Nonparametric Bayesian stochastic model updating with hybrid uncertainties

Masaru Kitahara [a, *], Sifeng Bi [b], Matteo Broggi [a], Michael Beer [a, c, d]

[a] *Leibniz Universität Hannover, Institute for Risk and Reliability, Callinstrasse 34, Hannover, Germany*
[b] *Beijing Institute of Technology, School of Aerospace Engineering, Beijing, China*
[c] *University of Liverpool, Institute for Risk and Uncertainty, Peach Street L69 7ZF, Liverpool, United Kingdom*
[d] *Tongji University, International Joint Research Center for Engineering Reliability and Stochastic Mechanics, Shanghai 200092, China*
* Correspondence author. E-mail address: masaru.kitahara@irz.uni-hannover.de (M. Kitahara).

Abstract: This work proposes a novel methodology to fulfil the challenging expectation in stochastic model updating to calibrate the probabilistic distributions of parameters without prior knowledge about the distribution format. To achieve this task, an approximate Bayesian computation model updating framework is developed by employing staircase random variables and the Bhattacharyya distance. In this framework, parameters with aleatory and epistemic uncertainties are described as staircase random variables. The discrepancy between model predictions and observations are then quantified by the Bhattacharyya distance-based approximate likelihood. In addition, a Bayesian updating using the Euclidian distance is performed as preconditioner to avoid non-unique solutions. The performance of the proposed procedure is demonstrated with two exemplary applications, a simulated shear building model example and a challenging benchmark problem for uncertainty treatment. These examples demonstrate feasibility of the combination of staircase random variables and the Bhattacharyya distance in stochastic model updating and uncertainty characterization.

Keywords:
Stochastic model updating; Approximate Bayesian computation; Nonparametric probability-box; Staircase random variable; Bhattacharyya distance

---

1. Introduction

It has been widely accepted uncertainties should be appropriately considered in the campaign of model updating. The uncertainties can be typically classified into two categories, i.e., aleatory and epistemic uncertainties [1,2]. Aleatory uncertainty is inherent variation or randomness, and therefore cannot be reduced, but it enables to be described as precise probability models. Conversely, epistemic uncertainty is due to lack of knowledge, and is not completely avoidable, although it can be reduced through model updating using available data.

The complexity of model updating depends on the presence of different level of uncertainties. Deterministic model updating generally accounts for the presence of only epistemic uncertainty, in which parameters are unknown-but-fixed constants and are described as non-probabilistic models, such as interval/convex models [3] and fuzzy set theory [4]. Deterministic model updating aims at a single set of parameter values and at generating a single model prediction with maximum fidelity with regard to the observation.

On the other hand, the presence of both aleatory and epistemic uncertainties simultaneously (i.e. hybrid uncertainties) is considered by stochastic model updating. In this circumstance, parameters are described as imprecise probability models, where parameters are indeed aleatory uncertainty but their distribution parameters, e.g., mean and variance, are epistemic uncertainty. Commonly used imprecise probability models include evidence theory [5], probability-box (also known as p-box) [6], and fuzzy probability model [7]. In particular, those parameterized ones such as parametric p-box have attracted the most attentions due to their simplicity and ease of applications. Stochastic model updating aims at not the single set of parameter values but a reduced space of epistemic uncertainty and at generating stochastic model predictions capable to represent uncertainty characteristics of multiple sets of observations.

The uncertainties during stochastic model updating, i.e., the reasons of the discrepancy between the model predictions and observations, can be summarized as follows:

- Parameter uncertainty. The input parameters of the numerical model, such as material properties and boundary conditions, are imprecisely determined.
- Modelling uncertainty. The numerical model always contains inevitable simplifications and approximations of the physical system.
- Measurement uncertainty. The measurements are driven by hard-to-control randomnesses, such as environmental noises and measurement system errors.

A wide range of stochastic model updating methods has been investigated, such as perturbation method [8,9], Monte Carlo approach [10,11], and Bayesian inference [12,13]. No matter which method is performed in stochastic model updating, it is significant to define a comprehensive uncertainty quantification (UQ) metric capable to quantify the statistical discrepancy between model predictions and observations due to the above uncertainties. The Euclidian distance, Mahalanobis distance, and Bhattacharyya distance are different levels of distance metrics. These distances have been investigated as UQ metrics and the Bhattacharyya distance is demonstrated to be able to capture a higher level of statistical information from the investigating sample sets [11]. Moreover, Bi et al. [14] developed a Bayesian model updating framework, in which the Bhattacharyya distance was used as the UQ metric to define an approximate likelihood function by the approximate Bayesian computation (ABC) method [15,16]. This framework has been demonstrated to be a comprehensive updating procedure with the capability to recreate wholly the distribution of target observations.

However, stochastic model updating in the literature including Bi et al. [14] is generally based on the parameterized imprecise probability models; hence, it relies upon the pre-hypothesis of the distribution format to propagate epistemic uncertainty into parameters. For instance, in the NASA UQ challenge problem 2014 [17], which has gained attentions as the real-size practical uncertainty quantification problem, prior information of the distribution format is fully provided to perform model updating. On the other hand, it is often happened that information of the distribution format is unknown due to scarce and incomplete available data for the parameters. Hence, it is desired to develop a nonparametric model updating framework, in which epistemic uncertainty is propagated into parameters without prior knowledge about the distribution format.

Crespo et al. [18] recently provided a family of random variables having a bounded support set and prescribed values for the first four moments. The variables are called staircase random variable since the density function is determined as piecewise constant functions on pre-defined subintervals partitioning the support set. In addition, moment constraints for the existence of such variables are obtained as a series of inequalities conditioned upon the support set. As a consequence, the staircase random variable enables to describe a wide range of density shapes, including very peaked and/or multimodal distributions. The staircase random variable belongs to the precise probability models. Nevertheless, its combination with the non-probabilistic models has a potential to provide a non-parameterized imprecise probability model, where parameters are indeed aleatory uncertainty but the first four moments are epistemic uncertainty.

The objective of this work is consequently to propose a novel methodology which fulfils the challenging expectation in stochastic model updating to calibrate the probabilistic distribution of parameters without prior knowledge about the distribution format. To achieve this objective, an ABC model updating framework is developed by employing the Bhattacharyya distance and staircase random variable. At the same time, a Bayesian updating using the Euclidian distance is performed as preconditioner to avoid non-unique solutions. This framework is independent of the distribution format of investigating parameters; thus, it shows clear advantages in calibrating parameters whose probabilistic distribution cannot be defined analytically. The proposed framework is demonstrated by a simple shear building model for illustration. Moreover, it is applied to the NASA UQ challenge problem 2014. We focus on solving Sub-problem A (uncertainty characterization), where distribution formats of the investigating parameters are given, however we redefine the problem ignoring them to demonstrate the performance of the proposed framework.

The rest of this paper is organized as follows. In Section 2, we describe the theoretical and methodological bases of the Bhattacharyya distance metric and staircase random variable. Section 3 outlines the novel development of the Bayesian updating with the staircase random variable, and the proposed two-step ABC updating framework. The principle and illustrative application is detailed in Section 4, using a simple shear building model for illustration, and in Section 5, concentrating on the demonstration of performance of the framework on the highly challenging NASA UQ problem. Finally, some conclusions are given in Section 6.

2. Theories and methodologies

2.1. Bhattacharyya distance metric

In the context of stochastic model updating, the investigating system is characterized as:

$$\mathbf{y} = h(\mathbf{x}) \tag{1}$$

where $\mathbf{x} = [x_1, x_2, \cdots, x_n]$ is a column vector of n input parameters; $\mathbf{y} = [y_1, y_2, \cdots, y_m]$ is a column vector of m output features; $h(\cdot)$ is the simulator. The simulator herein is usually presented as either a sophisticated numerical analysis code, e.g., finite element model, or a metamodel.

The uncertainties of the system are first characterized by input parameters described as precise probability models, non-probabilistic models, and imprecise probability models depending on the presence of different level of uncertainties (refer Section 1). The uncertainties are then propagated through the simulator into output features presenting various forms of uncertainty as well, such as probabilistic distributions, intervals, and fuzzy sets. In general, regardless of the form of uncertainty, randomly sampled values of parameters and features are used in stochastic model updating. Suppose the required sample size is N_{sim}, the simulator h is executed N_{sim} times to generate the sample set of the simulated features $\mathbf{Y}_{sim} \in \mathbb{R}^{N_{sim} \times m}$:

$$\mathbf{Y}_{sim} = [\mathbf{y}_1, \mathbf{y}_2, \cdots, \mathbf{y}_m], \text{with } \mathbf{y}_i = \left[y_{1i}, y_{2i}, \cdots, y_{N_{sim}i}\right]^T, \forall i = 1, 2, \cdots, m \tag{2}$$

In addition to the simulated features, observed features collected from the campaign of experiments or measurements are also required as the target of model updating. Suppose the number of observations is N_{obs}, the sample set of the observed features has a similar structure as Eq. (2), where only the number of rows is changed: $\mathbf{Y}_{obs} \in \mathbb{R}^{N_{obs} \times m}$. The objective of stochastic model updating can be then expressed as to minimize the discrepancy between $\mathbf{Y}_{obs}$ and $\mathbf{Y}_{sim}$ by updating uncertainty characteristics of the input parameters.

After the simulated and observed features are available, the UQ metric is defined to capture the discrepancy between $\mathbf{Y}_{obs}$ and $\mathbf{Y}_{sim}$. The classical Euclidian distance metric is expressed as:

$$d_E(\mathbf{Y}_{obs}, \mathbf{Y}_{sim}) = \sqrt{(\overline{\mathbf{Y}}_{obs} - \overline{\mathbf{Y}}_{sim})(\overline{\mathbf{Y}}_{obs} - \overline{\mathbf{Y}}_{sim})^T} \tag{3}$$

where $\overline{\mathbf{Y}}_{\blacksquare}$ is a row vector of means of the features. The Euclidian distance is a point-to-point distance between the centre of mass of two sample sets and is generally used in deterministic model updating. Comparatively, in stochastic model updating, it is more desirable to employ a more comprehensive metric capable to consider not only the means but also a higher level of statistical information, such as variances, covariances, and even distribution shapes.

The Bhattacharyya distance is herein proposed as a stochastic metric measuring the degree of overlap between distributions of two sample sets. Its original definition is given as:

$$d_B(\mathbf{Y}_{obs}, \mathbf{Y}_{sim}) = -\log\left[\int_{\mathbb{y}} \sqrt{p_{obs}(y) p_{sim}(y)}\, \mathrm{d}y\right] \tag{4}$$

where $p_{\blacksquare}(y)$ is the probability density function (PDF) of the feature sample; $\mathbb{y}$ is the m-dimensional feature space; $\int_{\mathbb{y}} \blacksquare\, \mathrm{d}y$ is the integration performed over the whole feature space. Differently from the

Euclidian distance, the Bhattacharyya distance considers not only the means but also the variances, covariances, and even the distribution shapes. However, the direct evaluation of Eq. (4) is often not feasible because precise estimation of the PDF is generally unavailable due to the very limited number of observations. Bi et al. [14] thus proposed a so-called binning algorithm to evaluate the probability mass function (PMF) of a discrete distribution, such that the discrete Bhattacharyya distance is used instead. The PMF is a function to map possible values of a discrete random variable to probabilities of their occurrence [19]. The discrete Bhattacharyya distance is evaluated as [20]:

$$d_B(\mathbf{Y}_{obs}, \mathbf{Y}_{sim}) = -\log\left\{\sum_{i_m=1}^{n_{bin}} \cdots \sum_{i_1=1}^{n_{bin}} \sqrt{p_{obs}\left(b_{i_1,i_2,\cdots,i_m}\right)p_{sim}\left(b_{i_1,i_2,\cdots,i_m}\right)}\right\} \tag{5}$$

where $p_{\blacksquare}\left(b_{i_1,i_2,\cdots,i_m}\right)$ is the PMF value of the bin $b_{i_1,i_2,\cdots,i_m}$. The bin has m subscripts because it is generated under a m-dimensional joint PMF space.

The binning algorithm for the PMF calculation consists of the following steps.

1) Define the common interval I_i of both $\mathbf{Y}_{obs}$ and $\mathbf{Y}_{sim}$ according to the ith feature $\mathbf{y}_i$, $\forall i = 1, 2, \cdots, m$, by finding the general maximum and minimum values of $\mathbf{y}_i$ in both $\mathbf{Y}_{obs}$ and $\mathbf{Y}_{sim}$;
2) Within the defined interval, arbitrary decide the number of bins n_{bin};
3) Count the joint probability mass for each bin $p_{\blacksquare}\left(b_{i_1,i_2,\cdots,i_m}\right)$. Note that, the total number of bins in the m-dimensional feature space is $N_{bin} = n_{bin}^m$.

The principle of n_{bin} in Step 2) is that a larger n_{bin} leads to employing more detailed information of the distribution characteristics and to a larger value of the Bhattacharyya distance, while it also leads to a larger computational cost. In Ref. [14], n_{bin} is recommended to be $n_{bin} = \left\lceil \frac{\max(N_{obs}, N_{sim})}{10} \right\rceil$, where $\lceil \blacksquare \rceil$ denotes the upper integer of the investigating values.

2.2. Staircase random variables

In this study, parameters with hybrid uncertainties are described as staircase random variables. The staircase random variable x is constrained to have a bounded support set $[\underline{x}, \overline{x}]$ and a pair of variables $\mathbf{\theta}_x = [\mu, m_2, m_3, m_4]$ consisting of the mean μ, variance m_2, third-order central moment m_3, and fourth-order central moment m_4. Moment constraints of the pair of variables $\mathbf{\theta}_x$ to realize the staircase random variable x associated with the support set $[\underline{x}, \overline{x}]$ are given as a series of inequalities: $\Theta = \{\mathbf{\theta}_x : g(\mathbf{\theta}_x) \leq 0\}$ [21,22], and are summarized in Table 1.

Table 1. Moment constraints of the pair of variables $\mathbf{\theta}_x$.

	Moment constraints	
Mean μ	$g_1 = \underline{x} - \mu$	$g_2 = \mu - \overline{x}$
Variance m_2	$g_3 = -m_2$	$g_4 = m_2 - v^{a}$
Third-order central moment m_3	$g_5 = m_2^2 - m_2(\mu - \underline{x})^2 - m_3(\mu - \underline{x})$	$g_6 = m_3(\overline{x} - \mu) - m_2(\overline{x} - \mu)^2 + m_2^2$
	$g_7 = 4m_2^2 + m_3^2 - m_2^2(\overline{x} - \underline{x})^2$	$g_8 = 6\sqrt{3}m_3 - (\overline{x} - \underline{x})^3$
	$g_9 = -6\sqrt{3}m_3 - (\overline{x} - \underline{x})^3$	
Fourth-order central moment m_4	$g_{10} = -m_4$	$g_{11} = 12m_4 - (\overline{x} - \underline{x})^4$
	$g_{12} = (m_4 - vm_2 - u^{a}m_3)(v - m_2) + (m_3 - um_2)^2$	$g_{13} = m_3^2 + m_2^3 - m_4m_2$

[a] $u = \underline{x} + \overline{x} - 2\mu$ and $v = (\mu - \underline{x})(\overline{x} - \mu)$.

Considering that the chosen support set $[\underline{x}, \overline{x}]$ is partitioned into n_b subintervals of equal length $\kappa = (\overline{x} - \underline{x})/n_b$, the staircase density function $f_x(x)$ is expressed as:

$$f_x(x) = \begin{cases} l_i & \forall x \in (x_i, x_{i+1}], \text{for } 1 \leq i \leq n_b \\ 0 & \text{otherwise} \end{cases} \tag{6}$$

where l_i is the staircase density height of the ith bin; $x_i = \underline{x} + (i-1)\kappa$ is the partitioning point of the ith bin. Note that, l_i satisfies that $l_i \geq 0$ for all bins and $\kappa \sum_{i=1}^{n_b} l_i = 1$. The staircase density heights $\boldsymbol{l}$ associated with Θ can be determined by solving a following convex optimization problem:

$$\hat{\boldsymbol{l}} = \underset{l \geq 0}{\operatorname{argmin}} \left\{ J(\boldsymbol{l}): \sum_{i=1}^{n_b} \int_{x_i}^{x_{i+1}} x l_i dx = \mu, \sum_{i=1}^{n_b} \int_{x_i}^{x_{i+1}} (x-\mu)^r l_i dx = m_r, r = 2,3,4 \right\} \tag{7}$$

where J is an arbitrary cost function. Eq. (7) can be written as:

$$\hat{\boldsymbol{l}} = \underset{l \geq 0}{\operatorname{argmin}} \{J(\boldsymbol{l}): \mathbf{A}(\boldsymbol{\theta}_x, n_b)\boldsymbol{l} = \mathbf{b}(\boldsymbol{\theta}_x), \boldsymbol{\theta}_x \in \Theta\} \tag{8}$$

where

$$\mathbf{A} = \begin{bmatrix} \kappa\mathbf{e} \\ \kappa\mathbf{c} \\ \kappa\mathbf{c}^2 + \kappa^3/12 \\ \kappa\mathbf{c}^3 + \kappa^3\mathbf{c}/4 \\ \kappa\mathbf{c}^4 + \kappa^3\mathbf{c}^2/2 + \kappa^5/80 \end{bmatrix}, \text{and } \mathbf{b} = \begin{bmatrix} 1 \\ \mu \\ \mu^2 + m_2 \\ m_3 + 3\mu m_2 + \mu^3 \\ m_4 + 4m_3\mu + 6m_2\mu^2 + \mu^4 \end{bmatrix}$$

where $\mathbf{c}$ is a column vector of the centre of the bin $c_i = (x_i + x_{i+1})/2$; $\mathbf{c}^n$ is the component wise nth power of $\mathbf{c}$; $\mathbf{e}$ is a unit vector.

Regarding with the cost function J, several optimality criteria, such as maximal entropy, minimal squared amplitude, and maximal log-likelihood can be employed. The cost function used in this study is expressed as:

$$J(\boldsymbol{l}) = \boldsymbol{l}^T \mathbf{I} \boldsymbol{l} \tag{8}$$

where $\mathbf{I}$ is the identity matrix. Employing this cost function yields a staircase random variable that minimizes the squared sum of the likelihood at the bins. Note that, while we do not investigate other cost functions, the choice of the cost function may affect the quality of model updating and thus it should be further investigated in the future work.

Convexity of the optimization problem in Eq. (7) enables to very efficiently calculate the staircase densities and to solve for practically smooth probability densities. These features makes the staircase random variable well suited for stochastic model updating where its repeated calculation is required. Moreover, the staircase random variable is independent of the distribution format and can describe a wide range of density shapes, including very peaked and/or multimodal distributions. This fulfills the expectation as a non-parameterized model for the proposed model updating framework, in which the epistemic uncertainty space is calibrated without prior knowledge about the distribution format of the parameters.

3. Nonparametric Approximate Bayesian computation

3.1. Bayesian model updating with staircase random variables

In this study, the well-known Bayesian inference is employed as the stochastic model updating methodology. The Bayesian inference is based on the Bayes' theorem [23]:

$$P(\boldsymbol{\theta}|\mathbf{Y}_{obs}) = \frac{P_L(\mathbf{Y}_{obs}|\boldsymbol{\theta})P(\boldsymbol{\theta})}{P(\mathbf{Y}_{obs})} \tag{9}$$

where $P(\boldsymbol{\theta})$ is the prior distribution of adjustable parameters $\boldsymbol{\theta}$, determined by the prior knowledge of the system and expert experience; $P(\boldsymbol{\theta}|\mathbf{Y}_{obs})$ is the posterior distribution of $\boldsymbol{\theta}$, representing the updated knowledge of $\boldsymbol{\theta}$ based on observations $\mathbf{Y}_{obs}$; $P(\mathbf{Y}_{obs})$ is the normalization factor (also known as the evidence) ensuring that the posterior distribution integrates to one; $P_L(\mathbf{Y}_{obs}|\boldsymbol{\theta})$ is the likelihood function of $\mathbf{Y}_{obs}$ for an instance of $\boldsymbol{\theta}$.

For the parameters with hybrid uncertainties described as staircase random variables, the first four moments $\boldsymbol{\theta}_x$ are considered as the adjustable parameters $\boldsymbol{\theta}$. Given the support set $[\underline{x}, \overline{x}]$, feasible intervals of $\boldsymbol{\theta}_x$ are determined based on the moment constraints $\boldsymbol{\theta}_x \in \Theta$ as:

$$\mu \in [\underline{x}, \overline{x}], m_2 \in \left[0, \frac{(\overline{x}-\underline{x})^2}{4}\right], m_3 \in \left[-\frac{(\overline{x}-\underline{x})^3}{6\sqrt{3}}, \frac{(\overline{x}-\underline{x})^3}{6\sqrt{3}}\right], m_4 \in \left[0, \frac{(\overline{x}-\underline{x})^4}{12}\right] \tag{10}$$

The prior distribution of $\boldsymbol{\theta}_x$ can be then discretely obtained as multiple sets of possible realizations of $\boldsymbol{\theta}_x$ within the feasible intervals satisfying the moment constraints. In this manner, only the support set of the parameters is required as prior knowledge of the system, however the feasible intervals in Eq. (10) are possible to be narrower if more detailed information about the parameters is available. At the same time, the parameters with only epistemic uncertainty are also capable to be handled in Bayesian updating and the parameters themselves are considered as the adjustable parameters. In this case, the prior distribution $P(\boldsymbol{\theta}) = P(\mathbf{x})$ are represented as auxiliary uniform distributions within the given intervals of the parameters.

One non-trivial component in Eq. (9) is the evidence $P(\mathbf{Y}_{obs})$, because the direct evaluation of the posterior PDF over the whole parameter space is quite difficult or even intractable especially for very peaked and/or multimodal distributions [24]. Hence, the well-known Bayesian inference algorithm, transitional Markov chain Monte Carlo (TMCMC) [25] is employed as an effective updating tool. TMCMC is essentially interpreted as an iterative approach sampling from a series of intermediate PDFs which will progressively converge to the true posterior distribution. The jth intermediate PDF is expressed as:

$$P_j \propto P_L(\mathbf{Y}_{obs}|\boldsymbol{\theta})^{\beta_j} P(\boldsymbol{\theta}) \tag{11}$$

where β_j is the so-called reduction coefficient. Its value starts from $\beta_0 = 0$ in the first iteration and progressively increases until reaching $\beta_m = 1$ in the last iteration. β_j is adaptively computed from samples of the previous step. Markov chains with the Metropolis-Hasting algorithm [26] propagate new samples starting from the ones with higher intermediate likelihood values, enabling to sample from the very complex posterior PDF. The readers can be referred to Refs. [25,27,28] for the details of TMCMC and its applications.

3.2. Two-step ABC updating framework

The likelihood function is the key component in Bayesian model updating, since it quantifies the degree of relevance of a model with a given instance of the adjustable parameters, by representing the possibility of the observations. Under the assumption of independence between observations, the likelihood function in Eq. (9) is theoretically defined as:

$$P_L(\mathbf{Y}_{obs}|\boldsymbol{\theta}) = \prod_{k=1}^{N_{obs}} P(\mathbf{Y}_k|\boldsymbol{\theta}) \tag{12}$$

where $P(\mathbf{Y}_k|\boldsymbol{\theta})$ is the PDF value of the kth observation $\mathbf{Y}_k$ conditional to the corresponding instance of the adjustable parameters $\boldsymbol{\theta}$. Eq. (12) requires to estimate the PDF for each of the N_{obs} observations, which introduce considerable computation burden. Moreover, precise estimation of the PDFs is only achieved by a large number of model evaluations to generate a large number of simulated features. Hence, the full likelihood evaluation can be almost infeasible for complex simulators.

The ABC method [15,16] is utilized to overcome the above issue by replacing the full likelihood function with an approximate likelihood function containing information of both the observations and adjustable parameters $\boldsymbol{\theta}$. In the approximate likelihood, any types of statistics can be employed to measure the discrepancy between the model predictions and observations, and thus it is natural to define it employing the distance metrics. Various functional formulas have been investigated in the literature, including the Gaussian [29], Epanechnikov [30], and sharp [28] functions. Regardless of the functional formulas, the basic principle of the approximate likelihood is that it should return a high value when the distance metric is small, while it penalizes the $\boldsymbol{\theta}$ instance when its corresponding distance metric is large. In this study, the approximate likelihood based on the Gaussian function is proposed as:

$$P_L(\mathbf{Y}_{obs}|\boldsymbol{\theta}) \propto exp\left\{-\frac{d^2}{\varepsilon^2}\right\} \tag{13}$$

where d is the distance metric; ε is the so-called width factor, which is a pre-defined coefficient controlling the centralization degree of the posterior distribution. A smaller ε corresponds to a more peaked posterior distribution, which is more likely to converge to the true value but requires more calculation for convergence. The choice of ε is hence based on specific applications and is commonly between 10^{-3} and 10^{-1} [29]. The distance-based approximate likelihood in Eq. (13) is a convenient connection between the distance metrics and Bayesian model updating with significantly reduced calculation cost. In addition, it provides a uniform framework for either deterministic or stochastic updating, simply driven by the employed metric is the Euclidian or Bhattacharyya distances.

By employing the Bhattacharyya distance metric, the proposed approximate likelihood enables to capture comprehensive uncertainty characteristics of both the model predictions and observations. Moreover, thanks to the features of the staircase random variable, the stochastic updating procedure is theoretically applicable regardless of the distribution format of the parameters, including the one which cannot be obtained analytically. Nevertheless, multimodality of the parameters, for instance, may lead to non-unique solutions (which will be further discussed in Section 4 through an illustrative example), and thus the direct application of the stochastic updating procedure cannot be utilized.

To cope with this issue, a two-step ABC updating framework is proposed as shown in Fig. 1. This framework starts from performing an outer Bayesian updating using the Euclidian distance. Step I is equivalent to a deterministic updating procedure with the target to identify all the possible solutions. The support sets of the parameters are then divided into appropriate sub-intervals, so that each sub-interval contain one possible solution. This preliminary procedure is necessary to avoid to end with a local solution in the main step. After that, comprehensive uncertainty characteristics of the parameters are further updated in step II via stochastic updating of the epistemic parameters, i.e., the first four moments of the staircase random variables, using the Bhattacharyya distance.

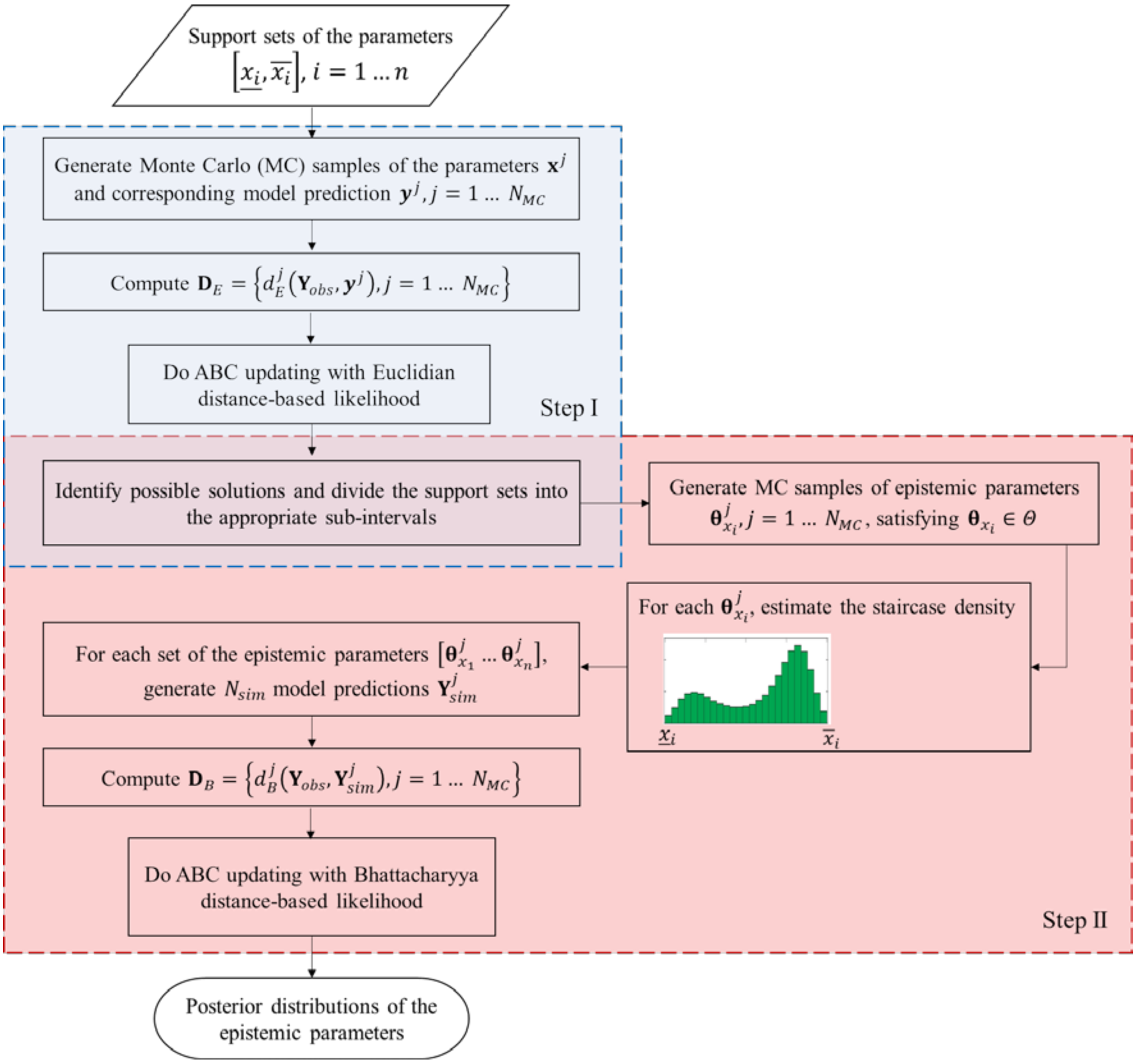

Fig. 1. Schematic of the two-step ABC updating framework.

4. Principle and illustrative application of the ABC updating framework

4.1. Problem description

The proposed two-step ABC updating framework with both the Euclidian and Bhattacharyya distances metrics is demonstrated on a two degree of freedom (DOF) shear building model shown in Fig. 2(a). This model was introduced by Beck and Au [24]. The first and second story masses are assumed as deterministic values with $m_1 = 16.531 \times 10^3$ kg and $m_2 = 16.131 \times 10^3$ kg. The first and second interstory stiffnesses are parameterized as $k_1 = \overline{k}x_1$ and $k_2 = \overline{k}x_2$, where $\mathbf{x} = [x_1, x_2]$ are the adjustable parameters to be identified, and $\overline{k} = 29.7 \times 10^6$ N/m is the nominal value of the stiffnesses.

In Beck and Au [24], the prior PDF $P(\mathbf{x})$ is given by uncorrelated lognormal distributions with most probable values 1.3 and 0.8 for x_1 and x_2, respectively, and unit standard deviations. Using the modal data, where the identified natural frequencies are $\tilde{f}_1 = 3.13$ Hz and $\tilde{f}_2 = 9.83$ Hz, the posterior PDF is formulated as:

$$P(\mathbf{x}|\mathbf{Y}_{obs}) = exp\left[-\frac{J(\mathbf{x})}{2\sigma^2}\right]P(\mathbf{x}) \tag{14}$$

where $\sigma = 1/16$ is the standard deviation of the prediction error and $J(\mathbf{x})$ is a modal measure-of-fit function given by:

$$J(\mathbf{x}) = \sum_{j=1}^{2} \lambda^2 \left[\frac{f_j^2(\mathbf{x})}{\tilde{f}_j^2} - 1\right]^2 \tag{15}$$

where $\lambda = 1$ is the weight and $f_j(\mathbf{x})$ is the jth natural frequency predicted by the model with the adjustable parameters $\mathbf{x}$. Fig. 2(b) illustrates the posterior PDF in Eq. (14) and it shows bimodality. It has been already demonstrated that this bimodal distribution can be achieved by several sampling methods, including the MCMC methods.

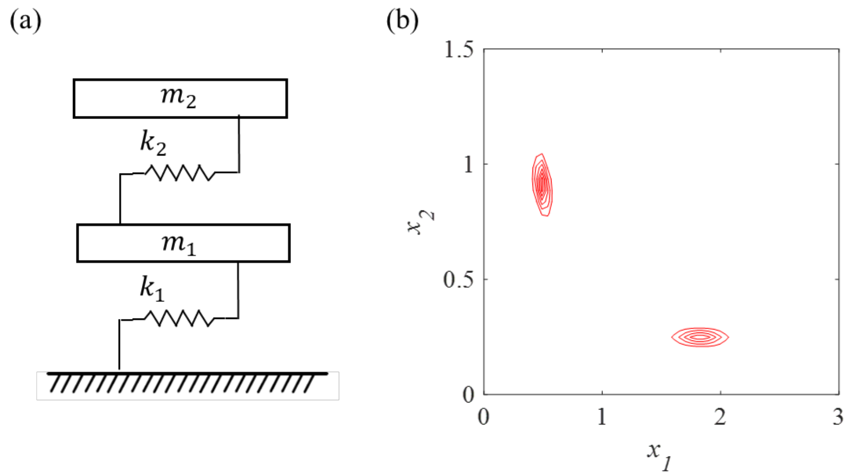

Fig. 2. (a) Two degree of freedom shear building model; (b) Posterior distribution in Ref. [24].

This problem can be interpreted as a deterministic updating of the parameters themselves using the single set of observations. However, its uncertain characteristics and observation data are hereby altered to demonstrate capabilities of the proposed stochastic updating framework with the presence of hybrid uncertainties. Both aleatory and epistemic uncertainties are involved in the model and are included by describing x_1 and x_2 as staircase random variables with given support sets. The natural frequencies f_1 and f_2 are taken as investigating features whose uncertainty is driven by the uncertain parameters x_1 and x_2. Their target probability distributions are assumed to be the posterior PDF in Eq. (14). In the altered problem setting, the identified natural frequencies $\tilde{f}_1$ and $\tilde{f}_2$ are not available, and thus the target probability distributions cannot be obtained analytically. Note that, in this case, the existing parametric stochastic updating procedures are not applicable. The support sets of x_1 and x_2 are detailed in Table 2.

Table 2. Uncertain parameters of the 2-DOF model.

Parameter	Uncertainty characteristic	Target distribution
x_1	$x_1 \in [0, 3.0]$	The marginal distribution of Eq. (14) for x_1
x_2	$x_2 \in [0, 1.5]$	The marginal distribution of Eq. (14) for x_2

The target of the updating procedure $\mathbf{Y}_{obs}$ is multiple sets of the features f_1 and f_2 obtained by assigning target probability distributions to x_1 and x_2, as shown in the last column of Table 2. The number of observations is $N_{obs} = 100$, generated by evaluating the model 100 times with parameters sampled from their assigned target distributions using the TMCMC method.

A single set of initial values of the first four moments $\boldsymbol{\theta}_{x_1}$ and $\boldsymbol{\theta}_{x_2}$ is set to be their possible realizations within the feasible intervals in Eq. (10), satisfying the moment constraints $\boldsymbol{\theta}_x \in \Theta$. These values are selected to be different from the target values, which return the staircase random variables matching with the target distributions, in order to illustrate how the imprecise model can produce outputs very different from the observations. Note that, those initial values are presented herein only for demonstration purpose as illustrated in Fig. 3 and Fig. 8. The two-step updating procedure is not really started from the initial values, but from the initial support sets of the parameters, as shown in the second column of Table 2.

Suppose the sample size is $N_{sim} = 1000$, N_{sim} parameter samples are generated from staircase densities with the initial values of the first four moments. The corresponding initial simulated output samples of f_1 and f_2 are obtained and illustrated in Fig. 3, along with the observed output samples. As shown in this figure, the objective of model updating herein is no longer a single updated point with maximum fidelity to a single observation, but updated distributions of the parameters which can represent the output samples as similar as the observed ones. Such parameter distributions can be estimated as the staircase densities with the updated first four moments. To achieve this objective, both the Euclidian and Bhattacharyya distances are employed as the UQ metrics in the ABC updating procedure.

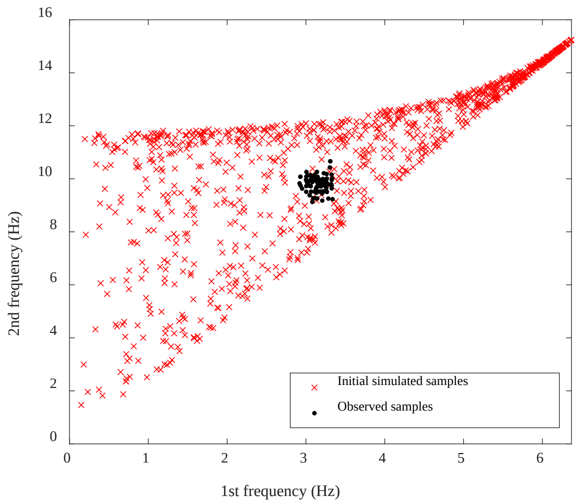

Fig. 3. Observed and initial simulated output samples.

4.2. Step I: determinisitic updating with the Euclidian distance metric

As shown in Figs. 2(b) and 3, the target distributions of the parameters show bimodality but the observed samples show unimodality; hence, it leads to non-unique solutions. In this circumstance, the direct application of the stochastic updating procedure may end with a local solution. To avoid the local solution, a deterministic updating of the parameters themselves is herein performed with the Euclidian distance metric to identify the possible solutions. There are two parameters x_1 and x_2, whose prior distributions are set to be uniform within the support sets in Table 2. When the Euclidian distance is taken as the metric, the geometric distance between a simulated sample and the centre of mass of the observed samples is measured, while the dispersion information of the observed samples cannot be considered.

In this section, the width factor in the distance-based likelihood is set as $\varepsilon = 0.1$, and totally eight TMCMC iterations are executed to reach convergence. Updated samples of the parameters from the posterior distribution are presented in Fig. 4, along with the target distributions. As shown in this figure, the updated samples clearly capture bimodality of the target distributions as same as the result in Ref [24]. It fulfils that deterministic updating focuses only on the means of the parameters and they are much easier to be properly updated compared with the higher moments. However, orientation and dispersion of the updated samples remain different from the target distributions. Hence, a more comprehensive metric is required in the second step to further reduce the discrepancy between the samples in Fig. 4 and the target distributions.

Based on the updated distributions of the parameters, their given support sets are divided into two sub-intervals, such that each sub-interval contain single possible solution. Table 3 presents the defined sub-intervals of the parameters.

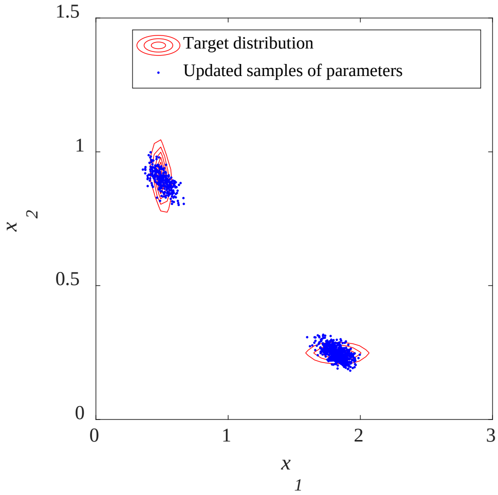

Fig. 4. Updated samples of parameters with Euclidian distance.

Table 3. Sub-intervals of parameters for stochastic updating.

Sub-interval	Uncertain characteristic	Epistemic parameters
I	$x_1 \in [0, 1.0]$	$\mu_1 \in [0, 1.0], m_{21} \in [0, 0.25], m_{31} \in \left[-\frac{1}{6\sqrt{3}}, \frac{1}{6\sqrt{3}}\right], m_{41} \in \left[0, \frac{1}{12}\right]$
	$x_2 \in [0.5, 1.5]$	$\mu_2 \in [0.5, 1.5], m_{22} \in [0, 0.25], m_{32} \in \left[-\frac{1}{6\sqrt{3}}, \frac{1}{6\sqrt{3}}\right], m_{42} \in \left[0, \frac{1}{12}\right]$
II	$x_1 \in [1.0, 3.0]$	$\mu_1 \in [1.0, 3.0], m_{21} \in [0, 1.0], m_{31} \in \left[-\frac{4}{3\sqrt{3}}, \frac{4}{3\sqrt{3}}\right], m_{41} \in \left[0, \frac{4}{3}\right]$
	$x_2 \in [0, 0.5]$	$\mu_2 \in [0, 0.5], m_{22} \in \left[0, \frac{1}{16}\right], m_{32} \in \left[-\frac{1}{48\sqrt{3}}, \frac{1}{48\sqrt{3}}\right], m_{42} \in \left[0, \frac{1}{192}\right]$

4.3. Step II: stochastic updating with the Bhattacaryya distance metric

This section presents the stochastic updating procedure with the Bhattacharyya distance metric. There are eight epistemic parameters in the updating procedure, i.e., $\boldsymbol{\theta}_{x_i} = \{\mu_i, m_{2i}, m_{3i}, m_{4i}\}$, for $i = 1, 2$, whose feasible intervals are computed by Eq. (10) for each sub-interval, as shown in the last column of Table 3. Prior distributions of $\boldsymbol{\theta}_{x_1}$ and $\boldsymbol{\theta}_{x_2}$ are discretely obtained as multiple sets of their possible realizations within the feasible intervals, satisfying the moment constraints $\boldsymbol{\theta}_x \in \Theta$.

The width factor in the likelihood is set as $\varepsilon = 0.01$ in this section. In addition, the numbers of bins in the binning algorithm and in staircase density estimation are set to be $n_{bin} = 10$ and $n_b = 50$, respectively. After 16 TMCMC iterations, the finally updated posterior distributions of the epistemic parameters are estimated for each sub-interval of the parameters, as illustrated in Figs. 5 and 6. In the figures, the third and fourth central moments are normalized as $m_3/m_2^{3/2}$ and m_4/m_2^2 (those values are also known as the skewness and kurtosis, respectively). In addition, the target and updated values of the epistemic parameters are given in these figures. The target values are computed by samples generated from the target distributions in Fig. 2(b), and the updated values are obtained by estimating

361 most probable values (MPVs) of the posterior distributions. As shown in these figures, the posterior distributions of all of the epistemic parameters are significantly updated compared with the prior feasible intervals. Moreover, the updated values show almost good agreement with the target values, except for the kurtoses, implying the Bhattacharyya distance metric is capable to capture not only mean information but also dispersion and distribution information of both the model predictions and observations.

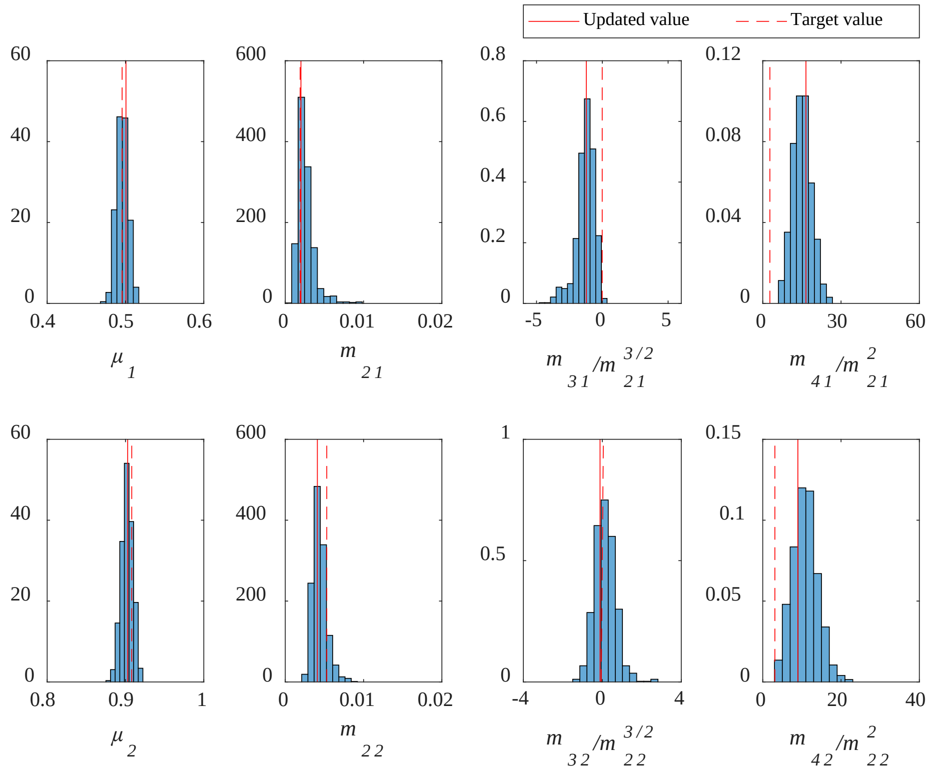

Fig. 5. Posterior distributions of epistemic parameters for sub-interval I.

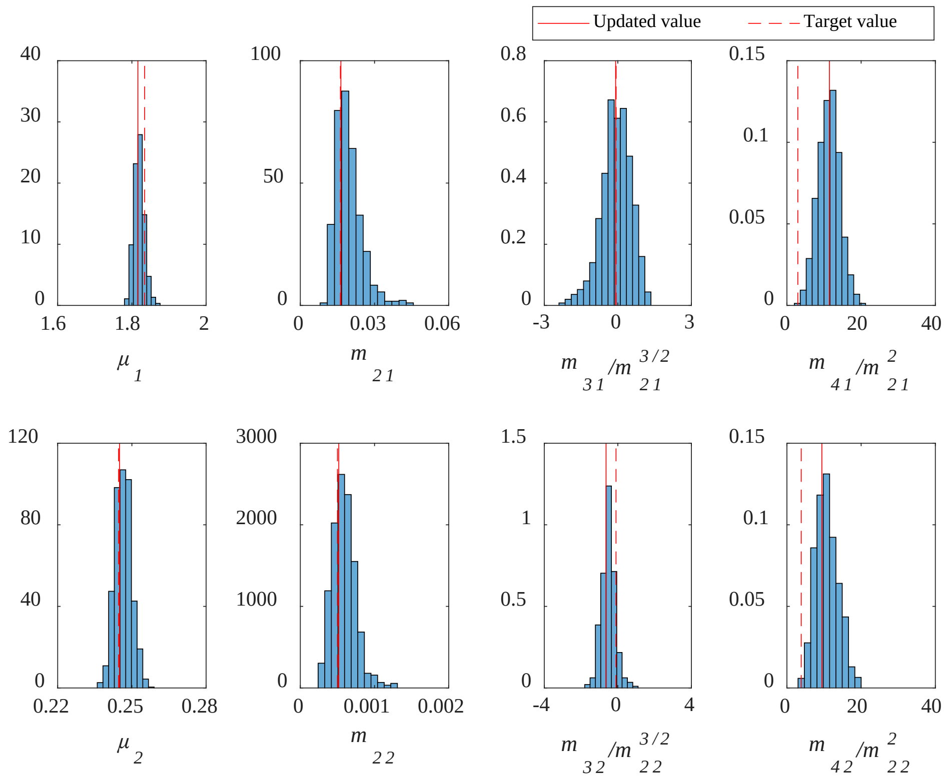

Fig. 6. Posterior distributions of epistemic parameters for sub-interval II.

Tables 4 and 5 present the target and updated values of the epistemic parameters. The updated values of the means and variances are quite close to the target and those of the skewnesses are also almost close to the target, even though those of the kurtoses still remain differences compared with the target. This fulfils the general experience in stochastic updating that the higher level of statistical information of is much more difficult to be precisely updated compared with the means. However, the proposed procedure enables to quantify even the higher level of statistical information, such as the variances and skewnesses.

Table 4. Updated epistemic parameters for sub-interval I.

Parameter	Mean μ	Variance m_2	Skewness $m_3/m_2^{3/2}$	Kurtosis m_4/m_2^2
x_1				
Target value	0.4957	0.0019	-0.0086	2.7313
Updated value	0.5006	0.0020	-1,2150	16,550
x_2				
Target value	0.9080	0.0053	-0.0531	3.0972
Updated value	0.9028	0.0041	-0,1150	8,9700

Table 5. Updated epistemic parameters for sub-interval II.

Parameter	Mean μ	Variance m_2	Skewness $m_3/m_2^{3/2}$	Kurtosis m_4/m_2^2
x_1				
Target value	1.8344	0.0162	0.0695	2.9410
Updated value	1.8162	0.0164	-0.0900	11.450
x_2				
Target value	0.2446	0.0005	-0.0984	3.8560
Updated value	0.2450	0.0005	-0.6400	9,4200

The updated samples of the parameters are generated from staircase densities with the updated values of the first four moments and illustrated in Fig. 7. It can be seen that the updated epistemic parameters for each sub-interval give the staircase densities providing accurate estimation of each mode of the parameters. More attention is paid to orientation and dispersion of the updated samples, which show good agreement with the target distributions, although some higher moment values are not precisely updated, as shown in Tables 4 and 6. It indicates that estimated errors in the higher moment values do not significantly affect to the sample distributions of each mode of the parameters obtained via staircase densities, compared with the means and variances. For comparison purpose, cumulative distribution functions (CDFs) of x_1 and x_2 are plotted in Fig. 8 for both the updated and target distributions. The CDFs for the updated distributions are obtained by combining the updated samples for both sub-intervals. The CDFs for the updated distributions exhibit a good match with those for the target distributions.

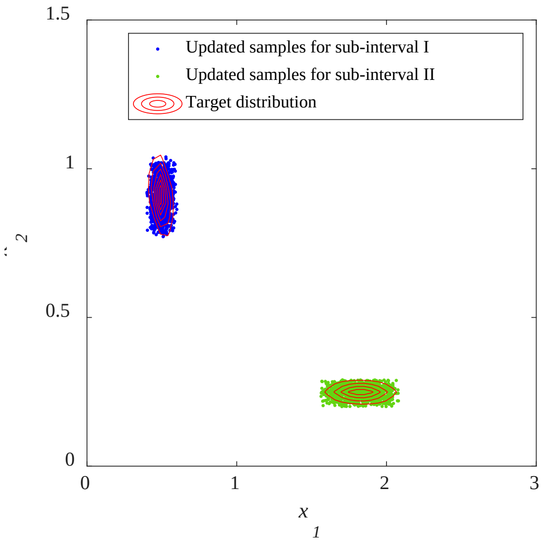

Fig. 7. Updated samples of parameters with Bhattacharyya distance.

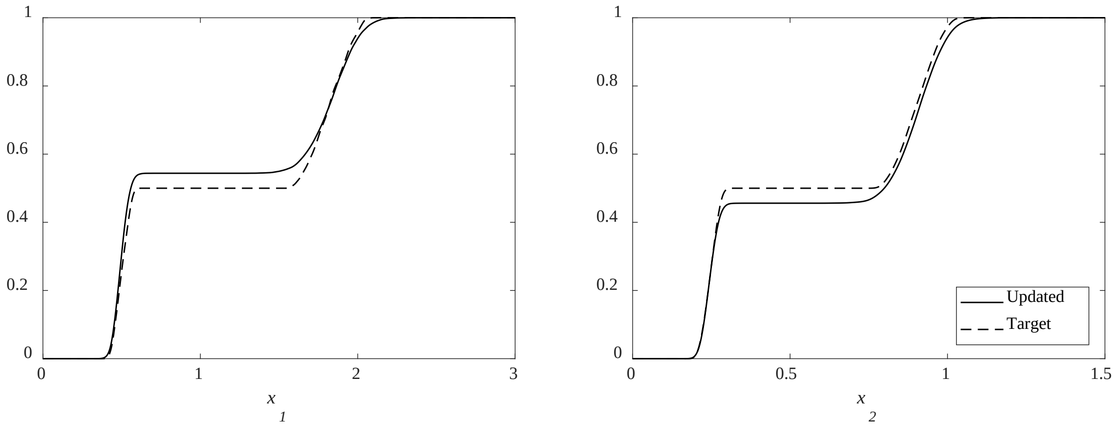

Fig. 8. Updated cumulative distribution functions of parameters.

Finally, Fig. 9 illustrates the final simulated output samples of f_1 and f_2, obtained by assigning the estimated staircase densities to x_1 and x_2, along with the initial simulated and target observed output samples. The updated simulated samples show a distribution identical to the target observed samples, implying the Bhattacharyya distance has the capability to recreate wholly the distribution of the target observations.

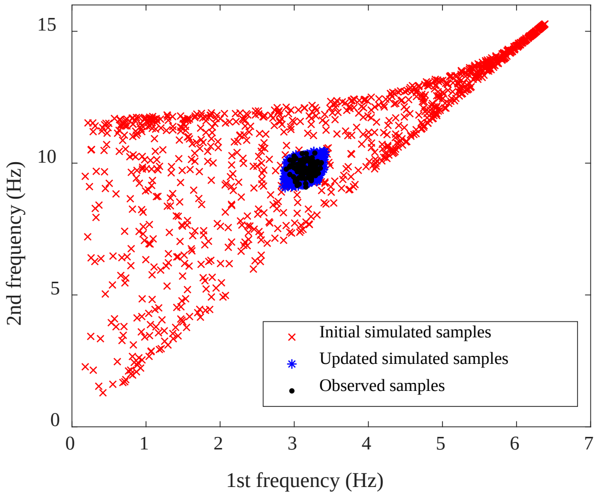

Fig. 9. Updated simulated output samples.

4.4. Summary

This example presented the combined application of the Euclidian and Bhattacharyya distances as the metrics in the two-step ABC updating procedure. Staircase random variables are employed in stochastic updating to calibrate the epistemic uncertainty space without prior information about the distribution format of parameters. The stochastic model updating procedure with the Bhattacharyya distance metric is demonstrated to be capable to estimate the detailed distributional properties of the parameters. However, a significant drawback of the updating procedure is revealed in situations, e.g., in Figs. 2(b) and 3, where the target distributions show multimodality but the observed features show unimodality. In this situation, the direct application of the stochastic model updating procedure may end with a local solution.

Consequently, the two-step procedure is proposed to overcome this drawback by performing deterministic updating with the Euclidian distance metric in the first step to identify every modes of the target distributions, and then to capture their detailed distributions in the second step. As a result, this example demonstrated that deterministic updating should be performed as a precondition of any stochastic updating procedure.

5. NASA UQ challenge problem 2014

5.1. Problem description

The NASA UQ challenge problem 2014 [17] is investigated herein to demonstrate the capabilities of the proposed updating framework for complex applications. The schematic in Fig. 10 illustrates the general structure of Sub-problem A, including the investigating parameters, outputs, and the UQ metric. As shown in Fig. 10, the simulator is provided in a black-box, which evaluates a scalar output y using five parameters: p_i, for $i = 1, \cdots, 5$.

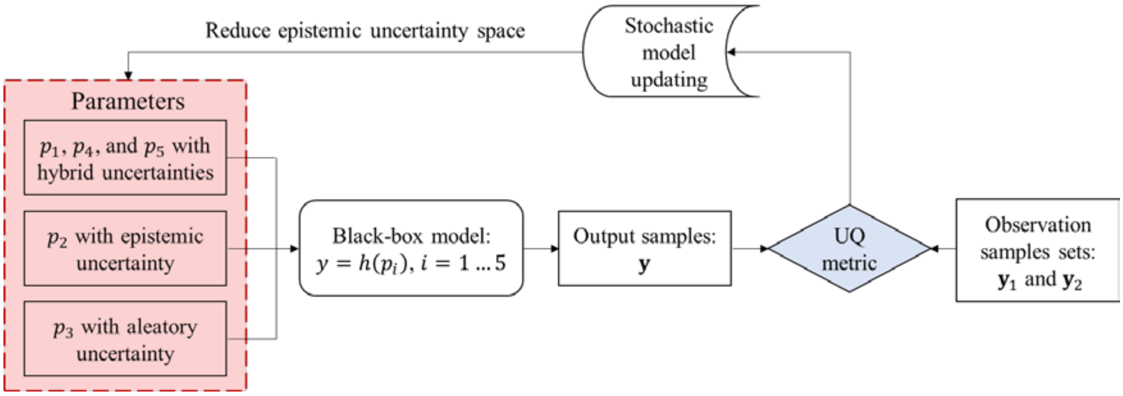

Fig. 10. Schematic of the NASA UQ challenge 2014 Sub-problem A.

Table 6 summarizes the detailed uncertainty characterization of the parameters in the original problem setting. p_1, p_4, and p_5 are parameters with hybrid uncertainties, p_2 is a parameter with only epistemic uncertainty, and p_3 is a parameter with only aleatory uncertainty represented by a fully prescribed uniform distribution with explicit mean and variance. During model updating, only p_1, p_2, p_4, and p_5 are considered and p_3 is omitted as it involves only irreducible aleatory uncertainty. More importantly, the distribution formats of p_1, p_4, and p_5 are fully provided, such that p_1 follows a unimodal beta distribution and p_4 and p_5 follow Gaussian distributions. Recently, true values of the epistemic parameters are released as shown in the last column of Table 6.

Table 6. Uncertain parameters of Sub-problem A in the NASA UQ challenge problem.

Parameter	Uncertainty characteristic	True value
p_1	Unimodal beta, $\mu_1 \in [0.6, 0.8]$, $m_{21} \in [0.02, 0.04]$	$\mu_1 = 0.6364$, $m_{21} = 0.0356$
p_2	Constant, $p_2 \in [0, 1.0]$	$p_2 = 1$
p_3	Uniform, $\mu_3 = 0.5$, $m_{23} = 1/12$	–
p_4, p_5	Gaussian, $\mu_i \in [-5.0, 5.0]$, $m_{2i} \in [0.0025, 4.0]$, $\rho \in [-1.0, 1.0]$[a], $i = 4,5$	$\mu_4 = 4$, $\mu_5 = -1.5$, $m_{24} = 0.04$, $m_{25} = 0.36$, $\rho = 0.5$

[a] ρ is the correlation coefficient.

On the other hand, in this study, the uncertainty characterization of p_1, p_4, and p_5 are redefined by ignoring distribution information to demonstrate the performance of the proposed framework. Table 7 presents the redefined uncertainty characteristics of these parameters. The support set of p_1 is based on the definition of the beta distribution. While only the support set is required to perform the updating framework, prior information about the mean μ_1 and variance m_{21} is remained to make the feasible intervals in Eq. (10) narrower. The support sets of p_4 and p_5 are set to cover more than 99.99 % confidence intervals of their true Gaussian distributions. The feasible intervals of the means and variances for these support sets are already narrower than the given intervals in Table 6. Note that, while these support sets can also be set to cover the p-box obtained from the original uncertainty characteristics, this leads too wide support sets to be precisely updated. Furthermore, the correlation between p_4 and p_5 is ignored, since the staircase random variable is a univariate random variable and thus cannot consider the correlation. However, this assumption is still reasonable because the target output is insensitive to the correlation coefficient ρ, which has been investigated by the previously published works [29-31] in other sub-tasks of the NASA UQ challenge problem 2014.

Table 7. Redefined uncertainty characteristics of p_1, p_4, and p_5.

Parameter	Uncertainty characteristic
p_1	$p_1 \in [0, 1]$, $\mu_1 \in [0.6, 0.8]$, $m_{21} \in [0.02, 0.04]$
p_4, p_5	$p_4 \in [3, 5]$, $p_5 \in [-4, 1]$

There are two observation sets $\mathbf{y}_1$ and $\mathbf{y}_2$, both containing 25 values respectively. Note that, in the original NASA UQ challenge problem, there were different tasks in Sub-problem A, where the first observations ($\mathbf{y}_1$) are supposed to be used for model updating in Task 1, and the remaining ($\mathbf{y}_2$) for model validation in Task 2. And in Task 3, all the 50 observations are used for model updating to improve the result. However, in this work, only Task 3 is addressed, because the comparison of the results using 25 or 50 observations is not our focus. As a consequence, totally 13 epistemic parameters, such as $\boldsymbol{\theta}_{p_i} = \{\mu_i, m_{2i}, m_{3i}, m_{4i}\}$, $i = 1,4,5$ and p_2 using the 50 observations.

5.2. Results assessment

The stochastic updating procedure with the Bhattacharyya distance metric is executed. Note that, the true distributions of all the parameters are unimodal; thus, an outer Bayesian updating with the Euclidian distance metric is not performed. The width factor in the likelihood is set as $\varepsilon = 0.01$ and the numbers of bins in the binning algorithm and in staircase density estimation are set to be $n_{bin} = 25$ and $n_b = 50$, respectively.

Fig. 11 illustrates the posterior histograms of the epistemic parameters. Those are converted to distributions with kernel density estimation (KDE) and the estimated distributions are also illustrated in this figure. It can be seen that all the epistemic parameters are successfully updated compared with their initial intervals.

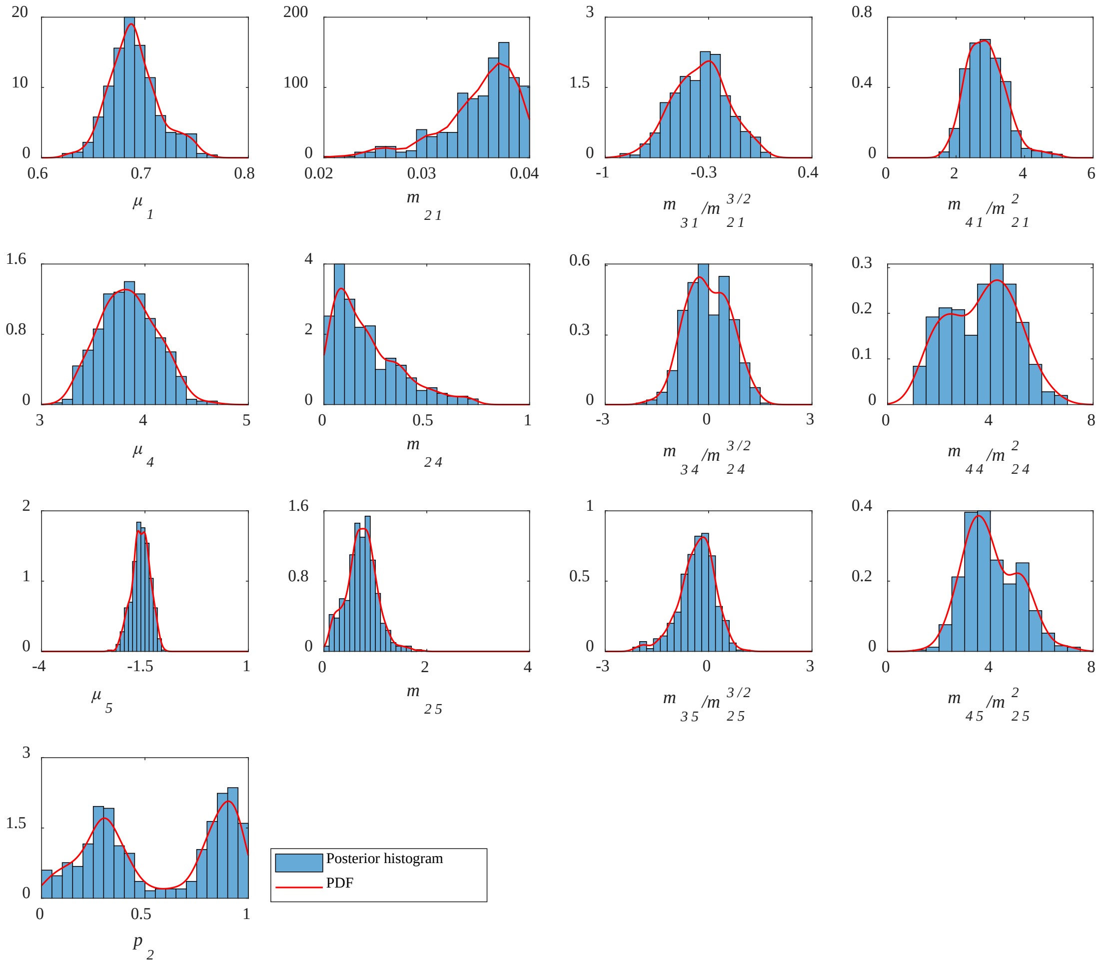

Fig. 11. Posterior histograms and PDFs estimated via KDE.

The updated values and intervals of $\boldsymbol{\theta}_{p_1}$, $\boldsymbol{\theta}_{p_4}$, $\boldsymbol{\theta}_{p_5}$, and p_2 are estimated from their posterior distributions, and their accuracy is assessed according to their true values as shown in Table 8. The posterior distributions are normalized so that their maximums are equal to one, as shown in Fig. 12. With this procedure, the posterior distributions are interpreted as Fuzzy sets [7], such that different levels of confidence will result to interval values of increased width. The crisp updated values of $\boldsymbol{\theta}_{p_1}$, $\boldsymbol{\theta}_{p_4}$, $\boldsymbol{\theta}_{p_5}$, and p_2 are obtained as the MPVs, i.e., the values corresponding to the alpha-level of one.

Table 8. The updated results of the epistemic parameters.

Epistemic parameter	Initial interval	True value	MPVs	0.9-level intervals
μ_1	$[0.6, 0.8]$	0.6364	0.6824	$[0.6793, 0.6933]$
m_{21}	$[0.02, 0.04]$	0.0356	0.0369	$[0.0366, 0.0386]$
$m_{31}/m_{21}^{3/2}$	$\left[-\frac{1}{6\sqrt{3}}, \frac{1}{6\sqrt{3}}\right]$	-0.3840	-0.3560	$[-0.3963, -0.2400]$
m_{41}/m_{21}^{2}	$\left[0, \frac{1}{12}\right]$	2.4886	2.4360	$[2.4007, 3.0787]$
μ_4	$[3, 5]$	4	3.8780	$[3.6517, 3.9517]$
m_{24}	$[0.0025, 1]$	0.04	0.0488	$[0.0480, 0.1236]$
$m_{34}/m_{24}^{3/2}$	$\left[-\frac{4}{3\sqrt{3}}, \frac{4}{3\sqrt{3}}\right]$	0.0068	-0.2620	$[-0.5241, 0.0499]$
m_{44}/m_{24}^{2}	$\left[0, \frac{1}{12}\right]$	2.9780	4.0560	$[3.7298, 4.7558]$
μ_5	$[-4, 1]$	-1.5	-1.7000	$[-1.7204, -1.4554]$
m_{25}	$[0.0025, 4]$	0.36	0.7920	$[0.6070, 0.8940]$
$m_{35}/m_{25}^{3/2}$	$\left[-\frac{4}{3\sqrt{3}}, \frac{4}{3\sqrt{3}}\right]$	0.0068	-0.0910	$[-0.4172, 0.0208]$
m_{45}/m_{25}^{2}	$\left[0, \frac{1}{12}\right]$	2.9780	3.6840	$[3.6830, 3.9185]$
p_2	$[0, 1]$	1	0.9050	$[0.8580, 0.9410]$

By employing the crisp updated values of $\boldsymbol{\theta}_{p_1}$, $\boldsymbol{\theta}_{p_4}$, and $\boldsymbol{\theta}_{p_5}$, the corresponding crisp updated distributions of p_1, p_4, and p_5 are estimated as staircase density functions. Fig. 13 plots the CDFs of the updated distributions, along with those of their true distributions. It can be seen that the updated CDFs show almost good agreement with their true distributions, while some differences still remain for p_4 and p_5. These differences are mainly due to no consideration of the correlation between p_4 and p_5. Nevertheless, the proposed updating procedure is demonstrated to be capable to estimate the probabilistic distributions of unknown parameters regardless of their distribution formats.

A more comprehensive assessment of the result is performed by estimating the p-boxes of the output. The initial intervals of $\boldsymbol{\theta}_{p_1}$, $\boldsymbol{\theta}_{p_4}$, $\boldsymbol{\theta}_{p_5}$, and p_2 result in a large p-box of the output, representing a large epistemic uncertainty space. The objective of model updating in this problem is to reduce the epistemic uncertainty space, so that the p-box of the output is accordingly reduced. In the ideal case, when the true values of the epistemic parameters are achieved form a perfect updating process, the resulting p-box of the output would be reduced to a single CDF, which perfectly coincides with the CDF of the observations. Based on the above motivation, three alpha-levels, namely 0.5, 0.75, and 0.9, are set for the normalized PDFs as shown in Fig. 12. The 0.9-level intervals are presented in the last column of Table 8, which are significantly reduced compared with the initial intervals. Along with the p-boxes, the updated crisp CDF with 1-alpha level is also estimated by employing the updated distributions of p_1, p_4, and p_5 as shown in Fig. 13 and the crisp updated value of p_2.

The p-boxes with the three alpha-levels and the updated CDF are illustrated in Fig. 14. The initial p-box with the original epistemic uncertainty space is significantly reduced through the updating procedure with different alpha-levels. An integrative comparison of Figs. 12 and 14 shows that the higher the alpha-level, the smaller the input epistemic intervals, and furthermore, the narrower the resulting p-box of the output, even though the differences in the three p-boxes are very small. The narrowest p-box with 0.9-alpha level can still envelop the target CDF of the observations. More importantly, the updated CDF shows good agreement with the CDF of the target observations. This outcome clearly demonstrates feasibility of the combination of the Bhattacharyya distance metric and staircase random variable in stochastic model updating and uncertainty characterization.

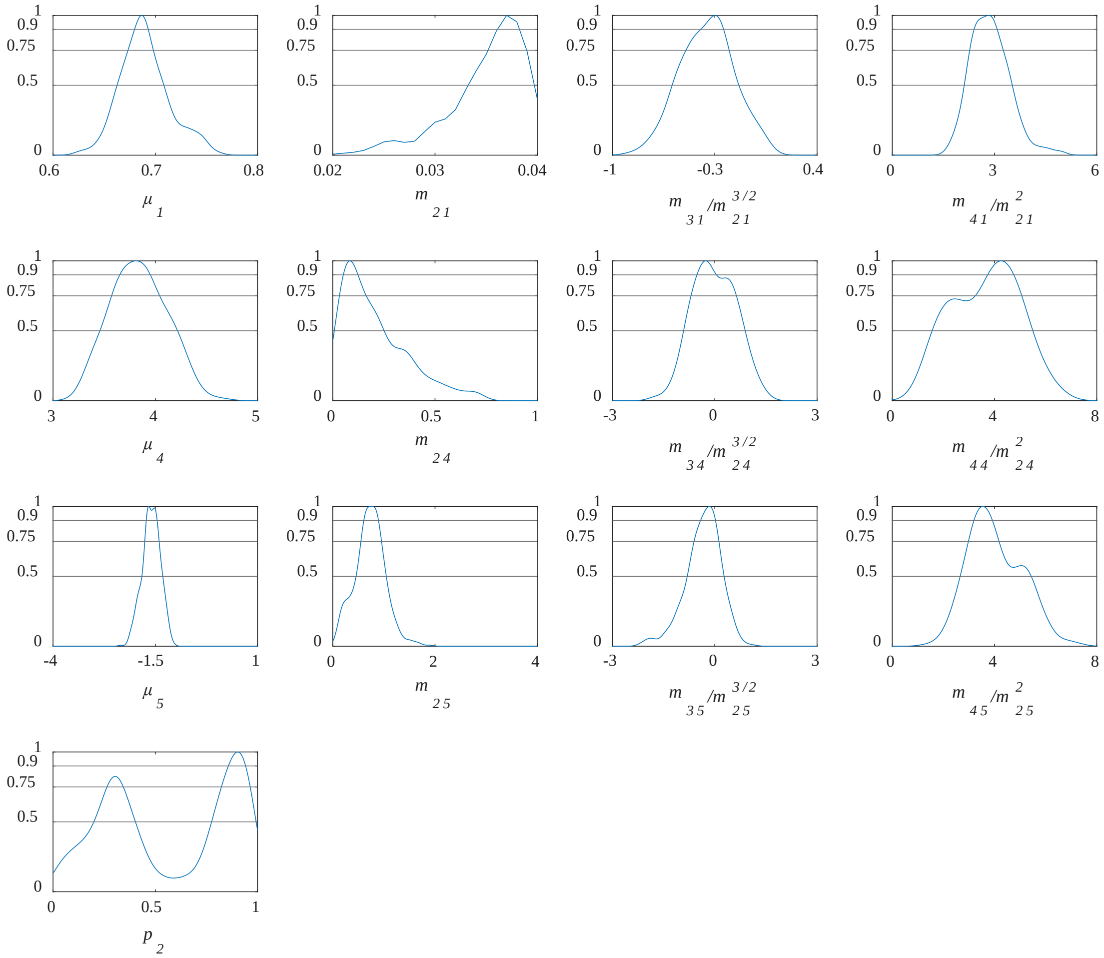

Fig. 12. Three truncation levels of the normalized posterior distributions.

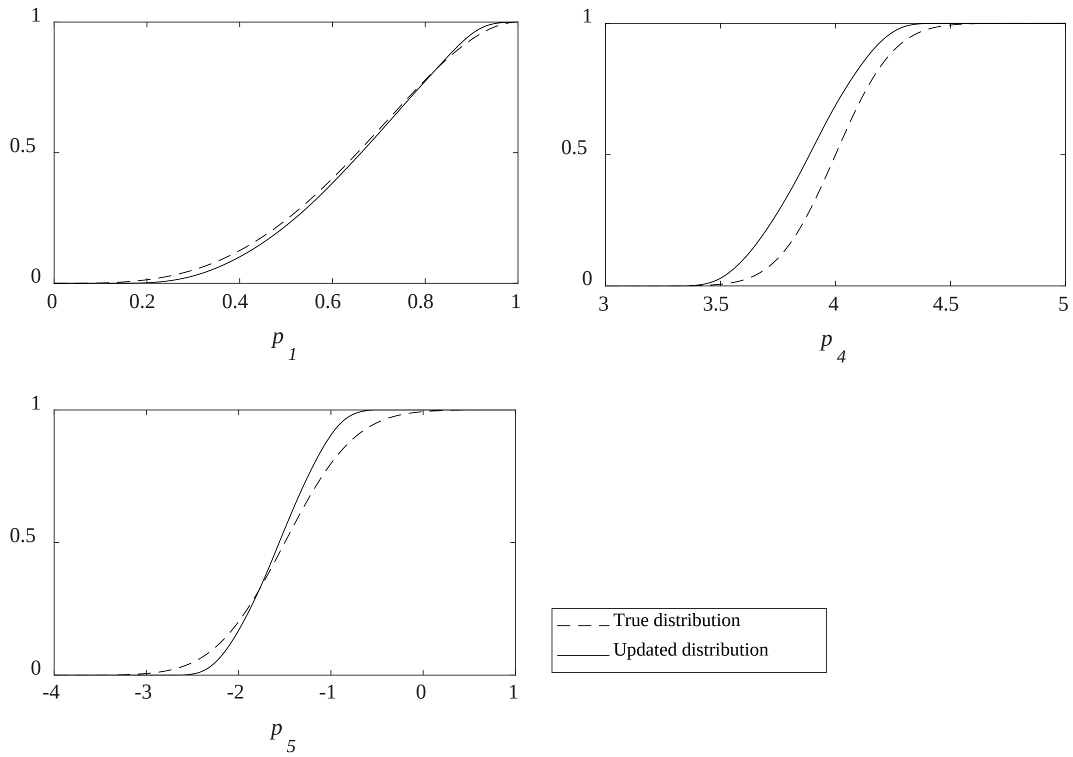

Fig. 13. Updated cumulative distribution functions of p_1, p_4, and p_5.

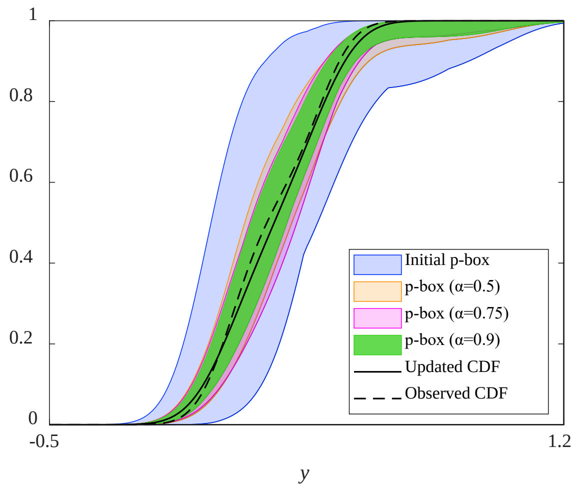

Fig. 14. Updated p-boxes with different alpha-levels.

6. Conclusions

The combined application of the Bhattacharyya distance metric and staircase random variable is demonstrated as key ingredients of the proposed ABC model updating framework. The application to the NASA UQ challenge problem reveals the feasibility of the Bhattacharyya distance and staircase random variable for stochastic updating and uncertainty characterization. The staircase densities act as a nonparametric connection between the epistemic uncertainty space and investigating outputs. In addition, the distance-based approximate likelihood function serves as a convenient connection between the Bayesian updating procedure and UQ metrics. By employing the Bhattacharyya distance, the proposed approximate likelihood enables to capture comprehensive uncertainty characteristics of both model predictions and observations. As a consequence, the proposed updating framework fulfills the challenging expectation in stochastic updating to calibrate the probabilistic distributions of parameters without prior knowledge about their distribution formats.

Despite the advantage on uncertainty characterization, the combination use of the Bhattacharyya distance metric and staircase random variable is not appropriate in an exclusive manner as revealed in the shear building model example. It needs to be complemented by the Euclidian distance metric in a two-step approach to avoid non-unique solutions in stochastic updating. That is to say, Euclidian distance-based deterministic updating should be the precondition before performing the stochastic updating procedure.

One of the perspectives of the proposed Bayesian updating framework is to be combined with structural health monitoring (SHM). In SHM, uncertainties not only in the input parameters but also in measurements need to be considered, because the measurements are driven under hard-to-control randomnesses, such as environmental noises and measurement system errors. Another challenging perspective focuses on uncertainties in modeling. The numerical model always contains unavoidable simplifications and approximations, such as the linearized representation of the nonlinear behaviors. These uncertainties make the updating procedure significantly difficult, and hence it should aim at robust calibration to achieve the maximum allowable uncertainty, while providing acceptable fidelity to the measurements. This extension of the proposed Bayesian model updating framework to cope with SHM will be addressed in the future work.

References

1. W. Oberkampf, J.C. Helton, C.A. Joslyn, S.F. Wojtkiewicz, S. Ferson, Challenge problems: uncertainty in system response given unertain parameters, Reliab. Eng. Syst. Safety. 85 (2004) 11-19.
2. C. Roy, W. Oberkampf, A comprehensive framework for verification, validation, and uncertainty quantification. Compt. Methods Appl. Mech. Eng. 200 (2011) 2131-2144.

3. R.E. Moore, R.B. Kearfott, M.J. Cloud, Introduction to interval analysis, SIMA Press, Philadelphia, 2009.
4. M. Faes, D. Moens, Recent trends in the modeling and quantification of non-probabilistic uncertainty, Arch. Comput. Methods Eng. 27 (2020) 633-671.
5. K. Sentz, S. Ferson, Combination of evidence in Dempster-Shafer theory, Tech. Rep. SAND2002-0835, Sandia National Laboratories, California, 2002.
6. S. Ferson, V. Kreinovich, L. Ginzburg, D.S. Myers, K. Sentz, Constructing probability boxes and Dempster-Shafer structures, Tech. Rep. SAND2002-4015, Sandia National Laboratories, California, 2003.
7. M. Beer, S. Ferson, V. Kreinovich, Imprecise probabilities in engineering analysis, Mech. Syst. Signal Process. 37 (2013) 4-29.
8. C. Mares, J.E. Mottershead, M.I. Friswell, Stochastic model updating: Part 1—theory and simulated example, Mech. Syst. Signal Process. 20 (2006) 1674-1695.
9. H.H. Khodaparast, J.E. Mottershead, M.I. Friswell, Perturbation methods for the estimation of parameter variability in stochastic model updating, Mech. Syst. Signal Process. 22 (2008) 1751-1773.
10. K.K. Sairajan, G.S. Aglietti, Robustness of system equivalent reduction expansion process on spacecraft structure model validation, AIAA J. 50 (2012) 2376-2388.
11. S. Bi, S. Prabhu, S. Cogan, S. Atamturktur, Uncertainty quantification metrics with varying statistical information in model calibration and validation, AIAA J. 55 (2017) 3570-3583.
12. B. Goller, M. Broggi, A. Calvi, G.I. Schuëller, A stochastic model updating technique for complex aerospace structures, Finite Elem. Anal. Des. 47 (2011) 739-752.
13. E. Patelli, Y. Govers, M. Broggi, H.M. Gomes, M. Link, J.E. Mottershead, Sensitivity or Bayesian model updating: a comparison of techniques using the DLR AIRMOD test data, Arch. Appl. Mech. 87 (2017) 905-925.
14. S. Bi, M. Broggi, M. Beer, The role of the Bhattacharyya distance in stochastic model updating, Mech. Syst. Signal Process. 117 (2019) 437-452.
15. M.A. Beaumont, W. Zhang, D.J. Balding, Approximate Bayesian computation in population genetics, Genetics 162 (2002) 2025-2035.
16. B.M. Turner, T. Van Zandt, A tutorial on approximate Bayesian computation, J. Math. Psychol. 56 (2012) 69-85.
17. L.G. Crespo, S.P. Kenny, D.P. Giesy, The NASA Langley multidisciplinary uncertainty quantification challenge, In: 16th AIAA Non-Deterministic Approaches Conf., 2014, 1-9.
18. L.G. Crespo, S.P. Kenny, D.P. Giesy, B.K. Stanford, Random variables with moment-matching staircase density function, Appl. Math. Model 64 (2018) 196-213.
19. G.R. Grimmett, D.R. Stirzaker, Probability and random processes, Oxford University Press, New York, 2001.
20. B.K. Patra, R. Launonen, V. Ollikainen, S. Nandi, A new similarity measure using Bhattacharyya coefficient for collaborative filtering in sparse data, Knowledge-Based Syst. 82 (2015) 163-177.
21. R. Sharma, S. Devi, G. Kapoor, N.S. Barnett, A brief note on some bounds connecting lower order moments for random variables defined on a finite interval, Int. J. Theor. Appl. Sci. 1 (2009) 83-85.

22. P. Kumar, Moment inequalities of a random variable defined over a finite interval, J. Inequ. Pure. Appl. Math. 3 (2002) 1-11.
23. J.L. Beck, L.S. Ktafygiotis, Updating models and their uncertainties. I: Bayesian statistical framework, J. Eng. Mech. 124 (1998) 455-461.
24. J.L. Beck, S.K. Au, Bayesian updating of structural models and reliability using Markov chain Monte Carlo simulation, J. Eng. Mech. 128 (2002) 380-391.
25. J. Ching, Y.C. Chen, Transitional Markov chain Monte Carlo method for Bayesian updating, model class selection, and model averaging, J. Eng. Mech. 133 (2007) 816-832.
26. W.K. Hasting, Monte Carlo sampling methods using Markov chains and their applications, Biometrica 57 (1970) 97-109.
27. W. Betz, I. Papaioannou, D. Straub, Transitional Markov chain Monte Carlo: Observations and improvements, J. Eng. Mech. 142 (2016).
28. R. Rocchetta, M. Broggi, Q. Huchet, E. Patelli, On-line Bayesian model updating for structural health monitoring, Mech. Syst. Signal Process. 103 (2018) 174-195.
29. E. Patelli, D.A. Alvarez, M. Broggi, M. de Angelis, Uncertanty management in multidisciplinary design of critical safety systems, J. Aerosp. Inf. Syst. 12 (2015) 140-169.
30. C. Safta, K. Sargsyan, H.N. Najm, K. Chowdhary, B. Debusschere, L.P. Swiler, M.S. Eldred, Probabilistic methods for sensitivity analysis and calibration in the NASA challenge problem, J. Aerosp. Inf. Syst. 12 (2015) 219-234.
31. R. Ghanem, I. Yadegaran, C. Thimmisetty, V. Keshavarzzadeh, S. Masri, J. Red-Horse, R. Moser, T. Oliver, P. Spanos, O.J. Aldraihem, Probabilistic approach to NASA Langley research center multidisciplinary uncertainty quantification challenge problem, J. Aerosp. Inf. Syst. 12 (2015) 170-188.