



# A bayesian model calibration under insufficient data environment

Jeonghwan Choo<sup>1,2</sup> · Yongsu Jung<sup>1</sup> · Ikjin Lee<sup>1</sup>

Received: 24 October 2021 / Revised: 4 February 2022 / Accepted: 10 February 2022 / Published online: 28 February 2022  
© The Author(s), under exclusive licence to Springer-Verlag GmbH Germany, part of Springer Nature 2022

## Abstract

In recent years, remarkable advances in computing performance and computer-aided engineering have enabled reliability-based design optimization (RBDO) to guarantee the target reliability of a product. For successful product development through RBDO, it is indispensable to clarify uncertainties of unknown model variables. In most cases, however, due to cost and time constraints, there are not enough test data, which can lead to a less reliable optimum. For this reason, the primary purpose of this study is to propose a pragmatic approach to perform an inverse uncertainty quantification or a statistical model calibration more accurately and efficiently under an insufficient data environment. Based on the Bayesian model calibration framework, the proposed method consists of two main steps: (1) prior distribution prediction using output (i.e., component) test data and (2) posterior distribution prediction using input (i.e., coupon) test data. In the prior distribution prediction step, the maximum likelihood estimate (MLE) is used to obtain the estimated statistical parameters, the distribution type of unknown model variables, and the Fisher information matrix (FIM) to calculate variances of the estimated statistical parameters. The posterior distribution prediction step utilizes the Bayes' theorem, which combines the prior distribution with the likelihood obtained by reflecting the input test data into the probability density of the estimated unknown model variable. During this process, each test data that is insufficient to directly model or indirectly predict the probability density of the unknown model variable can be integrated to address the crucial issue of the insufficient data effectively. Mathematical and engineering examples are utilized to validate the proposed method for quantification of unknown model variables.

**Keywords** Statistical model calibration · Maximum likelihood estimate · Fisher information matrix · Area metric · Bayesian inference · Epistemic uncertainty

## List of symbols

**d** A controllable design variable vector  
 $z^s(\cdot)$  Response function of a simulation model  
 **$\xi$**  A known model variable vector  
 $\varphi(\cdot)$  Function as a calibration metric  
 **$\theta$**  An unknown model variable vector  
 $L(\cdot)$  Likelihood function  
 **$\Theta$**  A vector of the statistical parameters of the unknown model variables  
 $l(\cdot)$  Log-likelihood function

$\zeta$  Distribution type of the unknown model variable  
 $p(\cdot)$  Posterior PDF  
 $\delta(\cdot)$  A discrepancy function between the experimental and the simulation model  
 $\pi(\cdot)$  Prior PDF  
 $\varepsilon$  A measurement error  
 $\hat{F}_{x_i}(\cdot)$  CDF of an unknown model variable at calibration site  $x_i$   
 **$\mathbf{y}^e$**  An output test data vector  
 $U_{a,m_v}^m(\cdot)$  PDF of area-metric calculated by taking  $m_v$  observations from the estimated probability model  
 **$\mathbf{x}^e$**  An input test data vector  
 $T_{n_v}(\alpha)$  Threshold of  $\alpha$  significance level for  $n_v$  validation samples  
 $y_i^e, y_j^e$  Output test data for calibration and validation, respectively  
 $n_c, n_v$  The quantity of output observations for calibration and validation, respectively

Responsible Editor: Chao Hu

✉ Ikjin Lee  
ikjin.lee@kaist.ac.kr

<sup>1</sup> Department of Mechanical Engineering, Korea Advanced Institute of Science and Technology, Daejeon 34141, South Korea

<sup>2</sup> CAE, Global Brake BU, Mando Corporation, Gyeonggi-do, Seongnam 13486, South Korea

$x_i^e, x_j^e$	Input test data for calibration and validation, respectively
$m_c, m_v$	The quantity of input observations for calibration and validation, respectively
$\bar{\mathbf{I}}$	Expected Fisher information matrix in entire observations
$k$	Number of unknown model variables
$\mathbf{I}$	Observed Fisher information matrix in entire observations
$u_j$	The u-value; CDF value at validation site $x_j$
$\mathbf{I}_1$	The Fisher information matrix in a single observation
$u_{a,m_v}^{(j)}$	$j$ -th area-metric calculated by taking $m_v$ observations from the estimated probability model
$a, b$	The model parameters of the given parametric PDF
$u_{a,m_v}^m$	Area-metric value calculated by taking $m_v$ observations from the estimation
$z^e(\cdot)$	Response function of experiment
$u_{a,m_v}^e$	Area-metric value calculated by taking $m_v$ observations from experiment
$F_X, F_Y$	Horizontal and vertical loads acting on a cantilever beam, respectively
$\lambda_1, \lambda_2, \lambda_3$	Principal stretches
$w, t$	Width and thickness of the cantilever beam, respectively
$I_1, I_2, I_3$	Strain invariants
$D$	The deflections of the cantilever beam
$s$	Strain energy
$E$	The elastic modulus of the cantilever beam
$W(\cdot)$	Function of strain energy potential
$L$	The length of the cantilever beam
$C_{10}, B_1$	The material parameters of the Neo-Hookean function

## 1 Introduction

In recent years, RBDO using a simulation model has played an essential role in reducing product development costs and time by a breakaway from heuristic rule-of-thumb design approaches. In order to elicit accurate RBDO results, the process of quantifying the uncertainties inherent in all models, called uncertainty quantification (UQ), is essential. This process is divided into two main categories depending on what model you want to quantify: one is estimating the probabilistic model of the quantity of interest (QoI) by propagating the uncertainty of model variables through the model, and the other is estimating the uncertainty of the model variables using measured data; the former is called forward UQ, and the latter is called inverse UQ or the statistical model calibration (Lee et al. 2019a; Ralph 2014).

The statistical model calibration aims to minimize the differences between the observed data and the prediction results through mathematical models that can describe the physical phenomena in a statistical sense. Several forms of notable statistical model calibration frameworks have been proposed to achieve this goal (Arendt et al. 2012a; Campbell 2006; Xiong et al. 2009). The framework proposed by Kennedy and O'Hagan, called the KOH framework, is widely used for various scenarios: bias correction, parameter calibration, or both (Kennedy and O'Hagan 2001; Jung et al. 2015). However, the statistical model calibration problem inherently corresponds to the ill-posed problem, which means that optimization solutions for the inverse problem are unstable and non-unique: it implies that the optimization solutions can be sensitive to the measurement errors and have multiple solution sets, respectively. To solve the ill-posed problem, a regularization approach to obtain a more stable approximate solution and a multiple local search approach from different initial points can be used (Lee et al. 2019b; Sun et al. 2015; Villaverde et al. 2019). However, the problems mentioned above are intrinsically arisen due to insufficient data. Such a statistical uncertainty, which appears using insufficient data, is also called epistemic or reducible uncertainty because it can be reduced by adding more data, unlike aleatory uncertainty, which is inherent variability such as material properties, loads, and boundary conditions (Bi 2018; Roy et al. 2011).

Since few specifications indicate precisely how many samples are large enough, such as MIL-HDBK-5H, which states that 100 and 299 samples are required to find a distribution of properties (MIL-HDBK-5H 1998), interval approaches (Pashazadeh et al. 2008; Rao et al. 2008), information theory-based model selection method (Lim et al. 2016), and the goodness-of-fit test (Youn et al. 2011) can be applied to quantify the uncertainty of the model variables. Unlike the aforementioned parametric methods, Kang et al. (2018) proposed KDE-bd and KDE-ebd methods, which combine bounds information with the kernel density estimation (KDE), which exhibits unusual distribution shapes when using extremely small information (e.g. less than 10 data). Moon et al. (2019) also used a bootstrapping method to reduce unnecessary conservativeness by selecting a bandwidth satisfying a user-specified quantile value in the bootstrap distribution of the bandwidth. However, some studies have shown that at least 100 initial samples are required to obtain reliable results (Linnet 2000; Picheny et al. 2009; Wehrens et al. 2000). There are also studies that indirectly considered epistemic uncertainty of a model variable in reliability analysis. Jung et al. (2015) proposed a validation method to consider the uncertainty of model variables through the hypothesis test utilizing the area metric and u-pooling methods. Xi et al. (2012) applied the Bayesian approach to model random fields in the insufficient data sets and to consider the uncertainty of model variables for reliability analysis,

respectively. Li et al. (2018) and Jung et al. (2021) focused on reflecting the epistemic uncertainty induced by insufficient data in the surrogate model to find the conservative optimum that satisfies the target reliability. In the research of Xi (2019), various scenarios were established according to the status of the model parameter and test data being dealt with, and reliability analysis was carried out considering epistemic uncertainty for both model parameter and model bias at the same time. In particular, Moon et al. (2017) proposed a target output distribution method, which is a reliability analysis method that integrates all uncertainties such as the simulation model bias, insufficient input test, and output test data, based on the Bayesian approach. In addition, there are researches to increase the efficiency in simultaneously conducting the model calibration and validation processes. Jiang et al. (2020) performed model calibration and bias correction in a sequential manner, and Hu et al. (2021) utilized a stochastic Kriging model by distinguishing aleatory and epistemic uncertainty.

The biggest challenge encountered in most real engineering cases is that there are insufficient input (i.e., coupon) test data or output (i.e., component) test data available, which could be used to characterize unknown model variables directly or indirectly. For this reason, the purpose of this research is to propose a statistical model calibration framework that can reduce epistemic uncertainty by utilizing all available test data in constructing the unknown model variables. To consider epistemic uncertainty caused by the insufficient data, the model calibration field has been shifted from the unknown model variable domain to the statistical parameter domain of the unknown model variable, and the Bayesian approach has been employed to aggregate both input and output test data available. In addition, by applying the output test data to the optimization-based model calibration (OBMC), which uses log-likelihood as a calibration metric, a reasonable prior distribution of the statistical parameters represented by FIM is obtained. Eventually, the likelihood reflecting the input test data can be multiplied by the prior distribution to obtain the posterior distribution, so that all test data can be used to quantify unknown model variable.

A brief review of existing statistical model calibration methods and the Fisher Information for the prediction of the prior distribution is covered in Sect. 2. In Sect. 3, the proposed method is explained in detail. Then, the proposed method is validated through mathematical and engineering examples in Sect. 4. Lastly, conclusions are discussed in Sect. 5.

## 2 Review of statistical model calibration

The model calibration attempts to maximize consistency of a simulation model and test results by adjusting calibration parameters or unknown model variables. In particular,

the statistical model calibration differs from a deterministic model calibration in that calibration parameters can be expressed in statistical distributions rather than a deterministic perspective (Arendt et al. 2012a; Sargsyan et al. 2015; Trucano et al. 2006). To perform the statistical model calibration, a specific formulation of the relationship between experiments and simulation models is required, and the most widely used KOH framework is defined as (Kennedy and O'Hagan 2001)

$$z^e(\mathbf{d}) = z^s(\mathbf{d}, \boldsymbol{\xi}, \boldsymbol{\theta}^*) + \delta(\mathbf{d}) + \varepsilon \quad (1)$$

where  $\mathbf{d}$  is a controllable design variable vector,  $\boldsymbol{\xi}$  is a known model variable vector,  $\boldsymbol{\theta}$  is an unknown model variable vector as a calibration parameter vector, and the asterisk in  $\boldsymbol{\theta}^*$  means the true value.  $z^e(\cdot)$ ,  $z^s(\cdot, \cdot, \cdot)$ ,  $\delta(\cdot)$ , and  $\varepsilon$  in Eq. (1) indicate the experimental response function, response function of a simulation model, discrepancy function, and the measurement error, respectively. In many applications, the discrepancy term may be ignored on the assumption that its expected value is zero or that the simulation model is accurate (Campbell 2006). Moreover, under the assumptions that test data are obtained from well-designed experiments and that the unknown model variables are dominant, Eq. (1) is simplified as (Campbell 2006; Jung et al. 2015; Ralph 2014)

$$z^e(\mathbf{d}) = z^s(\mathbf{d}, \boldsymbol{\xi}, \boldsymbol{\theta}^*) \quad (2)$$

The probabilistic model of the calibration parameter vector in Eq. (2) can be estimated using the given test data with the statistical model calibration method such as optimization-based or Bayesian-based approaches to be described in Sects. 2.1 and 2.3.

### 2.1 Optimization-based model calibration (OBMC)

OBMC attempts to solve an inverse problem for finding the calibration parameters satisfying Eq. (2) using optimization algorithms, and thus it can be formulated as an optimization problem to maximize agreement with observations as (Lee et al. 2019a)

$$\hat{\boldsymbol{\theta}} = \arg \max_{\boldsymbol{\theta}} \varphi(z^e(\mathbf{d}), z^s(\mathbf{d}, \boldsymbol{\xi}, \boldsymbol{\theta})) \quad (3)$$

where  $\varphi(\cdot, \cdot)$  denotes the calibration metric as an objective function of the optimization problem to quantify correspondence between the observations and the simulation responses. For this reason, various calibration metrics such as a normalized absolute error, a weighted sum of the square error, and the distance measures, have been suggested, and among them, the most commonly used calibration metric for the statistical model calibration is the likelihood function (Oh et al. 2016, 2019; Vakilzadeh et al. 2017). Assuming

that the probabilistic distribution type is known, as a parametric approach, the statistical parameter vector of a calibration parameter or unknown model variable defined as  $\Theta = [\mu_\theta, \sigma_\theta]^T$  is determined by the maximum likelihood defined as

$$\hat{\Theta}_{ML} = \arg \max_{\Theta} L(\Theta; \mathbf{y}^e) \tag{4}$$

where  $L(\cdot)$  represents a likelihood function defined by  $\prod_{i=1}^{n_c} f(y_i^e | \Theta)$ ;  $n_c$  is the number of output test data (observations) for calibration;  $y_i^e$  is an individual output test data;  $f(y_i^e | \Theta)$  stands for the conditional probability density function (PDF) given statistical parameter vector  $\Theta$ ; and  $\mu_\theta, \sigma_\theta$  are mean and standard deviation of an unknown model variable as the calibration parameters, respectively. The method is intuitive and can adequately find a probabilistic model by estimating statistical parameters, such as the mean and variance, but also has some drawbacks such as inaccuracy of estimation if underlying candidates are inadequate or there are insufficient test data available (Lee et al. 2019a; McFarland et al. 2008; Ralph 2014).

### 2.2 Asymptotic normality of MLE

The MLE has two significant properties: consistency and asymptotic normality (Fahrmeir et al. 1985). These features mean that as the number of samples increases based on the law of large numbers and the central limit theorem, the estimator approaches a normal distribution containing the true value as (Ly et al. 2017)

$$\sqrt{n_c}(\hat{\Theta}_{ML} - \Theta) \xrightarrow{d} N(0, \bar{\mathbf{I}}^{-1}(\Theta)) \tag{5}$$

where  $\hat{\Theta}_{ML}$  is an estimated calibration parameter vector through MLE; the letter d above the arrow indicates a convergence in distribution; and  $\bar{\mathbf{I}}(\Theta)$  is the expected FIM defined as the expectation for the negative second derivative of the log-likelihood, expressed as (Cavanaugh et al. 1996)

$$\bar{\mathbf{I}}(\Theta) = \mathbb{E} \left[ - \frac{\partial^2 \ln L(\Theta | \mathbf{y}^e)}{\partial \Theta^2} \right] \tag{6}$$

However, since the expected FIM is not always computable, the observed FIM, which can replace the expected FIM in many instances, is defined as the Hessian of the observed log-likelihood written as (Cavanaugh et al. 1996; Efron et al. 1978)

$$\mathbf{I}(\hat{\Theta}_{ML}) = \sum_{i=1}^{n_c} - \frac{\partial^2 \ln L(\Theta | y_i^e)}{\partial \Theta^2} \Big|_{\Theta = \hat{\Theta}_{ML}} \tag{7}$$

In Eq. (7), true values for the calibration parameter can be replaced by MLE as a consistent estimator (Cavanaugh et al. 1996; DeGroot et al. 2011). By applying the relationship of  $I(\cdot) = n_c I_1(\cdot)$  and Slutsky’s theorem to Eq. (5), the estimated calibration parameter vector converges in distribution to a normal distribution or a multivariate normal distribution as (Myung et al. 2005; Sourati et al. 2017)

$$\hat{\Theta}_{ML} \xrightarrow{d} N(\Theta, \mathbf{I}_1^{-1}(\Theta) / n_c) \tag{8}$$

Consequently, Eq. (8) shows that the estimation accuracy of MLE can be expressed in the form of Fisher information and that as  $n_c$  increases, the amount of information provided for the unknown model variables can also increase, reducing the estimation error.

### 2.3 Bayesian-based model calibration

The Bayesian inference, which is more suitable for the statistical model calibration under insufficient data environment since it can incorporate a prior information, constructs a probability distribution of a parameter satisfying Eq. (2) through rejection sampling based on the Bayes’ theorem and is defined as

$$p(\Theta; \mathbf{y}^e) = \frac{L(\mathbf{y}^e | \Theta) \cdot \pi(\Theta)}{\int_{\Theta} L(\mathbf{y}^e | \Theta) \cdot \pi(\Theta) d\Theta} \tag{9}$$

where  $p(\Theta; \mathbf{y}^e)$  is a posterior distribution as a PDF of the calibration parameter to be estimated based on the observed data;  $L(\mathbf{y}^e | \Theta)$  represents the likelihood that varies with the given candidate calibration parameter; and  $\pi(\Theta)$  denotes a prior distribution for the calibration parameter. Since the denominator in Eq. (9), which corresponds to a normalization constant, is not easy to compute and does not affect the shape of the posterior distribution, Eq. (9) can be expressed as (Arendt et al. 2012b; Sun et al. 2015)

$$p(\Theta; \mathbf{y}^e) \propto L(\mathbf{y}^e | \Theta) \cdot \pi(\Theta) \tag{10}$$

The noteworthy features of the method are that it can utilize expert knowledge as a prior distribution to resolve the insufficient data problem and update the posterior distribution by adding new data efficiently, unlike OBMC (Lee et al. 2019a). However, selection of an improper prior distribution has significant effects on the estimation results, and the use of time-consuming methods such as the Markov Chain Monte Carlo (MCMC) algorithm to sample the estimated calibration parameter vectors from the posterior distribution is a major impediment (Higdon et al. 2008; Honarmandi et al. 2020).

### 3 Statistical model calibration integrating obtainable input and output test data

This research aims to figure out how to mitigate epistemic uncertainty during the statistical model calibration procedure when the data is scarce. To this end, a practical method is proposed to integrate all available input and output test data which may not be sufficient to quantify unknown model variables directly or indirectly by adopting the Bayesian inference. In the absence of expert knowledge, an approach to select an appropriate prior distribution using the output test data is also suggested to implement the proposed method. In order to maintain the conservativeness of the model calibration due to lack of data, the statistical model calibration is performed in a statistical parameter domain of the unknown model variable, not in the unknown model variable domain. In detail, it means finding a probability distribution of the statistical parameter by considering the statistical parameter of the unknown model variable as a random variable rather than a deterministic variable.

### 3.1 Prior distribution selection using output test data

Most statistical model calibrations, often referred to as inverse UQ, commonly use output test data to characterize distributions of unknown model variables (Arendt et al. 2012b; Oh et al. 2016; Xi et al. 2012). Similarly, in this research, the OBMC procedure of finding statistical parameters of unknown model variables that maximize likelihood by utilizing the output test data can be formulated as

$$\hat{\Theta}_{ML}, \hat{\zeta}_{ML} = \arg \max_{\Theta, \zeta} L(\Theta, \zeta; \mathbf{y}^e) \tag{11}$$

where  $\zeta$  refers to a distribution type of the unknown model variable that best represents the output test data and is specific to five types with two parameters as shown in Table 1; and  $\Theta = [\mu_{\theta_1}, \sigma_{\theta_1}, \dots, \mu_{\theta_k}, \sigma_{\theta_k}]^T \in \mathbb{R}^{2k}$  is a statistical parameter vector for  $k$  unknown model variables, and each component is still treated as a deterministic variable, which can be expressed as a sample statistics, as shown in Table 1.

Due to the use of a limited number of output test data, the statistical parameters derived from Eq. (11) may have statistical uncertainties. These uncertainties can be defined by the asymptotic normality of MLE covered in Sect. 2.2, and  $\mathbf{I}(\hat{\Theta}_{ML})$  for  $k$  unknown model variables is defined as

$$-\nabla_{\Theta}^2 l(\Theta | \mathbf{y}^e) \Big|_{\Theta = \hat{\Theta}_{ML}} = \begin{bmatrix} -\frac{\partial^2 l(\Theta | \mathbf{y}^e)}{\partial \Theta_1^2} & -\frac{\partial^2 l(\Theta | \mathbf{y}^e)}{\partial \Theta_1 \partial \Theta_2} & \dots & -\frac{\partial^2 l(\Theta | \mathbf{y}^e)}{\partial \Theta_1 \partial \Theta_{2k}} \\ -\frac{\partial^2 l(\Theta | \mathbf{y}^e)}{\partial \Theta_2 \partial \Theta_1} & -\frac{\partial^2 l(\Theta | \mathbf{y}^e)}{\partial \Theta_2^2} & \dots & -\frac{\partial^2 l(\Theta | \mathbf{y}^e)}{\partial \Theta_2 \partial \Theta_{2k}} \\ \vdots & \vdots & \ddots & \vdots \\ -\frac{\partial^2 l(\Theta | \mathbf{y}^e)}{\partial \Theta_{2k} \partial \Theta_1} & -\frac{\partial^2 l(\Theta | \mathbf{y}^e)}{\partial \Theta_{2k} \partial \Theta_2} & \dots & -\frac{\partial^2 l(\Theta | \mathbf{y}^e)}{\partial \Theta_{2k}^2} \end{bmatrix} \Big|_{\Theta = \hat{\Theta}_{ML}} \tag{12}$$

**Table 1** PDFs with two distribution parameters

Distribution type	PDF $f_X(x; a, b)$	Relationship between two parameters and sample statistics
Normal	$\frac{1}{\sqrt{2\pi}b} \cdot \exp \left[ -\frac{1}{2} \left( \frac{x-a}{b} \right)^2 \right]$	$\mu = a$ $\sigma^2 = b^2$
Lognormal	$\frac{1}{\sqrt{2\pi}xb} \cdot \exp \left[ -\frac{1}{2} \left( \frac{\ln x - a}{b} \right)^2 \right]$	$\mu = \exp \left[ a + (b^2/2) \right]$ $\sigma^2 = (\exp [b^2] - 1) \cdot \exp [2a + b^2]$
Gamma	$x^{a-1} \cdot \frac{\exp \left[ -\frac{x}{b} \right]}{\Gamma(a) \cdot b^a}$	$\mu = ab$ $\sigma^2 = ab^2$
Gumbel	$\frac{1}{b} \cdot \exp \left[ -\frac{(x-a)}{b} - \exp \left[ -\frac{(x-a)}{b} \right] \right]$	$\mu = a + 0.5772b$ $\sigma^2 = b^2 \pi^2 / 6$
Weibull	$\frac{b}{a} \left( \frac{x}{a} \right)^{b-1} \exp \left[ -\left( \frac{x}{a} \right)^b \right]$	$\mu = a \cdot \Gamma(1 + 1/b)$ $\sigma^2 = a^2 [\Gamma(1 + 2/b) - \Gamma^2(1 + 1/b)]$



where  $l(\cdot)$  is a log-likelihood, and the subscript of the denominator refers to the order of statistical parameter components. Lastly, the estimated statistical parameters treated as random variables can be defined as a prior distribution in the form of a multivariate normal distribution represented as

$$\pi(\hat{\Theta}_{ML}, \hat{\zeta}_{ML}; \mathbf{y}^e) \sim N(\hat{\Theta}_{ML}, \mathbf{I}^{-1}(\hat{\Theta}_{ML})) \tag{13}$$

### 3.2 Posterior distribution updated by input test data

The estimated prior distribution based on the output test data discussed in Sect. 3.1 could be used as a reasonable alternative rather than non-informative prior such as a uniform distribution because it is based on given observations in the absence of the related literature information or expert knowledge. However, unnecessary conservativeness or inaccuracy of the prior distribution induced from insufficient output test data needs to be improved. To this end, based on the results estimated in Sect. 3.1, plausibility for an occurrence of input test data expressed in the form of a likelihood function is defined as

$$L(\mathbf{x}^e | \hat{\Theta}_{ML}, \hat{\zeta}_{ML}, \mathbf{y}^e) \tag{14}$$

where  $\mathbf{x}^e$  refers to the input test data vector, which means a limited number of realizations taken in the unknown model variable domain of  $z^s(\cdot, \cdot, \cdot)$ . Since  $\Theta$  represents a random variable vector, the likelihood of Eq. (14) is calculated by reflecting  $\mathbf{x}^e$  in the probability model of the unknown model variable estimated through Eq. (13), where the statistical parameter of the unknown model variable is obtained by sampling an appropriate amount from the prior distribution. In conclusion, the prior distribution is multiplied by the likelihood function and updated to the posterior distribution as

$$p(\hat{\Theta}, \hat{\zeta}; \mathbf{x}^e, \mathbf{y}^e) = L(\mathbf{x}^e | \hat{\Theta}_{ML}, \hat{\zeta}_{ML}, \mathbf{y}^e) \cdot \pi(\hat{\Theta}_{ML}, \hat{\zeta}_{ML}; \mathbf{y}^e) \tag{15}$$

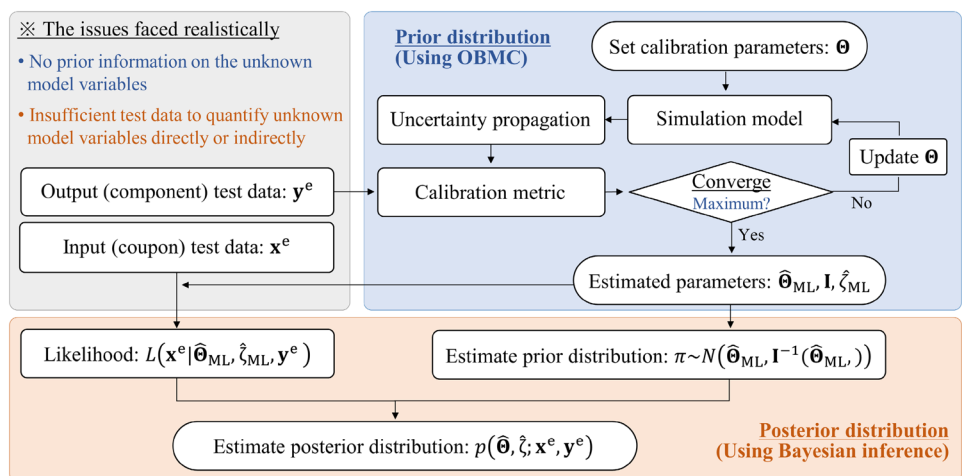
The proposed method using the Bayesian framework, as shown in Eq. (15), can readily reduce the epistemic uncertainty by integrating both input and output test data reasonably for the statistical model calibration. It is also expected that the predictive accuracy will be improved as the current inverse UQ results are updated in the most plausible direction by the likelihood of the input test data. The overall procedure for the proposed method is shown in Fig. 1.

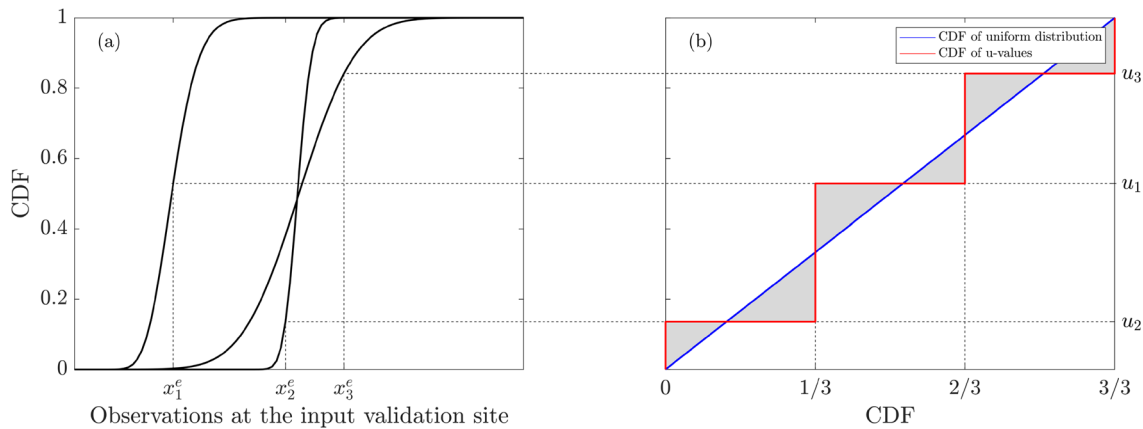
### 3.3 Statistical model validation for the calibration parameter

The validation metric is a measure of quantifying the similarity between the calibrated prediction and the observations. There are various measures such as the root mean square error, hypothesis testing, Bayes factor, and Kullback–Leibler divergence, which are sometimes also used as the calibration metrics (Liu et al. 2011; Oh et al. 2019; Xiong et al. 2009). In this study, the validity of the calibrated statistical model is verified by employing a hypothesis test using the probability distribution of the area-metric calculated by applying the u-pooling method (Jung et al. 2015). In addition, by propagating the probability distribution of the unknown model variable to the probability model of QoI based on Eq. (2) and calculating its likelihood, the degree of improvement in the predictive accuracy of the proposed method is quantitatively evaluated comparing it with the results obtained from the prior distribution.

The u-pooling method was devised by Ferson to calculate disparate observations collected under different conditions as one index called the area-metric shown in Fig. 2b based on the probability integral transform theorem as shown in Fig. 2a (Ferson et al. 2008; Ferson et al. 2009).

**Fig. 1** Flowchart of the proposed statistical model calibration method





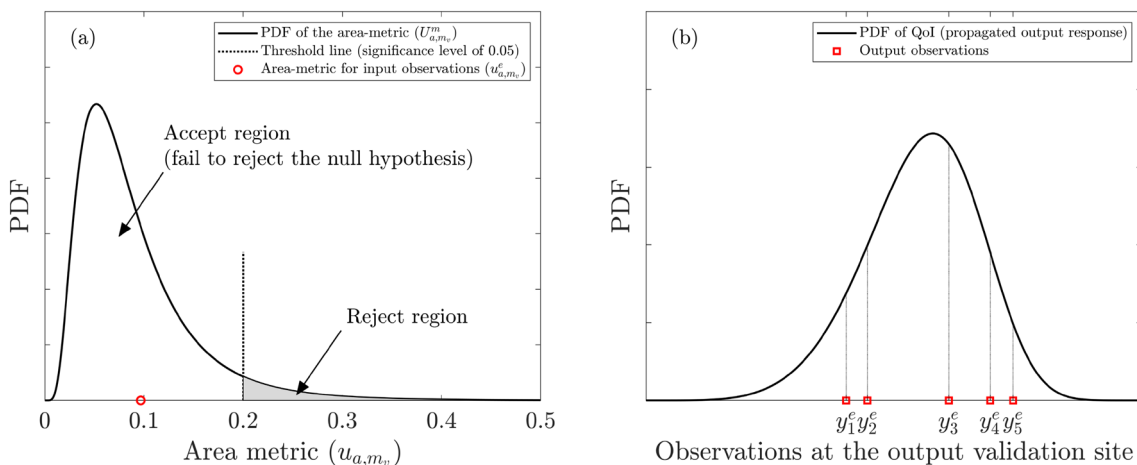
**Fig. 2** Main concepts of the validation metric: **a** evaluation of u-value, and **b** area-metric at 3 validation site

The  $u_j$  values pooled in Fig. 2a refer to the cumulative distribution function (CDF) values of the unknown model variable estimated in the calibration site and are calculated as

$$u_j = \hat{F}_{x_i}(x_j) \tag{16}$$

where  $\hat{F}_{x_i}(\cdot)$  denotes the estimated CDF of the unknown model variable at the calibration site  $x_i$ , while  $x_j$  refers to the validation site satisfying  $i \neq j$  (Campbell 2006). Since the  $u_j$  values calculated in this way must follow a standard uniform distribution assuming that the  $x_j$  values come from the identical mother distribution, the degree of inconsistency between the estimated probability distribution of the unknown model variable and the observations could be expressed as an area-metric, as shown in Fig. 2b (Li et al. 2014). However, since this limited number of observations given in the validation site causes epistemic uncertainty on

the calculated area-metric, a hypothesis test is performed to take this into account. As the first step for hypothesis testing, the  $u_j$  values are calculated by acquiring the same number of samples ( $m_v$ ) as the observations given in the validation site from the estimated probability distribution of unknown model variables. This process can be repeated multiple times (e.g., 1000) to secure randomness data ( $u_{a,m_v}^{(1)}, \dots, u_{a,m_v}^{(1000)}$ ) on the area-metric and expressed as  $U_{a,m_v}^m$ , the probability distribution of area-metric calculated using the validation samples on the estimated probability model, using KDE as shown in Fig. 3a. After that, a one-tailed test is conducted to accept or reject the null hypothesis that the estimated probability model is valid at a significance level of 5% using the area-metric ( $u_{a,m_v}^c$ ) calculated by reflecting the given observations at the validation site to  $\hat{F}_{x_i}(\cdot)$  (Nah et al. 2020; Son et al. 2020).



**Fig. 3** Multiple strategies for validity check: **a** hypothesis test to validate the estimated probability model, and **b** likelihood calculation for output response to evaluate the accuracy of the estimation

However, the adoption of the null hypothesis only means that there is not enough evidence to reject it, so an additional step is required to quantitatively confirm that the predictive accuracy of the proposed method has improved compared to the estimation result of the prior distribution. To this end, as shown in Fig. 3b, the probability distribution of unknown model variables estimated at each phase was propagated into the output response distribution to calculate the likelihood and compare them.

### 4 Numerical examples

In this section, two examples are implemented to validate the proposed method. Two significant features of the proposed method that should be carefully contemplated through the following examples are (1) accuracy of the MLE distribution represented by the Fisher information using output test data and (2) reduction of the epistemic uncertainty obtained by considering input test data. In addition, the adequacy of the statistical model validation approaches could be considered for each case where a statistical parameter of the unknown model variable is given as a single true value or as a limited number of test data of the unknown model variable.

#### 4.1 Mathematical example: cantilever beam

The mathematical example in this section aims to provide a good grasp of the overall procedure and major features of the proposed method. Therefore, the uniform cantilever beam illustrated in Fig. 4 is adopted to find a probability model of an unknown model variable using given deflection data (as the output test data) and modulus of elasticity data (as the input test data), and also the tip deflection corresponding to the QoI of this example is formulated as (Wu et al. 2001)

$$D(F_X, F_Y, w, t, E) = \frac{4L^3}{Ewt} \sqrt{\left(\frac{F_X}{w^2}\right)^2 + \left(\frac{F_Y}{t^2}\right)^2} \tag{17}$$

Fig. 4 Uniform cantilever beam exposed to horizontal and vertical loads

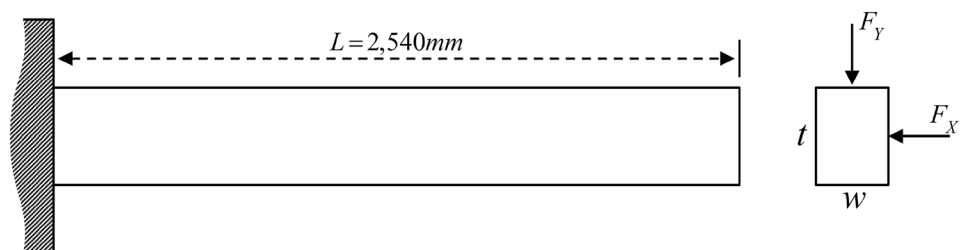


Table 2 Geometric dimensions and loads for the cantilever beam example

Known model variables	Symbol	Value	Unit
Width of the cross-section	$w$	62.19	[mm]
Thickness of the cross-section	$t$	98.76	[mm]
Horizontal load	$F_X$	4,448	[N]
Vertical load	$F_Y$	2,224	[N]

where  $\mathbf{d} = [F_X, F_Y, w, t]^T$  is defined as a known model variable vector, which is listed in Table 2, and  $E$  represents the modulus of elasticity as an unknown model variable. The calibration parameter  $\Theta = [\mu_E, \sigma_E]^T$  is derived in the form of a probability distribution of random variables by implementing the proposed method. In addition, the unknown model variable  $E$  is specified to follow a normal distribution, and the true values of the statistical parameters, which are the target values of the calibration parameters, are assumed to be  $\mu_E^* = 199,947$  MPa and  $\sigma_E^* = 9,997$  MPa (Hess et al. 2002).

#### 4.1.1 Prior distribution with the output test data (beam deflections)

Since the probability model of the unknown model variable is given as mentioned above, the deflection of the beam corresponding to the output test data can be obtained without limitation. Following the formulation of Eq. (11), the OBMC process using output test data vector denoted as  $\mathbf{y}^e = [D_1, D_2, \dots, D_{n_c}]^T$  can be expressed as

$$\hat{\Theta}_{ML}, \hat{\zeta}_{ML} = \underset{\Theta, \zeta}{\operatorname{argmin}} -l(\Theta, \zeta; \mathbf{y}^e) \tag{18}$$

where  $l(\cdot)$  denotes a log-likelihood function defined by  $\sum_{i=1}^{n_c} \ln [f(y_i^e | \Theta, \zeta)]$ . Although estimated statistical parameters  $\hat{\mu}_E$  and  $\hat{\sigma}_E$  obtained through the above optimization process are deterministic variables, the asymptotic normality property of the MLE addressed in Sect. 2.2 results in each estimation having a  $100 \cdot (1-\alpha)\%$  confidence interval (CI) represented as



$$\Theta_l^* \in \left[ \hat{\Theta}_l - z_{\alpha/2} \sqrt{\left(n_c I_1(\hat{\Theta}_l)\right)^{-1}}, \hat{\Theta}_l + z_{\alpha/2} \sqrt{\left(n_c I_1(\hat{\Theta}_l)\right)^{-1}} \right] \tag{19}$$

where the subscript  $l = 1, 2$  refers to the order that consists of statistical parameter vector;  $z_{\alpha/2}$  represents a critical point for  $\alpha$  significant level in the standard normal distribution, and as an example, the critical point  $z_{0.025}$  for the 95% CI is 1.96. Based on the given output test data, the variance–covariance matrix for the two estimated statistical parameters is expressed in Eq. (19) as Fisher information  $I(\cdot) = n_c I_1(\cdot)$  and defined as

$$-\nabla_{\Theta}^2 l(\Theta | y^e) \Big|_{\Theta = \hat{\Theta}_{ML}} = \begin{bmatrix} -\frac{\partial^2 l(\Theta | y^e)}{\partial \mu_E^2} & -\frac{\partial^2 l(\Theta | y^e)}{\partial \mu_E \partial \sigma_E} \\ -\frac{\partial^2 l(\Theta | y^e)}{\partial \sigma_E \partial \mu_E} & -\frac{\partial^2 l(\Theta | y^e)}{\partial \sigma_E^2} \end{bmatrix} \Big|_{\Theta = \hat{\Theta}_{ML}} \tag{20}$$

Finally, according to Eq. (13), the distribution type of the unknown model variable and the prior distribution expressed as the bivariate normal distribution can be obtained as

$$\pi(\hat{\Theta}_{ML}, \hat{\zeta}_{ML} | y^e) \sim N(\hat{\Theta}_{ML}, -\nabla_{\Theta}^2 l(\Theta | y^e) \Big|_{\Theta = \hat{\Theta}_{ML}}) \tag{21}$$

As the number of output test data increases, based on Eq. (19), the estimation performance of the MLE for each statistical parameter and the 95% CI based on the Fisher information is plotted as shown in Fig. 5. Fisher information refers to a measure of the amount of information observed by the random variables about the unknown parameters (Prokopenko et al. 2011), so more observations improve estimation accuracy of the MLE for true probability model and the 95% CI becomes narrower. Similarly, identification results for the

distribution type of the unknown model variable represented by the filled marker also tend to be more accurate with more observations. Figure 5 also shows that the identified distribution types could be different from the true ones (normal distributions for both parameters in this example) even when 300 test data were used as marked as hollow circles in the figure. Lognormal distributions are identified in such cases; however, when mean values are very large as in this example, the shapes of normal and lognormal distributions are very similar with each other so that the incorrectly identified distribution type would not affect calibration results. It is also confirmed that the estimated mean and standard deviation are 200,018 MPa and 10,220 MPa, respectively, which are very close to the true values. In addition, in terms of the convergence rate for the true values, the mean value converges faster than the variance value.

By Eq. (21), it can also be represented by a contour plot representing a bivariate normal distribution over the calibration domain, as shown in Fig. 6. It shows the prior distribution when 3 and 5 output test data ( $n_c$ ) are given. As shown in Figs. 5 and 6, as the number of data increases, the estimation accuracy improves. In other words, the estimated statistical parameter approaches the true value, and the estimated distribution with the 95% confidence level narrows. The above results show that if there is a large amount of output test data, OBMC can express the uncertainty of the unknown model variables that are sufficiently accurate. However, in actual engineering cases, it is difficult to obtain more than 5 output test data due to time and cost reasons (Jung et al. 2015; Son et al. 2020). So, in this problem, the estimation result using 3 output test data will be used to perform the next step.

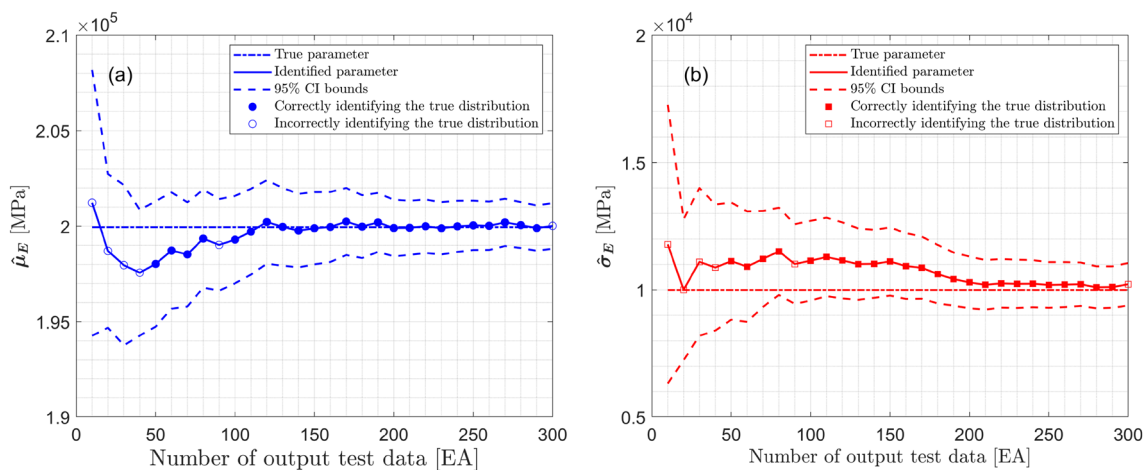
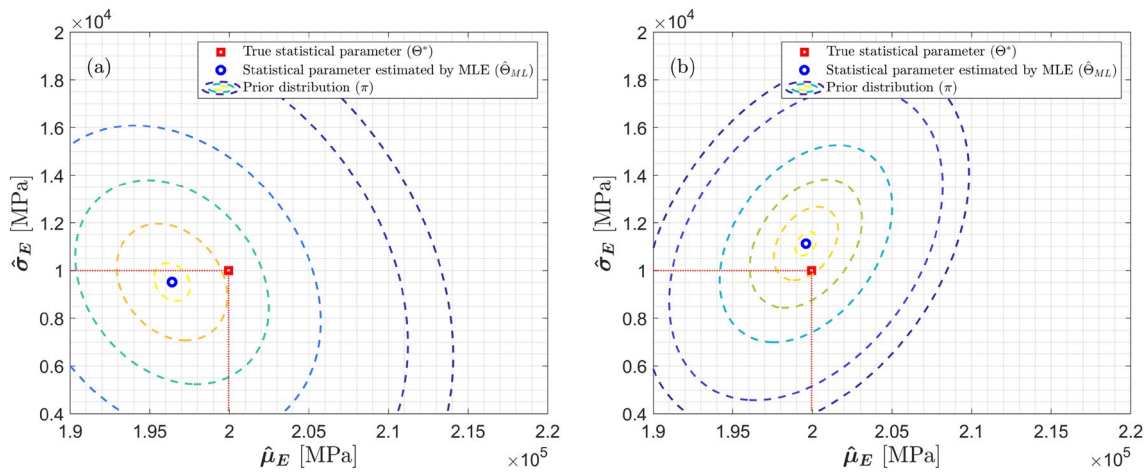


Fig. 5 Estimation performance of the Fisher information according to the quantity of output test data: **a** estimated mean with 95% CI bounds, and **b** estimated standard deviation with 95% CI bounds



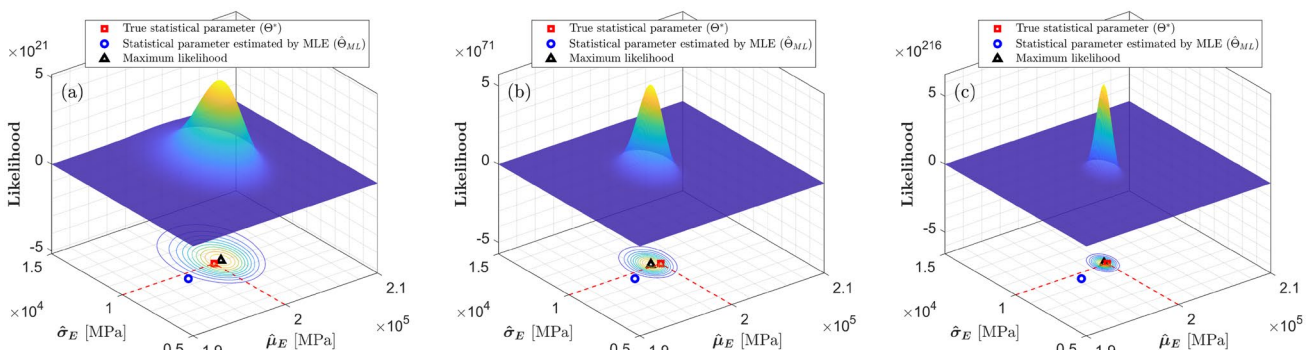
**Fig. 6** Comparison of the prior distribution estimation according to the quantity of output test data (lines in the contour plot represent the same probability of 10, 30, 50, 70, 90, and 95% in the outward direction): **a** considering 3 output test data, and **b** considering 5 output test data

**4.1.2 Posterior distribution with the input test data (modulus of elasticity obtained by coupon test)**

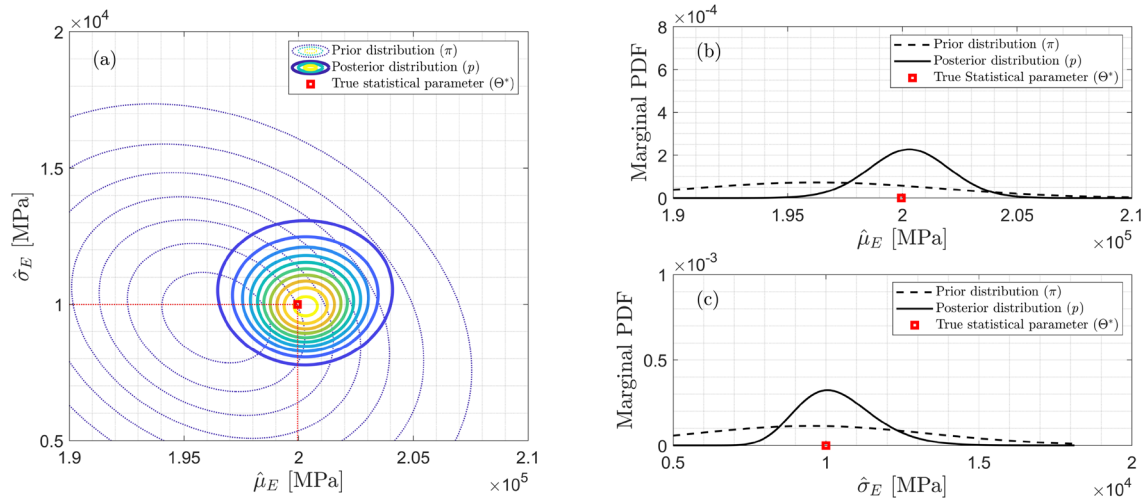
This section assumes a situation in which limited number of input test data are also available, denoted by  $\mathbf{x}^e = [E_1, E_2, \dots, E_{m_c}]^T$ , which will be considered in the statistical model calibration to reduce the epistemic uncertainty due to lack of data. The statistical parameters that can best represent the limited input test data are obtained in the form of likelihood shown in Eq. (14). In particular, the statistical parameters of the lognormal distribution are sampled by an appropriate quantity (e.g., 100,000) from the estimated prior distribution as random variables, and the likelihood values at each of the sampled statistical parameters are tallied over the statistical parameter domain as shown in Fig. 7. In Fig. 7, it can be seen that as the number of input test data increases, the likelihood is concentrated as a statistical parameter with a high plausibility that best represents the given input test data. In addition, in Fig. 7a, since the estimated distribution

type of the input model parameter used in the likelihood calculation is different from the true one as a lognormal distribution, the location of the maximum likelihood obtained through input test data (triangle symbol in black) differs from that obtained through the output test data (circular symbol in blue). However, Figs. 7b and c show that this difference can also be narrowed if the number of input test data increases.

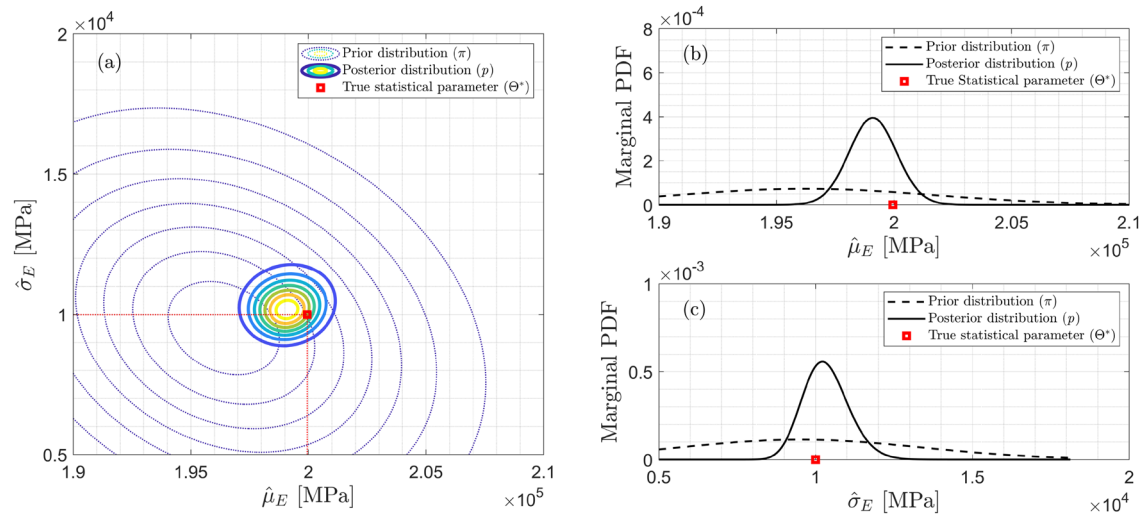
Finally, the likelihood distribution for 30, 100, and 300 input test data shown in Fig. 7 is multiplied by the prior distribution in Fig. 6a and is updated to the posterior distribution as shown in Figs. 8, 9, and 10, respectively. Through Figs. 8 and 10, it can be seen that the likelihood distribution of Fig. 7 plays an effective role in getting closer to the true statistical parameters by updating the prior distribution to the posterior distribution, and at the same time, the estimation accuracy also increases as the amount of the input test data increases.



**Fig. 7** Calculated likelihood according to the different number of additional input test data: **a** considering 30 additional input test data, **b** considering 100 additional input test data, and **c** considering 300 additional input test data



**Fig. 8** Comparison of prior and posterior distributions given 30 input test data: **a** joint PDF of statistical parameters, **b** marginal PDF of mean, and **c** marginal PDF of standard deviation

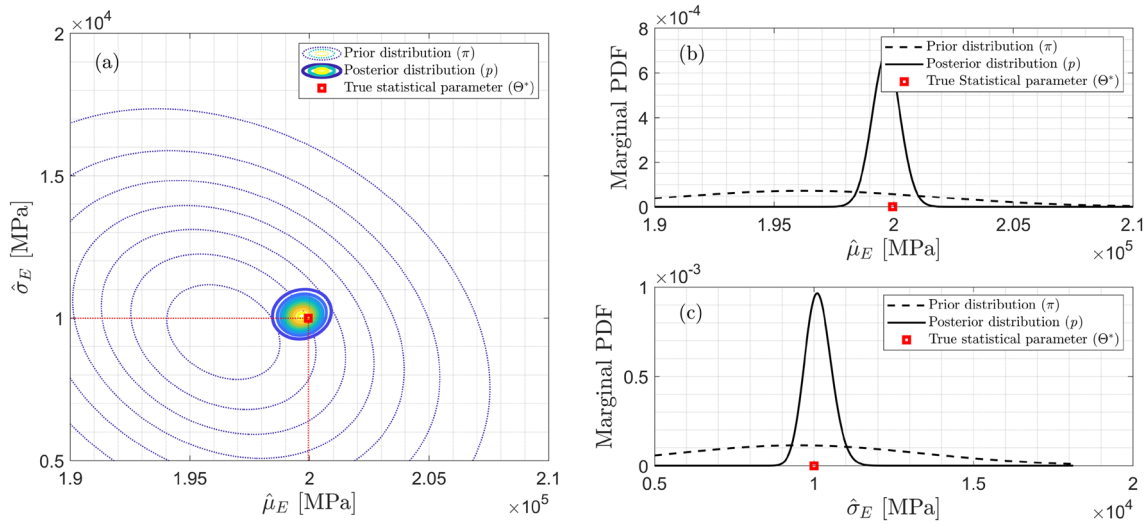


**Fig. 9** Comparison of prior and posterior distribution given 100 input test data: **a** joint PDF of statistical parameters, **b** marginal PDF of mean, and **c** marginal PDF of standard deviation

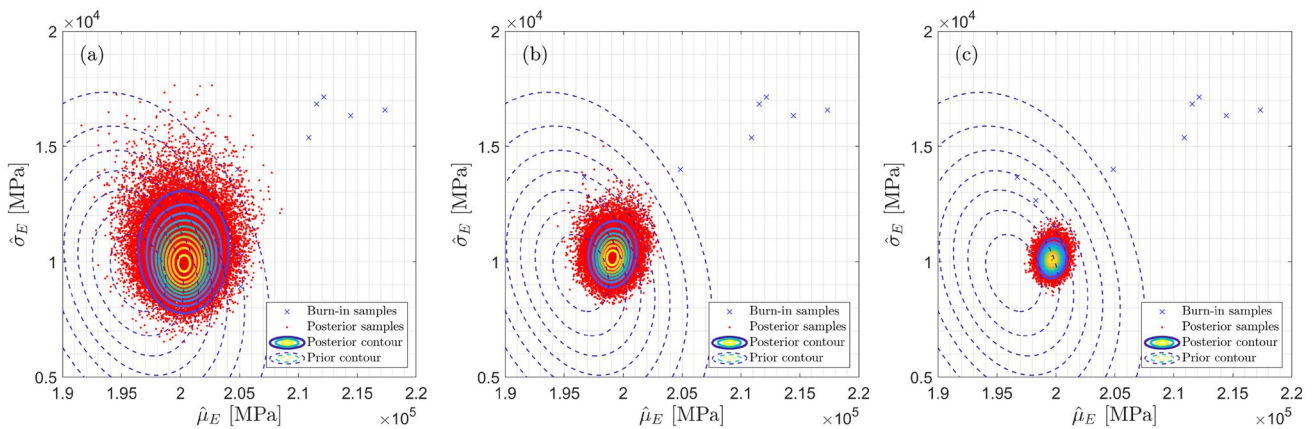
Subsequently, to characterize the probability distribution of the unknown model variable, samples of the statistical parameters following the posterior distribution of Figs. 8, 9, and 10 are extracted using the random walk Metropolis (RWM) algorithm, one of the MCMC methods (Vrugt 2016). The sampled 100,000 statistical parameters for each case in Fig. 11 are reflected in the distribution type of the unknown model variable estimated in Sect. 4.1.1, and 1,000 elastic moduli are sampled for each statistical parameter sample point, resulting in a total elastic modulus of 100 million. It can be represented in the form of a probability distribution as shown in Fig. 12 using KDE. As shown in Fig. 12a, considering additional input test data to the statistical

model calibration allows the final estimated distribution to approach the true distribution by adjusting the right-end and central density values of the estimated PDF using only the output test data. In addition, it can be confirmed that when the number of input test data increases, it becomes almost the same as the true distribution as shown in Fig. 12c.

In order to demonstrate the effectiveness of the proposed method, a comparison with the OBMC, an existing statistical model calibration method, was conducted. The estimation results for an unknown model variable by each method are shown in Fig. 13. The identical condition as shown in Fig. 12a was applied to the proposed method, while only 3 and 30 output test data were used



**Fig. 10** Comparison of prior and posterior distribution given 300 input test data: **a** joint PDF of statistical parameters, **b** marginal PDF of mean, and **c** marginal PDF of standard deviation



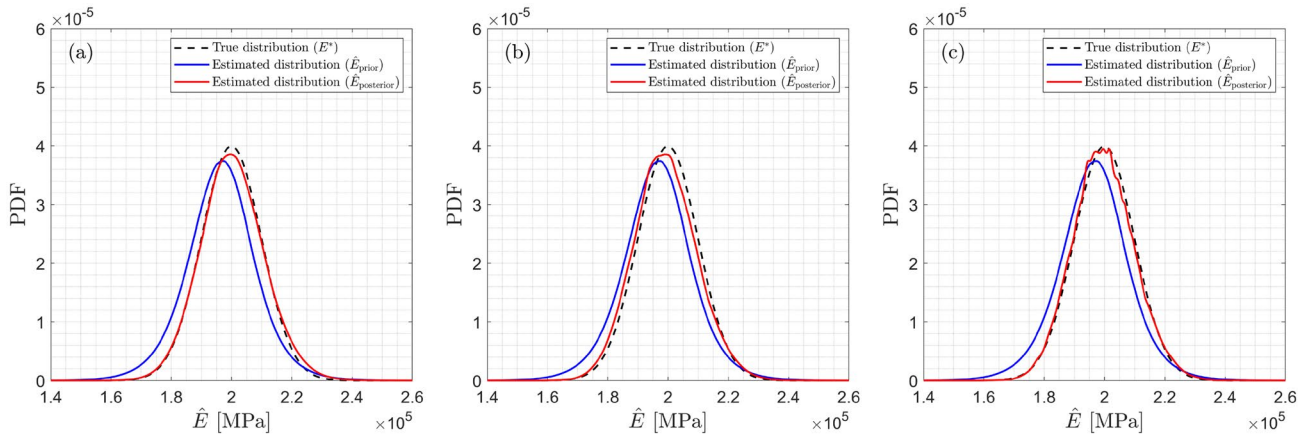
**Fig. 11** Bivariate scatter plots of the samples derived from the posterior distributions using the MCMC method: **a** considering 30 additional input test data, **b** considering 100 additional input test data, and **c** considering 300 additional input test data

in the OBMC method. Comparing the estimated PDF results derived from the prior distribution ( $\hat{E}_{\text{prior}}$ ) and the OBMC method ( $\hat{E}_{\text{OBMC}}$ ) in Fig. 13a, where 3 output test data are equally used, the former is more spread. This is because the statistical parameters of an unknown model variable are considered to be fixed in the OBMC method, whereas the proposed method statistically estimated them considering the parameter uncertainty under the initial lack of information. Therefore, if the data is insufficient, the proposed method that shows conservative estimation results would be more reasonable. In addition, in Fig. 13b, it can be seen that, despite the increase in the output test data, the estimation result by the OBMC method becomes more precise, but it is estimated differently from the true

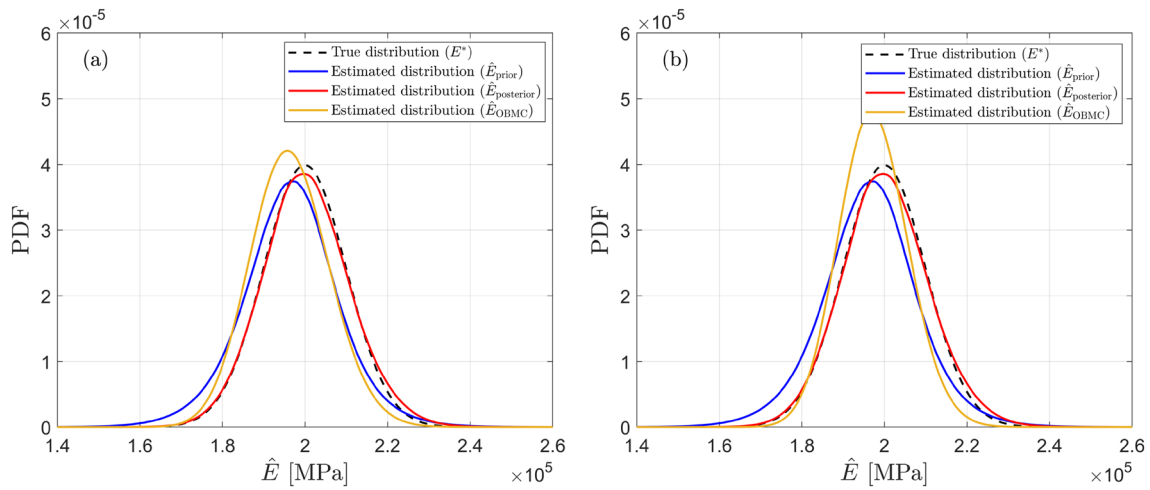
distribution ( $E^*$ ). Since the OBMC method, which is dependent on the given output test data, causes an overfitting problem as shown in Fig. 13b, it could be confirmed that using the input and output test data for the statistical model calibration as in the proposed method is effective in preventing overfitting (Bishop 2006; Deisenroth et al. 2020; Jiang et al. 2020).

Interestingly, even if the same 30 test data is used for the statistical model calibration, as shown in the contour plots of Fig. 14, the estimation accuracy may differ depending on the quantity of each input and output test data used. This means that it can be designed effectively according to the time, cost, and level of difficulty required for each test condition.

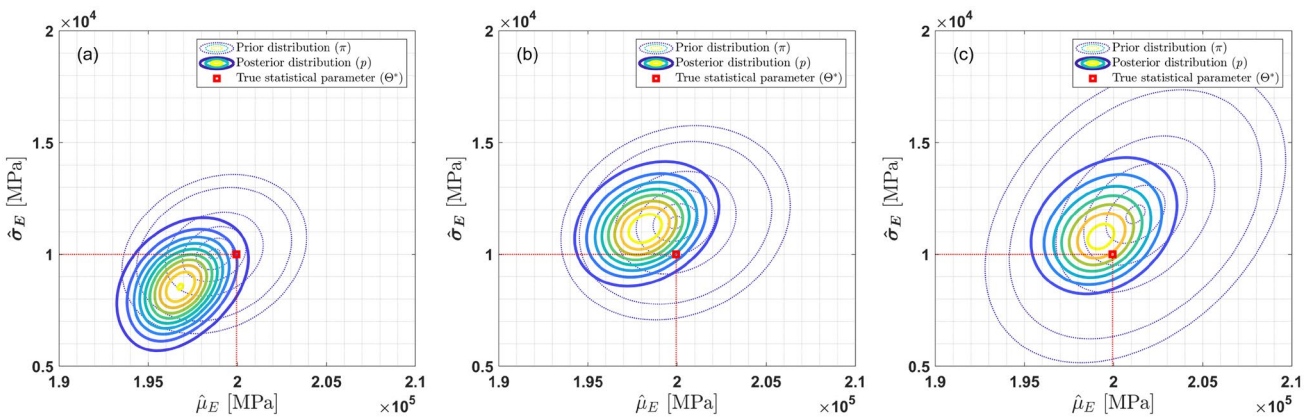




**Fig. 12** Comparison of estimation accuracy for an unknown model variable according to the number of input test data: **a** considering 30 additional input test data, **b** considering 100 additional input test data, and **c** considering 300 additional input test data

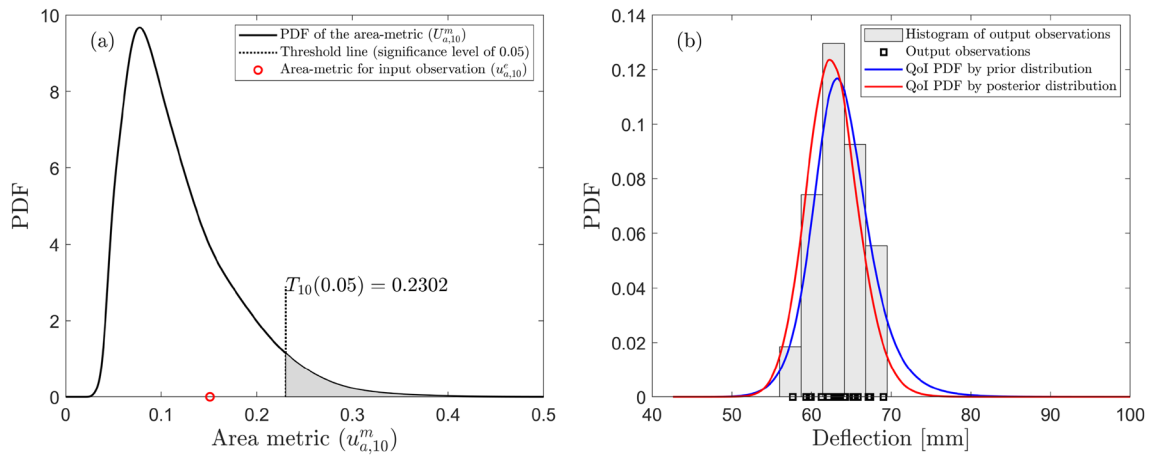


**Fig. 13** Estimation accuracy for an unknown model variable through the comparison with the existing statistical model calibration method (OBMC) using **a** 3 output test data, and **b** 30 output test data



**Fig. 14** Comparison of the estimation results based on a combination of the input and output test data with the same total quantity: **a** 20 output & 10 input, **b** 15 input & 15 output, and **c** 10 output & 20 input test data





**Fig. 15** Statistical model validation for the estimated calibration parameter under the conditions of Fig. 12a: **a** hypothesis test using additional 10 input test data, and **b** likelihood calculation of addi-

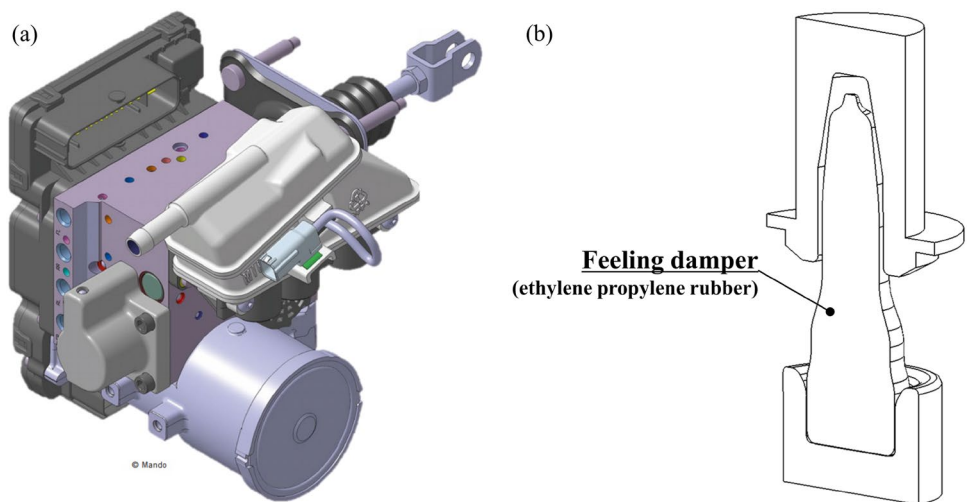
tional 20 output test data for the propagated QoI distribution based on the prior and posterior distribution

**4.1.3 Statistical model validation**

The probability model of the unknown model variable estimated through the aforementioned series of processes is evaluated for predictive accuracy through the statistical validation method covered in Sect. 3.3. The hypothesis test results to verify the estimated probability model of the elastic modulus are shown in Fig. 15a. 10 observations ( $m_v$ ) added from the true distribution of the elastic modulus were used, and  $U_{a,10}^m$ , which means the PDF of the area-metric calculated from the estimated distribution, was fitted through the results of 1,000 iterations and KDE. The null hypothesis that the prediction model is valid can be adopted because the area-metric ( $u_{a,10}^e$ ) calculated from the estimated model and the added 10 observations is 0.15, which is less than the threshold  $T_{10}(0.05)$  corresponding to the significance level of 5%.

In addition, to quantitatively prove the effectiveness of the proposed method, the likelihood calculation was performed in the QoI domain as shown in Fig. 15b, and the deflections diffused from the 20 observations ( $n_v$ ) added from the true distribution of the elastic modulus are used. The PDF values in the central part where the deflections are concentrated are higher in the QoI distribution propagated from the posterior distribution than the prior, and the actually calculated likelihood value also increases by 1.34 times from 1.15E-22 to 1.54E-22. Therefore, it was verified that the probability distribution of the calibration parameter was statistically valid and improved in terms of reducing uncertainty by integrating all available test data using the proposed method.

**Fig. 16** An electro-hydraulic braking system called the IDB system: **a** assembly diagram of the system, and **b** schematic diagram of the pedal feel simulator



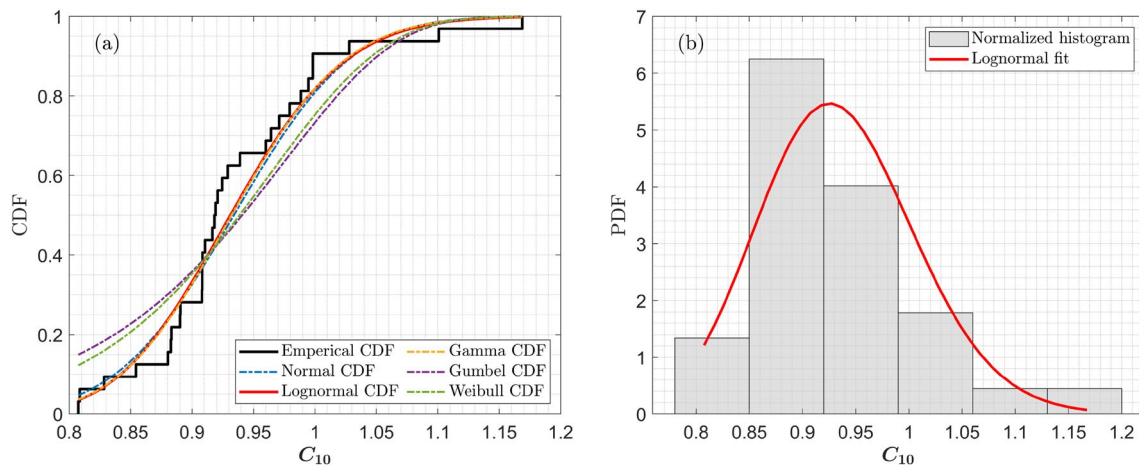
### 4.2 Engineering example: Pedal feel simulator, a component of an integrated dynamic brake (IDB)

The objective of this example is to examine whether the proposed method can also be effectively applied in real-world engineering problems. For this reason, the industrial engineering model introduced is the pedal feel simulator, which comprises the electro-hydraulic brake system shown in Fig. 16a. The model is shown in Fig. 16b is intended to reproduce the same pedal sensation to the driver of an electric vehicle or hybrid vehicle as the conventional hydraulic brake operated at the negative pressure of the engine (Wachter et al. 2019). As a rational material to artificially produce the existing sophisticated nonlinear pedal feeling, the ethylene propylene-based rubber has been generally used, and it is crucial to define the material property model accurately for the virtual validation of design specifications and performance prediction through the finite element analysis.

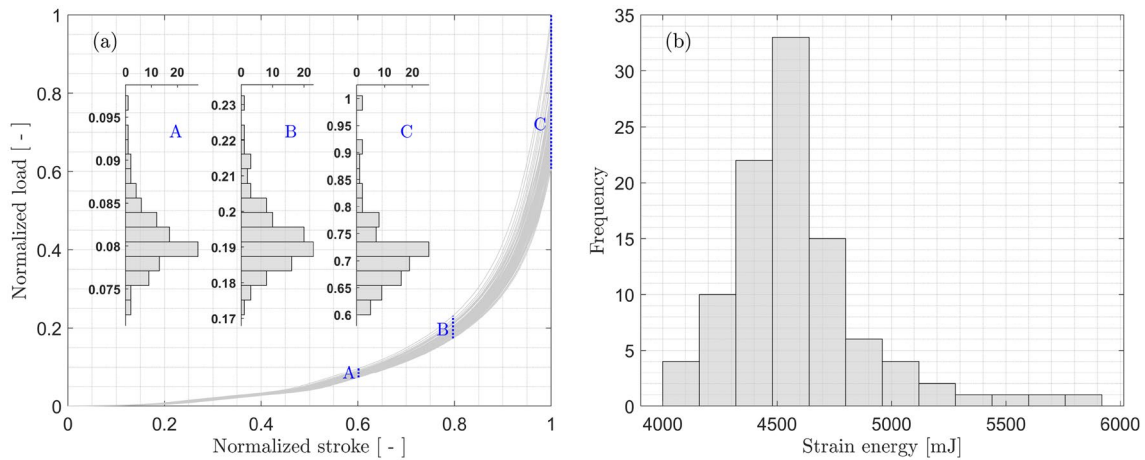
In order to express the large deformation of a material, which is called hyperelasticity, a phenomenological constitutive model defined as a function of strain energy potential ( $W$ ) with respect to the principal stretches ( $\lambda_1, \lambda_2, \lambda_3$ ) or the strain invariants ( $I_1, I_2, I_3$ ) is commonly used (Hossain et al. 2013; Steinmann et al. 2012). In this problem, the strain energy potential model of a Neo-Hookean form was adopted by considering the operating conditions and the Drucker stability of material for the stable performance of optimization in the subsequent prior distribution estimation, and is represented as (ABAQUS Documentation 2014; Marckmann et al. 2006; Romanov 2001)

$$W = C_{10}(\bar{I}_1 - 3) + \frac{1}{B_1}(J^{el} - 1)^2 \tag{22}$$

where  $C_{10}$  and  $B_1$  are the material parameters to be determined;  $J^{el}$  is the elastic volume ratio; and  $\bar{I}_1$  means the first deviatoric strain invariant defined as  $\bar{I}_1 = \lambda_1^2 + \lambda_2^2 + \lambda_3^2$  by deviatoric stretch  $\lambda_i$ . Assuming that the material is fully incompressible, the second term in Eq. (22) representing the volumetric part is negligible, so it is expressed only as the first term representing the deviatoric part. As a result, only  $C_{10}$  is the material parameter to be determined through the curve fitting with the provided coupon test results. In order to properly characterize the various behaviors of a material, coupon tests for various deformation modes such as uniaxial, equibiaxial, planar, and volumetric tests are necessary. In this problem, the null hypothesis that the 32 material properties obtained from the coupon test were extracted from the lognormal distribution was accepted at the 5% significance level through the Kolmogorov–Smirnov and Anderson–Darling Goodness-of-Fit test for the 5 candidate distributions as shown in Fig. 17a below. In addition, the distribution of material properties,  $C_{10}$ , is shown in Fig. 17b, and the sample statistics  $\bar{\mu}_{C_{10}}$  and  $\bar{\sigma}_{C_{10}}$  are 0.9340 and 0.0735, respectively. However, because the test conditions to be considered for reproducing only a specific deformation mode are quite difficult, sometimes the components test can be rather simple as in this problem (Kim et al. 2019; Moreira et al. 2013). Thus, the unknown model variable,  $C_{10}$ , is characterized by the proposed method using a number of output (component) test data and a small number of input (coupon) test data. Subsequently, the validation phase is carried out also using input and output test data.



**Fig. 17** Model variable (32 material properties obtained through coupon test) characterization: **a** Goodness-of-Fit test result, and **b** histogram with normalization set to PDF value

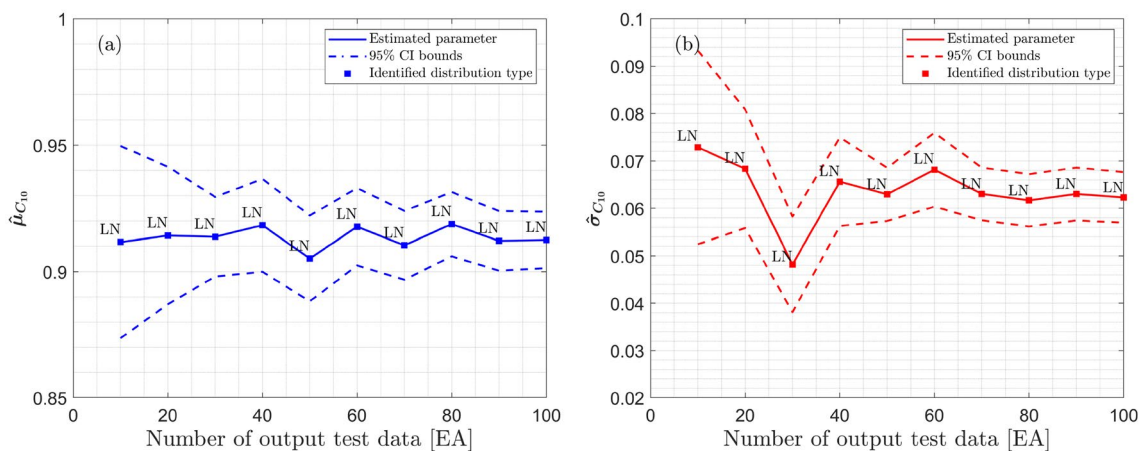


**Fig. 18** Experimental test results for one hundred feeling damper samples: **a** load – stroke curves with load histograms at checkpoints (A, B, and C), and **b** histogram of the strain energies representing each curve

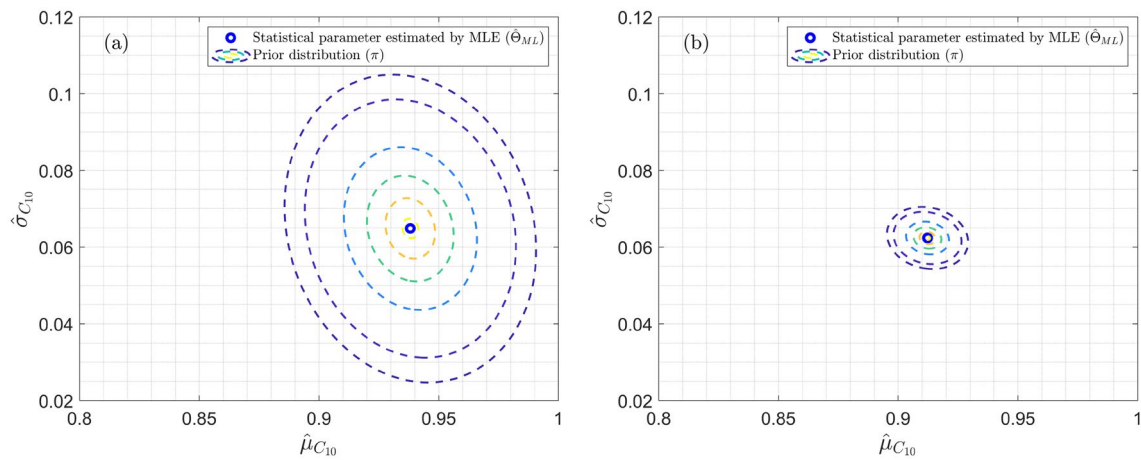
**4.2.1 Prior distribution with the output test data (strain energy stored in the feeling damper)**

As the given engineering model reproduces the desired pedal effort (i.e., applied load) by the compression of the feeling damper placed between the counterparts, as shown in Fig. 16b, the output performance of the unit is obtained in the form of a pedal effort curve according to the compression stroke. Compared to the coupon test, 100 individual feeling dampers were tested relatively easily using a universal testing machine, and the output test data are shown in Fig. 18 below. As shown in Fig. 18a, the load-stroke curves monotonically increase in direct proportion to variation of the  $C_{10}$  value. Based on these observations, for the convenience of the statistical model calibration

procedure, each curve could be defined as one quantitative value, strain energy ( $s$ ), and a histogram for a hundred strain energies is shown in Fig. 18b. As mentioned in Sect. 4.1.1, this output test data vector  $\mathbf{y}^e = [s_1, s_2, \dots, s_{n_c}]^T$  allows the prediction of the prior distribution of the statistical parameter vector  $\Theta = [\mu_{C_{10}}, \sigma_{C_{10}}]^T$  of the calibration parameter through the OBMC procedure. The numerical analysis model of the pedal feel simulator required during this procedure was modeled with 2-D axisymmetric elements (e.g., CAX4H, RAX2) using ABAQUS®, the commercial finite element code, and the DACEFIT was utilized to establish a Kriging surrogate model. The Kriging surrogate model was constructed with 5 samples generated by the Latin hypercube sampling by the maximin criterion with 1,000 iterations (Viana et al. 2013). The range of hyperparameter was set to [0.001, 20], and the Gaussian correlation function and zeroth-order polynomial



**Fig. 19** Estimation performance of the Fisher information according to the quantity of output test data: **a** estimated mean with 95% CI bounds, and **b** estimated standard deviation with 95% CI bounds



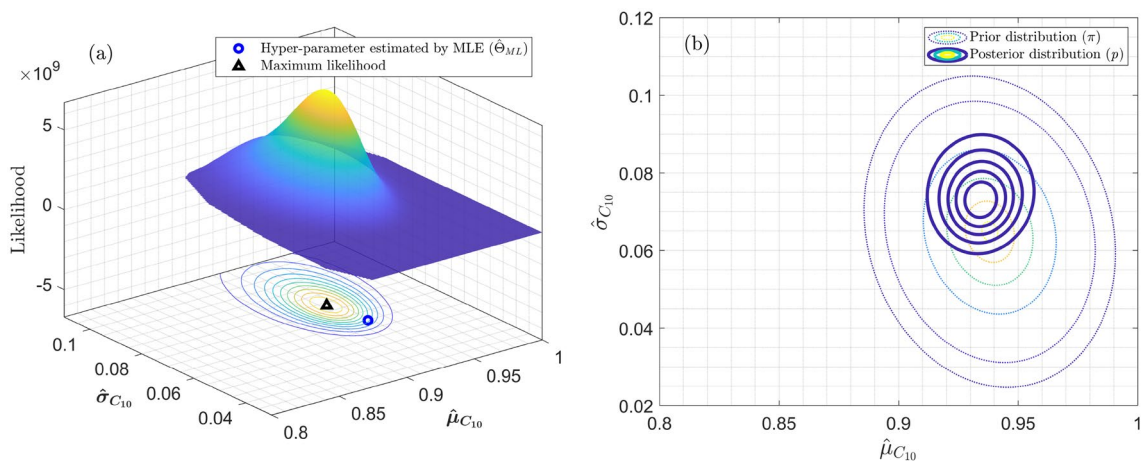
**Fig. 20** Comparison of the prior distribution estimation according to the quantity of output test data (lines in the contour plot represent the same probability of 10, 30, 50, 70, 90, and 95% in the outward direction): **a** considering 10 output test data, and **b** considering 100 output test data

regression function were applied. (Kang et al. 2019). The normalized leave-one-out cross-validation error was used as an index for evaluating the accuracy of the Kriging surrogate model and was 0.0017 for the model used in this example which is an acceptable accuracy level according to the previous studies (Blatman et al. 2010; Kalinina et al. 2020; Khalil et al. 2021).

The 95% CI for the estimated statistical parameters of a calibration parameter is found to narrow as the number of output test data utilized increases, as shown in Fig. 19. Furthermore, the probability distribution type of the calibration parameter is also predicted to be a lognormal distribution (tagged with a symbol of LN in Fig. 19) similar to the results in Fig. 17. Then, the estimated prior distribution is depicted in the form of a bivariate normal distribution by Eq. (21) as shown in Fig. 20, and estimation precision increases as more output test data from 10 to 100.

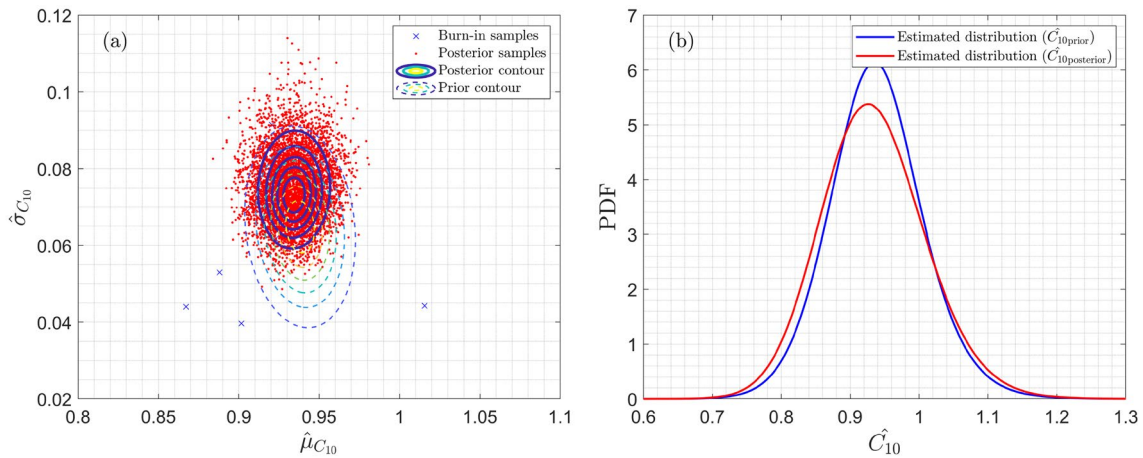
#### 4.2.2 Posterior distribution with the input test data (material parameter obtained by coupon test)

As in the procedure in Sect. 4.1.2, the prior distribution estimated from 10 ( $n_c$ ) of the given output test data is updated to the posterior distribution using input test data. In order to avoid the redundant use of input test data, the data required for each calibration and validation phase was divided into 20 ( $m_c$ ) and 12 ( $m_v$ ), respectively (Campbell 2006). The contour plot of the likelihood calculated by reflecting the 20 input test data in the lognormal distribution, the estimated distribution type of the unknown model variable in the previous step, is shown in Fig. 21a. Also, the 100,000 statistical parameter values of the lognormal distribution required for each probability calculation are sampled from the distribution in Fig. 20a. In Fig. 21a, the difference present in MLE using the input and output test data can be exhibited.

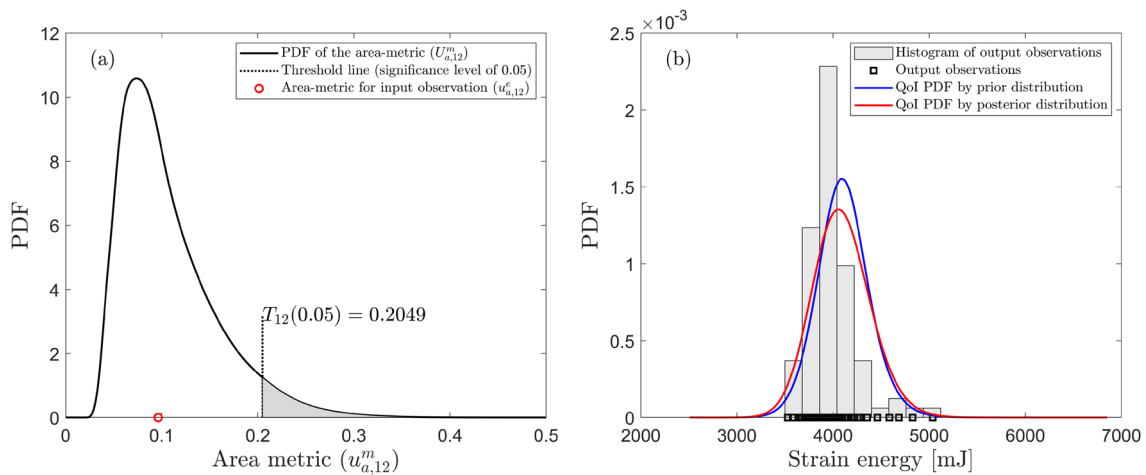


**Fig. 21** Improvement from the prior distribution to the posterior distribution: **a** likelihood contour calculated by reflecting input test data from the prior distribution, and **b** contour plots of the prior and posterior distribution





**Fig. 22** Characterization of the probability model of an unknown model variable through the MCMC: **a** statistical parameter sampling from the posterior distribution, and **b** calibrated statistical model



**Fig. 23** Statistical model validation for the estimated calibration parameter under the conditions of Fig. 22b: **a** hypothesis test using additional 12 input test data, and **b** likelihood calculation of additional 90 output test data for the propagated QoI distribution based on the prior and posterior distribution

Figure 21b, which shows the updated posterior distribution, confirms that consideration of input test data in the statistical model calibration can reduce the epistemic uncertainty about the statistical parameters of the unknown model variable  $C_{10}$ .

Ultimately, in order to derive the estimated probability model of  $C_{10}$ , 100,000 statistical parameter values were sampled with the RWM algorithm as shown in Fig. 22a, and KDE was performed for a total of 100 million samples by extracting 1,000  $\hat{C}_{10}$  at each point, as shown in Fig. 22b. From the results of Fig. 22a, it appears that the mode of mean distribution in the posterior distribution compared to the prior distribution slightly decreases, and that of the standard deviation slightly increases. These results are also expressed in the PDF plot of Fig. 22b and show that the mode of PDF derived from the

posterior distribution moves to the left and is somewhat widely distributed than that obtained from the prior distribution.

### 4.2.3 Statistical model validation

The hypothesis test was performed at a significance level of 5% using 12 input test data ( $m_v$ ) that were not used during the calibration phase in Sect. 4.2.2, and the result is shown in Fig. 23a. Since the area-metric value for the 12 observations is 0.1, which is less than the threshold  $T_{12}(0.05)$  of the selected significance level, it can be concluded that the model is valid under the given conditions. Furthermore, the likelihood was calculated using 90 observations ( $n_v$ ) at the validation site and the QoI distribution propagated from



the result of Fig. 23b. The likelihood values derived from the prior and posterior distribution increased by 13.7 times from 8.83E-275 to 1.21E-273, confirming that the proposed method was also valid in this example.

## 5 Conclusion

The biggest obstacle in the statistical model calibration that we face in reality is the lack of available data, and a practical method to solve this problem is presented in this research. The notable parts of the proposed method are that calibration is carried out in the statistical parameter domain of the unknown model variable to maintain the conservativeness of the estimation results under an insufficient data environment, and the epistemic uncertainty could be reduced by consolidating available input and output test data employing the Bayes' theorem. The output test data is applied to the OBMC to derive the prior distribution from the calculated MLE and FIM. Then, by multiplying the prior distribution and the likelihoods calculated using the input test data, the posterior distribution of the statistical parameters of the unknown model variable can be derived. Eventually, the probability model of the unknown model variable is obtained using MCMC and KDE methods from the estimation results of the statistical parameter domain. As the results of applying the proposed method to numerical and engineering examples handled in the real field, the intended effects were verified, so it is expected that the method will help solve the problem of insufficient data in the statistical model calibration. In addition, it is necessary to expand its application not only to univariate distributions but also to multivariate distributions, and in the future, as confirmed in Sect. 4.1.2 for the quantity of each test data, research will be conducted to suggest an optimal combination by considering the cost and time required to procure the test data.

## Declarations

**Conflict of interest** The authors declare that they have no conflict of interest.

**Replication of results** The results presented in this paper are obtained from our in-house MATLAB codes. For academic use only, contact the corresponding authors to get a copy of the source codes.

## References

ABAQUS Documentation. (2014) Dassault Systèmes Simulia Corp  
 Arendt PD, Apley DW, Chen W (2012a) Quantification of model uncertainty: Calibration, model discrepancy, and identifiability. *J Mech Des* 134(10):7390

- Arendt PD, Apley DW, Chen W, Lamb D, Gorsich D (2012b) Improving identifiability in model calibration using multiple responses. *J Mech Des* 10:7573
- Bi Z (2018) Finite element analysis applications: A Systematic and practical approach. Academic Press, Cambridge
- Bishop CM (2006) Pattern recognition and machine learning. Springer
- Blatman G, Sudret B (2010) An adaptive algorithm to build up sparse polynomial chaos expansion for stochastic finite element analysis. *Probabilistic Eng Mech* 25:183–197
- Campbell K (2006) Statistical calibration of computer simulations. *Reliab Eng Syst Saf* 91(10–11):1358–1363
- Cavanaugh JE, Shumway RH (1996) On computing the expected Fisher information matrix for state-space model parameters. *Stat Probab Lett* 26:347–355
- DeGroot MH, Schervish MJ (2011) Probability and statistics. Pearson Education. Hoboken
- Deisenroth MP, Faisal AA, Ong CS (2020) Mathematics for machine learning. Cambridge University Press
- Efron B, Hinkley DV (1978) Assessing the accuracy of the maximum likelihood estimator: Observed versus expected Fisher information. *Biometrika* 65(3):457–487
- Fahrmeir L, Kaufmann H (1985) Consistency and asymptotic normality of the maximum likelihood estimator in generalized linear models. *Ann Stat* 13(1):342–368
- Ferson S, Oberkampf WL (2009) Validation of imprecise probability models. *Int J Reliab Saf* 3(1):3–22
- Ferson S, Oberkampf WL, Ginzburg L (2008) Model validation and predictive capability for the thermal challenge problem. *Comput Methods Appl Mech Eng* 197(29–32):2408–2430
- Hess PE, Bruchman D, Assakkaf IA, Ayyub BM (2002) Uncertainties in material and geometric strength and load variables. *Nav Eng J* 114(2):139–166
- Higdon D, Nakhleh C, Gattiker J, Williams B (2008) A Bayesian calibration approach to the thermal problem. *Comput Methods Appl Mech Eng* 197(29–32):2431–2441
- Honarmandi P, Arroyave R (2020) Uncertainty quantification and propagation in computational materials science and simulation-assisted materials. *Integr Mater Manuf Innov* 9:103–143
- Hossain M, Steinmann P (2013) More hyperelastic models for rubber-like materials: consistent tangent operators and comparative study. *J Mech Behav Mater* 22(1–2):27–50
- Hu J, Zhou Q, McKeand A, Xie T, Choi SK (2021) A model validation framework based on parameter calibration under aleatory and epistemic uncertainty. *Struct Multidisc Optim* 63:645–660
- Jiang C, Hu Z, Liu Y, Mourelatos ZP, Gorsich D, Jayakumar P (2020) A sequential calibration and validation framework for model uncertainty quantification and reduction. *Comput Methods Appl Mech Eng* 368:113172
- Jung BC, Park J, Oh H, Kim J, Youn BD (2015) A framework of model validation and virtual product qualification with limited experimental data based on statistical inference. *Struct Multidisc Optim* 51:573–583
- Jung Y, Kang K, Cho H, Lee I (2021) Confidence-based design optimization for a more conservative optimum under surrogate model uncertainty caused by Gaussian process. *J Mech Des* 143(9):091701
- Kalinina A, Spada M, Vetsch DF, Marelli S, Wheelton C, Burgherr P, Sudret B (2020) Metamodeling for uncertainty quantification of a flood wave model for concrete dam breaks. *Energies* 13(14):3685
- Kang YJ, Noh Y, Lim OK (2018) Kernel density estimation with bounded data. *Struct Multidisc Optim* 57:95–113
- Kang K, Qin C, Lee BJ, Lee I (2019) Modified screening-based Kriging method with cross validation and application to engineering design. *Appl Math Model* 70:626–642
- Kennedy MC, O'Hagan A (2001) Bayesian calibration of computer models. *J R Stat Soc Ser B Methodol* 63(3):425–464

- Khalil M, Teichert GH, Alleman C, Heckman NM, Jones RE, Garikipati K, Boyce BL (2021) Modeling strength and failure variability due to porosity in additively manufactured metals. *Comput Methods Appl Mech Eng* 373:113471
- Kim S, Shin H, Rhim S, Rhee KY (2019) Calibration of hyperelastic and hyperfoam constitutive models for an indentation event of rigid polyurethane foam. *Compos B Eng* 163:297–302
- Lee G, Kim W, Oh H, Youn BD, Kim NH (2019a) Review of statistical model calibration and validation—from the perspective of uncertainty structures. *Struct Multidisc Optim* 60:1619–1644
- Lee G, Son H, Youn BD (2019b) Sequential optimization and uncertainty propagation method for efficient optimization-based model calibration. *Struct Multidisc Optim* 60:1355–1372
- Li M, Wang Z (2018) Confidence-driven design optimization using Gaussian process metamodeling with insufficient data. *J Mech Des* 140(12):121405
- Li W, Chen W, Jiang Z, Lu Z, Liu Y (2014) New validation metrics for models with multiple correlated responses. *Reliab Eng Syst Saf* 127:1–11
- Lim W, Lee TH, Kang S, Cho S (2016) Estimation of body and tail distribution under extreme events for reliability analysis. *Struct Multidisc Optim* 54:1631–1639
- Linnet K (2000) Nonparametric estimation of reference intervals by simple and bootstrap-based procedure. *Clin Chem* 46:867–869
- Liu Y, Chen W, Arendt P, Huang HZ (2011) Toward a better understanding of model validation metrics. *J Mech Des* 133(7):071005
- Ly A, Marsman M, Verhagen J, Grasman R, Wagenmakers E (2017) A Tutorial on Fisher information. *J Math Psychol* 80:40–55
- Marckmann G, Verron E (2006) Comparison of hyperelastic models for rubber-like materials. *Rubber Chem Technol* 79(5):835–858
- McFarland J, Mahadevan S (2008) Calibration and uncertainty analysis for computer simulations with multivariate output. *AIAA J* 46(5):1253–1265
- MIL-HDBK-5H (1998) *Metallic Material and Elements for Aerospace Vehicle Structures*. Tech. Rep. MIL-HDBK-5H, U.S. Department of Defense
- Moon M, Choi KK, Cho H, Gaul N, Lamb D, Gorsich D (2017) Reliability-based design optimization using confidence-based model validation for insufficient experimental data. *J Mech Des* 139(3):031404
- Moon M, Choi KK, Gaul N, Lamb D (2019) Treating epistemic uncertainty using bootstrapping selection of input distribution model for confidence-based reliability assessment. *J Mech Des* 141(3):031402
- Moreira DC, Nunes NCS (2013) Comparison of simple and pure shear for an incompressible isotropic hyperelastic material under large deformation. *Polym Test* 32(2):240–248
- Myung JI, Navarro DJ (2005) Information matrix. *Encyclopedia of Statistics in Behavioral Science* 2:923–924
- Nah JS, Lee J (2020) Reliability assessment of display delamination considering adhesive properties based on statistical model calibration and validation. *Int J Mech Mater Des* 16:191–206
- Oh H, Wei H, Han B, Youn BD (2016) Probabilistic lifetime prediction of electronic packages using advanced uncertainty propagation analysis and model calibration. *IEEE Trans Compon Packaging Manuf Technol* 6(2):238–248
- Oh H, Choi H, Jung JH, Yoon BD (2019) A robust and convex metric for unconstrained optimization in statistical model calibration-probability residual (PR). *Struct Multidisc Optim* 60:1171–1187
- Pashazadeh S, Sharifi M (2008) Reliability assessment under uncertainty using Dempster-Shafer and vague set theory. *IEEE Int Conf Comput Meas Syst Appl*, Istanbul, Turkey, July 14–16
- Picheny V, Kim NH, Haftka RT (2009) Application of bootstrap method in conservative estimation of reliability with limited samples. *Struct Multidisc Optim* 41:205–217
- Prokopenko M, Lizier JT, Obst O, Wang R (2011) Relating Fisher information to order parameters. *Phys Rev E* 84:041116
- Ralph CS (2014) *Uncertainty quantification: Theory, implementation, and applications*. Society for Industrial and Applied Mathematics, Philadelphia
- Rao SS, Annamdas KK (2008) Evidence-based fuzzy approach for the safety analysis of uncertain systems. *AIAA J* 46(9):2383–2387
- Romanov KI (2001) The drucker stability of a material. *J Appl Math Mech* 65(1):15–162
- Roy CJ, Oberkampf WL (2011) A Comprehensive framework for verification, validation, and uncertainty quantification in scientific computing. *Comput Methods Appl Mech Eng* 200:2131–2244
- Sargsyan K, Najm HN, Ghanem R (2015) On the statistical calibration of physical models. *Int J Chem Kinet* 47(4):246–276
- Son H, Lee G, Kang K, Kang Y, Youn BD, Lee I, Noh Y (2020) Industrial issues and solutions to statistical model improvement: a case study of an automobile steering column. *Struct Multidisc Optim* 60:1739–1756
- Sourati J, Akcakaya M, Leen TK, Erdogmus D, Dy JG (2017) Asymptotic analysis of objectives based on Fisher information in active learning. *J Mach Learn Res* 18:1–41
- Steinmann P, Hossain M, Possart G (2012) Hyperelastic models for rubber-like materials: consistent tangent operators and suitability for Treloar's data. *Arch Appl Mech* 82:1183–1217
- Sun NZ, Sun A (2015) *Model calibration and parameter estimation: For environmental and water resource systems*. Springer-Verlag, New York
- Trucano TG, Swiler LP, Igusa T, Oberkampf WL, Pilch M (2006) Calibration, validation, and sensitivity analysis: What's what. *Reliab Eng Syst Saf* 91(10–11):1331–1357
- Vakilzadeh MK, Yaghoubi V, Johansson AT, Abrahamsson TJS (2017) Stochastic finite element model calibration based on frequency responses and bootstrap sampling. *Mech Syst Signal Process* 88:180–198
- Viana FAC, Haftka RT, Watson LT (2013) Efficient global optimization algorithm assisted by multiple surrogate technique. *J Glob Optim* 56:669–689
- Villaverde AF, Fröhlich F, Weindl D, Hasenauer J, Banga JR (2019) Benchmarking optimization methods for parameter estimation in large kinetic models. *Bioinformatics* 35(5):830–838
- Vrugt JA (2016) Markov chain Monte Carlo simulation using the DREAM software package: Theory, concepts, and MATLAB implementation. *Environ Model Softw* 75:273–316
- Wachter E, Ngu TQ, Alirand M (2019) Virtual simulation of an electrohydraulic braking system. *ATZ Worldwide* 121:54–59. <https://doi.org/10.1007/s38311-019-0070-y>
- Wehrens R, Putter H, Buydens LMC (2000) The bootstrap: a tutorial. *Chemom Intell Lab Syst* 54:35–52
- Wu YT, Shin Y, Sues R, Cesare M (2001) Safety-factor based approach for probability-based design optimization. In: *Proceedings of the 42nd AIAA/ASME/ASCE/AHS/ASC Struct Struct Dyn Mater Conf*, number AIAA-2001-1522, Seattle, WA, 196, 199–342
- Xi Z (2019) Model-based reliability analysis with both model uncertainty and parameter uncertainty. *J Mech Des* 141(5):051404
- Xi Z, Jung BC, Youn BD (2012) Random field modeling with insufficient data sets for probability analysis. In: *Proc Annu Reliab Maintainab Symp*, Reno, NV, USA, January 23–26
- Xiong Y, Chen W, Tsui K, Apley DW (2009) A better understanding of model updating strategies in validating engineering models. *Comput Methods Appl Mech Eng* 198:1327–1337
- Youn BD, Jung BC, Xi Z, Kim SB, Lee WR (2011) A hierarchical framework for statistical model calibration in engineering product development. *Comput Methods Appl Mech Eng* 200:1421–1431