Mechanical Systems and Signal Processing 204 (2023) 110784

Contents lists available at ScienceDirect

Mechanical Systems and Signal Processing

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

Stochastic Model Updating with Uncertainty Quantification: An Overview and Tutorial

Sifeng Bi [a,*], Michael Beer [b,c,d], Scott Cogan [e], John Mottershead [c]

[a] Aerospace Centre of Excellence, Department of Mechanical and Aerospace Engineering, University of Strathclyde, Glasgow, UK
[b] Institute for Risk and Reliability, Leibniz University Hannover, Hannover, Germany
[c] Institute for Risk and Uncertainty, University of Liverpool, Liverpool, UK
[d] International Joint Research Center for Resilient Infrastructure & International Joint Research Center for Engineering Reliability and Stochastic Mechanics, Tongji University, Shanghai, China
[e] Department of Applied Mechanics, Femto-ST Institute, Besancon, France

ARTICLE INFO

Keywords:
Model updating
Uncertainty quantification
Uncertainty propagation
Bayesian updating
Model validation
Verification and validation

ABSTRACT

This paper presents an overview of the theoretic framework of stochastic model updating, including critical aspects of model parameterisation, sensitivity analysis, surrogate modelling, test-analysis correlation, parameter calibration, etc. Special attention is paid to uncertainty analysis, which extends model updating from the deterministic domain to the stochastic domain. This extension is significantly promoted by *uncertainty quantification* metrics, no longer describing the model parameters as unknown-but-fixed constants but random variables with uncertain distributions, i.e. *imprecise probabilities*. As a result, the stochastic model updating no longer aims at a single model prediction with maximum fidelity to a single experiment, but rather a reduced uncertainty space of the simulation enveloping the complete scatter of multiple experiment data. Quantification of such an imprecise probability requires a dedicated *uncertainty propagation* process to investigate how the uncertainty space of the input is propagated via the model to the uncertainty space of the output. The two key aspects, forward uncertainty propagation and inverse parameter calibration, along with key techniques such as P-box propagation, statistical distance-based metrics, Markov chain Monte Carlo sampling, and Bayesian updating, are elaborated in this tutorial. The overall technical framework is demonstrated by solving the NASA Multidisciplinary UQ Challenge 2014, with the purpose of encouraging the readers to reproduce the result following this tutorial. The second practical demonstration is performed on a newly designed benchmark testbed, where a series of lab-scale aeroplane models are manufactured with varying geometry sizes, following pre-defined probabilistic distributions, and tested in terms of their natural frequencies and model shapes. Such a measurement database contains naturally not only measurement errors but also, more importantly, controllable uncertainties from the predefined distributions of the structure geometry. Finally, open questions are discussed to fulfil the motivation of this tutorial in providing researchers, especially beginners, with further directions on stochastic model updating with uncertainty treatment perspectives.

* Corresponding author.
E-mail address: sifeng.bi@strath.ac.uk (S. Bi).

https://doi.org/10.1016/j.ymssp.2023.110784
Received 24 May 2023; Received in revised form 13 September 2023; Accepted 15 September 2023
Available online 23 September 2023

1. Background: Uncertainties, parameters, and categorisation

Numerical models and their simulations have been recognised to be significant, even irreplaceable, for the complete product life-cycle design and maintenance in almost all modern engineering fields. However, the critical role of models raises an equally critical problem: How much can we rely on the numerical model regarding the physical system? The fact is, unfortunately, there is always a discrepancy between the numerical simulation and the experimental measurement. Such an inevitable discrepancy inspired the development of *model updating* [1,2], since the early 1970 s, to calibrate the parameters or the model itself, such as to tune the model simulation toward the experimental measurement.

Various concepts/terms emphasising the trustfulness of numerical simulation have been developed in the community of computational mechanics engineering. The development of guidelines for the Verification and Validations (V&V) of Computational Solid Mechanics of ASME [3] sought to formalise the process of assessing the credibility of physics-based model simulations through dedicated metrics for classifying and quantifying different sources of uncertainties and their impacts on observable outcomes. A glossary is provided as follows to clarify and differentiate the various terms and definitions, with the expectation of bringing them together into a uniform framework.

- *Model updating*, also termed *model calibration*, is the process of adjusting physical or non-physical parameters in the computational model to improve agreement with experimental results.
- *Model verification* is the process of determining that a computational model accurately represents the underlying mathematical equation and its solution.
- *Model validation* is the process of determining the degree to which the model is an accurate representation of dedicated physical experiments from the perspective of its intended use.
- *Uncertainty quantification* is the process of characterising all uncertainties in the model or experiment and of quantifying their effect on the simulation or experimental outputs.
- *Uncertainty propagation* is the process of transferring the uncertainty characteristics from the input parameters to the output quantify of interest through the numerical model (or a specific pathway among multiple sub-models thereof).

From the perspective of uncertainty analysis, the discrepancy between models and physical systems is caused by various sources of uncertainties, and hence the main task of model updating becomes to investigate, categorise, and finally reduce the uncertainties during the modelling and experiment processes. The following three uncertainty sources are typically summarised in literature:

- **Parameter uncertainty**: In numerical modelling, the inherent properties of the physical system are imprecisely determined due to the lack of knowledge, especially for complex structural systems, novel composites, nonlinear dynamic systems, etc.;
- **Model form uncertainty**: the numerical models are developed with inevitable simplifications and approximations, i.e., linearisation of non-linear properties, simple elements instead of complex joints, etc.;
- **Experiment uncertainty**: although the measurements are regarded as the reference in model updating, they are unfortunately also uncertain because of the hard-to-control randomness in experiments, such as environmental noise, system errors, subjective judgment, etc.

The above uncertainties, in both models and experiments, lead to the typical single-simulation-single-test scenario being insufficient to capture the scattering feature of the simulation and experimental data. It also indicates that the deterministic model updating aiming at a single simulation with maximum fidelity to a single test is unnecessary, even misleading [4]. Fig. 1 provides an example of the potentially misleading situation when only one test point data and one simulation point data are employed during the deterministic model updating. As shown in Fig. 1(a), when the engineer has only these two single points, he would take the direct distance between them as his exclusive minimisation objective. However, as more data becomes available in Fig. 1(b), the truth is revealed that the single test and simulation points are not necessarily at the centre of the scatters of the simulation and test. The action in Fig. 1(b) just moved the simulation scatter even further apart from the test data. If no further validation is performed, such a model, only updated by the deterministic process, would lead to critical sequences in its applications.

In contrast, the stochastic model updating based on the multi-simulation-multi-test scenario requires the model parameters to be characterised as not merely unknown-but-fixed constant, but comprehensive characterisations including constants, intervals, precise and imprecise probabilities. A maximum fidelity regarding a single test is no longer the objective; on the contrary, stochastic model updating seeks to guarantee the robustness of the model while acknowledging the inevitable aleatory uncertainty, and is dedicated to reducing the epistemic uncertainty in the modelling and experiment process. The classification of uncertainties as to be epistemic or aleatory has been widely accepted. The epistemic uncertainty is led by the lack of knowledge, which is expected to be reduced by model updating. The aleatory uncertainty is derived from the inherent randomness of the system, and thus it cannot be reduced, whereas an appropriate representation is still required.

Note that, the uncertainty categorisation (being aleatory or epistemic) should be differentiated from uncertainty sources (from parameter, model form, and experiment). All three uncertainty sources are mixtures of both types of uncertainties, although the model form uncertainty and experimental uncertainty are dominated by epistemic uncertainty and aleatory uncertainty, respectively. The uncertainty in parameters is an equivalent combination of both the aleatory and epistemic uncertainties, since there are both fixed-but-unknown constants and random variables to be parameterised in the numerical model. Hence, a suitable parameterisation considering the source and type of uncertainty becomes a fundamental step of model updating. Fig. 2 provides a clear logic to perform

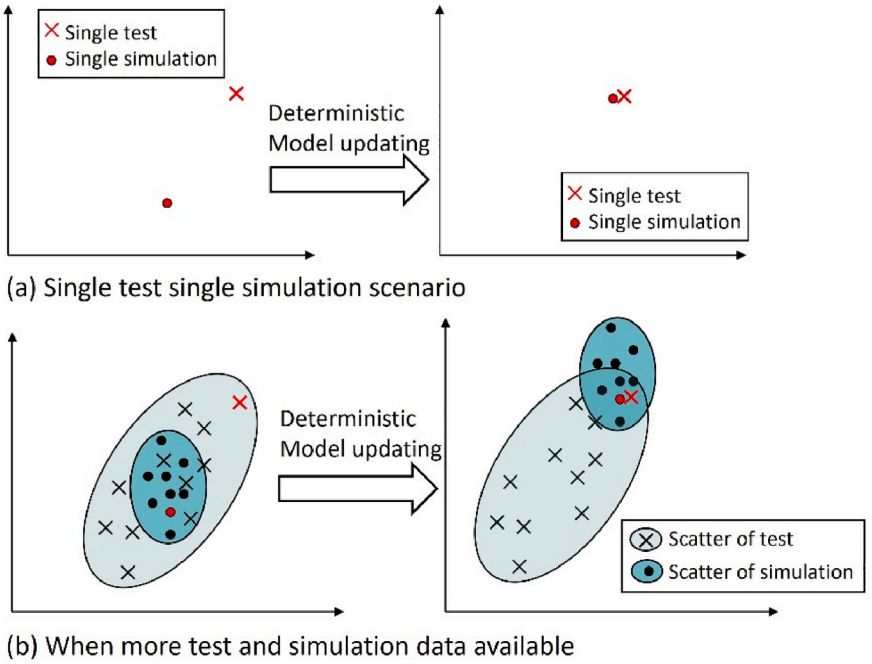


Fig. 1. Misleading of single test single simulation scenario during deterministic model updating.

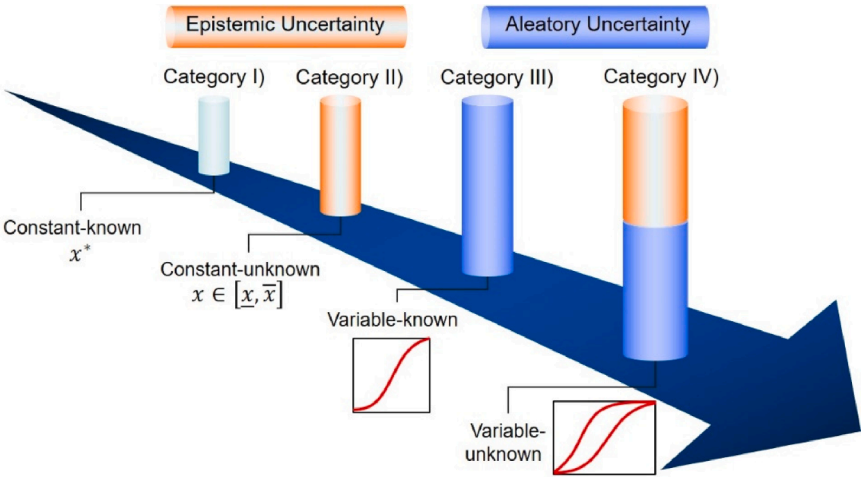


Fig. 2. Categorisation of parameters according to different types of uncertainties.

parameterisation by classifying parameters into four categories according to whether there are epistemic and/or aleatory uncertainties involved in the parameters.

Category I): The parameter with neither epistemic nor aleatory uncertainty is presented as a constant with a fully determined value.

Category II): The parameter with only epistemic uncertainty is presented as an unknown-but-fixed constant, falling within a pre-defined interval.

Category III): The parameter with only aleatory uncertainty is presented as a random variable with fully determined distribution properties, such as distribution format, mean, variance, etc. Such a fully determined distribution is termed as a "precise probability".

Category IV): The parameter with both epistemic and aleatory uncertainties is presented as a random variable whose distribution properties are not fully determined, i.e., the "imprecise probability". Such an imprecise probability is modelled by the so-called Probability-box (P-box), where an infinite number of Cumulative Distribution Function (CDF) curves constitute a specific region in the probability space.

The Category I parameters with no uncertainty are clearly an ideal case, which requires no treatment in model updating. The deterministic model updating falls exclusively within Category II parameters to determine a specific value within the pre-defined interval, such that the single model output would achieve the maximum fidelity regarding the single measurement. In contrast, the stochastic model updating is expected to handle all Categories II-IV parameters by reducing the epistemic uncertainty of both Categories II and IV parameters and appropriately characterising the aleatory uncertainty in Category III parameters. To achieve this objective, it is important to perform an appropriate measure of the difference between two uncertain parameters, especially for the Category IV parameters. More attention is consequently paid to the development of comprehensive metrics based on various statistical measures to capture as much as possible the different forms of uncertainty (epistemic, aleatory) associated with parameters.

The topic stochastic model updating is a synthetical technology system with key aspects, including sensitivity analysis, test-analysis correlation, and parameter calibration, where uncertainty analysis is deeply integrated as a critical role to promote the development of model updating from the deterministic domain to the stochastic domain. In sensitivity analysis, the forward uncertainty analysis, i.e. the uncertainty propagation, is important to determine how the uncertainty of the input parameters influences the uncertainty of the model predictions. In test-analysis correlation and parameter calibration, the inverse uncertainty analysis, i.e. the Uncertainty Quantification (UQ), is required to provide a quantitative measure of the discrepancy between model simulations and test measurements, such that both Category II and Category IV parameters in Fig. 2 can be quantified and calibrated.

Having the background of uncertainty sources, types, parameterisation, and categorisation introduced, the remaining part of this tutorial is organised as follows. A simple state-of-the-art of model updating is provided in Sec. 2 with emphasis on the development from the deterministic perspective to the stochastic perspective. An overall procedure for the stochastic model updating, as well as its key component, is elaborated in Sec. 3. Sec. 4 explains the key techniques for uncertainty propagation and parameter calibration, including double-loop P-box propagation, statistical distance-based UQ metrics, and Markov chain Monte Carlo Bayesian updating. Sec. 5 provides a step-by-step demonstration of the mentioned techniques by solving the NASA Langley Multidisciplinary UQ Challenge. Sec. 6 introduces a newly designed benchmark testbed, from which a measurement dataset with both experiment uncertainty and controllable parameter uncertainty is presented. Such a testbed and its measurement dataset would be an ideal case study to verify up-to-date uncertainty treatment techniques. Open questions and further perspectives on the stochastic model updating and uncertainty treatment are discussed in Sec. 7 to inspire further studies in this field.

2. A simple review: From deterministic to stochastic approaches

This section provides a simple state-of-the-art of model updating with representative literature for each key aspect or technique, such that the readers, especially the beginners, could draw a quick understanding of this topic emphasising the pathway from deterministic to stochastic developments.

A comprehensive review of model updating in structural dynamics, from a deterministic perspective, is provided by Mottershead and Friswell [1], who subsequently published the initial and fundamental monograph [5] on this topic covering the key aspects, e.g. model preparation, vibration experiment, sensitivity analysis, error localisation, parameter calibration, etc. Among the plentiful techniques for parameter calibration, the sensitivity-based method is one of the most popular approaches based on the linearisation of the generally nonlinear relation between the model parameters and structural dynamic features. Mottershead et al. [6] provide a tutorial literature for the sensitivity-based updating approach of finite element models with both demonstrative and industry-scale applications. The sensitivity-based approach is valid for typical modal features, e.g. natural frequencies and mode shapes, whose sensitivity can be theoretically derived from the stiffness and mass. These theoretical derivations become impractical for modern structural engineering where strong nonlinear dynamics or transient analysis are presented with large-scale structure systems.

Instead of the theoretical solutions of sensitivity, the numerical simulation approach, more specifically the Monte Carlo method, attracts more interest by providing a direct connection between the model parameters and any output features via multiple deterministic model evaluations. More importantly, the random sampling process has the natural adaptation with uncertainty analysis, because the samples are always obtained from probabilistic hypothesis [7]. This tendency is further promoted by the significant development of the computational technique, which makes it possible for large-size sampling, from which the statistical information of the variables is precisely estimated [8]. The advanced or efficient Monte Carlo based methods have been successfully applied in large scale structures, see e.g. Refs. [4,9].

In the first decades of this century, stochastic model updating has been developed into two genres: the frequentist and Bayesian approaches. The frequentist approach focuses on optimisation techniques to minimize the discrepancy between the existing measurements and the model simulations. In a stochastic sense, the minimizing object is, of course, not only the mean of the data, but also some frequentist properties, such as the distribution coefficients [8] and covariance [10]. For the Bayesian genre, Beck and Katafygiotis [11] proposed the fundamental framework of Bayesian updating, which was further developed via the Markov chain Monte Carlo (MCMC) sampling by Beck and Au [12]. Since it derives the posterior distribution of the calibrating parameters to estimate their actual values, the Bayesian approach naturally involves uncertainty treatment in the updating. It also has the superiority to capture uncertainty information in the presence of very rare measurement data. As a result, the Bayesian approach has been adopted as one of the most popular techniques in stochastic model updating. For example, in both editions of the NASA Langley UQ Challenge 2014 [13] and 2021 [14], the Bayesian updating with MCMC algorithm was adopted by most of the responding groups, e.g. Refs [15–17] to solve this problem. Comparisons between the sensitivity-based, frequentist, and Bayesian approaches can be found in Refs. [18,19].

In addition to the frequentist and Bayesian methodologies, recent developed imprecise probability techniques also have considerable potential to be applied in stochastic model updating, such as interval probabilities [20], evidence theory [21], info-gap theory [22], and fuzzy probabilities [23]. For the background of imprecise probability, the comprehensive review by Beer et al. [24] is

suggested for an overall understanding of this topic. A comprehensive review of the non-probabilistic approaches, especially the interval and fuzzy techniques are provided by Ref. [20]. Note that, regardless of which approach is adopted for uncertainty modelling and quantification, in the context of model updating, it is always critical to propose a suitable calibration metric to quantitatively measure the discrepancy between the model simulation and experimental measurement. Such a calibration metric, also known as the UQ metric, plays an important role in constructing the objective function in frequentist updating and the likelihood function in Bayesian updating. Pioneering work on this issue can be found in Refs. [25,26] where the statistical distance-based metrics, i.e. the Mahalanobis distance and Bhattacharyya distance, are introduced in stochastic model updating.

3. Stochastic model updating methodology

This section provides an overview of the complete methodology for stochastic model updating as illustrated in Fig. 3. The procedure, starting from the initial numerical model and the experimental setup, contains key steps such as feature definition, model parameterisation, surrogate modelling, parameter calibration, and model validation. The final outcome is the validated model according to multiple experimental measurements. Differing from the deterministic procedure, the objective herein is no longer a determined simulation with maximum fidelity regarding single test data, but calibrated probability properties of uncertain parameters which can represent the dispersion feature of the existing experimental measurements, i.e., a set of numerical predictions enveloping the existing experimental uncertainty.

From the perspective of uncertainty treatment, the presented technique route is committed to reducing the epistemic uncertainty, while an appropriate representation of the aleatory uncertainty is also required. Considering the three sources of uncertainties as described in Sec. 1, this technique route is designed to handle all three sources by 1) presenting the experiment uncertainty in multiple sets of measurement data; 2) parameterising the model form uncertainty and quantifying it together with 3) the parameter uncertainty in the mixture form of interval, distribution, and P-box (recall Fig. 2). In Fig. 3, the components surrounded by the light-yellow box within the dashed line are the ones different from the deterministic updating and require an additional extension for uncertainty treatment. The selected key steps are explained in the following subsections.

3.1. Feature definition

The output "feature" is defined as the quantity of interest that the engineer wants to predict from the numerical model. Different features clearly have different formats (e.g. scalar, vector, time-/frequency-domain sequence, and random process) with different uncertainty properties, and thus require different quantification and calibration methods. The different features also have different sensitivities according to the input parameters, leading to different results of the sensitivity analysis and different parameters to be calibrated in the following procedure. It is thus the first step of model updating to define suitable features based on practical

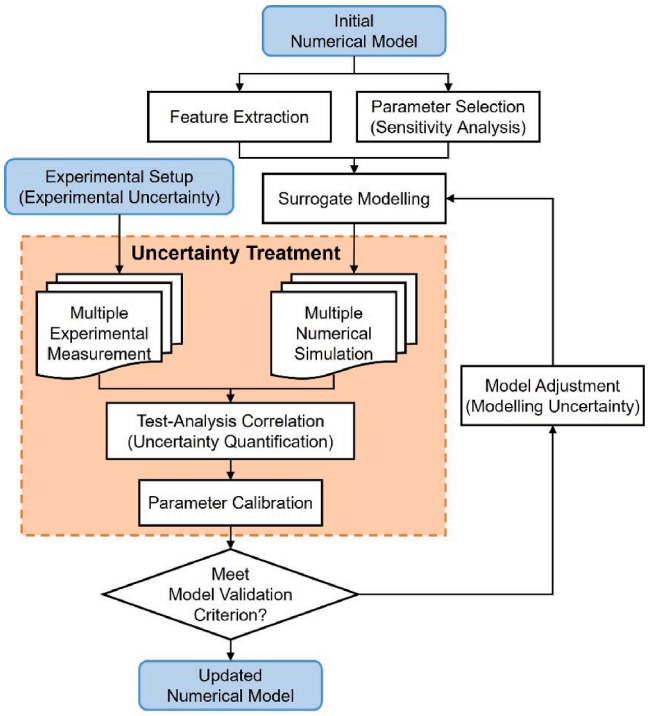


Fig. 3. Stochastic model updating methodology with uncertainty treatment.

requirements and the capacity of the experimental setup.

The most typical features in structural dynamics updating are the modal quantities, i.e., the natural frequencies and mode shapes. The absolute mean error is simply utilised to quantify the discrepancy between the simulated and experimental natural frequencies. And for mode shapes, the Modal Assurance Criterion (MAC) [27] is a typical tool to measure the correlation between the simulated and experimental eigenvectors. Besides the discrete modal quantities, it is common to encounter continuous quantities such as time domain system responses or the Frequency Response Function (FRF). Typical techniques to deal with these continuous quantities are to discretise the time-/frequency-domain curve into multiple feature values at key time/frequency points. And the so-called Signature Assurance Criterion (SAC) and Cross Signature Scale Factor (CSF) are calculated based on a similar principle of MAC. Ref. [28] presents an integrated application of SAC and CSF for a comprehensive comparison between two FRFs.

3.2. Parameterisation and sensitivity analysis

Parameterisation refers to the first definition and extraction of the input parameters from the numerical model. Always performed based on engineering judgement, the parameterisation results in a large number of parameters which could be uncertain or significant to the model output. Subsequently, the sensitivity analysis has been developed as a typical technique to measure the significance of input parameters with respect to the output features, thus aiding the selection of key parameters to be calibrated in the next step. The classical, also deterministic, technique is the Sobol's variance-based method [29]. For a comprehensive knowledge of the global sensitivity analysis inspired by Sobol's method, the well-written book by Saltelli et al. [30] is suggested to the readers.

When multiple categories of uncertain parameters (as shown in Fig. 2) are presented in a single problem, it is necessary to extend the sensitivity analysis from the deterministic perspective to the stochastic perspective. The uncertain parameters raise a question which cannot be addressed by the deterministic sensitivity analysis: Considering a parameter which is concluded to be insensitive based on its perturbation through the deterministic sensitivity analysis, what would happen if the parameter were found to be extremely uncertain? If its uncertainty boundary is clearly larger than the determined perturbation used in the deterministic analysis, how is one to measure its sensitivity in the presence of epistemic and aleatory uncertainties? The stochastic sensitivity analysis is committed to answering such a question - by ascertaining to what extent the uncertainty space of the features can be reduced, when the epistemic uncertainty space of the parameters is completely reduced. This requires additional techniques to propagate the uncertainty space from the input parameters to the output features, which will be addressed in Sec. 4 using the double-loop uncertainty propagation technique.

3.3. Surrogate modelling

A surrogate model is a fast-running script between the selected parameters (in Sec. 3.2) and the defined features (in Sec. 3.1), which is utilised to replace the time-consuming numerical model, e.g. a sophisticated finite element model. The surrogate model is especially useful for the sampling-based stochastic approach, where a large number of model evaluations are required for uncertainty propagation and quantification. There are numerous formats of surrogate models, including the polynomial function, radial basis function, support vector machine, Kriging function, and neural network, etc. The selection of a suitable surrogate model format is determined according to its efficiency, generality, and nonlinearity. Regardless of the selected model format, a certain number of training samples are always required to train the surrogate model. Since the surrogate model is expected to balance the trade-off between efficiency and precision, the size of training samples is expected to be as small as possible, while the precision of the surrogate model should be high enough. A suitable set of training samples is available through another related technique, i.e. the Design of Experiment (DoE), with the aim to efficiently and uniformly configure a spatial distribution of the samples within the complete parameter space. A comprehensive review of the existing techniques of surrogate modelling and DoE can be found in Ref. [31].

3.4. Parameter calibration

The parameter calibration, the core of model updating, is essentially an inverse procedure taking the discrepancy between the simulated and measured output features as a reference, and focusing on the principle and technique about how to calibrate the input parameters. To describe this task as an optimisation problem, the output discrepancy is employed to construct the objective function, which will be minimised by searching suitable values of parameters with their epistemic space as the optimisation constraint. Another calibration strategy is the Bayesian updating where the prior distribution of the parameters is expected to be updated with respect to the likelihood function of the existing measurements, and the updated posterior distribution is obtained with reduced epistemic uncertainty. Comparison studies between the sensitivity-based optimisation approach and the Bayesian approach are presented in Refs. [18,19] where both approaches are applied in a Benchmark model updating problem, i.e. the GARTEUR SM-AG19 (also known as AIRMOD) structure [32].

Nevertheless, either in the optimisation or the Bayesian process, the Test-Analysis Correlation (TAC) is important, not only because it significantly influences the calibration outcome, but also because it is the part mostly extended due to uncertainty treatment. TAC is the step to quantitatively measure the agreement (or lack thereof) between simulations and measurements, taking uncertainties into account. It, therefore, requires a comprehensive metric which can capture multiple uncertainty sources simultaneously. Some recently developed UQ metrics are based on statistical distances. A comprehensive comparison among the Euclidian, Mahalanobis, and Bhattacharyya distances in model updating and validation can be found in Ref. [25], where the Bhattacharyya distance is found to be more comprehensive to capture more sources of uncertainties. In this paper, special emphasis is paid to the novel UQ metric based on

the Bhattacharyya distance, whose performance in parameter calibration will be demonstrated in the example section.

3.5. Model validation

Model validation is an essential step to assess the predictive ability of the model before it can be practically utilised. A series of validation criteria with increasing requirements are provided in Ref. [6]:

- As the most basic criterion, the model should predict the existing measurements used for parameter calibration;
- The model should predict an independent set of measurements which is different from the ones used for parameter calibration;
- The model should predict the practical modification of the physical system by making the same modification on the model;
- The calibrated model, when utilised as a component of an assembly, should improve the prediction of the assembly model of the complete system.

Since model form uncertainty is inevitable for any numerical model, no matter how sophisticated it is, it is not rare to find the calibrated model cannot fulfil the above criteria. In another situation, the calibrated model can fulfil the criteria. However, the calibrated values of its parameters are found to be unphysical or out of the pre-defined boundaries. This is because the modelling uncertainty is so severe that it cannot be compensated by calibrating the parameters within their physical ranges. In this case, the numerical model must be adjusted to reduce the modelling uncertainty by, e.g., increasing the mesh resolution, using 3D elements to replace 2D elements, etc. A new round of parameter calibration is required for the new model until it fulfils the above criteria without unphysical parameter values.

4. Key techniques of forward uncertainty propagation and inverse parameter calibration

Throughout the model-updating process, key aspects such as sensitivity analysis, parameter calibration, and Test-Analysis Correlation are all influenced by whether or not the uncertainty can be comprehensively and efficiently quantified and propagated. This section is consequently focusing on three key techniques ensuring uncertainty treatment can be firmly implemented in the process, i.e. the double-loop P-box propagation, statistical distance-based UQ metrics, and Markov chain Monte Carlo (MCMC) Bayesian updating.

4.1. Double-loop P-box propagation

In the presence of multiple parameter categories, as shown in Fig. 2, it is a critical but complex task to investigate how the uncertainty properties are propagated from the input parameters to the output features through the model. As illustrated in Fig. 4, a double-loop framework employing both Monte Carlo simulation and optimisation processes are proposed, where the aleatory uncertainty is handled by the Monte Carlo simulation and the epistemic uncertainty is handled by the optimisation, in the outer and inner loop, respectively. Such a double-loop process is capable of decoupling the aleatory uncertainty and epistemic uncertainty in a clear logic.

The four categories of parameters, termed as p_{1-4}, have four different formats of Cumulative Distribution Functions (CDFs) as shown in the right part of Fig. 4. In the outer-loop, the Monte Carlo simulation is performed along the vertical axis of the four CDFs within the probability range [0, 1]. The randomly sampled probability value α corresponds to different objects on the horizontal axis, termed as *random sets* in the context, for different categories of parameters:

Category I: Different α values result in a fixed constant p^*;

Category II: Different α values result in a fixed interval $\left[\underline{p}, \overline{p}\right]$;

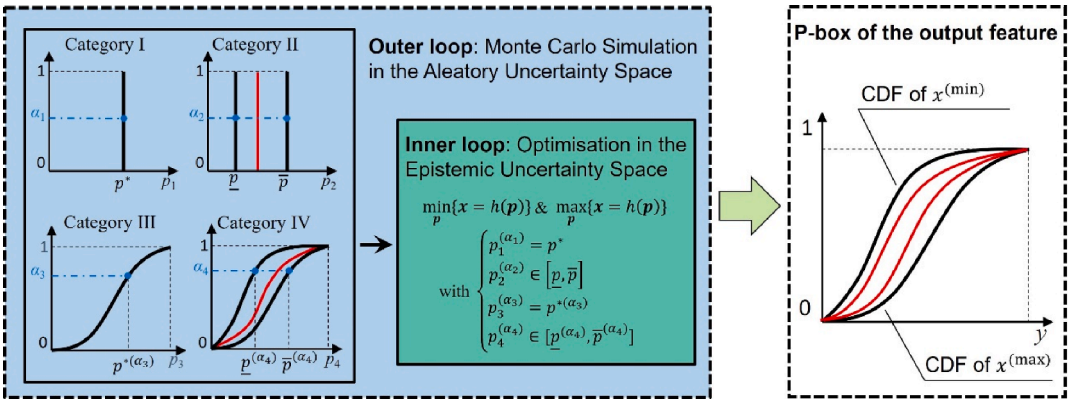


Fig. 4. Double-loop procedure for uncertainty propagation in the form of P-box.

Category III: Different α values result in a scalar with changing position $p^{*(\alpha)}$;

Category IV: Different α values result in an interval with changing bounds $\left[\underline{p}^{(\alpha)}, \overline{p}^{(\alpha)}\right]$.

The above random sets are utilised in the inner loop as the constraint of the optimisation problem, where the minimum and maximum of the output features are found through a constrained optimisation:

Find two sets of input parameters $\boldsymbol{P}^{(min)}$ and $\boldsymbol{P}^{(max)}$, respectively, minimising and maximising the output feature $\boldsymbol{x}$

$$\begin{aligned} &\min_{p}\{\boldsymbol{x} = h(\boldsymbol{P})\} \\ &\max_{p}\{\boldsymbol{x} = h(\boldsymbol{P})\} \end{aligned} \tag{1}$$

with the constraints

$$\begin{aligned} &p_1^{(\alpha_1)} = p^* \\ &p_2^{(\alpha_2)} \in \left[\underline{p}, \overline{p}\right] \\ &p_3^{(\alpha_3)} = p^{*(\alpha_3)} \\ &p_4^{(\alpha_4)} \in \left[\underline{p}^{(\alpha_4)}, \overline{p}^{(\alpha_4)}\right] \end{aligned} \tag{2}$$

where $h(\boldsymbol{P})$ is the numerical model through which the uncertainty propagation is performed. It can be a sophisticated Finite Element (FE) model, or a much faster surrogate model as explained in Sec. 3.3.

For a certain number of Monte Carlo samples in the outer loop, the same number of optimisations are executed to generate the minimum and maximum sample pairs of the feature $\boldsymbol{x}$. The CDFs of the minimum and maximum samples are estimated as the upper and lower bounds of the P-box in the right side of Fig. 4. This double-loop approach provides a clear logic for uncertainty propagation, and the precision can be very high when the number of Monte Carlo samples is large enough. This approach therefore has the potential to be employed as a standard method producing benchmark result that may be used to assess the feasibility of other approximate methods of uncertainty propagation. A large number of optimisations may raise the calculation cost. However, Ref. [33] has demonstrated that the overall calculation cost of the approach is acceptable since each single optimisation is performed on a very narrow interval and thus can be solved very quickly. This very narrow interval is driven by the random α value, which truncates only a small part of the original range of the Category IV parameter as shown in Fig. 4.

Another treatment that can significantly reduce the calculation time is parallel computing. The multiple optimisations in the inner loop are independent because the aleatory uncertainty value α is randomly sampled from the outer loop. Such a characteristic makes it immediately suitable for parallel computing. It will be shown in the case study section that the calculation time of the double-loop procedure is acceptable on a standard desktop computer.

Considering the objective of stochastic sensitivity analysis, the explicit task as explained in Sec. 3.2 is to quantify how much the uncertainty properties of output can be changed, when the epistemic uncertainty of input parameter is reduced. As shown in the left part of Fig. 4, the ideally complete reduction of epistemic uncertainty of Categories II and IV parameters implies their P-boxes are compressed to single curves. Such compression of the input parameter P-boxes clearly leads to compression of the output feature P-box. How much the compression occurs to the feature P-box becomes the sensitivity index of the input parameter uncertainty.

4.2. Statistic distance-based UQ metrics

Although providing a visual representation, the P-box itself is not a quantitative index. It is therefore necessary to develop metrics

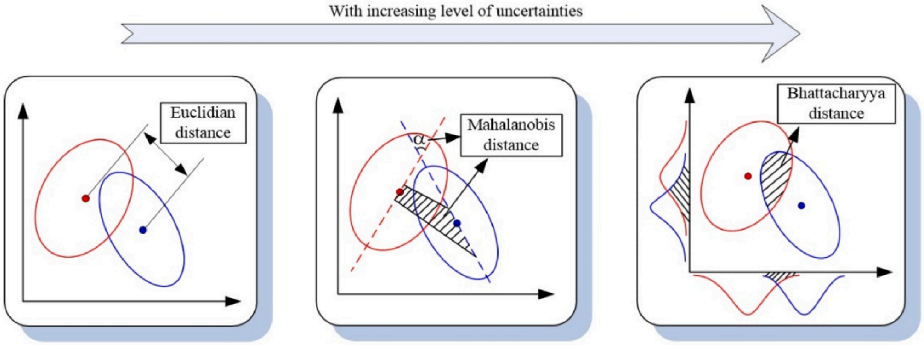


Fig. 5. Sketch of three statistical distances.

to measure the size of the P-box. Not only being useful in sensitivity analysis, but these metrics should also play key roles in the inverse parameter calibration by measuring the discrepancy between the model prediction and measurement data. Since the P-box is a region surrounded by two CDF curves, it is natural to quantify this region by comparing the difference between these two CDFs, i.e., two distributions. Hence, various statistical distances are proposed on this occasion, as sketched in Fig. 5.

The Euclidian distance, i.e., the absolute geometry distance between two single points, is probably the most common metric used in deterministic model updating. In the presence of multiple samples of the numerical and measured outputs, a common treatment is to measure the Euclidian distance between the centre points of the two sample sets (i.e. mean value of the random samples):

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

where $\mathbf{Y}_{exp}$ and $\mathbf{Y}_{sim}$ are the measured and simulated feature samples, respectively. The Euclidian distance becomes insufficient for stochastic model updating because the dispersion properties are completely ignored in Eq. (3).

The Mahalanobis distance is a weighted Euclidian distance, where the covariance matrix is employed as the weighting coefficient, expressed as

$$d_M\left(\mathbf{Y}_{exp}, \mathbf{Y}_{sim}\right) = \sqrt{\left(\overline{\mathbf{Y}_{exp}} - \overline{\mathbf{Y}_{sim}}\right)\mathbf{C}^{-1}\left(\overline{\mathbf{Y}_{exp}} - \overline{\mathbf{Y}_{sim}}\right)^T} \tag{4}$$

where $\mathbf{C}$ is the "pooled" covariance matrix of both the simulation and measurement samples. It is defined as [25]

$$\mathbf{C} = \frac{(m-1)\mathbf{C}_{sim} + (n-1)\mathbf{C}_{exp}}{m+n-2} \tag{5}$$

where m and n are the numbers of simulated and measured samples, respectively. $\mathbf{C}_{sim}$ and $\mathbf{C}_{exp}$ are the covariance matrices of the simulated and measured samples. Although the covariance information is considered, it turns out that the Mahalanobis distance is not a suitable metric for parameter calibration. The reason is that, from Eqs. (4) and (5), sample sets with large covariances would always lead to a small Mahalanobis distance. When it is employed as a calibration metric, the model updating algorithm tends to enlarge the variance so as to reach a minimised distance. Nevertheless, the Mahalanobis distance can be used as a validation metric to assess whether or not the updated samples are consistent with an independent set of reference data.

The Bhattacharyya distance is a statistical distance measuring the overlap between two random distributions with the expression as

$$d_B\left(\mathbf{Y}_{exp}, \mathbf{Y}_{sim}\right) = -\log\left[\int_{\mathbb{y}} \sqrt{P_{exp}(y)P_{sim}(y)}\,\mathrm{d}y\right] \tag{6}$$

where P_{exp} and P_{sim} are the Probability Density Functions (PDFs) of the measured and simulated variables, respectively. Clearly the Bhattacharyya distance is more comprehensive for uncertainty quantification, since the complete distribution information is involved. However, the explicit PDF is not always available in practical applications due to the limited data from experiments. It is therefore suggested to utilize the Probability Mass Function (PMF) of a discrete distribution to replace the PDF in Eq. (6). The PMF can be obtained by estimating the discrete distribution based on the available measurement samples via the so-called Binning Algorithm [26]. This is a technique to achieve the balance between the limitation of existing measurements and the high requirement of uncertainty information. For theoretical completeness, the Binning Algorithm is simply recalled here as a similar procedure to plot a histogram of random samples. The first step is to determine a public range containing all samples of both $\mathbf{Y}_{exp}$ and $\mathbf{Y}_{sim}$. In the second step, it is important to select a suitable number of bins n_{bin} based on the number of samples in $\mathbf{Y}_{exp}$ and $\mathbf{Y}_{sim}$, and the edge of the bins, i.e. the grid, is subsequently determined by uniformly cutting the public range into n_{bin} parts. The next step is to count how many samples falling within each bin, and the PMFs of $\mathbf{Y}_{exp}$ and $\mathbf{Y}_{sim}$ are obtained, respectively. Note that, when multiple dimensional features are considered, say m dimensions, the PMF is actually a m-dimensional table. The number of bins for each single dimension is n_{bin}, and thus the total number of bins in the PMF table is n_{bin}^m. As a result, the Bhattacharyya distance, between two m-dimensional PMFs, is calculated as

$$d_B\left(\mathbf{Y}_{exp}, \mathbf{Y}_{sim}\right) = -\log\left[\sum_{i_m}^{n_{bin}} \cdots \sum_{i_1}^{n_{bin}} \sqrt{p_{exp}\left(b_{i_1,i_2,\cdots,i_m}\right)p_{sim}\left(b_{i_1,i_2,\cdots,i_m}\right)}\right] \tag{7}$$

where $p\left(b_{i_1,\ i_2,\ \ldots,\ i_m}\right)$ is the PMF value of the bin $b_{i_1,i_2,\cdots,i_m}$. Clearly, when m is large, Eq. (5) requires a huge memory storage, which delays the calculation speed. As explained in Sec. 3.1, the techniques of feature definition are always employed in practical applications to extract a single quantity of interest from multiple features. Also, the margin PDM of a specific variable from the whole feature space is always employed in practical applications. The practical example in Sec. 6 will demonstrate the selection of different margin distributions to configure a feasible Bhattacharyya distance-based calibration metric. The Bhattacharyya distance for a one-dimensional PDM is much simplified from Eq. (7) to

$$d_B\left(\mathbf{Y}_{exp}, \mathbf{Y}_{sim}\right) = -\log\left[\sum_{i}^{n_{bin}} \sqrt{p_{exp}(b_i)p_{sim}(b_i)}\right] \tag{8}$$

The pre-defined n_{bin} has significant influence on the calculated distance value. A larger n_{bin} leads a larger value of the distance. Considering the extreme case, on the one hand when $n_{bin} = 1$, all samples would always fall into the single bin, and thus the Bhattacharyya distance value would be zero for any random samples. On the other hand, when n_{bin} is large enough, there would be no sample from $\mathbf{Y}_{exp}$ and $\mathbf{Y}_{sim}$ falling within a public bin of the PMFs, leading the Bhattacharyya distance to be infinite. A general principle to determine n_{bin} is to select a value making the distance to be sensitive with regard to any change of the investigated samples. Although this value is generally case-dependent, an estimation of n_{bin} is suggested as

$$n_{bin} \cong \max\left(n_{exp}, n_{sim}\right)/5 \tag{9}$$

4.3. Bayesian updating with Markov chain Monte Carlo algorithm

The Bayesian updating strategy has a natural connection with uncertainty treatment because it starts from the prior distribution with incomplete knowledge (i.e. epistemic uncertainty), with the aid of the additional observation data, and endeavours to reduce the epistemic uncertainty such as to obtain the posterior distribution of the quantity to be estimated. The Bayesian updating is based on the Bayes' Theorem

$$P\left(\theta|\mathbf{Y}_{exp}\right) = \frac{P_L\left(\mathbf{Y}_{exp}|\theta\right)P(\theta)}{P\left(\mathbf{Y}_{exp}\right)} \tag{10}$$

with the key elements described as follows.

- $P(\theta)$ is the prior distribution of the calibrating quantity θ. Note that, the calibrating quantity θ is not necessarily the model parameter p itself. In most cases of stochastic model updating, θ is the distribution coefficient of the parameters, e.g., mean, variance, correlation coefficient, etc.
- $P\left(\theta|\mathbf{Y}_{exp}\right)$ is the posterior distribution of θ. It is essentially a conditional probability distribution with additional knowledge extracted from the experimental measurements, $\mathbf{Y}_{exp}$.
- $P_L\left(\mathbf{Y}_{exp}|\theta\right)$ is the likelihood function, i.e., the probability of the existing measurements, $\mathbf{Y}_{exp}$, conditional to an instance of the calibrating quantity θ;
- $P(\mathbf{Y}_{exp})$ is the evidence, also known as the "normalisation factor", guaranteeing the integration of the posterior distribution $P\left(\theta|\mathbf{Y}_{exp}\right)$ equal to one.

The complete implementation of Bayesian updating relies on two key aspects: 1) the definition of the likelihood function and 2) the procedure to update the posterior distribution with the MCMC algorithm.

4.3.1. Definition of the likelihood function

The likelihood function, $P_L\left(\mathbf{Y}_{exp}|\theta\right)$, is important because not only is it the criterion of sample selection for each Markov chain within the MCMC algorithm (explained in the following subsection), but also, more importantly, it contains information on both the existing measurement and the parameters to be calibrated. The different definitions of the likelihood with varying levels of uncertainty information involved determine whether the overall model updating is a stochastic or a deterministic procedure. In the presence of multiple and independent measurements, supposing the number is n_{exp}, the complete likelihood of such multiple measurements is calculated as

$$P_L\left(\mathbf{Y}_{exp}|\theta\right) = \prod_{k=1}^{n_{exp}} P(\mathrm{y}_k|\theta) \tag{11}$$

Eq. (11) requires the explicit distribution of each measurement, $P(\mathrm{y}_k|\theta)$, whose precise estimation requires a large number of random samples, i.e. a large number of model evaluations. What's more, this random sampling process for each $P(\mathrm{y}_k|\theta)$ needs to be repeated for n_{exp} times, because of the involvement of multiple sets of experimental data. To reduce the calculation burden, it is important to propose an approximate likelihood, as long as it still contains two aspects of information: 1) the discrepancy between measurements and simulation; and 2) the parameters to be calibrated. Ref. [26] proposes the approximate likelihood function embedding the difference between the model simulations and the experimental measurements:

$$P_L\left(\mathbf{Y}_{exp}|\theta\right) \propto \exp\left\{-\frac{d\left(\mathbf{Y}_{exp}, \mathbf{Y}_{sim}\right)^2}{\sigma^2}\right\} \tag{12}$$

where $d(\mathbf{Y}_{exp}, \mathbf{Y}_{sim})$ is the distance-based UQ metrics defined in Sec. 4.2. No longer requiring the explicit PDFs, Eq. (12) clearly has much reduced calculation cost. More importantly, it provides a convenient connection between the Bayesian updating framework and the UQ metrics. No matter the Euclidian, Mahalanobis, or Bhattacharyya distance is employed as UQ metric, there is no need to change the uniform Bayesian updating framework, nor the following MCMC algorithm. However, the parameter calibration result can be significantly different, when different distance-based UQ metrics are employed, as it can be demonstrated in the following tutorial example.

The coefficient σ in Eq. (12) is the pre-defined standard deviation of the likelihood function, which is used to control the centralisation degree of the posterior distribution. The practical effect of σ in the MCMC sampling process is that it determines the "searching width" of each iteration. In most cases, a small σ is desired because it results in a peaked posterior distribution which is more likely to indicate the explicit estimation result. But a too small a σ could lead to searching over a very narrow space and the whole MCMC procedure would require more iterations and possibly a trap in a local region. A practical suggested range of σ is given as $[10^{-3}, 10^{-1}]$. It is suggested to adjust this according to the mean of the distance-based metrics.

4.3.2. *Updating of the posterior distribution with MCMC algorithm*

Recall the Bayesian Theorem in Eq. (10), in order to obtain the posterior distribution, the normalisation factor $P(\mathbf{X}_{exp})$ should be evaluated, using the following theoretical definition:

$$P(\mathbf{X}_{exp}) = \int P_L(\mathbf{Y}_{exp}|\theta)P(\theta)\mathrm{d}\theta \tag{13}$$

However, this is prohibitive in practical applications, since it requires the direct integration on the explicit distribution of the likelihood, which is generally unavailable. The normalisation factor is therefore one of the difficulties of Bayesian updating, especially for high-dimensional distribution with complex distribution format. As an alternative, the well-known MCMC algorithm is used here to avoid direct integration of Eq. (13). The most common version of MCMC in model updating is the so-called Transitional MCMC (TMCMC) proposed in Ref. [34]. Not directly targeting the final posterior distribution, it suggests an iterative sampling procedure taking a series of intermediate PDFs as target in each of the iteration steps, where the PDF in the final step converges to the posterior PDF. The principle of TMCMC algorithm is expressed as

$$P^{(j)}(\theta) = P_L(Y_{\mathrm{exp}}|\theta)^{\beta_j}P(\theta) \tag{14}$$

where β_j is the weighting power falling within the range [0, 1], and monotonically increasing following the iterations. In the 1st iteration step when $\beta_1 = 0$, $P^{(1)}(\theta)$ is actually the prior distribution. Supposing totally n iterations are executed before convergency, when $\beta_n = 1$ in the last step, $P^{(n)}(\theta)$ equals to the multiplication between the likelihood and the prior distribution, which results in the final posterior distribution. Such treatment is especially useful when the target posterior distribution is multimodal or highly-peaked, and causes the target distribution to be reached by single direct MCMC sampling.

A sketch flow of the TMCMC logic is illustrated in Fig. 6. The TMCMC algorithm consists of single MCMC in each iteration step, to generate Markov chains starting from each existing samples derived from the intermediate PDF in the previous iteration. The candidate chains (i.e., samples) are accepted or rejected according to the Metropolis-Hastings (MH) criterion.

A simple summary of the MH algorithm is given as follows. It employs a "proposal distribution" to generate the candidate samples θ^*. Considering in the (j)-th iteration, the proposal distribution $q\left(\theta^*|\theta^{(j)}\right)$ is defined as a joint normal distribution taking the previous

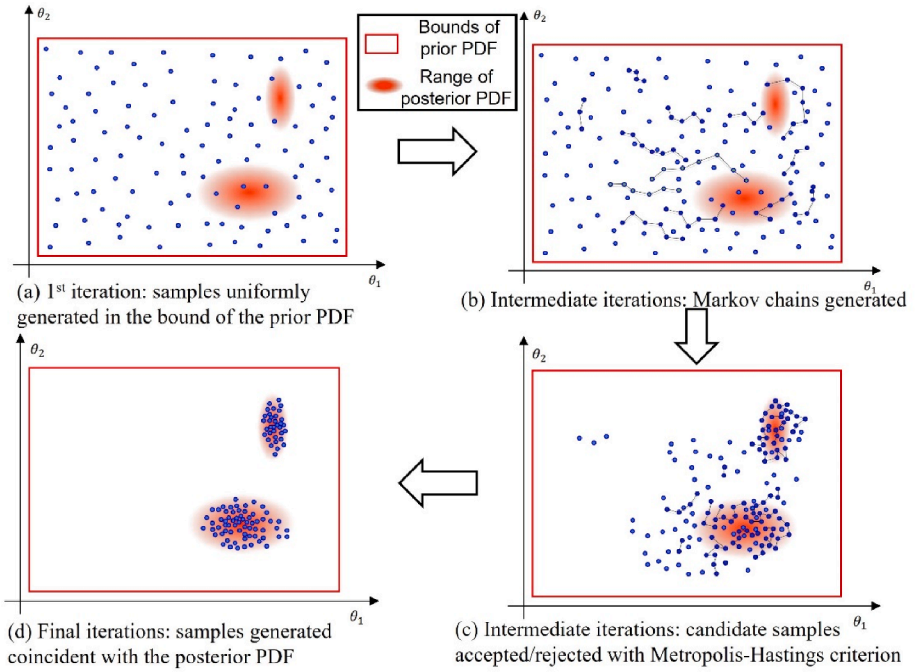


Fig. 6. Flow of the TMCMC process with intermediate PDFs and MH criterion.

sample $\theta^{(j)}$ as the mean value:

$$\theta^* \sim q\left(\theta^* \middle| \theta^{(j)}\right) = N\left(\theta^{(j)}, \Sigma\right) \tag{15}$$

where the covariance matrix Σ controls the searching width of the proposal distribution. Σ is adaptively determined according to the total number of samples N_{sam} and their mean $\overline{\theta}$ in this iteration. An empirical equation to determine the Σ is given by the original paper [34]

$$\Sigma = \gamma^2 \sum_{i=1}^{N_{sam}} \widehat{w}(\theta_i) \bullet \left[\{\theta_i - \overline{\theta}\} \times \{\theta_i - \overline{\theta}\}^T\right] \tag{16}$$

Eq. (16) involves a weighting function $\widehat{w}(\theta_i)$ of each sample θ_i and a scaling coefficient γ whose optimal value is determined to be 0.2 in the original paper. Without going into the very detail, calculation of $\widehat{w}(\theta_i)$ is omitted here, the readers are suggested to refer to Ref. [34].

After the candidate sample θ^* is obtained from Eq. (15), the main task of MH algorithm is to determine whether to accept θ^* or reject it. The criteria α and u are defined using the likelihood $P(\mathbf{Y}_{exp}|\theta)$ and standard uniform distribution, respectively

$$\begin{aligned} \alpha &= \min\left[\frac{P\left(\mathbf{Y}_{exp}\middle|\theta^*\right)}{P\left(\mathbf{Y}_{exp}\middle|\theta^{(j)}\right)}, 1\right] \\ u &\sim \ Uniform(0,1) \end{aligned} \tag{17}$$

The candidate sample θ^* is accepted/rejected by the following logic

$$\theta^{(j+1)} = \begin{cases} \theta^*, \alpha \geq u \\ \theta^{(j)}, \alpha < u \end{cases} \tag{18}$$

A simple understanding of the above logic is that, when the likelihood of the candidate sample is larger than that of the current sample, the candidate sample is accepted; when its likelihood is smaller than the current sample, however, it is not immediately rejected. An additional chance is given to the candidate sample if the ratio between the likelihoods of the candidate sample and the current sample is larger than the random value u. Such a logic ensures that the MCMC sampling converges towards the target region and simultaneously does not overlook any possible regions in the whole searching space.

After the MH selection is implemented on all N_{sam} candidate samples, the (j)-th iteration is finished and N_{sam} new samples are transmitted to the next iteration. The entire TMCMC process requires each iteration step to first calculation the increment of the weighting power $\Delta\beta_j$. The optimal way to determine $\Delta\beta_j$ is to keep the Coefficient of Variance (CoV) of the data set $P\left(D|\theta_i^{(j)}\right)^{\Delta\beta_j}$ as close as possible to 100%. If there are N_{sam} number of samples in the (j)-th iteration $(\theta_1^{(j)}, \theta_2^{(j)}, \cdots, \theta_{N_{sam}}^{(j)})$, then the corresponding CoV of the likelihood data set is

$$COV = \frac{\sigma\left\{P(D|\theta_i)^{\Delta\beta_j}\right\}}{\mu\left\{P(D|\theta_i)^{\Delta\beta_j}\right\}} = 100\% \tag{19}$$

The optimal value of $\Delta\beta_j$ is obtained as the root of the following equation

$$f\left(\Delta\beta_j\right) = \sigma\left\{\exp\left(P(D|\theta_i) \bullet \Delta\beta_j\right)\right\} - \mu\left\{\exp\left(P(D|\theta_i) \bullet \Delta\beta_j\right)\right\} \tag{20}$$

The iterative process terminates when $\beta_{j+1} = \beta_j + \Delta\beta_j \geq 1$. The TMCMC pseudocode describing the main steps and iterative logic is presented in **Algorithm 1**.

Algorithm 1: TMCMC Iterative Sampling

Set iteration index $j = 1$ and $\beta_j = 0$
Take N_{sam} initial samples from prior distribution: $\boldsymbol{\theta} = \left[\theta_1^{(1)}, \theta_2^{(1)}, \cdots, \theta_{N_{sam}}^{(1)}\right] \sim P(\theta)$
while $\beta_j < 1$ **do**
  Set $j = j + 1$
  Compute optimal $\Delta\beta_j$ based on the N_{sam} samples using Eq. (20)
  Compute P^j: $P^j \propto P_L(\mathbf{Y}_{exp}|\theta)^{\beta_j} \bullet P(\theta)$
  Resample with weighted probability $\theta_j \sim \widehat{w}(\theta_i)$
  Generate single Markov chain from each sample
    for $i = 1 : N_{sam}$
      Sample candidate from proposal distribution: $\theta^* \sim q(\theta_i^*|\theta_i)$
      Accept or reject θ_i^* following the Metropolis-Hastings algorithm
    end for
end while

5. Tutorial example of the NASA UQ Challenge 2014

The NASA UQ Challenge 2014 [13] is an interdisciplinary competition problem launched by the NASA Langley Research Center. The challenge problem is developed based on an aerodynamics control experiment of a laboratory-scale aircraft model. Although deriving from a discipline-specific application, this problem is extracted as a discipline-independent system with the core aerodynamic control process structured as a black-box model. This problem is hence investigated by different parties with multiple results released and compared, e.g., in Refs. [15,26,35]. This section aims at demonstrating the key techniques for forward uncertainty propagation and inverse parameter calibration by solving the relevant tasks in the NASA UQ Challenge problem. Readers are encouraged to reproduce the results following this tutorial example and compare it with other published results in the literature.

5.1. Problem description

The complete NASA UQ Challenge 2014 contains multiple subproblems of uncertainty characterisation, sensitivity analysis, reliability analysis, robust design, etc. This example is not aiming to solve the whole problem, but focusing on two subtasks: 1) to propagate the P-boxes of the input parameters to the P-box of the output using the double-loop propagation technique; 2) to calibrate the uncertainty model of the input parameters based on the available observations using the Bayesian MCMC updating technique.

Fig. 7 illustrates the structure of the NASA UQ Challenge 2014, which involves a black-box model $y = h(p_i)$ with five input parameters p_{1-5} and one output feature y. The uncertainty characterisation of the parameters is presented in Table 1. Considering the parameter categorisation (recall Fig. 2), p_2 is an unknown-but-fixed constant in the range [0, 1], hence it belongs to Category II. p_3 follows the uniform distribution with fully determined distribution properties, hence it belongs to Category III. $p_{1,4,5}$ are random variables following certain distributions but their distribution coefficients (mean, variance, correlations, etc.) are not fully determined, hence they belong to Category IV. This example contains the following two tasks:

1) **Uncertainty propagation**: Perform the forward procedure to investigate how the uncertainty properties are propagated from the multiple categories' parameters to the features. This involves the propagation of both the aleatory uncertainty and epistemic uncertainty, simultaneously.
2) **Parameter calibration**: Perform the inverse procedure based on a set of given observation samples to reduce the epistemic uncertainty of the parameters in turn. These experimental samples, provided by the Challenge host, are generated from the "true" distributions of the input parameters. The task here is to find the true values of the distribution coefficients from the given range in the last column of Table 1. Since the aleatory uncertainty is irreducible, only the parameters with epistemic uncertainty, i.e. $p_{1,2,4,5}$, require to be calibrated in this subtask.

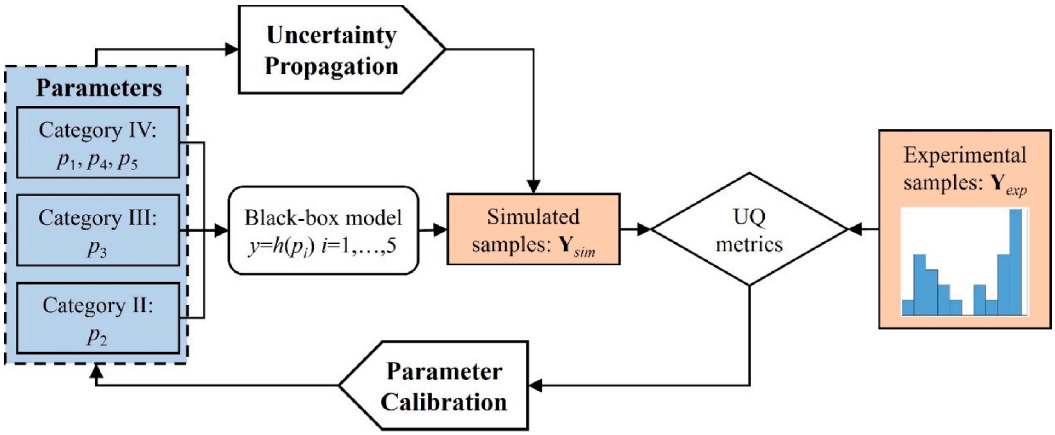


Fig. 7. Structure and relative tasks of the NASA UQ Challenge 2014.

Table 1
Uncertainty characteristics of the input parameters.

Category	Parameter	Distribution	Uncertainty characteristics
IV	p_1	Beta	$\mu_1 \in [0.6, 0.8]$ $\sigma_1^2 \in [0.02, 0.04]$
	p_4, p_5	Gaussian	$\mu_i \in [-5.0, 5.0]$ $\sigma_i^2 \in [0.0025, 4.0]$ $\rho \in [-1.0, 1.0],\ i = 4,\ 5$
III	p_3	Uniform	$a_3 = 0, b_3 = 1.0$
II	p_2	Constant	$p_2 \in [0.0, 1.0]$

5.2. Forward uncertainty propagation

According to the double-loop procedure in Sec. 4.1, the first step is to determine the P-boxes of the input parameters. Clearly parameters from different categories have varying formats of the P-boxes. The P-box of the Category II parameter has the simplest shape appearing as a rectangle, whose left and right edges are determined as the lower and upper bounds of the parameter's interval. This is because the Category II parameters have only epistemic uncertainty. Any possible value within its interval corresponding to an impulse function with the amplitude as 1.0. An infinite number of impulse functions constitute a rectangle with length as 1.0 and width to be the same as the parameter interval. As an example, the P-box of p_2 is illustrated in Fig. 8(b).

The P-box of the Category IV parameter is more complex since both epistemic and aleatory uncertainties are involved. The determination of this type of P-boxes is based on the fact that the shape and position of a single CDF curve are driven by its mean and variance. The mean value controls the horizontal position of the CDF curve: A CDF with a small mean is generally located at the left part of the coordinate plane; the one with a large mean is located at the far-right. The variance controls the degree of tilt: A CDF with a smaller variance is generally "steeper"; the one with a larger variance is generally flatter. Based on this principle, it is easy to conclude that the P-box is always enveloped by the four extreme CDF curves:

- The CDF with minimum mean and minimum variance $P(\mu_{min}, \sigma^2_{min})$;
- The CDF with maximum mean and minimum variance $P(\mu_{max}, \sigma^2_{min})$;
- The CDF with minimum mean and maximum variance $P(\mu_{min}, \sigma^2_{max})$;
- The CDF with maximum mean and maximum variance $P(\mu_{max}, \sigma^2_{max})$.

For example, the P-box of p_4 and p_5 is obtained by comparing the CDFs of the four Gaussian distributions $N(-5.0, 0.0025)$, $N(-5.0, 4.0)$, $N(5.0, 0.0025)$, and $N(5.0, 4.0)$. The whole region enveloped by the four CDFs is illustrated in Fig. 8(c). For the determination of the P-box of p_1, an additional process is required to transfer the mean and variance to the shape coefficients (α and β) of the Beta distribution, following the equations

$$\alpha = \mu\left(\frac{\mu(1-\mu)}{\sigma^2} - 1\right) \tag{21}$$

$$\beta = (1-\mu)\left(\frac{\mu(1-\mu)}{\sigma^2} - 1\right) \tag{22}$$

The four edge CDFs with maximum/minimum mean and variance, presented in shape coefficients, are $Beta(6.6, 4.4)$, $Beta(3.0, 2.0)$, $Beta(5.6, 1.4)$, $Beta(2.4, 0.6)$. The P-box of p_1 is illustrated Fig. 8(a).

As long as the P-boxes of the input parameters are obtained, the double-loop propagation is implemented to estimate the P-box of the output. In the outer-loop, the number of Monte Carlo samples of the probability value α is set to $N_{sample} = 1000$. Different probability samples result to the constant interval of p_2 to be [0, 1] as shown in Fig. 8(b). However, for p_1, p_4, and p_5, the different sampled probability values along the vertical axis in Fig. 8(a, c) truncate varying intervals in the horizontal axis of the input parameters. Such intervals of p_1, p_2, p_4, and p_5 are taken as constraint to solve the inner-loop optimisation as given in Eq. (1).

The fmincon.m optimiser in MATLAB© is employed to find the 1000 pairs of minimum and maximum samples. Their histograms and fitted PDFs are illustrated in Fig. 9(a). Clearly, the distribution of the output feature y is no longer a standard distribution. In contrast, it appears as an implicit and multimodal function because of the complicated black-box model between the inputs and outputs. The actual shape the output P-box is illustrated in Fig. 9(b). The implicit distribution of the minimum/maximum samples result in the so-called "distribution-free" P-box, because the boundary CDF curve is not that of a standard distribution, e.g. Gaussian or

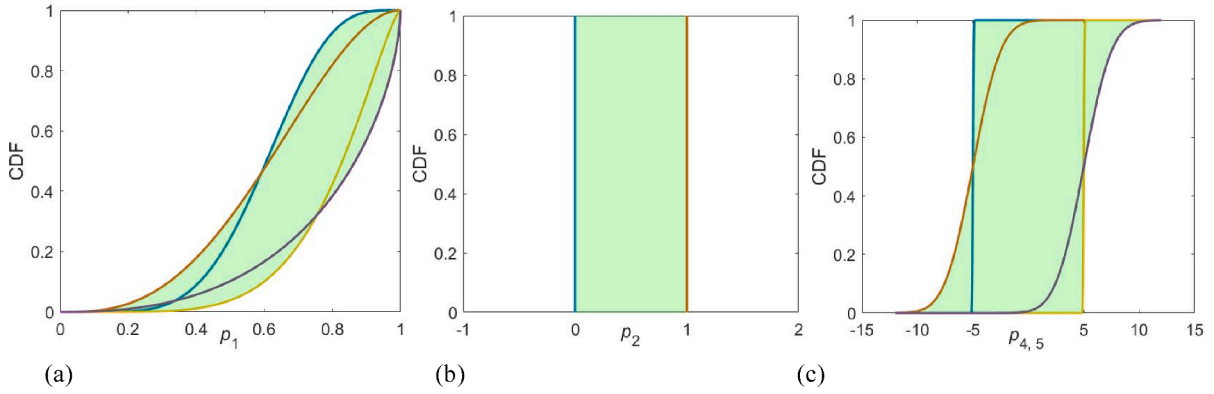
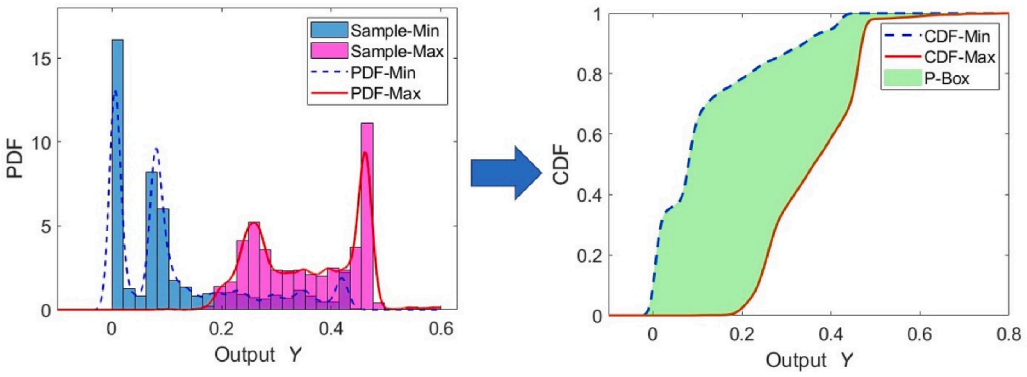


Fig. 8. The P-boxes of the four input parameters: (a) P-box of p_1 with uncertain Beta distributions, (b) P-box of p_2 within a given interval, (c) P-box of p_4 and p_5 with uncertain Gaussian distributions.



(a) Histograms and fitted PDF curves (b) P-box enveloped by boundary CDF curves

Fig. 9. The distributions of the minimum and maximum samples and the P-box of the output feature.

Beta. This demonstrates the double-loop propagation is capable of drawing the whole figure of the complete P-box no matter its boundary CDFs are explicit or implicit. This is a significant advantage over most of the simplified propagation techniques where only the probabilistic moments, e.g., mean, or variance, can be estimated.

For calculation efficiency, the Monte-Carlo-Optimisation double-loop approach is already much more efficient than the exclusive Monte Carlo double-loop approach. When merely Monte Carlo sampling is employed in both the outer- and inner-loop, assuming 1000 samples are required to estimate each CDF distribution and totally 1000 CDFs are required to estimate the P-box region, a total 10^6 model evaluations are required. The Monte-Carlo-Optimisation double-loop approach is further beneficial from the independence among each optimisation. Hence, parallel calculation is employed to dramatically reduce the calculation time. The above double-loop P-box propagation, with $N_{sample} = 1000$, takes only 428 s on a standard desktop with Intel(R) Core(TM) i7 CPU (8 cores for parallel calculation).

5.3. Parameter calibration

In this section the process of parameter calibration is described. It includes a study of effects of using different distance-based likelihood functions, consideration of non-unique parameter estimates and the P-box propagation of such estimates to reveal the reduction of epistemic uncertainty brought about by the calibration process.

5.3.1. Cross-comparison of results from different statistical distance-based likelihoods

The problem host provides 50 observation samples of the output y, termed as Y_{exp}, generated from the unknown true values within the ranges in the last column of Table 1. The task here is, starting from the available Y_{exp}, to inversely calibrate the distribution coefficients of the input parameters, so as to turn the predicted output Y_{sim} as close as possible to the observation samples Y_{exp}. The calibration metrics based on various statistical distances have significant influence on the calibrated results, since different distances (as defined in Sec. 4.2) involves different levels of uncertainty information from the simulation and observation samples.

This example has totally eight coefficients, $\theta_1 - \theta_8$, to be calibrated. Their physical representations, prescribed ranges, and true values (afterwards released by the problem host) are listed in Table 2. It is shown that some true values appear at, or very close to, the edge of the prescribed ranges, e.g. $\theta_3 = 1.0$ for its range.

$[0.0, 1.0]$, $\theta_5 = 0.04$ for its range $[0.0025, 4.0]$. Some prescribed ranges are given to be large and with general bounds, e.g. θ_8 is given the range $[-1.0, 1.0]$, which covers all possible values as a correlation coefficient. Such settings lead to a very challenging calibration

Table 2
Prescribed features and updating results of the calibrating coefficients.

Parameter	Description	Calibrating Coefficient	Original interval	True value	Calibrated values θ		
					Euclidian distance	Mahalanobis distance	Bhattacharyya distance
p_1	Mean	θ_1	$[0.6, 0.8]$	0.6364	0.709	0.700	0.649
	Variance	θ_2	$[0.02, 0.04]$	0.0356	0.035	0.037	0.039
p_2		θ_3	$[0.0, 1.0]$	1.0	0.564	0.541	0.997
p_4	Mean	θ_4	$[-5.0, 5.0]$	4.0	−2.617	0.847	4.096
	Variance	θ_5	$[0.0025, 4.0]$	0.04	1.640	2.543	1.404
p_5	Mean	θ_6	$[-5.0, 5.0]$	−1.5	0.498	3.633	−4.700
	Variance	θ_7	$[0.0025, 4.0]$	0.36	1.633	2.187	0.186
p_4, p_5	Correlation Coefficient	θ_8	$[-1.0, 1.0]$	0.5	0.103	−0.586	0.634

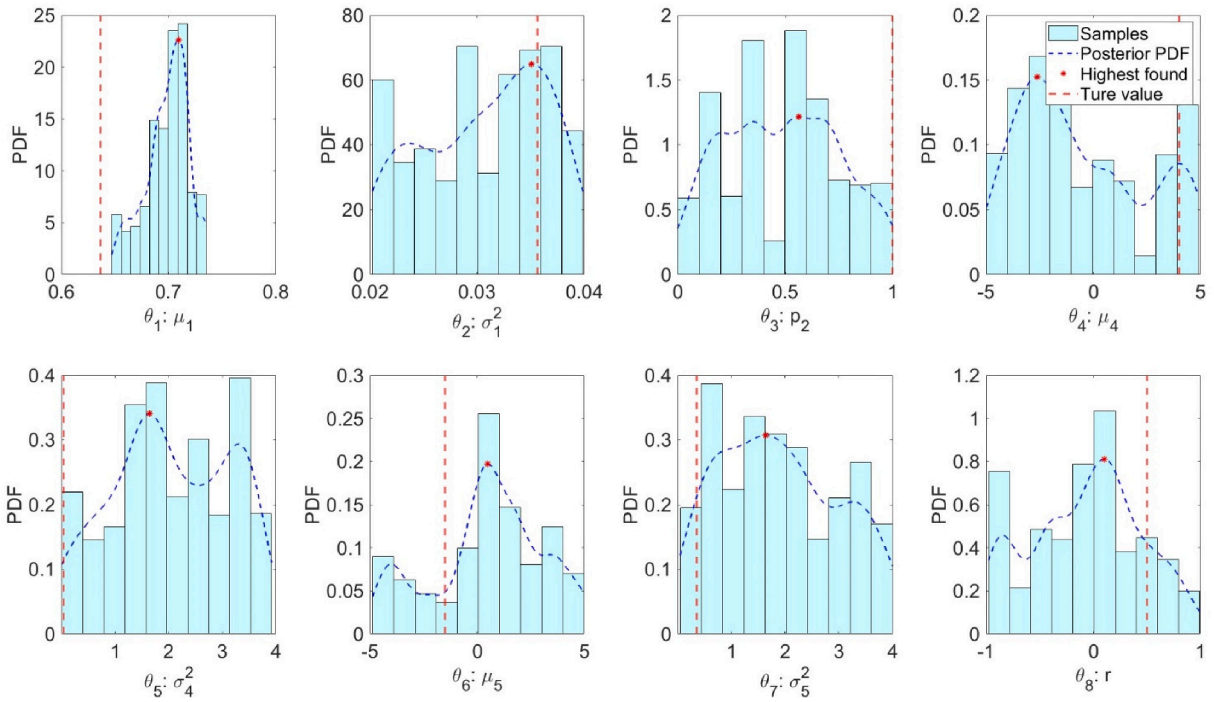


Fig. 10. Updated posterior distributions of $\theta_1 - \theta_8$ using the Euclidean distance.

problem. This example employs all three distances, i.e. the Euclidian distance, Mahalanobis distance, and Bhattacharyya distance, to define the likelihood function in the Bayesian updating framework, and embeds them into the same framework with the TMCMC algorithm. In each TMCMC iteration step, the number of samples is set as 1000. The posterior PDFs are fitted based on these 1000 samples by the Kernel Density Estimation technique [36]. The histograms and fitted PDFs calibrated by the different statistical

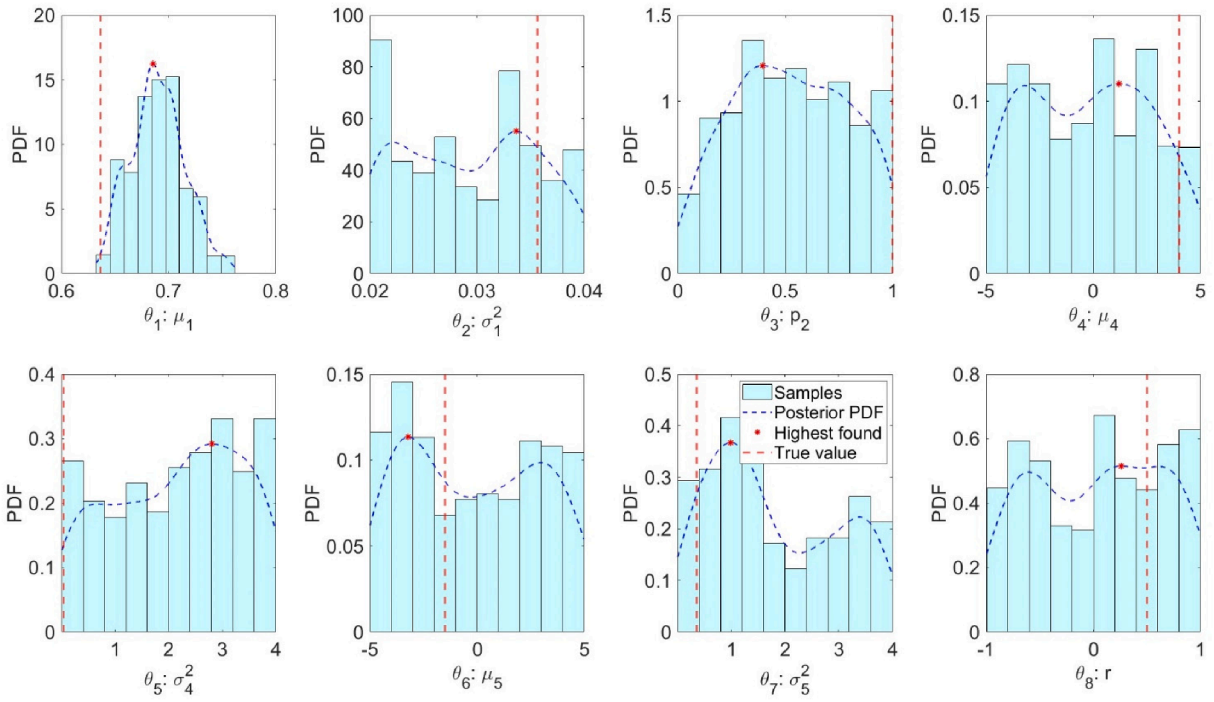


Fig. 11. Updated posterior distributions of $\theta_1 - \theta_8$ using the Mahalanobis distance.

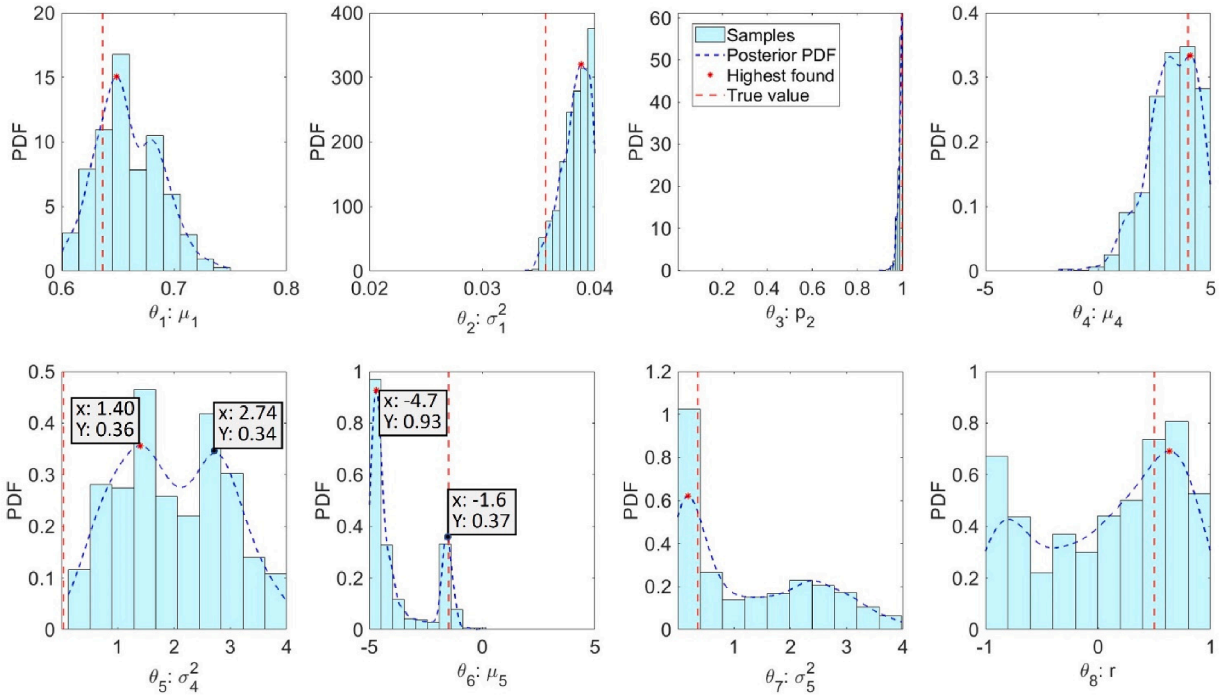


Fig. 12. Updated posterior distributions of $\theta_1 - \theta_8$ using the Bhattacharyya distance.

distances are illustrated in Figs. 10-12, respectively. The updated values of $\theta_1 - \theta_8$ are obtained by taking the maximum value of PDF curves, and are listed in Table 2.

Fig. 10 illustrates the updating results from the Euclidian distance, where θ_2, i.e. the standard deviation of p_1, is the only quantity with a good updating result. θ_3 and θ_7 have flat-shaped distributions, which are still nearly uniform, implying they are not obviously calibrated from their prior distribution. The remaining coefficients, $\theta_{1,4,5,6,8}$ have peaks far from the position of their true values, showing the updating process does not converge to the correct answer. Such results are expected because the Euclidian distance-based likelihood function only involves the distance between means. No dispersion information is considered in such likelihood functions, and hence it is natural that the variances and correlations of the parameters cannot be updated.

The updating results employing the Mahalanobis distance are illustrated in Fig. 11. The general updating effect is even worse than the results from the Euclidian results, since most posterior distributions, e.g. those of $\theta_{2-6,8}$, are still nearly uniform. The maximums of such posterior distributions are not prominent enough to be identified as the updated values. Only the posterior distributions of θ_1 and θ_7 have relatively obvious peaks, although the locations of the peaks are not close to the true values. Explanation of such unsatisfactory results is that, although the Mahalanobis distance-based likelihood function involves the covariance information, such information is not sensitive to the overlap between the simulation and observation samples. Even if the TMCMC algorithm is capable of reducing the Mahalanobis distance, it is not necessarily able to search the true values of the distribution coefficients of the input parameters.

As illustrated in Fig. 12, the Bayesian updating with the Bhattacharyya distance-based likelihood function obtains very different posterior distributions, compared with the ones in Figs. 10 and 11. The most obvious and extreme example is the posterior PDF of θ_3, which is extremely centralised to the edge of its range, i.e. the true value to be reached. In contrast, the posterior PDF of θ_3 in Figs. 10 and 11 do not achieve a good result since they are still nearly uniform without any tendency towards the target true value. The similar effect is observed for θ_2, θ_4, and θ_7, which are more clearly peaked than the ones in Figs. 10 and 11. It is shown by the comparison of Figs. 10-12 and in Table 2 that the updating precision of the result with the Bhattacharyya distance is clearly higher than that of the Euclidian and Mahalanobis distances.

5.3.2. *Non-uniqueness solutions and output distributions*

However, special attention should be paid on θ_5 and θ_6 in Fig. 12, whose posterior PDFs have two peaks. One of the peaks of θ_6 is quite close to its true value, while the other one is far from it but with a higher probability. If only one maximum point is selected from the PDF curve, clearly for θ_6 the value -4.7 will be selected as the calibrated result as shown in Table 2. However, another peak with -1.6 should not be ignored since it is another potential local solution. And actually, it is the true value. The possible local solutions of $\theta_5 = 1.4 or 2.74$, $\theta_6 = -4.7 or -1.6$ are further investigated as shown in Table 3. Four configurations of all possible local solutions of θ_5 and θ_6 are listed in Table 3. Each configuration along with the calibrated values of other coefficients, see θ_{1-4} and $\theta_{7,8}$ in the last column of Table 2, is employed to generate 1000 samples of the input parameters p_{1-5}, which are substituted into the model $y = h(p_i)$ to obtain the predicted samples Y_{sim}. The Bhattacharyya distances between the predicted samples and the observations samples are calculated and listed in Table 3. It is shown that the distance values of all four configurations are similar, and they are quite different

Table 3
Bhattacharyya distances between the predicted samples Y_{sim} and the observation samples Y_{exp}.

Different configurations Index	θ_5	θ_6	Bhattacharyya distance after calibration	Bhattacharyya distance before calibration
#1	1.4	−4.7	0.692	1.0156
#2	1.4	−1.6	0.671	
#3	2.74	−4.7	0.689	
#4	2.74	−1.6	0.650	

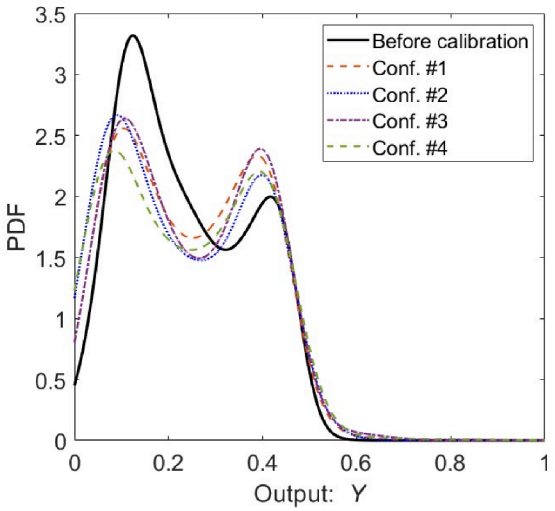


Fig. 13. Calibrated PDFs of the output samples with different configurations of θ_5 and..θ_6

Table 4
Statistical distances between the observation output samples and the predicted output samples before and after the Bayesian updating with the TMCMC algorithm.

Likelihood function with	Before calibration	After calibration
Euclidian distance	0.0116	0.0025
Mahalanobis distance	0.0847	0.0126
Bhattacharyya distance	1.0156	0.6921

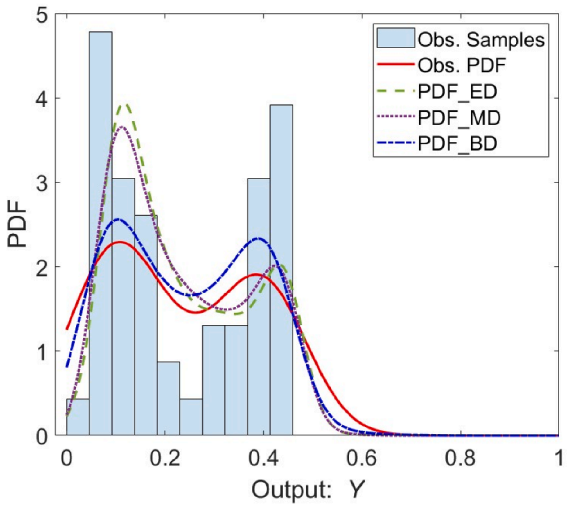


Fig. 14. Comparison of the output PDFs according to different statistic distances.

from the distance value before calibration. This demonstrates that all the local solutions of θ_5 and θ_6 lead to the similar samples of the model output. This is to say, the capability of the TMCMC algorithm is demonstrated because it is capable of minimising the Bhattacharyya distance. The phenomenon that some calibrating quantities cannot converge to their true values is caused by the non-unique nature of the problem itself. A vivid illustration is presented in Fig. 13 where the fitted output PDF curves derived from the four coefficient configurations are plotted together with the initial PDF. It is observed that the four PDFs after calibration are similar, while all of them have apparent discrepancy from the initial PDF.

The following assessment focuses on the model outputs. In practical applications where the true values of the input parameters are unknown, the investigation of the output becomes the more direct, even the sole measure to evaluate the updating effect. Table 4 presents the distances between the predicted samples Y_{sim} and the observation samples Y_{exp}, regarding different likelihood functions constructed by the Euclidian, Mahalanobis, and Bhattacharyya distances. This table mainly demonstrates the capability of the TMCMC algorithm. Regardless which statistic distance is employed, the TMCMC algorithm is capable of reducing the distance values within the Bayesian updating framework. Besides the comparison of distance values, the actual distributions, i.e. PDF curves, of the model output y, along with the histogram of the 50 provided observations, are illustrated in Fig. 14. The observation histogram and the solid red PDF curve serve as the target of the updating process. It is observed that the PDF calibrated by the Bhattacharyya distance (labelled as PDF_BD) is closer to the target than the PDFs calibrated by other two distances (PDF_ED and PDF_MD). Such result fulfils the expectation that the Bhattacharyya distance is a more relevant metric to measure the discrepancy between two probabilistic distributions.

5.3.3. *Uncertainty reduction in the form of P-box reshaping*

This subsection provides a further explanation of how to generate the final updating result from the posterior distributions, and how to understand the updating results from the perspective of uncertainty reduction and P-box presentation.

For clarity, the posterior distributions of $\theta_1 - \theta_8$ according to the Bhattacharyya distance-based likelihood function in Fig. 12 are normalised and re-plotted in Fig. 15. Note that the posterior distribution curves of $\theta_1 - \theta_8$ are not the real PDFs in the statistical sense. Actually, $\theta_1 - \theta_8$ are not random variables, hence they do not have any probabilistic distributions. The posterior distribution curves here are used to estimate the final updated values/ranges of $\theta_1 - \theta_8$.

To truncate the intervals from the posterior distribution, a so-called "truncation level", $\alpha \in [0, 1]$, is defined. It is an indicator of the degree of epistemic uncertainty reduction from the whole uncertainty space, i.e., how much the original intervals in Table 5 would be reduced. As illustrated in Fig. 15, four truncation levels, $\alpha = 0.5, 0.7, 0.9, 1.0$, are implemented on the normalised posterior distributions, and the obtained intervals are listed in Table 5. The extreme case is, when $\alpha = 1.0$, the truncation outcome is no longer an

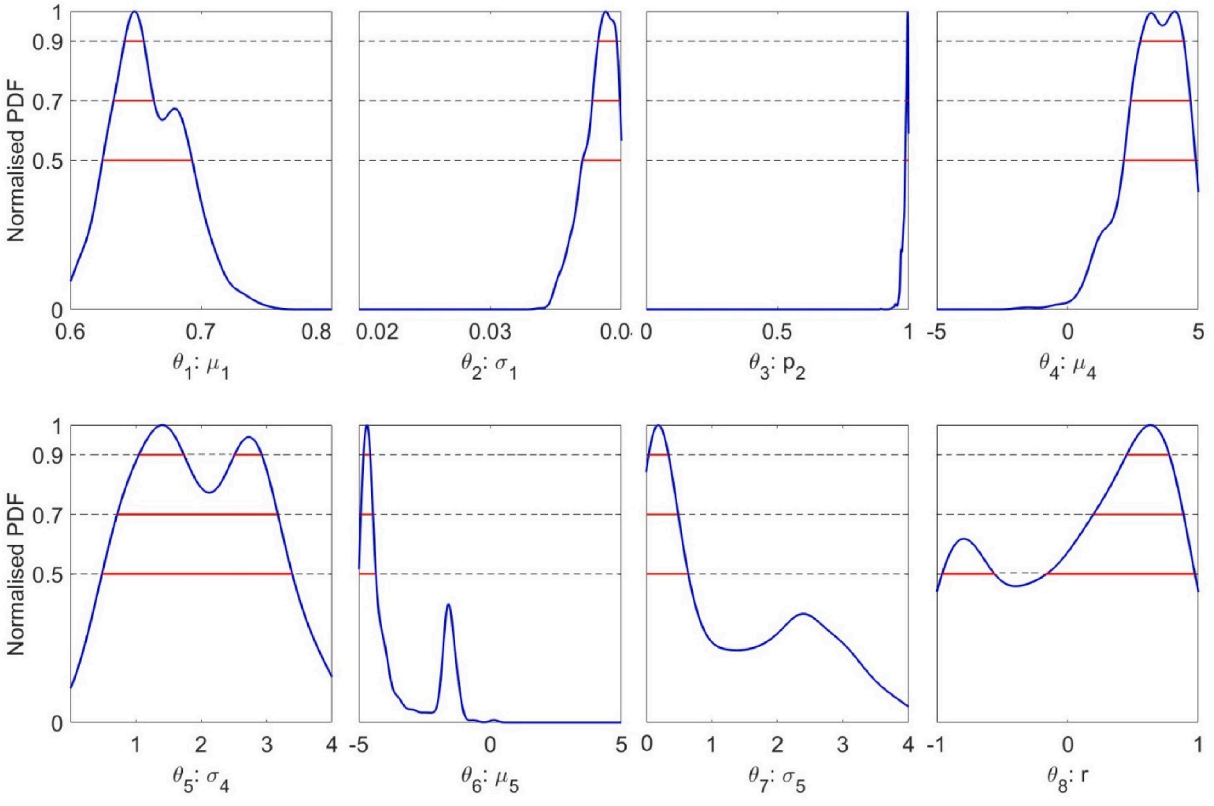


Fig. 15. Normalised posterior distributions employed to estimate ranges according to various truncation levels.

Table 5
Truncated intervals of the calibrating quantities according to various levels of truncation factor.

Calibrating Quantity	Original interval	Truncated intervals after calibration			
		0.5-level	0.7-level	0.9-level	1.0-level
θ_1	[0.6, 0.8]	[0.625, 0.693]	[0.633, 0.664]	[0.641, 0.656]	0.649
θ_2	[0.02, 0.04]	[0.0371, 0.04]	[0.0378, 0.0399]	[0.0382, 0.0396]	0.0388
θ_3	[0.0, 1.0]	[0.988, 1.0]	[0.992, 0.999]	[0.996, 0.998]	0.997
θ_4	[−5.0, 5.0]	[2.15, 4.89]	[2.41, 4.69]	[2.78, 4.44]	4.096
θ_5	[0.0025, 4.0]	[0.49, 3.39]	[0.72, 3.17]	[1.05, 1.74]&[2.51, 2.92]	1.404
θ_6	[−5.0, 5.0]	[−5.0, −4.34]	[−4.92, −4.67]	[−4.82, −4.58]	−4.700
θ_7	[0.0025, 4.0]	[0.0025, 0.643]	[0.0025, 0.487]	[0.0425, 0.339]	0.186
θ_8	[−1.0, 1.0]	[−0.958, −0.560]&[−0.159, 0.972]	[0.197, 0.884]	[0.452, 0.778]	0.634

interval but a constant value located at the peak of the curve. This is the same approach used in Sec. 5.3.1 to estimate the specific values of $\theta_1 - \theta_8$, and this is why the last columns of Table 2 and Table 5 are the same.

Special attention should be paid to the 0.9-level truncated interval of θ_5 and the 0.5-level truncated interval of θ_8. From Fig. 15, it is observed that these intervals have multiple segments, because of the multimodal feature of the posterior distributions. Such a truncation approach is based on the understanding that the posterior distribution of the calibrating quantity in Bayesian updating essentially represents the degree of the trust when this quantity is assigned to different values. Taking the 0.9 truncation level of θ_5 in Fig. 15 as an example, any values of θ_5 with a posterior PDF higher than 0.9 should be reserved, rather than only keeping the highest peak. Note that this should not be confused with the local/global solutions in optimisation, where the only highest peak is regarded as the global solution while any lower peaks are assumed as local solutions and to be disregarded. This is one of the main differences between the Bayesian philosophy and optimisation techniques.

Fig. 16 provides a clear understanding of the stochastic updating effects from the perspective of epistemic uncertainty reduction. The different truncated intervals of $\theta_1 - \theta_8$ in Table 5 result in different P-boxes of the input parameters, which are subsequently propagated by the double-loop approach in Sec. 4.1 to generate the P-boxes of the model output. The 0.0 α-level means that no truncation is implemented, hence the whole original intervals of $\theta_1 - \theta_8$ are employed to propagate the largest P-box containing the most epistemic uncertainty. Since the truncation α-level increases, the output P-box is progressively "pinched" implying the epistemic uncertainty is progressively reduced.

In the ideally perfect case, when the truncation α-level reaches 1.0, the P-box will be finally pinched into a single CDF curve, and such a curve should be exactly the same as the CDF fitted from the experimental data. However, in practical applications such result is mostly impossible, because it requires the numerical model to be a perfect representation of physical systems without any model form errors. It is interesting to be observed that, in Fig. 16, some parts of the experimental CDF have gone beyond the edge of the α-0.9P-box. This phenomenon is normal because, as seen from Fig. 12, the final updated values of θ_2, θ_5, and θ_6 are different from their true values. It is unnecessary to expect the P-box to completely converge towards the experimental CDF. Nevertheless, the fact that the P-boxes with high truncation α-levels can still envelop the experimental CDF curve demonstrates the updated model prediction is coincident

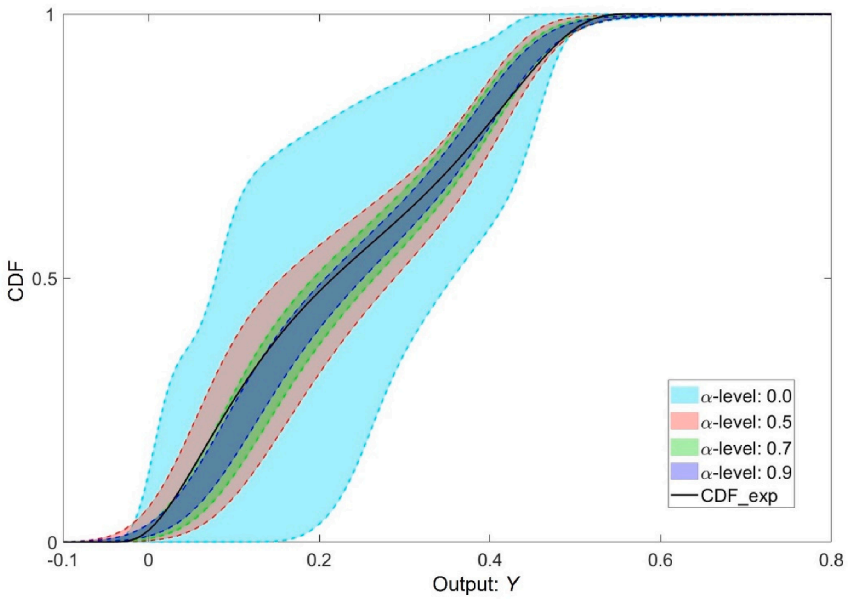


Fig. 16. P-boxes of the output when various truncation levels are implemented to the posterior distributions.

with the experimental observations given the tolerance of a certain amount of epistemic uncertainty, which is actually a measure of the *robustness*, a key criterion of stochastic updating effect.

6. Practical example of an uncertain benchmark testbed

6.1. Design, experiment, and parameterisation of the model

An uncertain benchmark testbed is dedicatedly designed for stochastic model updating with the emphasis on involving multisource uncertainties. It is a testbed with not a single structure, but a series of nominally identical lab-scale airplane models whose geometry characteristic is intentionally changing. Such configurations ensure the involvement of multisource uncertainties, since the experimental observations would contain not only the experimental uncertainty but also the controllable uncertainty from the structures themselves. It is a fact that measurement errors may arise due to electronic noise. Also, inexperienced operators may process signals incorrectly, resulting in seemingly noisy FRFs and bias. However, it is well recognized that uncertainty associated with well carried-out experiments is tiny in comparison to modelling uncertainty. In this paper we address the problem of model updating on the assumption that the effects of operational measurement error are negligibly small in comparison to the modelling error and the artificially controlled structural uncertainty.

The testbed is constructed based on the prototype airplane model with its geometry details illustrated in Fig. 17. Totally 30 airplane models are manufactured with most of the geometric sizes exactly following the specification in Fig. 17, except the half wingspan a and wing tip chord b. Here these two parameters a and b are designed as random variables following a pre-defined distribution, with their nominal mean values equal to 300 mm and 25 mm, respectively. The actual 30 wing pieces with varying sizes are illustrated in Fig. 18 (a). The vibration test is performed on each structure to measure the airplane's natural frequencies and mode shapes. During the test, the airplane model is placed in a free suspension condition, and with both non-contact excitation (loudspeaker) and laser vibrometer as shown in Fig. 18(b). Such free-suspension setup and non-contact techniques are adopted to reduce the installation randomness among each 30 experiments. The final obtained measurement dataset (Table A1) would contain not only the experimental uncertainty caused

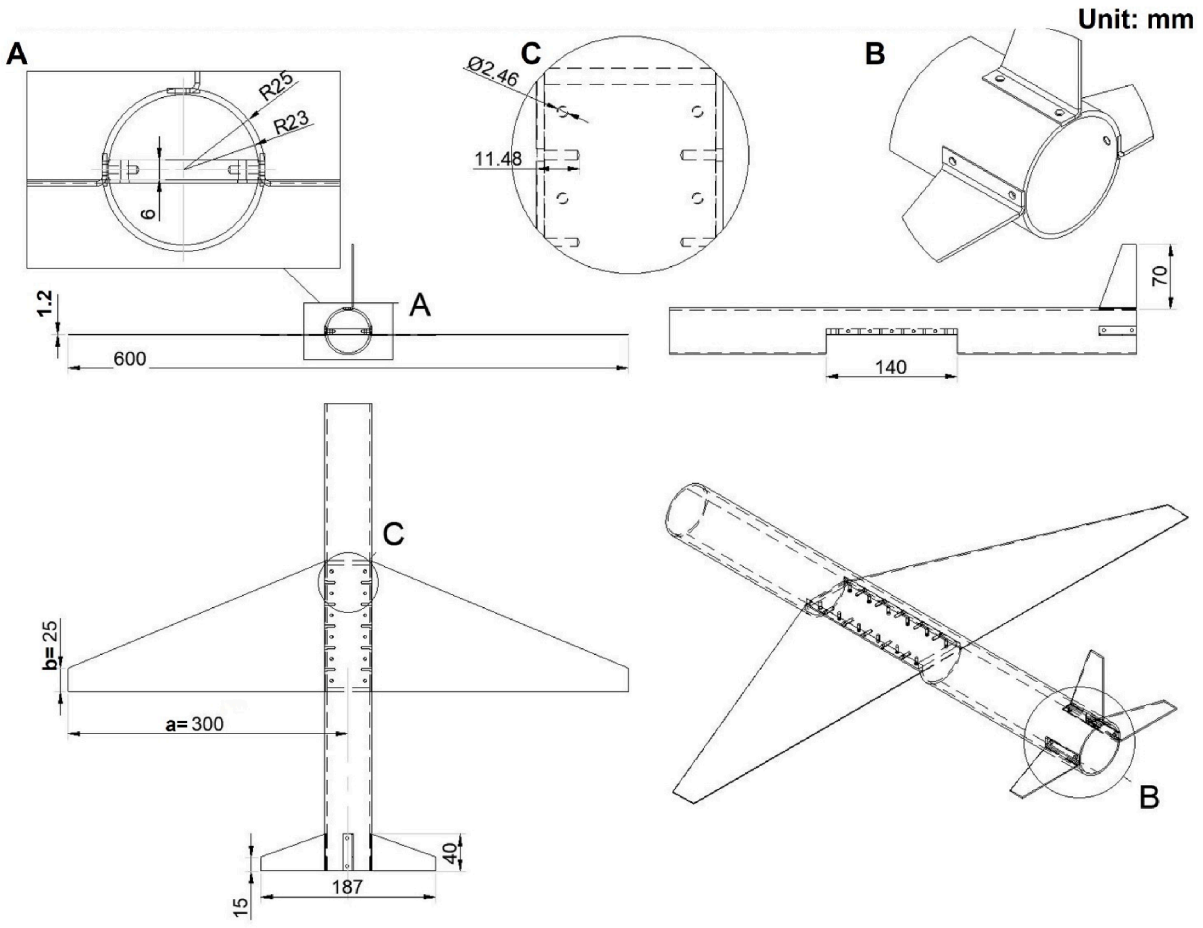


Fig. 17. Geometry details of the benchmark airplane model.

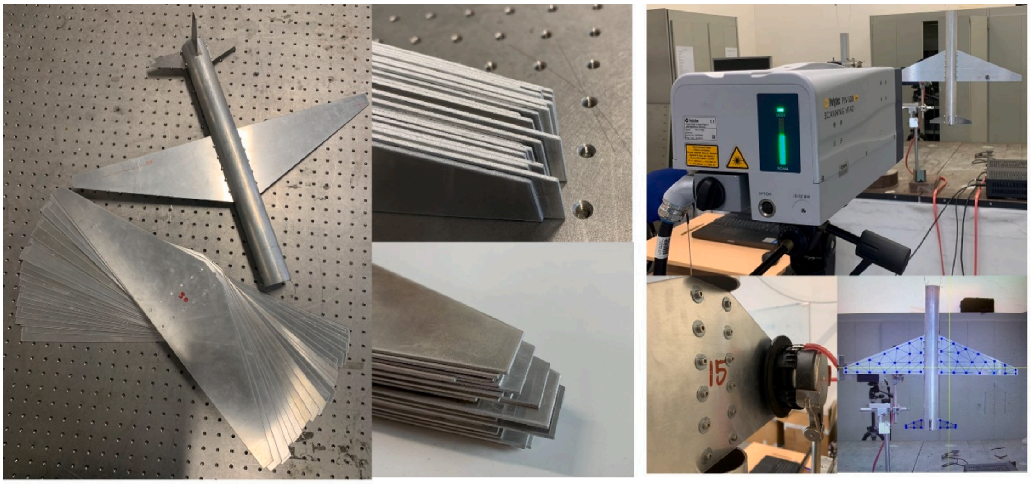


Fig. 18. 30 manufactured airplane models and the vibration test.

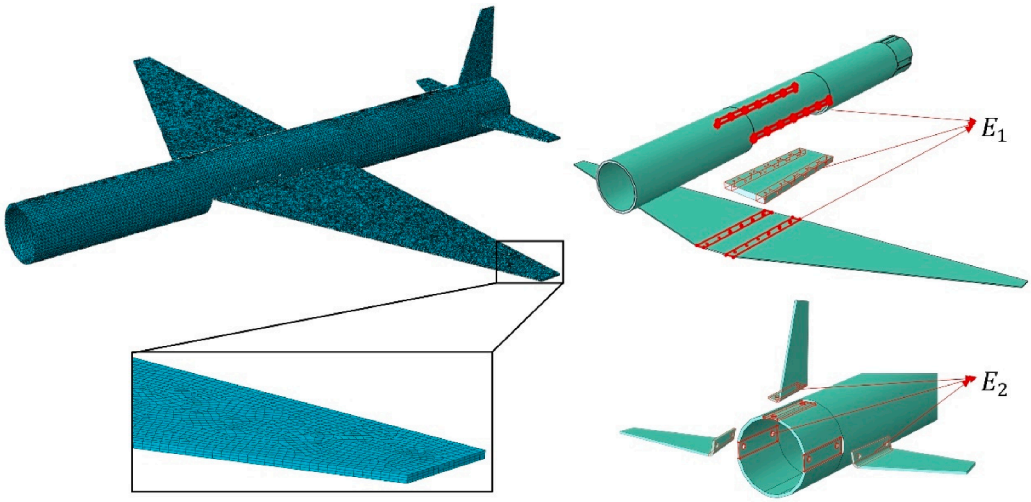


Fig. 19. (a) Initial FE model with hexahedron elements; (b) decomposition of the joint components.

by environmental noise, observation error, sensor tolerance, etc., but also the controllable and "synthetic" uncertainty deriving from the pre-defined distribution of the geometry parameters a and b.

The initial FE model is built with 3D hexahedron elements as shown in Fig. 19(a). For parameterisation, the first choice is the geometry size a and b, because they are prescribed as uncertain parameters. The next concern is the elements of the joint parts among the fuselage, the wing, and the tail wings, as shown in Fig. 19(b), where considerable model form error would exist because of the

Table 6
Parameterisation of the FE model and their uncertainty characteristics.

Category	Parameter	Description	Uncertainty model	Uncertainty characteristics
IV	a	Half wingspan	Joint Gaussian	$\mu_a \in [290, 310](\text{mm})$ $\sigma_a \in [0, 5](\text{mm})$
	b	Wingtip chord		$\mu_b \in [20, 30](\text{mm})$ $\sigma_b \in [0, 5](\text{mm})$ $\rho \in [-1.0, 1.0]$
II	E_1	Young's modulus of fuselage/wing joint	Fixed-but-unknown constant	$E_1 \in [0.5, 0.9](10^{11} Pa)$
	E_2	Young's modulus of fuselage/tail joint		$E_2 \in [0.5, 0.9](10^{11} Pa)$

simplification of the bolted connection. The Young's modulus of joints elements are hence selected parameters to be updated. The prescribed uncertainty characteristics of the above selected parameters are listed in Table 6.

6.2. Calibration results

6.2.1. Preliminary investigation of the measured natural frequencies

The 30 sets of measured natural frequencies are listed in the A ppendix Table A1. They are expected to involve multiple sources of uncertainties deriving from experimental randomness, signal processing, observation system errors, and more dominantly, the pre-designed uncertainty feature of the wing geometry parameters a and b. It is hence important to have a preliminary investigation of the distribution feature of the frequencies. Fig. 20 presents both the marginal distribution of each frequency and the joint 2-dimensional distributions of any two frequencies. The PDF fitting technique employed here is the Kernel Density Estimation. It is observed that the first four natural frequencies have similar marginal distributions and their joint distributions show a strong linear relation. However, the 5th natural frequency has different character not only in its marginal distribution but also in its joint distributions with other four frequencies, as shown in the bottom line of the plot matrix in Fig. 20.

The above phenomenon can be explained by the mode shapes of the different modes as illustrated in Fig. 21. It is observed that the first four modes are all about the bending of the wing, while the 5th mode is the torsion mode. The different mode patterns lead the varying correlations between different natural frequencies. Another possible reason is that the uncertain wing tip chord b (recall Fig. 17) has a more significant influence on the torsion mode (the 5th mode) rather than the other bending modes (the 1st-4th modes). As a result, the 5th natural frequency has a larger dispersion than the other four frequencies.

The complicated distribution features of the first five natural frequencies fulfil our expectation that the datasets contain multiple sources of uncertainties, and it presents a challenge in reproducing the distribution features via the updating of the uncertainty model of the input parameters.

6.2.2. Influence of the likelihood function to the calibration results

Since the different statistical distances (Euclidian, Mahalanobis, and Bhattacharyya distances) are compared in detail in the NASA UQ Challenge example, this example will only employ the Bhattacharyya distance to construct the likelihood function in the Bayesian updating process. Nevertheless, the complicated distribution feature of the frequencies implies that the different configurations employing the Bhattacharyya distance to construct the final likelihood functions has significant influence to the calibration results. And hence should be further investigated as follows.

6.2.3. Likelihood based on even mean of the Bhattacharyya distances

The likelihood is first constructed as the even mean values of the Bhattacharyya distances between each of the simulated and

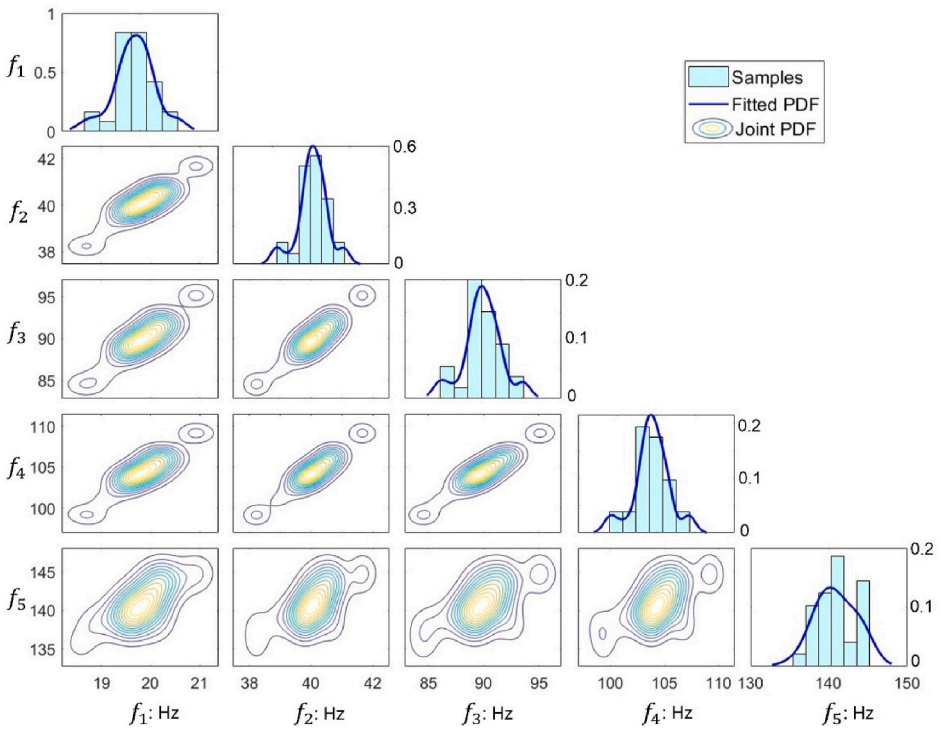


Fig. 20. Marginal and joint distributions of the first five natural frequencies in the measurement dataset.

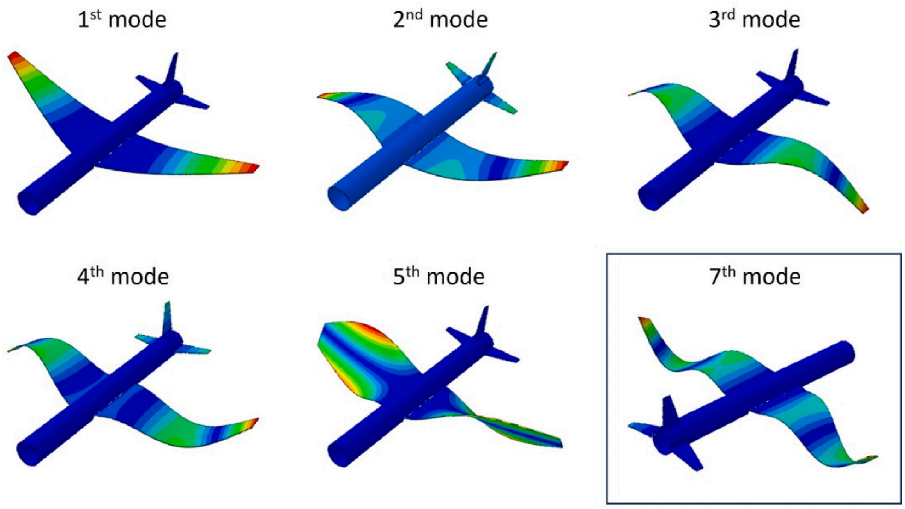


Fig. 21. Mode shapes calculated from the preliminary FE model (the first five modes are used for calibration in Sce. 6.2, and the 7th mode will be used for validation in Sec. 6.3).

measured natural frequencies.

$$BD_{even} = (BD_1 + BD_2 + BD_3 + BD_4 + BD_5)/5 \tag{23}$$

where $BD_i = d_B\left(f_i^{(exp)}, f_i^{sim}\right), i = 1, \cdots, 5$, is the Bhattacharyya distance of the marginal distribution of each frequencies, calculated based on Eq. (8). The likelihood function is then calculated as

$$P_{L_{even}} = \exp\left\{-\frac{BD_{even}^2}{\sigma^2}\right\} \tag{24}$$

The Bayesian TMCMC process is performed employing the above definition of the likelihood function with the width coefficient σ set as 0.05. According to Table 6, there are totally seven uncertainty characteristics coefficients to be calibrated, namely, the mean μ and standard deviation σ of a and b, the correlation coefficient r between a and b, and the constant value of E_1 and E_2. The posterior distributions of the seven calibrating coefficients are illustrated in Fig. 22. Their updated values are obtained as the highest point found from the posterior PDF functions, as shown in Fig. 22, and listed in Table 7.

For the practical application, it is more direct to compare the distribution of the predicted natural frequencies, which is illustrated in Fig. 23. It is interesting to compare the updating effect among different frequencies. For $f_2 - f_4$, their updated PDFs are quite close to the PDFs of the measured data, implying the Bayesian updating process works well for these three frequencies. However, the updated PDFs of f_1 still presents a discrepancy from the target measurement data. And especially for f_5, its PDF after the updating process is still nearly the same as its prior PDF before the updating process. This means the current Bayesian updating process is ineffective for f_5. Such phenomenon can be explained by the above investigation of the distribution feature of the five natural frequencies in Fig. 20, where f_5 exhibits a different dispersion feature compared with $f_1 - f_4$. Because the likelihood function is defined based on the even mean value of all five frequencies, the strong linearity among $f_1 - f_4$ lead the updating process is dominated by the error from $f_1 - f_4$ but not f_5. This reveals the motivation to adjust the likelihood function according to the specific linearity and proportion among the output features.

6.2.4. *Likelihood based on weighted mean of the Bhattacharyya distances*

A simple adjustment of the original likelihood is to rearrange the weightings among the frequencies when calculating the overall

Table 7
Updated values of the uncertainty coefficients of the input parameters.

Parameter	Description	Uncertainty Coefficient	Original interval	Calibrated values θ	
				Even likelihood	Weighted likelihood
a	Mean	μ_a(*10^{2}mm)	[2.9, 3.1]	3.022	3.026
	Standard deviation	σ_a(*10 mm)	[0, 0.5]	0.109	0.450
b	Mean	μ_b(*10^{2}mm)	[2, 3]	2.352	2.480
	Standard deviation	σ_b(*10 mm)	[0, 0.5]	0.495	0.106
a and b	Correlation Coefficient	r	[−1.0, 1.0]	−0.713	−0.151
E_1	Fixed-but-unknown constant	E_1(*10^{11}Pa)	[0.5, 0.9]	0.522	0.757
E_2	Fixed-but-unknown constant	E_2(*10^{11}Pa)	[0.5, 0.9]	0.542	0.726

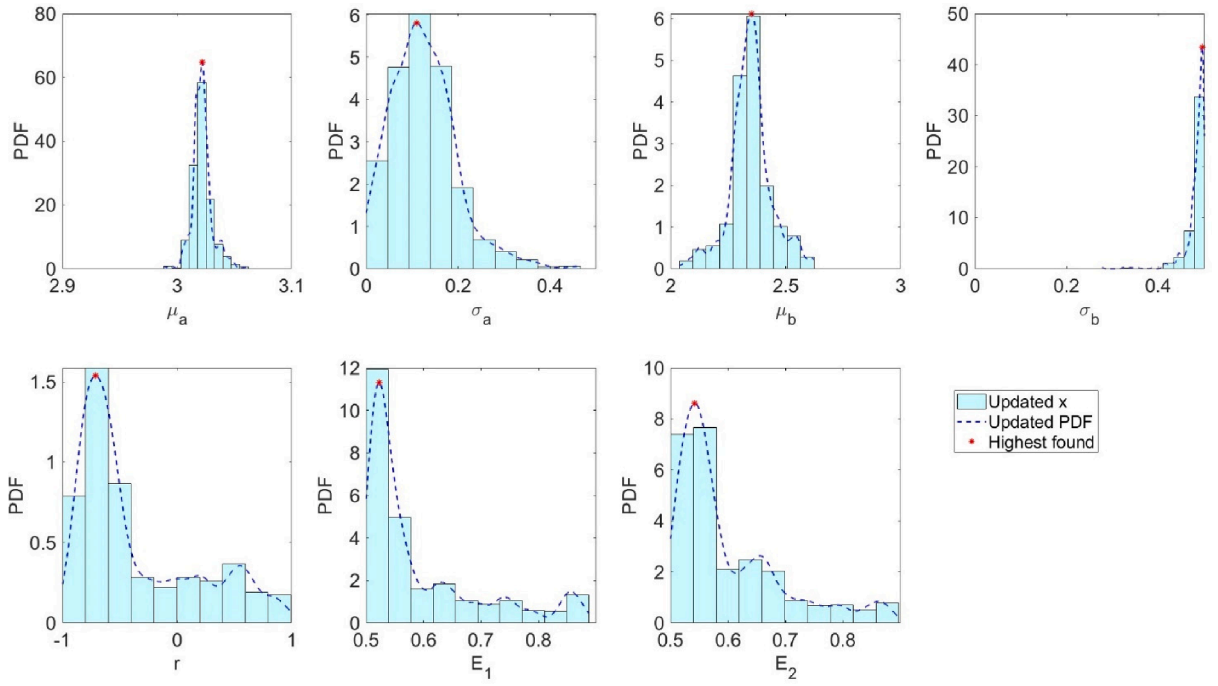


Fig. 22. Posterior distributions of the uncertainty coefficients of the input parameters updated with the even mean likelihood function.

Bhattacharyya distance

$$BD_{weight} = (BD_1 + BD_2 + BD_3 + BD_4 + 2{*}BD_5)/6 \tag{25}$$

Compared with the even mean value in Eq. (21), the weighting of f_5 is increased from 20% to 33.3%, and correspondingly the weighting of f_{1-4} is decreased from 20% to 16.7%. Such adjustment leads to an immediate effect on the updated PDFs of the predicted frequencies, as shown in Fig. 24. Especially, the updated PDF of f_5 is closely coincident with the measured PDF, which is an obvious

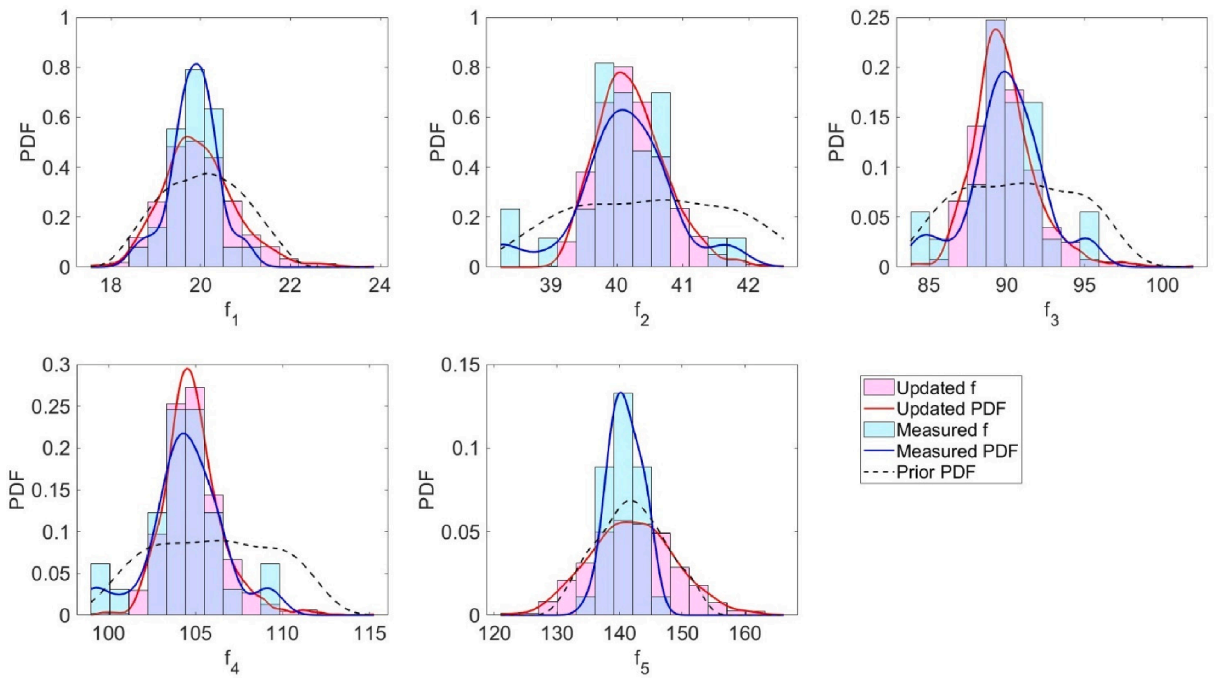


Fig. 23. Comparison among the measured, initial, and updated PDFs of the natural frequencies with even mean likelihood function.

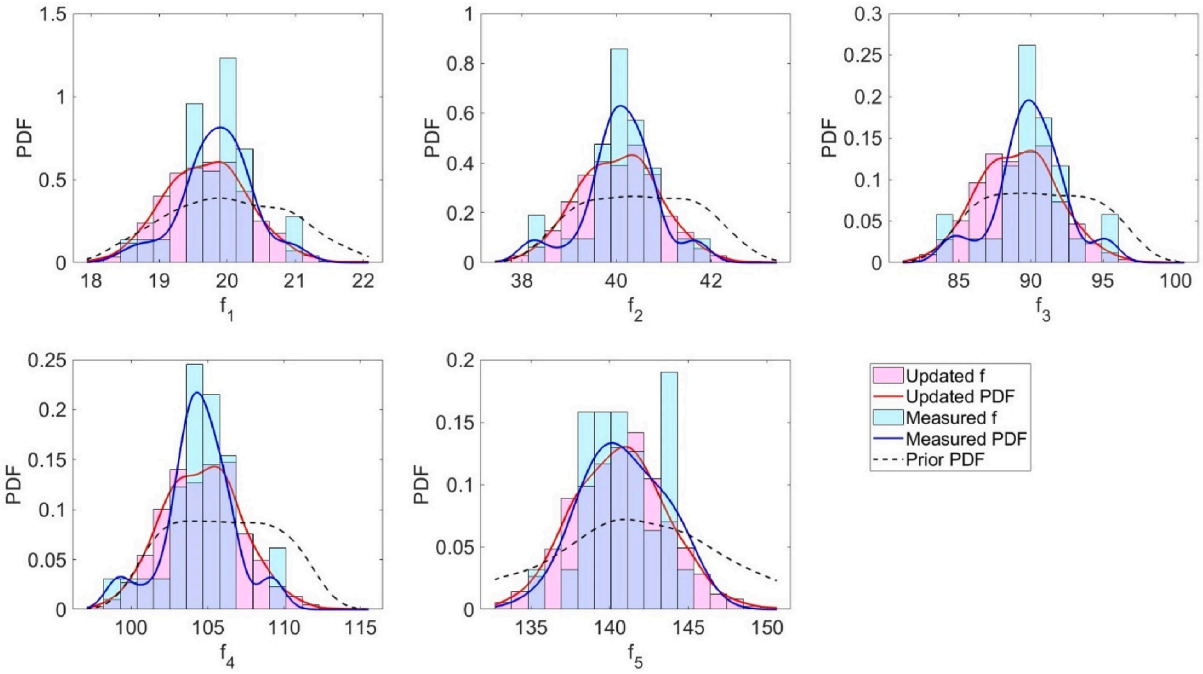


Fig. 24. Comparison among the measured, initial, and updated PDFs of the natural frequencies with weighted mean likelihood function.

change compared with the one in Fig. 23. Simultaneously, however, the consistency of the updated PDFs of f_{1-4} is not as good as the ones in Fig. 23. This phenomenon reveals the trade-off between the updating effect of f_{1-4} and f_5. Such a trade-off origins from the imperfectness of the FE model, which is inevitable, but a sophisticated model or a more appropriate parameterisation would relieve such phenomenon.

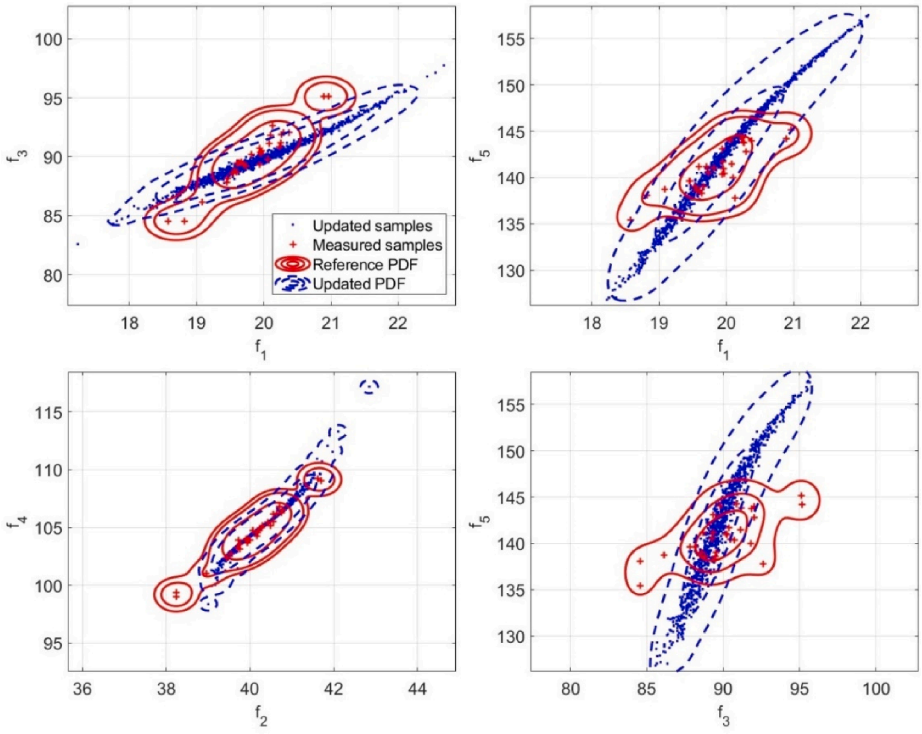


Fig. 25. Scatters of the measured and updated frequencies with the even mean likelihood function.

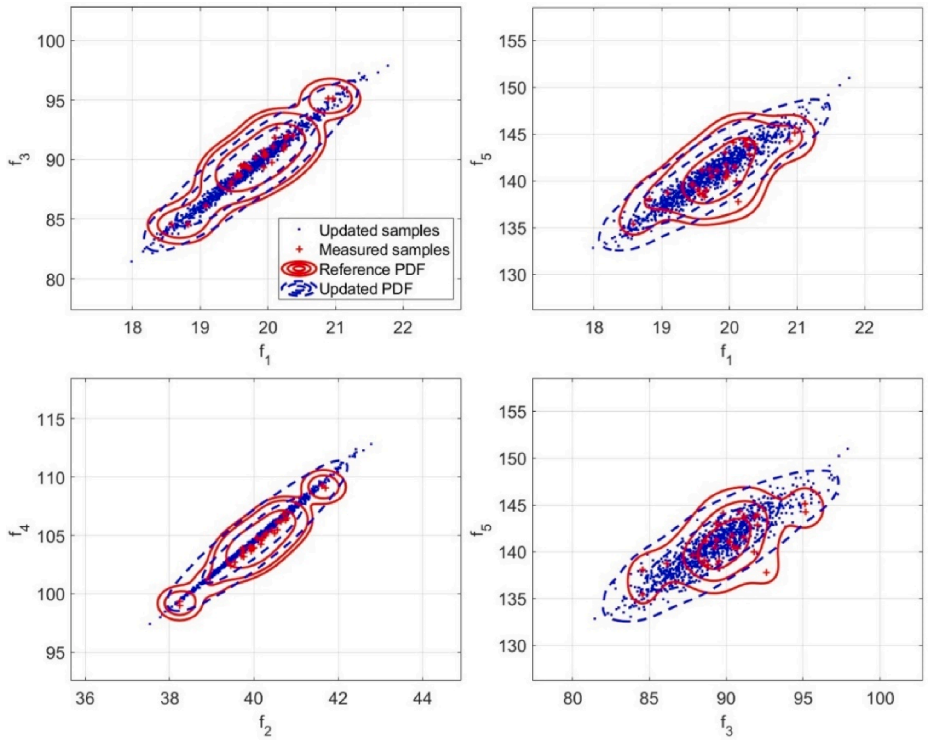


Fig. 26. Scatters of the measured and updated frequencies with the weighted mean likelihood function.

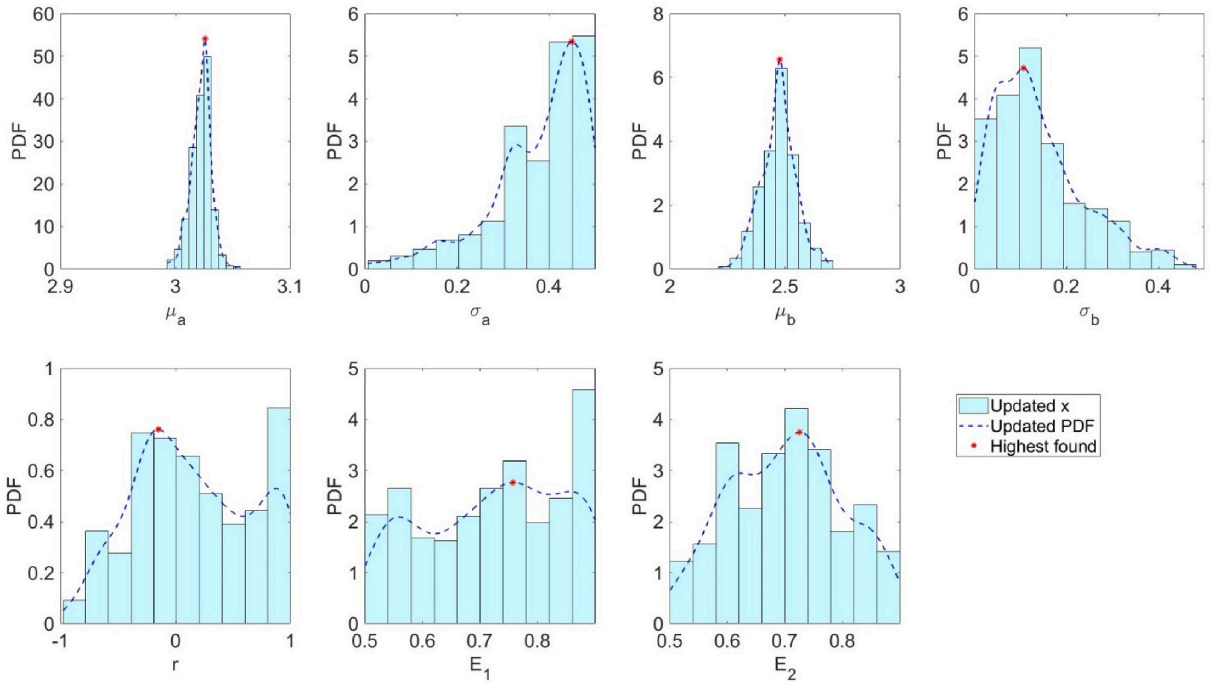


Fig. 27. Posterior distributions of the uncertainty coefficients of the input parameters updated with the weighted mean likelihood function.

The different updating effects can be further vividly revealed by the comparison between Fig. 25 and Fig. 26, where the scatters (and their fitted 2D joint PDFs) are presented. Similarly, special attention should be paid on the scatters involving f_5 (the subfigures in the left part of both Fig. 25 and Fig. 26). It is observed that, when the even mean likelihood is employed in Fig. 25, the updated scatters of $f_1 vs. f_5$ and $f_3 vs. f_5$ are still obviously inconsistent with the target measured scatters. While when the weighted mean likelihood is employed in Fig. 26, the updated scatters of $f_1 vs. f_5$ and $f_3 vs. f_5$ coincide much more with the measurement data. Note that, this comparison is not necessarily to conclude that the weighted mean likelihood function is superior upon the even mean likelihood function. It is merely a demonstration that the customised likelihood function would act as a steering of the Bayesian updating process when one specific quantity of interest is paid more attention than the others.

The different updating effects of the different likelihood functions is a good demonstration of the ability of the TMCMC searching algorithm. Different definitions of the likelihood functions are actually the different searching criteria in the Bayesian updating process. By assigning a higher weighting to a specific frequency, the TMCMC algorithm is capable of searching particular regions in the parameter space leading to a better fitting effect for this specific frequency. This means the TMCMC algorithm is sensitive and effective to a given searching direction.

Now come back to the updating effect of the input parameters. The different likelihood function results in a very different posterior distributions of the input parameters coefficients, as shown Fig. 27, compared with the ones in Fig. 22. This reveals the challenging feature of the example and leads to the discussion of the open questions of the benchmark testbed in the following subsection.

6.3. Validation results

The above parameter updating results are obtained by employing the first five frequencies as the reference data. This subsection provides the model validation result according to the second validation criterion introduced in Sec. 3.5. The 7th mode (see Fig. 21) is employed here as the independent data to assess whether the updated model can predict its distribution feature. The reason for selecting the 7th mode rather than the 6th mode is that the 5th and 6th FE modes are the anti-symmetric and symmetric wing twisting modes respectively, very close to each other in frequency. In practical experiments, the two are combined and a 6th test mode is therefore non-existent.

Updating results from both the even mean likelihood and the weighted mean likelihood in Sec. 6.2 are investigated. The mean, standard deviation, and correlation coefficient of a and b are extracted from Table 7. Based on the pre-setting of the joint Gaussian distribution, 30 samples of a and b are obtained. Together with the obtained constant values of E_1 and E_2 (also in Table 7), these input parameters are sent to the FE model and 30 samples of f_7 are obtained. The histograms and fitted PDFs of f_7 are illustrated and compare with the experimental samples in Fig. 28. The means and variances of these samples are compared in Table 8.

It is observed from Table 8 that both the even-likelihood result and the weighted-likelihood result have a good match for the mean (with the relative errors less than 1.5%). This implies that, from a deterministic model updating point of view, the validation result is already satisfactory. For the variance, however, the error is higher than the mean. But it is interesting to note that the variance error of

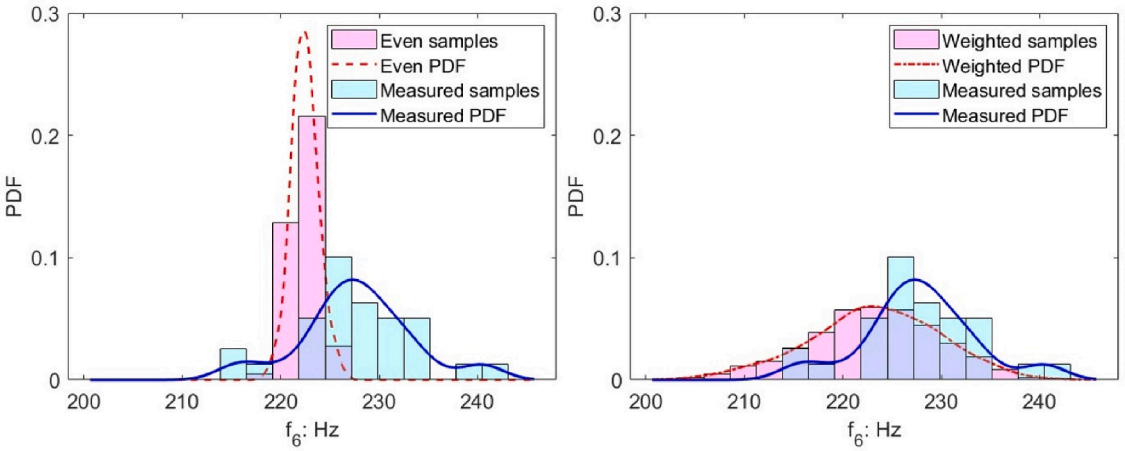


Fig. 28. Validation result of the 7th natural frequency.

Table 8
Comparison of the FE data and the experimental data of the 7th natural frequency (with percentage error in the brackets).

The 7th frequency	Experimentaldata (Hz)	FE data (Hz) Even-likelihood	Weighted-likelihood
Mean	227.74	225.66 (-0.91%)	224.41 (-1.47%)
Variance	33.01	2.98 (-91.97%)	40.45 (22.52%)

the weighted likelihood data is much less than the one of the even-likelihood data. This phenomenon can be clearly reflected by Fig. 28 where the distribution of the weighted-likelihood samples fits with the experimental samples much better than the even-likelihood result. Although the fitting degree is not as good as the results of $f_1 - f_5$ in Fig. 23 and Fig. 24, it does not influence the conclusion that the validation using f_7 is acceptable given the mean and variance errors of the weighted-likelihood data have been clearly reduced.

6.4. Summary and open questions

The fact is that, no matter the posterior distributions in Fig. 22 or Fig. 27, they are not fully converged to the pre-set true distribution coefficients of a and b. This reveals the challenging feature of the benchmark testcase and some summaries and possible directions to improve the updating performance are given as follows.

- The capacity of the TMCMC algorithm is demonstrated in this example as being able to produce output features with probabilistic characteristics coinciding with the measurements, subject to specific emphasis on certain frequencies, e.g. f_5.
- One highlight of the Bayesian updating framework is the flexibility of the customised likelihood function with different weightings assigned to the output features, which effectively steers the direction of the TMCMC sampling process.
- Non-uniqueness is much more common for stochastic model updating than the deterministic tasks, especially when variances and correlations are involved. The ability to differentiate the global solutions from multiple local solutions will be one of the key expectations for advanced updating algorithms.
- The trade-off of the updating effects among $f_1 - f_5$ originates from the imperfectness of the FE model in representing the physical system as well as the practical experimental randomness. Parameterisation is a critical step to set an appropriate compensation between the modelling uncertainty and experimental uncertainty. Remodelling with new mesh configuration and selection of new parameters other than E_1 and E_2 would relieve the compensation effect, and lead to different updating effects.

7. Perspectives and future challenges

The following section presents the latest advancements and challenges in stochastic model updating. The goal is to provide readers with insights that can inspire further exploration of this topic.

Distribution-free model updating: The techniques and examples presented in this tutorial are all under the condition that the probabilistic distribution format of the input parameters is already known. This is a strong assumption, especially in a situation where limited observation data is available. The NASA UQ Challenge 2021 [37] is developed to confront such a distribution-free problem with limited observation. Various techniques have been developed, e.g. staircase density function [38], Gaussian mixture model, and Bate mixture model [16], to parameterise the unknown distributions into undermined coefficients. This is actually a transformation of the distribution-free problem into the distribution-based problem by this additional parameterisation step. Further techniques capable of quantifying the unknown distribution directly without introducing more coefficients would be a significant contribution to this problem. The distribution-free model updating would be significantly facilitated by hybrid uncertainties quantification, non-parametric imprecise probabilities and distribution-free P-box technique.

Non-probabilistic model updating: Interval and fuzzy methods have been extensively studied as complements to probabilistic UQ, making them promising options for use in stochastic model updating. Both approaches are set-theoretical approaches, which do not engage probabilities for assessing likelihood or belief of some events. Hence, they require less information for quantification in comparison to probabilistic approaches, which is helpful when data are very limited. The interval approach [39,40] focuses on estimating and propagating bounds for the quantity of interest. Applied to modelling of epistemic uncertainty, intervals become tighter as available information grows in quantity and quality. Fuzzy sets can be understood as a generalisation of intervals, representing a nested set of intervals and be used as an instrument to process intervals of varying size [24,41]. As such, fuzzy sets are suitable for quantifying linguistic variables and approximate statements or expressions. When using interval and fuzzy techniques for model updating, a different method for uncertainty propagation is needed compared to sampling-based distribution propagation. This is typically an optimisation approach to find the bounds. To fully utilise the cost-effective nature of intervals and fuzzy sets in stochastic model updating, a precise and efficient propagation process would be the key.

Nonlinear model updating: Modern engineering systems often exhibit nonlinearity due to factors such as large deformations, nonlinear boundary conditions, or geometrically nonlinear effects. Nonlinearities can create intricate interactions between model parameters, cause significant variability in the sensitivities of model parameters, and often result in non-unique and localised solutions. Nonlinearity presents a challenge that requires more advanced model updating techniques [42]. First, a more sophisticated parameterisation is necessary to isolate the nonlinearity from the entire system and distinguish specific parameters related to it, e.g. see the work Ref [43]. Second, advanced test techniques are important for nonlinear model updating to provide full-field measurements [44] with sufficient data to reveal nonlinearity from large-scale industrial systems.

Robust model updating: When dealing with uncertainty, it is essential to also prioritise robustness in model updating in addition to precision. Robustness means being able to accept uncertainty and is a way to measure how much uncertainty can be tolerated while still maintaining acceptable model precision. Multi-objective optimisation is a common solution when it comes to making a trade-off choice [45] between robustness and precision in stochastic model updating. It is hence important to define a clear and quantitative metric of robustness before it can be implemented in the multi-objective optimisation process, although a common agreement of robustness metrics has been far from achieved given the fact that varying conceptual definitions are used in the literature [46].

Credibility of simulation outside the validation domain: When using numerical models, analysts are often confronted with the

situation whereby simulation predictions are used to inform decision-making in a domain where experimental data are unavailable. In such cases, the numerical model is trusted to extrapolate in the forecasting domain simply because it has been shown to be trustworthy in the validation domain, where experimental data are available. Although the inherent dangers of this prediction extrapolation are widely recognised, few methods are available to actually establish simulation credibility for a given application. Credible model extrapolation is hence a significant perspective of model updating to reflect the uncertainty from the validation domain to the extrapolation domain.

Parameterisation and stochastic sensitivity analysis: The benchmark testbed example in Sec. 6 highlights the importance of the initial parameterisation in achieving the proper compensation between modelling uncertainty and experimental uncertainty. A well-executed parameterisation of the FE model sets the stage for effective model updating, which in turn enhances the model's robustness and extrapolation performance. To achieve this, more advanced techniques like stochastic and global sensitivity analysis [33,47] are necessary. These techniques not only take into account the functional importance of the parameter but also accurately model its relative uncertainty according to its statistical nature.

Model updating with data science: The tide of data science has clearly influenced model updating but actually has yet to reach a level of revolution. The fact is that vast literature on data science and model updating is still limited to training a surrogate model using machine learning or reinforcement learning. However, a nature contradiction is that the data science-based model relies on massive data, which is exactly missing for practical engineering observations. Data science can support other aspects of model updating, not just surrogate modelling. One exciting possibility is to fully replace the optimisation and sampling algorithm with an intelligent, self-renewing algorithm by engaging deeply in the inverse parameter calibration process. Also, the stochastic pattern identification and feature extraction would be the significant perspectives of data science to boost the embrace of uncertainty quantification in stochastic model updating.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

Data will be made available on request.

Acknowledgement

This work is supported by the National Natural Science Function of China under grant 12102036.

Appendix

Table A1

Table A1
Measured natural frequencies of the airplane models with 30 varying wing pieces.

Index	f_1(Hz)	f_2(Hz)	f_3(Hz)	f_4(Hz)	f_5(Hz)
1	19.5939	39.7633	88.6695	103.149	141.1495
2	19.7038	39.9739	89.3527	103.9124	141.2092
3	19.0858	38.9641	86.1415	100.9968	138.7494
4	20.2534	40.5603	91.1972	105.4919	143.7954
5	19.562	39.7552	88.8154	103.5576	138.5737
6	20.2339	40.4956	90.8475	105.1642	144.2686
7	20.289	40.7772	92.0356	106.4934	142.8112
8	19.9444	40.287	90.3861	104.9847	141.7157
9	19.9382	40.2902	90.7258	105.1755	140.3442
10	20.3694	40.7582	92.0218	106.2751	143.9561
11	19.9688	40.0731	90.4805	104.7616	140.455
12	19.6028	39.7311	89.4728	103.8851	138.6102
13	18.8157	38.2389	84.5431	98.9893	138.0854
14	18.5723	38.237	84.5388	99.3751	135.4328
15	20.2594	40.6322	91.8662	106.1835	143.7322
16	20.0747	40.4262	91.1133	105.433	141.4829
17	19.8147	40.1952	90.1828	104.754	141.0558
18	19.4615	39.5471	88.1474	102.7056	139.6227
19	20.1035	40.5532	91.8229	106.1403	139.9808
20	20.8916	41.614	95.1423	109.2518	144.2192
21	19.9619	40.1781	90.2592	104.5704	140.9626
22	19.5338	39.7062	88.5632	103.2578	139.0521
23	20.9609	41.6971	95.1049	109.0708	145.1307

(*continued on next page*)

Table A1 (*continued*)

Index	f_1(Hz)	f_2(Hz)	f_3(Hz)	f_4(Hz)	f_5(Hz)
24	19.7298	39.9948	89.2391	103.9312	140.5159
25	19.4527	39.4478	87.8373	102.4027	139.5921
26	19.6318	39.9219	89.4969	104.0277	138.2781
27	20.133	40.6985	92.6301	106.8184	137.7652
28	19.6585	39.909	89.5363	103.9684	139.0352
29	19.9378	39.9397	89.3377	103.6612	143.0868
30	20.055	40.1601	89.7844	104.257	144.0743
Mean	19.8531	40.0842	89.9764	104.4215	140.8914
Variance	0.2611	0.5772	5.8244	5.0873	5.7667

References

[1] J.E. Mottershead, M.I. Friswell, Model updating in structural dynamics: A survey, Journal of Sound and Vibration 167 (1993) 347–375, https://doi.org/10.1006/jsvi.1993.1340.

[2] J.E. Mottershead, M. Link, M.I. Friswell, C. Schedlinski, Model Updating, in: R. Allemang, P. Avitabile (Eds.), Handb. Exp. Struct. Dyn., Springer New York, New York, NY, 2020: pp. 1–53. https://doi.org/10.1007/978-1-4939-6503-8_18-1.

[3] ASME, Guide for Verification & Validation in Computational Solid Mechanics, 2006.

[4] A. Calvi, Uncertainty-based loads analysis for spacecraft: Finite element model validation and dynamic responses, Computers and Structures 83 (2005) 1103–1112, https://doi.org/10.1016/j.compstruc.2004.11.019.

[5] M.I. Friswell, J.E. Mottershead, Finite element model updating in structural dynamics, kluwer academic press, Dordrecht (1995), https://doi.org/10.1007/978-94-015-8508-8.

[6] J.E. Mottershead, M. Link, M.I. Friswell, The sensitivity method in finite element model updating: A tutorial, Mechanical Systems and Signal Processing 25 (2011) 2275–2296, https://doi.org/10.1016/j.ymssp.2010.10.012.

[7] G.I. Schuëller, Computational stochastic mechanics - recent advances, Computers and Structures 79 (2001) 2225–2234, https://doi.org/10.1016/S0045-7949(01)00078-5.

[8] C. Mares, J.E. Mottershead, M.I. Friswell, Stochastic model updating: Part 1-theory and simulated example, Mechanical Systems and Signal Processing 20 (2006) 1674–1695, https://doi.org/10.1016/j.ymssp.2005.06.006.

[9] B. Goller, M. Broggi, A. Calvi, G.I. Schuëller, A stochastic model updating technique for complex aerospace structures, Finite Elements in Analysis and Design 47 (2011) 739–752, https://doi.org/10.1016/j.finel.2011.02.005.

[10] Y. Govers, M. Link, Stochastic model updating-Covariance matrix adjustment from uncertain experimental modal data, Mechanical Systems and Signal Processing 24 (2010) 696–706, https://doi.org/10.1016/j.ymssp.2009.10.006.

[11] J.L. Beck, L.S. Katafygiotis, Updating models and their uncertainties. I: Bayesian statistical framework, Journal of Engineering Mechanics 124 (1998) 455–461, https://doi.org/10.1061/(ASCE)0733-9399(1998)124:4(455).

[12] J.L. Beck, S.K. Au, Bayesian updating of structural models and reliability using markov chain monte carlo simulation, Journal of Engineering Mechanics 128 (2002) 380–391, https://doi.org/10.1061/(ASCE)0733-9399(2002)128:4(380).

[13] L.G. Crespo, S.P. Kennyy, D.P. Giesyz, The NASA langley multidisciplinary uncertainty quantification challenge, in, 16th AIAA Non-Deterministic Approaches Conf. (2014) 1–9, https://doi.org/10.2514/6.2014-1347.

[14] L.G. Crespo, S.P. Kenny, Synthetic validation of responses to the NASA langley challenge on optimization under uncertainty, Mechanical Systems and Signal Processing 164 (2022), 108253, https://doi.org/10.1016/J.YMSSP.2021.108253.

[15] 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, https://doi.org/10.2514/1.I010256.

[16] S. Bi, K. He, Y. Zhao, D. Moens, M. Beer, J. Zhang, Towards the NASA UQ challenge 2019: Systematically forward and inverse approaches for uncertainty propagation and quantification, Mechanical Systems and Signal Processing 165 (2022), 108387, https://doi.org/10.1016/j.ymssp.2021.108387.

[17] A. Lye, M. Kitahara, M. Broggi, E. Patelli, Robust optimization of a dynamic black-box system under severe uncertainty : A distribution-free framework, Mechanical Systems and Signal Processing 167 (2022), 108522, https://doi.org/10.1016/j.ymssp.2021.108522.

[18] 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, Archive of Applied Mechanics 87 (2017) 905–925, https://doi.org/10.1007/s00419-017-1233-1.

[19] S. Bi, M. Beer, J. Zhang, L. Yang, K. He, Optimization or bayesian strategy? performance of the bhattacharyya distance in different algorithms of stochastic model updating, ASCE-ASME j. Risk uncertain. Eng. Syst. Part B, Mechanical Engineering 7 (2021), https://doi.org/10.1115/1.4050168.

[20] M. Faes, D. Moens, Recent trends in the modeling and quantification of non-probabilistic uncertainty, Archives of Computational Methods in Engineering 27 (2020) 633–671, https://doi.org/10.1007/s11831-019-09327-x.

[21] G. Filippi, M. Vasile, Introduction to Evidence-Based Robust Optimisation, in: M. Vasile (Ed.), Optim. Under Uncertain. with Appl. to Aerosp. Eng., Springer International Publishing, Cham, 2021: pp. 541–573. https://doi.org/10.1007/978-3-030-60166-9_17.

[22] Y. Ben-Haim, info-gap value of information in model updating, Mechanical Systems and Signal Processing 15 (2001) 457–474, https://doi.org/10.1006/mssp.2000.1377.

[23] J.-Y. Wu, G. Gumbs, M.-F. Lin, Combined effect of stacking and magnetic field on plasmon excitations in bilayer graphene, Springer Science & Business Media 89 (16) (2014), https://doi.org/10.1103/PhysRevB.89.165407.

[24] M. Beer, S. Ferson, V. Kreinovich, Imprecise probabilities in engineering analyses, Mechanical Systems and Signal Processing 37 (2013) 4–29, https://doi.org/10.1016/j.ymssp.2013.01.024.

[25] S. Bi, S. Prabhu, S. Cogan, S. Atamturktur, Uncertainty quantification metrics with varying statistical information in model calibration and validation, AIAA Journal 55 (2017) 3570–3583, https://doi.org/10.2514/1.J055733.

[26] S. Bi, M. Broggi, M. Beer, the role of the bhattacharyya distance in stochastic model updating, Mechanical Systems and Signal Processing 117 (2019) 437–452, https://doi.org/10.1016/j.ymssp.2018.08.017.

[27] R.J. Allemang, The modal assurance criterion - twenty years of use and abuse, Sound Vib. 37 (2003) 14–21.

[28] S. Bi, M. Ouisse, E. Foltête, Probabilistic approach for damping identification considering uncertainty in experimental modal analysis, AIAA Journal 56 (2018) 4953–4964, https://doi.org/10.2514/1.J057432.

[29] B. Bulca K. Arslan Surfaces given with the Monge patch in E4 J. Math. Physics, Anal. Geom. 9 2013 435 447 1061-7590/93/04407-008.

[30] A. Saltelli, M. Ratto, T. Andres, F. Campolongo, J. Cariboni, D. Gatelli, M. Saisana, S. Tarantola (Eds.), Global Sensitivity Analysis. The Primer, Wiley, 2007.

[31] R. Yondo, E. Andrés, E. Valero, A review on design of experiments and surrogate models in aircraft real-time and many-query aerodynamic analyses, Progress in Aerospace Science 96 (2018) 23–61, https://doi.org/10.1016/j.paerosci.2017.11.003.

[32] M. Link, M. Friswell, Generation of validated structural dynamic models - Results of a benchmark study utilising the GARTEUR SM-AG19 testbed, Proc. 2002 Int. Conf. Noise Vib. Eng. ISMA. 17 (2002) 1005–1017.
[33] S. Bi, M. Broggi, P. Wei, M. Beer, the bhattacharyya distance: Enriching the p-box in stochastic sensitivity analysis, Mechanical Systems and Signal Processing 129 (2019) 265–281, https://doi.org/10.1016/j.ymssp.2019.04.035.
[34] J. Ching, Y.C. Chen, Transitional markov chain monte carlo method for bayesian model updating, model class selection, and model averaging, Journal of Engineering Mechanics 133 (2007) 816–832, https://doi.org/10.1061/(ASCE)0733-9399(2007)133:7(816).
[35] 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, https://doi.org/10.2514/1.I010273.
[36] S.J. Sheather, M.C. Jones, A reliable Data-Based bandwidth selection method for kernel density estimation, Journal of the Royal Statistical Society, Series B 53 (1991) 683–690, https://doi.org/10.1111/j.2517-6161.1991.tb01857.x.
[37] L.G. Crespo, S.P. Kenny, The NASA langley challenge on optimization under uncertainty, Mechanical Systems and Signal Processing 152 (2021), 107405, https://doi.org/10.1016/j.ymssp.2020.107405.
[38] M. Kitahara, S. Bi, M. Broggi, M. Beer, Nonparametric bayesian stochastic model updating with hybrid uncertainties, Mechanical Systems and Signal Processing 163 (2022), 108195, https://doi.org/10.1016/j.ymssp.2021.108195.
[39] M. Faes, M. Broggi, E. Patelli, Y. Govers, J. Mottershead, M. Beer, D. Moens, A multivariate interval approach for inverse uncertainty quantification with limited experimental data, Mechanical Systems and Signal Processing 118 (2019) 534–548, https://doi.org/10.1016/j.ymssp.2018.08.050.
[40] Y. Zhao, J. Yang, M.G.R. Faes, S. Bi, Y. Wang, The sub-interval similarity: A general uncertainty quantification metric for both stochastic and interval model updating, Mechanical Systems and Signal Processing 178 (2022), 109319, https://doi.org/10.1016/j.ymssp.2022.109319.
[41] B. Möller, W. Graf, M. Beer, Fuzzy structural analysis using α-level optimization, Computational Mechanics 26 (2000) 547–565, https://doi.org/10.1007/s004660000204.
[42] P. Bussetta, S.B. Shiki, S. da Silva, Nonlinear updating method: a review, J. Brazilian Soc. Mech. Sci. Eng. 39 (2017) 4757–4767, https://doi.org/10.1007/s40430-017-0905-7.
[43] G. Kerschen, J.-C. Golinval, F.M. Hemez, Bayesian model screening for the identification of nonlinear mechanical structures, Journal of Vibration and Acoustics 125 (2003) 389–397, https://doi.org/10.1115/1.1569947.
[44] X. Wang, M. Szydlowski, J. Yuan, C. Schwingshackl, A multi-step Interpolated-FFT procedure for full-field nonlinear modal testing of turbomachinery components, Mechanical Systems and Signal Processing 169 (2022), 108771, https://doi.org/10.1016/j.ymssp.2021.108771.
[45] S. Atamturktur, Z. Liu, S. Cogan, H. Juang, Calibration of imprecise and inaccurate numerical models considering fidelity and robustness: a multi-objective optimization-based approach, Structural and Multidisciplinary Optimization 51 (2015) 659–671, https://doi.org/10.1007/s00158-014-1159-y.
[46] S. Moritz Göhler, T. Eifler, T.J. Howard, robustness metrics: Consolidating the multiple approaches to quantify robustness, Journal of Mechanical Design 138 (2016), https://doi.org/10.1115/1.4034112.
[47] Z. Yuan, P. Liang, T. Silva, K. Yu, J.E. Mottershead, Parameter selection for model updating with global sensitivity analysis, Mechanical Systems and Signal Processing 115 (2019) 483–496, https://doi.org/10.1016/j.ymssp.2018.05.048.