

The role of the Bhattacharyya distance in stochastic model updating

Sifeng Bi*, Matteo Broggi, and Michael Beer

Institute for Risk and Reliability, Leibniz Universität Hannover, Hannover 30167, Germany

Abstract: The Bhattacharyya distance is a stochastic measurement between two samples and taking into account their probability distributions. The objective of this work is to further generalize the application of the Bhattacharyya distance as a novel uncertainty quantification metric by developing an approximate Bayesian computation model updating framework, in which the Bhattacharyya distance is fully embedded. The Bhattacharyya distance between sample sets is evaluated via a binning algorithm. And then the approximate likelihood function built upon the concept of the distance is developed in a two-step Bayesian updating framework, where the Euclidian and Bhattacharyya distances are utilized in the first and second steps, respectively. The performance of the proposed procedure is demonstrated with two exemplary applications, a simulated mass-spring example and a quite challenging benchmark problem for uncertainty treatment. These examples demonstrate a gain in quality of the stochastic updating by utilizing the superior features of the Bhattacharyya distance, representing a convenient, efficient, and capable metric for stochastic model updating and uncertainty characterization.

Keywords: uncertainty quantification; stochastic model updating; model validation; Bayesian updating; approximate Bayesian computation

1 Introduction

It has been widely accepted that uncertainties should be appropriately considered in the campaign of model updating and validation. Uncertainty quantification (UQ) metrics are consequently significant to provide an elaborate measurement of the uncertainty information in stochastic model updating methodologies.

In the context of UQ, the system parameters can be categorized according to the involvement of aleatoric or/and epistemic uncertainties [1,2]:

- I) parameters without any epistemic uncertainty, appearing as either explicit constants or random variables with fully determined uncertainty characteristics such as distribution type, mean, variance;
- II) parameters with only epistemic uncertainty, modeled as constants but with unknown exact value bounded by a given interval;
- III) parameters with both aleatoric and epistemic uncertainties, modeled as random variables with only vaguely determined uncertainty characteristics.

* Corresponding author. Email address: sifeng.bi@irz.uni-hannover.de

As Category I parameters contain no epistemic uncertainties, they are investigated by neither deterministic nor stochastic model updating techniques. The deterministic updating methodologies generally investigate only Category II parameters aiming at a single set of crisp parameter values and at generating a single model prediction with maximum fidelity with regard to the observation. Comparatively, both Categories II and III parameters are considered by stochastic updating methodologies, whose target is not the single set of the parameters themselves, but a reduced space of the epistemic uncertainties, e.g. reduced intervals of Category II parameters and reduced bounds of the Cumulative Probability Function (CDF) of Category III parameters.

A wide range of stochastic model updating methodologies has been investigated in the literature, such as perturbation methods [3,4], interval updating [5,6] and fuzzy theory [7,8]. No matter which methodology is performed in stochastic model updating, it is significant to define a comprehensive UQ metric which is capable of quantifying the uncertain discrepancy between two random data samples, e.g. the observed data and the numerical model predictions.

As an alternative to the classical Euclidian distance, Khodaparast et al. [9] propose to use the Frobenius norm to measure the difference between the covariance matrices of the experimental and simulated data. By minimizing the objective function combining the Euclidian distance between the means and the Frobenius norm between the variances, this approach achieves satisfied updating result for both mean and variance with high efficiency. Another stochastic distance concept is the Bhattacharyya distance [10], which has been proposed as a potential UQ metric capable of capturing a higher level of statistical information from the investigating variables [11]. This metric is clearly more comprehensive for uncertainty treatment, however, its application in stochastic model updating is quite limited in the current literature. This limited application is not only caused by a relative high calculation cost but also, more critically, caused by its stochastic feature when integrating the product of the probability density functions (PDFs) of the random variables (which will be further discussed in Sec. 2.1). The objective of this work is consequently to generalize the Bhattacharyya distance as a universal metric for uncertainty treatment by developing a stochastic model updating approach, in which the Bhattacharyya distance is fully embedded and actively operated.

Bi et al. [11] performed a comprehensive comparison among the Euclidian, Mahalanobis, and Bhattacharyya distances as updating metrics via a direct Monte Carlo approach. However, this approach is essentially based on a single and non-directional searching technique with low efficiency. In the case of complex problems with high-dimensional parameter space, this approach would fail to search the global solutions. To solve this issue, in this paper, the Bayes' theorem [1] and transitional Markov chain Monte Carlo (TMCMC) algorithm [12] are employed to construct a Bayesian updating framework, in which the Bhattacharyya distance is embedded as the metric. In this framework, the approximate Bayesian computation (ABC) [13] is proposed to develop an approximate but efficient likelihood function constructed on the distance between model predictions and experimental observations. The distance-based ABC is of central importance in this framework since it acts as the connection between the Bhattacharyya distance metric and the Bayesian updating tool.

The present development is particularly motivated by the NASA UQ challenge problem [2], since it has revealed directions for improvement of current model updating technologies when they meet real-size practical problems. Herein,

we demonstrate the advantages of the Bhattacharyya distance metric in association with the proposed ABC updating approach, specifically on the NASA UQ challenge. We focus on solving Sub-problem A (uncertainty characterization) out of the overall challenge problem. Sub-problem A is closely related to model updating, and its results can be significantly influenced by different UQ metrics. Three of the works previously published on solving the NASA UQ challenge problem, namely the papers by Patelli et al. [14], Ghanem et al. [15], and Safta et al. [16], are used as reference works to assess the updating results. All of these three works are also employing Bayesian updating methodologies. However, since the employed UQ metrics and the associated individual updating strategies vary, the results exhibit considerable discrepancies. This underlines the significance of using a powerful UQ metric. The proposed approach using the Euclidian and the Bhattacharyya distances in sequence as metrics in a two-step updating procedure shows clear advantages in this context.

In Sec. 2 we describe the theoretical and methodological basis of the Bhattacharyya distance evaluation and the Bayesian model updating. Sec. 3 outlines the novel developments of the distance-based ABC likelihood function, and the proposed two-step ABC updating framework. The principle and illustrative applications are detailed in Sec. 4, using a simple spring-mass example for illustration, and in Sec. 5, concentrating on to the demonstration of the performance of the framework on the highly challenging NASA UQ problem. Conclusions are drawn in Sec. 6.

2 Theories and Methods

2.1 Formulations of the Bhattacharyya Distance in UQ

In the context of stochastic model updating with UQ, the investigating system is characterized using three components: input parameters $\boldsymbol{\theta}$, output features \mathbf{x} , and simulator $h(\cdot)$:

$$\mathbf{x} = h(\boldsymbol{\theta}) \quad (1)$$

where $\mathbf{x} = [x_1, x_2, \dots, x_m]$; $\boldsymbol{\theta} = [\theta_1, \theta_2, \dots, \theta_n]$; m and n are respectively the number of outputs and inputs. The simulator is usually presented as either a sophisticated code package (e.g. finite element model) or a simplified function (e.g. response surface model).

The uncertainties of the system are first characterized by uncertain parameters in various categories (refer Sec. 1), and then, propagated through the simulator into uncertain outputs or features presenting various forms of uncertainty as well, e.g. probabilistic distributions, intervals, and fuzzy sets. In general, independently of the form of the uncertainties, randomly sampled values of parameters and features are used in UQ algorithms. Suppose the required size of the samples is N_{sim} , the simulator h is executed N_{sim} times to generate the simulated feature sample $\mathbf{X}_{sim} \in \mathbb{R}^{N_{sim} \times m}$:

$$\begin{aligned} \mathbf{X}_{sim} &= [\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_m]; \\ \mathbf{x}_i &= [x_{1i}, x_{2i}, \dots, x_{N_{sim}i}]^T, \quad \forall i = 1, \dots, m. \end{aligned} \quad (2)$$

In addition to the simulated features, observed features are required as the target of model updating. Suppose the observation size 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: $\mathbf{X}_{obs} \in \mathbb{R}^{N_{obs} \times m}$. The objective of stochastic model updating can be expressed as to minimize the discrepancy between \mathbf{X}_{obs} and \mathbf{X}_{sim} , considering not only their means but also their dispersion information such as variances and covariance, by updating the uncertainty characteristics of parameters.

In the following, more attention is given to the possible metrics capable to capture the discrepancy between \mathbf{X}_{obs} and \mathbf{X}_{sim} . The very classical metric based on the Euclidian distance concept is expressed as:

$$d_E(\mathbf{X}_{obs}, \mathbf{X}_{sim}) = \sqrt{(\bar{\mathbf{X}}_{obs} - \bar{\mathbf{X}}_{sim})(\bar{\mathbf{X}}_{obs} - \bar{\mathbf{X}}_{sim})^T} \quad (3)$$

where $\bar{\mathbf{X}}_{\blacksquare}$ is the row vector of the means of the features. The Euclidian distance is a point-to-point distance more generally used in the deterministic updating methodologies where a single-simulation-single-observation procedure is utilized. In the presence of random samples, it is more desirable to employ a more comprehensive metric capable to account for not only the effect of the mean but also the variance, covariance, and even the difference in distribution shape.

The Bhattacharyya distance is thus proposed herein as a stochastic metric measuring the degree of overlap between distributions of two random variables. The Bhattacharyya distance is defined as

$$d_B(\mathbf{X}_{obs}, \mathbf{X}_{sim}) = -\log \left[\int_{\mathbf{x}} \sqrt{p_{obs}(\mathbf{x})p_{sim}(\mathbf{x})} d\mathbf{x} \right] \quad (4)$$

where $p_{\blacksquare}(\mathbf{x})$ is the PDF of the sample; \mathbf{x} is the m -dimensional space, implying $\int_{\mathbf{x}} \blacksquare d\mathbf{x}$ is the integration performed over the whole feature space. Note that, for most systems, the number of features is $m > 1$, thus the integration is performed on multi-dimensional joint PDFs. This operation can be quite time consuming and complex especially in high-dimensional feature space and requires a highly detailed knowledge of the distributional characteristics, achievable only through a high number of observations. As a matter of fact, one of the classical methods to estimate PDF from a finite data sample is the kernel density estimation (KDE) [17]. However, a converged estimation in KDE is usually unavailable because of the very limited number of observations, especially for applications where experiments are difficult or expensive. In such situations, the so-called binning algorithm is proposed to evaluate the probability mass function (PMF) of a discrete distribution, so that the discrete Bhattacharyya distance is used instead. The PMF is a function which maps the possible values of a discrete random variable to probabilities of their occurrence [18]. The Bhattacharyya distance between two discrete distributions using two discrete PMFs is evaluated as [19]

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

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

In order to calculate the PMF, the binning algorithm consists of the following steps. An example of a two-dimensional case following this binning algorithm is illustrated in Fig. 1.

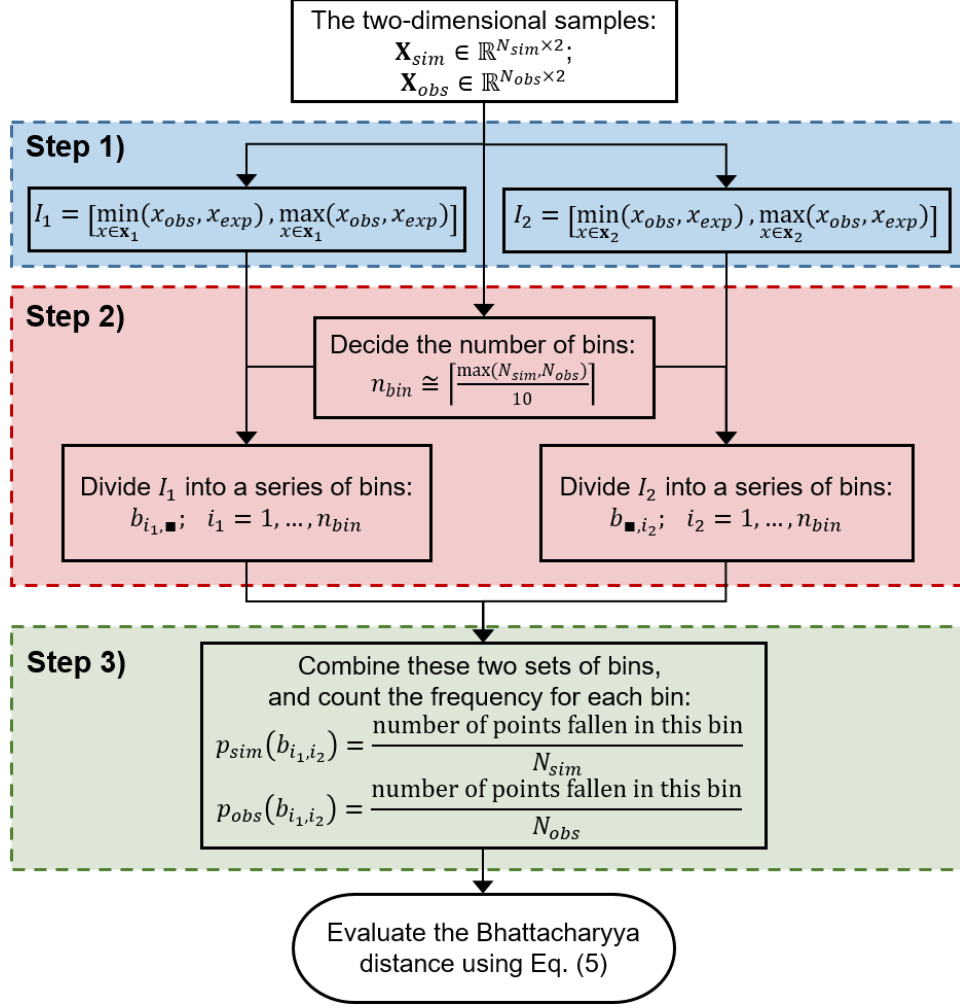


Fig. 1: Schematic of the binning algorithm for a two-dimensional case

- 1) Define the common interval I_i of both \mathbf{X}_{obs} and \mathbf{X}_{sim} according to the i^{th} feature $\mathbf{x}_i, \forall i = 1, \dots, m$, by finding the general maximum and minimum of \mathbf{x}_i in both \mathbf{X}_{obs} and \mathbf{X}_{sim} ;
- 2) Within the defined interval, decide the number of bins $n_{bin} \cong \left\lceil \frac{\max(N_{sim}, N_{obs})}{10} \right\rceil$, where $\lceil \blacksquare \rceil$ denotes the upper integer of the investigating values;
- 3) Count the joint probability mass for each bin $p_{\blacksquare}(b_{i_1, i_2, \dots, i_m})$. Note that the total number of bins in the joint-PMF space is $N_{bin} = n_{bin}^m$.

The principle of the n_{bin} in Step 2) is that a smaller n_{bin} leads to a smaller value of the Bhattacharyya distance. In the extreme case when $n_{bin} = 1$, Eq. (5) would always return the value $d_B(\mathbf{X}_{obs}, \mathbf{X}_{sim}) = 0$ for arbitrary \mathbf{X}_{obs} and \mathbf{X}_{sim} . This principle requires that N_{sim} or N_{obs} should be at least larger than 10, however, this is a practical requirement for N_{sim} in most applications.

Regarding the N_{obs} , it is a common obstacle for any stochastic updating approach that the number of observations can be quite limited. To overcome this issue, Ref. [11] proposes is a random sampling method to obtain the so-called *semi-*

experimental sample. Suppose an observation sample with quite reduced size is available in practical applications, the first step of this random sampling method is to estimate the mean and variance of this small sample. The second step is to assume a suitable distribution of the observation population. Based on the Law of Large Number and general engineering experience, the most classical assumption could be the Gaussian distribution. Some other distribution types can also be proposed according to the property of the application. As long as the mean, variance, and distribution type are determined, the random sampling can be performed to generate a large sample, which still involves the distributional property of the original small observation sample, and also with sufficient data points in this semi-experimental sample.

The Bhattacharyya distance is applicable to any sample set regardless of its distribution function. This makes it especially appropriate for stochastic model updating where the distributions of the features cannot be exactly determined. Another advantage of the metric is that it provides a scalar measurement for all the features simultaneously, which fulfill the expectation as an elaborate, quantifiable, and uniform measurement of multiple uncertain features. However, the Bhattacharyya distance's quantification of the distance in terms of the means or center of mass of the sets is not as direct as with the Euclidian distance. As a matter of fact, as long as two sample sets have no overlap, the Bhattacharyya distance return a value that is infinite, and is therefore insensitive to the relative position of the center of mass of these two samples. This drawback of the Bhattacharyya distance will be further discussed in Sec. 3.

2.2 Bayesian Model Updating

The stochastic nature of the Bhattacharyya distance makes it difficult to be implemented by the deterministic updating tools, e.g. the derivative-based optimizations [11]. Therefore, the well-known Bayesian updating framework along with the ABC method is proposed, employing the Bhattacharyya distance as an updating metric. The foundation of Bayesian updating is

$$P(\boldsymbol{\theta}|\mathbf{X}_{obs}) = \frac{P_L(\mathbf{X}_{obs}|\boldsymbol{\theta})P(\boldsymbol{\theta})}{P(\mathbf{X}_{obs})} \quad (6)$$

with the key elements in Bayes' theorem:

- $P(\boldsymbol{\theta})$ is the prior distribution representing the initial knowledge about the parameters $\boldsymbol{\theta}$;
- $P(\boldsymbol{\theta}|\mathbf{X}_{obs})$ is the posterior distribution representing the updated knowledge based on the observation data;
- $P(\mathbf{X}_{obs})$ is the normalization factor ensuring the posterior distribution integrates to 1;
- $P_L(\mathbf{X}_{obs}|\boldsymbol{\theta})$ is the likelihood function of \mathbf{X}_{obs} for an instance of the parameters $\boldsymbol{\theta}$.

One challenging component of the Bayesian updating framework is the normalization factor $P(\mathbf{X}_{obs})$, also known as evidence, as the direct integration of the posterior PDF over the whole parameter space is quite difficult or even intractable for very peaked or multi-modal distributions [20]. As a well-known simulation algorithm, the Transitional Markov Chain Monte Carlo (TMCMC) [12] along with the Metropolis-Hasting algorithm [21] are employed as an effective updating tool. This algorithm is essentially an iterative approach sampling from a series of intermediary PDFs which progressively converge to the posterior distribution. The j^{th} intermediary PDF is expressed as

$$P_j \propto P_L(\mathbf{X}_{obs}|\boldsymbol{\theta})^{\beta_j} P(\boldsymbol{\theta}) \quad (7)$$

where the exponent of the likelihood β_j is the so-called reduction coefficient. Its value starts from $\beta_0 = 0$ in the first iteration and progressively increases until reaches $\beta_m = 1$ in the last step. β_j is adaptively computed from the samples of the previous step. Markov chains with the Metropolis-Hasting algorithm propagates new samples starting from these with higher intermediary likelihood values, allowing for sampling from very complex posterior distributions. The readers are referred to Refs. [12,22] for details of TMCMC algorithm, and more applications of this algorithm can be found in fields from stochastic model updating [16,23] to structural health monitoring [24].

3 Approximate Bayesian Computation with Stochastic Distance Metrics

The likelihood is the key component in a Bayesian updating framework, since it quantifies the degree of relevance of a model with a given instance of parameters, by representing the possibility of the observations. Under the assumption of independence between observations, the likelihood in Eq. (6) is theoretically defined as

$$P_L(\mathbf{X}_{obs}|\boldsymbol{\theta}) = \prod_{k=1}^{N_{obs}} P(x_k|\boldsymbol{\theta}) \quad (8)$$

where $P(x_k|\boldsymbol{\theta})$ is the probability density value at x_k , for a given instance of the parameters $\boldsymbol{\theta}$. Note that, the PDF should be estimated respectively for every $\boldsymbol{\theta}$ instance. However, the total number of $\boldsymbol{\theta}$ instances is generally large throughout a complete Bayesian updating procedure, see e.g. Refs. [14,16]. In addition, even in a single PDF estimation, the classical method such as KDE requires also a large number of simulated features. Consequently, an analytical formula of the likelihood as in Eq. (8) demands a huge number of model evaluations, and for complex simulators with high-dimensional parameter/feature, it can be almost impossible to evaluate.

The Approximate Bayesian Computation (ABC) method [25,26] is utilized to overcome the above obstacle by replacing the full likelihood with an approximate and efficient function containing the information of the observations and the instance of $\boldsymbol{\theta}$. In this context, it is natural to construct the approximate likelihood employing the distance metrics. Various functional formulas have been investigated in the literature for the ABC method, such as the Gaussian [24], sharp [16], and Epanechnikov [27] functions. Nevertheless, the basic principle of distance-based ABC development is that the proposed likelihood formulae should return a high value when the distance metric is small, while the formulae penalizes the $\boldsymbol{\theta}$ instance when its corresponding distance metric is large. In this work, the approximate likelihood based on the Gaussian function and is proposed as

$$P_L(\mathbf{X}_{obs}|\boldsymbol{\theta}) \propto \exp\left\{-\frac{d^2}{\varepsilon^2}\right\} \quad (9)$$

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 of the parameter. Based on a series of tests in various applications, ε is determined to lie in the interval $[10^{-3}, 10^{-1}]$ [14]. 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. When ε is much smaller than

the identified lower bound, integration and sampling from such very peaked posterior distribution would be very computationally expensive, or even intractable.

The distance-based ABC likelihood proposed in Eq. (9) serves as a convenient connection between the novel distance metrics and the Bayesian updating procedure with significantly reduced calculation cost, while the updating precision is not negatively affected as it will be demonstrated in the example sections. More importantly, the distance-based ABC formula provides a uniform framework for either deterministic or stochastic updating, simply driven by the employed metric is Euclidian or Bhattacharyya distance.

The overall two-step ABC updating framework is illustrated in Fig. 2. It is specially designed to cope with the drawback of the Bhattacharyya distance as mentioned in Sec. 2.1. Before performing the ABC updating, the initial distance between \mathbf{X}_{obs} and \mathbf{X}_{sim} is assessed. If the initial \mathbf{X}_{sim} is too far from \mathbf{X}_{obs} , the Bhattacharyya distance evaluated in Eq. (4) will be infinite, and thus cannot be directly utilized for the distance-based likelihood. In this case, a preliminary step is employed to ensure the feasibility of the overall framework, i.e. to force an overlap between \mathbf{X}_{obs} and \mathbf{X}_{sim} . Step I is equivalent to a deterministic updating procedure with the target to update only the mean of the parameter, while the comprehensive uncertainty characteristics of the parameter are further updated in Step II via the stochastic updating with Bhattacharyya distance-based likelihood.

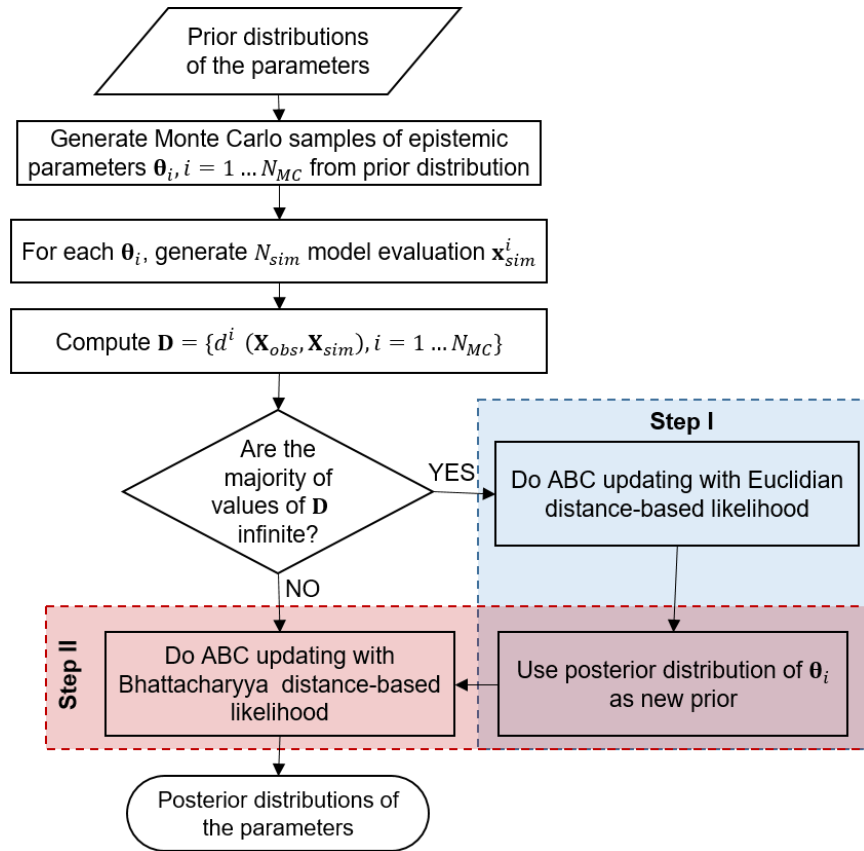


Fig. 2: Schematic of the two-step ABC updating procedure

4 Principle and illustrative application of the ABC updating procedure

4.1 Problem description

The two-step strategy of the proposed ABC updating with both the Euclidian and Bhattacharyya distances as metrics is demonstrated on a three Degree of Freedom (DOF) spring-mass system shown in Fig. 3. This numerical system has been utilized in Ref. [3,4,28] for various model updating approaches, however, its parameters and uncertainty characteristics are here altered to demonstrate the capabilities of the proposed approach. The stiffness coefficients k_1 , k_2 and k_3 are the uncertain parameters to be investigated, while the remaining parameters (i.e. k_4 to k_6 and the three masses m_1 to m_3) are fixed constants: $k_{4-6}=5.0\text{N/m}$, $m_1=0.7\text{kg}$, $m_2=0.5\text{kg}$, and $m_3=0.3\text{kg}$. The three natural frequencies f_1 , f_2 and f_3 are taken as the investigating outputs whose uncertainty are driven by the uncertain parameters k_1 - k_3 . Both aleatoric and epistemic uncertainties are involved in this system and are included by modeling k_1 - k_3 as independent Gaussian random variables, where the mean and standard deviation are not fixed but unknowns lying within given intervals. According to the parameter categorization in Sec. 1, k_4 - k_6 and m_1 - m_3 belong to Category I while k_1 - k_3 are Category III parameters. The intervals of the means μ and standard deviations σ associated to k_1 - k_3 are detailed in Table 1.

The target of the updating procedure \mathbf{X}_{obs} is a set of features of the frequencies f_1 - f_3 obtained by assigning target values to the parameter means and variances, as shown in Table 1. The size of the observation sample is $N_{obs} = 100$, generated by evaluating the model 100 times with parameters sampled from their assigned Gaussian distributions with the target means and variances.

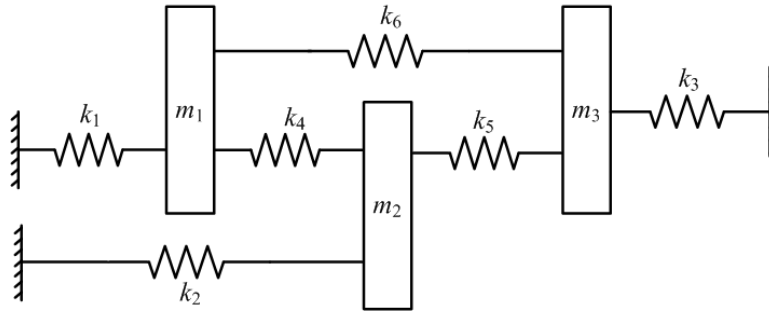


Fig. 3: The three degree-of-freedom spring mass system

Table 1: Parameters and epistemic inputs of the uncertain mass-spring system

Category	Parameter	Uncertainty characteristic	Target value of epistemic input
III	k_1	Gaussian, $\mu_1 \in [3.0, 7.0], \sigma_1 \in [0.0, 0.5]$	$\mu_1 = 4.0, \sigma_1 = 0.3$
	k_2	Gaussian, $\mu_2 \in [3.0, 7.0], \sigma_2 \in [0.0, 0.5]$	$\mu_2 = 5.0, \sigma_2 = 0.1$
	k_3	Gaussian, $\mu_3 \in [3.0, 7.0], \sigma_3 \in [0.0, 0.5]$	$\mu_3 = 6.0, \sigma_3 = 0.2$
I	k_4 - k_6, m_1 - m_3	Deterministic, no updating required	--

In addition to the target values in Table 1, a single set of initial values is set to be a possible realization of means and standard deviations different from the target values (but also within the prior intervals), in order to show how the

imprecise model can produce outputs far from the observation sample. The initial values are presented herein only for demonstration purpose as illustrated in Figs. 4 and 6. The Bayesian updating is not really started from the initial values, but from the prior distributions of the epistemic inputs, as shown in the 3rd column of Table 1. The corresponding initial and observation output samples of f_1 and f_2 are illustrated in Fig. 4. In the following figures, an estimated ellipse is provided for each scatter to assist the comparison of the distributional properties among the samples. This ellipse is obtained as a contour of a fitted distribution based on the scatter, using the Gaussian Mixture Modelling (GMM) algorithm [29]. In the following context, the height level of this contour is defined as 0.05 times of the maximal value of the fitted PDF. Here the height level ratio (0.05) falls within the interval [0, 1]. Clearly, a larger ratio (i.e. 0.1) leads to a smaller contour of the PDF. The determination of this ratio is flexible for different applications, as long as a clear differentiation among the samples is achieved. A similar usage of the ellipse and scatter can be found in Refs. [4,28].

As shown in Fig. 4, the objective of model updating herein is clearly no longer a single updated point with maximum fidelity to a single observation point, but the updated means and standard deviations of the parameter distributions which can represent an output sample as similar as the observational one. To achieve this objective, both the Euclidian and Bhattacharyya distances are employed as metrics in the ABC updating procedure.

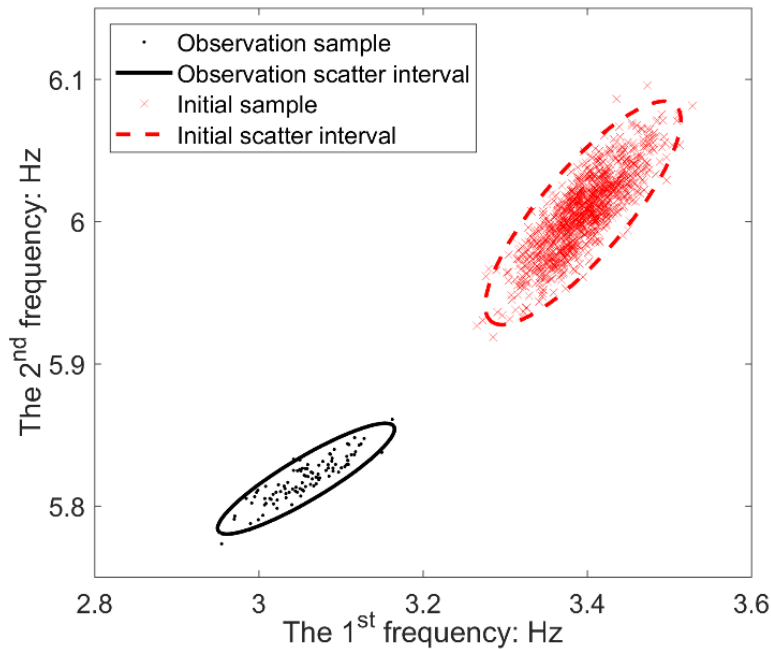


Fig. 4: The observation sample and initial sample in the plane of the 1st and 2nd frequencies

4.2 Step I: Updating with the Euclidian Distance Metric

There are six epistemic inputs in the updating procedure, i.e. $\{\mu_i, \sigma_i\}, i = 1, \dots, 3$, whose prior distributions are set to be uniform based on the intervals in Table 1. When the Euclidian distance is taken as the metric, the geometric distance between the centre of mass of the samples is measured, while the dispersion and distribution information of the samples cannot be considered. Hence the results of Step I conform the expectation in Sec. 3 that only the parameter means is

updated. As shown in Fig. 5, the posterior distributions of μ_1 - μ_3 accurately converge to the target values, while the posterior distributions of σ_1 - σ_3 does not exhibit distinct changes compared with their prior uniform distributions.

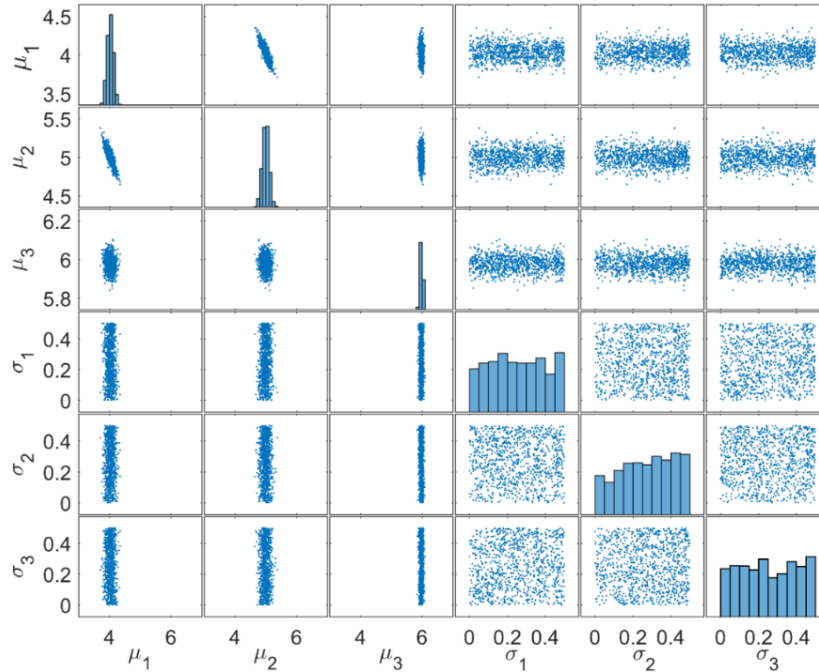


Fig. 5: The posterior distributions of the six inputs after updating with Euclidian distance

Table 2: Updated inputs using both Euclidian and Bhattacharyya distances as metrics

Input	Target value	Updated value	
		with Euclidian metric	with Bhattacharyya metric
μ_1	4.0	4.0314	4.0386
μ_2	5.0	5.0124	5.0102
μ_3	6.0	6.0262	6.0253
σ_1	0.3	--	0.3067
σ_2	0.1	--	0.0937
σ_3	0.2	--	0.1773

Table 2 presents the updated values of μ_1 - μ_3 which are obtained by estimating the means of the posterior distributions. As the distributions of σ_1 - σ_3 are not clearly changed, the first updating step with the Euclidian distance metric is proved to be incapable of updating the variance. In this section, the width coefficient in ABC is set as $\varepsilon = 0.01$, and totally eight TMCMC iterations are executed to reach convergence. The initial, second (i.e. Sample iter.#2), and final (i.e. Sample iter.#8) samples are presented in Fig. 6. As explained in Sec. 4.1, the initial sample is generated based on a single set of initial value of the epistemic inputs to show how the imprecise model can produce outputs far from the observation sample. However, the intermediate samples (e.g. Sample iter.#2) are no longer randomly generated, but obtained from a certain set of epistemic inputs, which are estimated as the mathematical expectations of the intermediate-posterior distributions after each iteration.

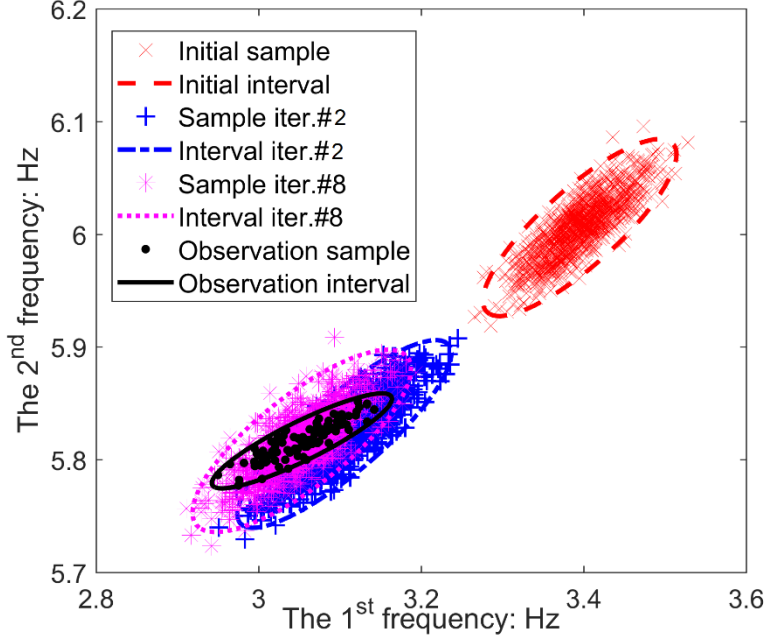


Fig. 6: The updated samples with Euclidian distance

As shown in Fig. 6, the effect of this updating step is to progressively shift the initial sample toward the observation sample. The centre point of the final sample coincides well with the centre of mass of the observations, thus the Euclidian distance metric has been minimized, corresponding to the maximization of the likelihood. Notice, however, as the parameter variances remain unchanged, the orientation and dispersion of Sample iter.#8 remain different from the observations and unchanged from the initial samples. Therefore, a more comprehensive metric is required in the second updating step to further reduced the discrepancy between Sample iter.#8 and the observation sample.

4.3 Step II: Updating with the Bhattacharyya Distance Metric

This section presents the second updating step where the Bhattacharyya distance is employed. The posterior distributions obtained in the first step are taken as prior distributions in this step. After 11 TMCMC iterations, the finally updated posterior distributions of the six inputs are obtained, as illustrated in Fig. 7. The mean values of these distributions are listed in the last column of Table 2. Considering the means $\mu_1-\mu_3$, the updated values with the Bhattacharyya distance metric are similar to the values with the Euclidian metric. However, the posterior distributions of $\mu_1-\mu_3$ in Fig. 7 are much more peaked than the ones in Fig. 5. This is caused by the treatment to introduce the outcome of the first step as the input of the second step, such that the distributions of $\mu_1-\mu_3$ are further updated to be more centralized to the target values.

More attention is paid to the results of $\sigma_1-\sigma_3$, whose posterior distributions in Fig. 7 are significantly updated compared with their prior uniform distributions. The estimated means of these posterior distributions are also listed in Table 2, and it is shown that the updated values are quite close to the target. A representation of the updating process of the standard deviations is illustrated in Fig. 8. Sample iter.#0 is the starting sample in this step, identical to the last sample of the updating with the Euclidian distance (compare to Fig. 6). While the center of mass of the samples in iter.#0 and the

observation sample were already coincident with each other, their orientations and dispersions remain different. In the following iterations, the intermediary sample, e.g. sample iter.#2, is rotated toward to the orientation of the observations, and the dispersion is also reduced progressively. After the iterative updating process, the final sample shows a distribution identical to the target sample, implying the Bhattacharyya distance metric has successfully captured the dispersion information of the observations and the model predictions.

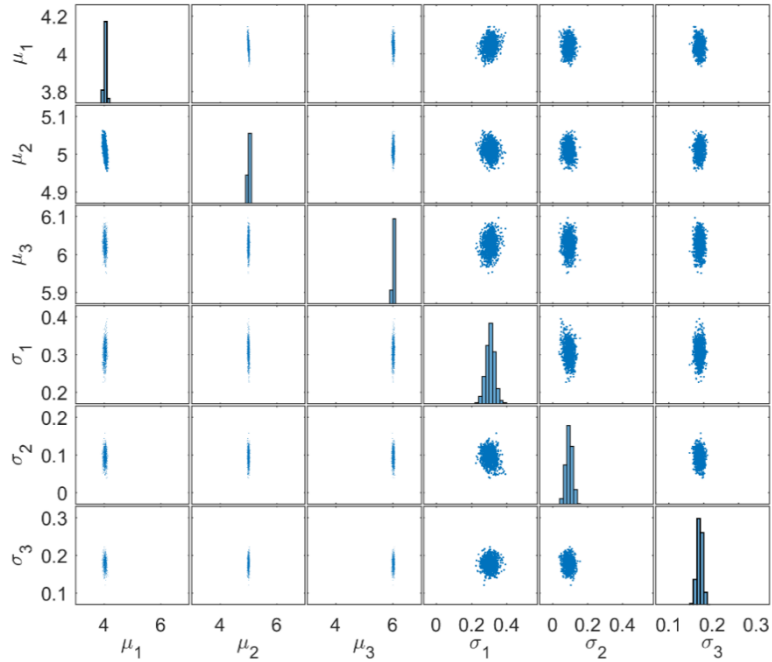


Fig. 7: The posterior distributions of the six inputs after updating with Bhattacharyya distance

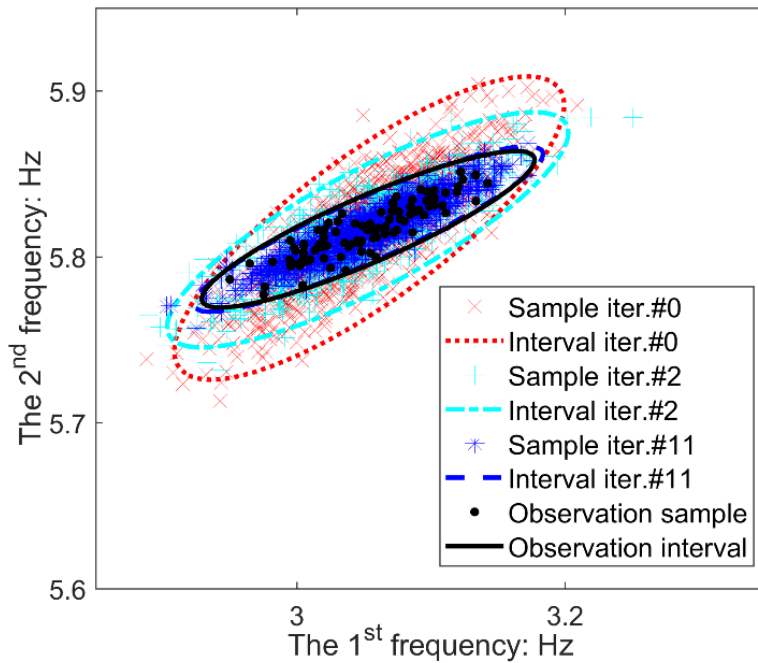


Fig. 8: The updated samples with Bhattacharyya distance

4.4 Summary

This example presented the combined application of the Euclidian and Bhattacharyya distances as updating metrics in a two-step ABC updating procedure. These two steps have the uniform framework and the only difference between them is the approximate likelihood constructed using the Euclidian or Bhattacharyya distance. The Bhattacharyya distance is demonstrated to be a more comprehensive UQ metric with the capability to recreate wholly the distribution of the target observations. However, a significant drawback of the Bhattacharyya distance is revealed in situations, e.g. in Fig. 4, where the initial sample is completely apart from the observation sample; that is to say, when the PDFs of these two samples have no superposition. In this situation, the value of the Bhattacharyya distance is always infinite regardless the varying discrepancies between these two samples.

Consequently, the two-step procedure is designed to overcome this drawback of the Bhattacharyya distance by employing the Euclidian distance in the first step to shift the initial sample toward the observation sample, and then to capture the actual distribution of the outputs in the second step. As a consequence, it is a good practice that a deterministic updating is used as a precondition of any stochastic updating. In other words, the classical Euclidian distance metric using the prediction and observation means should be kept as the backbone in any stochastic updating procedure.

5 The NASA UQ challenge problem

5.1 Problem Description

In this section, the NASA UQ challenge problem is investigated to demonstrate the capabilities of the Bhattacharyya distance metric in model updating for complex applications. The schematic in Fig. 9 illustrates the general structure of Subproblem A out of the overall problem including the model parameters, output, and the proposed UQ metrics.

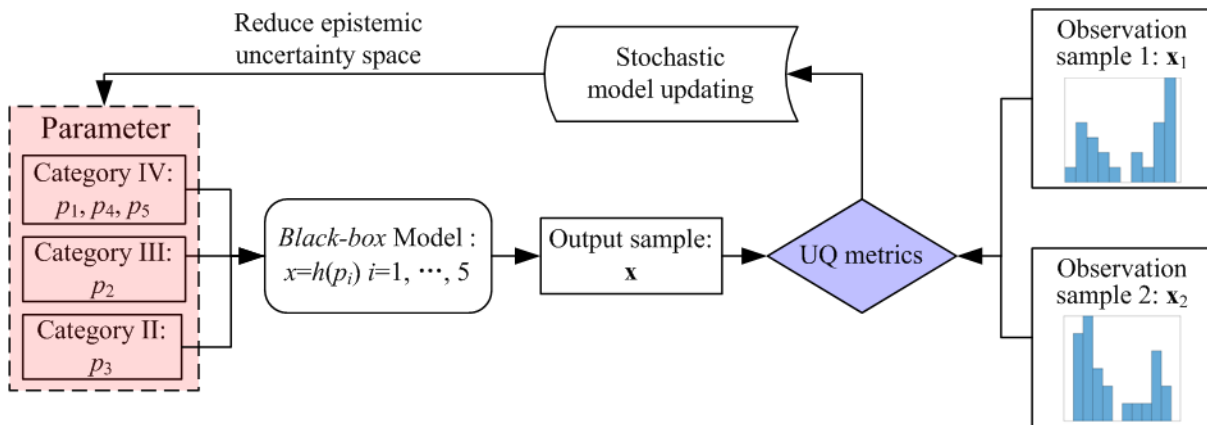


Fig. 9: Schematic of the NASA UQ challenge Subproblem A

As shown in Fig. 9, the system model is provided in a *black-box*, which evaluates a scalar output x using five parameters: p_1 - p_5 . According to the parameter categorization in Sec. 1, Table 3 presents the detailed uncertainty characterization of the parameters. p_1 , p_4 , and p_5 are Category III parameters with both aleatoric and epistemic

uncertainties, p_2 belongs to Category II with only epistemic uncertainty and p_3 is assumed to obey 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 aleatoric uncertainties, which are irreducible. The updating inputs relative to epistemic uncertainty of the parameters are listed in the last column of Table 3. In the following context, the “inputs” specifically refer to the epistemic uncertainty characteristics θ_1 - θ_8 , while the “parameters” denote the variables p_1 - p_5 used by the black-box model to evaluate the output x .

There are two observation sets of the output x , both containing 25 values respectively. Both samples are corresponding to a set of prescribed (but unknown in the original introduction of the problem) values of θ_1 - θ_8 . Three published works [14–16] on the NASA UQ challenge problem are referred here for a comparison to assess the feasibility and potential superiority of the Bhattacharyya distance metric in association with the ABC updating framework. Note that, in the original NASA UQ challenge problem, there were different tasks in Subproblem A, where the first 25 observations (\mathbf{x}_1) are supposed to be used for model updating in Task 1, and the remaining (\mathbf{x}_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, the first two tasks are skipped and only Task 3 is addressed, since the comparison of the results using 25 or 50 observations is not the focus herein. Instead, we pay more attention on the difference performances of the Euclidian and Bhattacharyya distance-based metrics, which is fully compared in the following section.

Table 3: The uncertain parameters of Subproblem A in the NASA UQ challenge problem

Category	Parameter	Uncertainty characteristics	Epistemic input
III	p_1	Unimodal Beta, $\mu_1 \in [0.6, 0.8]$ $\sigma_1^2 \in [0.02, 0.04]$	$\theta_1 = \mu_1, \theta_5 = \sigma_1^2$
	p_4, p_5	Gaussian, $\mu_i \in [-5.0, 5.0], \sigma_i^2 \in [0.0025, 4.0],$ $\rho \in [-1.0, 1.0], i = 4, 5$	$\theta_3 = \mu_4, \theta_4 = \mu_5,$ $\theta_6 = \sigma_4^2, \theta_7 = \sigma_5^2, \theta_8 = \rho$
II	p_2	Constant, $p_2 \in [0.0, 1.0]$	$\theta_2 = p_2$
I	p_3	Uniform, $\mu_3 = 0.5, \sigma_3^2 = 1/12$	--

5.2 Results Assessment

The two-step ABC updating procedure is executed to successively employ Euclidian and Bhattacharyya distances as the metrics. As a matter of fact, the initial sample of simulated output, based on the given initial input intervals, has enough overlap with the observation sample. Hence the Step I in Fig. 2 is not necessary for this problem. Nevertheless, these two steps are still completely executed in this example to demonstrate how the Euclidian distance alone, i.e. a mean-value only updating, is not enough. The detailed results of these two steps are compared in Sec. 5.2.1, so that the Bhattacharyya distance’s advantage upon the Euclidian distance is further revealed. The final result after the second step is also compared with the published results to assess the feasibility and potential improvements. A more comprehensive

assessment of the outcome is performed in Sec. 5.2.2 employing the now available true values of the inputs and p -boxes of the updated output.

5.2.1 Comparison with the published results

Fig. 10 illustrated a detailed comparison between the posterior histograms with the Euclidian and Bhattacharyya distance metrics. The samples and estimated PDFs with the Euclidian distance metric are denominated with the suffix “_ED”, while the ones with the Bhattacharyya distance metric are denominated with the suffix “_BD”.

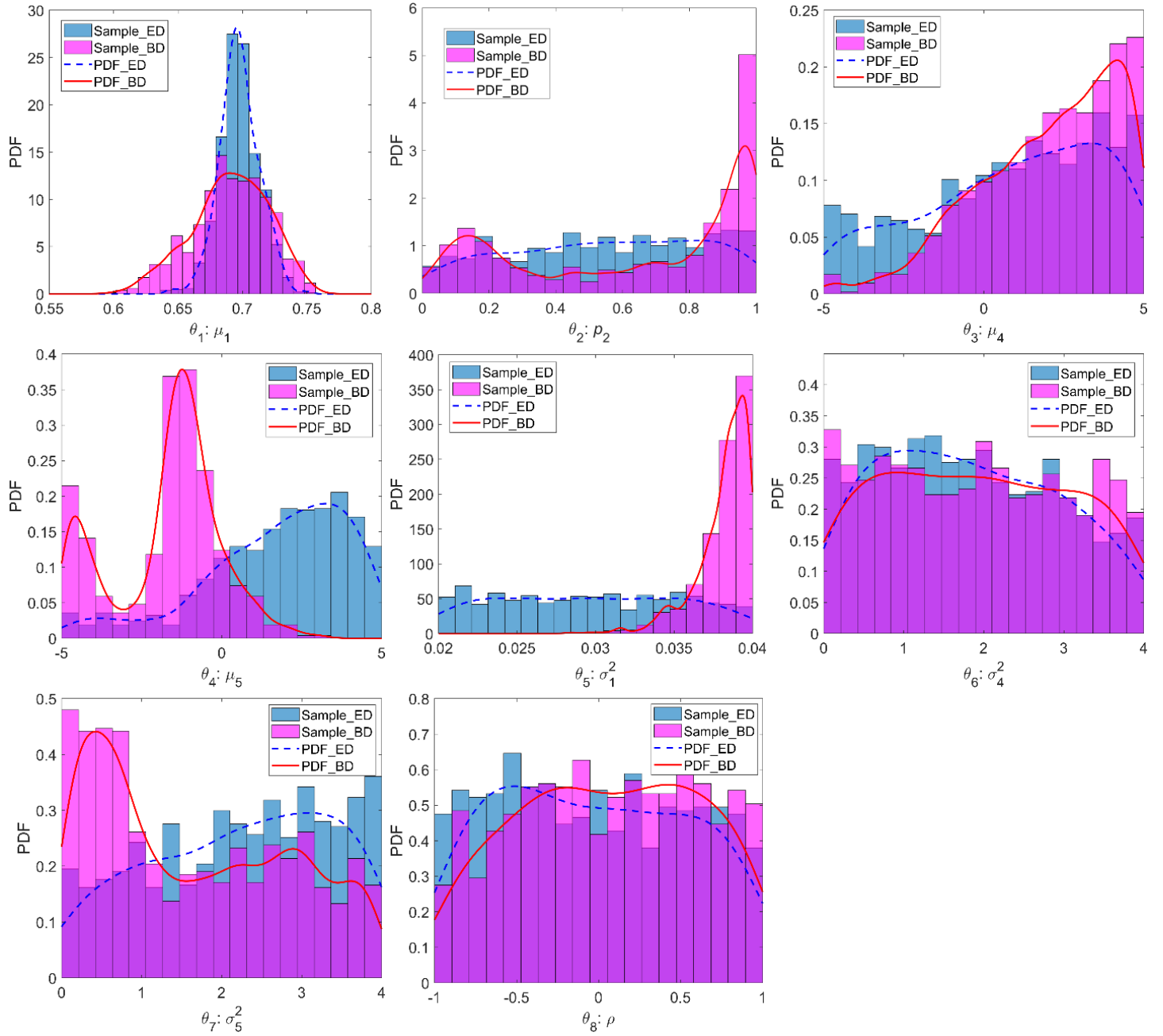


Fig. 10: Posterior distributions with the Euclidian and Bhattacharyya distances, PDFs are estimated via KDE

Considering the results with the Euclidian distance metric, only the posterior distributions of θ_1 and θ_4 are obviously updated. The distributions of the other inputs, especially the last four concerning the variances and correlation, remain approximately uniform and unchanged. As already proved in the previous example, only the input directly involved in the determination of the center of the mass of the samples, namely θ_1 and θ_4 herein, can be updated with the Euclidian distance metric. Similar histograms or PDFs, especially for the variances and correlation, are obtained in the works of Patelli et al. [14] and Ghanem et al. [15], implying the variance information is difficult to be updated by the classical Bayesian approaches with classic distance metrics.

As done previously, the posterior distributions of the first step are employed as prior distributions for the second step involving the Bhattacharyya distance. Fig. 10 clearly illustrates the further updated histograms and PDFs after the second step, showing clear improvements from the results of the first step, not only for the variances (i.e. θ_5 and θ_7) but also for the means/constant (i.e. θ_2 , θ_3 , and θ_4). Especially for θ_4 , the very discrepant PDF_ED and PDF_BD demonstrate that the Bhattacharyya distance metric is capable of calibrating (or correcting) the mean distributions when the results with the Euclidian distance metric are inappropriate. This phenomenon is different from the results in the 3-DOF example where the means and variances appear to be independent as they are successively updated in the first and second steps with the Euclidian and Bhattacharyya distance metrics, respectively. This difference is caused by the special distribution of the observation sample in this problem. The observation sample in the 3-DOF example is unimodal, while in the NASA challenge problem it is bimodal (refer to Fig. 9). The bimodal distribution results in interdependence between the mean and variance when evaluating the Bhattacharyya distance, which is one of the challenging features of this problem.

The final results with the Bhattacharyya distance metric are very similar to what was obtained by Safta et al. [16] except for θ_4 (i.e. the mean of p_5). For this input, the posterior distribution in Ref. [16] extremely converges to the lower bound, while the result herein is bimodal with one mode close to the lower bound and a second, more predominant, mode near the center of the interval. A potential explanation of this discrepancy is that the NASA UQ challenge problem contains multiple parameters with highly varying uncertainty characteristics, but end up with only one scalar output. This multi-input-single-output problem is more likely to produce non-unique solutions, as demonstrated by the fact that most of the published literature on this problem, e.g. Ref. [14–16], do not converge to a unique solution. Nevertheless, the proposed ABC approach with the Bhattacharyya distance metric demonstrates its ability on variance updating and capturing the complex non-linear relationship between inputs and outputs.

With regards the comparison of the computational expenses of the employed method, more attention is paid to the work of Safta et al. [16] since it shows very similar and highly accurate results. Safta’s method employed the full likelihood in Eq. (8). In order to evaluate the PDF of the output x for each instance of θ , a large number of simulated sample is required. As reported in Ref. [16], the simulation size for each PDF estimation is at least $N_{sim} = 10^5$, and totally 2×10^{11} model evaluations are executed in the complete updating process, leading to a huge computational burden in Safta’s method. Consequently, Ref. [16] also proposes an ABC treatment to replace the expensive full likelihood and use an approximate likelihood based on the Euclidian norm of selected sample statistics such as mean or/and quantiles.

However, in Safta’s ABC approach, the approximated likelihood’s ability of variance updating is not as obvious as the full likelihood (see figure 8 in Ref. [16]).

Comparatively, the ABC approach in this paper has a simple principle, employing only one distance metric in the Bayesian approach, but using the Bhattacharyya distance instead of the Euclidian distance in the ABC likelihood. Moreover, the evaluation of the Bhattacharyya distance in this example utilizes a sample with $N_{sim} = 10^3$. The overall procedure takes in total 1.3×10^7 model evaluations the TCMCMC algorithm is converged. Hence the calculation amount has been significantly reduced compared with Safta’s method by a factor 10^4 while achieving similar updating performances.

5.2.2 True values of the inputs and p -boxes of the output

In this paper, the updated values of θ_1 - θ_8 are obtained from their posterior distributions, and their correctness is assessed according to the recently released true values[†] as shown in Table 4. The posterior histograms are converted to distributions with KDE and are normalized so that their maximums are equal to one, as shown in Fig. 11. With this procedure, the posterior distributions are interpreted as Fuzzy sets [8], so that different levels of confidence will result to interval values of increased width. The “crisp” updated values of θ_1 - θ_8 are determined as the values corresponding to the maximum of the PDFs, i.e., the value corresponding to an alpha-level of 1. Therefore, these updated values are denominated as maximal probability (max- p) values in Table 4, together with the relative errors with respects to the true values. The max- p values of θ_2 and θ_4 are omitted in Table 4, since their posterior distributions remain approximately uniform as shown in Fig. 11, mainly because the output is insensitive to θ_2 and θ_4 , which has been addressed by the previously published works [14–16] in other sub-tasks of the NASA UQ challenge problem [2].

A more comprehensive assessment of the result is performed by estimating the p -boxes of the output. The initial intervals of θ_1 - θ_8 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 intervals of the inputs, so that the p -box of the output is accordingly reduced. In the ideal case, when the true input values are achieved from a perfect updating process, the resulting p -box would be reduced to a precise CDF curve, which should perfectly coincide with the observation CDF of the output. Based on the above motivation, three alpha-levels (namely 0.5, 0.75, and 0.9) are set for the normlized PDFs in as shown in Fig. 11. Table 4 presents the 0.9-level intervals (i.e., the higher confidence level), which are significantly reduced compared with the initial intervals.

[†] Data available online at <https://uqtools.larc.nasa.gov/nda-uq-challenge-problem-2014> [retrieved 2017]

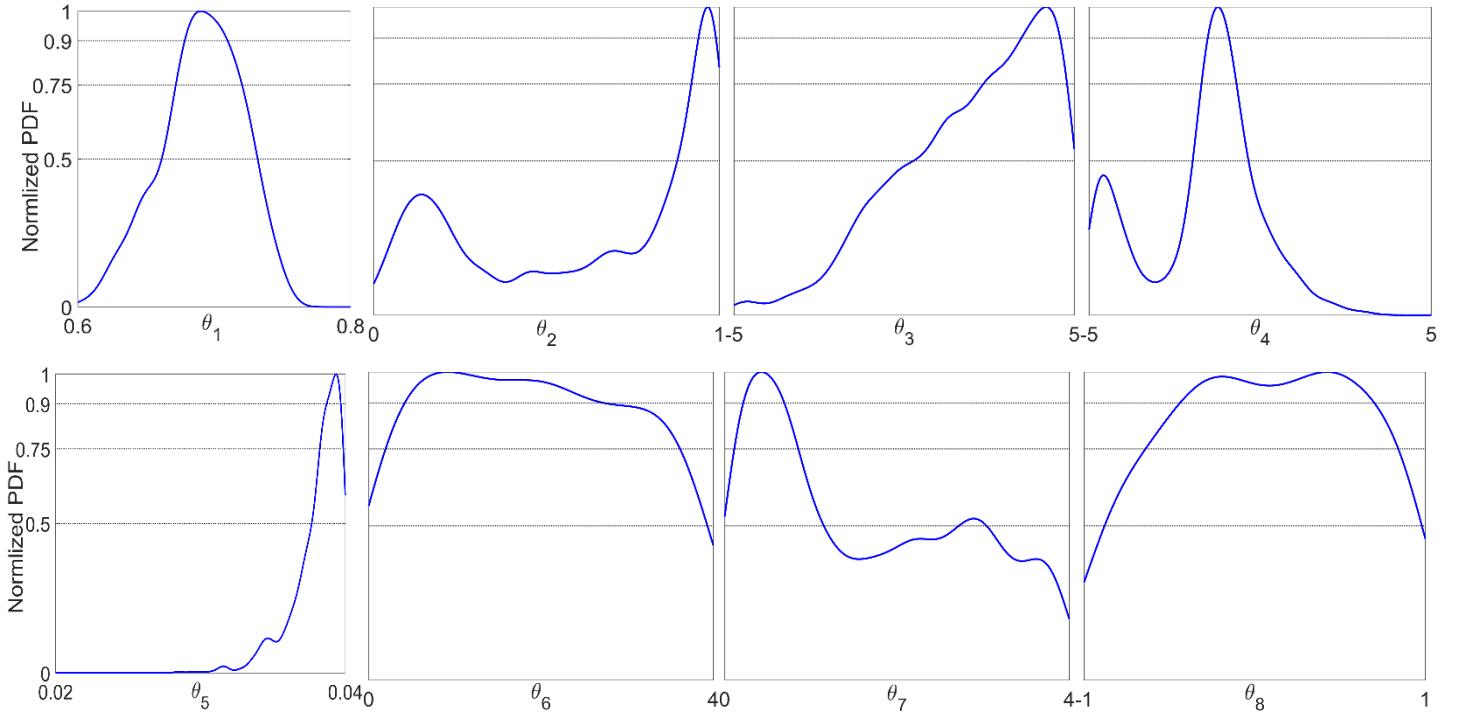


Fig. 11: Three truncation levels of the normalized posterior PDFs of θ_1 - θ_8

Table 4: The updated results of the epistemic inputs with Bhattacharyya metric

Input	Initial interval	True value	Max- p values ^a	0.9-level intervals
$\theta_1: \mu_1$	[0.6, 0.8]	0.6364	0.6901 (8.438)	[0.6783, 0.7097]
$\theta_2: p_2$	[0, 1]	1	0.9666(-3.34)	[0.9399, 0.9902]
$\theta_3: \mu_4$	[-5, 5]	4	4.1734 (4.335)	[3.4493, 4.5812]
$\theta_4: \mu_5$	[-5, 5]	-1.5	-1.2316(17.9)	[-1.5306, -0.9106]
$\theta_5: \sigma_1^2$	[0.02, 0.04]	0.0356	0.0394 (10.67)	[0.0387, 0.0397]
$\theta_6: \sigma_4^2$	[0.0025, 4]	0.04	--	[0.4190, 2.7209]
$\theta_7: \sigma_5^2$	[0.0025, 4]	0.36	0.4258 (18.72)	[0.2157, 0.6914]
$\theta_8: \rho$	[-1, 1]	0.5	--	[-0.4370, 0.7008]

^aPercent errors in the parentheses, with the absolute mean error as 10.57%.

The resulting p -boxes with the three alpha-levels are illustrated in Fig. 12. The p -box_ori represents the original epistemic uncertainty space, and it is significantly reduced by the Bayesian updating approach as compared with the updated p -boxes with different alpha-levels. An integrative comparison of Figs. 11 and 12 shows that the higher the alpha-level, the smaller the input epistemic interval, and furthermore, the narrower the resulting p -box of the output. Even so,

the narrowest p -box with 0.9-level can still envelop the observational CDF. This outcome clearly demonstrates feasibility of the Bhattacharyya distance metric in stochastic model updating and uncertainty characterization.

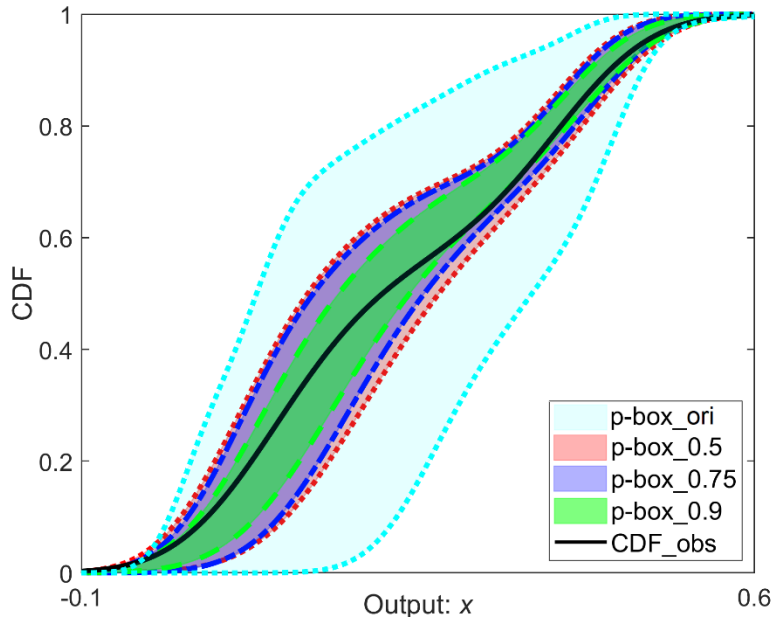


Fig. 12: Updated p -boxes with different alpha-levels compared with the smoothed CDF of the 50 observation

5.3 Summary

Discrepant results in the current literature clearly demonstrate the complexity and non-unique nature of the NASA UQ problem, which presents a more challenging task than the 3-DOF example in Sec. 4. The challenging nature of the NASA UQ problem is addressed by the following aspects:

- 1) The parameters involve multiple uncertainty models such as undetermined constants, undetermined variables with various distribution types;
- 2) The five-parameter-one-output problem (and the complicated functional relationship thereof) is more likely to produce non-unique solutions;
- 3) The bimodal distribution of the output leads to interdependence between the parameter means and variances;
- 4) The parameters have extremely different sensitivities, which can disturb the updating procedure.

Despite the above difficulties, the Bhattacharyya distance is proved to be a powerful UQ metric in the ABC updating approach. The selected published works and the recently released true values of the NASA UQ challenge problem provide a reliable assessment of the results, from which the summaries are concluded as follows.

- I) The Euclidian distance metric fails to solve this challenging problem, while the Bhattacharyya distance metric provides satisfied result for both mean and variance updating;
- II) The ABC updating framework with the Bhattacharyya distance metric provides similar (or better) results compared with the published approach employing the full likelihood, and the calculation cost is significantly reduced;

III) To exactly represent the true input values is an arduous (even unnecessary) task. However, the assessment of output p -boxes presents desirable outcome which fulfills the objective, i.e. to minimize the epistemic uncertainty space according to the available observation data.

6 Conclusions and perspectives

The Bhattacharyya distance is demonstrated to be a powerful UQ metric as a key ingredient of the proposed two-step ABC model updating framework. The comparison with the published works on NASA UQ challenge problem demonstrates the feasibility of the novel UQ metric for stochastic updating and uncertainty characterization. The approximate likelihood function built upon the concept of the distance acts as a convenient connection between the stochastic Bayesian updating framework and various distance-based metrics. The Bhattacharyya distance metric, designed as a universal tool, can simply replace the Euclidian distance metric in the stochastic Bayesian updating framework, as a stochastic counterpart to the deterministic concept.

Despite the advantage on uncertainty characterization, the Bhattacharyya distance metric is not appropriate to be used in an exclusive manner as revealed in the 3-DOF example. It needs to be complemented by the Euclidian distance metric in a two-step approach in order to achieve high-quality results. That is to say, the classically Euclidian distance together with observational means should be the primary consideration even in stochastic updating processes.

One of the perspectives of the Bhattacharyya distance as a UQ metric is to perform the stochastic sensitivity analysis, by quantifying and ranking the importance of input parameters, according to the uncertainty properties of the system outputs. This issue requires a comprehensive UQ metric to how much the uncertainty space of the outputs can be reduced, when the epistemic uncertainty space of the parameters is reduced. Another challenging perspective focuses on the imprecision of the experimental data. The Bhattacharyya distance in this work is based on histograms of the data samples with fixed bins. However, imprecision on data would need to be captured by the histogram with unfixed bins, and the counting numbers in each bin would also be changeable. This situation requires the Bhattacharyya distance to provide an interval-valued quantity, but no long a fixed value as utilized in this current work. This extension of the Bhattacharyya distance to cope with imprecision on data will be addressed in the future work.

Acknowledgement

This work is supported by the Alexander von Humboldt Foundation, which is greatly appreciated by the first author.

Reference

- [1] M.C. Kennedy, A. O'Hagan, Bayesian calibration of computer models, *J. R. Stat. Soc. Ser. B (Statistical Methodol.* 63 (2001) 425–464. doi:10.1111/1467-9868.00294.

- [2] L.G. Crespo, S.P. Kenny, D.P. Giesy, The NASA Langley Multidisciplinary Uncertainty Quantification Challenge, in: 16th AIAA Non-Deterministic Approaches Conf., 2014: pp. 1–9. doi:10.2514/6.2014-1347.
- [3] 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. doi:10.1016/j.ymsp.2008.03.001.
- [4] 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. doi:10.1016/j.ymsp.2005.06.006.
- [5] H.H. Khodaparast, J.E. Mottershead, K.J. Badcock, Interval model updating with irreducible uncertainty using the Kriging predictor, *Mech. Syst. Signal Process.* 25 (2011) 1204–1206. doi:10.1016/j.ymsp.2010.10.009.
- [6] Y. Govers, H.H. Khodaparast, M. Link, A comparison of two stochastic model updating methods using the DLR AIRMOD test structure, *Mech. Syst. Signal Process.* 52 (2015) 105–114. doi:10.1016/j.ymsp.2014.06.003.
- [7] M. Stein, M. Beer, V. Kreinovich, Bayesian approach for inconsistent information, *Inf. Sci. (Ny)*. 245 (2013) 96–111. doi:10.1016/j.ins.2013.02.024.
- [8] M. Beer, S. Ferson, V. Kreinovich, Imprecise probabilities in engineering analyses, *Mech. Syst. Signal Process.* 37 (2013) 4–29. doi:10.1016/j.ymsp.2013.01.024.
- [9] H.H. Khodaparast, J.E. Mottershead, Efficient Methods in Stochastic Model Updating, in: ISMA2008 Int. Conf. Noise Vib. Eng., 2008: pp. 1855–1869.
- [10] A. Bhattacharyya, On a Measure of Divergence between Two Multinomial Populations, *Indian J. Stat.* 7 (1946) 401–406. doi:10.1038/157869b0.
- [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. doi:10.2514/1.J055733.
- [12] J. Ching, Y.-C. Chen, Transitional Markov Chain Monte Carlo Method for Bayesian Model Updating, Model Class Selection, and Model Averaging, *J. Eng. Mech.* 133 (2007) 816–832. doi:10.1061/(ASCE)0733-9399(2007)133:7(816).
- [13] B.M. Turner, T. Van Zandt, A tutorial on approximate Bayesian computation, *J. Math. Psychol.* 56 (2012) 69–85. doi:10.1016/J.JMP.2012.02.005.
- [14] E. Patelli, D.A. Alvarez, M. Broggi, M. de Angelis, Uncertainty Management in Multidisciplinary Design of Critical Safety Systems, *J. Aerosp. Inf. Syst.* 12 (2015) 140–169. doi:10.2514/1.I010273.
- [15] 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. doi:10.2514/1.I010271.
- [16] 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. doi:10.2514/1.I010256.

- [17] S.J. Sheather, M.C. Jones, A reliable data based bandwidth selection method for kernel density estimation, *J. R. Stat. Soc.* 53 (1991) 683–690. doi:10.2307/2345597.
- [18] G.R. Grimmett, D.R. Stirzaker, *Probability and Random Processes*, Oxford university press, New York, 2001.
- [19] 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. doi:10.1016/j.knosys.2015.03.001.
- [20] J.L. Beck, S.-K. Au, Bayesian updating of structural models and reliability using Markov Chain Monte Carlo simulation, *J. Eng. Mech.* 124 (2002) 380–391.
- [21] W.K. Hastings, Monte Carlo sampling methods using Markov chains and their applications, *Biometrika.* 57 (1970) 97. doi:10.1093/biomet/57.1.97.
- [22] 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. doi:10.1061/(ASCE)0733-9399(2002)128:4(380).
- [23] 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. doi:10.1007/s00419-017-1233-1.
- [24] 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. doi:10.1016/j.ymsp.2017.10.015.
- [25] I. Park, R. V. Grandhi, Quantifying Multiple Types of Uncertainty in Physics-Based Simulation Using Bayesian Model Averaging, *AIAA J.* 49 (2011) 1038–1045. doi:10.2514/1.J050741.
- [26] E.C. DeCarlo, B.P. Smarslok, S. Mahadevan, Segmented Bayesian Calibration of Multidisciplinary Models, *AIAA J.* 54 (2016) 1–15. doi:10.2514/1.J054960.
- [27] M.A. Beaumont, W. Zhang, D.J. Balding, Approximate Bayesian computation in population genetics, *Genetics.* 162 (2002) 2025–2035.
- [28] Y. Govers, M. Link, Stochastic model updating-Covariance matrix adjustment from uncertain experimental modal data, *Mech. Syst. Signal Process.* 24 (2010) 696–706. doi:10.1016/j.ymsp.2009.10.006.
- [29] G. McLachlan, D. Peel, *Finite Mixture Models*, 2000. doi:10.1198/tech.2002.s651.