



From inference to design: A comprehensive framework for uncertainty quantification in engineering with limited information

A. Gray^a, A. Wimbush^a, M. de Angelis^{a,*}, P.O. Hristov^{a,*}, D. Calleja^a,
E. Miralles-Dolz^a, R. Rocchetta^{b,*}

^a Institute for Risk and Uncertainty, University of Liverpool, UK

^b Department of Mathematics and Computer Science, Technical University of Eindhoven, The Netherlands

ARTICLE INFO

Communicated by John E Mottershead

Keywords:

Bayesian calibration
Probability bounds analysis
Uncertainty propagation
Uncertainty reduction
Epistemic uncertainty
Optimisation under uncertainty

ABSTRACT

In this paper we present a framework for addressing a variety of engineering design challenges with limited empirical data and partial information. This framework includes guidance on the characterisation of a mixture of uncertainties, efficient methodologies to integrate data into design decisions, and to conduct reliability analysis, and risk/reliability based design optimisation. To demonstrate its efficacy, the framework has been applied to the NASA 2020 uncertainty quantification challenge. The results and discussion in the paper are with respect to this application.

1. Introduction

Often complex systems must safely operate in extreme environments, under varying, or uncertain operational conditions. Ensuring satisfactory operation of novel systems subject to uncertain conditions is the aim of *uncertainty quantification*, the process of analysing the uncertainties in the interaction between complex systems and their environments [1]. It is common in modern engineering for a black-box, deterministic model to be used to inform design decisions, and uncertainties are always present in these models. Many model parameters may have an uncertain, unknown, or unobservable value. These models, used in the design process, are often constructed in isolation of the acquisition of experimental data. Experimental data may be scarce or poor-quality because of the novelty of the mission, inaccuracy of experimental measurements, or prohibitive cost of collecting data. Uncertainty quantification must be performed rigorously where possible, and diligently otherwise, so as to maximize the reliability of designs, decisions, and policies.

In engineering, the design process always, whether explicitly acknowledged or not, is about reducing uncertainty. All engineering design problems share some of the following four common challenges when dealing with their uncertainty.

- **Variability** – All engineering systems are expected to operate in varying environmental conditions. The design of these systems must be sensitive to the expected variability in the systems anticipated operation environment.

* Corresponding authors.

E-mail addresses: ander.gray@liverpool.ac.uk (A. Gray), a.wimbush@liverpool.ac.uk (A. Wimbush), marco.de-angelis@liverpool.ac.uk (M. de Angelis), p.hristov@liverpool.ac.uk (P.O. Hristov), d.calleja@liverpool.ac.uk (D. Calleja), enmidol@liverpool.ac.uk (E. Miralles-Dolz), r.rocchetta@tue.nl (R. Rocchetta).

<https://doi.org/10.1016/j.ymssp.2021.108210>

Received 12 March 2021; Received in revised form 29 May 2021; Accepted 3 July 2021

Available online 25 August 2021

0888-3270/© 2021 The Author(s). Published by Elsevier Ltd. This is an open access article under the CC BY license

(<http://creativecommons.org/licenses/by/4.0/>).

- *Inference* – When designing novel systems, often data is sparse or lacking entirely. During the design phases of engineered systems its common for more data to become available progressively. The design strategy needs to be able to incorporate this data as it becomes available, but doesn't excessively constrain the design by over-fitting to any data that does become available.
- *Ignorance* – During the design of engineering systems computational models are often used to support decision makers. These models may include parameter values that are unknown, or there may be unknown parametric dependencies. An effective design strategy must express and conserve this uncertainty, and make minimal assumptions about it, to ensure design decisions are robust to ignorance.
- *Decision Making* – The design process of complex systems involves multiple stages of progressive refinement. Each of these stages necessitates a decision about the collection of data which can be used to improve the design. These decisions must be made with limited information, and must accommodate the uncertainty present in the available data. The design strategy should provide as much flexibility to decision makers as possible. Constraining the design too early in the design process can be costly to change later.

The approach presented within this paper is a harmonised framework, borrowing from the most recent advances in uncertainty science, to form a complete solution sequence for dealing with these four types of uncertainty that are always present within engineering design. This approach supports the characterisation of these uncertainties simultaneously and can be used to perform an iterative design optimisation. This new framework provides support to practitioners to: make decisions about further data acquisition, assess the reliability of prototype designs, optimise the design against severe failures and uncertainties, and refine design decisions.

Within this work¹ we propose a new Bayesian calibration method to construct an uncertainty model which captures both epistemic and aleatory uncertainties using a multi-dimensional second-order probability distribution. The stochastic area metric (Kantorovich-Rubinstein/Wasserstein distance) [3,4], which measures the discrepancy between two non-parametric probability distributions is used in an approximate Bayesian updating procedure to replace the likelihood. This extension allows for a single-step calibration of correlated second-order probability distributions. We used *sliced-normal* distributions, recently introduced in [5], to efficiently generate samples of the updated posterior and to describe data-enclosing sets to visualise dependencies in high dimensional spaces. We used probability bounds analysis, a value-of-information approach, to investigate system output sensitivity to the epistemic parameters. We compare this approach to a variance-based sensitivity analysis, and use this comparison to identify which parameters should be prioritised. Second-order Monte-Carlo and probability bounds analysis are both adopted for forward propagation. Finally, we developed a reliability- and risk-based design optimisation scheme inclusive of both aleatory and epistemic uncertainties to identify candidate system designs which are robust to these unknowns. The use of probability bounds analysis throughout the framework allows for consistent characterisation and propagation of both aleatory and epistemic uncertainties.

A new NASA Langley challenge on optimisation under uncertainty was launched in 2020 [6]. The challenge combines problems of statistical inference, with complex multi-dimensional and black-box engineering systems. This challenge emulates common fundamental problems across science and engineering, and is intended to push the limits of the state-of-art within the field of uncertainty quantification. Because of the complexity of the challenge problem no single approach in the literature currently satisfies all requirements, and solutions have to strike a balance between robustness and efficiency. This new challenge builds upon the previous NASA uncertainty quantification and robust design challenge launched in 2014 [7] where a black-box model was used to provide a discipline-independent platform for bench-marking and testing of cutting edge methods for uncertainty propagation, sensitivity analysis, model calibration, and robust optimisation. Several authors tackled the original challenge using a variety of approaches [8–12]. Among these, [8] proposed a random set theoretical approach combined with approximate Bayesian computations and meta modelling, upon which our work builds and takes inspiration from.

The rest of the paper is organized as follows: In Section 2 we present the problem statement and discuss some of the most relevant challenges. Section 3 presents the theoretical preliminaries and reviews the existing literature. We highlight the novel contribution of this work in Section 4 and the numerical implementation of the proposed solution sequence is described in Section 5. We apply the proposed solution method to the NASA challenge problem and Section 6 presents the numerical results for the six problems and sub-problem. Section 8 closes the paper with a discussion on the results, failed attempts and possible research directions.

2. Problem statement

Consider a system of interest which is modeled as a set of interconnected subsystems, with uncertain parameters (a, e) . Here, a is vector of the aleatory inputs of size n_a such that $a \in A \subseteq \mathbb{R}^{n_a}$, and e is vector of the epistemic inputs, such that $e \in E_0 \subseteq \mathbb{R}^{n_e}$. The true value of the parameters $e \in E_0$ and the probability distribution of the variables $a \sim F_A$ are both unknown. The subsystem is modeled by the function $y(a, e, t)$, where $y: \mathbb{R}^{n_a} \times \mathbb{R}^{n_e} \times [0, T] \rightarrow \mathbb{R}$ and t is time. The integrated system is modeled by $z(a, e, \theta, t)$, where $z: \mathbb{R}^{n_a} \times \mathbb{R}^{n_e} \times \mathbb{R}^{n_\theta} \times [0, T] \rightarrow \mathbb{R}^{n_z}$. Hence, the output of the $y(a, e, t)$ is a function of time, and the output of the integrated system, $z(a, e, \theta, t)$, are n_z functions of time. The functions y and z are given as discrete time histories, e.g., $y(t) = [y(0), y(dt), \dots, y(n_t dt)]$ where $n_t dt = T$.

This model is the subject of a design optimisation procedure, where the goal of the procedure is to select a design vector θ of size n_θ in the space $\Theta \subseteq \mathbb{R}^{n_\theta}$, which ensures a high probability of system compliance with given reliability requirements, $g_j, j = 1, \dots, n_g$; these requirements may be the stability of the system, efficiency and energy consumption, structural integrity, maximum thermal or elastic stress, etc. The reliability requirements are modeled via a vector of reliability functions $g: \mathbb{R}^{n_a} \times \mathbb{R}^{n_e} \times \mathbb{R}^{n_\theta} \rightarrow \mathbb{R}^{n_g}$, where a g_j defines the

¹ This paper is an extension to the conference paper [2].

compliance of the system with respect to requirement $j = 1, \dots, n_g$. If $g_j(a, e, \theta) > 0$, a design θ results in failure of requirement j given the uncertain scenario (a, e) .

Two data sets of discrete time histories $D_1 = \{y_t^{(i)}, t = 0, \dots, T\}_{i=1}^{n_1}$ and $D_2 = \{z_t^{(i)}, t = 0, \dots, T\}_{i=1}^{n_2}$ are made available at different stages of the design process. The quantities n_1 and n_2 denote the number of time-histories for a subsystem and the integrated system, respectively. For the sake of notation simplicity, the dependency of the signals z and y with respect to the uncertain factors (a, e) and θ has been dropped in D_1 and D_2 .

The computational models and the data sets must be combined to tackle different challenges in the prototyping of reliable design. Specifically, this work will focus on the following six tasks:

- Model calibration and uncertainty quantification for the subsystem
- Epistemic uncertainty reduction
- Reliability analysis of a baseline design
- Integrated system prototype via reliability-based design optimisation
- Model updating and final design tuning
- Risk-based design optimisation

The problem statement presented in this section, as previously described emulates many real-life design challenges. The focus of this work is the NASA 2020 challenge on optimisation under uncertainty [6], but our approach is general. In summary, the specific setup of the NASA challenge problem considers $n_z = 2$ dynamic system responses, $n_e = 4$ epistemic parameters, $n_a = 5$ inputs affected by aleatory uncertainty, and $n_\theta = 9$ parameters defining the design of a feedback controller with a non-collocated sensor/actuator pair. The design vector θ , must satisfy $n_g = 3$ reliability requirements on the stability, settling time, and energy consumption of the system-controller. Two data sets D_1 and D_2 containing $n_1 = n_2 = 100$ length $n_t = 5000$ realizations of the system responses are made available at different stages of the design process. Fig. 1 provides a visual summary of the NASA problem statement whilst Table 1 summarises the dimension of the parameters, the availability of data, and the relevance of the factors for the six tasks.

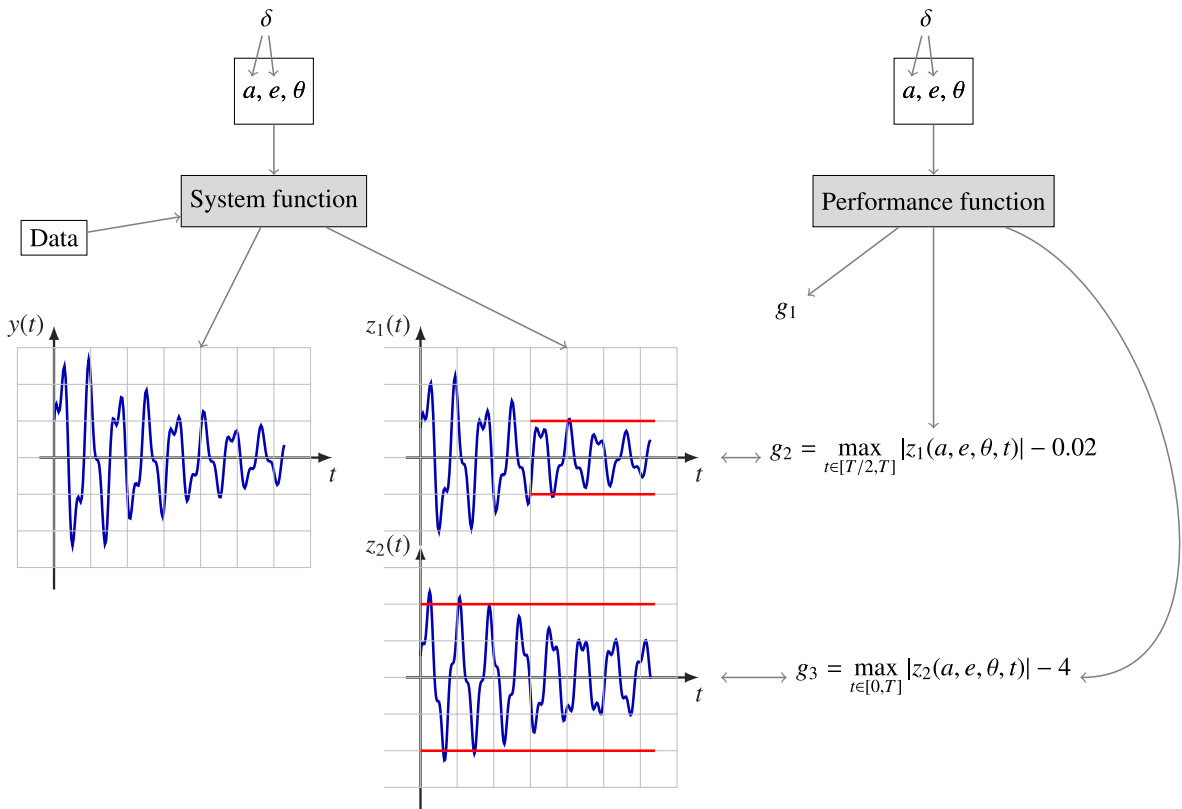


Fig. 1. Summary of the model functions, inputs and outputs. Left: System functions produce time series outputs in 1 (y) or 2 (z) dimensions. Right: The z functions are then used to compute the values of the performance functions for design evaluation.

Table 1
Summary of input parameters and output quantities.

Symbol	Description	Data	Task	I/O	Domain
a	Aleatory parameters	No	Calibration	Input	$[0, 2]^5$
e	Epistemic parameters	No	Sensitivity	Input	$[0, 2]^4$
θ	Design parameters	No	optimisation	Input	\mathbb{R}^9
y, z	System response	Yes	Calibration, Sensitivity	Output	\mathbb{R}^{5000}
g	Performance output	Yes (indirectly)	Propagation, optimisation	Output	\mathbb{R}^3

2.1. Challenges

There are *four* features of the NASA 2020 challenge which characterise its complexity and distinguish it from previous challenge problems:

- i) the calibration of a mixed uncertainty model—containing epistemic parameters and an unknown correlated probability distribution simultaneously—in high dimensions;
- ii) the inverse black-box inference problem, where an uncertainty model of the inputs is sought however only data for the output is provided;
- iii) the time-history of the model output is auto-correlated;
- iv) very limited data was provided compared to the size of the output making inferential uncertainty large, and the propensity to over-fitting significant.

3. Literature review

3.1. Models of uncertainty

Most in the uncertainty quantification community agree that uncertainty is characterised by two classes: *aleatory* and *epistemic* uncertainty. These two classes behave differently, and require a careful and separate consideration in their modelling. This distinction between the nature of uncertainty is echoed in the NASA challenge problem statement, which clearly states that their computational model has both aleatory and epistemic inputs. In this section we describe these concepts, and review various mathematical frameworks for modelling systems with mixed aleatory and epistemic uncertainty.

Aleatory uncertainty arises from natural stochasticity, this is sometimes referred to as randomness. The specific random values of an aleatory parameter cannot be predicted exactly, however its overall dispersion follows a specific probability distribution which may in aggregate be inferred. Aleatory uncertainty is therefore thought to be perfectly modelled by probability theory, where the probability is interpreted as a *frequency*. This type of uncertainty is *irreducible* in the sense that gathering more data may give a better characterisation of its distribution, but cannot reduce it to a scalar.

Epistemic uncertainty, which arises from imperfect knowledge is typically understood to be reducible by empirical effort (gaining more knowledge). Epistemic parameters are constant, but unknown, and are not naturally varying. Epistemic uncertainty may also be modelled with probability, following the *Bayesian* interpretation, where a probabilistic statement is a subjective degree of belief assigned to possible events. The Bayesian interpretation makes no distinction between the frequency of observed events or subjective belief about unobserved events. Kolmogorov's axioms, that define the calculus of probability, do not distinguish between these either, therefore characterising physical systems with a combination of both (observed data and subjective belief) can allow biases to yield erroneous results.

Intervals [13] offer a more natural way to model imperfect knowledge, or epistemic uncertainty. These intervals may be bounded at one-end, both-ends or open. Intervals make no statement about the mass inside the interval and hence are a relaxation of probability theory. However, they are not ideally suited to characterising aleatory uncertainty. Because of the natural ability of intervals to express epistemic uncertainty, and the ability of probability to express aleatory uncertainty, combining these two frameworks yields the theory of *imprecise probabilities* (IP). IP is a theory of uncertainty which models both variability and imprecision. This provides a framework for more rigorous characterisation of the limits of knowledge, and a distinction between what is not known, and what is varying. IP is effectively a generalisation of probability theory for computing with *sets* of probabilities, relaxing the Kolmogorov's additivity axiom. Imprecise uncertainty models include *probability bounds analysis* [14–18], *Dempster-Shafer evidence theory* [19], *random set theory* [20–22], *fuzzy set theory* [23–26], *polymorphic uncertainty theory* [27,28], *info-gap theory* [29] and others.

The probability box (p-box) is an imprecise uncertainty model which has recently grown in popularity. A p-box bounds a set of probability distributions when any of the following three constraints are satisfied: (i) there are two bounding CDFs, (ii) interval bounds on the moments (μ_x, σ_x) are defined, and (iii) a set of distribution families \mathbb{F} (e.g. normal) is given. That is, a distribution $F(x)$ is a member of a p-box if:

$$i) \underline{F}(x) \leq F(x) \leq \bar{F}(x)$$

- ii) $\mu_X \in [\underline{\mu}_X, \bar{\mu}_X]$ and $\sigma_X \in [\underline{\sigma}_X, \bar{\sigma}_X]$
- iii) $F(x) \in \mathbb{F}$

Some of these constraints may be missing, or informed from the others. For example, bounds on the moments may be retrieved from the CDF and distribution family information. Alternatively, the CDF bounds may be informed through moment and range information [30], by making use of the Chebyshev, Markov, and Cantelli inequalities. Or, the distribution family may be missing entirely, with such p-boxes sometimes being referred to as *non-parametric*.

The standard method for propagating p-boxes is p-box arithmetic, or probability bounds analysis [31,17,32]. This requires an interface between the model source code and compiler because most programming languages do not support interval computation. Efforts to overwrite machine level arithmetic for interval computation are underway [33], but currently these approaches are not commonly used. In the NASA challenge, and generally in engineering, we are restricted to studying a black-box model. Therefore, the choice of the appropriate model of uncertainty is limited to methods primarily based on sampling. In recent literature we have seen the emergence of methods based on sampling that simulate the aforementioned imprecise models of uncertainty [34–39], including a number of non-intrusive methods involving multiple Monte-Carlo or optimisation operations [40–43]. Each of these authors have shown that *probability bounds analysis*, *evidence theory*, and *random set theory* can be performed via sampling without altering the rules of code compilation, demonstrating the applicability of such frameworks to black-box models.

A popular mixed uncertainty model, particularly in combination with Bayesian methods, is the second-order distribution. These attempt to model the aleatory and epistemic components. These structures arise naturally in so-called *double-loop Monte-Carlo* simulation, where first a probability distribution is picked at random from a specific family, and then its values are simulated with Monte Carlo. This yields an ensemble of distributions. These structures grew in popularity in the 1990s [44,45] as more computational resources became available. However, they are rarely discussed in recent literature, most likely due to the many criticisms: the method is generally very expensive since multiple Monte-Carlo loops are required; they are thought to severely underestimate epistemic uncertainty [46,47]; and, modelling dependencies may also be troublesome. In [47, Section 5.1], an example is given for the sum of N probability distributions in the interval $[a, b]^N$. When propagating epistemic uncertainty, Monte Carlo gets increasingly less accurate as the number of variables increases. Given that the sample variance of the sum linearly decreases with the number of variables, it gets ever more improbable to sample close to the endpoints of the box where the exact bounds hold. Despite this, second-order distributions are a flexible method for modelling epistemic and aleatory uncertainty with sampling. Some of the criticisms of second-order distributions may be alleviated by converting them to p-boxes by enveloping the ensemble, referred to in the challenge document as an inner approximation of a p-box [17,18].

In this work we develop a *probability bounds analysis* via sampling strategy to address the NASA challenge problems using an inner approximation of a p-box constructed from a second-order distribution. This is well suited to deal efficiently with the propagation problem of mixed uncertainty.

3.2. Calibration of mixed uncertainty models

Inverse problems cover a class of mathematical challenges where the objective is to derive unobserved function inputs from a set of output observations [48–50]. In the context of probability modelling this involves finding a distribution of input parameters which matches some output distribution. Model calibration is an example of such an inverse problem. Probabilistic model calibration can be performed with Bayesian [51,52] or frequentist methods [53].

Bayesian approaches to model calibration have a number of desirable features, for example, they are generally applicable and can incorporate prior expert judgment where data is scarce. However, there are a number of limitations to these approaches. The calibration problem may be ill-posed [51] with multiple optimal solutions. Furthermore, computational challenges arise from high-dimensional, high-peaked or multi-modal distributions, or when computationally expensive models [54] are required to evaluate candidates. Accounting for both aleatory and epistemic uncertainty is particularly challenging. A number of recent papers have also shown that non-additive measures are a requirement for valid inferential problems, and additive models used for epistemic uncertainty, commonly employed in Bayesian approaches, lead to false confidence [55,56]. False confidence occurs when there is a large probability of assigning high confidence to a false assertion. Martin and Liu, [57,58] propose an inferential framework to combat false confidence.

If inferential uncertainty is large due to limited data, or if both epistemic and aleatory uncertainty are present, imprecise probabilities are required. A number of recent authors have proposed methods for the calibration of imprecise probabilities [59,25,26,60]. The literature on the topic is sparse and is mainly under the Bayesian paradigm [51,61]. There is also work on fuzzy logic and arithmetic [25,26]. An example of fuzzy arithmetic to solve inverse problem in finite element models can be found in [62]. Interval approaches to the inverse problem have also appeared in recent literature [63,64]. There is work on interval characterization too [65]. However, the black box nature of the problem necessitates the use of sampling methods, making robust imprecise inference difficult.

Updating over time histories is a common problem; this introduces complications due to the auto-correlation present between time steps. One way to account for this auto-correlation within model updating is to model time histories by stochastic processes or random fields [66]. Due to the high dimensionality, approximations or dimensionality reduction methods are often required to compactly represent random fields, for example, Gaussian processes [67,68], Fourier transformations, and Karhunen-L eve expansions. Compact representations of interval fields [69], as well as imprecise random fields using parameterised kernels [70] have been proposed.

Within the NASA challenge problem, we are attempting to calibrate an uncertainty model (*UM*) for the system. This problem is high-dimensional, there is no knowledge of the input space other than bounds, the model accounts for both aleatory and epistemic uncertainty, and the available data is limited. This motivates the use of an approach, described in Section 5, which avoids the limitations of previous methods described.

3.3. Approximate Bayesian computation

Approximate Bayesian methods are a way to solve inference problems in the absence of a likelihood function which is replaced by a sufficient statistical metric. Bi et al. [71] propose a Bayesian method for calibrating a set of distributions using stochastic metrics to replace the likelihood [72], allowing for an updating procedure over a space of distributions. Similarly, an approximate Bayesian computation using the Wasserstein distance was applied by Bernto et al. [73] making the method scalable. Engquist et al. [50] analytically and numerically characterise two major effects of the quadratic Wasserstein distance as the measure of data discrepancy in computational solutions of inverse problems. Approximate Bayesian methods utilise a class of random-walk algorithms, Markov Chain Monte-Carlo (MCMC), that incrementally improve the selection of inputs that best match the output data when propagated through the model. A recent tutorial paper for some of these MCMC methods is available [74]. These algorithms use a simple proposal probability distributions seeded by samples of the previous iteration to sample new candidate points. If these candidate input points yield output points closer (as defined by the chosen distance metric) to the target output data then they are kept, otherwise they are discarded. Most common amongst these methods is the use of Metropolis–Hastings [51,75]. Transitional Markov Chain Monte-Carlo (TMCMC) was first proposed by Ching et al. [76] adapting the earlier Metropolis–Hastings algorithm. TMCMC constructs a set of independent Markov chains, making parallelisation possible. This is essential when the model is computationally expensive to evaluate, or when the dimensionality of the problem is large.

Since the introduction of TMCMC several improvements have been proposed, such as: a recalculation of weights at each iteration, a burn-in at the beginning of each chain to improve intermediary posterior sample convergence, and an adaptive scaling of the proposal distribution [77]. Further improvements have been identified with enhanced selection of task dependent distance metrics within the ABC regime [73,72]. Similar approaches include the Metropolis-adjusted Langevin algorithm [78], the Metropolis preconditioned Crank–Nicolson method [79], Hamiltonian MCMC [80], relativistic MCMC [81], and many others [82–86].

An appropriate likelihood metric is not readily available for the NASA challenge problem. Therefore approximate Bayesian methods with TMCMC sampling provides a means of calibration which is efficient and capable of handling the high dimensionality of the output.

3.4. Sensitivity analysis

Sensitivity analysis “*is the study of how uncertainty in the output of a model (numerical or otherwise) can be apportioned to different sources of uncertainty in the model input*” [87]. In general there are a number of motivations for one to employ sensitivity analysis, these may be: to rank uncertainties in order of magnitude of impact; to identify where greater empirical effort may best be placed to reduce uncertainty; to estimate the effect of different modeler assumptions; or to justify simplifying their model and removing inputs that have little effect. There are a number of approaches to do sensitivity analysis that are concerned with different models of uncertainty. These approaches are broadly classified as variance-based, Bayesian sensitivity analysis, and value-of-information. Variance based sensitivity uses a variance decomposition scheme. This class of sensitivity analyses attempts to attribute portions of the variance in the model output to different inputs, characterised by probability distributions [88,89]. Within variance-based methods the most commonly employed approach is to propagate samples of the input distributions to the output domain multiple times and to derive sample statistics about the resulting distribution. This approach attempts to rank input parameters based on their respective contribution to the output variance, either in isolation or by some order of combination.

Bayesian sensitivity is also probabilistic, but uses the machinery of Bayesian analysis [90]. This approach is sometimes referred to as a robust Bayesian analysis, where the modeller explores the effect of different prior distributions to their best available estimates on the posterior of interest. This approach seeks to identify which of the modeller’s prior beliefs have the greatest impact on the posterior [91,92].

Value-of-information sensitivity analysis is concerned with estimating the reduction in uncertainty of the output should more evidence become available about each of the inputs. This approach is usually done comparing a metric of uncertainty, such as the width of the output interval or the Hartley-like measure, before and after *pinching* the uncertainty of one input or multiple inputs at the same time. Several methods have been proposed in the literature that explore this using different metrics [3,8,89].

3.5. Reliability- and risk-based design optimisation

The goal of reliability-based design optimisations (RBDO) is to identify a design that minimizes (or constrain to an acceptable level) the failure probability for some performance requirement(s). A traditional approach to RBDO involves two nested loops, an outer-loop explores the design space, the inner-loop calculates the failure probability for a candidate design [93]. However, nested loop methods are often very inefficient. This inefficiency is due to the time-consuming estimation of the failure probability which involves approximating a multidimensional integral over the failure domain for each candidate. Efficient RBDO can be achieved by: (i) replacing the nested loop with a single-loop [94] or decoupled approach [95], or (ii) reducing the computational costs of the inner loop. The cost of the inner loop can be reduced by utilising an approximate analytical method such as the first- and second- order

reliability method to estimate the failure probability; or, by using efficient Monte-Carlo methods such as subset simulation [96], line sampling [34], or multi-level surrogate models [97]. Other contributions have explored inverting these nested loops for the propagation of non-additive probabilities [22].

In addition to these computational issues with RBDO, when we maximize the reliability we obtain no guarantees on the severity of failures. This may yield sub-optimal solutions, where the probability of failure is small but the severity of any failure may be very significant. In other words, the tails of the distribution of the reliability performance function can extend deep into the failure domain. In practice, the analyst might want to minimize the severity of failure given by the extent of the tail of the reliability functions. The average severity of failures can be obtained by computing the conditional value-at-risk scores. This can replace the failure probability in the objective function of the optimisation procedure. Minimization of superquantile, the conditional value-at-risk, yields designs with a small probability of failure and that are robust against severe failures. This type of design problem is known as risk-based design optimisation. Beck et al. [98], presents a comparison of risk-based, reliability-based, and robust design optimisation approaches, see also [99]. Robust methods tend to yield designs that are less sensitive to uncertainty, whereas RBDO methods seek designs with a maximum level of reliability, whilst risk-based design methods target designs that optimally compromise cost, probability of safe operations, and severity of failure. For more details about the superquantile approach on failure probabilities see [100], and for recent risk-based design methods see [101].

The majority of the existing reliability- and risk- BDO methods rely on a precise characterization of a probabilistic model, which is necessary to estimate a design's reliability or risk. The issue with defining a precise probabilistic model is the sensitivity of the failure probability and severity to the choice of model. Selecting a good model of the uncertainty can be challenging, especially for high-dimensional problems, or when dependencies are unknown, or data is lacking. Where a lack of data is affecting the analysis, a distributionally robust reformulation of the RBDO problem or a non-probabilistic hybrid reformulation of the reliability problem can be done [102,103]. Other methods are available to circumvent these issues in the definition of the probability model. These include performing risk-based and RBDO directly on given data [104], set-valued characterization of the uncertain factors and minimizing the non-probabilistic reliability index [103], or using Scenario-RBDO methods to enclose design solution data while offering epistemic guarantees [105]. Only a few works attempted to solve RBDO in the presence of mixed aleatory and epistemic uncertainty. Most of these have been in attempts to address the previous NASA challenge problem on UQ (2014), [9,8,10].

4. Novel contributions

In this paper we present a complete and harmonised framework for dealing with uncertainty in engineering design. It is complete in the sense that it addresses many major UQ tasks found in any engineering design mission, from calibration to design optimisation. This framework combines some of the latest methods in uncertainty quantification, and is tested on the high dimensional, low-data NASA challenge problem in an mixed uncertainty setting.

The novel contributions within the framework we present include methods for:

- calibration of second order distributions on time history data,
- construction of a mixed uncertainty model with a focus on characterisation of inferential uncertainty (Avoiding over-fitting),
- efficient simulation of posterior samples using sliced-normal distributions,
- probability bounds sensitivity analysis via sampling, and
- RBDO under mixed parametric uncertainty.

5. The proposed approach

In this section, we present a novel solution framework to address some of the challenges discussed in the previous sections. We sketch some of the mathematical foundations that are required to understand the results presented in this document. We will keep a practical mindset oriented towards the problems of interest.

5.1. Notation

Extensive use of mathematical notation will be made throughout the paper. The use of notation will be consistent with the symbols already introduced in the challenge manifesto [106]. For more details about the symbols herein used and their meaning the reader is referred to the Nomenclature in [Appendix A](#).

5.2. Constructing a multivariate parametric distribution

In this framework, we model the aleatory uncertainty with a parametric distribution F_A whose parameters must be inferred from data as additional epistemic inputs. In principle any parametric family may be selected for the marginals, with the dependency structure being defined by a parametric copula family. The selection of these families may be problem specific. If a problem has bounded aleatory inputs, and no information about the inputs shape in the marginals, then a Beta distribution family is a good choice [107]. For our application we introduce a joint distribution based on the Beta marginal. This is because the family is bounded, like our aleatory inputs A_i for the NASA challenge, and produces a wide variety of shapes with two parameters. For our problem, each a_i is bounded in the range $[0, 2]$, so we scale the standard Beta model:

$$F_{A_i}(a_i; \alpha_i, \beta_i) = B\left(\frac{a_i}{2}; \alpha_i, \beta_i\right) \tag{1}$$

where $i = 1, \dots, n_a, a_i \in [0, 2]$, and B is the beta distribution's CDF with shape parameters α and β . Since F_A is n_a dimensional, a total of $2n_a$ parameters are required to define the shape of all of the marginal distributions. The dependence structure of F_A may be modelled by selecting an n_a dimensional parametric copula. The Gaussian copula is a typical choice which is parameterised by a semi-positive definite correlation matrix Σ . A Gaussian copula C_Σ is defined in terms of the (mean zero) multivariate Gaussian CDF F_G parameterised by correlation matrix Σ :

$$C_\Sigma(u_1, u_2, \dots, u_{n_a}) = F_G(\Phi^{-1}(u_1), \Phi^{-1}(u_2), \dots, \Phi^{-1}(u_{n_a}); \Sigma) \tag{2}$$

where Φ^{-1} is the inverse CDF of a standard normal distribution, $u_i \in [0, 1]$, and with n_a being the dimension of the copula. With n_a dimensions, an $n_a \times n_a$ correlation matrix is required to define the copula for F_A , the off diagonals of which may be varied to change the dependence structure. Since there are $\frac{n_a(n_a-1)}{2}$ unique off-diagonal elements in an $n_a \times n_a$ correlation matrix, a total of $\frac{n_a(n_a-1)}{2}$ parameters are required to define the dependence of F_A .

For this problem, since $n_a = 5$, there are $2n_a = 10$ beta parameters required along with $\frac{n_a(n_a-1)}{2} = 10$ parameters for the correlation matrix. This gives a total of 20 additional epistemic parameters $e_{5:24} \in E_B \subseteq \mathbb{R}^{20}$ that are required to completely specify the joint distribution F_A :

$$F_A(a_1, \dots, a_5; e_{5:24}) = C_{e_{15:24}}(F_{A_1}(a_1; e_{5:6}), \dots, F_{A_5}(a_5; e_{13:14})) \tag{3}$$

where Sklar's theorem is used, and $E_0 \times E_B$ is the collection of 24 parameters defining the entire joint distribution, with $e_{1:4}$ defining the original epistemic parameters, $e_{5:14}$ defining the shape of the marginals, and $e_{15:24}$ the dependence structure (see Fig. 2). We begin the indexing of this vector from 5 and call it e because they are epistemic parameters, and will be added to the original 4 epistemic parameters of the computational model in the calibration stage, where inference can be performed on this entire $e_{1:24}$ epistemic vector. Fig. 2 shows the density f_A for a particular choice of E_B .

A Gaussian copula family requires a 5×5 correlation matrix as an input, the off diagonal elements of which may be varied to create different dependence structures. A valid correlation matrix must be semi-positive definite. It has however been shown in [108] that a semi-positive definite correlation matrix may be parameterised in terms of partial correlations, which all independently take values in $[-1, 1]$. Thus ten partial correlations, which may all take values between $[-1, 1]$ may be included in the calibration procedure to define the dependence structure for A .

A multivariate imprecise structure can quite easily be constructed from F_A , by expressing epistemic uncertainty about parameters E_B ; for example when E_B is an interval (compact space), F_A is a multivariate p-box, and when E_B is distributional (probability space), F_A will be a second order distribution. The parameters of E_B will be inferred in the calibration procedure.

The UM , as has been described, is an ensemble of aleatory models. Fig. 2 shows several visualisations of the probabilistic model, given a sample of the epistemic space. This is demonstrative of the flexibility of the Beta marginal and Gaussian copula model.

Since the epistemic space is expanded in this framework, its dimensionality will increase. For example, in the context of the NASA challenge, we go from a $n_e = 4$ dimensional space to an $n_e = 24$ dimensional space. Therefore we redefine $e \in E = E_0 \times E_B \subseteq \mathbb{R}^{24}$ for the subsequent steps in this framework. For the sake of simplicity, we use e without any subscripts, whenever this does not cause any confusion. Whenever e is used as an input to define an aleatory distribution this refers only to components $e_{5:24}$, as in Eq. (3). In contrast, whenever e is supplied to the system or performance functions, it will denote $e_{1:4}$.

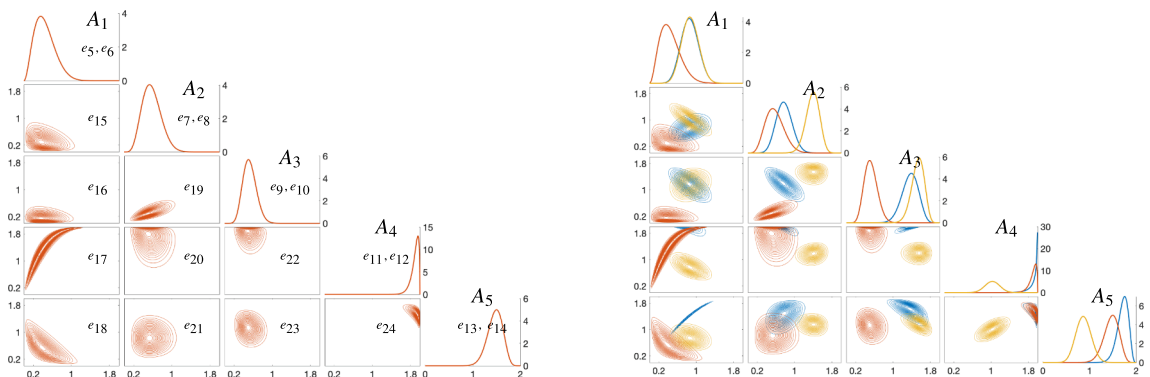


Fig. 2. Construction of the aleatory joint distribution F_A . Left: the contribution of each epistemic parameter in $e_{5:24} \in E_B$ to each of the marginals and joint correlations are shown. Right: three examples of possible F_A models produced by three different epistemic vectors $e_{5:24}$ are shown, demonstrating its flexibility.

5.3. Uncertainty calibration

Under this parametrisation, a standard calibration technique may be performed over e . In this work we use Bayesian model updating to perform the calibration of our uncertainty model. In Bayesian model updating, one would like to construct a probability distribution over e that is reflective of the provided experimental data. For this, we select a prior distribution which reflects the analysts knowledge about the parameter space e before observing any data, and is updated such that the updated distribution (a posterior) matches the experimental data in the output domain. This posterior is found by solving Bayes law:

$$P(e|D) = \frac{P(D|e)P(e)}{P(D)} \quad (4)$$

where $P(e|D)$ is the posterior, constituting a calibrated model. $P(e)$ is the prior, which is often selected to be a uniform distribution when nothing about the parameters except their ranges is known. $P(D|e)$ is the likelihood, and is the probability of observing data D given a point e in the input domain. Defining and evaluating the likelihood is often the most difficult part of Bayesian inference, since it requires evaluating the model at point e and comparing its prediction to the data in a consistent way. This will be discussed in the following subsection. $P(D)$ is a normalising constant and involves solving an integral, which in high dimensions is usually intractable. However since it does not depend on the parameters e , it is just a constant and so a method for sampling from an unnormalized probability distribution, such as MCMC, may be used to draw samples from $P(e|D)$.

5.3.1. Approximate likelihood for second order distributions

For a multidimensional output, and under the assumption of independence, the likelihood is a function $\mathbb{R}^{n_1} \times \mathbb{R}^{n_2} \times E \rightarrow \mathbb{R}$ defined as:

$$P(D|e) = \prod_{t=1}^{n_t} f(y_t|e) \quad y_t \in D \quad (5)$$

where $f(y_t|e)$ is the probability density of the data at time point t , evaluated for a model output from point e .

Note that there is some loss of information with this independence assumption. Accurate estimation of the data density is very costly, and would require a large amount of samples and possibly a density estimator such as Kernel Density Estimation. We therefore follow a similar approach outlined in [71], where the likelihood is replaced by an approximate likelihood based on a stochastic distance metric. A stochastic distance metric defines distances between probability distributions (i.e. defines how dissimilar they are). This framework fits well in our problem, since both our predictive distribution (for a fixed e) and the experimental data are distributional. We may therefore define a higher likelihood for points in e which produce similar distributions to D , and penalise points which produce distributions which are ‘‘far’’ to D w.r.t our metric. We define an approximate likelihood based on the Gaussian distribution for each time point, as:

$$f(y_t|e) \propto \exp\left(-\frac{d^2}{\epsilon^2}\right) \quad (6)$$

where d is a distance metric between distributions and ϵ is the so-called width factor which must be preselected by the analyst. The width factor controls the ‘‘spikiness’’ of the posterior, the smaller ϵ is the tighter the posterior is but also harder to sample from. The larger ϵ is, the more spread out the posterior is, leading to a looser inference. Bi et al. [71] suggest that $\epsilon \in [10^{-3}, 10^{-1}]$.

A common distance metric is the Euclidean distance, which would correspond to the distance between the means. If the Euclidean distance is selected for d , this corresponds to the standard approximate Bayesian computing. However, this discards important information about the shapes and dependencies. A distance metric which can capture the entire shape of the distributions is ideal. For d we have selected to use the area metric [109]: the horizontal integral, or the area, between the empirical distribution functions of D and samples of the predictive distribution. The area metric is defined as:

$$d(F_P, F_{y_t}) = \int_{-\infty}^{+\infty} |F_P(x|e) - F_{y_t}(x)| dx \quad (7)$$

where $F_P(x|e)$ and F_{y_t} are the empirical cumulative distribution functions of the prediction and data at time t respectively. The area is also commonly known as the L_1 Wasserstein metric [73]. The area metric has some favourable properties for updating.

- It is unbounded, therefore it can quantify the discrepancy between two distributions no matter their relative locations or overlap.
- It returns the Euclidean distance for scalars (i.e. degenerate distributions).
- It is sensitive to the entire shape of the distributions, and not only their moments.
- It can be computed efficiently, and from sample data.
- It can be generalised to higher dimensions.
- It satisfies all the mathematical properties of a metric.
- It applies to non-parametric distributions.
- It is units preserving (returns the physical units of the distribution).

Alternative metrics commonly used in updating include the Kolmogorov–Smirnov distance or the Bhattacharyya distance, but these metrics are bounded, therefore require a 2-step updating procedure [71]. Since the area metric is unbounded we can perform the Bayesian updating in a single step. Eq. (6) is computed for each time point in the output, the likelihood for a sample of the epistemic space e , is computed by taking their product.

This requires n_t evaluations of the integral Eq. (7), one for each time step $t = 0, 1, 2, \dots, n_t dt$, and add significant cost to the Bayesian updating. A dimensionality reduction technique can be used to reduce the dimension n_t and the computational cost of the calibration, see e.g. PCA or Karhunen–Lóeve expansion [110]. In this work, we reduce the dimensionality n_t by Fourier transformations of the discrete-time stochastic processes in D . We then select a subset of harmonics ($\ll n_t$) in the Fourier domain that are sufficient to explain the variability in the data.

Samples of the posterior may be simulated with a MCMC algorithm. For this work we used TMCMC [76], an MCMC algorithm which works well for high dimensional problems such as this. This procedure is expensive, since an entire Monte Carlo loop is required for a single evaluation of the likelihood. We therefore draw fewer samples from the posterior using TMCMC, and propose an efficient method for drawing further samples from the posterior based on sliced-normal distributions (see Section 5.4.1), to increase the accuracy of the uncertainty propagation. This is explored in the following section.

5.4. Uncertainty propagation

5.4.1. Sliced-normal sampling

In this framework, we use a sliced-normal distribution to approximate the Bayesian posterior [5]. This serves as a density estimator that is used to generate a larger posterior sample set for propagation. This allows for a more robust propagation of uncertainty through this sampling approach when the generation of a sufficient sample set directly from the Bayesian posterior is prohibitive computationally.

A sliced-normal is generated by extending the parameter space from the $n_p = 24$ physical space of the data vector $e \in E$ to an n_q dimensional feature space vector $q \in \mathbb{R}^{n_q}$, where $n_q = \binom{n_p + d}{n_p} - 1$ and d is a specified degree of freedom for a monomial expansion function $q(e; d)$. For example, for a two-dimensional vector in physical space (x, y) , a monomial expansion in $d = 2$ degrees of freedom would produce a five-dimensional feature space vector $q(x, y; 2) = (x, y, xx, xy, yy)$. This has the effect of ‘bending’ the physical space such that the distribution of data more closely resembles a feature space Gaussian distribution.

The mean μ_q and covariance matrix Σ_q of the data in feature space can be estimated from a length N_e set of posterior samples of e_j , $j = 1, 2, \dots, N_e$ in a number of ways. We compare two methods in terms of their level set coverage, the maximum-likelihood, and worst-case maximum methods described in [5].

Once the mean and covariance μ_q, Σ_q of this feature space Gaussian distribution have been determined, the feature space density f_q of a mapped point $q(e; d)$ can be used as an estimate of the unnormalised physical space density f_e :

$$f_e(e) \propto f_q(q; \mu_q, \Sigma_q). \tag{8}$$

The constant missing in Eq. (8) is difficult to compute because it requires the solution of a multidimensional integral over a polynomial manifold given by $q(e; d)$. However, the unnormalised densities are sufficient for producing samples from a TMCMC process. Evaluating the density of each point e using the sliced-normal is efficient since it only requires a monomial expansion and a density calculation on a single Gaussian. Because of this, the TMCMC process can proceed at a much faster pace than with the Bayesian posterior, allowing for many thousands of samples to be drawn for forward propagation.

5.4.2. Sliced-normal validation

If the sliced-normal is calibrated appropriately, the level sets drawn from the relative densities of the data allows one to maintain the coverage criterion of a suitable confidence structure. To check the coverage of this structure, the distribution of the p-values must stochastically dominate a uniform CDF. A p-value represents the probability of observing an event at least as extreme as some proposal. The p-values can be used to define confidence level sets $C(\alpha)$ that should contain a future sample with probability $\geq \alpha$. This is represented by:

$$C(\alpha) = \{e \in E; \beta(e; e_j) \geq 1 - \alpha, j = 1, \dots, N_e\}. \tag{9}$$

where, α is the desired confidence level, and $\beta(e; e_j) = B^{-1}(\alpha_B; k + 1, N_e - k)$ is the inverse upper Clopper–Pearson bounding distribution evaluated at a second confidence level α_B , with k defined as the number of samples having a Mahalanobis distance greater than the equivalent distance for e .

The minimum confidence level required to contain a remaining sample in a leave-one-out validation process is given by $1 - \beta(e; e_j)$. The available sample set is shuffled and split in two, one set for training $e_{1:N_e-1}$ and one for testing e_{N_e} . The minimum required confidence to cover e_{N_e} is then calculated. The training and sample sets are then reshuffled, and the process is repeated a number of times in order to estimate the distribution of minimum confidence levels, which is visualised in the left of Fig. 12. This provides evidence that the sliced-normal distribution is capable of covering the samples with a minimum frequency. Therefore the approach used here to calculate the optimal sliced-normal parameters matches the maximum-likelihood approach of [111].

5.5. The double loop MC method

The Monte-Carlo (MC) simulation is one of the most widely used methods to propagate aleatory uncertainty. With MC it is possible to treat the model under study as a black box, and makes this method relatively easy to utilise in almost any problem.

When a mixture of aleatory and epistemic uncertainty is present in the input, characterised by second-order distributions or p-boxes, it is still use the Monte-Carlo with a modification. A number of different modifications have been proposed, including double-loop Monte-Carlo [39], generalized importance sampling and interval Monte Carlo method [112], slicing method [43], multi-level meta-models [97] and, for a special classes of problems, analytical methods [113].

The double-loop MC approach is composed of two nested loops. The outer loop explores E whilst the inner loop is used to propagate the aleatory uncertainty. Each sample of the outer loop allows us to define a random multivariate distribution, F_A , which is then sampled in the inner loop. Repeating this process produces an ensemble of F_A . Fig. 3 summarises this process of sampling from a second-order distribution using double loop MC. First, N_e epistemic samples, e , are obtained from the epistemic box E , either with Latin hyper-cube sampling or with the available samples from the calibration procedure. Each of these epistemic samples defines an epistemic point $e_{1:4,j}$, the green dot in Fig. 3, and an j th aleatory distribution defined by a set of marginal distributions (parameters $e_{5:14,j}$), and a copula (parameters $e_{15:24,j}$), the orange dot in Fig. 3. The aleatory distribution is then sampled by correlated inverse transform sampling. This involves drawing samples from the copula, the blue dot in Fig. 3, and then by inverting them through the marginal CDFs. These transformed samples will then follow the correlated multivariate distribution. In the case of a Gaussian copula, a well known method can be applied to generate its samples, i.e. first generating samples of a multivariate Gaussian (using Cholesky decomposition) and then performing an isoprobabilistic transform through the standard normal CDFs.

The N_a samples of F_A are evaluated through the model with the epistemic point $e_{1:4,j}$, producing a distribution in the output space. For each iteration $j = 1, \dots, N_e$ in the outer loop, a new random F_A is created along with a new vector $e_{1:4}$, which populate two

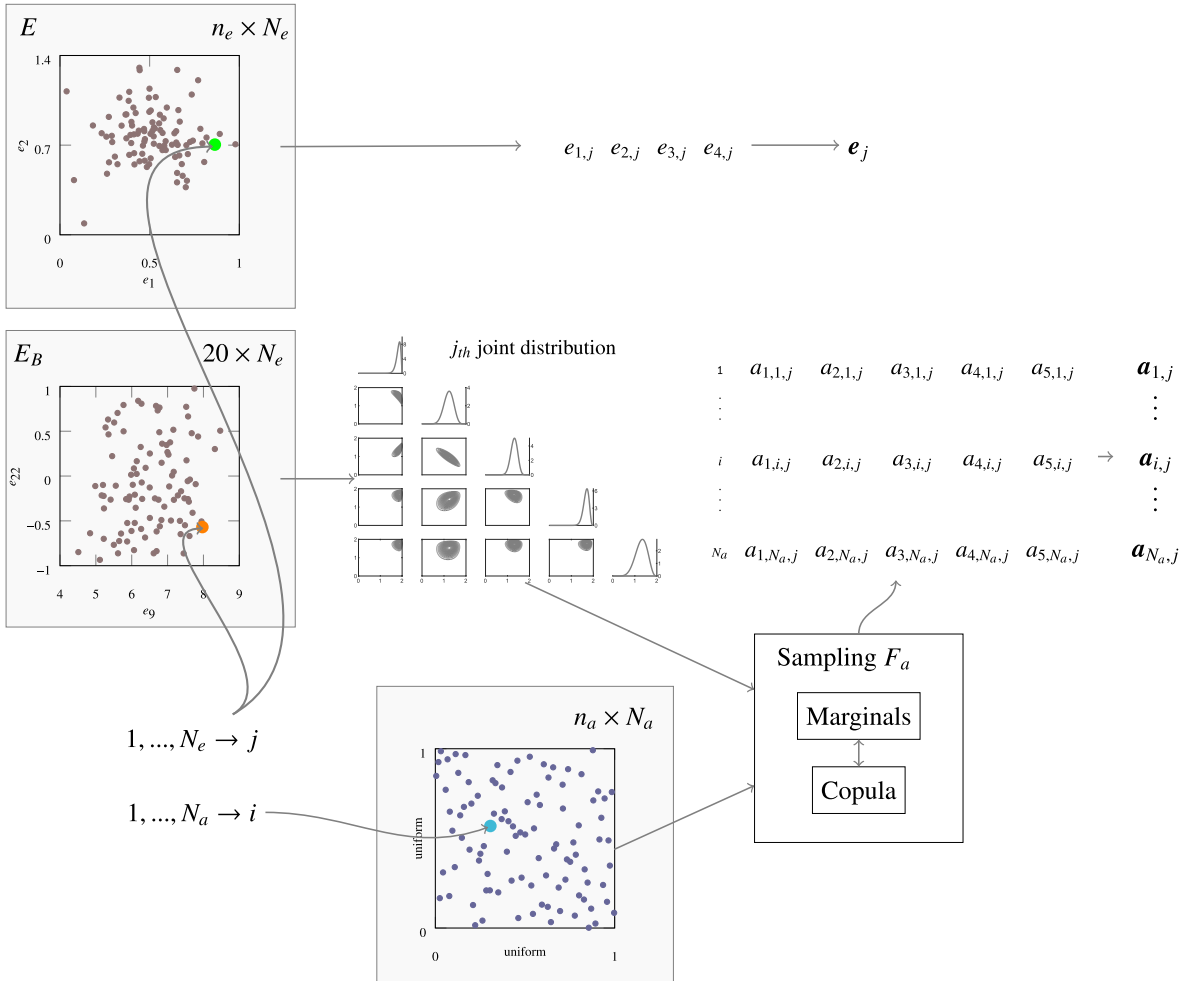


Fig. 3. A visual summary of the double-loop sampling process for the second-order distribution model of the uncertainty. The process loops over all the indices i (inner loop) and over all indices j (outer loop).

matrices: a matrix of size $n_e \times N_e$ with the epistemic samples, and a matrix of size $n_a \times N_a \times N_e$ with the aleatory samples (see Fig. 4). The black-box model (system and performance functions) is evaluated on these matrices, producing the output shown in Fig. 4. A p-box is then constructed by enveloping the MC ensemble predictions over the N_e dimension. The rows of the output matrices (e.g. y, z and g) are samples from a unique probability distribution which can then be enveloped to create a p-box.

In this work we generate samples of e according to a sliced-normal model of the posterior, see Section 5.4.1.

5.5.1. Interval Monte Carlo

Representing the aleatory space with a p-box presents challenges for sampling. However, this can be overcome with an interval Monte Carlo approach. The interval Monte Carlo method can be conveniently adopted if the input p-boxes are non-parametric, i.e., only upper and lower CDF bounds are available. It also slackens a distributional assumption on F_A , e.g., a Beta assumption. This allows us to simulate samples from distributions in the p-box which may be non-Beta. This can be useful to check the validity of the double loop MC results if a non-parametric, distribution-free, model is adopted.

The *interval Monte Carlo* [114], also known as *slicing method*, *alpha-cut method*, or *random set sampling* [22] is a simulation procedure to propagate p-boxes. Unlike the double loop Monte-Carlo, the slicing method samples intervals from the real-line input p-box rather than point-valued realizations from a random distribution F_A . Intervals are propagated through a black-box model by computing the minimum and the maximum of the output quantities of interest.

In order to illustrate the slicing method we will make use of two indices: $i = 1, \dots, n_a$ to index the i^{th} dimension of the aleatory vector, here denoted by x , and $j = 1, \dots, N_a$ to index the j^{th} simulated sample, i.e. a slice. The slicing method operates as follows:

1. Produce uniform correlated samples, $\alpha_j \in (0, 1]^{n_a}$, $j = 1, \dots, N_a$, from a copula function $\alpha_j \sim \mathcal{C}$.
2. Use the inverse CDFs $\bar{F}_X^{-1}(\alpha)$ and $F_X^{-1}(\alpha)$ of the bounds of the p-box to generate a *slice*. The slice of the joint correlated p-box is also called a *focal element*, and it is obtained with the following Cartesian product:

$$\mathcal{J}_j := \prod_{i=1}^{n_a} \left[F_{X_i}^{-1}(\alpha_{ij}), \bar{F}_{X_i}^{-1}(\alpha_{ij}) \right], j = 1, \dots, N_a \tag{10}$$

Notice that for the slicing method, a focal element \mathcal{J} is equivalent to an MC sample vector.

3. An inner approximation of the interval valued response of the output from the performance function $g : \mathbb{R}^{n_a} \rightarrow \mathbb{R}^{n_g}$ is computed as follows:

$$\bar{g}_j = \max_x g(x), \min_x g(x), \text{ subject to } x \in \mathcal{J}_j \subseteq \mathbb{R}^{n_a}, j = 1, \dots, N_a, \tag{11}$$

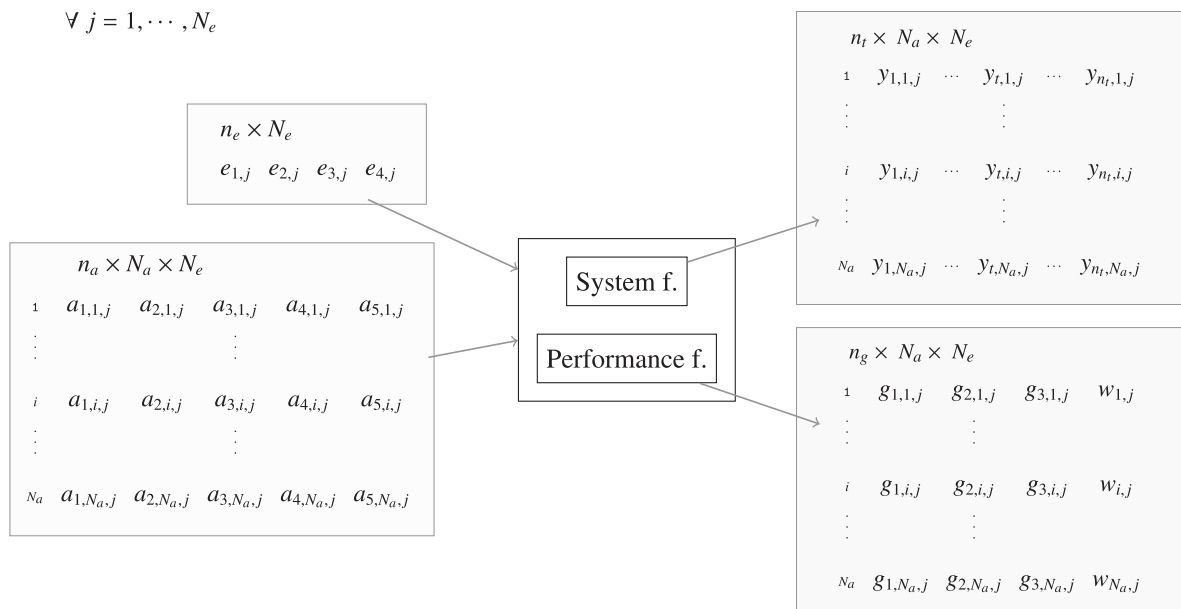


Fig. 4. Input and output tensors through the system and performance function from the second-order distribution samples. For example, in the NASA problem $n_a = 5$, $n_e = 4$, $n_t = 5000$, and $n_g = 3$; while the number of samples N_a and N_e varies depending on the analysis.

4. The p-box of each individual component of the output can be obtained via simulation using the indicator function $\mathbf{1}(c)$, which is 1 if the condition c is true and 0 otherwise:

$$F_G(g) = \frac{1}{N_a} \sum_{j=1}^{N_a} \mathbf{1}(g \leq \underline{g}_j), \quad \text{and} \quad \bar{F}_G(g) = \frac{1}{N_a} \sum_{j=1}^{N_a} \mathbf{1}(g \leq \bar{g}_j), \tag{12}$$

Different techniques can be used to approximate Eq. (11), for instance, sampling methods, the vertex method, or global optimisation.

5.6. Sensitivity analysis

With the means of sampling defined, we can propagate the uncertainty in the input through the model. In many practical cases, it is important to understand the effect of different input uncertainties on the output. In this work, we utilised two different sensitivity analyses methods: variance-based and value-of-information. The theory behind these approaches can be found in Section 3.4. The motivation for utilising two distinct methods was to investigate the effect of different measures of sensitivity and to enhance the robustness of our conclusions.

5.6.1. Value-of-information sensitivity analysis

The value-of-information (VoI) sensitivity analysis aims to assess the possible reduction in uncertainty about the output if extra knowledge about the input were available [3]. This is usually done by comparing the uncertainty in the output before and after reducing the uncertainty in one or more inputs through a process called *pinching*. In this section we propose a VoI sensitivity analysis based on *partial pinching*. Our proposed partial pinching procedure replaces the original uncertain input set with a tighter subset.

This choice is motivated by the need to address the challenge problem, where directional refinements of the epistemic parameters are requested. A directional refinement for a single parameter consists of either: lowering the upper bound, or raising the lower bound of a model parameter.

We face two issues in addressing the NASA challenge uncertainty reduction problem.

1. The output of the system response is stochastic and highly multidimensional.
2. The system is a complex nonlinear black-box function.

The combination of these two issues makes the sensitivity study uniquely challenging within the literature.

VoI sensitivity within in the *probability bounds analysis* (PBA) paradigm is naturally equipped to deal with stochastic systems, however additional research is required to address VoI sensitivity analysis within black-box models.

The VoI score for any uncertainty contraction can be computed with the following simple expression,

$$S^{Vol} = 1 - \frac{unc([F]_R)}{unc([F]_B)}, \tag{13}$$

where, B is the baseline scenario, R corresponds to the scenario with an input partially or totally pinched (or reduced), $[F]$ is the p-box system output, and $unc()$ is a measure of uncertainty of the output.

The result of the expression in Eq. (13) is an estimate of the expected value any additional empirical information about the input might give in terms of its corresponding reduction in uncertainty of the output. This score belongs to the $[0, 1]$ interval. When the pinching is applied to each input quantity in turn we obtain a relative rank. The highest ranked input is the largest contributor to the output uncertainty. The sum of the VoI scores for each input pinch, unlike the factorization used in variance-based sensitivity analyses, does not generally add up to 1.

Because we are in the context of PBA, the obvious choice for the $unc()$ measure is the area between the upper and lower bounds of the output p-box. This measure is the previously described stochastic area metric (Wasserstein distance), now applied to the bounds of a p-box:

$$unc\left(\left[F\right]\right) := W\left(\left[F\right]\right) := \int_{-\infty}^{+\infty} |\bar{F}(x) - \underline{F}(x)| dx. \tag{14}$$

In this work, the sensitivity analysis is conducted to understand how the input epistemic vector \mathbf{e} influences the output of the system functions y, z_1 , and z_2 . Because a sensitivity score is always computed associating a reduction in the input space (epistemic vector) to a reduction in the output space (system response), we introduce the following notation to uniquely denote each sensitivity score.

$$S_{e_i \rightarrow \gamma}^{Vol} := 1 - \frac{W\left([F]_{e_i}\right)}{W\left([F]_B\right)}, \quad i = 1, \dots, n_e, \tag{15}$$

where $\gamma \in \{y, z_1, z_2\}$ is used as a place holder symbol for the output.

Since the dimension of the output vectors $y, z_1, z_2 \in \mathbb{R}^{5000}$ is large, we analyse the sensitivities in the Fourier domain, where the dimensions can be reduced to about 60 harmonics. The combined sensitivity scores of Eq. (15) are obtained in the Fourier space, by summing up the area metric corresponding to each of the 60 harmonics. This procedure looks at the marginal output distributions individually, so it cannot capture the effect of dependence between the harmonics.

5.6.2. Numerical strategy to compute the VoI scores

As pointed out in [115], PBA is global sensitivity analysis. However, the presence of a black-box code poses an additional challenge: it is not possible to propagate an interval vector of parameters while guaranteeing that the output is *inclusion monotone*. While the interval extension of the function representing the black-box may itself be monotone, this property cannot be ensured via sampling.

An interval function f is inclusion monotone if for any two interval vectors $[x], [y] \subseteq \mathbb{R}^d$ holds,

$$[x] \subseteq [y] \Rightarrow f([x]) \subseteq f([y]), \text{ component wise.} \tag{16}$$

With sampling the above rule is violated, because coverage (minimum and maximum are included in the bounds) is only ensured when the function is evaluated with interval arithmetic rules. Therefore, it is not possible to consistently compute sensitivity scores using PBA with sampling, potentially leading to negative scores. The more samples the function is evaluated on, the better the approximation. The use of a global optimiser to obtain the bounds on likely a better one, each focal element would also lead to good approximation, although with enormous computational effort. To overcome this limitation, we devise a sampling strategy that emulates the inclusion property of Eq. (16). This strategy is based on a caching algorithm that stores the result of a large number of point valued input/output simulations. While the proposed strategy emulates the property of Eq. (16) of interval analysis, its bounding capability is still tied to sampling, so it is still an inner approximation. Convergence to the target bounds cannot be guaranteed in a black-box system, nevertheless an approximation can be computed with a significantly high amount of samples, in the order of thousands. The strategy has its highest efficiency when the epistemic uncertainty is predominant over the aleatory one, and gets less efficient as the focal elements get thinner. In other terms, this strategy works well when the p-boxes are fat as opposed to thin. The caching algorithm computes the bounds of each focal element on stored samples produced upfront, it emulates inclusion monotonicity and removes the need for re-sampling.

The algorithm implementing this strategy is summarised in Fig. 5. From left to right: (i) first, each p-box is discretised into a number of focal elements with equal mass, (ii) [lower] these are then combined by Cartesian product under random set independence, this results in the gray box shown over the blue scatter points, (iii) [higher] is the partial pinching (halving) of e_1 , (iv) the caching algorithm executes the propagation using the samples contained in each gray box [of (ii) and (iii)], and (v) finally, the output p-box for the baseline case (blue) and the pinched scenario (purple), corresponding to the lower halving of e_1 .

5.6.3. Variance-based sensitivity analysis

From the probabilistic perspective to uncertainty quantification, the model’s output can be regarded as a random variable, Y , with a

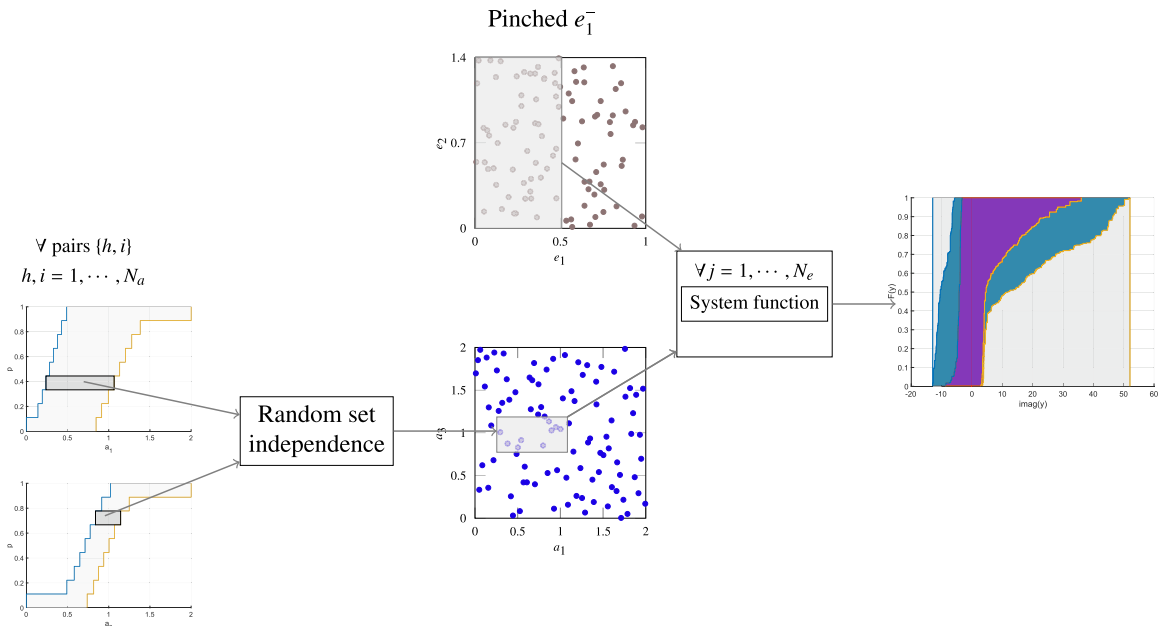


Fig. 5. Visual summary of value-of-information sensitivity procedure with partial pinching and probability bounds analysis. Left to right: focal elements drawn from input p-boxes, caching algorithm for pinching, propagation through the system function to the output probability box, and value-of-information sensitivity score.

corresponding (joint) probability distribution. In scenarios where the variance of this random variable is regarded as a good measure of output uncertainty, the importance ranking of the inputs can be done via variance-based sensitivity analysis. We will denote the epistemic vector $e_{1..4}$ by X , and the output of the model Y .

Since the models considered here are deterministic the variance of Y will be entirely due to the uncertainty in the input values. This leads to the notion that fixing one of the inputs, X_i , to a given value, x_i , and re-running the code results in Y having a smaller variance. Let $\mathbb{V}_{\sim i}[Y|X_i]$ denote the conditional variance of Y taken over all factors, but X_i (denoted $X_{\sim i}$) and given $X_i = x_i$. This conditional variance can be used as a measure of how influential the fixed parameter is. However, the conditional variance is also a function of x_i and is of little practical use in determining input importance. This problem can be resolved by taking the expected value of the conditional variance over all possible values of x_i , that is $\mathbb{E}_i[\mathbb{V}_{\sim i}[Y|X_i]]$. Then, by the law of total variance, the *the first order effect* of X_i on Y is given by $\mathbb{V}_i[\mathbb{E}_{\sim i}[Y|X_i]] = \mathbb{V}[Y] - \mathbb{E}_i[\mathbb{V}_{\sim i}[Y|X_i]]$. To improve interpretability, the first order effect can be normalized by the unconditional variance of the output, to give what is known as the first order *Sobol' index*, [88]

$$S_i = \frac{\mathbb{V}_i[\mathbb{E}_{\sim i}[Y|X_i]]}{\mathbb{V}(Y)} \tag{17}$$

A high value of the Sobol' index for the given variable, means that its contribution to the output variance, without regard to any interactions with other inputs, is important.

Often, it would be desirable to inspect the importance of interactions among inputs. In such cases, the *higher order* Sobol' indices can be constructed in much the same way as the first order ones. Let $p \subseteq \{1, \dots, n_e\}$ be a set of all inputs under investigation. Then, the $|p|^{th}$ order index is given by

$$S_p = \frac{\mathbb{V}_p[\mathbb{E}_{\sim p}[Y|\mathbf{X}_p]]}{\mathbb{V}(Y)} - \sum_{j \subset p} S_j \tag{18}$$

where $|\cdot|$ denotes the cardinality of a set. For a model with d inputs, it is easy to show that there are $2^d - 1$ possible combinations for p - too many to be considered one by one, even for moderately sized d . To summarise the effect of interactions, one can compute the *total Sobol' index* [116],

$$S_{T_i} = 1 - \frac{\mathbb{V}_{\mathbf{x}_{\sim i}}[\mathbb{E}_{x_i}[Y|\mathbf{X}_{\sim i}]]}{\mathbb{V}(Y)} \tag{19}$$

This metric captures the effect of the i^{th} input and all of its interactions. An efficient way to compute Sobol' indices is discussed at length in [89].

For a model with independent inputs, the sensitivity indices provide a decomposition for the variance of the model's output. That is, they obey:

$$\sum_{i=1}^d S_i + \sum_{i < j}^d S_{i,j} + \sum_{i < j < k}^d S_{i,j,k} + \dots + S_{1,2,\dots,d} = 1 \tag{20}$$

In such models it also holds that $0 < S_i \leq S_{T_i} < 1$, where equality can only arise in the case of a perfectly additive model. When the inputs are correlated, the results of the decomposition in Eq. (20) may include contributions from inputs that are outside of those currently analysed. In this case, the alternative decomposition introduced in [117] can be used. In this decomposition, which is a generalisation of that in Eq. (20), the covariance among inputs is taken into account. The reader is referred to [118] for further discussion of variance-based SA with correlated inputs.

For the case where the output is a multidimensional variable in the time domain, we can estimate the sensitivity indices for each time step and then compute a mean value

$$\bar{S}_{T_i} = \frac{1}{n_t} \sum_t^{n_t} S_{T_i}(t)$$

and a corresponding standard error

$$\mathcal{S}_{T_i} = \sqrt{\frac{1}{n_t^2(n_t - 1)} \sum_t^{n_t} (S_{T_i}(t) - \bar{S}_{T_i})^2}$$

of an aggregated sensitivity metric across time.

Standard error is used to produce error bars on each total/local sensitivity index. This provides a sense of the amount of variability in the value of each index across the time dimension.

5.7. Reliability analysis and reliability-based design optimisation

Calibration, propagation and sensitivity analysis all allow for a more accurate characterisation of the uncertainty model, and this is motivated by a desire to use this information to improve design decisions and prototyping. To do so, we use reliability analysis and an iterative design optimisation process. Both of these require assessment of the reliability performance of a design, and this framework provides a mean of doing so in the presence of a mixture of both aleatory and epistemic uncertainty.

The reliability of a design θ with respect to a set of requirements g_1, \dots, g_{n_g} can be assessed using the *worst-case reliability function*:

$$w(a, e, \theta) = \max_{j \in \{1, \dots, n_g\}} g_j(a, e, \theta), \tag{21}$$

where $g_j : \mathbb{R}^{n_\theta} \times \mathbb{R}^{n_a + n_e} \rightarrow \mathbb{R}$ are the *reliability functions* describing the compliance of the system with respect to requirement $j = 1, \dots, n_g$. A design θ will satisfy all requirements if $w(a, e, \theta) < 0$ and requirement j is satisfied if $g_j(a, e, \theta) < 0$. If e is known and only aleatory uncertainty is affecting the study, the failure probability $P_f(\theta)$ of a design gives the precise probability that the θ complies with the reliability requirements. Traditionally, $P_f(\theta)$ is defined as a multidimensional integral of the probability density of a computed over the failure domain. If both epistemic and aleatory uncertainty affect the study, a generalized expression for the joint failure probability is given by

$$P_f(e, \theta) = \mathbb{P}[w(a, e, \theta) \geq 0], \tag{22}$$

and for the individual failure probabilities is defined as follows:

$$P_{f,j}(e, \theta) = \mathbb{P}[g_j(a, e, \theta) \geq 0], \quad j = 1, \dots, n_g, \tag{23}$$

where the probability of failure is a now function of e . Note that to estimate $\mathbb{P}[\cdot]$ as a precise probability, the distribution of the aleatory variables and the value of e must be fixed. If e is only known to be contained in some set or the distribution of the aleatory variables is imprecise, the probability of failure will be an interval. A sample-based estimate of $P_f(e, \theta)$, obtained from a set of n samples a_1, a_2, \dots, a_n from a precise $F_A(a; e_{5.24})$, is given by:

$$\hat{P}_f(e, \theta) = \frac{1}{n} \sum_{i=1}^n \mathbf{1}_{c_i}, \tag{24}$$

where $\mathbf{1}_{c_i}$ is the indicator function for the condition c_i , defined as $w(a_i, e, \theta) \geq 0$.

The minimization of the probability in Eq. (23) is often used as the optimisation criterion in traditional RBDO methods. However, the resulting designs do not preclude severe failures since the positive values of $w(a, e, \theta)$ can be arbitrarily large. The extent by which $w(a, e, \theta)$ exceeds the reliability threshold $w(a, e, \theta) = 0$ measures the severity of failure. The analyst might want to consider not only a minimization of the failure probability, but also the shape of the upper tail of the distribution of $w(a, e, \theta)$. This design criterion can be enforced by minimizing the conditional tail expectation given by:

$$\sigma(e, \theta) = \mathbb{E}[w(a, e, \theta) | w(a, e, \theta) \geq 0], \tag{25}$$

where $\sigma(e, \theta)$ is the conditional expectation of $w(a, e, \theta)$ over the failure domain, i.e., the average magnitude of the violations. For a presentation of sample-based approximation of $\sigma(e, \theta)$ and a discussion on confidence bounds and the computational issues involved with it see, e.g., [119].

Note that the model of the uncertainty f_a is a key component in the procedure and a large sample size is generally required to improve accuracy and convergence of the failure probability and tail expectation estimates. In addition, if epistemic uncertainty is present a precise estimation of the reliability cannot be obtained. To account for this, epistemic bounds on the joint and individual failure probabilities, P_f and $P_{f,j}(e, \theta)$, respectively, of a proposed design θ are computed as follows:

$$R_w(\theta) = \left[\min_{e \in E} P_f(e, \theta), \max_{e \in E} P_f(e, \theta) \right], \tag{26}$$

$$R_j(\theta) = \left[\min_{e \in E} P_{f,j}(e, \theta), \max_{e \in E} P_{f,j}(e, \theta) \right], \quad j = 1, 2, \dots, n_g. \tag{27}$$

If the resulting probability interval is vacuous, i.e., $R_j(\theta) = [0, 1]$, $j = 1, \dots, n_g$, the epistemic uncertainty in the problem is preventing us from answering any questions about the reliability of the design θ .

Similarly, the severity score of the integrated system for all requirements, $s(\theta)$, and for individual requirements, $s_j(\theta)$, are computed as follows:

$$s_w(\theta) = \max_{e \in E} \sigma(e, \theta) \times P_f(e, \theta), \tag{28}$$

$$s_j(\theta) = \max_{e \in E} \sigma_j(e, \theta) \times P_{f,j}(e, \theta), \quad j = 1, 2, \dots, n_g. \quad (29)$$

Note that the value of the product $\sigma(e, \theta) \times P_f(e, \theta)$ can be regarded as a classical measure of the risk for the design θ for an $e \in E$, that is, it quantifies not only the probability of incurring a failure but also the expected magnitude of that failure given by $\sigma(e, \theta)$. Hence, the severity score $s_w(\theta)$ can be interpreted as the worst-case risk computed over all possible values of the epistemic parameters $e \in E$. While the width of R_j indicates the degree of epistemic uncertainty, s_j is an unbounded value from above. Increasing the epistemic uncertainty in a model will strictly increase s_j , allowing for relative comparisons, but a single severity score does not have a clear interpretation.

Two levels of approximation jointly affect the estimates of $R_w(\theta)$ and $s(\theta)$:

- i) sample-based approximation of $\mathbb{P}[\cdot]$ and $\mathbb{E}[\cdot]$.
- ii) approximation for the minimum and maximum operators over the epistemic space E .

The problem in ii) is aggravated by the presence of strong non-linearities in $g_j, P_f(e, \theta)$ and $\sigma(e, \theta)$ that, combined with poor coverage of the epistemic domain leads to inner approximations of the bounds of these operators. The accuracy of these inner approximations can be improved by the methods outlined in Section 5.4.1 and Section 5.5.

These numerical issues further complicate the reliability-based design in the presence of mixed aleatory-epistemic uncertainties. In this work, we seek reliable designs by solving the following optimisation programs:

$$\min_{\theta \in \Theta} \max_{e \in E} P_f(e, \theta), \quad (30)$$

$$\min_{\theta \in \Theta} s_w(\theta). \quad (31)$$

The optimisation program in Eq. (30) minimizes the right epistemic bound of $R_w(\theta)$ whilst the one in Eq. (31) minimizes the severity for all requirements. These programs are non-convex and highly discontinuous in Θ , which precludes the use of gradient-based optimisation methods. The approach we adopted instead, is based on repeated runs of a genetic algorithm (GA). The fittest candidates in the final θ populations for the two programs were pooled together and ranked according to the resulting reliability and severity scores, given by Eqs. (26)–(29). Note that, although samples are very cheap to produce thanks to the SN approach, the computational cost needed to estimate the severity and epistemic bounds on the reliability of a design can be very high. This problem is particularly serious for small failure probabilities and numerically expensive reliability functions g . In these situations, decoupled RBDO approaches, numerically efficient methods for efficient estimation of low-failure probabilities and multi-level high fidelity models can be considered to reduce the computational cost of the RBDO.

5.8. Risk-based design optimisation

Traditionally, risk-based design optimisation methods seek system designs by minimizing a combination of probability and severity of failures. In this work, we look at a risk-based design from a new perspective. A risk-based design θ_{risk} is a design that only accounts for a portion of the remaining epistemic space. A risk parameter $r \in [0, 100]$ quantifies a reduction in epistemic volume. For example, a risk $r = 10$ means that we only retain 90% of the volume of the epistemic space. We will refer to the retained volume as $E(r)$. Clearly, we have that $E(r) \subseteq E(0) = E$ for any $r > 0$. The risk-based design θ_{risk} must be compared to the optimal RBDO design (θ_{final}) and the resulting reliability scores must be used to either adopt or reject a risk-based design solution. For a more detailed description of the problem refer to the NASA challenge [6Problem F]. In this work, we investigate two metrics to quantify a gain from taking a risk r . The gain metrics are defined as follows:

$$l_{p_j}(r, \theta) = 1 + \left(\max_{e \in E} P_f(e, \theta_{final}) - \max_{e \in E(r)} P_f(e, \theta) \right), \quad (32)$$

$$l_s(r, \theta) = 1 + \left(\max_{e \in E} P_f(e, \theta_{final}) \times \sigma(e, \theta) - \max_{e \in E(r)} P_f(e, \theta) \times \sigma(e, \theta) \right), \quad (33)$$

where $l_{p_j}(r, \theta)$ and $l_s(r, \theta)$ measure the improvement in reliability (or severity) achieved by a candidate θ , in comparison to the best available design θ_{final} , when a risk r is taken. A $l_{p_j}(r, \theta) < 1$ means that the worst-case reliability of a candidate design θ is lower than the one of final design computed over E , i.e.,

$$\max_{e \in E} P_f(e, \theta_{final}) < \max_{e \in E(r)} P_f(e, \theta).$$

Similarly $l_s > 1$ implies a reduction in severity with respect to θ_{final} . Given a risk r , a risk-based design θ_{risk} results in superior reliability compared to θ_{final} if $l_{p_j}(r, \theta_{risk}) > l_{p_j}(r, \theta_{final})$ or, in words, if the maximum failure probability of the risk-based design computed over the risk-taking epistemic space $E(r)$ is lower than the one of the final design computed over the same epistemic volume

design. A risk-based design will result superior performance with respect to the severity score if $I_s(r, \theta_{risk}) > I_s(r, \theta_{final})$.

6. The 2020 NASA black-box challenge

In this section, it will be assumed that the reader is familiar with the NASA challenge [6]. A distinction between aleatory and epistemic uncertainty is made in the challenge statement (see Section 2), thus the two kinds will be treated differently. Such a distinction carries strong practical implications for the solution strategy. In this section we implement the solution strategy described in Section 5 to address the NASA challenge. Fig. 6 shows a visual representation of the NASA challenge for reference. The relation of the different system models, their inputs, and interactions are shown in the context of each of the problem tasks presented in Section 2. Here, we describe the solution sequence in corresponding order of these problem tasks.

6.1. Problem A: Model calibration & uncertainty quantification

Problem A of the challenge is to calibrate the model’s input for the limited experimental data that is provided. No aleatory distribution is provided, so the beta model described in Section 5.2 is adopted to attempt to envelope the true distribution F_A . Therefore, a calibrated uncertainty model will constitute a set within E describing a five-dimensional probability box along with a reduced volume for the four epistemic inputs in $[0, 2]^4$. The calibration data provided represents 100 realisations of a discrete-time stochastic process. The provided data is neither Gaussian nor stationary. In this section, the calibration of UM is performed against the first set of data $D_1 = \{y_t^{(i)}, t = 0, \dots, 5000\}_{i=1}^{100}$. UM_y is the resulting UM based on D_1 .

UM_z is the second calibrated UM based on $D_2 = \{z_t^{(i)}, t = 0, \dots, 5000\}_{i=1}^{100}$. From our perspective, D_2 was unavailable to us until the initial design decision had been made based on UM_y ; this mimics the experience of a real-world design process.

As previously described, the problem of calibrating F_A and the four-dimensional epistemic box was overcome by selecting the parametric distribution Eq. (3), and updating its parameters along with $e_{1:4}$. Under this parametrisation 20 additional dimensions are added to the original four epistemic inputs in the updating procedure. A summary of these parameters and their meaning is given in Table 2.

6.1.1. Uncertainty calibration

Under the parametrisation presented previously, an approximate Bayesian calibration technique as described in Section 5.3 may be performed over E . This method requires prior $P(e)$ and a likelihood, $P(D|e)$ to be defined. The prior was selected to be a uniform distribution bounded to the ranges given in Table 2. The likelihood was estimated using the procedure described in Section 5.3.1.

The first calibration procedure was performed over D_1 and defines UM_y . This calibration involves drawing realisations of E to produce a distribution of F_A from which samples can be generated. These were then propagated through the system model y and compared against the available data D_1 , as shown in Fig. 7.

Calculating the likelihood over the time domain would require 5001 evaluations of the integral Eq. (7), and would add significant cost to the Bayesian updating. To reduce the number of comparisons that must be done, this procedure may be carried out in the Fourier domain. Of the entire spectrum of harmonics, we determined that the output signal was well represented with the first 30, these are shown in Fig. 8. Therefore, if a Fourier transformation is performed of the experimental data and of the predictive distribution only 60 (30 real and 30 imaginary) evaluations of Eqs. (6) and (7) are required.

With a prior and a likelihood now defined, we generate 520 samples of the posterior by TMCMC. Fig. 9 shows the calibrated model

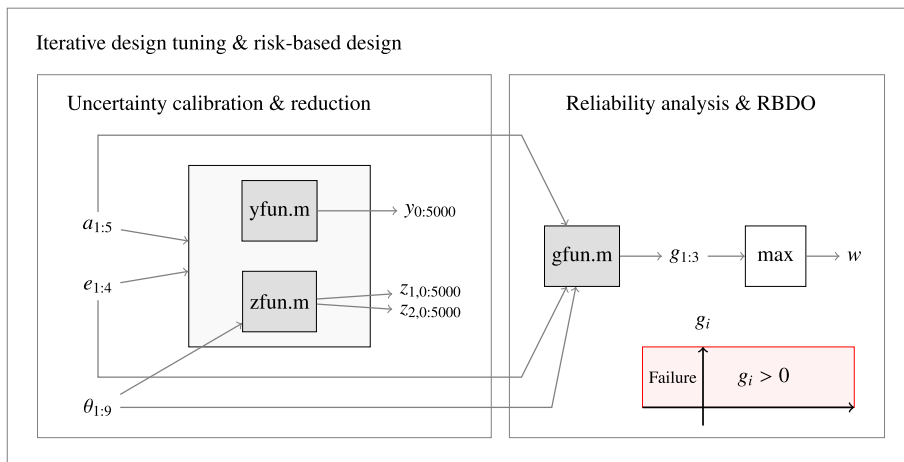


Fig. 6. Schematic of the NASA challenge problem structure and their relation to inputs (a, e), outputs (y, z), Matlab black-box functions (p -function), designs (θ), performance functions ($g_{1:3}$), and worst-case performance function w .

Table 2
Summary of expanded epistemic parameters and their prior bounds.

	Epistemic inputs to the model	Parameters for marginals of F_A	Partial correlations of F_A
Parameter	$e_{1:4}$	$e_{5:14}$	$e_{15:24}$
Prior bounds	[0, 2]	[0, 100]	[-1, 1]

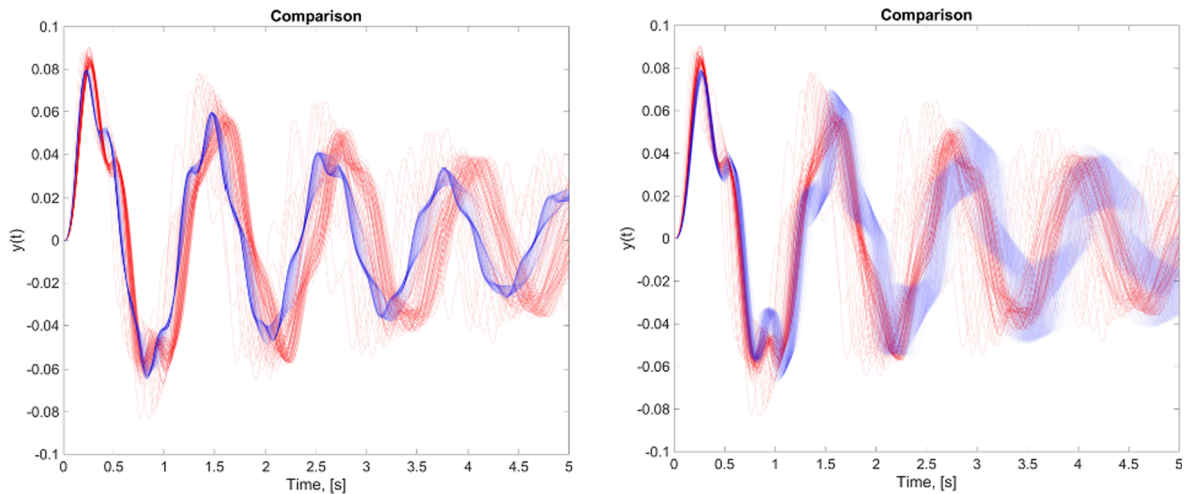


Fig. 7. Two samples (left and right) of the prior for E and 1000 samples from the resulting F_A , evaluated in y_{fun} with the corresponding time histories shown in blue. These are compared with D_1 , shown in red. Note the difference in characteristics between the two stochastic processes, emphasising the flexibility of our uncertainty model.

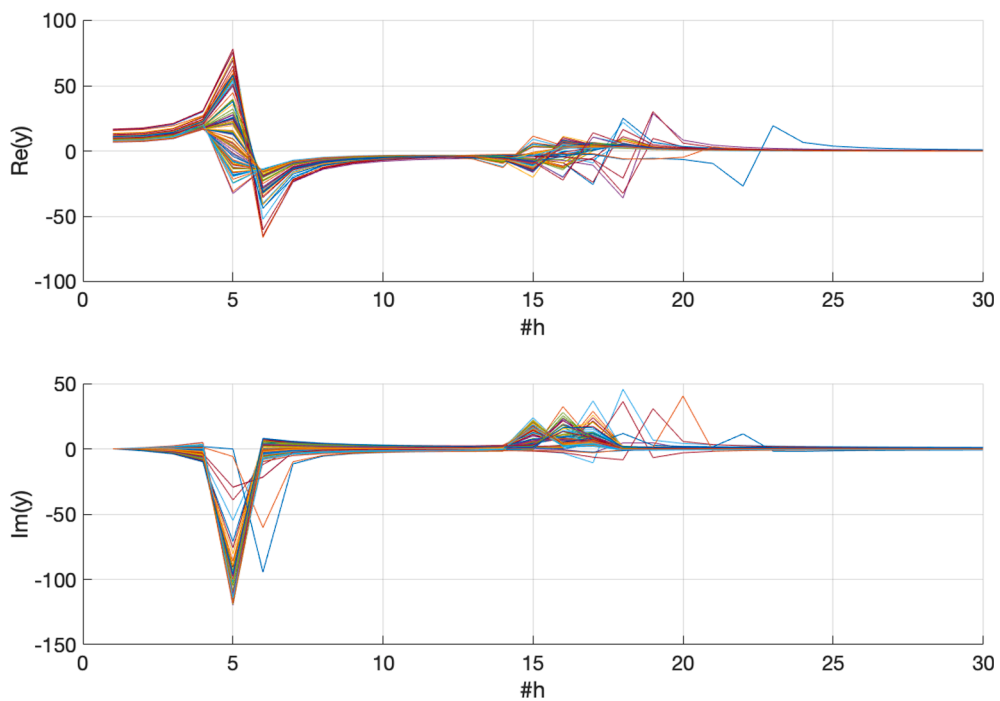


Fig. 8. Fourier transform of the dataset D_1 , showing the first 30 harmonics in the real (top) and imaginary (bottom) domains.

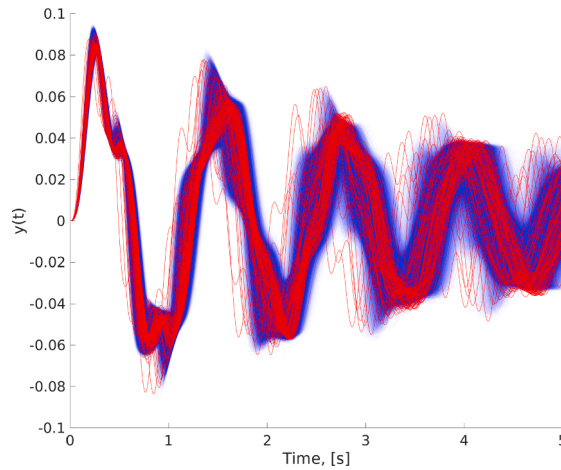


Fig. 9. Realisations of the calibrated model UM_y (blue) against data from D_1 (red). All of the samples of the second-order distribution have been plotted simultaneously.

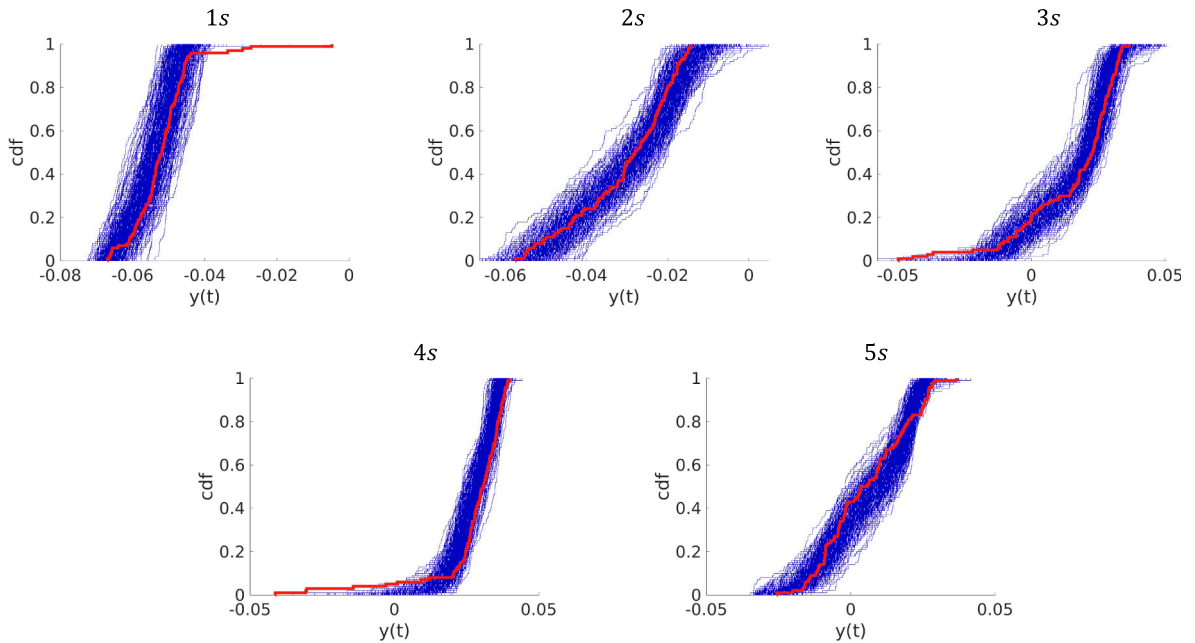


Fig. 10. Comparison of the predictive p-box (blue) from the calibrated model depicted in Fig. 9 to the empirical CDF of the experimental data (red) at various times.

output against D_1 . The output of these samples produces a p-box at each point in time, shown in Fig. 10. It can be seen that the tails of the data are captured by most, but not all, of the predictive p-boxes. There could be several reasons for this. The first, and we believe the main reason, is due to the very high dimensionality of the inputs e . Because the limited number of posterior samples being propagated do not fill the 24-dimensional space very densely, the bounds of the second-order distribution may not have converged. This is one of the main drawbacks of using second-order Monte-Carlo as a propagation method: that a very large number of samples are needed for high dimensions. We have however proposed a cheap method for generating a large number of samples of this 24-dimensional posterior, see Section 5.4.1. The second reason may be that our choice of stochastic metric in Eq. (7) is not very sensitive to the tails of the distribution [109]. If a metric which gives higher weight to the tails of distributions is selected, for example a higher-order Wasserstein distance, Anderson–Darling distance, or Kolmogorov Smirnov, distance we believe this could be improved. This would however introduce further computational complexity.

The calibration procedure for the unknown probability distribution F_A results in a five-dimensional p-box, which envelops the true distribution. Fig. 11 shows this calibrated p-box, with marginal p-boxes shown along the diagonal panels and pairwise bivariate p-boxes in the off-diagonal panels. A multivariate p-box is an extension of p-boxes to higher dimensions, and bound a set of multivariate

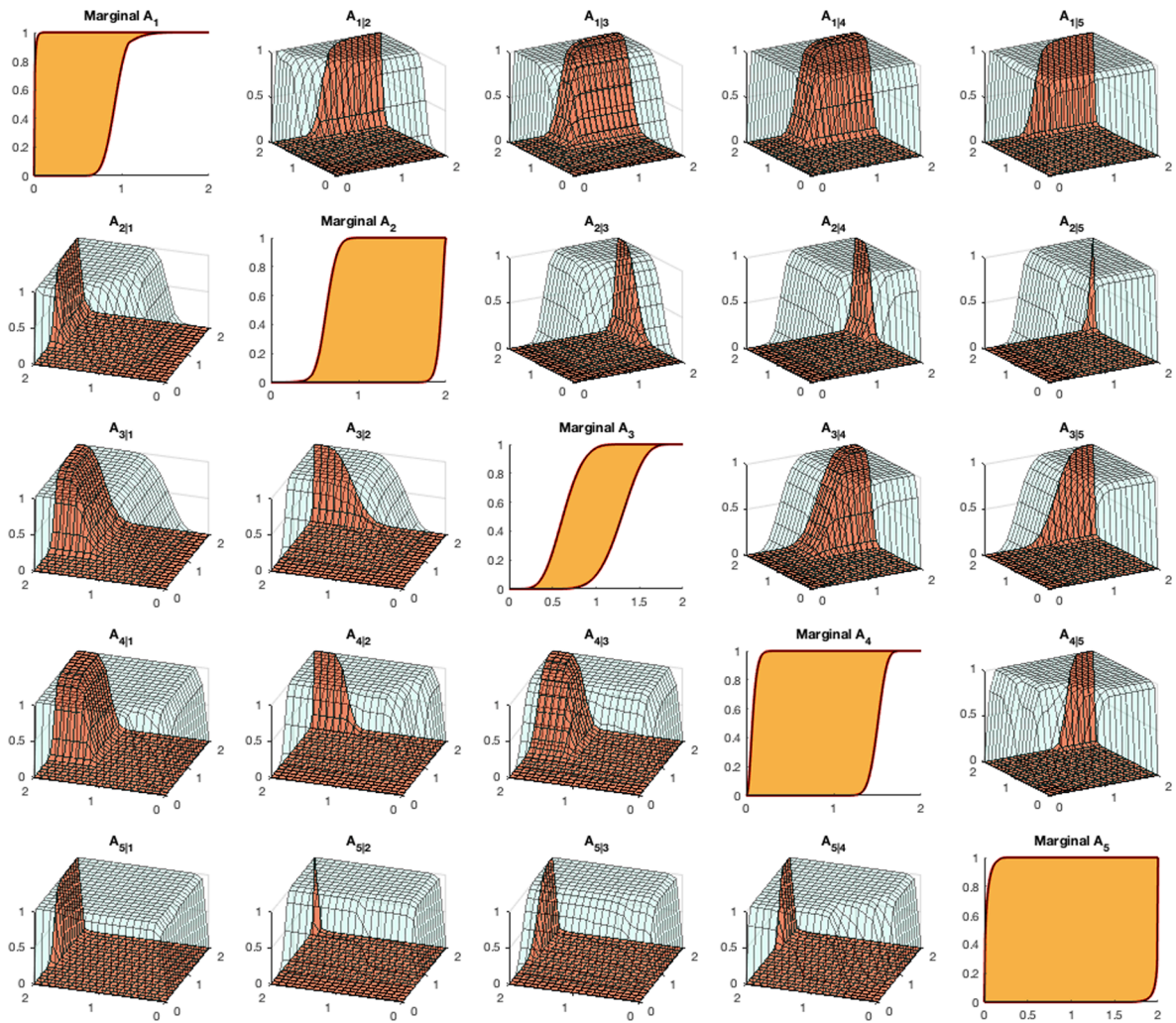


Fig. 11. Representation of the five-dimensional p-box for F_A . The marginal p-boxes are shown on the diagonal panels, with off-diagonal panels showing bivariate p-boxes for parameter pairs. Upper bounds for the bivariate p-boxes are shown in transparent blue, with the lower bound in red.

CDFs. It can be seen that, the marginal p-box for a_5 is very wide and close to an interval, therefore according to UM_y the marginal for a_5 can be nearly any distribution in $[0, 2]$. In contrast, the probability distribution for the variable a_3 is much narrower. This result suggests a higher importance of the variable a_3 when prescribing an uncertainty model for the data y_t .

6.1.2. Sliced-normal approximation of posterior for propagation

Due to the cost of sampling the posterior, because it required expensive MCMC, a means of efficiently generating additional samples is required to reduce sample based approximation errors. We thus use a sliced normal distribution as an approximation of the posterior, based on the 520 TCMC samples. The density of the sliced normal is less expensive to evaluate than the posterior, allowing us to efficiently draw more samples. The sliced normal was fit to the 24 dimensional posterior with two degrees of freedom, producing a feature space expansion into 324 dimensions. Generating 10^4 sliced-normal samples takes roughly 90 min when parallelised using a 4-core personal computer. By comparison, the 520 posterior samples takes approximately two days on a 40-core HPC machine.

The suitability of this approximation is dependent on its ability to appropriately represent the posterior density. For the two feature-space methods (maximum likelihood and worst-case) of calculating the sliced-normal parameters presented in [5], we evaluated the ability of the corresponding sliced-normals to appropriately assign relative likelihoods to the Bayesian posterior samples through leave-one-out validation. This was assessed by calculating a possibilistic transform of the sliced normal, and evaluating the alpha cut required to contain the remaining sample. The cumulative distribution of these alpha levels should dominate a unit uniform distribution, indicating that the sliced normal is correctly ordering the samples based on relative likelihood.

Fig. 12 depicts the coverage properties for possibilistic structures based on the two methods on the left, with a representation of the alpha cut sets generated in a two-dimensional space shown on the right. The coverage comparison was performed over the four epistemic dimensions only due to the computational cost of the worst-case method. It can be seen that both methods have similar

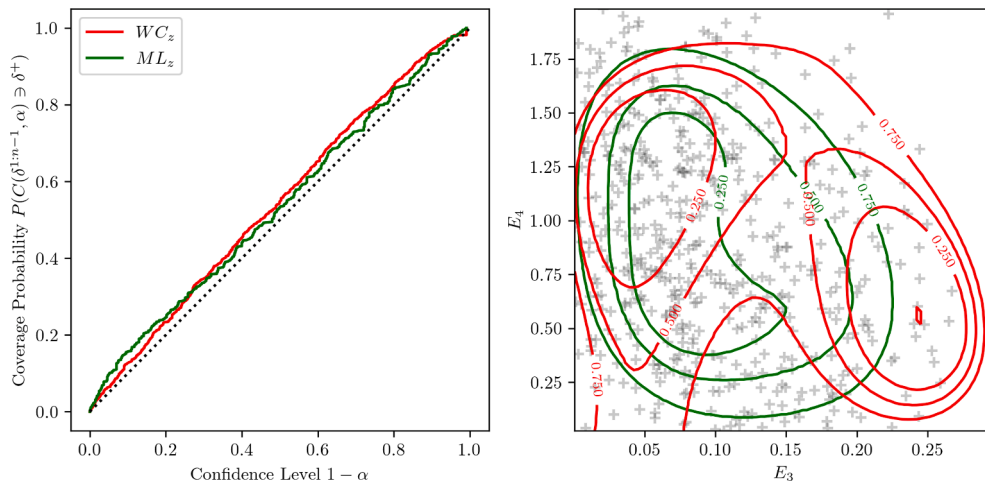


Fig. 12. Left: Probability of alpha cut sets containing a future sample against desired confidence level, with a comparison of two methods of generating sliced-normals, worst-case optimisation, WC_z , and maximum-likelihood optimisation, ML_z . Both measure density in feature space with a confidence of $\alpha_B = 0.99$, and $n_p = 4, d = 2, N_e = 520$. Right: Two-dimensional comparison of the alpha cut sets generated by both methods for the space of E_3 and E_4 in the NASA challenge problem. The alpha level contours for each method are 0.25, 0.5, and 0.75.

predictive capability. The right side of the figure indicates that the confidence level sets for this data are not unique, depending on the approach used to generate the sliced normal. Since the worst-case method is significantly cheaper to calculate we adopted this approach for generating further posterior samples.

6.1.3. Geometry of epistemic space

The sliced-normal distribution can be used to define level sets on the expanded epistemic space. For visualisation purposes these sets are calculated for pairs of epistemic parameters to highlight the ability of sliced normals to capture a variety of dependence

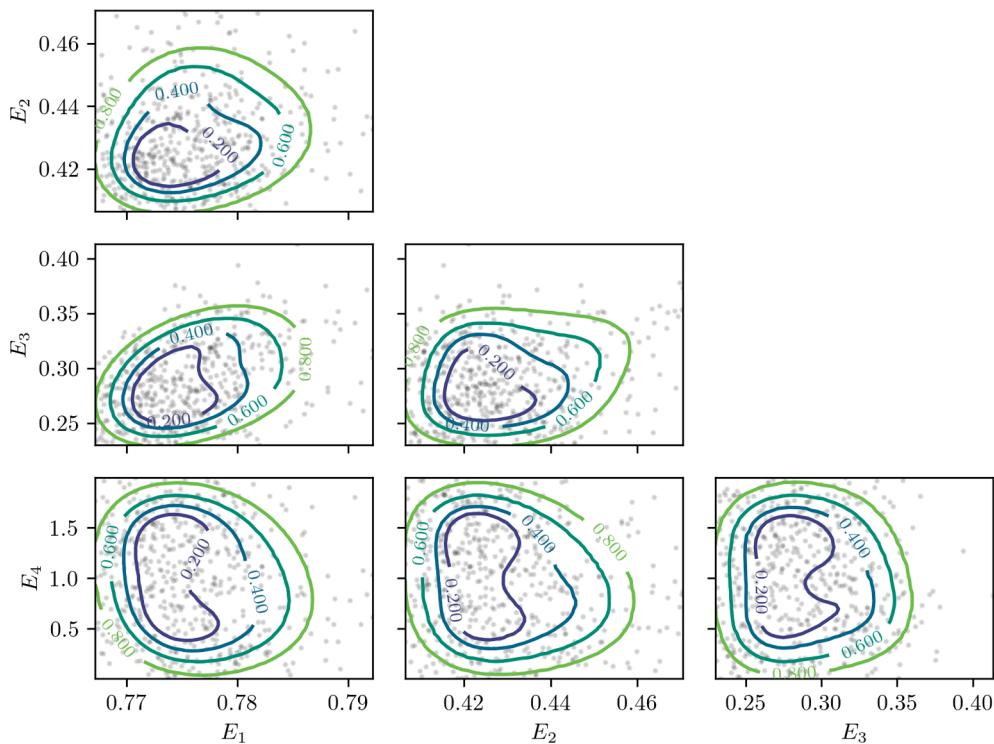


Fig. 13. The pairwise representations of the geometry of the epistemic space in terms of the sliced-normal distribution at 0.2, 0.4, 0.6 and 0.8 confidence levels.

structures. A representation of the resulting geometry of the epistemic space is shown at defined confidence levels with contours in Fig. 13. Similar sets can be produced for the full 24 dimensional space, though the data sparsity and projection of a 24-dimensional set onto two dimensions makes visualisation difficult. Fig. 14.

6.2. Problem B: Uncertainty reduction

In this section we describe the path taken to update the uncertainty model with the aim of improving its predictive capability. The task is divided in two stages.

- The first stage is to rank the epistemic inputs by their importance, see Section 5.6.
- The second stage is to identify a set of changes, to reduce the bounds of the epistemic inputs, such that the expected reduction in the output uncertainty is as large as possible.

The tasks of ranking the epistemic parameters according to their effect on the predictive capability of the computational model, and determining those whose reduction will increase the predictive capability of the model is a challenging one. This cannot be done using naïve implementations of variance-based sensitivity analysis. Therefore, we describe two different approaches to evaluate these rankings. One using a modified variance based sensitivity analysis approach described in Section 5.6.3, the other using a value-of-information approach described in Section 5.6.1. Different aspects of variance-based and value-of-information sensitivity analysis are highlighted which makes them suitable for tackling the problem at hand. Despite the fact that the two methods differ philosophically and computationally, they are shown to achieve similar results.

6.2.1. Ranking via variance-based sensitivity analysis

As mentioned in Section 5.6.3, variance-based SA (VBSA) is an appropriate ranking approach whenever the variance of the model output is the primary measure of its uncertainty.

In this procedure the epistemic parameters are the only ones to be ranked, and they are given uniform distributions in UM_0 . We

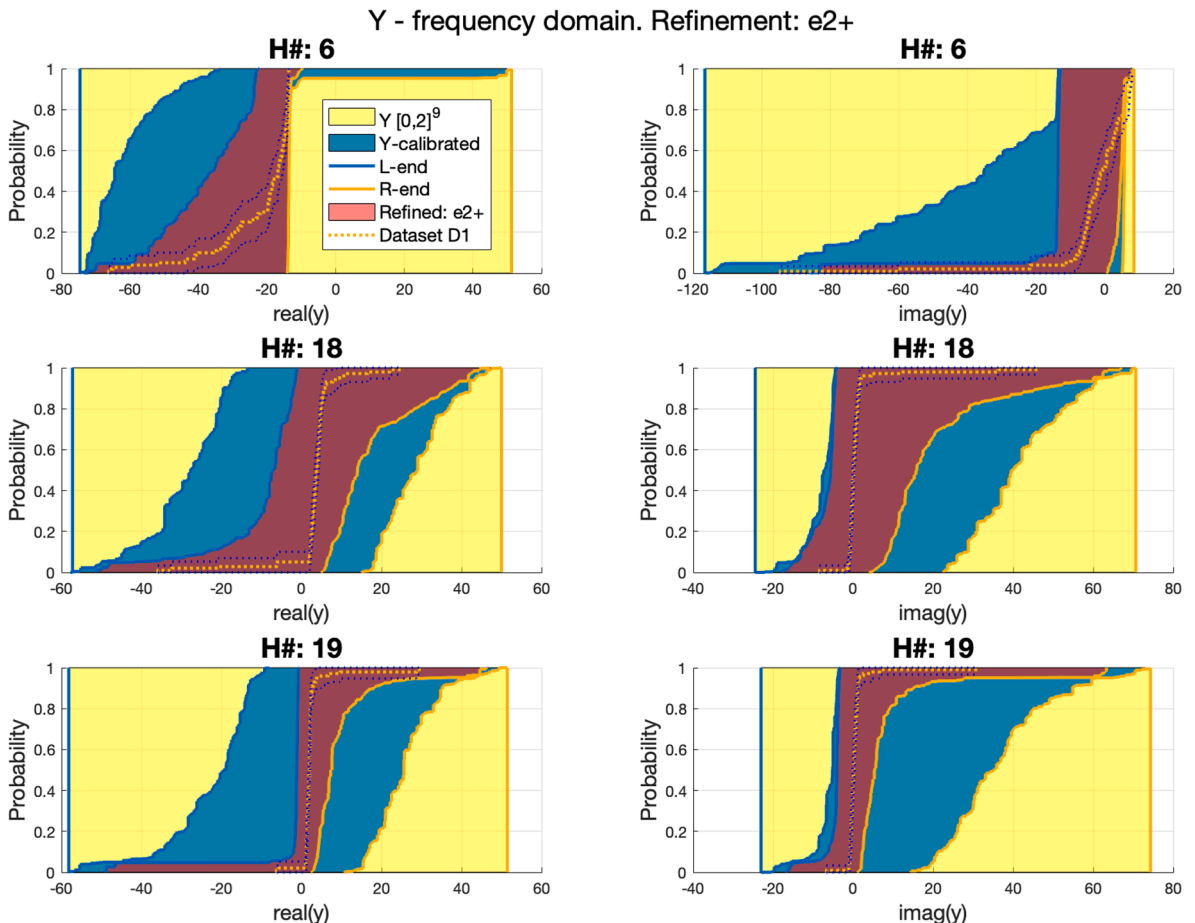


Fig. 14. Refinement e_{2+} for harmonics 6, 18, 19.

interpret the variance of these uniform distributions as a metric for their uncertainty. By re-framing the interpretation in this way we are able to carry out the VBSA on the $n_a + n_e$ -dimensional input space directly. This is in contrast to the full uncertainty model, described in and Section 6.1, where a more expressive, p-box, model of the uncertainty is used.

We carry out the VBSA analysis in both the original and calibrated input spaces and compare the results below. For the prior space analysis, $n = 5000$ Latin hyper-cube (LH) points were sampled from the marginal distributions of each of the $n_a + n_e = 9$ inputs. First-order and total sensitivity indices were obtained at each time step and the results were averaged across the entire time window. It is important to note that here, as in Section 6.1.1, a simplifying assumptions is made about the time dimension of the model in that separate time steps are considered independent (see [120], for a discussion on time-dependent approaches to VBSA). The results are summarised on the left of Table 3.

We used a similar procedure for the sensitivity analysis with respect to the calibrated uncertainty model, UM_y . In order to obtain input combinations in the constrained domain, the TMCMC samples from the updating procedure were used to compute the bounds of each parameter. A new set of $n = 5000$ LH samples was obtained in these bounds and the first order, and total indices were recomputed. The results are shown on the right in Table 3. For both prior and posterior analyses the aleatory variables were left free to vary within their ranges, so as to allow for any interactions between the aleatory and epistemic sets of variables to manifest themselves in the total sensitivity indices.

Under both UM_0 and UM_y the second epistemic input, e_2 was identified as the most important, followed by e_1 , e_3 and e_4 in order. From the values of the total indices in Table 3, it becomes clear that the model output is sensitive to interactions between e_2 and one or more other inputs. In the present, variance-centric setting, these results suggest that e_2 is the epistemic parameter with the highest potential to improve the predictive capability of the model.

6.2.2. Ranking via value-of-information sensitivity analysis

The ranking of the epistemic parameters is done by scoring each of the 8 possible refinements $\{e_i^-, e_i^+\}$, $i = 1 : 4$, where e_i^- denotes a request for the lower bound of the e_i to be increased, whilst e_i^+ denotes a request for the upper bound of e_i to be decreased. For UM_0 : $e_i^- = [0, 1]$ and $e_i^+ = [1, 2]$. The ranking is done by comparing the amount of epistemic uncertainty in the output y_t before and after the refinement, and simultaneously checking that the data D_1 still falls within the bounds. The comparison is enabled by the area metric of Eq. (14), which quantifies the amount of *area* contained between two distributions or datasets.

For the purpose of the sensitivity analysis, the calibrated uncertainty model of Problem A, consisting of correlated second-order probability distributions is converted to its equivalent bounded counterpart. So the 24-dimensional epistemic space is *boxed* in a hyper-rectangle containing all the samples of the updated posterior. This has allowed us to treat the calibrated inputs as probability boxes for the sake of the uncertainty propagation. When the area metric, Eq. (14), is applied to the two bounding distributions of a p-box, the resulting area metric quantifies the amount of epistemic uncertainty carried in the p-box. After the refinement, the area of the p-box shrinks (or does not change) to reflect the fact that less epistemic uncertainty is projected to the output following the refinement. Denoting $[F_Y]_0$ and $[F_Y]_{e_i^*}$ the p-box of the output before and after the refinement respectively, their corresponding area metrics W_0 and $W_{e_i^*}$ are used to compute the sensitivity scores as:

$$S_{e_i \rightarrow y}^{Vol} = 1 - \frac{W_{e_i^-} + W_{e_i^+}}{2 W_0}. \tag{34}$$

In order for the computation of Eq. (7) to succeed, the p-boxes $[F_Y]_{e_i}$ need to be fully nested in $[F_Y]_0$, i.e. $[F_Y]_{e_i} \subseteq [F_Y]_0$. When this does not hold, negative sensitivity scores can show up. When the uncertainty propagation is not rigorous it is not obvious how to avoid this problem. A meta-model strategy has been developed within this work to ensure that the nesting always occurs, as described in Section 5.6.1 (see Table 4).

Since the model of the subsystem has a time-varying output, VoIs S_{e_i} were computed in the reduced Fourier space looking at the first 30 harmonics. A summary of the ranking is provided in Table 5. Similar to the analysis of the VBSA, in Table 5, UM_0 and UM_y are respectively the UMs before and after the calibration.

Both VoI and variance-based methods largely agree on the top-ranked parameters e_2 carrying the highest sensitivity. A slight difference between the two methods holds for the ranking of parameters e_3 and e_1 , but are still ranked above e_4 .

Table 3

Main and total effect indices for the epistemic inputs before, (UM_0), and after, (UM_y), calibration. Inputs ranked by their importance in descending order.

	UM_0		UM_y	
	Main	Total	Main	Total
e_2	0.0497	0.3535	0.0264	0.1706
e_1	0.0204	0.1401	0.0092	0.0179
e_3	0.0127	0.0455	0.0063	0.0131
e_4	0.0076	0.0011	0.0045	0.0002

Table 4
Summary of the change in Sobol' indices when the ranges of $e_{1,\dots,4}$ are reduced.

	e_1	e_2	e_3	e_4
			<i>First order indices</i>	
e_i^+	0.0048	0.0095	0.0006	0.0000
e_i^-	0.0040	0.0103	0.0006	0.0000
			<i>Total indices</i>	
e_i^+	0.0340	0.1375	0.0124	0.0000
e_i^-	0.0333	0.1455	0.0128	0.0000

Table 5
Vol sensitivity scores for each refinement.

$S_{e_i^-}$	e_1	e_2	e_3	e_4
e_i^+	0.0618	0.1410	0.1381	0.0394
e_i^-	0.0828	0.1348	0.0294	0.0130

6.2.3. Parameter refinement

The purpose of ranking the epistemic parameters by their influence on the predictive capability of the model. As part of the NASA challenge, the hosts offered up to 4 refinements to the bounds of the epistemic parameters. As outlined in Section 2, there are 2 possible updates for the i^{th} epistemic parameter. Namely, increasing its lower bound (denoted e_i^-) and/or decreasing its upper bound (denoted e_i^+), thus we had 8 possible options to choose from. Having a smaller budget than the possible selection of actions is common in real-world applications, where the analyst is likely to have constraints on the computational or experimental budget. Here, based on the two ranking approaches discussed earlier we show how each of them can be used to inform the choice of refinements.

6.2.4. Parameter refinement via variance-based analysis

In order to choose the parameters and bounds to refine, we conducted two sets of sensitivity analysis. In the first set, the epistemic parameters had their upper bounds lowered one at a time, and in the second set, each lower bound was raised, again, one parameter at a time. The main and total effects of these analyses were then compared to their $UM-y$ counterparts and the parameter-bound combination with the largest absolute difference in the indices were selected for refinement. The aleatory variables were left unchanged throughout the analyses, to enable effects due to variability in A to propagate to the final results.

The motivation behind the above approach hinges on the fact that, in variance-based sensitivity analysis, a variable can have a Sobol' index equal to 0 in one of two ways: either the variable is not included in the model (the so-called dummy variable), or it has no variation, i.e. it is a fixed parameter. Thus, if the interval for each epistemic parameter was reduced to a number, its Sobol' index would become 0. For this reason, one would expect to observe a noticeable change in the values of sensitivity indices of important parameter-bound combinations.

For each analysis the respective bound was moved to the median value of the parameter's range. This particular point was chosen as

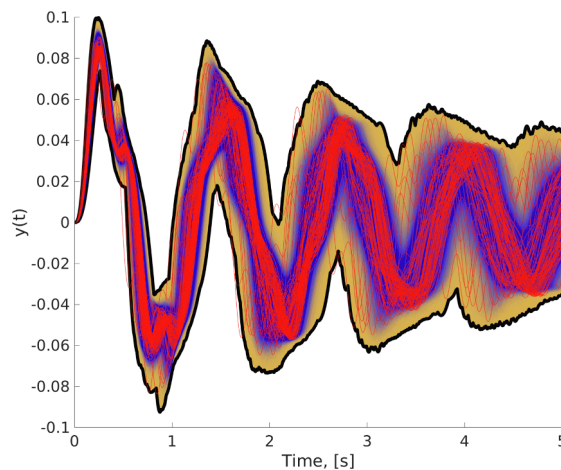


Fig. 15. Shows the time response of $y(t)$ for the calibrated model after the second calibration post parameter refinement. Shown in blue are samples from the Bayesian updating, with the bounds from the sliced-normal in yellow, and the experimental data in red.

no information was provided on how much the bounds would be shrunk by, if requested. To escape the median assumption, one could carry out a sensitivity study on the bounds themselves, at an increased computational and post-processing expense. Such an analysis will yield highly correlated variates which need to be accounted for in the interpretation of results.

We compared the results of the sensitivity study against a straightforward variance reduction analysis, in which the average output variance across the time domain was calculated with- and without- bound reduction. The analyses showed a clear preference for e_2 to be refined, with a subtle, but consistent advantage of e_2^- over e_2^+ . A summary of the results is shown in Table 5.

6.2.5. Parameter refinement through value-of-information

The parameter refinements are scored as in Eq. (34) by computing the relative amount of *reduced* uncertainty that each refinement carries compared to the total. Eq. (34), where $S_{e_i^-} + S_{e_i^+} = S_{e_i}$, weighs the importance of each partial score S_{e_i} towards the whole score S_{e_i} . The results are summarised in Table 5.

Out of the 8 possible refinements and with a maximum of $k \leq 4$, we have requested one reduction for this problem. The VoI and conditional variance-based analyses for parameter refinement largely agree in the important refinements to be made. Following the requested parameter refinement, an additional iteration of the updating procedure was conducted. The new outputs of the post-refinement calibrated model, UM_{y-2} , is shown for a number of output time points in Fig. 16 and for the entire process in Fig. 15. The improvement, i.e reduction in the epistemic uncertainty of the output, is obvious.

6.3. Problem C: Reliability analysis

In this section we seek to evaluate the reliability of a baseline design θ_{base} according to the current UM. This task can be divided in the following sub-tasks:

- evaluate the range of the failure probability for joint requirements and each individual requirement, Eqs. (26) (27);
- rank the epistemic uncertainties according to the contraction of $R_w(\theta)$ that might result from their reduction;
- identify representative points (a, e) with a comparatively large likelihood near the failure domain, using (a, e) to characterize qualitatively different transitions to failure, and the corresponding time responses of the integrated system; and,
- evaluate the severity of each individual requirement violation and all violations.

The challenge of the reliability assessment is to balance the computational cost with the precision of the reliability estimates. The severity estimators are very sensitive to a lack of samples. It is also really difficult to obtain precise lower failure probability bounds due to the epistemic uncertainty at the tails of our calibrated model. We question the interpretation of small failure probabilities ($< 10^{-2}$) yielded by advanced Monte-Carlo methods such as importance sampling. Given the paucity of data $n_1 = 100$, and the explicit ac-

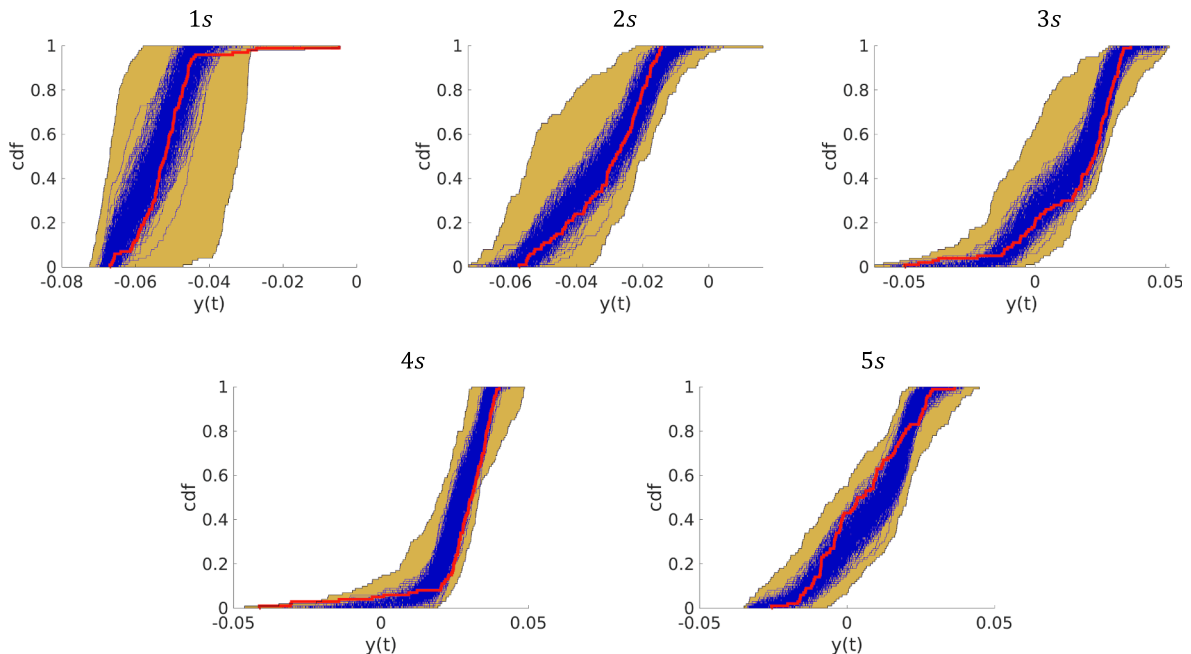


Fig. 16. Calibrated p-boxes at various times in the process are shown. The second order samples from the Bayesian calibration are in blue, with the bounds from the sliced-normal in yellow, and the experimental data in red. It can be seen that the extra samples from the sliced-normal leads to a more converged p-box bound, and a better fitting to data.

counting for the epistemic uncertainties, it is difficult to confidently obtain such small failure probabilities.

6.3.1. Range of failure probabilities for individual and all requirements

The range of failure probability $R_w(\theta)$, for each individual requirement $g_{1:3}$ for the baseline design, θ_{base} , are estimated via double loop MC sampling as described in Section 5.4. Table 6 present (first row) the results of the above analysis with respect to $R_w(\theta)$, $R_j(\theta)$, $s_w(\theta)$ and $s_j(\theta)$. The reliability performance of θ_{base} is estimated using $N_e = 10^4$ epistemic samples from the sliced normal of UM_{y-2} and $N_a = 200$ aleatory samples from the corresponding realizations of the joint probability distributions F_A . The most uncertain reliability scores for θ_{base} are in respect of the, settling time [0, 0.99], stability [0, 0.67], and energy consumption [0, 0.27]. Note that the failure probability bounds for the joint requirement $R_w(\theta)$ is poorly informative due to the large epistemic uncertainty in UM_{y-2} .

6.3.2. Rank of epistemic uncertainty on R_w

In this section we are asked to rank the epistemic uncertainties according to the contraction of $R_w(\theta)$ that might result from their reduction. To do so we use the VoI metric, Eq. (13), which measures the reduction in area of the p-box of $g_{1:3}$ for a fixed value of $e_{1:4}$, relative to the area obtained with all parameters varying in their posterior range. Additionally, any reduction in the range of R_w was also taken into account. The performance requirements $g_{1:3}$ showed high sensitivity towards the epistemic uncertainty in the UM. Only a modest contraction of $R_{1:3}$ was found from the reduction of epistemic inputs, with parameters e_1 and e_2 ranking highest.

We note that variance-based sensitivity analysis can also be used to assess the effect of $e_{1:4}$ on R_w . However, such an analysis would require the problem to be slightly reformatted to enable VBSA to work with the interval R_w . We chose not to pursue this strategy further, postponing more in-depth investigations for future work.

6.3.3. Identify realizations of (a, e) with large likelihood near the failure domain

To find the likely realisation near the failure domain we:

- isolated a focal element for each failing requirement $g_{1:3} = 0$, so three in total, and identified the realizations in the input domain falling inside these three focal elements.
- looked at the p-box lower bound of each requirement, and isolated the samples near the failure domain.
- then identified three examples of time responses for each type of failure.
- A parallel plot analysis was conducted to highlight components which have a large effect on failure probability.

The parameters defining the marginal distributions of a_1 and a_2 were identified as major contributors to high failure probabilities for the three failure conditions. In contrast, none of the original epistemic parameters was found to be influential for resulting in high probability of failure.

6.3.4. Severity of requirement violation

We estimate the severity of each individual requirement and joint requirements via the double loop Monte-Carlo propagation procedure presented in Section 5.4. The accuracy of this estimator is more sensitive to the number of aleatory samples when compared to the epistemic bounds of the failure probability, i.e. an outlier can alter substantially the value of the severity. The requirements ranking for the most severe failures are reversed in order, compared to the reliability ones, see Table 6. Here, the energy consumption potentially leads to the most severe failures $s_3(\theta_{base}) = 0.7154$.

6.4. Problem D: Reliability-based design

6.4.1. Optimality criteria and computational approach

For the reasons outlined in Section 5.7, a new candidate design was selected according to the criteria of minimizing failure probability and severity of the integrated system. We selected genetic algorithms as a global gradient-free optimisation strategy to search for a candidate design θ_{new} . The set of possible candidate designs Θ was not provided in the challenge. As such, we elected to search within the space defined by the hyper-cube $\theta_{base}(1 \pm \alpha)$, with $\alpha > 0$ centred on the baseline design.

Table 6

Summary of the RBDO results for UM_{y-2} and UM_z estimated using $N_e = 10^4$ and $N_a = 200$.

Design	$R_w(\theta)$	$R_1(\theta)$	$R_2(\theta)$	$R_3(\theta)$	$s_w(\theta)$	$s_1(\theta)$	$s_2(\theta)$	$s_3(\theta)$
$\theta_{base}(UM_{y-2})$	[0,0.99]	[0,0.67]	[0,0.99]	[0, 0.269]	0.7393	0.1197	0.0171	0.7154
$\theta_{base}(UM_z)$	[0,0.7]	[0,0.7]	[0,0.51]	[0,0.19]	0.4511	0.1522	0.0106	0.3649
$\theta_{new}^*(UM_{y-2})$	[0,0.980]	[0,0.359]	[0,0.976]	[0,0.0348]	0.0572	0.0527	0.0070	0.0170
$\theta_{new}^*(UM_z)$	[0,0.12]	[0,0]	[0,0.12]	[0,0.045]	0.0126	0	3.69×10^{-4}	0.0125
$\theta_{final}^*(UM_{y-2})$	[0,0.985]	[0,0]	[0,0.985]	[0,0]	0.0039	0	0.0039	0
$\theta_{final}^*(UM_z)$	[0,0.075]	[0,0]	[0,0.075]	[0,0]	1.27×10^{-4}	0	1.27×10^{-4}	0

6.4.2. Analysis for θ_{new}

The first four rows of Table 6 present the results of the optimisation. The reliability metrics are obtained as described for the baseline θ_{base} and updated θ_{new} designs. A substantial improvement of the severity can be observed, along with a modest reduction of upper bound on the system failure probability. It is also interesting to note that the baseline design represents a case that highlights the importance of the severity metric. The transition to failure is marked by an extremely sharp increase in severity at the boundary. Subsequent designs remove this sharp change, allowing for a less extreme transition to failure. Fig. 17 presents the resulting probability boxes for the baseline design and θ_{new} in blue and red, respectively.

Fig. 18, shows the composite failure regions and safe regions projected on to a 2-d space of uncertain factors. The figure shows three designs in the projected space. These figures are indicative of the relatively improved performance of the designs, shown by the removal of failure limit-states in the aleatory space for each θ . The transparent markers are aleatory samples from multiple probabilistic models F_A defined by epistemic vectors $e_{5:24}$.

6.5. Problem E: Model update & design tuning

Upon finding a new design point θ_{new} , the last problem of the challenge is to perform a final improvement to the UM and design. After providing θ_{new} to the challenge hosts, 100 realisations of the subsystems Z_1 and Z_2 were provided for calibration. The calibration approach pursued here is identical to that described in Section 6.1.1, where the new data was used alongside the original data to improve the UM. A sensitivity analysis, identical to that in Section 6.2.2 was done to request 3 more parameter refinements to E . These refinements were selected, with the analysis performed with respect the subsystems Z_1, Z_2 and Y , and we ranked parameters again as in Table 7. Once the challenge hosts returned with 3 reductions to E , the UM was calibrated a final time over this reduced space with the new data. Figs. 19–21 show the results of this calibration, along with bounds from a sliced-normal distribution.

The updated model was then used within the RBDO described in Section 6.4 to obtain θ_{final} . Table 6 presents the reliability scores for θ_{final} and Fig. 18 shows a 2-d projection of the individual failure regions. Similarly to θ_{new} , the most uncertain reliability score for θ_{final} is with respect to the settling time requirement. However, the epistemic uncertainty affecting this score decreased substantially with the new model of the uncertainty, i.e., from [0,0.985] to only [0,0.075]. Fig. 22 shows bounds of the uncertainty model on Z_1 and Z_2 over the three designs and calibrated uncertainty models.

6.6. Problem F: Risk-based design

- F-1) Propose a metric to quantify the gain, l , resulting from taking the risk $r = \hat{r}$.
- F-2) Find a design point that maximizes $l(\hat{r})$ and denote it as θ_r^- . Explain the process used to choose the portion of E being ignored.
- F-3) Compare θ_{final} and θ_r^- using the figures of merit above.
- F-4) Evaluate $l(r, \theta_{final})$ and $l(r, \theta_r^-)$ for $r = [0, 0.5, 1, \dots, 10]$. determine the optimal trade-off between performance gain $l(r)$ and risk for each design point (if any).
- F-5) Use the results of this sub-problem to detail a rationale that might be used to justify either adopting or rejecting a risk-based design

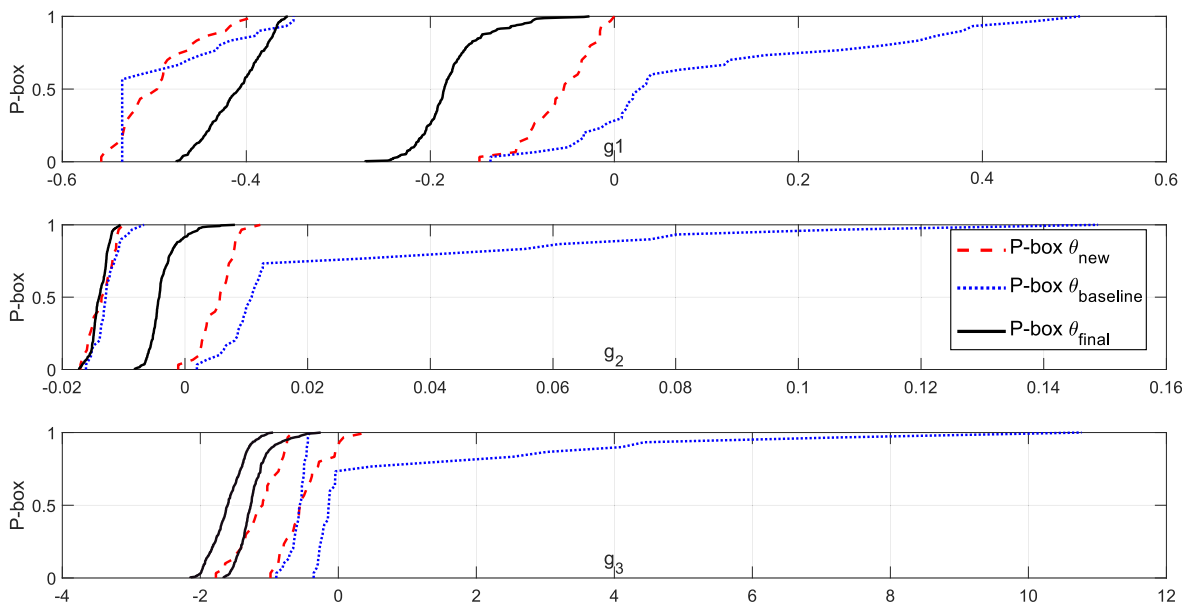


Fig. 17. A comparison of the p-boxes for $g_{1:3}$ (top to bottom panel) for the baseline design (blue), the design θ_{new} (red) and θ_{final} (black solid line).

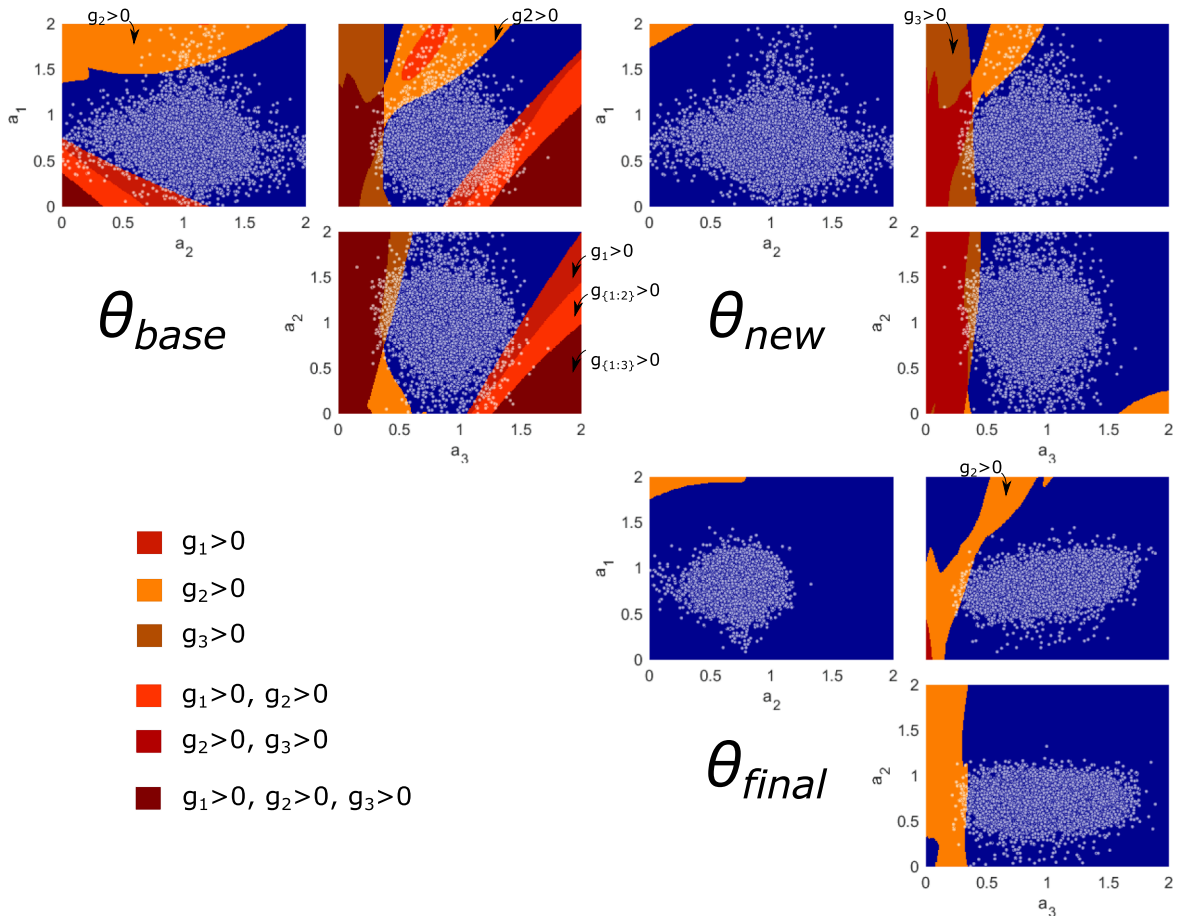


Fig. 18. Projections of the individual safe regions (blue) and failure regions (otherwise) induced by θ_{base} (3 panels on the top-left), θ_{new} (3 panels on the top-right), and θ_{final} . White markers show realization of a_1, a_2 and a_3 for the proposed models UM_y (top panels) and UM_x (bottom panel).

Table 7
Summary of VoI sensitivity ranking for UM-z.

		e_1	e_2	e_3	e_4
Y, Z_1, Z_2	e^-	0.406	0.306	0.208	0.137
	e^+	0.679	0.169	0.429	0.230

This task is highly related to the definition of risk for non-probabilistic spaces. The NASA challenge problem gives a way to define risk as a reduction of the volume [6]. However, this way of defining risk does not uniquely identify possible subsets from which the volume can be computed, because there is an infinite number of subsets of E that have a given volume. So a procedure to establish a way to generate subsets of the updated epistemic space is needed. For example, we can make a 5% reduction by removing volume from the boundary of E while enforcing some compactness condition. Alternatively, in some applications, the best design improvements may be obtained by removing volume from the geometric center, e.g., discarding failure domains that are internal to the space defining E .

Thus, one of the main challenges in problem F is the definition of an appropriate epistemic volume removal scheme, or in other words, a good justification about the way we are accepting a risk r . The way we remove epistemic volume inevitably affects the resulting risk-based design. If a probabilistic characterization of the epistemic volume E , is adopted (e.g. using a posterior distribution) a risk-based analyst would probably feel tempted to reduce epistemic volume by removing the regions with the lowest probability mass. In contrast, a non-probabilistic characterization of the epistemic volume does not associate a probability to the different regions of $E_0 \times E_B$.

This appears to be a more rational choice, however, it is not clear how to reduce volume in this case, e.g. from the boundaries or the geometric center. Both procedures entail highly subjective choices.

We consider two different schemes for the removal of volume from the updated epistemic space for parameters $e_{1:4}$. For the first risk-taking scheme, we arbitrarily remove volume from the boundary of the convex hull and reduction is calculated from the ratio

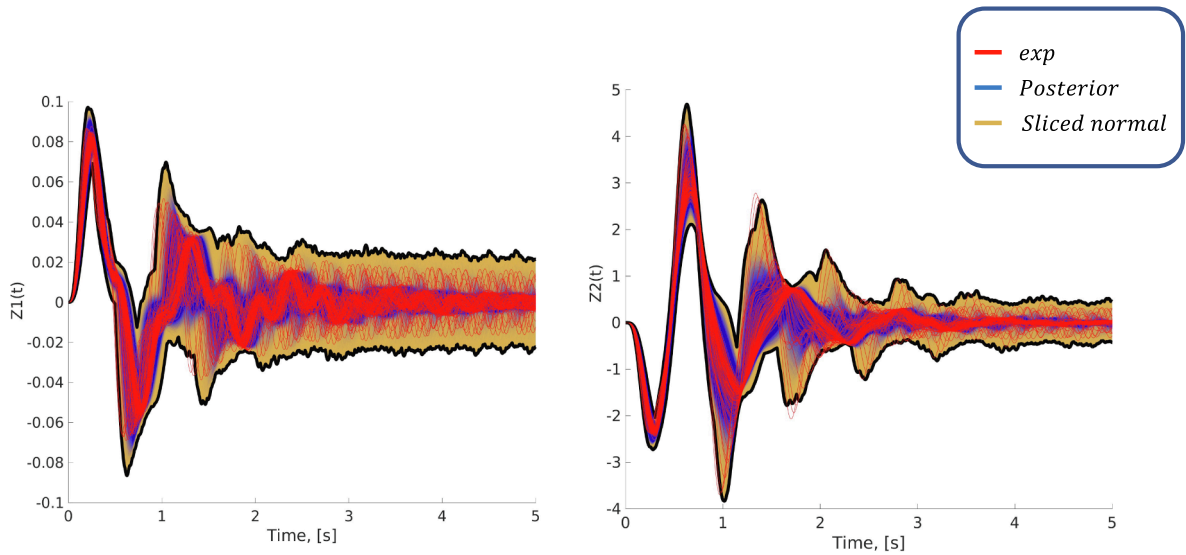


Fig. 19. Shows the system time response $Z_1(t)$ and $Z_2(t)$, for design θ_{new} . The experimental data is shown in red, the posterior simulated samples in blue, and the orange region is the envelope of the sliced-normals simulated samples.

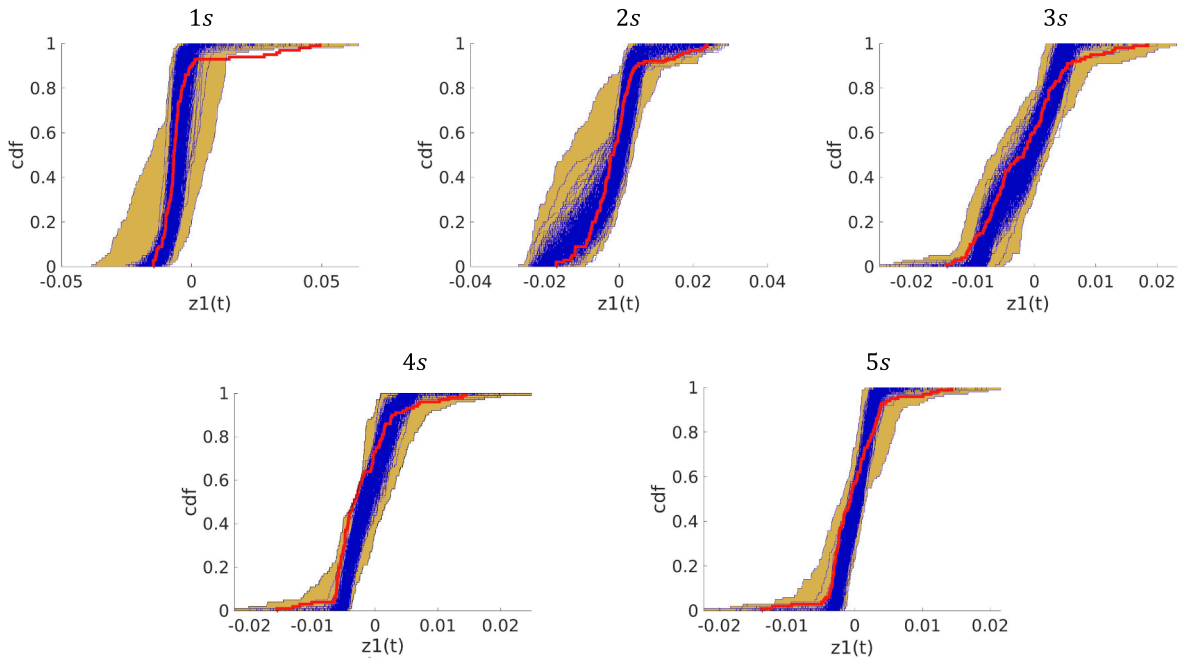


Fig. 20. Shows the cumulative distribution of the response, Z_1 , for the newly updated model and θ_{new} at time points $t = [1, 2, 3, 4, 5]$ s. The experimental data is shown in red, the posterior simulated samples in blue, and the orange region is the envelope of the sliced-normals simulated samples.

between the two volumes. This is equivalent to a removal of epistemic regions corresponding to the support of the tails of the posterior distributions of e . This operation is iterative and is referred to as *convex hull peeling* [121,122]. For computational reasons, instead of computing the convex hull in the full 24-dimensional space, we compute its pairwise projections in (e_i, e_j) . The Cartesian product of these convex hulls conservatively includes the 24-dimensional convex hull. Any samples that define this convex hull are then removed and the process is repeated.

Fig. 23 displays the procedure and the resulting nested compact epistemic volumes. The black marked lines represent the volume for a risk $r = 0$, i.e., the convex hull containing all 10^4 sliced-normal samples. The red solid lines are nested volumes obtained by an iterative removal of the samples on the boundary of the hull. For the second removal scheme we consider a selective removal of a

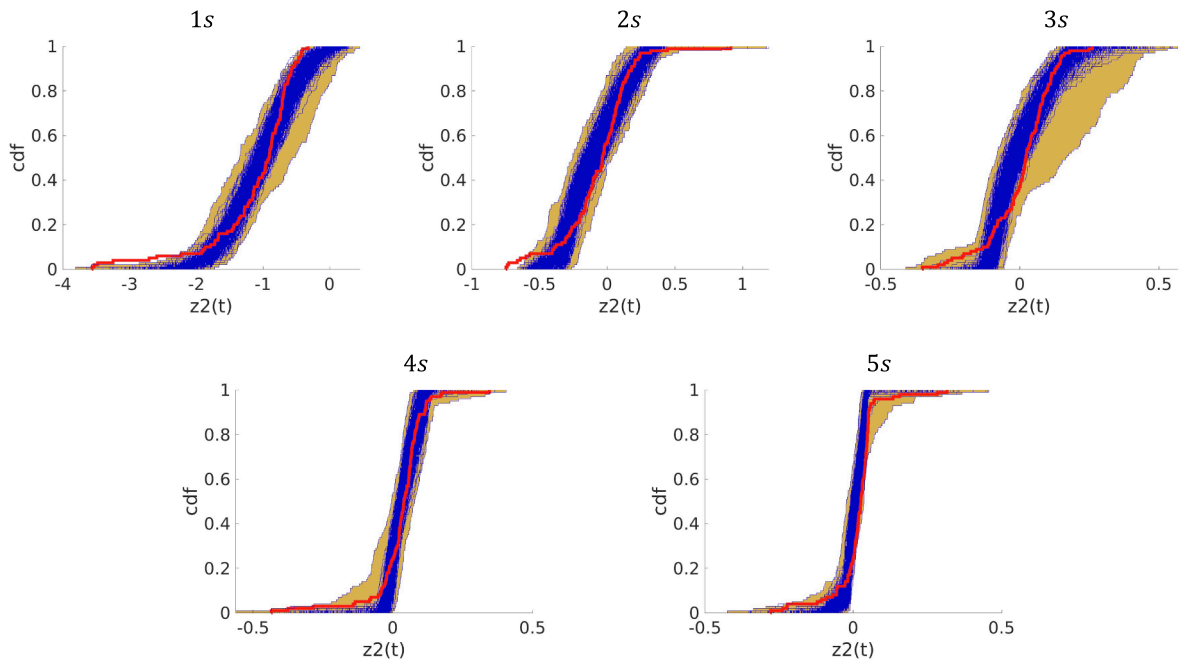


Fig. 21. Shows the cumulative distribution of the response, Z_2 , for the newly updated model and θ_{new} at time points $t = [1, 2, 3, 4, 5]$ s. The experimental data is shown in red, the posterior simulated samples in blue, and the orange region is the envelope of the sliced-normals simulated samples.

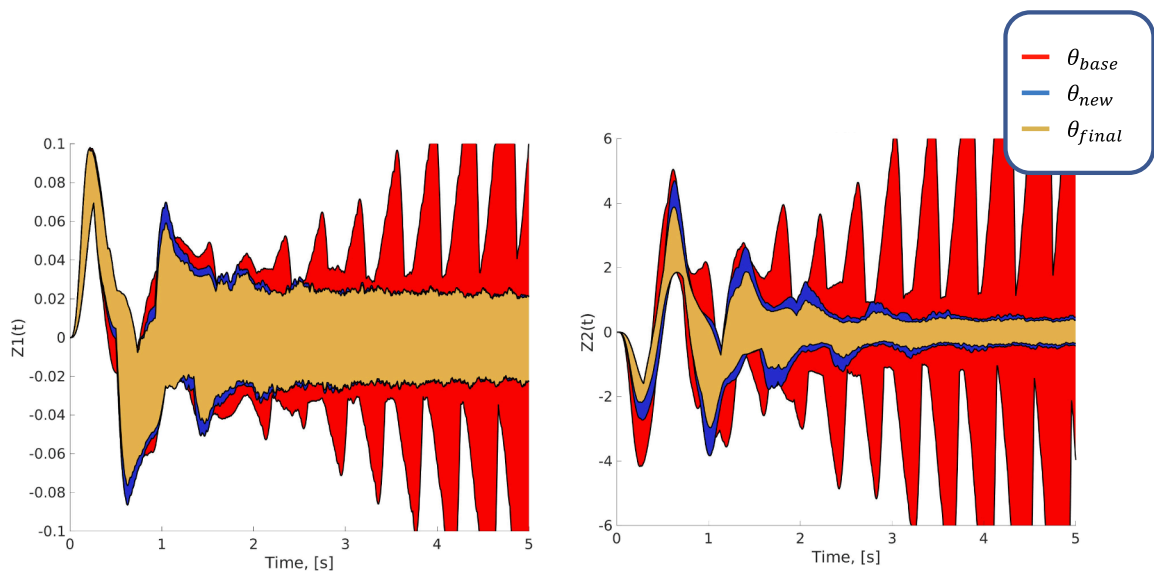


Fig. 22. Shows a comparison of the envelopes of the system time responses $Z_1(t)$ and $Z_2(t)$, for designs θ_{base} , θ_{new} , and θ_{final} .

volume having the highest failure probability and severity.

We considered seven candidates θ_{risk} and compared their reliability performance to θ_{final} and θ_{new} .

The seven risk-based candidates have been selected to be optimal in specific sub-regions of E , indicating that they are potentially a suitable design choice depending on the approach taken for volume reduction.

Fig. 24, shows the reliability results of the different designs by removing volume as displayed in **Fig. 23.** The metrics of gain, $l_s(r)$ and $l_p(r)$, of the nine designs are presented for an increasing reduction in the volume of the resulting set within the original (calibrated) four-dimensional epistemic space E_0 . Note that θ_{final} results in a gain that is the highest for most of the volume reduction

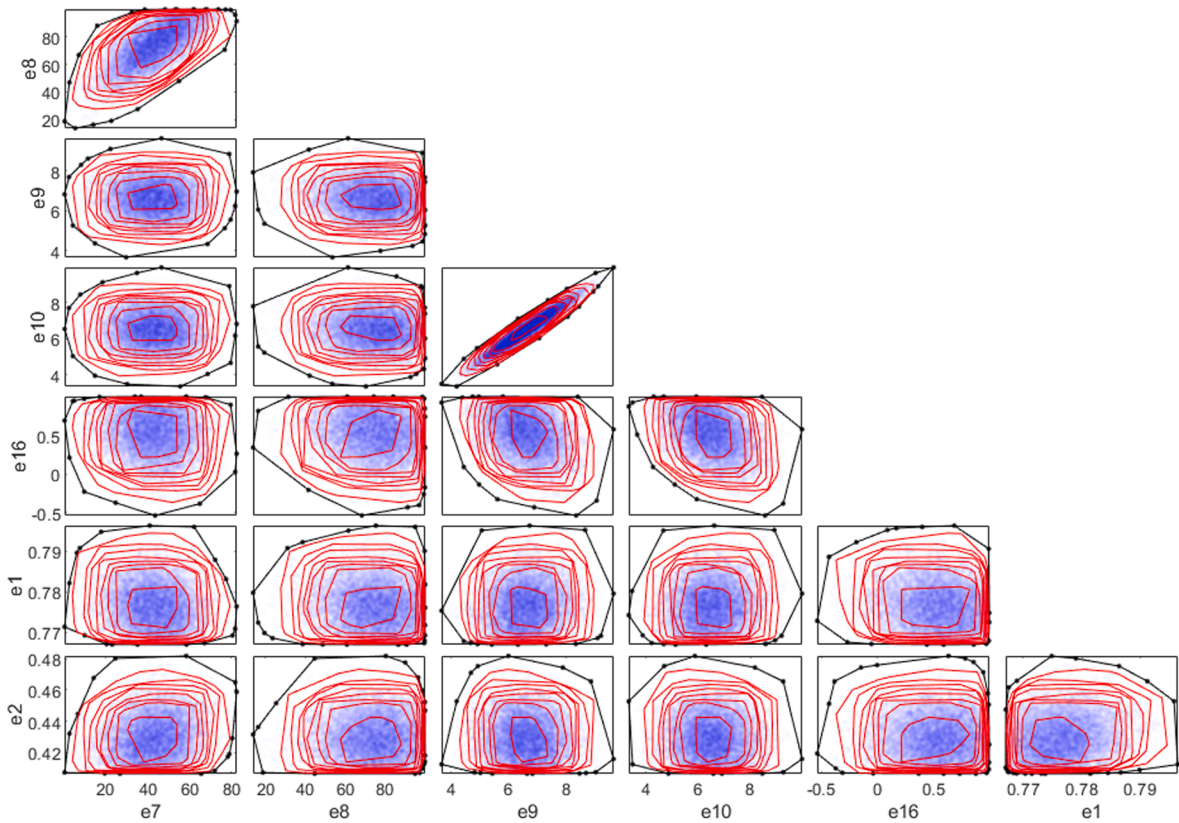


Fig. 23. Pairwise plot showing an illustrative subset of the expanded epistemic space, selected to graphically highlight the dependence, or lack thereof, between a number of the epistemic parameters. The incremental volume reduction resulting from the convex hull peeling procedure is shown across several parameter pairs. The black marked lines show the convex hull projection of the sliced-normal samples (the blue markers). The red lines show convex hulls obtained for increasing risk levels r and defining nested epistemic volumes.

Table 8

Ranking of the θ_{base} , θ_{new} , θ_{final} and seven risk based designs according to their reliability and severity dominance over percentage of epistemic volume. The percentage of volumes are approximated as the retro of sliced-normal samples where a design’s reliability result superior to the reliability of the others.

θ	final	new	baseline	risk-1	risk-2
Best P_f vol. [%]	35.01	5.67	0	29.35	3.92
Best s vol. [%]	33.55	3.96	0	28.88	3.63
θ	risk-3	risk-4	risk-5	risk-6	risk-7
Best P_f vol. [%]	4.34	10.75	1.04	1.36	8.56
Best s vol. [%]	5.10	10.88	1.68	3.86	8.46

procedure. Only after removing approximately 70 % of the volume, do some of the risk-based designs result in a better severity, see e.g. θ_{risk-7} . The numerical results for θ_{final} and two of the seven risk-based designs θ_{risk} (with increasing risk levels) are presented in Table 9. The results show that the final design solution for Problem E is very robust to the considered uncertainties, and only when the epistemic uncertainty becomes very small, can we identify designs that have a better performance.

For the second approach, we discard a volume from E where the failure probability and severity of the candidate designs result in the highest score. To this aim, we associate labels to the 10^4 sliced-normal samples that correspond to the index of the design having the best performance. Each sample of e is labeled based on the best-performing design for of this sample, i.e., the index of the design having the lowest probability of failure, or severity, with respect to e . This yields nine classes, seven for the risk-based design, one for θ_{new} , and one for θ_{final} . The ratio between the number of samples within a class and the total number of samples, gives an approximation of the volume where each of the 9 candidates outperforms the others. Table 8 shows a summary of these approximated volumes.

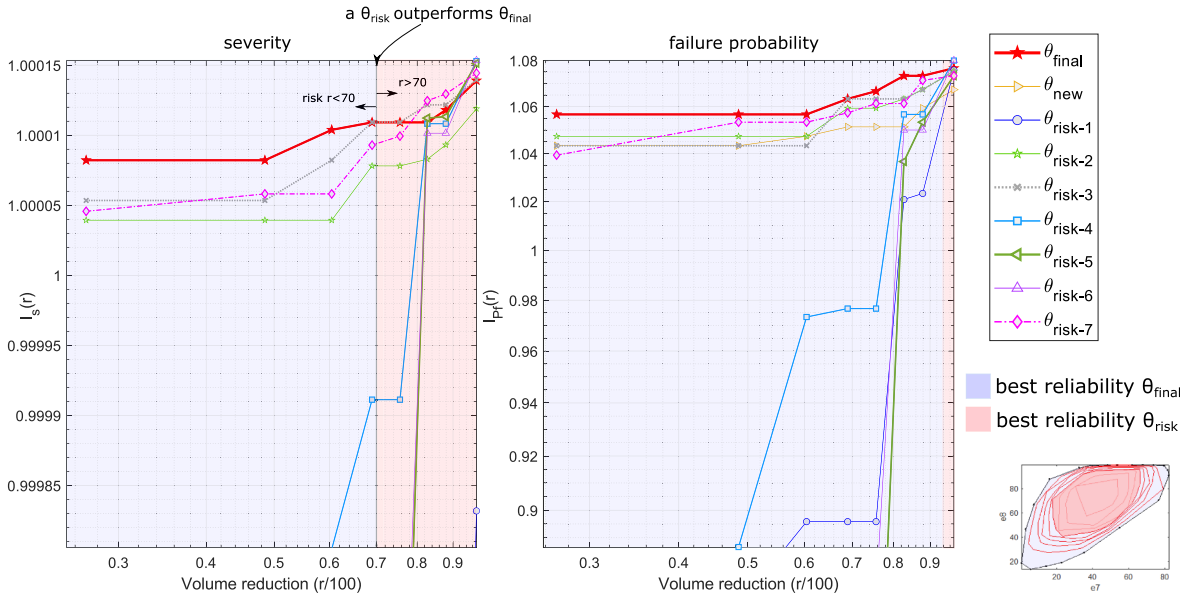


Fig. 24. For increasing levels of risk r (volume reduction), we show a comparison between the gain in severity $I_s(r, \theta)$ (left) and gain in reliability $I_p(r, \theta)$ (right) for the different designs listed in the legend. Annotated as a change in shading is the threshold of volume for which one design outperforms θ_{final} .

Table 9

Summary of the reliability scores of the final design and two risk-based designs for an increasing volume reduction given by r . Results estimated using $N_e = 10^4$ and $N_a = 300$.

Design	Risk	$R_w(\theta)$	$R_1(\theta)$	$R_2(\theta)$	$R_3(\theta)$	$s_w(\theta)$	$s_1(\theta)$	$s_2(\theta)$	$s_3(\theta)$
θ_{final}^*	$r = 0\%$	[0,0.0833]	[0,0]	[0,0.0833]	[0,0]	1.56×10^{-4}	0	1.56×10^{-4}	0
θ_{final}^*	$r = 26.9\%$	[0,0.026]	[0,0]	[0,0.026]	[0,0]	0.74×10^{-4}	0	0.74×10^{-4}	0
θ_{final}^*	$r = 82.7\%$	[0,0.01]	[0,0]	[0,0.01]	[0,0]	0.47×10^{-4}	0	0.47×10^{-4}	0
θ_{risk-1}^*	$r = 0\%$	[0,0.73]	[0,0.28]	[0,0.73]	[0,0]	0.029	0.029	0.003	0
θ_{risk-1}^*	$r = 26.9\%$	[0,0.33]	[0,0.11]	[0,0.33]	[0,0]	0.01	0.01	0.001	0
θ_{risk-1}^*	$r = 82.7\%$	[0,0.0625]	[0,0.025]	[0,0.0625]	[0,0]	0.0022	0.0022	1.36×10^{-4}	0
θ_{risk-7}^*	$r = 0\%$	[0,0.096]	[0,0]	[0,0.096]	[0,0]	2.77×10^{-4}	0	2.77×10^{-4}	0
θ_{risk-7}^*	$r = 26.9\%$	[0,0.044]	[0,0]	[0,0.044]	[0,0]	1.1×10^{-4}	0	1.1×10^{-4}	0
θ_{risk-7}^*	$r = 82.7\%$	[0,0.022]	[0,0]	[0,0.022]	[0,0]	0.32×10^{-4}	0	0.32×10^{-4}	0

7. Discussion

7.1. Issues/problems

The proposed solution faced a number of issues during development, many of which stem from the black-box nature of the problem. Bayesian updating allowed for flexibility in specifying an imprecise model, but this was extremely costly and there is no good method for selecting the optimum epsilon parameter in the likelihood (6), which must be selected heuristically. Part of the flexibility of the model stems from the ability to capture dependence through the correlation matrix between the marginal beta distributions. However, these correlations did not update to a notable degree during Bayesian updating. This may reflect the imprecision in attempting to assign a precise distribution to the aleatory space, and supports the use of imprecise methods, but it seems to have introduced conservatism where there may have been alternative approaches that could have made more use of the dependence between samples.

The sliced normal distributions provided a means of efficiently drawing additional samples whilst maintaining dependence structures, but the expressiveness was limited by computational cost. The sliced normal used in this approach had 234 dimensions based on a two degree of freedom expansion (in a 24 dimensional problem). Additional degrees of freedom would have improved the ability to capture multi-modality accurately, for example. But this was not possible due to the computational cost of such an expansion. A more efficient means of dealing with dimensionality when using sliced normals could have allowed for a more accurate representation of the Bayesian posterior. This problem gets worse the more dimensions your problem has.

A sensitivity analysis approach based on sampling initially failed due to the potential for samples drawn from the reduced p-box falling outside the original when resampling occurred. This had the potential to cause negative sensitivity indices for some reductions. This was remedied by restricting resampling to subsets of the original sample set and enforcing inclusion monotonicity, but this approach is limited by the extent to which epistemic uncertainty defines the reduced p-box. If epistemic uncertainty is greatly reduced after reduction, the resulting set of sub-samples may be too small to make practical use of.

The risk based design approach suggested a means of taking a risk where the epistemic space was reduced. In this solution, there was no unique means of producing a given reduction of the epistemic space. This could lead to potential false confidence, since reduction could preferentially target epistemic space which is associated with failure states. This may not reflect any true risk, and the means of taking such a risk should be defined before attempting any risk based design optimisation. This was remedied by proposing a number of candidate designs that are optimal for different risk approaches.

7.2. Limitations

The approach also has the following limitations:

- Time correlation information is not used for calibration: some information is lost in the Bayesian updating when taking the product of the marginal likelihoods (independence assumption).
- A prior must be selected for the second order model, and it is difficult to make use of prior knowledge in this case.
- The second order and sliced normal distributions suffer from increasing dimensionality, computational cost needs to be considered.
- Engineering problems are rarely true black boxes, in which case intrusive analysis can provide more robust results.
- Epistemic bounds on black-box models may not be rigorous.
- Optimisation requires a specified means of volume reduction.
- The probabilistic model is inherently imprecise.

Although a reasonable approach for limited data, it should be noted that unless the aleatory uncertainty follows the same distribution that was selected for the probabilistic model (beta/Gaussian copula in this work), then this method will not yield a precise distribution, even with a large data set. This limitation may be solved by selecting a non-parametric distribution for the second order probabilistic model.

7.3. Future work

This work may be extended in the following way, and may solve some of the limitations previously described.

- An uncertainty metric which accounts for dependence

Accounting for dependencies in the output data would give a more accurate calibration. In this work independence had to be assumed since the selected uncertainty metric is one dimensional. A possible candidate is the multivariate Wassertsein metric, a popular distance in Optimal Transport Theory [123]. However an efficient method to compute this distance between multivariate distribution functions is sought.

- Non-parametric second order distributions

The p-boxes produced in this work will not reduce to a precise distribution, even with a large amount of data. This is because a parametric distribution family was used in the probabilistic model. Using non-parametric distributions would solve this limitation and give a more flexible calibration.

- A non-probabilistic solution to this problem

A possibilistic approach would allow for unique means of epistemic volume reduction and would also prevent false confidence in reliability statements due to the consonant nature of possibilistic structures. The methods for doing so are currently under development, and the most appropriate means of assigning necessity to the input space based on the black box output are not yet defined. One potential solution is to represent the Bayesian posterior with a possibilistic transform of the sliced-normal distribution. This was investigated over the course of this work and is currently under development.

- Investigate impact of other priors, Jeffreys, Haldane etc.

A uniform prior was used here for all model parameters, but the black box nature of the problem and the lack of context gives no justification for any individual prior.

- Testing for false confidence

Failure regions that are finite may lead to false confidence when a probabilistic approach is taken to reliability analysis. Assessing the impact of this would be a useful aid in approaching design problems using the approach described in this manuscript. This may highlight where false-confidence is likely to cause concern, and where non-probabilistic approaches may alleviate these problems.

- Efficient RBDO under mixed uncertainties

There are several efficient reliability optimization methods that work well in situations where only aleatory uncertainty is considered, e.g., decoupled RBDO methods, multi-level surrogates. However, efficient optimization methods for mixed uncertainties are not as mature. Genetic algorithms provided a solution, but at significant computational expense. Resolving this would greatly expand the applicability of this approach.

- Probabilistic model free RBDO

A probabilistic model free optimisation under uncertainty method may be applied into this problem, where the reliability is computed directly from the experimental data in a robust way [105]. This would remove the need for calibration (problem A).

- Use of surrogate modelling (re dimensionality/ expense)

The use of surrogate modelling would allow for more samples to be used and for the reliability to be computed to a higher accuracy. Surrogate models would also be required for the methods of this paper to be applied on more computationally expensive models.

8. Conclusions

In this paper we present a harmonised framework for dealing with uncertainty in engineering design. The framework described within this document builds upon recent advances within the UQ literature. We have consolidated a plethora of methodologies and analysis procedures into one workable framework for engineering design with black box models. The framework deals with the four common sources of uncertainty in engineering design: (i) limited data, (ii) variability, (iii) unknown parameter values, and (iv) uncertain dependencies. We frame the design challenge, inclusive of these sources of uncertainty, as uncertainty/risk analytic problems:

- uncertainty characterisation,
- calibration,
- sensitivity analysis and uncertainty reduction,
- reliability analysis,
- design optimisation, and
- risk-, reliability-based design optimisation.

Our novel sequence of solutions quantifies each uncertainty, and provides support to practitioners to: make decisions about further data acquisition, assess the reliability of prototype designs, optimise the design against severe failures and uncertainties, and refine design decisions.

To demonstrate the applicability, efficacy, and efficiency of the proposed solution strategy we apply our framework to the NASA 2020 challenge; which combines problems of statistical inference, with complex multi-dimensional and black-box engineering systems. The challenge emulates common fundamental problems across science and engineering, so our success in addressing this challenge showcases the flexibility of our approach.

CRedit authorship contribution statement

A. Gray: Conceptualization, Methodology, Software, Data curation, Visualization, Writing - original draft, Writing - review & editing. **A. Wimbush:** Conceptualization, Methodology, Software, Data curation, Visualization, Writing - original draft, Writing - review & editing. **M. de Angelis:** Conceptualization, Methodology, Software, Data curation, Visualization, Writing - original draft, Writing - review & editing. **P.O. Hristov:** Conceptualization, Methodology, Software, Data curation, Visualization, Writing - original draft, Writing - review & editing. **D. Calleja:** Conceptualization, Writing - review & editing. **E. Miralles-Dolz:** Conceptualization, Writing - review & editing. **R. Rocchetta:** Conceptualization, Methodology, Software, Data curation, Visualization, Writing - original draft, Writing - review & editing.

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.

Acknowledgements

The authors would like to thank Professor Scott Ferson, Chair of Risk and Uncertainty at the Institute for Risk and Uncertainty, University of Liverpool, in particular for his guidance, support, mentorship and seemingly bottomless knowledge of uncertainty analysis.

This research was funded in part by EPSRC and ESRC CDT in Risk and Uncertainty, EP/L015927/1, EPSRC “*Digital twins for improved dynamic design*” EP/R006768/1, the Digital Lifecycle Twins for predictive maintenance “ITEA3-2018-17030-DayTiMe” grant, and EUROfusion Consortium and Euratom research and training programme under grant agreement No 633053. The views and opinions expressed herein do not necessarily reflect those of the funding organisations that have supported this work.

Appendix A. Nomenclature

$a \in A$	vector of n_a aleatory variables
$e_{1..4} \in E_0$	vector of n_e epistemic parameters
$\theta \in \Theta$	vector n_θ design variables
$A \subset \mathbb{R}^{n_a}$	Events space
$E \subset \mathbb{R}^{n_e}$	Set of epistemic parameters
$\Theta \subset \mathbb{R}^{n_\theta}$	Space of candidate designs
$y(t)$	Output of the subsystem
$z(t)$	Output of the integrated system
$g \in \mathbb{R}^{n_g}$	Vector of n_a reliability performance functions
$w \in \mathbb{R}$	Worst-case reliability performance function
f_u	Joint density of uncertain factors u
F_u	Cumulative distribution of uncertain factors u
$[F(u), \bar{F}(u)]$	Probability box of u
UM	An uncertainty model described by a pair (f_a, E)
D_i	Data set of n_i system responses
$P_f(e, \theta)$	Precise probability of failure
$R_w(\theta)$	Epistemic bounds on the probability of failure for all requirements
$R_j(\theta)$	Epistemic bounds on the probability of failure for requirement j
$s(\theta)$	Severity function for all requirements
$e_{1..24} \in E$	Augmented epistemic vector
N	Number of samples
ϵ	Tuning parameter for the updating
$d(F_{u1}; F_{u2})$	Value of the stochastic area metric
$e_{5..24} \in E_B$	Vector of epistemic parameters defining f_a
E_B	Augmented epistemic space
P_{fj}	Failure probability for requirement j
$\hat{P}_j(\theta)$	Estimator of the failure probability
$l(\theta, r)$	Gain metric for a risk r
ϕ	Unnormalised gaussian density function
$q(u; n_d)$	Monomial expansion of u of degree n_d

References

- [1] R. Ghanem, D. Higdon, H. Owhadi, *Handbook of Uncertainty Quantification*, vol. 6, Springer, 2017.
- [2] A. Wimbush, A. Gray, R. Rocchetta, M. de Angelis, P. Hristov, E. Miralles-Dolz, D. Calleja, *Bayesian Calibration and Probability Bounds Analysis Solution to the Nasa 2020 UQ Challenge on Optimization Under Uncertainty*, in: *Proceedings of the 30th European Safety and Reliability Conference and the 15th Probabilistic Safety Assessment and Management Conference*, 2020.
- [3] S. Ferson, W. Troy Tucker, *Sensitivity analysis using probability bounding*, *Reliability Engineering and System Safety* 91 (10–11) (2006) 1435–1442.
- [4] A. Ramdas, N. Garcia, M. Cuturi, *On Wasserstein Two Sample Testing and Related Families of Nonparametric Tests*, 2015.
- [5] L.G. Crespo, B.K. Colbert, S.P. Kenny, D.P. Giesy, *On the quantification of aleatory and epistemic uncertainty using Sliced-Normal distributions*, *Systems & Control Letters* 134 (2019), 104560.
- [6] L.G. Crespo, S.P. Kenny, *The NASA langley challenge on optimization under uncertainty*, *Mechanical Systems and Signal Processing* 152 (2021) 107405, ISSN 0888-3270, doi:<https://doi.org/10.1016/j.ymssp.2020.107405>.
- [7] L.G. Crespo, S.P. Kenny, D.P. Giesy, *The NASA langley multidisciplinary uncertainty quantification challenge*, in: *16th AIAA Non-Deterministic Approaches Conference*, 2014, <https://doi.org/10.2514/6.2014-1347>. URL: <https://arc.aiaa.org/doi/abs/10.2514/6.2014-1347>.
- [8] E. Patelli, D.A. Alvarez, M. Broggi, M.d. Angelis, *Uncertainty management in multidisciplinary design of critical safety systems*, *Journal of Aerospace Information Systems* 12 (1) (2015) 140–169, doi:10.2514/1.1010273.
- [9] A. Srivastava, A.K. Subramaniyan, L. Wang, *Hybrid Bayesian solution to NASA Langley Research Center multidisciplinary uncertainty quantification challenge*, *Journal of Aerospace Information Systems* 12 (1) (2015) 114–139, <https://doi.org/10.2514/1.1010266>.
- [10] R. Ghanem, I. Yadegaran, C. Thimmisetty, V. Keshavarzadeh, S. Masri, J. Red-Horse, R. Moser, T. Oliver, P. Spanos, O.J. Aldraihem, *Probabilistic approach to NASA Langley Research Center multidisciplinary uncertainty quantification challenge problem*, *Journal of Aerospace Information Systems* 12 (1) (2015) 170–188, <https://doi.org/10.2514/1.1010271>.
- [11] J. McFarland, B.J. Bichon, D.S. Riha, *A Probabilistic Treatment of Multiple Uncertainty Types: NASA UQ Challenge*, in: *16th AIAA Non-Deterministic Approaches Conference*, doi:10.2514/6.2014-1500, xxxx.

- [12] C. Safta, K. Chowdhary, K. Sargsyan, H.N. Najm, B. Debuschere, L.P. Swiler, M.S. Eldred, Uncertainty quantification methods for model calibration, validation, and risk analysis, in: 16th AIAA Non-Deterministic Approaches Conference, 1497, 2014.
- [13] R.E. Moore, Interval Analysis, vol. 4, Prentice-Hall Englewood Cliffs, 1966.
- [14] W.L. Oberkampf, W.T. Tucker, J. Zhang, L. Ginzburg, D.J. Berleant, S. Ferson, J. Hajagos, R.B. Nelsen, Dependence in probabilistic modeling, Dempster-Shafer theory, and probability bounds analysis., Tech. Rep., Sandia National Laboratories, 2004.
- [15] S. Ferson, V. Kreinovich, L. Ginzburg, D. Myers, K. Sentz, Constructing Probability Boxes and Dempster-Shafer Structures, Technical Report Sandia National Labs doi:10.2172/809606.
- [16] M. Beer, S. Ferson, V. Kreinovich, Imprecise probabilities in engineering analyses, *Mechanical Systems and Signal Processing* 37 (1–2) (2013) 4–29.
- [17] S. Ferson, V. Kreinovich, L. Grinzburg, D. Myers, K. Sentz, Constructing probability boxes and Dempster-Shafer structures, Tech. Rep., Sandia National Lab. (SNL-NM), Albuquerque, NM (United States), 2015.
- [18] S. Ferson, J.G. Hajagos, Arithmetic with uncertain numbers: rigorous and (often) best possible answers, *Reliability Engineering & System Safety* 85 (1–3) (2004) 135–152.
- [19] G. Shafer, *A Mathematical Theory of Evidence*, Princeton University Press, Princeton, 1976.
- [20] J. Goutsias, R.P. Mahler, H.T. Nguyen, *Random Sets: Theory and Applications*, vol. 97, Springer Science & Business Media, 2012.
- [21] A. Beresteanu, I. Molchanov, F. Molinari, Partial identification using random set theory, *Journal of Econometrics* 166 (1) (2012) 17–32.
- [22] D.A. Alvarez, F. Uribe, J.E. Hurtado, Estimation of the lower and upper bounds on the probability of failure using subset simulation and random set theory, *Mechanical Systems and Signal Processing* 100 (2018) 782–801.
- [23] L. Zadeh, Fuzzy sets, *Information and Control* 8 (3) (1965) 338–353, ISSN 0019-9958, doi: 10.1016/S0019-9958(65)90241-X.
- [24] D. Dubois, H. Prade, Random sets and fuzzy interval analysis, *Fuzzy Sets and Systems* 42 (1) (1991) 87–101, ISSN 0165-0114, doi: 10.1016/0165-0114(91)90091-4, mathematical Aspects of Fuzzy Sets.
- [25] T. Denoeux, Maximum likelihood estimation from uncertain data in the belief function framework, *IEEE Transactions on Knowledge and Data Engineering* 25 (1) (2011) 119–130.
- [26] D. Hose, M. Hanss, Consistent inverse probability and possibility propagation, in: 11th Conference of the European Society for Fuzzy Logic and Technology (EUSFLAT 2019), Atlantis Press, 2019, pp. 1–8.
- [27] W. Graf, M. Götz, M. Kaliske, Analysis of dynamical processes under consideration of polymorphic uncertainty, *Structural Safety* 52 (2015) 194–201.
- [28] M. Fina, P. Weber, W. Wagner, Polymorphic uncertainty modeling for the simulation of geometric imperfections in probabilistic design of cylindrical shells, *Structural Safety* 82 (2020) 101894, ISSN 0167-4730, doi:https://doi.org/10.1016/j.strusafe.2019.101894.
- [29] Y. Ben-Haim, *Info-gap Decision Theory: Decisions under Severe Uncertainty*, Elsevier, 2006.
- [30] S. Ferson, V. Kreinovich, L. Ginzburg, D.S. Myers, K. Sentz, Constructing probability boxes and Dempster-Shafer structures, Sandia National Laboratories (2002).
- [31] R.C. Williamson, T. Downs, Probabilistic arithmetic. I. Numerical methods for calculating convolutions and dependency bounds, *International Journal of Approximate Reasoning* 4 (2) (1990) 89–158.
- [32] M. Faes, M. Daub, S. Marelli, E. Patelli, M. Beer, Engineering analysis with probability boxes: a review on computational methods.
- [33] N. Gray, M. De Angelis, S. Ferson, Computing with uncertainty: introducing puffin the automatic uncertainty compiler, in: Proceedings of the 3rd International Conference on Uncertainty Quantification in Computational Sciences and Engineering (UNCECOMP 2019), Institute of Structural Analysis and Antiseismic Research School of Civil Engineering National Technical University of Athens (NTUA) Greece, 2019.
- [34] M. de Angelis, E. Patelli, M. Beer, Advanced line sampling for efficient robust reliability analysis, *Structural Safety* 52 (2015) 170–182, <https://doi.org/10.1016/j.strusafe.2014.10.002>.
- [35] J. Sadeghi, M. de Angelis, E. Patelli, Robust propagation of probability boxes by interval predictor models, *Structural Safety* 82 (2020), 101889.
- [36] R. Rocchetta, Q. Gao, M. Petkovic, Soft-constrained interval predictor models and epistemic reliability intervals: A new tool for uncertainty quantification with limited experimental data, *Mechanical Systems and Signal Processing* 161 (2021) 107973, ISSN 0888-3270, doi:https://doi.org/10.1016/j.ymsp.2021.107973, URL: <https://www.sciencedirect.com/science/article/pii/S088832702100368X>.
- [37] B.-S. Yang, K.J. Kim, Application of Dempster-Shafer theory in fault diagnosis of induction motors using vibration and current signals, *Mechanical Systems and Signal Processing* 20 (2) (2006) 403–420, ISSN 0888-3270, doi: 10.1016/j.ymsp.2004.10.010.
- [38] S. Sun, G. Fu, S. Djordjević, S.-T. Khu, Separating aleatory and epistemic uncertainties: Probabilistic sewer flooding evaluation using probability box, *Journal of Hydrology* 420 (2012) 360–372.
- [39] R. Rocchetta, M. Broggi, E. Patelli, Do we have enough data? Robust reliability via uncertainty quantification, *Applied Mathematical Modelling* 54 (2018) 710–721.
- [40] S. Marelli, B. Sudret, UQLab: A framework for uncertainty quantification in Matlab, in: *Vulnerability, Uncertainty, and Risk: Quantification, Mitigation, and Management*, 2014, pp. 2554–2563.
- [41] E. Patelli, R. Ghanem, D. Higdon, H. Owhadi, COSSAN: a multidisciplinary software suite for uncertainty quantification and risk management, *Handbook of Uncertainty Quantification* (2016) 1–69.
- [42] S. Wojtkiewicz, M. Eldred, R. Field Jr, A. Urbina, J. Red-Horse, Uncertainty quantification in large computational engineering models, in: 19th AIAA Applied Aerodynamics Conference, 1455, 2001.
- [43] R. Rocchetta, E. Patelli, A post-contingency power flow emulator for generalized probabilistic risks assessment of power grids, *Reliability Engineering & System Safety* 197 (2020) 106817, ISSN 0951-8320, doi:https://doi.org/10.1016/j.res.2020.106817.
- [44] A.C. Cullen, H.C. Frey, C.H. Frey, *Probabilistic Techniques in Exposure Assessment: A Handbook for Dealing with Variability and Uncertainty in Models and Inputs*, Springer Science & Business Media, 1999.
- [45] H.C. Frey, D.S. Rhodes, Characterization and simulation of uncertain frequency distributions: Effects of distribution choice, variability, uncertainty, and parameter dependence, *Human and Ecological Risk Assessment: An International Journal* 4 (2) (1998) 423–468.
- [46] S. Ferson, K. Sentz, Epistemic uncertainty in agent-based modeling, in: 7th International Workshop on Reliable Engineering Computing, 2016, pp. 65–82.
- [47] M. De Angelis, S. Ferson, E. Patelli, V. Kreinovich, Black-box propagation of failure probabilities under epistemic uncertainty, in: Proceedings of the 3rd International Conference on Uncertainty Quantification in Computational Sciences and Engineering (UNCECOMP 2019), Institute of Structural Analysis and Antiseismic Research School of Civil Engineering National Technical University of Athens (NTUA) Greece, 2019.
- [48] A. Tarantola, *Inverse Problem Theory and Methods for Model Parameter Estimation*, SIAM, 2005.
- [49] G.E. Box, G.C. Tiao, *Bayesian Inference in Statistical Analysis*, vol. 40, John Wiley & Sons, 2011.
- [50] B. Engquist, K. Ren, Y. Yang, The quadratic Wasserstein metric for inverse data matching, *Inverse Problems* 36 (5) (2020), 055001, <https://doi.org/10.1088/1361-6420/ab7e04>. ISSN 1361-6420.
- [51] J.L. Beck, L.S. Katafygiotis, Updating models and their uncertainties. I: Bayesian statistical framework, *Journal of Engineering Mechanics* 124 (4) (1998) 455–461.
- [52] J.L. Beck, S.-K. Au, Bayesian Updating of Structural Models and Reliability using Markov Chain Monte Carlo Simulation, *Journal of Engineering Mechanics* 128 (4) (2002) 380–391.
- [53] J. Sadeghi, M. de Angelis, E. Patelli, Frequentist history matching with interval predictor models, *Applied Mathematical Modelling* 61 (2018) 29–48.
- [54] R. Rocchetta, M. Broggi, Q. Huchet, E. Patelli, On-line Bayesian model updating for structural health monitoring, *Mechanical Systems and Signal Processing* 103 (2018) 174–195, ISSN 0888-3270, doi:https://doi.org/10.1016/j.ymsp.2017.10.015.
- [55] M.S. Balch, R. Martin, S. Ferson, Satellite conjunction analysis and the false confidence theorem, *Proceedings of the Royal Society A* 475 (2227) (2019) 20180565.
- [56] R. Martin, False confidence, non-additive beliefs, and valid statistical inference, *International Journal of Approximate Reasoning* 113 (2019) 39–73.
- [57] R. Martin, C. Liu, *Inferential Models: Reasoning with Uncertainty*, vol. 145, CRC Press, 2015.

- [58] R. Martin, An imprecise-probabilistic characterization of frequentist statistical inference, 2021.
- [59] M. Faes, M. Broggi, E. Patelli, Y. Govers, J. Mottershead, M. Beer, D. Moens, Inverse quantification of epistemic uncertainty under scarce data: Bayesian or Interval approach?, in: *13th International Conference on Applications of Statistics and Probability in Civil Engineering*, KOR, 2019.
- [60] H.H. Khodaparast, Y. Govers, I. Dayyani, S. Adhikari, M. Link, M. Friswell, J. Mottershead, J. Sienz, Fuzzy finite element model updating of the DLR AIRMOD test structure, *Applied Mathematical Modelling* 52 (2017) 512–526, ISSN 0307-904X, doi: 10.1016/j.apm.2017.08.001.
- [61] J. Berger, *Statistical Decision Theory and Bayesian Analysis*, Springer Series in Statistics, Springer, New York, ISBN 9781475742862, 2013.
- [62] Y. Serhat Erdogan, P. Gundes Bakir, Inverse propagation of uncertainties in finite element model updating through use of fuzzy arithmetic, *Engineering Applications of Artificial Intelligence* 26 (1) (2013) 357–367, ISSN 0952-1976, doi: 10.1016/j.engappai.2012.10.003.
- [63] H.H. Khodaparast, J.E. Mottershead, K.J. Badcock, Interval model updating with irreducible uncertainty using the Kriging predictor, *Mechanical Systems and Signal Processing* 25 (4) (2011) 1204–1226, ISSN 0888-3270, doi: 10.1016/j.ymssp.2010.10.009.
- [64] 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.
- [65] Modeling correlation and dependence among intervals.
- [66] C. Soize, A computational inverse method for identification of non-Gaussian random fields using the Bayesian approach in very high dimension, *Computer Methods in Applied Mechanics and Engineering* 200 (45–46) (2011) 3083–3099.
- [67] D.A. Henderson, R.J. Boys, K.J. Krishnan, C. Lawless, D.J. Wilkinson, Bayesian emulation and calibration of a stochastic computer model of mitochondrial DNA deletions in substantia nigra neurons, *Journal of the American Statistical Association* 104 (485) (2009) 76–87.
- [68] I. Vernon, M. Goldstein, R. Bower, Galaxy formation: Bayesian history matching for the observable universe, *Statistical Science* 29 (1) (2014) 81–90.
- [69] M. Faes, J. Cerneels, D. Vandepitte, D. Moens, Identification of interval fields for spatial uncertainty representation in finite element models, *Proceedings of the ECCOMAS Congress* 30 (2016) 27–30.
- [70] M. Faes, D. Moens, Imprecise random field analysis with parametrized kernel functions, *Mechanical Systems and Signal Processing* 134 (2019), 106334.
- [71] 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.
- [72] 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.
- [73] E. Bernton, P.E. Jacob, M. Gerber, C.P. Robert, Approximate Bayesian computation with the Wasserstein distance, arXiv preprint arXiv:1905.03747.
- [74] A. Lye, A. Ciccirello, E. Patelli, Sampling methods for solving Bayesian model updating problems: A tutorial, *Mechanical Systems and Signal Processing* 159 (2021) 107760, ISSN 0888-3270, doi:https://doi.org/10.1016/j.ymssp.2021.107760, URL: https://www.sciencedirect.com/science/article/pii/S0888327021001552.
- [75] S.-K. Au, J.L. Beck, Estimation of small failure probabilities in high dimensions by subset simulation, *Probabilistic Engineering Mechanics* 16 (4) (2001) 263–277, ISSN 0266-8920, doi: 10.1016/S0266-8920(01)00019-4.
- [76] J. Chang, Y.-C. Chen, Transitional Markov chain Monte Carlo method for Bayesian model updating, model class selection, and model averaging, *Journal of Engineering Mechanics* 133 (7) (2007) 816–832.
- [77] W. Betz, I. Papaioannou, D. Straub, Transitional Markov Chain Monte Carlo: Observations and Improvements 142 (5) (2016) 1–10, [https://doi.org/10.1061/\(ASCE\)EM.1943-7889.0001066](https://doi.org/10.1061/(ASCE)EM.1943-7889.0001066).
- [78] T. Xifara, C. Sherlock, S. Livingstone, S. Byrne, M. Girolami, Langevin diffusions and the Metropolis-adjusted Langevin algorithm, *Statistics & Probability Letters* 91 (2014) 14–19, ISSN 0167-7152, doi:https://doi.org/10.1016/j.spl.2014.04.002.
- [79] E.D. Nino-Ruiz, A numerical method for solving linear systems in the preconditioned Crank–Nicolson algorithm, *Applied Mathematics Letters* 104 (2020) 106254, ISSN 0893-9659, doi:https://doi.org/10.1016/j.aml.2020.106254.
- [80] R.M. Neal, et al., MCMC using Hamiltonian dynamics, *Handbook of Markov Chain Monte Carlo* 2 (11) (2011) 2.
- [81] X. Lu, V. Perrone, L. Hasenclever, Y.W. Teh, S. Vollmer, Relativistic monte carlo, in: *Artificial Intelligence and Statistics*, PMLR, 2017, pp. 1236–1245.
- [82] D. Foreman-Mackey, D.W. Hogg, D. Lang, J. Goodman, emcee: the MCMC hammer, *Publications of the Astronomical Society of the Pacific* 125 (925) (2013) 306.
- [83] S.L. Cotter, G.O. Roberts, A.M. Stuart, D. White, MCMC methods for functions: modifying old algorithms to make them faster, *Statistical Science* 28 (3) (2013) 424–446, <https://doi.org/10.1214/13-STS421>.
- [84] M. Thompson, R.M. Neal, Covariance-adaptive slice sampling, arXiv preprint arXiv:1003.3201.
- [85] R.D. Morey, J.N. Rouder, M.S. Pratte, P.L. Speckman, Using MCMC chain outputs to efficiently estimate Bayes factors, *Journal of Mathematical Psychology* 55 (5) (2011) 368–378.
- [86] C. Andrieu, J. Thoms, A tutorial on adaptive MCMC, *Statistics and Computing* 18 (4) (2008) 343–373.
- [87] A. Saltelli, S. Tarantola, F. Campolongo, M. Ratto, *Sensitivity Analysis in Practice: A Guide to Assessing Scientific Models*, vol. 1, Wiley Online Library, 2004.
- [88] I.M. Sobol, Global sensitivity indices for nonlinear mathematical models and their Monte Carlo estimates, *Mathematics and Computers in Simulation* 55 (1–3) (2001) 271–280.
- [89] A. Saltelli, Making best use of model evaluations to compute sensitivity indices, *Computer Physics Communications* 145 (2) (2002) 280–297.
- [90] J.O. Berger, E. Moreno, L.R. Pericchi, M.J. Bayarri, J.M. Bernardo, J.A. Cano, J. De la Horra, J. Martín, D. Ríos-Insúa, B. Betrò, et al., An overview of robust Bayesian analysis, *Test* 3 (1) (1994) 5–124.
- [91] R. Weiss, An approach to Bayesian sensitivity analysis, *Journal of the Royal Statistical Society: Series B (Methodological)* 58 (4) (1996) 739–750.
- [92] L.C. McCandless, P. Gustafson, A. Levy, Bayesian sensitivity analysis for unmeasured confounding in observational studies, *Statistics in Medicine* 26 (11) (2007) 2331–2347.
- [93] I. Enevoldsen, J. Sørensen, Reliability-based optimization in structural engineering, *Structural Safety* 15 (3) (1994) 169–196, [https://doi.org/10.1016/0167-4730\(94\)90039-6](https://doi.org/10.1016/0167-4730(94)90039-6).
- [94] Z. Meng, B. Keshetegar, Adaptive conjugate single-loop method for efficient reliability-based design and topology optimization, *Computer Methods in Applied Mechanics and Engineering* 344 (2019) 95–119, ISSN 0045-7825, doi:https://doi.org/10.1016/j.cma.2018.10.009.
- [95] Z. Meng, H. Zhou, G. Li, D. Yang, A decoupled approach for non-probabilistic reliability-based design optimization, *Computers & Structures* 175 (2016) 65–73, ISSN 0045-7949, doi:https://doi.org/10.1016/j.compstruc.2016.06.008.
- [96] H.-S. Li, S.-K. Au, Design optimization using Subset Simulation algorithm, *Structural Safety* 32 (6) (2010) 384–392, ISSN 0167-4730, doi: 10.1016/j.strusafe.2010.03.001, modeling and Analysis of Rare and Imprecise Information.
- [97] R. Schöbi, Surrogate models for uncertainty quantification in the context of imprecise probability modelling, Ph.D. thesis, ETH Zürich, Zürich, Switzerland, 2017.
- [98] A.T. Beck, W.J. de Santana Gomes, A comparison of deterministic, reliability-based and risk-based structural optimization under uncertainty, *Probabilistic Engineering Mechanics* 28 (2012) 18–29, ISSN 0266-8920, doi:https://doi.org/10.1016/j.proengmech.2011.08.007, computational Stochastic Mechanics – CSM6.
- [99] A. Beck, W. Gomes, R. Lopez, L. Miguel, A comparison between robust and risk-based optimization under uncertainty, *Structural and Multidisciplinary Optimization* 52, doi:10.1007/s00158-015-1253-9.
- [100] R. Rockafellar, J. Royset, On buffered failure probability in design and optimization of structures, *Reliability Engineering & System Safety* 95 (5) (2010) 499–510, ISSN 0951-8320, doi: 10.1016/j.res.2010.01.001.
- [101] W. Li, C. Li, L. Gao, M. Xiao, Risk-based design optimization under hybrid uncertainties, *Engineering with Computers* (2020) 1–13, <https://doi.org/10.1007/s00366-020-01196-4>.
- [102] W. Xie, On distributionally robust chance constrained programs with Wasserstein distance, *Mathematical Programming* (2018) 1436–4646, <https://doi.org/10.1007/s10107-019-01445-5>.

- [103] Z. Meng, Z. Zhang, H. Zhou, A novel experimental data-driven exponential convex model for reliability assessment with uncertain-but-bounded parameters, *Applied Mathematical Modelling* 77 (2020) 773–787, ISSN 0307-904X, doi: 10.1016/j.apm.2019.08.010.
- [104] R. Rocchetta, L.G. Crespo, A scenario optimization approach to reliability-based and risk-based design: soft-constrained modulation of failure probability bounds, *Reliability Engineering & System Safety* 216 (2021), 107900. ISSN 0951-8320, <https://doi.org/10.1016/j.res.2021.107900>.
- [105] R. Rocchetta, L.G. Crespo, S.P. Kenny, A scenario optimization approach to reliability-based design, *Reliability Engineering & System Safety* 196 (2020) 106755, ISSN 0951-8320, doi:<https://doi.org/10.1016/j.res.2019.106755>.
- [106] L.G. Crespo, S.P. Kenny, The NASA Langley UQ Challenge on Optimization Under Uncertainty, ESREL 2020.
- [107] K. Bury, *Statistical Distributions in Engineering*, Cambridge University Press, 1999.
- [108] H. Joe, Generating random correlation matrices based on partial correlations, *Journal of Multivariate Analysis* 97 (10) (2006) 2177–2189.
- [109] S. Ferson, W.L. Oberkampf, L. Ginzburg, Model validation and predictive capability for the thermal challenge problem, *Computer Methods in Applied Mechanics and Engineering* 197 (29–32) (2008) 2408–2430.
- [110] C. Soize, A computational inverse method for identification of non-Gaussian random fields using the Bayesian approach in very high dimension, *Computer Methods in Applied Mechanics and Engineering* 200 (45) (2011) 3083–3099, ISSN 0045-7825, doi: 10.1016/j.cma.2011.07.005.
- [111] B.K. Colbert, L.G. Crespo, M.M. Peet, Improving the uncertainty quantification of sliced normal distributions by scaling the covariance matrix, in: *Proceedings of the American Control Conference, 2020*.
- [112] W. Zhu, N. Chen, J. Liu, M. Beer, A probability-box-based method for propagation of multiple types of epistemic uncertainties and its application on composite structural-acoustic system, *Mechanical Systems and Signal Processing* 149 (2021) 107184, ISSN 0888-3270, doi:<https://doi.org/10.1016/j.ymsp.2020.107184>.
- [113] J. Sadeghi, M. De Angelis, E. Patelli, Analytic probabilistic safety analysis under severe uncertainty, *ASCE-ASME Journal of Risk and Uncertainty in Engineering Systems, Part A: Civil Engineering* 6 (1) (2020) 04019019.
- [114] H. Zhang, R.L. Mullen, R.L. Muhanna, Interval Monte Carlo methods for structural reliability, *Structural Safety* 32 (3) (2010) 183–190.
- [115] S. Ferson, J. Hajagos, W.T. Tucker, Probability bounds analysis is a global sensitivity analysis, in: *International Conference on Sensitivity Analysis of Model Output (SAMO)*, 2004.
- [116] T. Homma, A. Saltelli, Importance measures in global sensitivity analysis of nonlinear models, *Reliability Engineering & System Safety* 52 (1) (1996) 1–17.
- [117] H. Rabitz, Global sensitivity analysis for systems with independent and/or correlated inputs, *Procedia - Social and Behavioral Sciences* 2 (6) (2010) 7587–7589.
- [118] G. Li, H. Rabitz, P.E. Yelvington, O.O. Oluwole, F. Bacon, C.E. Kolb, J. Schoendorf, Global sensitivity analysis for systems with independent and/or correlated inputs, *The Journal of Physical Chemistry A* 114 (19) (2010) 6022–6032.
- [119] S. Sarykalin, G. Serraino, S. Uryasev, Value-at-Risk vs. Conditional Value-at-Risk in Risk Management and Optimization, chap. Chapter 13, Institute for Operations Research and the Management Sciences (INFORMS), 2014, pp. 270–294, doi:10.1287/educ.1080.0052.
- [120] J. Guo, S. Du, Y. Wang, S. Zeng, Time-dependent global sensitivity analysis for long-term degeneracy model using polynomial chaos, *Advances in Mechanical Engineering* 6 (2014), 719825.
- [121] R.Y. Liu, J.M. Parelius, K. Singh, et al., Multivariate analysis by data depth: descriptive statistics, graphics and inference, (with discussion and a rejoinder by liu and singh), *The Annals of Statistics* 27 (3) (1999) 783–858.
- [122] V. Barnett, The ordering of multivariate data, *Journal of the Royal Statistical Society: Series A (General)* 139 (3) (1976) 318–344.
- [123] C. Villani, *Optimal Transport: Old and New*, vol. 338, Springer, 2009.