information in Table 1 are utilized to provide search bounds. In this example, we use relatively loose bounds by multiplying a factor 0.1–10. For example, based on $\mu_\kappa=0.0628$ and $\sigma_\kappa=0.0099$ in Table 1, the bounds applied in MLE optimiation are 0.00628–0.628 for μ_κ and 0.00099–0.099 for σ_κ.

For the purpose of demonstration, we provide in Table 4 only the results of the model updating parameters under Scenario 1 with high level data sufficiency. These results provide statistical representations of model updating parameters, which will be used further to characterize the uncertainty of model response in both prediction and validation.

5.5. Studying the predictive capability of updated models

We use the results from Formulation (2) as an example to demonstrate how the predictive capability of an updated model can be studied. As a reference, we first show the results from the 'Original' model without model updating (Eq. (15)), but with the consideration of uncertainty of model parameters κ and ρ as observed from the MPC data. Fig. 6 shows the predicted response with uncertainty yielded by the 'Original' model at two selected configuration sites:

Table 4
Estimated model updating parameters (Scenario 1).

Model #	κ			ρ		ε	$\delta(\mathbf{x})$
	μ_κ		σ_κ	μ_ρ	σ_ρ	σ_ε	$\beta_{\delta 0},\beta_{\delta 1},\ldots,\beta_{\delta 4}$
(1)	0.0579		0.00099	387,026	12,266	9.8001	N/A
(2)	0.0493		0.00099	399,822	22,210	5.2399	14.751, 118.593, −0.010, −663.605, 0
	$\beta_{\kappa 0}$	$\beta_{\kappa 1}$					
(3)	0.0508	0.025850	0.00110	391,171	20,549	5.4833	14.293, 176.377, −0.010, −606.117, 0

Config 1 and Config 5.1 (defined in Table 2). Uncertainty of the prediction is represented by the 95% prediction intervals (PIs). Also plotted in the figures are the validation data collected (EN data for Config 1 at left; AC data for Config 5.1 at right). It is observed that for this particular problem, even without model updating, uncertainty predictions based on the MPC data encompass the physical observations quite well. However, the magnitude of the prediction uncertainty at Config 5.1 (the accreditation site) appears to be much larger than the true dispersion observed from data.

Figs. 7 and 8 show the predictions using the updated model based on Formulation (2) under data Scenario 1 and Scenario 2, respectively. Similar to Fig. 6, we only show data at Config 1 and Config 5.1. In Fig. 7, the EN data is used for model updating while the EN data ('in-sample' test) and the AC data ('out-sample' test) are used separately to validate the updated model (Scenario 1). In Fig. 7, the AC data is used for model updating while EN ('out-sample' test) and AC ('in-sample' test) are used separately as validation data (Scenario 2).

In comparing the left plot in Fig. 7 with the left plot in Fig. 6, and then the right plot in Fig. 8 with the right plot in Fig. 6, we observe that after using the MLE method for model updating, the predictions with uncertainty quantification match much better with what observed in the physical data, i.e., the uncertainty bandwidth is significantly reduced to match with the dispersion of physical data. Such an improvement is accredited to using MLE as the optimization criterion for determining the model updating parameters. The right plot in Fig. 7 and the left plot in Fig. 8 (both for out-sample tests) show the predictive capability of the updated models. It is found that the predictions in the out-sample tests are not as good as those in the in-sample tests, and somewhat worse than those made by the 'Original' model (Fig. 6).

5.6. Model validation metrics

As shown in the proposed model updating framework (Fig. 1), a model validation metric is needed to assess the validity of either the original model $y^m(\cdot)$ or the updated model $y^{m'}(\cdot)$. The metric is also useful for comparing the effectiveness of multiple model updating formulations. Although various validation metrics are widely studied in the literature[13,33,34], most of them are limited to comparing the mean prediction of a computer model and the mean performance from the physical observations [9]. This differs from the principle of model updating in this work, i.e., to match the dispersion of physical experiments with the distributions of computer model output.

In this paper, we adopt the u-pooling method recently developed by Ferson et al. [9] for model validation. A nice feature of the u-pooling method is that it allows to integrate or pool all available physical experiments over a validation domain at different input settings $\mathbf{x}$ into a single aggregate metric. First, a value u_i is obtained for each experiment by calculating the CDF at y_i^e, i.e., $u_i=F_{\mathbf{x}_i^e}(y_i^e)(i=1,\ldots,N^e)$, where y_i^e represents a physical observation at the experimental site $\mathbf{x}_i^e(i=1,\ldots,N^e)$. $F_{\mathbf{x}_i^e}(y)$ represents the corresponding CDF in our case generated by the updated model $y^{m'}(\mathbf{x},\Theta)$ at $\mathbf{x}_i^e$. The distribution of u_i could be characterized after pooling all values of u_i for all physical experiments. According to Ferson et al., if each physical observation y_i^e hypothetically comes from the same 'mother' distribution $F_{\mathbf{x}_i^e}(y)$, all u_i's are expected to constitute a standard uniform distribution on [0,1]. An illustration of the u-pooling method with three ($N^e=3$) experimental sites is given in Fig. 9. By comparing the empirical distribution of u_i to that of the standard uniform distribution, the area difference (depicted as the shaded region in Fig. 9) can be used to quantify the mismatch between the dispersion of physical experiments and the distributions of model output. The larger the difference, the less agreement, and therefore the less accuracy of a model.

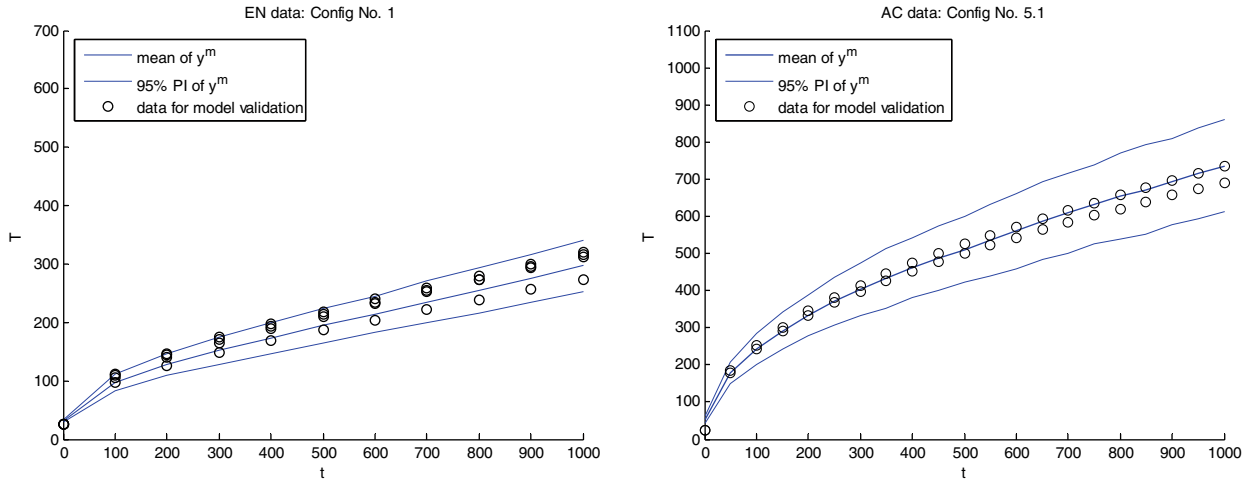

Fig. 6. Prediction through the 'Original' model at Config 1 and Config 5.1.

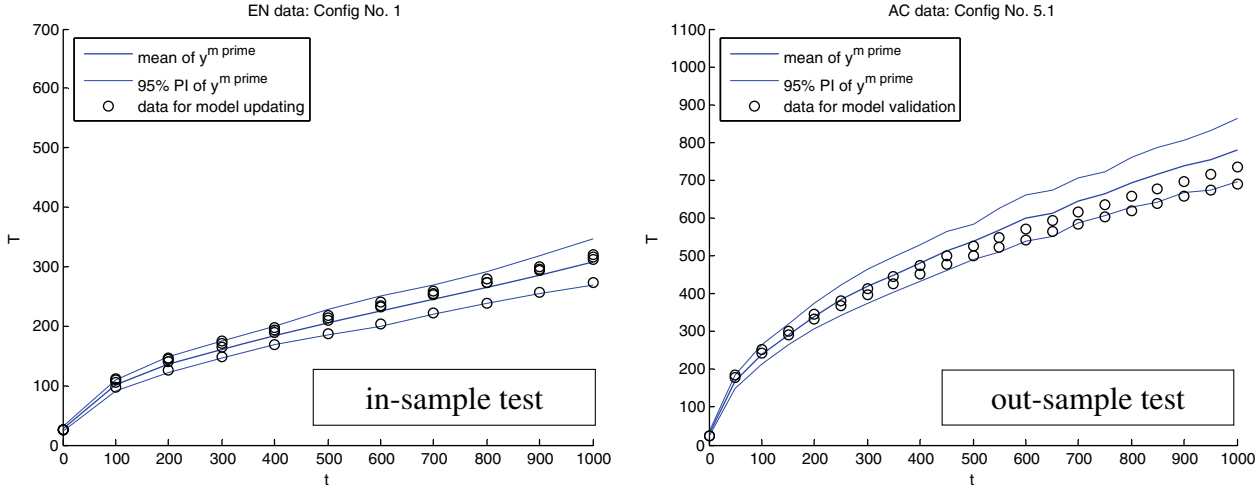

Fig. 7. Prediction through Formulation (2) (Scenario 1).

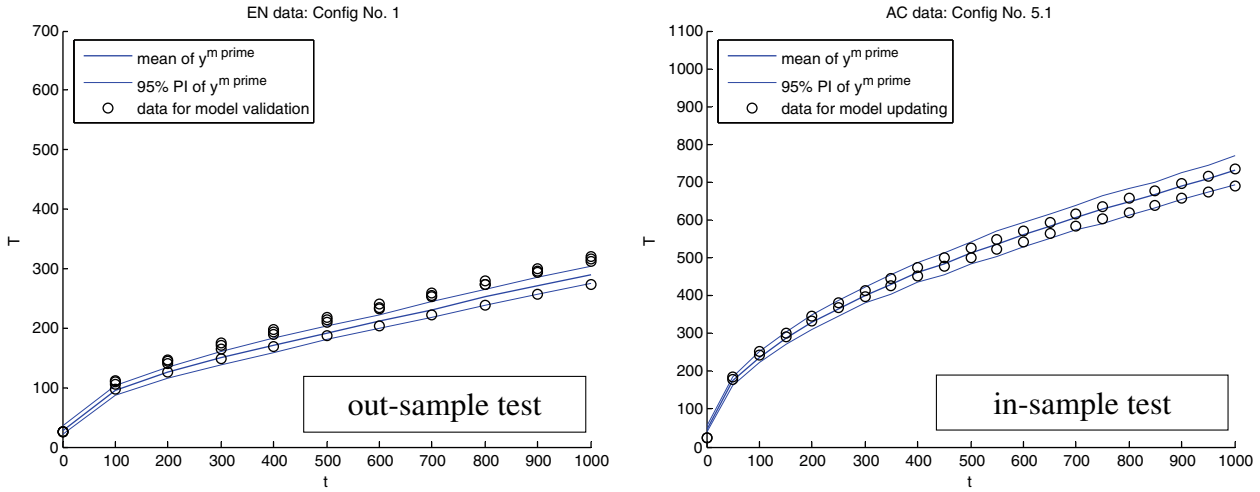

Fig. 8. Prediction through Formulation (2) (Scenario 2).

In addition to the u-pooling metric, we also employ the Rooted Mean Square Error (RMSE), a traditional accuracy metric, to assess the goodness-of-fit in terms of the mean prediction. In calculating the RMSE, the residual error e_i is the difference between y_i^e and the mean of the updated model $y^{m'}(\mathbf{x}, \Theta)$ at $\mathbf{x}_i^e$, i.e., $e_i = y_i^e - E_\Theta y^{m'}(\mathbf{x}_i^e, \Theta)$.

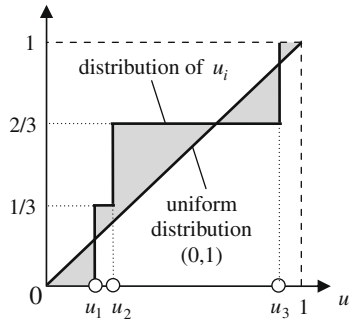

Fig. 9. Illustration of the u-pooling method.

5.7. Comparing model updating formulations based on validation metrics

For comparing the validity (accuracy) and the predictive capability of the three different model updating formulations, we provide in Tables 5–7 the model validation results under different scenarios, all with the high data sufficiency level. Both the 'u-pooling' and the RMSE metrics are considered in comparison; smaller values are desired for both metrics. Under Scenario 1 (Table 5) and Scenario 2 (Table 6), one set of data (either EN or AC) is used for model updating, called in-sample test; while the other set is used for verifying the predictive capability, called out-sample test. Under Scenario 3 (Table 7), since both EN and AC data are used in model updating, validation is performed over the joint data set, which is essentially an 'in-sample' test. For comparison, the metrics are also evaluated for the original model, with the uncertainty of parameters κ and ρ characterized directly from the material property characterization (MPC) data without updating the model itself. The results are summarized in the row marked with 'Original'.

Table 5
Summary of model validation metrics (Scenario 1).

Model #	Validation data: EN (in-sample test)		Validation data: AC (out-sample test)	
	u-Pooling	RMSE	u-Pooling	RMSE
Original	0.634	16.96	0.830^{**}	29.13^{*}
(1)	0.566	15.12^{*}	1.138	33.36
(2)	0.521^{**}	15.07^{**}	0.901^{*}	19.43^{**}
(3)	0.579	15.16	1.041	31.39

Table 6
Summary of model validation metrics (Scenario 2).

Model #	Validation data: EN (out-sample test)		Validation data: AC (in-sample test)	
	u-Pooling	RMSE	u-Pooling	RMSE
Original	0.634^{**}	16.96^{**}	0.830	29.13^{*}
(1)	0.891^{*}	17.87^{*}	0.540^{*}	11.27^{**}
(2)	0.813^{*}	18.14	0.471^{**}	11.24^{**}
(3)	1.002	19.53	0.463^{**}	11.40

Table 7
Summary of model validation metrics (Scenario 3).

Model #	Validation data: EN + AC (in-sample test)	
	u-Pooling	RMSE
Original	0.456^{*}	22.84
(1)	0.420^{**}	14.98^{**}
(2)	0.388^{**}	14.29^{*}
(3)	0.429	14.60^{*}

Table 8
Summary of predicted failure probability (Scenario 3).

	Original	Model #		
		(1)	(2)	(3)
$P\{T^{pred}(\mathbf{x}^*) > 900\}$	0.26	0.060	0.028	0.092

In Tables 5 and 6, the best and the second best 'u-pooling' or RMSE values in any single column across different model formulations are marked out with '**' and '*' respectively. In Table 5, when the EN data is used for model updating, Formulation (2) is found to be the best in the in-sample test in terms of both 'u-pooling' and RMSE values. In the out-sample test, Formulation (2) ranks the second best under 'u-pooling' and the best under RMSE. This indicates that when the EN data is used for model updating, Formulation (2), which uses constant calibration parameter $\boldsymbol{\theta}$ and bias function $\delta(\mathbf{x})$, can best adapt to the data with acceptable extrapolation capability. In Table 6, when the AC data is used for model updating, Formulation (2) again wins over others in the in-sample test. However, in the out-sample test, no model updating formulation is superior to the 'Original' model, which indicates that, although an updated model favors the data it used, the extrapolation should be treated with caution. In Table 7, where EN data and AC data are both incorporated in model updating, Formulation (2) is the best over other formulations and the 'Original' model. Overall, Formulation (2) achieves the best performance over other formulations and the 'Original' model. One common observation in Tables 5–7 is that all three model updating formulations are better than the 'Original' in all in-sample test columns. It is interesting to note that Formulations (3), in which one calibration parameter is considered as a function of one model input, do not bring a significant benefit as we expect based on the physical principle.

5.8. Comparison of the predicted failure probability in regulatory test

Based on the study in Section 5.7 and the fact that the application site $\mathbf{x}^*$ is close to the domain of AC data, it is determined that both the EN and AC data sets should be used for updating the model (i.e., Scenario 3) which will be further used to make the final prediction at $\mathbf{x}^*$ in the regulatory test. All data in EN and AC (i.e., high level data sufficiency) is considered. Model prediction introduced in Section 4.4 is used for evaluating the regulatory requirement stated in Eq. (16). To assess the failure probability, 1000 Monte Carlo simulations are used for propagating the parameter uncertainty determined by the model updating parameters which are identified by the MLE approach. Table 8 shows the estimated failure probabilities using different model updating formulations including the 'Original' model. The specified threshold value (0.01) is exceeded in all cases.

It is found that our estimations of failure probabilities are consistent with the results reported in the other works [20,26] on the thermal challenge problem. Considering that Formulation (2) achieves the best overall accuracy over others, our best estimation of failure probability is 0.028.

6. Closure

In this work we examine various strategies for model updating and study its relationship with model validation activities. The maximum likelihood estimation (MLE) method is introduced as an alternative approach to the traditional Bayesian method to estimate the model updating parameters, so that it seeks optimal distribution parameters underlying model updating parameters

through MLE. Unlike the traditional Bayesian approach which treats calibration parameters as fixed but unknown due to lack of knowledge, the MLE based approach treats calibration parameters as intrinsic random to account for the uncertainty due to experiment-to-experiment variability. Other differences of the two methods are summarized in Section 4.3 and will not be repeated here.

Through the thermal challenge example, we demonstrate that model updating can be used to refine a computer model based on the physical observations gathered before the validation metrics are applied and the model is used for prediction. Our presented model updating formulations explicitly account for various sources of uncertainties in the associated process. We illustrate that without running into numerical complexity, the MLE based model updating method is easier to implement and interpret compared to the existing Bayesian methods. Using the newly developed u-pooling method by Ferson et al, we show that the validation metric can be applied to both the original and the updated models to assess the accuracy and predictive capability of different model updating formulations. Our study also provides insights into the potential benefits and limitations of using model updating for improving the predictive capability of a model. Through in-sample and out-sample tests based on different data sets, we observe that the model updating approach used in this work improves the agreement between the model and the physical experiment data. However, when applying the updated model at a region that is far from the domain where data is used for model updating, the extrapolation capability of the updated model is not guaranteed. By comparing our approach to the existing works on the thermal challenge problem, we point out the differences of various methods in utilizing available data, the model updating formulations adopted, and the solution method employed. Even though our method is different from other works in the literature for solving the thermal challenge problem, we find the conclusion we reach on device failure probability is very close to other estimations reported. As for which model updating formulation is the most appropriate, we think it is problem dependent and should be selected by exercising the model validation metrics as demonstrated. While model updating is shown to be useful for improving the accuracy of a model, as the process is fully data-driven, we believe the method should be used with caution, especially when used for extrapolation.

Due to the nature of the MLE method, its effectiveness and accuracy may be downgraded when the data amount is extremely small. In our test with the 'low level' data sufficiency for the thermal challenge problem, it is found that the bandwidth of the prediction uncertainty could be degenerated to fairly small values. To mitigate this problem, prior knowledge may be used to specify more conservative bounds of model updating parameters to prevent them from running into 'absurd' values. Another potential weakness of the MLE based model updating approach might be associated with the numerical instability when optimizing the likelihood function, especially when a complex model updating formulation that involves many parameters is considered. To mitigate this issue, sensitivity analysis could be performed prior to MLE optimization to leave out parameters that are insensitive to model output and the likelihood function.

In this research, none of the model updating strategies studied account for the uncertainty due to insufficient data. Our future research is to investigate how we may quantify the impact of lack of data and provide decision support to resource allocation in planning physical experiments.

Acknowledgements

The grant support from National Science Foundation (NSF) to this collaborative research between Northwestern University (CMMI-0522662) and Georgia Tech (CMMI-0522366) is greatly appreciated.

References

[1] W.L. Oberkampf, T.G. Trucano, C. Hirsch, Verification, validation, and predictive capability in computational engineering and physics, Appl. Mech. Rev. 57 (5) (2004) 345–384.
[2] N. Leoni, C.H. Amon, Bayesian surrogates for integrating numerical, analytical, and experimental data: application to inverse heat transfer in wearable computers, IEEE Trans. Compon. Pack. Technol. 23 (1) (2000) 23–32.
[3] M.C. Kennedy, A. O'Hagan, Bayesian calibration of computer models, J. Roy. Statist. Soc., Ser. B 63 (2001) 425–464.
[4] R. Ghanem, A. Doostan, On the construction and analysis of stochastic models: characterization and propagation of the errors associated with limited data, J. Comput. Phys. 217 (2) (2006) 63–81.
[5] V.J. Romero, Validated model? Not so fast. The need for model "conditioning" as an essential addendum to model validation, in: 48th AIAA/ASME/ASCE/AHS/ASC Structures, Structural Dynamics, and Materials Conference, Honolulu, Hawaii, April 23–26, 2007.
[6] A.C. Tamhane, D.D. Dunlop, Statistics and Data Analysis: From Elementary to Intermediate, Prentice-Hall, Upper Saddle River, NJ, 2000.
[7] K.J. Dowding, M. Pilch, R.G. Hills, Formulation of the thermal problem, Comput. Methods Appl. Mech. Engrg. 197 (29–32) (2008) 2385–2389.
[8] R.G. Hills, M. Pilch, K.J. Dowding, I. Babuska, R. Tempone, Model validation challenge problems: tasking document, Comput. Methods Appl. Mech. Engrg. 197 (29–32) (2008) 2375–2380.
[9] S. Ferson, W.L. Oberkampf, L. Ginzburg, Model validation and predictive capability for the thermal challenge problem, Comput. Methods Appl. Mech. Engrg. 197 (29–32) (2008) 2408–2430.
[10] R.G. Easterling, J.O. Berger, Statistical Foundations for the Validation of Computer Models, Presented at Computer Model Verification and Validation in the 21st Century Workshop, Johns Hopkins University, 2002.
[11] T. Hasselman, K. Yap, C. Lin, J. Cafeo, A case study in model improvement for vehicle crashworthiness simulation, in: 23rd International Modal Analysis Conference, Orlando, Florida, January 31–February 3, 2005.
[12] W. Chen, Y. Xiong, K.-L. Tsui, S. Wang, Some metrics and a Bayesian procedure for validating predictive models in engineering design. ASME Design Technical Conference, Design Automation Conference, Philadelphia, PA, September 10–13, 2006.
[13] S. Wang, W. Chen, K. Tsui, Bayesian Validation of Computer Models. Revision resubmitted to Technometrics, 2008.
[14] Z. Qian, C.F.J. Wu, Bayesian hierarchical modeling for integrating low-accuracy and high accuracy experiments, in: Twelfth Annual International Conference on Statistics, Combinatorics, Mathematics and Applications, Auburn, AL, December 2–4, 2005.
[15] D.M. Bates, D.G. Watts, Nonlinear Regression Analysis and Its Application, Wiley, New York, 1988.
[16] T.G. Trucano, L.P. Swilera, T. Igusab, W.L. Oberkampf, M. Pilch, Calibration, validation, and sensitivity analysis: What's what, Reliab. Engrg. Syst. Safe. (91) (2006) 1331–1357.
[17] L.-E. Lindgren, H. Alberg, K. Domkin, Constitutive modelling and parameter optimization, in: 7th International Conference on Computational Plasticity, Barcelona, Spain, April 7–10, 2003.
[18] M.J. Bayarri, J.O. Berger, D. Higdon, M.C. Kennedy, A. Kottas, R. Paulo, J. Sacks, J.A. Cafeo, J. Cavendish, C.H. Lin, J. Tu, A Framework for Validation of Computer Models, Foundations for Verification and Validation in 21st Century Workshop, Johns Hopkins University, 2002.
[19] D. Higdon, M. Kennedy, J. Cavendish, J. Cafeo, R.D. Ryne, Combining field observations and simulations for calibration and prediction, SIAM J. Sci. Comput. 26 (2004) 448–466.
[20] F. Liu, M.J. Bayarri, J. Berger, R. Paulo, J. Sacks, A Bayesian analysis of the thermal challenge problem, Comput. Methods Appl. Mech. Engrg. 197 (29–32) (2008) 2457–2466.
[21] J.M. McFarland, S. Mahadevan, L.P. Swiler, A.A. Giunta, Bayesian calibration of the QASPR simulation, in: 48th AIAA/ASME/ASCE/AHS/ASC Structures, Structural Dynamics, and Materials Conference, Honolulu, Hawaii, April 23–26, 2007.
[22] J.M. Aughenbaugh, J.W. Herrmann, Updating uncertainty assessments: a comparison of statistical approaches, in: ASME International Design Engineering Technical Conferences and Computers and Information in Engineering Conference, Las Vegas, Nevada, September 4–7, 2007.
[23] L.J. Loeppky, D. Bingham, W.J. Welch, Computer Model Calibration or Tuning in Practice, Technometrics, submitted for publication.
[24] M. Abramowitz, I.A. Stegun, Handbook of Mathematical Functions, 10th ed., Dover, New York, 1972.
[25] S.H. Lee, H.S. Choi, B.M. Kwak, Multi-level design of experiments for statistical moment and probability calculation, Struct. Multidiscip. Optimiz. (2007).
[26] D. Higdon, C. Nakhleh, T. Gattiker, B. Williams, A Bayesian calibration approach to the thermal problem, Comput. Methods Appl. Mech. Engrg. 197 (29–32) (2008) 2431–2441.
[27] R.G. Hills, K.J. Dowding, L. Swiler, Thermal challenge problem: summary, Comput. Methods Appl. Mech. Engrg. 197 (29–32) (2008) 2490–2495.

[28] J. McFarland, S. Mahadevan, Multivariate significance testing and model calibration under uncertainty, Comput. Methods Appl. Mech. Engrg. 197 (2008) 2467–2479.
[29] M.D. Brandyberry, Thermal problem solution using a surrogate model clustering technique, Comput. Methods Appl. Mech. Engrg. 197 (29–32) (2008) 2390–2407.
[30] B.M. Rutherford, Computational modeling issues and methods for the 'regulatory problem' in engineering – solution to the thermal problem, Comput. Methods Appl. Mech. Engrg. 197 (29–32) (2008) 2480–2489.
[31] R.G. Hills, K.J. Dowding, Multivariate approach to the thermal challenge problem, Comput. Methods Appl. Mech. Engrg. 197 (29–32) (2008) 2442–2456.
[32] F. Liu, M.J. Bayarri, J. Berger, R. Paulo, J. Sacks, A Bayesian analysis of the thermal challenge problem, Comput. Methods Appl. Mech. Engrg. 197 (29–32) (2008) 2457–2466.
[33] S. Mahadevan, R. Rebba, Validation of reliability computational models using Bayes networks, Reliab. Engrg. Syst. Safe. 87 (2005) 223–232.
[34] W. Oberkampf, M. Barone, Measures of agreement between computation and experiment: validation metrics, J. Comput. Phys. 217 (1) (2006) 5–36.