Towards the NASA UQ Challenge 2019: Systematically forward and inverse approaches for uncertainty propagation and quantification

Sifeng Bi[*, 1], Kui He[1], Yanlin Zhao[2], David Moens[3], Michael Beer[4, 5, 6], Jingrui Zhang[1]

1. Beijing Institute of Technology, School of Aerospace Engineering, Beijing 100081, China
2. University of Science and Technology Beijing, School of Mechanical Engineering, Beijing 100083, China
3. KU Leuven, Department of Mechanical Engineering, Leuven, Belgium
4. Institute for Risk and Reliability, Leibniz University Hannover, Hannover 30167, Germany
5. Institute for Risk and Uncertainty, University of Liverpool, Liverpool L69 7ZF, the UK
6. International Joint Research Center for Engineering Reliability and Stochastic Mechanics, Tongji University, Shanghai 200092, China

*Corresponding author: Sifeng Bi (sifeng.bi@bit.edu.cn)

Abstract: This paper is dedicated to exploring the NASA Langley Challenge on Optimization under Uncertainty by proposing a series of approaches for both forward and inverse treatment of uncertainty propagation and quantification. The primary effort is placed on the categorization of the subproblems as to be forward or inverse procedures, such that dedicated techniques are proposed for the two directions, respectively. The sensitivity analysis and reliability analysis are categorized as forward procedures, while modal calibration & uncertainty reduction, reliability-based optimization, and risk-based design are regarded as inverse procedures. For both directions, the overall approach is based on imprecise probability characterization where both aleatory and epistemic uncertainties are investigated for the inputs, and consequently, the output is described as the probability-box (P-box). Theoretic development is focused on the definition of comprehensive uncertainty quantification criteria from limited and irregular time-domain observations to extract as much as possible uncertainty information, which will be significant for the inverse procedure to refine uncertainty models. Furthermore, a decoupling approach is proposed to investigate the P-box along two directions such that the epistemic and aleatory uncertainties are decoupled, and thus a two-loop procedure is designed to propagate both epistemic and aleatory uncertainties through the systematic model. The key for successfully addressing this challenge is in obtaining on the balance among an appropriate hypothesis of the input uncertainty model, a comprehensive criterion of output uncertainty quantification, and a computational viable approach for both forward and inverse uncertainty treatment.

Keywords: uncertainty quantification, uncertainty propagation, reliability analysis, risk-based design, NASA Challenge, reliability-based optimization

1 Introduction and problem pre-investigation

The NASA Challenge on Optimization under Uncertainty [1] is proposed by researchers in the NASA Langley Research Center in 2019 to represent the difficulties that are frequently encountered in the practical development of safety-critical vehicle systems. The Challenge 2019 can be regarded as an enhanced version of the previous edition of the NASA Multidisciplinary Uncertainty Quantification (UQ) Challenge 2014 [2]. The challenges are derived from the severe operating conditions that the flight vehicles are experiencing and the intractable trade-off between the vehicles' performance requirement and their safety constraints. Furthermore, the quantitative data representing the

operating conditions may be very sparse, strongly irregular, and contaminated by observation noise, leading that the gross data cannot be directly used in the modelling and design stage. The above difficulties are presented as the challenging tasks of model calibration, sensitivity analysis, uncertainty reduction, reliability-based optimization, and robust design, which are key aspects integrating the NASA Challenge 2019 problems.

The Challenge 2019 consists of a series of interactional tasks that are expected to be encountered in practical modelling and design processes. The overview of the Challenge 2019 is illustrated in Fig. 1, where Subproblems B1) Sensitivity Analysis and C) Reliability Analysis are categorized as the forward procedure, Subproblems A) Model Calibration, B3) Uncertainty Reduction, D) Reliability-based Design, E) Design Tuning, and F) Risk-based Design are categorized as the inverse procedure.

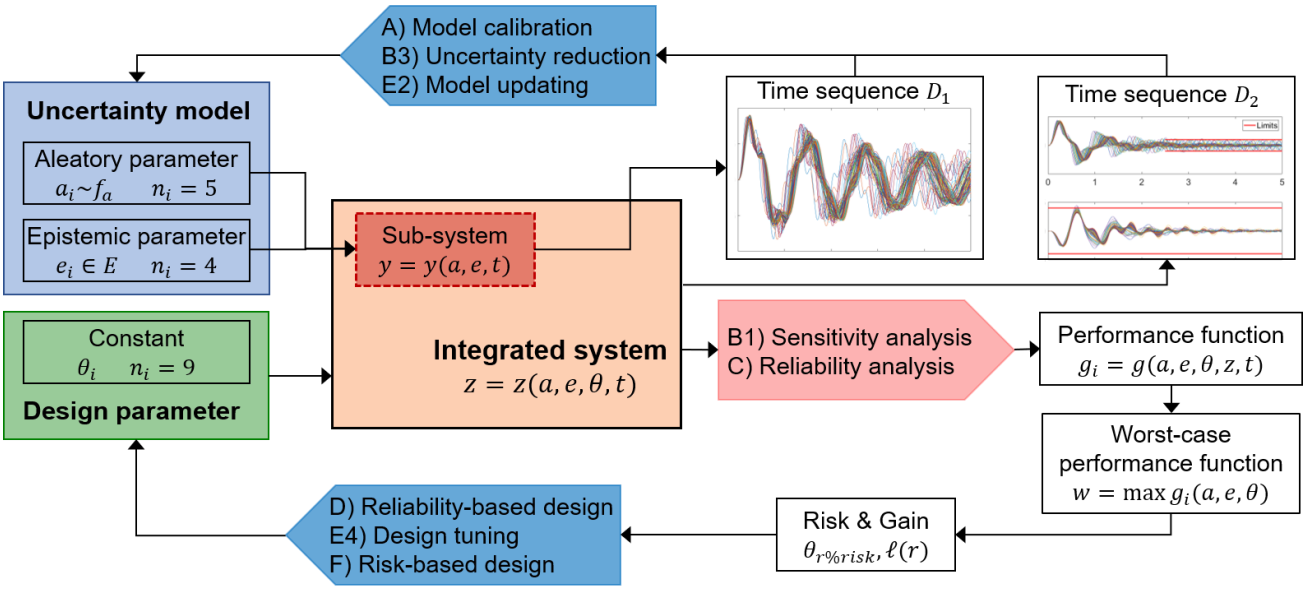

Fig. 1: Overview of the NASA Langley challenge on optimization under uncertainty

In this problem, the physical system is represented as a black-box model, i.e. the integrated system in Fig. 1, where a sub-system $y(a, e, t)$ is extracted to derive all sources of uncertainties within the integrated system. The uncertainties in this work are investigated following the classical categorization as to be aleatory or epistemic. Briefly explained, the aleatory uncertainty refers to the natural randomness of the system, and it is presented as variables following probabilistic distributions $a_i \sim f_a$. The epistemic uncertainty is the one caused by the lack of knowledge, and this type of uncertainty can be reduced as more information becomes available from the physical system. The epistemic uncertainty is presented in this problem as unknown-but-fixed constants situated within predefined intervals $e_i \in E$. The aleatory variables a_i and epistemic variables e_i are the input of the sub-system $y(a, e, t)$, whose output is a time-domain sequence. The distribution f_a and the predefined interval E are referred as the Uncertainty Model (UM). However, the prior knowledge of the UM is extremely limited in this problem, since only a wide boundary [0, 2] is given, and no further information of the distribution format or distribution coefficients is provided for a_i. One of the main tasks in this problem is to refine the UM regarding a set of observation of $y(a, e, t)$ as shown in Fig. 1. The observation set D_1 is presented to be very limited and irregular, furthermore with the lack of any distribution properties of aleatory variable a_i, rendering the model refinement extremely challenging.

The above-discussed model refinement task (including Subproblem A, B3, E2) belongs to the inverse procedure, because it is based on the output observation to calibrate the UM of the input. The other tasks of this problem are the

forward procedure including Subproblems B1) sensitivity and C) reliability analysis. The integrated system $z(a,e,\theta,t)$ takes not only the uncertain parameters a_i and e_i but also the design parameter θ_i as inputs. The outputs, also presented as time-domain sequences, are further employed to define the performance and worst-case performance function. Since both aleatory and epistemic variables are involved in the system, the output quantity of interest is presented as an infinite collection of probability distributions, i.e. the probability-box (P-box). The objective of the forward procedure herein is to investigate how the aleatory and epistemic uncertainties are propagated from the inputs to the outputs, and how the aleatory and epistemic uncertainties influence the failure probability of the system. This task requires a decoupling approach to investigate the aleatory and epistemic uncertainties respectively, and requires a comprehensive uncertainty quantification approach to measure the P-box of time-domain sequences.

The last task is the optimization design considering reliability and risk regarding uncertainties, including Subproblems D) reliability-based design, E4) Design tuning, and F) Risk-based design. It is required to search for a new design based on the provided baseline design, whereby the reliability of the system should be integrated into the objective function. The risk-based design requires the trade-off between the portion of the epistemic uncertainty one can neglect and the gain that can be obtained. This optimization task is performed in the presence of both aleatory and epistemic uncertainties, and hence the key challenge is the definition of the optimal criterion and the development of the computational viable optimization algorithm.

Based on the pre-investigation above, the featured challenging points of the overall problem are summarized as follows.

- The output features of the system are not scalars but time-domain sequences, rendering the classical model updating approaches developed for scalars no longer applicable on this occasion;
- The provided observation sequence data D_1 has only 100 samples and is presented in a complicated curve shape in the time domain, rendering the uncertainty information can be hardly extracted from the gross data;
- The aleatory variable a_i is supposed to follow a probability distribution, but no distribution information is provided other than a general boundary $[0, 2]$, rendering the inverse model calibration task extremely challenging;
- Both aleatory and epistemic uncertainties are involved in the output features, rendering the failure probability of the system to be also governed by a P-box, which dramatically increases the calculation burden for the reliability-based optimization;
- The intractable trade-off between the risk to neglect a portion of epistemic uncertainty and the gain obtained from taking this risk is disturbed by the aleatory uncertainty, rendering the deterministic optimization to be prohibitive for the risk-based optimization.

To cope with the above challenges, the following theoretic developments are proposed in this paper to further develop both the quantification and propagation strategies:

- A feature extraction approach to first process the irregular and noisy time-domain sequences, and then to define a comprehensive UQ metric to extract as much as possible uncertainty information from the limited data;
- A parameterization hypothesis to represent the imprecise probability of the aleatory variables, such that the prior uncertainty model can be parameterized and calibrated through a model updating procedure;
- A two-level decoupling approach to investigate the P-box, and to propagate both aleatory and epistemic

uncertainties from the parameters to the output quantity of interest;

- The definition of the optimal criterion for the Subproblem D) Reliability-based design, the risk and gain criteria for the Subproblem F) Risk-based design, considering both aleatory and epistemic uncertainties to make sure the calculation cost is acceptable.

The above theories and methods are explained in Section 2. More attention and details are presented in Section 3 to give an elaborate presentation of the application of the method, and a detailed discussion of the results of each subproblem. Section 4 gives the conclusion and perspectives gained from addressing the NASA Challenge 2019 based on the developed systematical UQ approaches.

2 Theoretical development

Form the pre-investigation, the subproblems are categorized into forward procedure and inverse procedure. The sensitivity analysis and reliability analysis belong to the forward procedure, where the key investigation lies on how the aleatory and epistemic uncertainties propagate from the system inputs to the output quantity of interest. A forward approach to decouple the influence of aleatory and epistemic uncertainties in the P-box is proposed in Section 2.1. The uncertainty reduction, reliability-based and risk-based design belong to the inverse procedure where the UQ metrics and optimal criteria should be first defined based on the output features and inversely to tune the prior models and design. The UQ metrics definition and feature extraction are explained in Section 2.2, and subsequently, the inverse algorithm for uncertainty reduction and optimization are explained in Section 2.3.

2.1 Forward decoupling approach for uncertainty propagation

2.1.1 Parameterization hypothesis of the input P-box

Considering the readability, we employ the same terminology as the one used in the original document [1] of Challenge 2019. The model of the sub-system is expressed as

$$y(t) = y(a, e, t) \tag{1}$$

where $a \in \mathbb{R}^{n_a}$ and $e \in \mathbb{R}^{n_e}$ are the aleatory and epistemic variables, respectively; $y(t)$ is the output feature of the sub-system presented as a time-domain sequence. The uncertainty model (UM) of a is denoted as $a \sim f_a$ where f_a is the Probability Density Function (PDF); the UM of e is denoted as $e \sim E$, where E is the interval of the unknown-but-fixed parameter. While a and e are prescribed to be variables exclusively with aleatory or epistemic uncertainty, the output $y(t)$ is governed by both aleatory and epistemic uncertainties, and hence it is described as P-box.

Compared with e, the UM of a is more limited since the functional form of f_a is unknown. A parameterization process is hence necessary to characterize the unknown distribution using a series of quantitative coefficients. The parameterization approach is nonunique, and we propose to employ the Beta Mixture Model (BMM) to represent the UM of a. The basic Beta distribution is a continuous probability distribution defined on the interval [0, 1] with two positively real shape coefficients A and B:

$$\mathrm{Beta}(x, A, B) = \frac{\Gamma(A+B)}{\Gamma(A)\Gamma(B)} x^{A-1}(1-x)^{B-1} \tag{2}$$

where $\Gamma(.)$ is the gamma function $\Gamma(\blacksquare) = \int_0^\infty x^{\blacksquare - 1} e^{-x} dx$. The BMM is defined as the sum of N basic Beta

distributions, which is expressed as

$$\mathrm{BMM(x)} = \sum_{i=1}^{N} \beta_i \mathrm{Beta}(x, A_i, B_i) \tag{3}$$

where N is the number of basic Beta distribution employed, β_i is the weighting coefficients satisfying $\sum_{i=1}^{N} \beta_i = 1$. There are two reasons for selecting the BMM as the UM of a: 1) the format flexibility and 2) the interval definiteness. First, the BMM is qualified to present multiple distribution formats in the given interval. For example, as shown in Fig. 2, a BMM with two components can represent PDF curves with different numbers of poles and increasing/decreasing features. Second, since the variable a has been limited on the interval [0, 2] by the problem host, the distributions with unlimited boundaries, e.g. Gaussian distribution, are inapplicable. The standard Beta distribution is limited on [0, 1], and hence it is convenient to control the interval of BMM by adjusting the weighing coefficient β_i. In this problem, we propose the BMM with two components and adjusted weighting coefficients expressed as:

$$f_{a_i}\left(a, A_1^{(i)}, B_1^{(i)}, A_2^{(i)}, B_2^{(i)}\right) = \frac{1}{4}\mathrm{Beta}\left(\frac{1}{2}a, A_1^{(i)}, B_1^{(i)}\right) + \frac{1}{4}\mathrm{Beta}\left(\frac{1}{2}a, A_2^{(i)}, B_2^{(i)}\right) \tag{4}$$

The reason for proposing two components is that: On the one hand, the BMM with only one component is reduced to the normal Beta distribution and hence not flexible enough to fit the potential true distribution; On the other hand, the BMM with too many components will lead to too much unknown coefficients leading the updating process very difficult. As a result, the choice of the two-component BMM has a balance of the flexibility and complexity. As illustrated in Fig. 2, by changing the 4 coefficients, the two-component BMM can present different shapes of the PDF curves. This distribution hypothesis will make sure the boundary of a_i falls within in [0, 2], and the UM of a_i is parameterized using four parameters, $A_1^{(i)}, B_1^{(i)}, A_2^{(i)}, B_2^{(i)}$, which brings convenience for the following uncertainty reduction and model updating procedures.

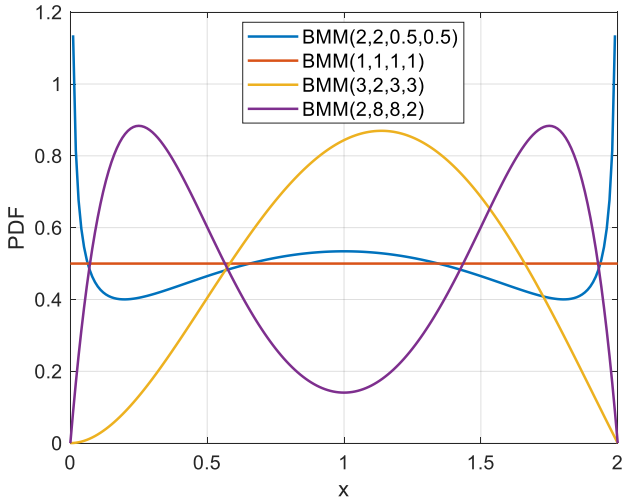

Fig. 2: Various formats of BMM PDFs with different configurations of shape coefficients

2.1.2 The two-loop approach for uncertainty propagation in the form of P-box

This subsection proposes a two-loop strategy to propagate both the aleatory uncertainty in a_i and the epistemic uncertainty in e_i to the output quantity through the black-box model, which will be used to solve the forward reliability analysis task in the Challenge. If only the aleatory uncertainty of the input is considered, the output quantities $y(a, e, t)$ and $z(a, e, \theta, t)$ are characterized as aleatory variables following certain probabilistic

distributions. However, because of the involvement of the epistemic uncertainty from e_i, the output quantities contain both the aleatory and epistemic uncertainties, and hence should be characterized as imprecise probabilities. In other words, they will be described as P-box in the proposed approach.

As the first step, in order to propagate both types of uncertainties, we propose a two-directional approach to investigate the uncertainty properties of a_i and e_i, as illustrated in Fig. 3. The uncertainty property of a_i is presented as a single Cumulative Distribution Function (CDF) curve as shown in Fig. 3(a), where the CDF curve is determined by the BMM distribution as proposed in Section 2.1.1. We observe the CDF curve along the vertical direction and truncate a range α from the [0, 1] probability space. Here the α presents a portion of aleatory uncertainty according to the single CDF curve, which projects to the horizontal axis obtaining an interval of the aleatory variable: $[\underline{a_i}, \overline{a_i}]$. The interval presents exclusively the aleatory uncertainty since it is the projection of a single CDF.

In Fig. 3(b), let's consider the uncertainty property of the epistemic variable e_i. Since e_i is prescribed as an unknown-but-fixed constant fallen within an interval, its "generalized" CDF is presented as a unit-impulse function at the position e_i^*. Considering the its interval $[\underline{e_i}, \overline{e_i}]$, infinite number of impulse functions constitute the shaded area, which can be regarded as a "generalized" P-box of the epistemic variable. Now consider a single probability point α^* and observe the P-box along the horizontal direction. Any α^* value results in the same interval projected on the horizontal axis $[\underline{e_i}, \overline{e_i}]$, which is governed only by the epistemic uncertainty.

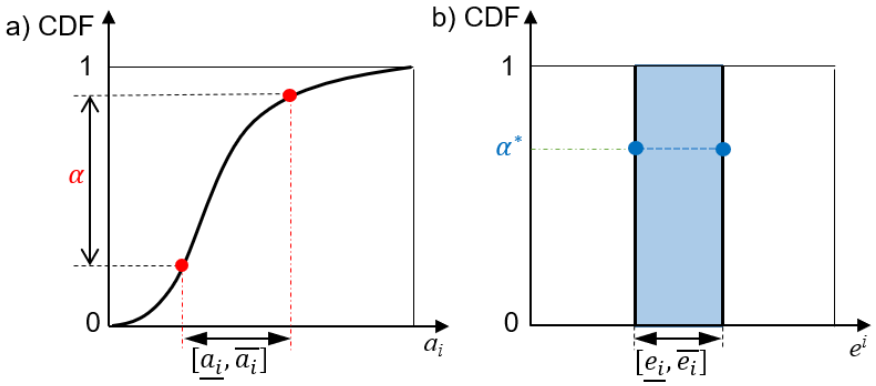

Fig. 3: The two-directional approach to investigate the P-box: (a) along the vertical direction for aleatory uncertainty; (b) along the horizontal direction for epistemic uncertainty

Based on the above two-directional investigation, the forward propagation approach is executed as a two-loop strategy to propagate the aleatory and epistemic uncertainties, respectively, by first a Monte Carlo sampling process for the probability point α^*, and an optimization process performed on each α^* to search the minimum and maximum of the output feature on the input interval $[\underline{e_i}, \overline{e_i}]$. This propagation approach was first proposed by the authors in Ref. [3] for sensitivity analysis. But this approach is proved to be efficient also for the reliability analysis in this problem.

The two-loop strategy is illustrated in Fig. 4. In the outer loop, the Monte Carlo sampling is executed on the complete aleatory space [0, 1] along the vertical axis, where a set of samples of the probability value α^* is obtained. For each of the aleatory value α^*, the corresponding interval $[\underline{e_i}, \overline{e_i}]$ is projected on the horizontal axis presenting exclusively the epistemic uncertainty. Within the epistemic interval, the inner loop is performed through an optimization process, where the $y_{max} = \underset{e \in E}{\operatorname{argmax}}\, y(x)$ and $y_{min} = \underset{e \in E}{\operatorname{argmin}}\, y(x)$ will be found. The function

$y(x)$ in the inner loop represents the sub-system model, as illustrated in Fig. 1, to calculate the time sequence D_1 from the aleatory variables a_i and epistemic variables e_i. Let n_α denotes the number of samples obtained from the outer loop, then the whole procedure requires totally n_α optimizations, which leads to considerable calculation cost. However, the optimizations are independent to each other, making the parallel calculation applicable to the inner loop, which significantly reduced the calculation time on the consumer multi-core computer.

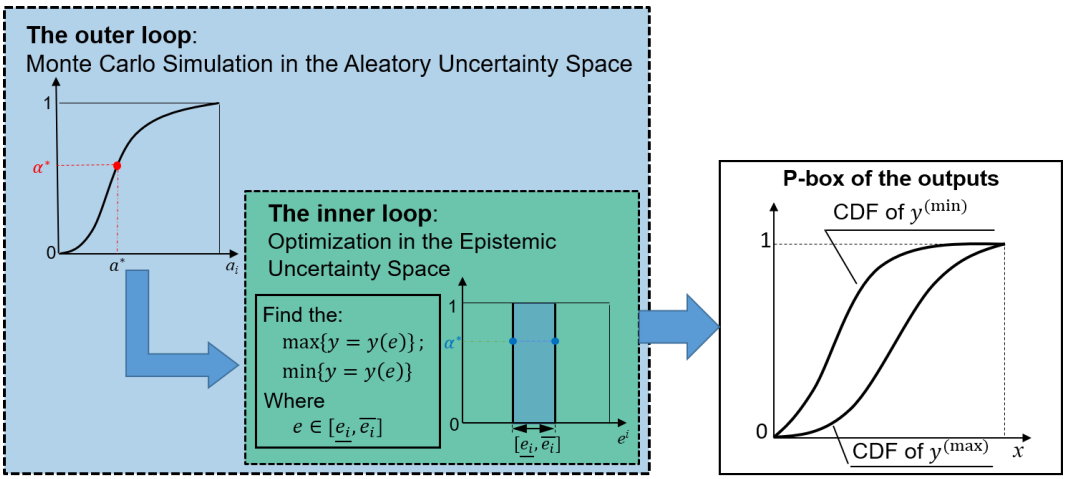

Fig. 4: The two-loop decoupling uncertainty propagation approach

The two-loop procedure results in two sets of the output samples, i.e. the maximum $y^{(\max)}$ and the minimum $y^{(\min)}$, with the number of samples n_α. The distributions of the maximum and minimum samples are fitted, respectively, and the CDF of these two distributions are plotted in the same plane, as illustrated in Fig. 4, where the P-box of the output feature is propagated from the input variables. The proposed strategy makes it possible for the forward reliability analysis to evaluate the range of the system failure probability when the epistemic uncertainty is involved.

2.2 Feature extraction and uncertainty quantification for time-domain variables

2.2.1 The Empirical Mode Decomposition

In this problem, the outputs of the system model are not scalars but time-domain sequences. The provided observations are limited (with only 100 samples), irregular (with multiple numbers of poles), and chaotic, see the data sets D_1 and D_2 in Fig. 1, rendering the direct usage of the data to extract uncertainty information prohibitive. In this case, pretreatment of the "gross" observation data is necessary, and we propose to employ the Empirical Mode Decomposition (EMD) method in this problem. The EMD, also known as Hilbert-Huang transform [4], is a featured method widely applied to time-domain signal processing. The purpose of EMD is to decompose a signal into so-called Intrinsic Mode Functions (IMF). An IMF is a component decomposed from the original signal that fulfils the following constraints: 1) An IMF is a function run through the horizontal axis multiple times with the number of poles and the number of zero crossings to the axis must be either equal or be differ at most by one; and (2) the envelop defined by the multiple maximums and minimums should be symmetric according to the horizontal axis.

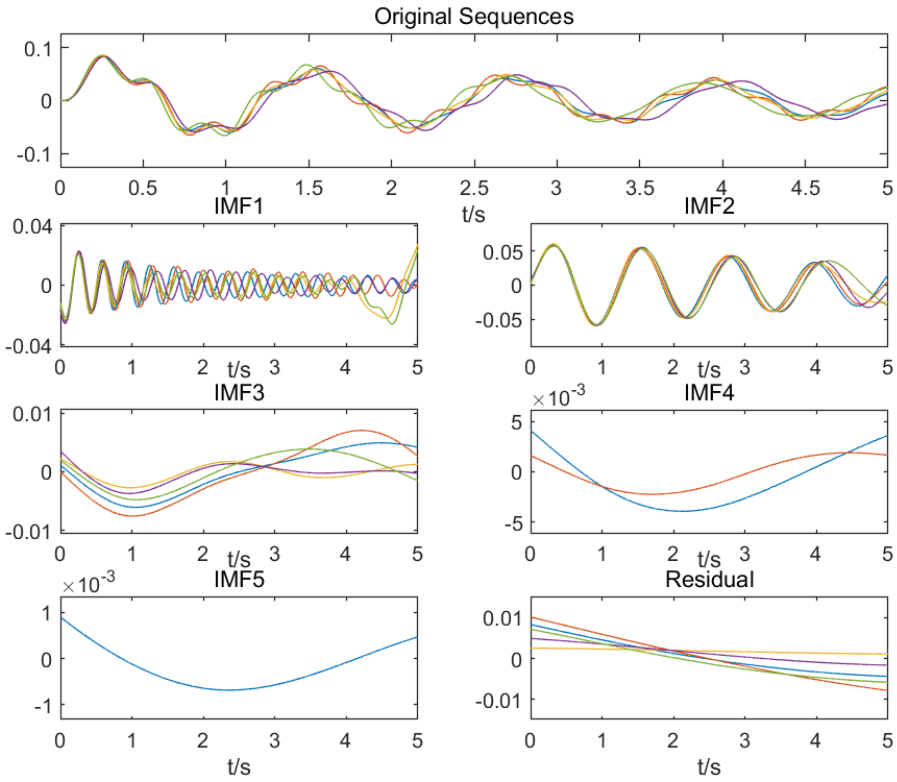

Fig. 5: The EMD results of 5 sets of observation

Therefore, compared to the original function, the IMF is more regular and has more explicit extrema so that the uncertainty feature will be captured more easily and comprehensively cover the whole-time domain. Fig. 5 illustrates the EMD results based on 5 original sequences which are randomly selected from the 100 samples. As a result, the uncertainty information will be extracted from the EMD components using the recently Uncertainty Quantification (UQ) metrics defined in the following subsection.

2.2.2 The Bhattacharyya distance: A comprehensive UQ metrics

The UQ metric is defined as a quantity measuring the difference between the provided observations and the simulation of the system model. The UQ metrics on this occasion is required to be as comprehensive as possible, such that the uncertainties in both the observation and simulation can be captured. The Bhattacharyya distance, the statistical distance measuring the overlap between two distributions [5], has been proved to be an efficient UQ metric in the application of stochastic model updating [6], especially, when solving the previous edition of the NASA Challenge 2014. The definition of the Bhattacharyya between two distributions is given as

$$d_B(y_{obs}, y_{sim}) = -\log\left\{\int_{\mathbb{Y}} \sqrt{P(y_{obs})P(y_{sim})}\mathrm{d}y\right\} \tag{5}$$

where the $P(y_{obs})$ and $P(y_{sim})$ are the PDF of the observational and simulated quantity. However, note that the quantity can only be a scalar such that the PDF can be estimated based on a number of samples. This means the original definition of the Bhattacharyya distance is only applicable to scalars. In the presence of continuous time-domain sequences, Eq. (5) cannot be directly employed. Consequently, a further feature extraction approach to transfer the time-domain sequence into several discrete quantities is proposed as illustrated in Fig. 6.

It is suggested to set a series of "time points" on the complete time range. The time $t_0 = 2.5s$ is chosen as an

example in Fig. 6. The left and right parts of Fig. 6 present respectively the observation and simulation curve sets of the 2nd IMF components, extracted by EMD method as shown in Fig. 5. The scalar samples under the timepoint 2.5s are obtained, which defines the disperse range of the sequence curves. Compared with the disperse range, it is more important to investigate the distribution property of the scalar samples within this range, where the Bhattacharyya distance can be naturally employed. Then more time points from 0*s* to 5*s* are chosen, and finally the mean value of the Bhattacharyya distances at those time points are defined as the quantification metric which describes the discrepancy between observations and predicted responses.

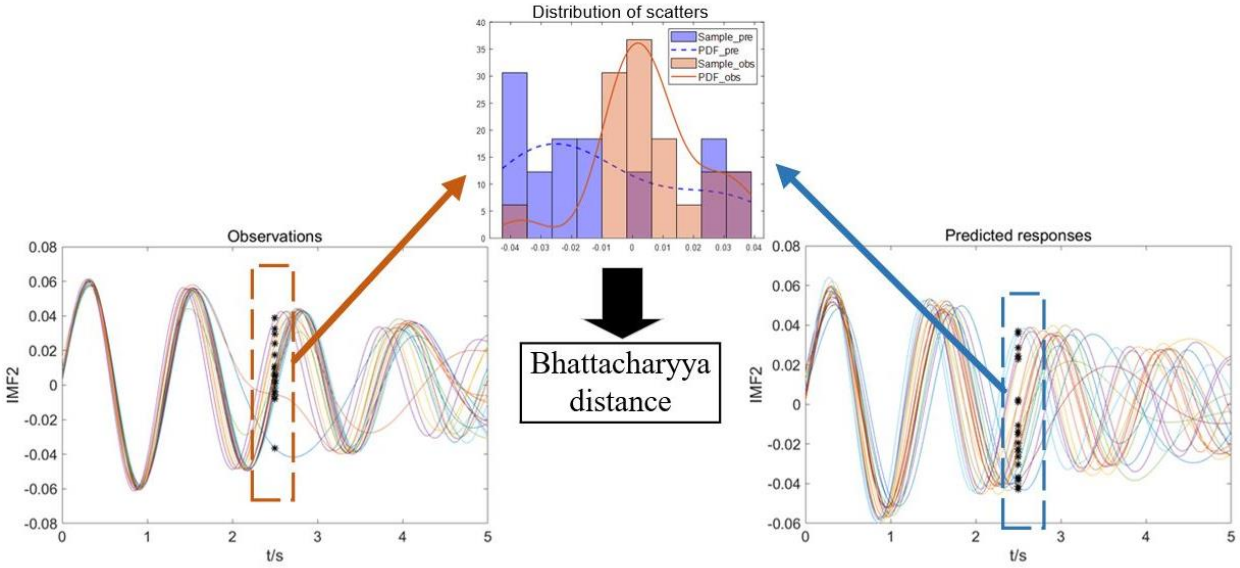

Fig. 6: Feature extraction approach on the sequence to employ the BD

2.3 Inverse uncertainty reduction and optimization

2.3.1 Bayesian model updating

In Subproblems A), B3), E2) and E4), it is required to update the Uncertainty Model (UM) of the input variables based on given observation data. This task is regarded as the inverse procedure. Various approaches are available for such an inverse procedure, for example, different optimization algorithms to minimize the discrepancy between the observation and model simulation. However, we propose to employ the Bayesian updating framework [6], since this approach is known to be specifically efficient for problems with limited observation data and prior knowledge of the input data. The foundation of Bayesian updating given as

$$P(x|y_{obs}) = \frac{P_L(y_{obs}|x)P(x)}{P(y_{obs})} \tag{6}$$

where $P(x)$ is the prior distribution of the calibrating parameter representing the prior knowledge to the system. Note that, the calibrating parameter x is not necessarily to be the actual input variable of the system model. For example, the actual input variables of the model $y(t) = y(a, e, t)$ is the aleatory variable a_i and epistemic variable e_i, and the calibrating parameter x refers to the distribution coefficients of a_i, i.e. $A_1^{(i)}, B_1^{(i)}, A_2^{(i)}, B_2^{(i)}$ in Eq. (4). It is also important to not confuse the prior distribution of $A_1^{(i)}, B_1^{(i)}, A_2^{(i)}, B_2^{(i)}$, and the actual probability distribution of a_i.

$P(x|y_{obs})$ in Eq. (6) is the posterior distribution of the calibrating parameter x, representing the updating result

based on the provided observation data. $P(y_{obs})$ is the normalization factor ensuring the posterior distribution integrates to one. $P_L(y_{obs}|x)$ is the likelihood function of y_{obs} for an instance of the parameters x. In this framework, the Approximate Bayesian Computation (ABC) method is utilized to provide a simplified version of $P_L(y_{obs}|x)$ using the Bhattacharyya distance-based feature:

$$P_L(y_{obs}|x) \propto exp\left\{-\frac{d^2}{\varepsilon^2}\right\} \tag{7}$$

where ε is the width factor determined to lie in the interval $[10^{-3}, 10^{-1}]$; the d is the distance-based UQ metric, in this paper specifically, will be the Bhattacharyya distance mentioned in Section 2.2.2. Furthermore, the Transitional Markov Chain Monte Carlo (TMCMC) along with the Metropolis-Hasting algorithm is employed to sample from some extremely complex posterior distributions. The TMCMC algorithm has been developed as a widely used approach for Bayesian updating. Not actually being the main scope of this paper, the detailed information of the algorithm can be found in the originating paper Ref. [7]. And the tutorial paper [8] is also suggested for the reader for more practical applications of the TMCMC algorithm.

2.3.2 Optimization tools selection

In Subproblems D1), E2), E4) and F2), the task of reliability- and risk-based design are presented with the requirement to find a new design point based on a meaningful optimal criterion, which is essentially an inverse and optimization process. The challenging features of the task lie on two aspects: 1) The optimization must be executed in a high dimension space, where the design parameter θ is a nine-dimensional variable with no boundary given; 2) the optimal criteria can only be defined based on the complicated and implicit P-box, leading the optimization extremely calculation-consuming. In this case, the optimization algorithm is required to have global search ability and derivative-free property. The Particle Swarm Optimization (PSO) method first proposed by Eberhart and Kennedy [9] is employed here because of its high speed of convergence and adaptability to the nonlinear functions in multidimensional space.

For the Subproblem F) Risk-based design requires to define the Risk to represent the portion of the epistemic uncertainty to be neglected, and the corresponding Gain resulting from taking the risk. The definition of the objective function, the gain, and the risk leading the optimization tremendously complex and expensive, when the search space is built with both aleatory and epistemic uncertainties. A surrogate optimization algorithm is consequently proposed to further release the calculation cost. This algorithm is a global optimization method proposed by Regis and Shoemaker [10] based on adaptive or sequential learning Radial Basis Function (RBF) which is perfectly suitable for the time-consuming objective function. The detailed definition of the objective function, the gain, and the risk will be elaborated in the following Subsection 3.6 where a clear explanation is expected only after the successful execution of the foregoing Subproblems A-E).

3 Problem investigation and outcome analysis

3.1 Subproblem A): Model calibration & UQ for the subsystem

3.1.1 Stochastic model updating process

In Subproblem A), the main objective is to calibrate the UM of both the aleatory variables a_i and epistemic variables e_i based on the limited number of observations D_1. This UM calibration task belongs to the classical topic known as stochastic model updating, which is regarded as the inverse procedure. Three key steps are required to

solve this problem:

i) Parameterization hypothesis of the aleatory variable a_i based on the Beta Mixture Model (BMM) described in Section 2.1.1;

ii) Feature extraction based on the Bhattacharyya distance as described in Section 2.2;

iii) Parameter calibration employing the Bayesian updating framework as described in Section 2.3.1.

The epistemic variable e_i is prescribed as the unknown-but-fixed constant within the pre-defined interval [0, 2]. And hence the e_i is directed calibrated in the Bayesian updating process. For the aleatory variable a_i, the BMM hypothesis is applied, and hence the UM of each a_i is parameterized for calibrating parameters $A_1^{(i)}, B_1^{(i)}, A_2^{(i)}, B_2^{(i)}$. The UM parameterization hypothesis for both aleatory and epistemic variables are summarized in Table 1, where the total number of calibrating parameters are 24 (four parameters for each of the five a_i plus four e_i). In Bayesian updating, the prior distributions of these calibrating parameters are assumed to be uniform on the pre-defined intervals as shown in Table 1.

Table 1: Prior distributions of aleatory and epistemic variables

Parameter	Uncertainty model hypothesis	Calibrating parameters and intervals
$a_i,\ i = 1, \ldots, 5$	$a_i \sim BMM(A_1^{(i)}, B_1^{(i)}, A_2^{(i)}, B_2^{(i)})$	$A_1^{(i)}, B_1^{(i)}, A_2^{(i)}, B_2^{(i)} \in [0.01, 30]$
$e_i,\ i = 1, \ldots, 4$	Unknown-but-fixed constant	$e_i \in [0, 2]$

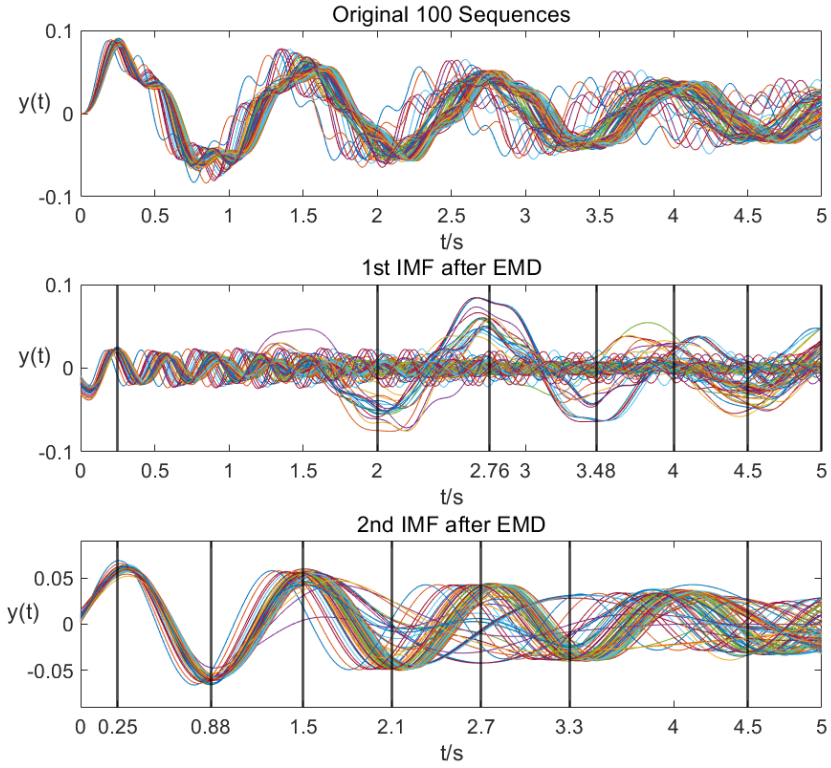

Fig. 7: The selected time points to investigate the 1st and 2nd IMF of the sequences

The EMD method and the Bhattacharyya distance are employed to quantify the discrepancy between observations and simulated sequences. All the 100 sequence curves are employed in the updating process. As

illustrated in Fig. 7, the 1st and 2nd EMD components are employed at the selected time points ([0.25s, 2.00s, 2.76s, 3.48s, 4.00s, 4.50s, 5.00s] for IMF1 and [0.25s, 0.88s, 1.50s, 2.10s, 2.70s, 3.30s, 4.50s] for IMF2), implying 14 sets of feature samples are employed to calculate the Bhattacharyya distances between the observation and simulation. There is no unique standard to select the featured time points. The general rule is that the most obvious local maxima and minima throughout the time sequence should be covered. The mean of the 14 Bhattacharyya values is employed to evaluate the likelihood function in Eq. (7), and after 12 iterations the TMCMC algorithm converged.

After the Bayesian updating process, the posterior distributions of the 24 calibration parameters are presented in Fig. 8. For the ahead 20 parameters ($A_1^{(i)}, B_1^{(i)}, A_2^{(i)}, B_2^{(i)}, i = 1, \dots, 5$), their exact values are estimated by taking the maximum pole of the posterior PDF, representing the position where the likelihood gets the maximum value in the Bayesian updating. The estimated values are listed in Table 2, which are essentially the distribution coefficients of the BMM model. As a result, the actual BMM distributions of the aleatory variables a_i are plotted in Fig. 9.

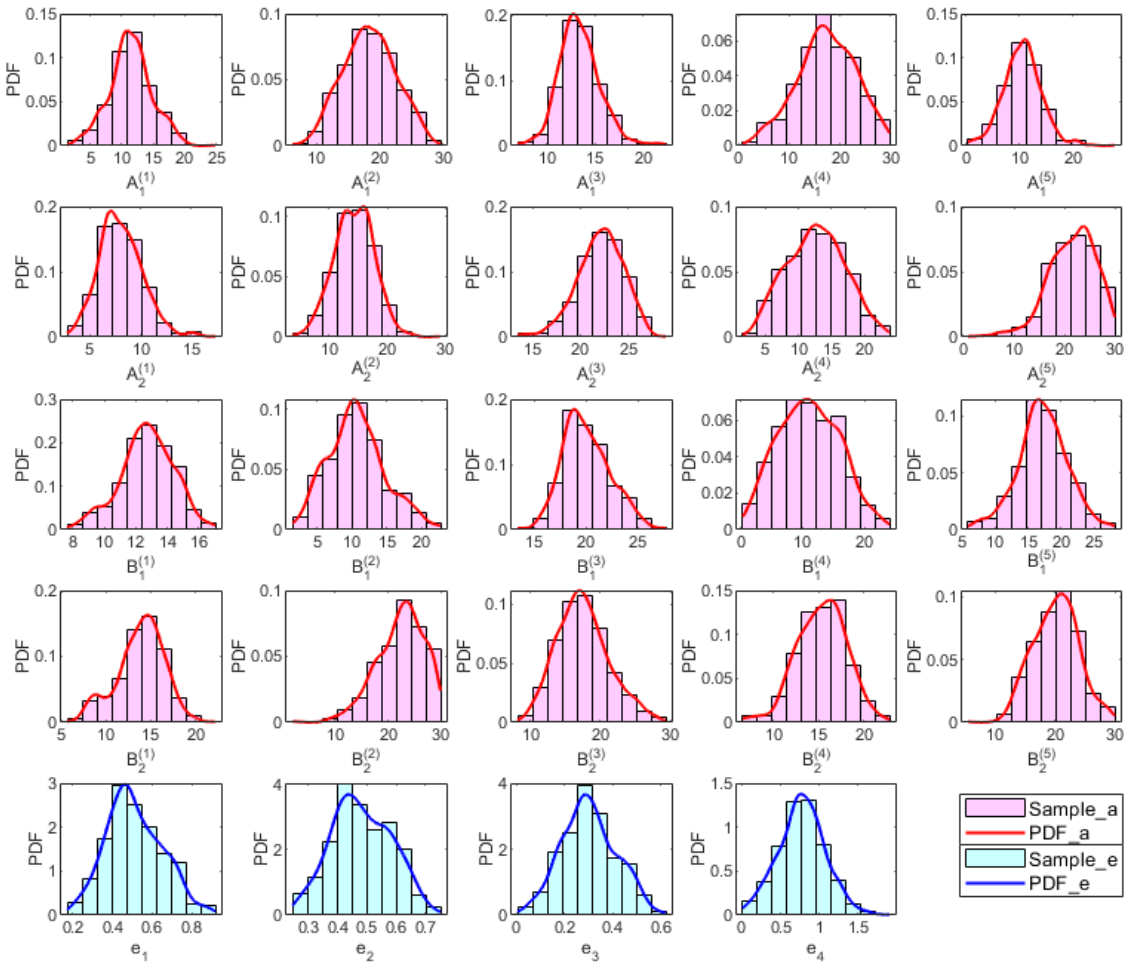

Fig. 8: Posterior distribution of the 24 calibrating parameters

Table 2: The calibrated UM of the input variables

Uncertainty Model of the input variables								
	Aleatory variables					Epistemic variables		
	a_1	a_2	a_3	a_4	a_5		Original interval	Reduced interval
$A_1^{(i)}$	10.8476	17.8747	12.7699	16.5293	11.0813	e_1	[0, 2]	[0.2103, 0.8185]
$B_1^{(i)}$	12.6593	10.2992	18.8625	10.6599	16.5631	e_2	[0, 2]	[0.2528, 0.7105]
$A_2^{(i)}$	7.0381	15.8893	22.5654	12.6301	23.7887	e_3	[0, 2]	[0.0521, 0.5510]
$B_2^{(i)}$	14.6622	23.4767	17.0233	16.2537	21.0530	e_4	[0, 2]	[0.0788, 1.4256]

For the last four calibrating parameters, i.e. the four epistemic variables of the model ($e_i, i = 1, \ldots, 4$), further treatment is required to estimate their intervals, as illustrated in Fig. 10. The posterior PDFs are estimated employing the well-known Kernel Density Estimation (KDE) approach based on the posterior samples obtaining from the Bayesian updating. The PDFs are normalized within the range [0, 1]. It is suggested to define a so-called "cutting ratio", denoted as α, which lies in the range [0.0, 1.0]. As shown in Fig. 10, a larger α results in a smaller truncated interval. In the extreme case, when $\alpha = 0.0$, we will get the largest interval in the bottom of the subfigures in Fig. 10; when $\alpha = 1.0$, the single value of e_i is obtained located at the maximum pole as shown in Fig. 10. For the epistemic variables, we expect to obtain larger intervals to make sure the reduced epistemic intervals can still include the true values of e_i. Consequently, $\alpha = 0.1$ is employed and the resulting intervals are listed as in Table 2. Note that, although we expect to obtain relatively large intervals of e_i, compared with their pre-defined interval [0, 2], the updated intervals in Table 2 are nevertheless very narrow, implying the epistemic uncertainty space has been dramatically reduced by the Bayesian updating framework.

Till now, we already get the explicit distributions of the aleatory variables a_i and the reduced intervals of the epistemic variables e_i. In other words, the UM of the input uncertain variables has been updated based on the provided observation sequences D_1.

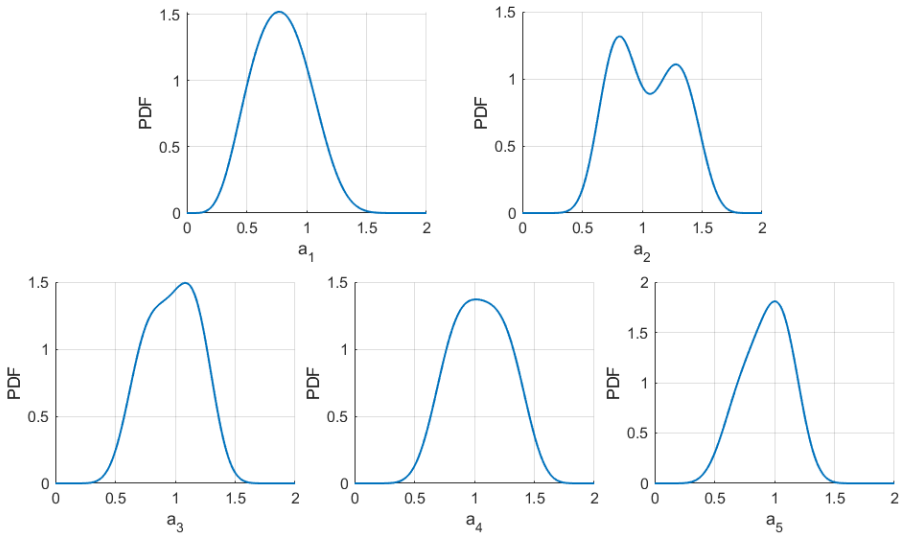

Fig. 9: Posterior PDF of aleatory variables

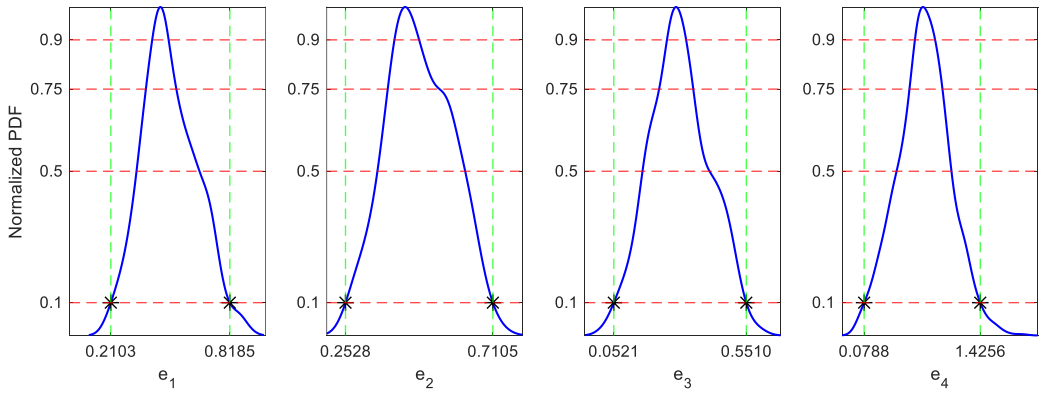

Fig. 10: Normalized PDFs of 21-24 calibrating parameters (the epistemic variables of the model)

3.1.2 Assessment of the UM calibration results

In this section, further assessment of the calibrating results is performed focusing on the output features of the sub-system model $y(a,e,t)$. Fig. 11 illustrates the cross-comparison among the observation sequences, original sequences, and calibrated sequences. The original sequences are obtained by randomly sampling input variables a_i and e_i from the original boundary [0, 2]. The comparison shows that the original output sequences are obviously disparate from the observations, while the calibrated sequence shows relative similarity with the target data.

To provide a more informative assessment of agreement between the calibrated sequences and the observation sequences, we select specific time points on the time-domain and take a similar treatment as the one performed in Fig. 7. But here the time points are taken directly from the original sequences (no longer from the IMFs after EMD), and the distribution properties of the captured values are compared in Fig. 12. The left part of Fig. 12 presents the two-dimensional scatters in the plane of values according to time points 1.0s and 2.0s; the right part of Fig. 12 presents the two-dimensional scatters in the plane of values according to time points 3.0s and 4.0s.

The samples in Fig. 12 are extracted from the output time sequences according to different time points. For example, the observation scatter has 100 samples because there are 100 sequences in the observation data set D_1. We employ the Gaussian Mixture Model (GMM) to fit the joint distribution of the scatters, and the contour with 0.05 normalized height is plotted as the estimated intervals. In other words, the interval in Fig. 12 represents the 95% confidence interval of the scatter based on the estimated GMM distribution. This is why some samples fall outside the estimated intervals. For the observation interval in the left part of Fig. 12, only one component of GMM is used, and hence the interval appears as a standard ellipse. For the other intervals, two components of GMM are employed, and hence the intervals have complicated shapes. Both planes show that the original scatters are obviously different from the target scatters (the observation data), while the calibrated scatters have been tuned to be relatively close to the observations.

The quantitative assessment of the calibration effect is performed by evaluating Bhattacharyya distances between the original, calibrated, and observation sequences, as listed in Table 3. The first two columns of the tables present the Bhattacharyya distances of original sequences vs. observations and calibrated sequences vs. observations. It is shown no matter the distances in specific time points or the mean are clearly reduced by the Bayesian calibration process.

The last column of Table 3 is presented to address Subproblem A2) which requires to analyze the impact of n_1 on the calibration effect. The n_1 is defined in the original document [1] as the number of curve samples (from the observation $D1$) employed in the calibration process. The above calibration results are obtained based on all the

provided samples, that it, $n_1 = 100$. We repeat the Bayesian updating by employing 50 samples, and the Bhattacharyya distances of the results are presented in the last column. It is natural that as the observation samples are more limited, the uncertainty, i.e. distribution information, is more difficult to be captured, and hence the Bhattacharyya distances between the calibrated sequences and the observation are relatively larger than the ones using 100 observation samples.

Table 3: The Bhattacharyya distances between observations and simulated sequences

	t=1s	t=2s	t=3s	t=4s	t=5s	Mean
BD between original sequences and observations	0.1761	0.2756	0.1118	0.3081	0.1292	0.2255
BD between calibrated sequences and observations	0.0495	0.0594	0.0943	0.1028	0.0616	0.0749
BD with reduced $n_1 = 50$	0.1711	0.0536	0.0862	0.0591	0.0641	0.0806

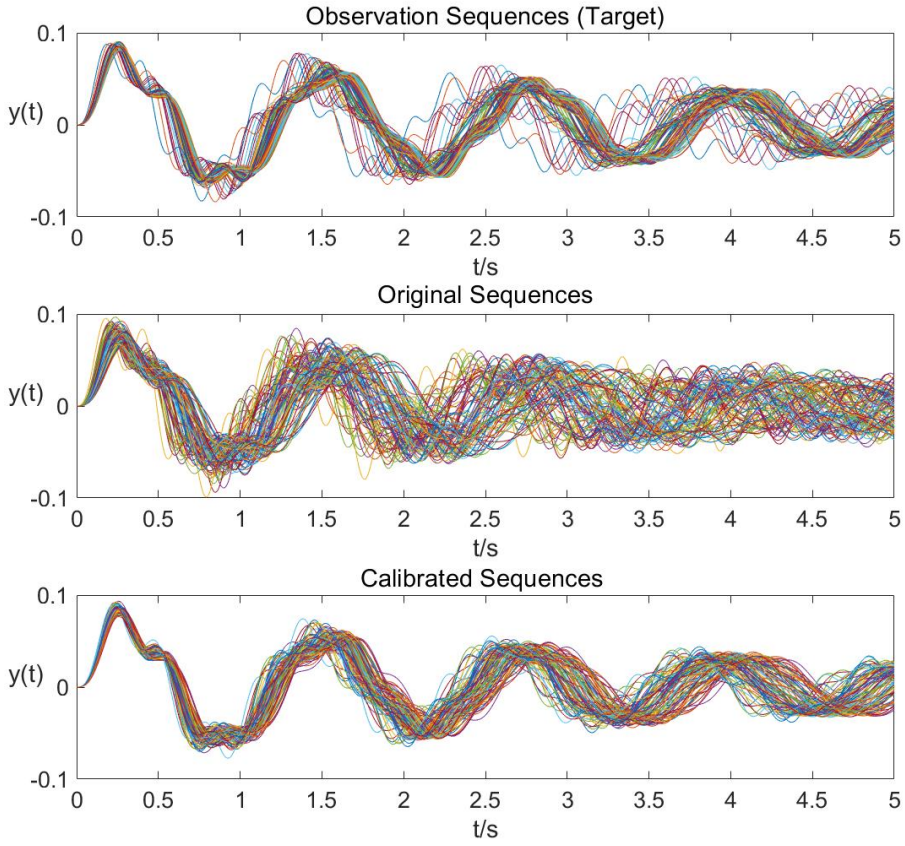

Fig. 11: The cross-comparison among the observations, original sequences, and calibrated sequences

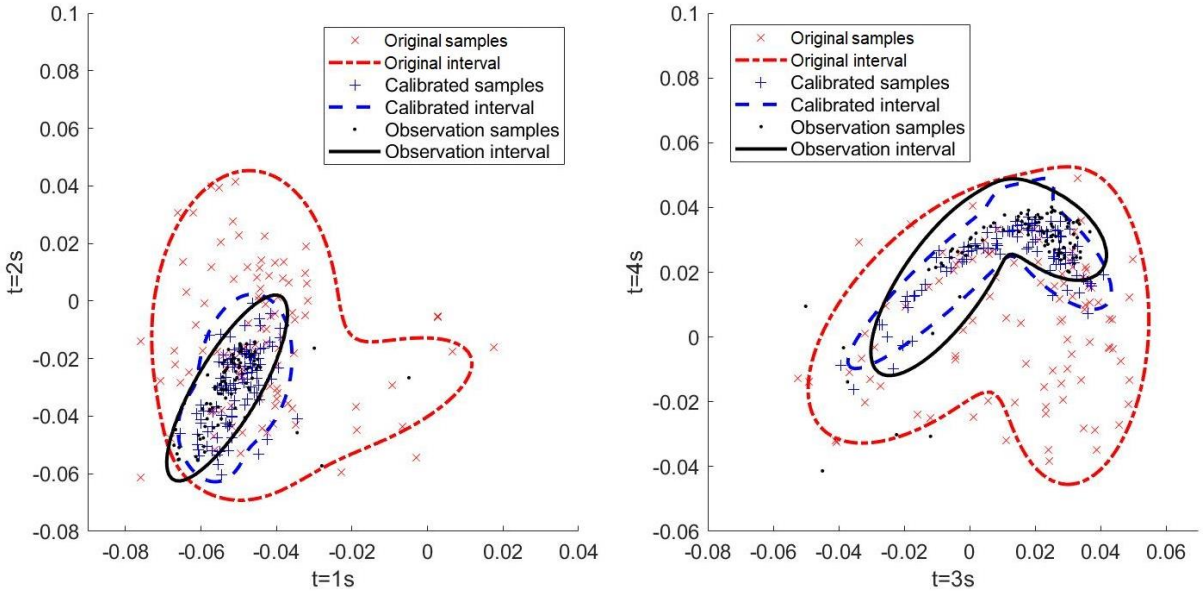

Fig. 12: Scatters of the original, calibrated, and observation values of the sequences on specific time points

3.2 Subproblem B): Uncertainty reduction

This subproblem includes three tasks:

i) Sensitivity analysis to rank the epistemic variables e_{1-4} according to their influence on the system model prediction;

ii) Interaction with the problem host to ask for uncertainty reduction based on the result of sensitivity analysis;

iii) Repeat the model calibration in Subproblem A) to further updated the UM of the input variables.

To address the 1st task, it is important to define a ranking criterion based on how much the uncertainty space of the output feature can be reduced when the epistemic uncertainty of each variable e_i is reduced. Considering the sub-system model $y(a,e,t)$, since a_i involves aleatory uncertainty and e_i involves epistemic uncertainty, the output feature $y(t)$ will be a quantity that involves both aleatory and epistemic uncertainty, and presented as a time-domain sequence. As a result, the degree of dispersion of the sequence along the time-domain can be regarded as a criterion of the influence of the epistemic uncertainty. We use $D(t)$ to denote the degree of dispersion of the sequence in time-domain, and it is evaluated as follows

$$D(t) = \max_n y(t) - \min_n y(t) \tag{8}$$

where n is the number of samples of the sequences. In other words, $D(t)$ is presented as an envelope of the total samples of time-domain sequences. When epistemic uncertainties of all e_{1-4} are not reduced, we obtain the largest envelope, denoted as $D_0(t)$. $D_0(t)$ is obtained by randomly generating 10000 samples of e_{1-4} from their original boundary [0, 2], and 10000 samples of a_{1-5} from their calibrated distributions as shown in Fig. 9. The sensitivity of each e_i is defined as the reduction from the full envelope $D_0(t)$ to the $D_i(t)$ obtained by reducing the original interval of e_i to an explicit value. The sensitivity is expressed as

$$\Delta D_i(t) = D_0(t) - D_i(t) \quad 0s \le t \le 5s \tag{9}$$

$D_i(t)$ is obtained by fixing e_i to its explicit value and keeping other three still free within the boundary [0, 2].

The explicit value of e_i is determined based on their posterior PDFs in Fig. 10 by taking the cutting ratio $\alpha = 1.0$. In other words, the kept value represents the maximum possible point of e_i based on the model calibration in Section 3.1. The $\Delta D_i(t)$, as well as the envelops before and after the reduction are illustrated in Fig. 13. Scales of the boundaries can be referred from the left-vertical axis of each subfigure, and scales of the sensitivity $\Delta D_i(t)$ can be referred from the right-vertical axis of each subfigure. It can be observed that the envelop boundaries before reduction (plotted in red dot line) and after reduction (plotted in blue and solid line) are nearly overlap for e_3 and e_4. As a comparison, the boundaries before and after reduction for e_2 are clearly different. This means the sensitivity of e_2 is clearly larger than the other epistemic variables. As a quantitative measure, we evaluate the average value of $\Delta D(t)$ in 5 seconds, denoted as $\overline{\Delta D(t)}$, and present it in Table 4. It's obvious that the $\overline{\Delta D(t)}$ for e_2 is the largest. The final ranking regarding $\overline{\Delta D(t)}$ is obtained as $e_2 > e_1 > e_3 \approx e_4$. Note that the $\overline{\Delta D(t)}$ for e_3 and e_4 are nearly zero, hence no meaningful order of e_3 and e_4 is provided based on the current definition of ranking criterion.

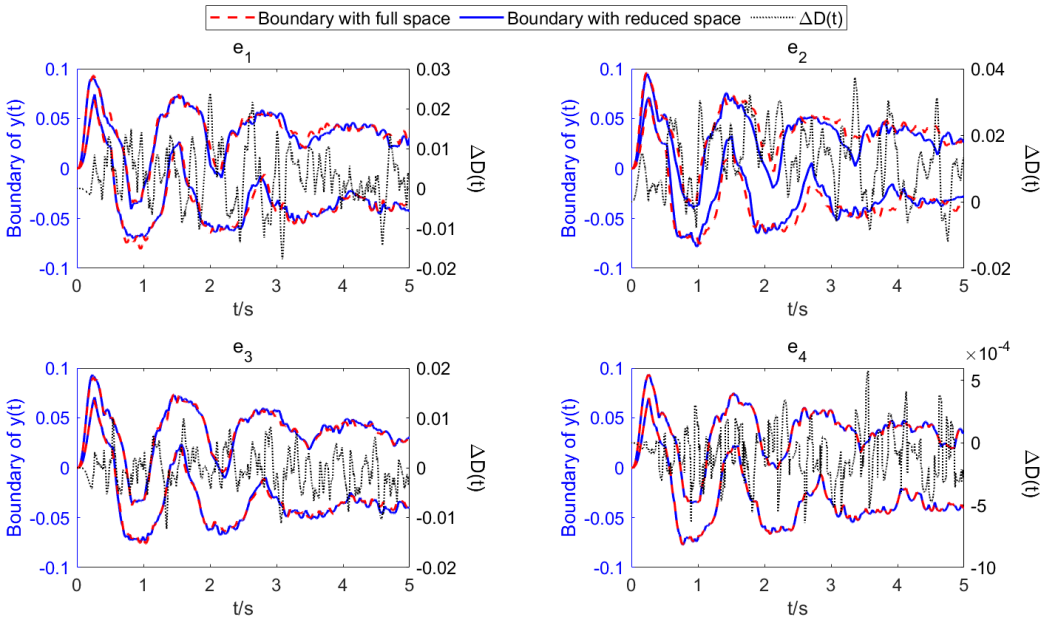

Fig. 13: The boundaries of sequences (left y-axis) and the $\Delta D(t)$ (right y-axis) before and after epistemic uncertainty reduction

Table 4: $\overline{\Delta D(t)}$ of each epistemic variable

Epistemic variable	e_1	e_2	e_3	e_4
$\overline{\Delta D(t)}$	0.0029	0.0125	-0.0004	-0.0001

After the sensitivity analysis, we request to the Challenge host for the refined UM (i.e. the further reduced interval of e_i) and then we repeat the Bayesian updating process using the new interval. The further calibrated BMM distributions of the aleatory variables a_i are illustrated in Fig. 14, with the comparison with the ones before the UM is refined. The estimated values of the BMM coefficients of a_i and the reduced intervals of e_i are listed in Table 5. The resulting Bhattacharyya distances are presented in Table 6. Because of the new interval, the Bayesian updating process obtains a better calibration effect, which can be reflected by the even smaller Bhattacharyya distances compared with the calibrated ones in Subproblem A).

Fig. 15 illustrates the comprehensive comparison among the scatters from four sets of data, namely, the original,

calibrated before reduction of UM, calibrated after the reduction of UM, and the observation data. The generation of the samples and intervals follows the same treatment as the one taken for Fig. 12. It can be observed that, especially from the left plane, the calibrated scatters after the reduction of UM are further close to the target observations, which once again demonstrates the significance of the interval reduction to the calibration effect.

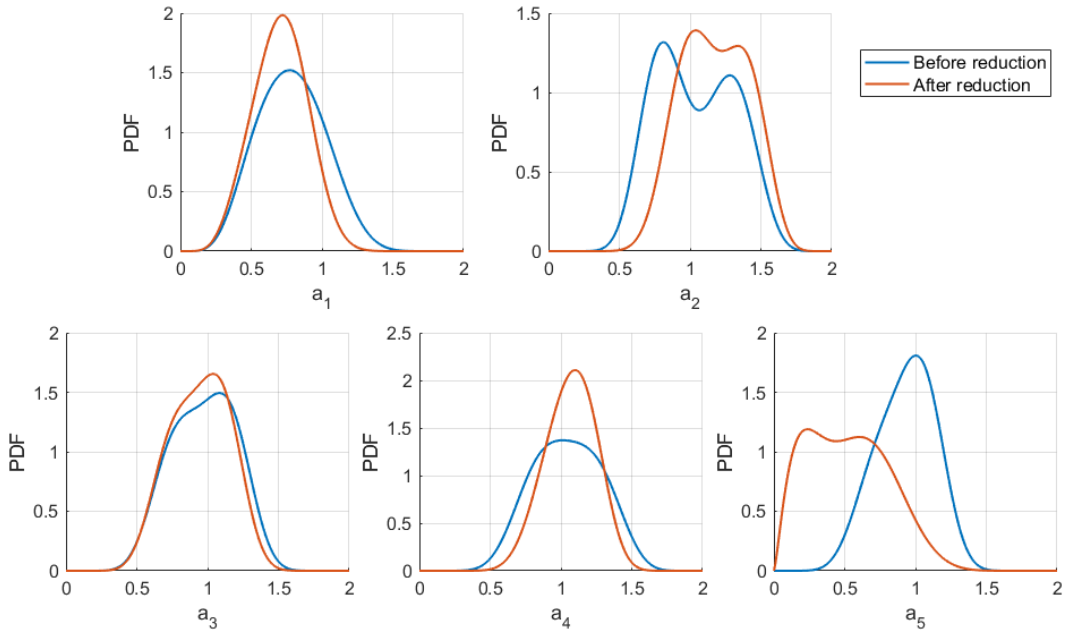

Fig. 14: The calibrated BMM distributions of a before and after the uncertainty reduction of UM

Table 5: Calibrated UM of the input variables after the reduction of UM

Uncertainty models of the input variables							
	Aleatory variables					Epistemic variables	
	a_1	a_2	a_3	a_4	a_5		Reduced interval
$A_1^{(i)}$	7.2873	21.8143	25.3483	17.2455	2.3861	e_1	[0.3346, 0.6453]
$B_1^{(i)}$	16.3419	10.3400	20.8685	18.2005	12.3553	e_2	[0.5433, 0.7312]
$A_2^{(i)}$	14.8134	19.4945	13.9479	26.6700	7.2420	e_3	[0.0393, 0.5555]
$B_2^{(i)}$	22.4664	19.7355	21.0413	19.1393	12.4983	e_4	[0, 1.0372]

Table 6: The Bhattacharyya distances before and after the reduction of UM

	t=1s	t=2s	t=3s	t=4s	t=5s	mean
BD between calibrated sequences and observations with original interval	0.0495	0.0594	0.0943	0.1028	0.0616	0.0749
BD between calibrated sequences and observations with reduced interval	0.0461	0.0747	0.0624	0.0852	0.0536	0.0586

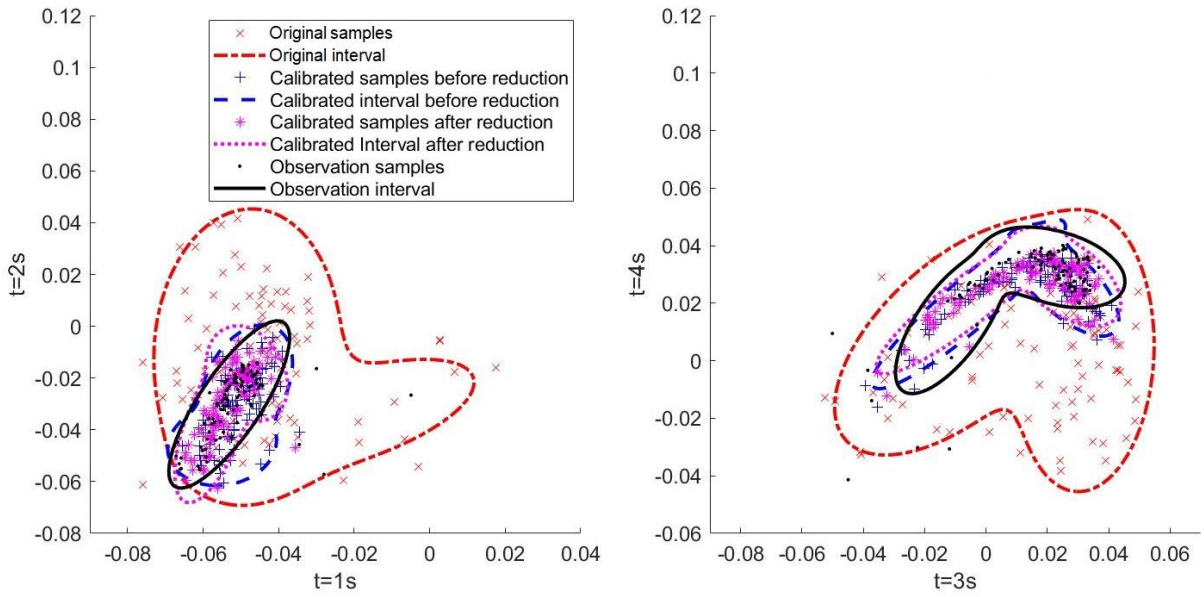

Fig. 15: The samples of responses before and after the reduction

3.3 Subproblem C): Reliability analysis of baseline design

From this subproblem, the integrated system model $z(a,e,\theta,t)$ is considered. Based on the outputs of this model, a series of performance functions are defined. $g_1(a,e,\theta)$ is a built-in function provided along with the problem. g_2 and g_3 are respectively defined as

$$g_2 = \max_{t\in[\frac{T}{2},T]} |z_1(a,e,\theta,t)| - 0.02 \tag{10}$$

$$g_3 = \max_{t\in[0,T]} |z_2(a,e,\theta,t)| - 4 \tag{11}$$

where z_1 and z_2 are the output features of the integrated system model $z(a,e,\theta,t)$ presented in time-domain. The worst-case performance function is defined as

$$w(a,e,\theta) = \max_{i=1,2,3} g_i(a,e,\theta) \tag{12}$$

Based on the performance functions, the reliability analysis is performed with regard to three metrics, namely, the failure probability for each individual requirement g_i, i.e. $R_i(\theta)$; the failure probability for all requirements g_{1-3}, i.e. $R(\theta)$; and the severity of each individual requirement violation g_i, i.e. S_i, defined as follows, respectively.

$$R_i(\theta) = \left[\min_{e\in E} \mathbb{P}\left[g_i(a,e,\theta) \geq 0\right], \max_{e\in E} \mathbb{P}\left[g_i(a,e,\theta) \geq 0\right]\right] \tag{13}$$

$$R(\theta) = \left[\min_{e\in E} \mathbb{P}\left[w(a,e,\theta) \geq 0\right], \max_{e\in E} \mathbb{P}\left[w(a,e,\theta) \geq 0\right]\right] \tag{14}$$

$$S_i = \max_{e\in E} \mathbb{E}[g_i | g_i \geq 0]\, \mathbb{P}[g_i \geq 0] \tag{15}$$

where $\mathbb{P}[\cdot]$ is the probability operator, $\mathbb{E}[\cdot\,|\,\cdot]$ is the conditional expectation; E is the whole epistemic uncertainty

space of the epistemic variables e_i. To calculate those metrics, we employ a two-loop strategy to search the minimum and maximum of the failure probability on the whole space E (see Fig. 4). To ensure the fidelity and computational feasibility, the Monte Carlo sampling size in the outer loop is set to be 1000, and the failure probability for each realization in the inner loop is also determined by Monte Carlo simulation with 1000 samples on the BMM distribution of a_i. The corresponding range of failure probability and severity are listed in Table 7.

Table 7: The result of reliability metrics based on current UM and $\theta_{baseline}$

Range of failure possibility				Severity		
R_1	R_2	R_3	R	S_1	S_2	S_3
[0.021, 0.288]	[0.009, 0.392]	[0, 0.016]	[0.025, 0.404]	0.0952	0.0019	0.0055

Table 8: The range of the failure probability with the reduction of the specific epistemic variables (% percentage contraction ratio compared with the full interval before any reduction)

Reduction variable	Range of R_1	Range of R_2	Range of R_3	Range of R	Mean contraction ratio
No reduction	[0.021, 0.288]	[0.009, 0.392]	[0, 0.016]	[0.025, 0.404]	--
e_1	[0.027, 0.225] (-25.8%)	[0.026, 0.290] (-31.1%)	[0, 0.011] (-31.3%)	[0.043, 0.304] (-31.1%)	-29.9%
e_2	[0.018, 0.210] (-28.1%)	[0.018, 0.342] (-15.4%)	[0, 0.014] (-12.5%)	[0.029, 0.347] (-16.1%)	-18.0%
e_3	[0.015, 0.219] (-23.6%)	[0.018, 0.303] (-25.6%)	[0, 0.012] (-25.0%)	[0.036, 0.307] (-28.5%)	-25.7%
e_4	[0.021, 0.246] (-15.7%)	[0.008, 0.389] (-0.5%)	[0, 0.015] (-6.3%)	[0.028, 0.393] (-3.7%)	-6.6%

The next task in the subproblem is the sensitivity analysis to rank the epistemic variables e_{1-4} according to the ratio of contraction of $R(\theta)$ that might result from the reduction of the epistemic variable e_i. Like what has been done in Subproblem B), we respectively reduce the epistemic uncertainty of each variable from its full interval to the explicit value meanwhile keeping the other 3 variables unchanged. The approach to determine the explicit value is the same as the one taken in Subsection 3.2. Then calculate the range of failure probability using the reduced epistemic intervals and the results are shown in Table 8. The first row presents the full interval when no epistemic uncertainty is reduced. The following four rows are the range of the failure probability according to the reduction of a specific epistemic variable. Clearly, the reduction of any epistemic variable leads to the contraction of the range, compared with the full ranges before reduction. The last column of Table 8 provides the mean contraction ratio,

implying the ranking of epistemic variables according to the contraction of $R(\theta)$ should be $e_1 > e_3 > e_2 > e_4$.

3.4 Subproblem D): Reliability-based design

In this subproblem, it is required to find a new design θ_{new} to improve the system's reliability. It's essentially an optimization process and thus it's important to choose an optimality criterion that can represent reliability and also have computational feasibility. To make sure the optimal solution can actually improve the system's reliability, the worst-case failure probability is taken as the basis to define the optimality criterion. However, directly using the metrics in Subproblem C) will bring a huge computational complexity, because the probability of failure should be described as a P-box. Each evaluation of the forward procedure through the two-loop strategy to estimate the P-box is extremely calculation consuming, leading the normal optimization computationally prohibitive. Consequently, we propose a simplified version of the objective function defined as:

$$f(\theta) = \max_{e \in E, a \sim f_a} \mathbb{P}[w(a, e, \theta) \geq 0] \tag{16}$$

This objective function is defined as the maximum of the failure probability according to the worst-case performance function in Eq. (12). In the reliability-based design, our objective is to minimize the objective function, i.e. to minimize the maximum of the failure probability. In the simplified objective function, instead of the two-loop strategy, the aleatory variables a_i and epistemic variables e_i are sampled subsequently in a single loop. The first step is to generate a certain number (denoted as N_e) of samples of e_i. For each sampled e_i, generate a certain number (denoted as N_a) of a_i, and a failure probability can be calculated based on these N_a samples of a_i. The maximum value from the N_e failure probabilities is finally identified as the maximum failure probability. By minimizing the failure probability, i.e. the objective function, the new design θ with improved system reliability will be found. The Particle Swarm Optimization (PSO) algorithm is employed here to find the new design. Furthermore, the parallelization calculation strategy is employed herein to dramatically reduce the calculation time, because the optimizations are independent to each other. However, when employing the PSO algorithm, a search boundary of θ is required. The search boundary is set to be $\pm 50\%$ of the provided θ_{baseline} to ensure the convergence.

Table 9: The result of reliability metrics based on θ_{new}

Range of failure possibility				Severity		
R_1	R_2	R_3	R	S_1	S_2	S_3
[0, 0.021]	[0, 0.036]	[0, 0.025]	[0, 0.038]	0.0068	0.0001	0.0020

After determining the θ_{new}, it's necessary to repeat the reliability analysis as performed in Subproblem C) employing the optimized θ_{new} instead of $\theta_{baseline}$, so that the optimization effect on the system reliability can be assessed. The new range of the failure probability and severity are listed in Table 9. Compared with the results before optimization (in Table 7), a huge contraction of the range of failure probability is observed. The width of the range of R_1, R_2, and R are reduced -91.2%, -92.3%, -90.31%. However, it's also noted that the length of R_3 is not reduced. But considering the big contraction of R, it is reasonable to conclude that the optimum θ_{new} improves the system reliability significantly. The output features (Z_1 and Z_2) regarding θ_{baseline} and θ_{new} are shown in Fig. 16, where the improvement of system reliability is illustrated intuitively. Based on the Eqs. (10) and (11), the system

failure thresholds (0.02 and 4) are plotted in the figure. The left subfigure shows the system responses before optimization where multiple sample curves of Z_1 have exceeded the threshold, and some sample curves of Z_2 are tending to divergency. In contrast, the new design after optimization in the right subfigure presents satisfactory steady and convergence in the whole time-domain for both Z_1 and Z_2.

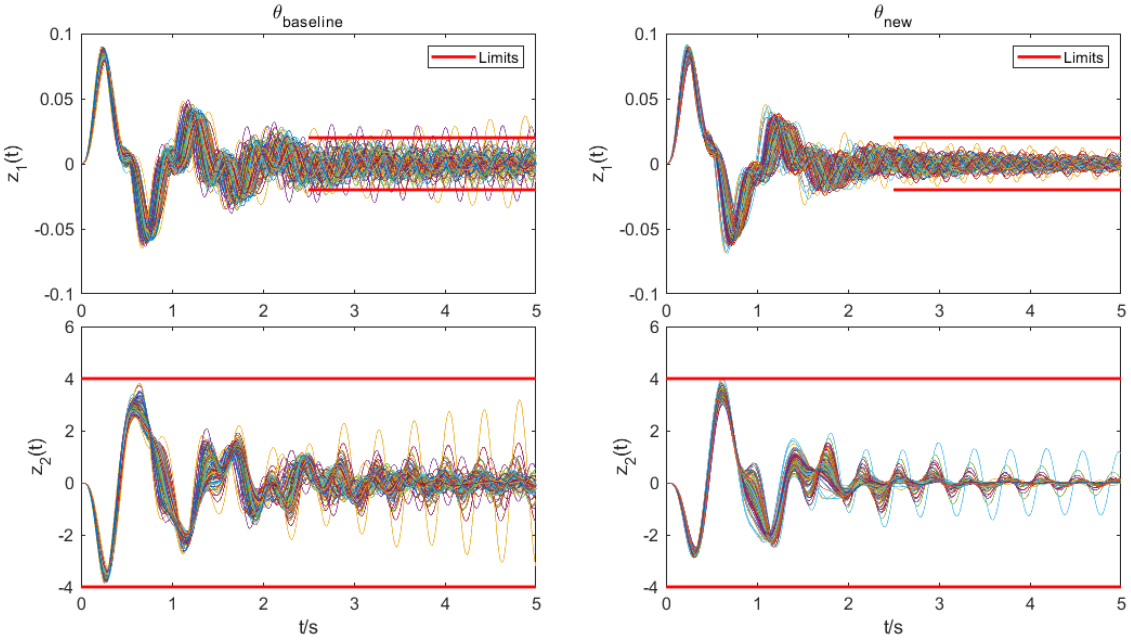

Fig. 16: Model outputs $z_1(t)$ and $z_2(t)$ regarding θ_{baseline} (left) and θ_{new} (right)

3.5 Subproblem E): Model update and design tuning

This subproblem contains the following tasks:

i) Interact with the host to provide the new design θ_{new} and ask for the corresponding new set of observations, namely D_2. Repeat the process in Subproblem A) to further update the UM using the new observation set D_2.

ii) Based on the further updated UM, interact with the host to ask for further refinements of the epistemic intervals of e_i. Repeat the processes in Subproblems A) and D) to further updated the UM and optimize the design. Denote the resulting design as θ_{final}.

iii) Repeat the reliability analysis in Subproblem C), and compare the reliability metrics according to $\theta_{baseline}$, θ_{new}, and θ_{final}.

Note that, the previous model updating in Subproblem A) was performed employing the output observations D_1 of the sub-system model $y(a, e, t)$, while the updating task here is employing the output observations D_2 of integrated system model $z(a, e, \theta, t)$, as plotted in the left subfigure of Fig. 17. The EMD method is applied to both sequences of Z_1 and Z_2. The Bayesian updating framework with the TMCMC algorithm is employed again, and the process converged after 13 iterations. The BMM coefficients of the aleatory variables a_i and reduced intervals of the epistemic variables e_i are presented in Table 10. Compare the results here and the ones in Subproblem B) (Table 5), it is observed the reduced intervals of e_i are not same, where the intervals of e_1 and e_3 are shifted, and the widths of intervals of e_2 and e_4 are further reduced, implying the new observations D_2 has a clear influence on the updating results.

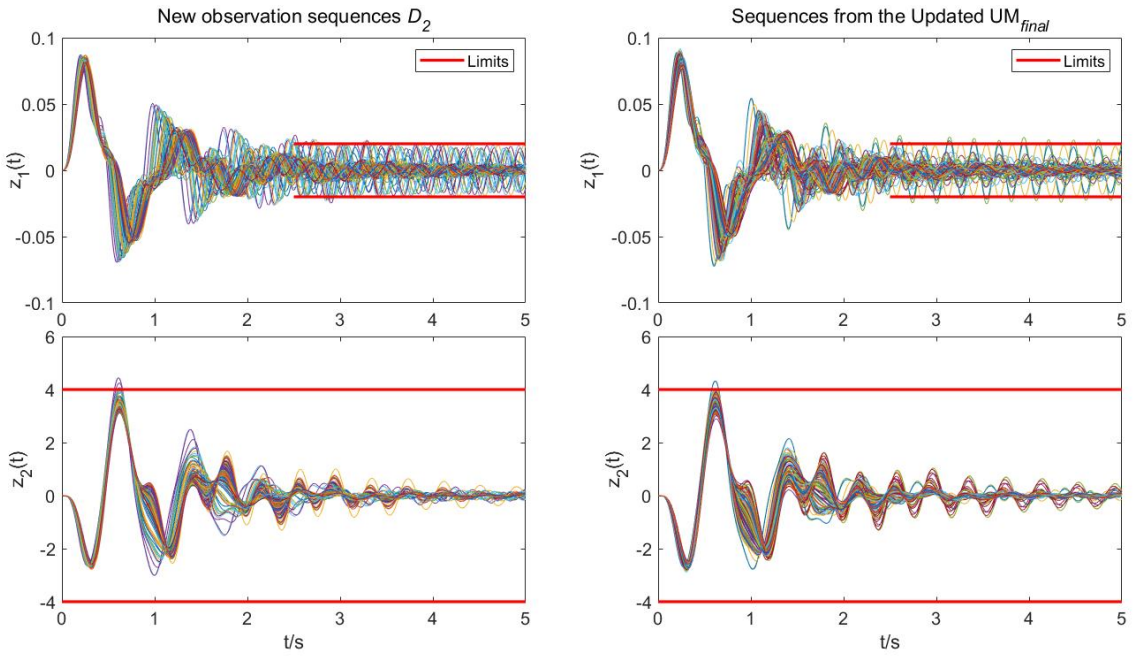

Fig. 17: Sequences z_1 and z_2 from the observation set D_2 (left) and from the updated UM$_{final}$ (right)

Table 10: The updated UM of the input variables using new observation D_2

Uncertainty models of the input variables							
	Aleatory variables					Epistemic variables	
	a_1	a_2	a_3	a_4	a_5		Reduced interval
$A_1^{(i)}$	11.5641	12.5038	2.5797	24.2593	13.8089	e_1	[0.6794, 0.9482]
$B_1^{(i)}$	21.9746	18.0192	5.9776	5.1945	6.3405	e_2	[0.5495, 0.6750]
$A_2^{(i)}$	16.8877	17.0460	20.4014	4.2724	5.0556	e_3	[0.4334, 0.8168]
$B_2^{(i)}$	29.0872	9.8369	21.4181	8.3106	15.454	e_4	[0.4928, 0.9174]

In task ii), we obtained the new (reduced) intervals of all e_{1-4} from the host, together with the new observation D_2, another round of model updating is repeated. The results, denoted as UM_{final}, are presented in Table 11. Comparing the results in Table 10 and Table 11, it is observed that the final reduced intervals of e_{1-4} are extremely narrow, implying the explicit values of the epistemic variables can be expected now, based on the supplementary data from the host. The estimated BMM distribution of the aleatory variables a_{1-5} are illustrated in Fig. 18. Clear differences are observed by comparing the PDF curves with the ones in Fig. 14, especially for a_{4-5}. The potential reason for the complicated PDF curves of a_{4-5} can be: on the one hand, there is inevitable compensation effect among a_{4-5} and other aleatory and epistemic variables; on the other hand, the shape coefficients of the BMM model can be very sensitive to the updated PDF curves.

Table 11: UM_{final} of the input variables using new observation D_2 and new intervals from the host

Uncertainty models of the input variables							
	Aleatory variables					Epistemic variables	
	a_1	a_2	a_3	a_4	a_5		Reduced interval
$A_1^{(i)}$	6.5137	12.2259	4.7597	0.8439	1.3977	e_1	[0.8353, 0.8791]
$B_1^{(i)}$	11.2953	15.9885	7.8330	14.9097	24.5739	e_2	[0.4440, 0.5377]
$A_2^{(i)}$	24.4351	5.9262	20.2813	23.0357	18.8409	e_3	[0.2351, 0.3126]
$B_2^{(i)}$	16.8815	4.7437	20.2739	17.6006	26.0315	e_4	[0.9902, 1.0520]

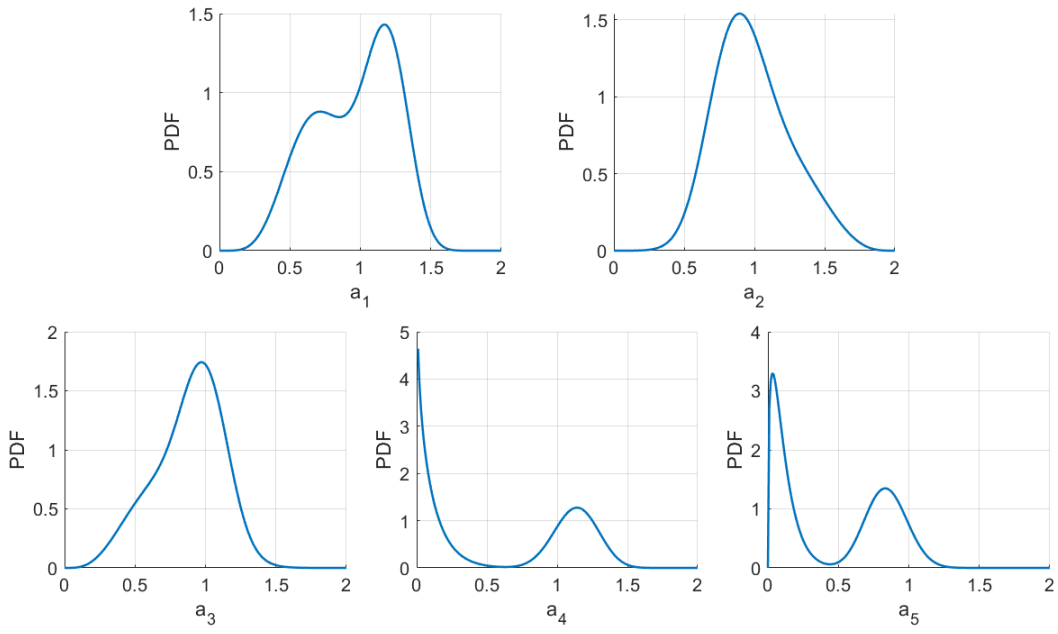

Fig. 18: The final updated BMM distributions of $a_{1\text{-}5}$ in UM_{final}

Task ii) also requires the reliability-based design starting from the θ_{new} and employing the UM_{final} obtained from task i). The resulting design, denoted as θ_{final}, is obtained and employed to calculate the final feature sequences Z_1 and Z_2 as illustrated in the right subfigure of Fig. 17. In task iii), the reliability metrics of $\theta_{baseline}$, θ_{new}, and θ_{final} are calculated according to Eqs. (13-15) and presented in Table 12. It is observed that from $\theta_{baseline}$ to θ_{new}, and finally to θ_{final} to system reliability is gradually improved. The ranges of the failure probability are more and more reduced by these three designs, and the severities are successively decreased.

Table 12: Reliability metrics evaluated based on the designs $\theta_{baseline}$, θ_{new}, and θ_{final}

Employed design	Range of failure possibility				Severity		
	R_1	R_2	R_3	R	S_1	S_2	S_3
$\theta_{baseline}$	[0.021, 0.288]	[0.009, 0.392]	[0, 0.016]	[0.025, 0.404]	0.0952	0.0019	0.0055

θ_{new}	[0, 0.036]	[00.34, 0.120]	[0.040, 0.140]	[0.059, 0.146]	0.0105	0.0005	0.0346
θ_{final}	[0, 0.027]	[0.025, 0.076]	[0, 0.007]	[0.027, 0.078]	0.0069	0.0002	0.0019

3.6 Subproblem F): Risk-based design

In this Subproblem, we focus on the definition of the risk and the gain regarding the influence of the epistemic uncertainty on the system reliability. Considering the remaining epistemic uncertainty space of the epistemic variables e_i in Subproblem E), the risk r% is defined as the portion of the epistemic space E to be neglected, where $r \in [0, 100]$. The gain should be defined as the improvement of the system performance benefitting from the retained $100 - r\%$ epistemic space. Clearly, the more the epistemic space is neglected (more risk), the more gain should be obtained. We define the gain here as the improvement degree of the system reliability in the presence of the retained epistemic uncertainty. Based on the definition above, the gain l can be expressed as

$$l(r) = \frac{P_E}{P_{E^*}} = \frac{\max\limits_{e \in E} \mathbb{P}[w(a, e, \theta) \geq 0]}{\max\limits_{e \in E^*(r)} \mathbb{P}[w(a, e, \theta) \geq 0]} \tag{17}$$

where $\max\limits_{e \in E} \mathbb{P}[w(a, e, \theta) \geq 0]$ is the maximum of the failure probability based on the worst-case performance function defined in Eq. (12); $E^*(r)$ is the retained epistemic space after taking the risk r%. The portion to be neglected is determined as follows:

$$E_{cut}(r, e^*) = \left[e^* \pm 0.5 \cdot \left(e_{upper} - e_{lower}\right) \cdot r\%\right] \tag{18}$$

where e_{upper} and e_{lower}, respectively, are the upper bound and lower bound of the epistemic interval; and e^* is the epistemic point leading the system failure probability to take the maximum value:

$$e^* = \underset{e \in E}{\operatorname{argmax}}\, \mathbb{P}[w(a, e, \theta) \geq 0]\,. \tag{19}$$

The retained epistemic space is calculated as $E^*(r) = E - E_{cut}(r, e^*)$. The process of taking the risk is shown in Fig. 19 so that it will be easier to understand. Clearly, the maximum failure probability in the full epistemic space P_E is larger than the one evaluated in the retained space P_{E^*}. The gain $l(r)$ in Eq. (17) will be larger than one.

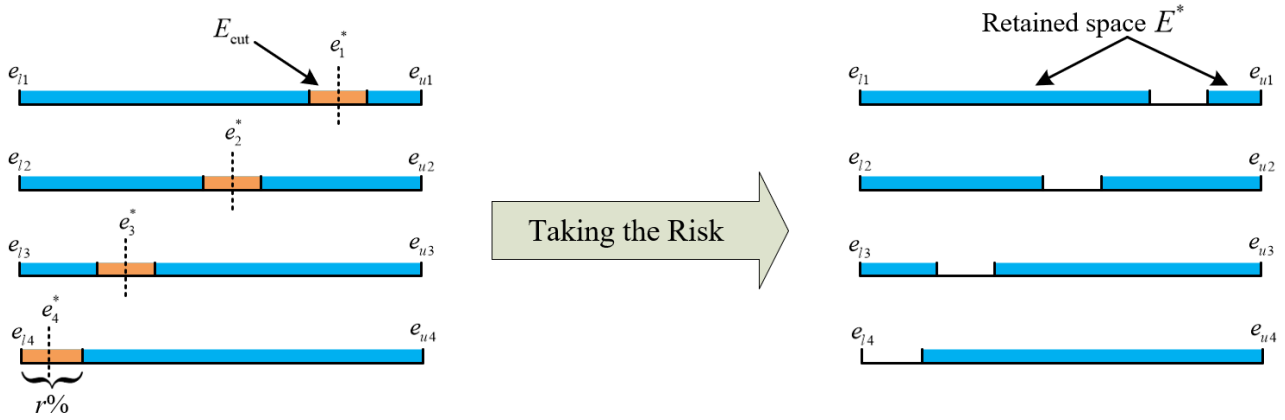

Fig. 19: Schematic of the process of taking the risk

After the risk and the gain are defined, the risk-based design is performed to find a design point, denoted as $\theta_{r\%risk}$, that maximizes the gain $l(r)$. This is an even more complicated optimization process compared with the optimization performed in the previous subproblems. As illustrated in Fig. 20, it contains not only the two-loop procedure for forward reliability analysis, but also a sub-optimization to find the e^* maximizing the failure probability in the epistemic space E, according to each instance of the design θ.

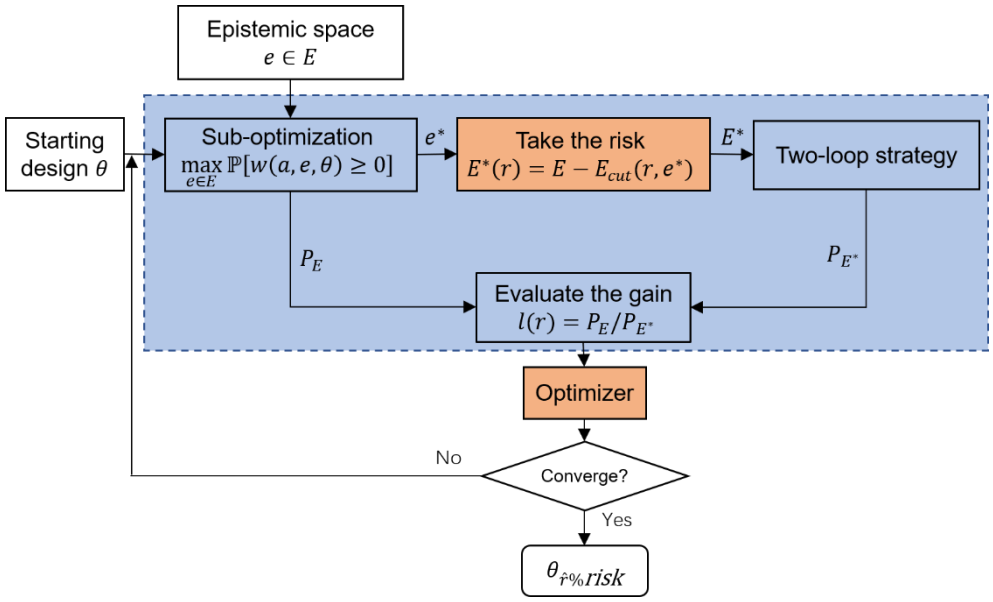

Fig. 20: Schematic of the risk-based optimization process

The sub-optimization is performed employing the PSO algorithm as the previous reliability-based optimizations. For the outer loop to find the θ, the main calculation burden originated from the complexity when evaluating the gain $l(r)$. In order to release the huge calculation cost, we employ the Surrogate Optimization Algorithm which interpolates an adaptive Radial Basis Function (RBF) to substitute the complex objective function. This technique is found to be well applicable for the multi-loop complicated optimization problem not only because of the adaptively trained RBF surrogate model but also the parallel calculation feature for the two-loop Monte Carlo sampling process.

Table 13: Reliability metrics according to θ_{final} and $\theta_{0.05\%risk}$ with both full epistemic space and retained space

Metrics	θ_{final}		$\theta_{0.05\%risk}$	
	Full space	Retained space	Full space	Retained space
R_1	[0, 0.027]	[0, 0.024]	[0, 0.051]	[0, 0.048]
R_2	[0.025, 0.076]	[0.025, 0.073]	[0.030, 0.120]	[0.034, 0.106]
R_3	[0, 0.007]	[0, 0.007]	[0, 0.007]	[0, 0.007]
R	[0.027, 0.078]	[0.026, 0.074]	[0.033, 0.120]	[0.035, 0.109]
S_1	0.0071	0.0060	0.0125	0.0118
S_2	0.0002	0.0002	0.0002	0.0002
S_3	0.0019	0.0011	0.0043	0.0027

The risk-based optimization is first performed for the prescribed risk level $r = 0.05$. It is required to repeat the processes in Subproblem C) to calculate the reliability metrics using θ_{final} and $\theta_{r\%risk}$ in both the full epistemic space E and the retained space $E^*(0.05\%)$. The results are presented in Table 13, where it is observed that, no matter for θ_{final} or $\theta_{0.05\%risk}$, the system performance in the retained epistemic space is improved than the ones in the full space. Because the ranges of the system failure probability (R_{1-3} and R) in the retained space are reduced and also the severities of in the retained space are clearly increased. This represents the gain after taking the risk. Another perspective to investigate the results is the comparison between the result under θ_{final} and the one under $\theta_{0.05\%risk}$. It is observed that both the failure probabilities (R_{1-3} and R) and the severities (S_{1-3}) under $\theta_{0.05\%risk}$ are higher than the ones under θ_{final}. This phenomenon is normal because θ_{final} is the design optimized under the whole epistemic space E, while $\theta_{0.05\%risk}$ is the local design within the reduced epistemic space $E^*(0.05\%)$. It is natural that the local optimization cannot get the system performance as good as the global one.

The last task is to investigate the trade-off between the gains and the risks on various levels under specific design. Multiple risk-based optimizations are performed to find the $\theta_{r\%risk}$ for risk levels of [0, 0.05, 0.5, 1, 5, 10]. The gains $l(r)$ are evaluated on these different risk levels employing both θ_{final} and $\theta_{r\%risk}$. The results are presented in Table 14. First, based on the vertical comparison between the results regarding θ_{final} and $\theta_{r\%risk}$, the gains regarding $\theta_{r\%risk}$ is larger than the one regarding θ_{final}. This demonstrates the optimization effect of the proposed risk-based optimization as shown in Fig. 20, since for the same risk level the gain has been maximized when this new design $\theta_{r\%risk}$ is applied to the system.

The trade-off between the gain and the risk is revealed by the horizontal comparison along each row of Table 14. It is observed that with the increasing of the risk levels (from 0% to 10%), the gain is indeed increased. As a conclusion to support decision making, it is necessary to adopt risk 0.05% because taking such a small risk will lead to a huge increasing of gain, i.e. growth of 12% from $l(\theta_{0\%risk})$ to $l(\theta_{0.05\%risk})$. But the risk 10% is suggested to be rejected, since such a large increase of risk (from 0.05% to 10%) only results in a small growth of 2% of the gain. This trade-off tendency is illustrated in Fig. 21, where the gain is increased significantly when the risk is taken on a tiny level in the preliminary stage, implying it is worthy to take such risk. When the risk is larger than 1%, it is shown that the increase of the gain is limited and hence the further risk should be rejected because of the flat trade-off curve between the gain and the risk.

Table 14: The gain based on different values of risks according to both θ_{final} and $\theta_{r\%risk}$

$r\%$	0	0.05	0.5	1	5	10
$l(r)$ regarding θ_{final}	1.0048	1.0552	1.0772	1.1202	1.1448	1.1348
$l(r)$ regarding $\theta_{r\%risk}$	1.0443	1.1609	1.1764	1.1660	1.1769	1.1812

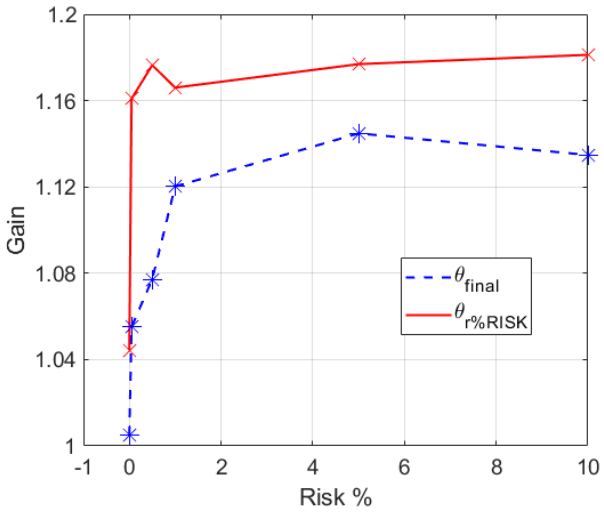

Fig. 21: The increase tendency between the risk and gain

3.7 Summary of the calculation cost

This subsection summarizes the calculation cost of the proposed techniques for each subproblem. The TMCMC algorithm employed in the Bayesian updating task has a nature feature that the randomly generated chains are independent with each other, making the parallel computation especially suitable here to reduce the time cost. Besides the parallel computation, the PSO algorithm and the adaptive REF-based surrogate optimization algorithm are employed in the reliability-based and risk-based optimization.

The overall problem is solved on a small-scale work station with 64 CPU cores (AMD EPYC 7742) to make full use of the parallel computation. The elapsed time as well as the Black-box function evaluation counts of each subproblem are listed in Table 15. It is observed that the Bayesian updating requires more elapsed time than the optimization processes because the Bayesian updating executes a huge number of the black-box functions *yfun* and *zfun*. The elapsed time of the optimization processes are acceptable, especially for the complicated risk-based optimization, because of the employment of the adaptive RBF-based surrogate optimization algorithm.

Table 15: The calculation cost of the proposed approach for each subproblem

Subproblems	Tasks	Elapsed time*	Function evaluation counts		
			yfun	*zfun*	*gfun*
A	Bayesian updating	19031s	4.48e10	--	--
B	Bayesian updating	14183s	3.92e10	--	--
C	Reliability Analysis	260s	--	--	1.0e6
D	Reliability-based Optimization	1725s	--	--	4.96e6
E	Bayesian updating	17046s	--	3.36e10	--
	Reliability-based Optimization	1584s	--	--	4.54e6
	Bayesian updating	27843s	4.20e10	4.20e10	0
F	Risk-based Optimization	19869s	--	--	5.0e7

*Work station with 64 CPU cores (AMD EPYC 7742)

4 Conclusion

This work addresses the NASA Optimization Challenge 2019 by proposing a series of systematically forward and inverse approaches for both uncertainty propagation and quantification, including key techniques such as UQ metrics based on Bhattacharyya distance, feature extraction based on EMD, UM parameterization based on BMM, the decoupling two-loop strategy for uncertainty propagation, and the inverse Bayesian updating approach. The challenging features of the problem originate from 1) the limited and irregular outputs, 2) the absent distribution information of the inputs, 3) the simultaneously involved aleatory and epistemic uncertainties in the reliability-based optimization, and 4) the intractable trade-off between the gain and risk in the risk-based optimization. Five key tasks are derived as uncertainty reduction, reliability-based design, risk-based design, sensitivity analysis, and reliability analysis, where the first three are categorized as the inverse procedure, and the two latter are categorized as the forward procedure. The proposed forward uncertainty propagation approach is designed to be capable of decoupling the aleatory and epistemic uncertainty in a two-loop strategy, where the P-box is investigated along with two directions and the influence of the aleatory and epistemic uncertainties on the sensitivity and reliability are assessed separately. The inverse uncertainty reduction approach first employs the Bhattacharyya distance and EMD method to preprocess the provided observations, after which the BMM coefficients are calibrated through the Bayesian updating framework. The PSO algorithm and parallel computation are employed in the reliability- and risk-based optimization to release the calculation burden caused by the inverse procedure performed along the P-box.

The key in addressing this problem lies in the principle to ***compromise*** with the inevitable uncertainties. We change our perspective by regarding the uncertainties as a concomitant partner but not a heinous evil that must be completely eliminated. It is not necessary, also impossible, to completely reduce the uncertainties during simulation-based design. This is reflected in the parameterization hypothesis of the aleatory distribution using the BMM method. Given the fact of limited outputs and the absence of prior distribution knowledge of inputs, we admit that the pursuing of the "true" aleatory distributions and epistemic constants becomes a mission impossible. But an appropriate hypothesis of parameterization and the comprehensive feature extraction allows the inverse procedure to reduce the

uncertainty meanwhile ensures the reserved uncertainty from the inputs can represent similar dispersion properties of the outputs, where a balance is achieved between the fidelity and robustness during the reliability- and risk-based design.

Acknowledgement

This work is supported by the Beijing Institute of Technology Research Fund Program for Young Scholars, which is greatly appreciated.

References

[1] Crespo, L. G., Kenny, S. P., "The NASA Langley Challenge on Optimization Under Uncertainty," Mechanical Systems and Signal Processing, 2021, 152, 107405.

[2] Crespo, L. G., Kennyy, S. P., Giesyz, D. P., "The NASA Langley Multidisciplinary Uncertainty Quantification Challenge," *16th AIAA Non-Deterministic Approaches Conference*, 1–9.

[3] Bi, S., Broggi, M., Wei, P., Beer, M., "The Bhattacharyya Distance: Enriching the P-Box in Stochastic Sensitivity Analysis," Mechanical Systems and Signal Processing, 2019, 129, 265–281.

[4] Huang, N. E., Shen, Z., Long, S. R., Wu, M. C., Shih, H. H., Zheng, Q., Yen, N.-C., Tung, C. C., Liu, H. H., "The Empirical Mode Decomposition and the Hilbert Spectrum for Nonlinear and Non-Stationary Time Series Analysis," Proceedings of the Royal Society of London. Series A: Mathematical, Physical and Engineering Sciences, 1998, 454(1971), 903–995.

[5] Bi, S., Prabhu, S., Cogan, S., Atamturktur, S., "Uncertainty Quantification Metrics with Varying Statistical Information in Model Calibration and Validation," AIAA Journal, 2017, 55(10), 3570–3583.

[6] Bi, S., Broggi, M., Beer, M., "The Role of the Bhattacharyya Distance in Stochastic Model Updating," Mechanical Systems and Signal Processing, 2019, 117, 437–452.

[7] Ching, J., Chen, Y. C., "Transitional Markov Chain Monte Carlo Method for Bayesian Model Updating, Model Class Selection, and Model Averaging," Journal of Engineering Mechanics, 2007, 133(7), 816–832.

[8] Lye, A., Cicirello, A., Patelli, E., "Sampling Methods for Solving Bayesian Model Updating Problems: A Tutorial," Mechanical Systems and Signal Processing, 2021, 159, 107760.

[9] Kennedy, J., Eberhart, R. C., "Particle Swarm Optimization," *Proceedings of the 1995 IEEE International Conference on Neural Networks*, 1942–1948.

[10] Regis, R. G., Shoemaker, C. A., "A Stochastic Radial Basis Function Method for the Global Optimization of Expensive Functions," INFORMS Journal on Computing, 2007.