Robust optimisation of a dynamic Black-box system under severe uncertainty: A Distribution-free framework

Adolphus Lye[a,*], Masaru Kitahara[b,*], Matteo Broggi[b], Edoardo Patelli[c,**]

[a] *Institute for Risk and Uncertainty, University of Liverpool*
[b] *Institute for Risk and Reliability, Leibniz Universität Hannover*
[c] *Centre for Intelligent Infrastructure, Department of Civil and Environmental Engineering, University of Strathclyde*

Abstract

In the real world, a significant challenge faced in designing critical systems is the lack of available data. This results in a large degree of uncertainty and the need for uncertainty quantification tools so as to make risk-informed decisions. The NASA-Langley UQ Challenge 2019 seeks to provide such setting, requiring different discipline-independent approaches to address typical tasks required for the design of critical systems.

This paper addresses the NASA-Langley UQ Challenge by proposing 4 key techniques to provide the solution to the challenge: 1) a distribution-free Bayesian model updating framework for the calibration of the uncertainty model; 2) an adaptive pinching approach to analyse and rank the relative sensitivity of the epistemic parameters; 3) the probability bounds analysis to estimate failure probabilities; and 4) a Non-intrusive Stochastic Simulation approach to identify an optimal design point.

Keywords: Uncertainty Quantification, Model Class Selection, Sensitivity Analysis, Staircase Density Function, Robust optimisation, Non-intrusive Imprecise Stochastic Simulation

*Both authors have equally contributed towards the solution of the problem and are equally addressed as the "first author" of the present work

**Corresponding author
Email address: edoardo.patelli@strath.ac.uk (Edoardo Patelli)

1. Introduction

1.1. Research Context

The design of critical safety systems is often associated with the availability of limited data. Despite such challenge posed, the system needs to be designed in order to cope with the unavoidable uncertainty. Such uncertainty can be classified as either aleatory or epistemic uncertainty [1, 2]. Aleatory uncertainty is often considered as the irreducible uncertainty that is caused by the inherent randomness of the system [3] and generally modelled as random variables according to some distribution function [4, 5]. On the other hand, epistemic uncertainty is caused by a lack of or limited knowledge which can be theoretically reduced or eliminated through, for instance, data collection [6]. An epistemic parameter is generally represented by a fixed value within a bounded set whose intervals reflect the level of knowledge on the parameter [3]. The lower the level of knowledge, the larger the interval of this bounded set. It is important to note that the aleatory and epistemic uncertainty can refer to the same physical quantity and, therefore, such classification becomes fuzzy. In fact, the aleatory uncertainty can be seen as the remaining uncertainty after a campaign, aimed at reducing the epistemic uncertainty, is performed.

The design of systems under uncertainty requires the availability of robust and efficient tools for uncertainty characterisation and quantification. In order to check the availability of discipline independent tools and applicability of such tools, NASA Langley proposed a new UQ Challenge problem in 2019 [3] with the purpose of modelling the dynamic behaviour of a system, analysing its operational reliability, and devising an improved design configuration for the system under uncertainty. This UQ Challenge problem follows from the success of the previous edition in 2013 [7].

In this challenge, a "Black-box" computational model of a physical system is used to evaluate and improve its reliability. Unlike the previous challenge [7], the Uncertainty Model (UM) to the respective aleatory input parameters are completely unknown and they are to be derived by the participants. In

addition, the response of the system is time-dependent providing a realistic setting under which different tasks will be addressed. This is because in the real-world, prior distributional knowledge to such models associated with the parameters of interest are usually unavailable.

This paper is part of a Special Issue providing the solution for the NASA-Langley UQ Challenge problem. Therefore, for the sake of the content length, a detailed explanation to the problem is not shown. Instead, a summary of the challenge and the description of the notations is provided.

1.2. The NASA-Langley UQ Challenge problem (2019)

The system is characterised by a design point $\boldsymbol{\theta}$ with 9 real components (i.e. $\boldsymbol{\theta} \in \mathbb{R}^{n_\theta}$), and an uncertain model δ comprising of elements $\boldsymbol{a}$ and $\boldsymbol{e}$ [3]. $\boldsymbol{a}$ denotes the vector of 5 aleatory parameters real components while $\boldsymbol{e}$ denotes the vector of 4 epistemic parameters.

The aleatory space A is represented as $\boldsymbol{a} \sim f_{\boldsymbol{a}}$ whereby $f_{\boldsymbol{a}}$ is the joint density function. The initial aleatory space is $A_0 = [0, 2]^5$. The epistemic space E is represented as $\boldsymbol{e} \sim E$. The initial epistemic space is $E_0 = [0, 2]^4$. Hence, the UM for δ is fully characterized by: $\langle f_{\boldsymbol{a}}, E \rangle$.

The system of interest consists of a set of interconnected subsystems for which δ is concentrated in one of these subsystems. This subsystem is modelled by a Black-box model function $\hat{y} = yfun(\boldsymbol{a}, \boldsymbol{e}, t)$, where $t \in [0, 5]$ s is the time parameter. The output of this subsystem is represented as a discrete time history: $y^l(t) = [y^l(0), y^l(dt), \ldots, y^l(5000 \cdot dt)]$, where $l = 1, \ldots, 100$, and $dt = 0.001$ s. This yields a total of 5001 data of $y^l(t)$ per given l and the entire time history data is denoted as $D_1 = \{y^l(t)\}_{l=1,\ldots,100}$.

The goal of this challenge can be summarised as follows [3]:

A. To create an UM for δ;

B. To decide a limited number of refinements (up to 4) on the epistemic variables;

C. To perform a reliability analysis on a given design point $\boldsymbol{\theta}$;

D. To identify a new $\boldsymbol{\theta}$ with improved reliability;

E. To improve the UM for δ and $\boldsymbol{\theta}$ given observations of the integrated system.

It needs to be highlighted that Task F of the challenge is not addressed in this paper.

2. Task A: Model Calibration and Uncertainty Quantification of the Subsystem

2.1. Modelling strategy and hypotheses

The Bayesian model updating technique is adopted to calibrate the UM using the available data D_1. This provides a probabilistic approach through which the joint distribution function $f_{\boldsymbol{a}}$ can be identified. Usually, Bayesian model updating is not adopted to reduce epistemic uncertainty when represented by intervals. However, the uncertainty of $\boldsymbol{e}$ can be quantified by modelling the intervals as uniform distributions and then computing the posterior distribution. At this point, it is important to note that the posterior distributions are used to define new intervals as already successful proposed in [8]. Therefore, the epistemic parameters e_{i_e}, for $i_e = 1, \ldots, 4$, are assumed to be independent between one another and their respective priors modelled by a non-informative Uniform distribution with bounds defined by the epistemic space E_0 (i.e. see Section 1.2).

In this section, different strategies are adopted to represent the aleatory uncertainty and metrics are used to compare between distributions. As a result, we have obtained a conservative (i.e. low-risk) approach; and a more aggressive (i.e. high-risk) approach aimed at reducing the uncertainty of the UM.

2.2. Bayesian Model updating

Bayesian model updating is a probabilistic model updating approach whose mathematical formulation follows the Bayes' rule introduced by [9, 10]:

$$P(\boldsymbol{\Theta}|\boldsymbol{D}, M) = \frac{P(\boldsymbol{D}|\boldsymbol{\Theta}, M) \cdot P(\boldsymbol{\Theta}|M)}{P(\boldsymbol{D}|M)} \tag{1}$$

where $\boldsymbol{\Theta}$ is the vector of inferred parameters, $\boldsymbol{D}$ denotes the vector of observed data used for model updating, and $M = \{\hat{y}, f_{\boldsymbol{a}}\}$ is the model class which best represents the observed data $\boldsymbol{D}$. The different components in Eq. (1) are the prior distribution (i.e. $P(\boldsymbol{\Theta}|M)$), the likelihood function (i.e. $P(\boldsymbol{D}|\boldsymbol{\Theta}, M)$), the posterior distribution (i.e. $P(\boldsymbol{\Theta}|\boldsymbol{D}, M)$) and finally the evidence (i.e. $P(\boldsymbol{D}|M)$). The term of interest is $P(\boldsymbol{\Theta}|\boldsymbol{D}, M)$ which describes our updated knowledge of $\boldsymbol{\Theta}$ after observing the data $\boldsymbol{D}$ and it is generally presented as an non-normalised distribution by neglecting the normalisation constant.

Advanced sampling techniques are used to sample from such non-normalised distributions [11]. In this work, the Transitional Markov Chain Monte Carlo (TMCMC) sampler is implemented whose algorithm is based on the Adaptive Metropolis-Hastings technique [12] and utilises "transitional" distributions P^j from which samples are obtained sequentially. Details to the algorithm can be found in [11, 13]. The motivations behind the use of TMCMC in this problem are attributed to the following: 1) the algorithm is able to sample from complex-shaped posteriors via "transitional" distributions P^j; 2) it can sample from high-dimensional posteriors (i.e. up to 24 dimensions) [14]; and 3) it computes the evidence $P(\boldsymbol{D}|M)$ which makes the algorithm useful in model selection problems [13].

Due to the large data-set provided (i.e. D_1), it becomes computationally expensive to use a full likelihood function to perform an actual Bayesian computation [15]. For this reason, Approximate Bayesian Computation (ABC) [16, 17] is adopted and defined as [15]:

$$P(\boldsymbol{D}|\boldsymbol{\Theta}, M) \propto exp\left(-\frac{d}{\epsilon}\right)^2 \tag{2}$$

where d is the stochastic distance metric which quantifies the difference between the distribution of the observed data $\boldsymbol{D}$ and the model output of $\hat{y}$, while ϵ is the width factor of the approximate Gaussian function.

2.3. Proposed Approach

To determine the UM for $\boldsymbol{a}$, 5 possible distribution types for $f_{\boldsymbol{a}}$ are identified and listed in Table 1. It needs to be highlighted that the choice of the Staircase Density Function (SDF) presents a distribution-free approach contrary to the other choice of distributions presented in the table. A key strength of SDFs lies in its flexibility in describing a wide range of density shapes, including highly-skewed and/or multi-modal distributions. This makes them highly applicable in modelling the marginal distributions of the aleatory variables whose density shapes are unknown a priori. In this analysis it has been assumed that: 1) the marginal distribution of all the aleatory uncertainties belong to the same distribution class; 2) no dependency exists between all a_{i_a} (i.e. no correlation matrix used).

Aleatory model	Distribution type	Prior distribution parameters
$f_{\boldsymbol{a}}^1$	Beta($\alpha_{i_a}, \beta_{i_a}$)	α_{i_a} (Shape parameter 1): $U[0, 100]$ β_{i_a} (Shape parameter 2): $U[0, 100]$
$f_{\boldsymbol{a}}^2$	Truncated Normal(μ_{i_a},σ_{i_a}) $[TN(a_{i_a}; \mu_{i_a}, \sigma_{i_a})]$	μ_{i_a} (Mean of a_{i_a}): $U[0, 2]$ σ_{i_a} (Standard deviation of a_{i_a}): $U[0.01, 2]$
$f_{\boldsymbol{a}}^3$	Truncated Lognormal $[TLN(a_{i_a}; \mu_{i_a}, \sigma_{i_a})]$	μ_{i_a} (Mean of $log(a_{i_a})$): $U[-10, 10]$ σ_{i_a} (Standard deviation of $log(a_{i_a})$): $U[0.01, 5]$
$f_{\boldsymbol{a}}^4$	Truncated Gamma $[TG(a_{i_a}; \alpha_{i_a}, \beta_{i_a})]$	α_{i_a} (Shape parameter): $U[0, 10]$ β_{i_a} (Scale parameter): $U[0, 10]$
$f_{\boldsymbol{a}}^5$	Staircase Density Function $[SDF(a_{i_a}; \mu_{i_a}, (m_2)_{i_a}, (m_3)_{i_a}, (m_4)_{i_a})]$	μ_{i_a} (Distribution mean of a_{i_a}): $U[0, 2]$ $(m_2)_{i_a}$ (2^{nd} central moment of a_{i_a}): $U[0, 1]$ $(m_3)_{i_a}$ (3^{rd} central moment of a_{i_a}): $U[-\frac{4}{3\sqrt{3}}, \frac{4}{3\sqrt{3}}]$ $(m_4)_{i_a}$ (4^{th} central moment of a_{i_a}): $U[0, \frac{4}{3}]$

Table 1: Distribution type with the non-informative Uniform prior bounds of its corresponding parameters for each aleatory model $f_{\boldsymbol{a}}$.

For the case of $f_{\boldsymbol{a}}^1$ to $f_{\boldsymbol{a}}^4$, the distribution parameters add an additional 10 inferred parameters, while $f_{\boldsymbol{a}}^5$ adds an additional 20 inferred parameters. Each of these parameters are assigned a non-informative Uniform prior with bounds,

stated in Table 1, chosen to ensure sufficient degrees of freedom in the model calibration. It is also assumed that these parameters are independent from one another. This brings the total number of inferred parameters to 24 for the case of the SDF, and 14 for the rest of the distributions.

2.3.1. Distribution-based Approach

For the case of $f_{\boldsymbol{a}}^1$ to $f_{\boldsymbol{a}}^4$, there is a need to reduce the size of the data to reduce the computation cost in evaluating $P(\boldsymbol{D}|\boldsymbol{\Theta}, M)$. To achieve this, the Fast Fourier Transformation (FFT) procedure is performed on D_1 for each l according to [14, 18, 19]:

$$y^l(t) = \sum_{q=0}^{5000} C_q^l \cdot exp\left[-i \cdot q \cdot \omega_0 \cdot t\right] \tag{3}$$

where $\omega_0 = \frac{2 \cdot \pi}{5001}$, and C_q^l is the numerical coefficient with real and imaginary components denoted as $Re(C_q^l)$ and $Im(C_q^l)$ respectively. From which, the amplitude A_q^l and phase angles ϕ_q^l are obtained as follows:

$$A_q^l = \sqrt{Re(C_q^l)^2 + Im(C_q^l)^2} \tag{4}$$

$$\phi_q^l = atan2\left[\frac{Im(C_q^l)}{Re(C_q^l)}\right] \tag{5}$$

To remove the periodicity associated with the values of ϕ_q^l, we introduced a phase shift such that a factor of 2π rad is added or subtracted whenever the jump between consecutive phase angles is greater than π rad. This is achieved using the *unwrap* function in MATLAB to ensure the jump between any consecutive phase angles is always less than π rad. In doing so, it ensures the monotonic behaviour of ϕ_q^l and simplifies its subsequent computation for $P(\boldsymbol{D}|\boldsymbol{\Theta}, M)$. When this is done, we obtain the frequency spectra of A_q^l and ϕ_q^l as shown in Figure 1 where it can be observed that beyond frequencies $\omega > 5.80\ Hz$, the values of A_q^l do not show any additional perturbations for all l, thereby allowing those data to be discarded. Hence, only 30 values of ω between 0 Hz and 5.80 Hz are considered for both A_q^l and ϕ_q^l. Let this set of values of ω be denoted as ω_n,

for $n = 1, \ldots, 30$. This effectively reduces the total number of data for model calibration from 500100 to 6000 (i.e. 3000 for A_q^l and 3000 for ϕ_q^l).

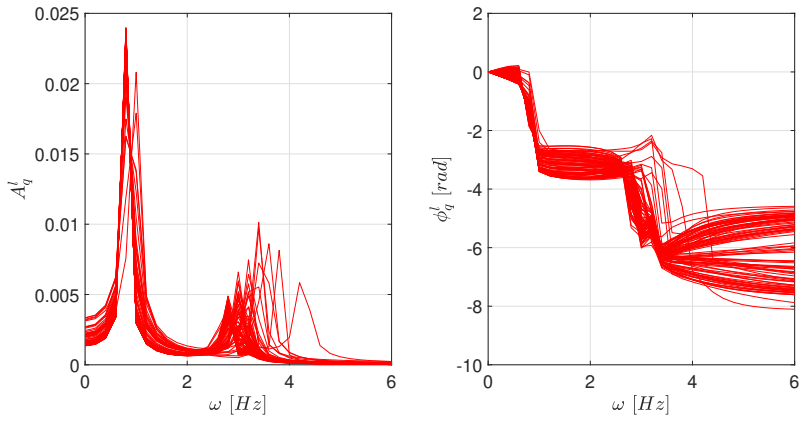

Figure 1: Illustration of the frequency spectra obtained from D_1 via FFT.

To account for the variability of A_q^l and ϕ_q^l at each ω_n, the stochastic distance metric d for $P(\boldsymbol{D}|\boldsymbol{\Theta}, M)$ is the Wasserstein distance defined as [20]:

$$d_W = \int_{-\infty}^{\infty} |F_D(x) - F_{\hat{y}}(x)| \cdot dx \tag{6}$$

whereby $F_D(x)$ and $F_{\hat{y}}(x)$ are the respective Empirical Cumulative Distribution Functions (ECDFs) of the data (i.e. A_q^l and ϕ_q^l) and the stochastic model output of $\hat{y}$ at a given f_n, while x is the variable denoting either A_q^l or ϕ_q^l. In essence, d_W quantifies the enclosed area between both ECDFs. The smaller d_W is, the higher the degree of similarity between the ECDFs of the data and the stochastic prediction by $\hat{y}$ [21]. Using Eq. (2), $P(\boldsymbol{D}|\boldsymbol{\Theta}, M)$ is defined as:

$$P(\boldsymbol{D}|\boldsymbol{\Theta}, M) = \prod_{n=1}^{30} exp\left[-\left(\frac{d_{W,n}^A}{\epsilon_n^A} \right)^2 - \left(\frac{d_{W,n}^\phi}{\epsilon_n^\phi} \right)^2 \right] \tag{7}$$

whereby the values of ϵ_n^A and ϵ_n^ϕ are approximated by the standard deviations of A_q^l and ϕ_q^l respectively at ω_n. Independence is assumed between data sets to reduce computational costs in computing $P(\boldsymbol{D}|\boldsymbol{\Theta}, M)$. However, it needs to be highlighted that in reality, there exists dependencies between the identified

ω_n for each l^{th} sequence. To compute $P(\boldsymbol{D}|\boldsymbol{\Theta}, M)$, 100 model evaluations by $\hat{y}$, per given set of model inputs $\{\boldsymbol{a}, \boldsymbol{e}\}$, are needed to construct $F_{\hat{y}}(x)$.

2.3.2. Distribution-free Approach

For the case of $f_{\boldsymbol{a}}^5$, the distribution is defined by the SDF as [22]:

$$f_{\boldsymbol{a}} = \begin{cases} h_{i_b} \ \ \forall \ \boldsymbol{a} \in ((i_b - 1) \cdot \kappa, i_b \cdot \kappa]^5 & \text{, for } 1 \leq i_b \leq N_b \\ 0 & \text{, otherwise} \end{cases} \tag{8}$$

where $N_b = 50$ is the number of bins, h_{i_b} is the height of the SDF in the ${i_b}^{th}$ bin, and $\kappa = \frac{2}{N_b}$ is the length of each sub-interval. It needs to be noted that $h_{i_b} \geq 0$ for all N_b bins and that their values are obtained by solving the following convex optimisation problem:

$$\hat{h}_{i_b} = \underset{h_{i_b} \geq 0}{argmin} \left\{ J(h) : \sum_{i_b=1}^{N_b} \int_{(i_b-1)\cdot\kappa}^{i_b\cdot\kappa} z \cdot h_{i_b} \cdot dz = \mu_{i_a}, \sum_{i_b=1}^{N_b} \int_{(i_b-1)\cdot\kappa}^{i_b\cdot\kappa} (z - \mu_{i_a})^r \cdot h_{i_b} \cdot dz = (m_r)_{i_a}, r = 0, 2, 3, 4 \right\} \tag{9}$$

where $J(h)$ is the cost-function. Details to $J(h)$ and Eq. (9) can be found in [22, 23].

To avoid a potential error in the implementation of the likelihood function for the Distribution-based approach being brought forward, a different set-up for $P(\boldsymbol{D}|\boldsymbol{\Theta}, M)$ is used in this analysis, where a different stochastic distance is employed and the data is analyzed in the time domain. Unlike in the Distribution-based approach, the discrete Bhattacharyya distance [15] is employed as the stochastic distance metric d:

$$d_B = -log \left\{ \sum_{i_{N_t}=1}^{n_b} \cdots \sum_{i_1=1}^{n_b} \sqrt{p_D(b_{i_1,\ldots,i_{N_t}}) \cdot p_{\hat{y}}(b_{i_1,\ldots,i_{N_t}})} \right\} \tag{10}$$

where $p_D(b_{i_1,\ldots,i_{N_t}})$ and $p_{\hat{y}}(b_{i_1,\ldots,i_{N_t}})$ are the Probability Mass Function (PMF) values of the data from D_1 and the stochastic model output from $\hat{y}$ respectively within the bin $b_{i_1,\ldots,i_{N_t}}$, and $n_b = 20$ is the number of bins used to compute the Bhattacharyya distance. It needs to be highlighted that each bin has $N_t = 5001$

coordinates as it is generated within a N_t-dimensional joint PMF space. Because of this, the resulting joint PMF space has an excessive number of dimensions for a direct evaluation of $P(\boldsymbol{D}|\boldsymbol{\Theta}, M)$. This brings the need for a dimension-reduction procedure which is employed through the following steps [24]:

1. Define the window length $L_w = 50$ and divide the data set $\{y^l(t)\}_{l=1,\ldots,100}$ into $\left\lceil \frac{N_t}{L_w} \right\rceil$ distinct intervals where $\lceil \bullet \rceil$ is the ceil operator;
2. Compute the Root Mean Squared (RMS) values of each interval $\boldsymbol{R} = \left[R_1, \ldots, R_{\lceil \frac{N_t}{L_w} \rceil}\right]$ and generate the sample set of the RMS values $\boldsymbol{R}_D \in \mathbb{R}^{100 \times \lceil \frac{N_t}{L_w} \rceil}$ where:

$$\boldsymbol{R}_D = \left[\boldsymbol{R}_D^1, \ldots, \boldsymbol{R}_D^{\lceil \frac{N_t}{L_w} \rceil}\right], \text{ with } \boldsymbol{R}_D^\nu = [R_{1,\nu}, \ldots, R_{100,\nu}]^T$$

   for $\nu = 1, \ldots, \left\lceil \frac{N_t}{L_w} \right\rceil$ while $\boldsymbol{R}_{\hat{y}} \in \mathbb{R}^{N_{sim} \times \lceil \frac{N_t}{L_w} \rceil}$ where $N_{sim} = 1000$ the number of model evaluations by $\hat{y}$ per given set of model inputs $\{\boldsymbol{a}, \boldsymbol{e}\}$. It needs to be highlighted that the matrix structure of $\boldsymbol{R}_{\hat{y}}$ is similar to that of $\boldsymbol{R}_D$ with the exception of the number of row elements;
3. Evaluate d_B between sample sets $\boldsymbol{R}_D^\nu$ and $\boldsymbol{R}_{\hat{y}}^\nu$ for all ν;
4. Obtain the corresponding RMS values R_{d_B} and use it as the distance metric.

Consequently, $P(\boldsymbol{D}|\boldsymbol{\Theta}, M)$ is re-expressed as:

$$P(\boldsymbol{D}|\boldsymbol{\Theta}, M) = exp\left(-\frac{R_{d_B}}{\epsilon_B}\right)^2 \tag{11}$$

where $\epsilon_B = 0.01$.

2.4. Results

For all set-ups, $N_s = 500$ samples are obtained from the resulting $P(\boldsymbol{\Theta}|\boldsymbol{D}, M)$. Based on the analysis done for all aleatory models $f_{\boldsymbol{a}}$, 2 models are chosen on the basis of their quality of the results and for the subsequent purpose of comparison: $f_{\boldsymbol{a}}^1$ and $f_{\boldsymbol{a}}^5$. It needs to be highlighted, that $f_{\boldsymbol{a}}^1$ is chosen given its relatively higher value of the evidence $P(\boldsymbol{D}|M)$ compared to the other aleatory models used in the Distribution-based approach as shown in Table 2.

Aleatory model	$f_{\boldsymbol{a}}^1$	$f_{\boldsymbol{a}}^2$	$f_{\boldsymbol{a}}^3$	$f_{\boldsymbol{a}}^4$
$P(\boldsymbol{D}\|M)$	3.2229×10^{-7}	2.1952×10^{-7}	5.0815×10^{-10}	1.0180×10^{-9}

Table 2: Results of the evidence computed via TMCMC for each choice of model for $f_{\boldsymbol{a}}$.

To create the UM based on the information from the Bayesian model updating results, the following procedure is undertaken: For the aleatory space, the histograms of the distribution parameters of the given $f_{\boldsymbol{a}}$ are obtained from $P(\boldsymbol{\Theta}|\boldsymbol{D}, M)$. These histograms are converted into probability distribution functions using Kernel density estimation with a Gaussian kernel [25] and are normalised such that the distribution peak equals to 1. An illustration is provided using the distribution parameters for $f_{\boldsymbol{a}}^1$ as an example in Figure 2. From these results, the posterior distributions are interpreted as Fuzzy sets where different levels of confidence $L_c \in [0, 1]$ would yield intervals of varying width [26]. Here, intervals at $L_c = 0.5$ level of confidence are considered for both the $f_{\boldsymbol{a}}^1$ and $f_{\boldsymbol{a}}^5$ distribution parameters. The resulting intervals obtained would serve as shape parameter inputs of the respective aleatory model $f_{\boldsymbol{a}}$. This yields the Probability-boxes (i.e. P-boxes) [27, 28] of $f_{\boldsymbol{a}}^1$ and $f_{\boldsymbol{a}}^5$ which are illustrated in Figure 3.

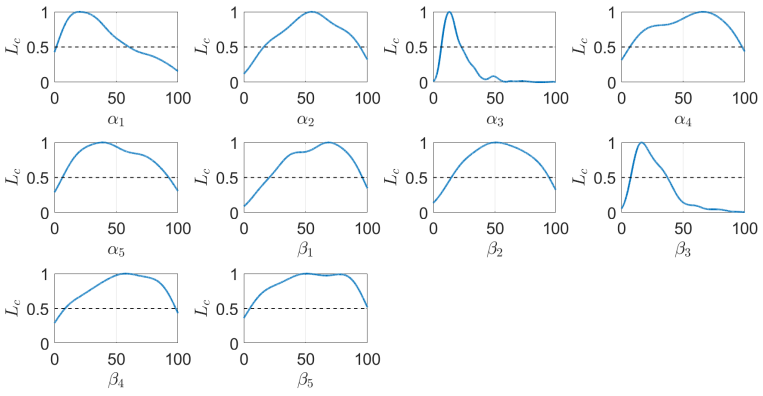

Figure 2: Illustration of the resulting distribution functions to the respective shape parameters of the joint Beta distribution (i.e. $f_{\boldsymbol{a}}^1$) obtained via Kernel density estimates.

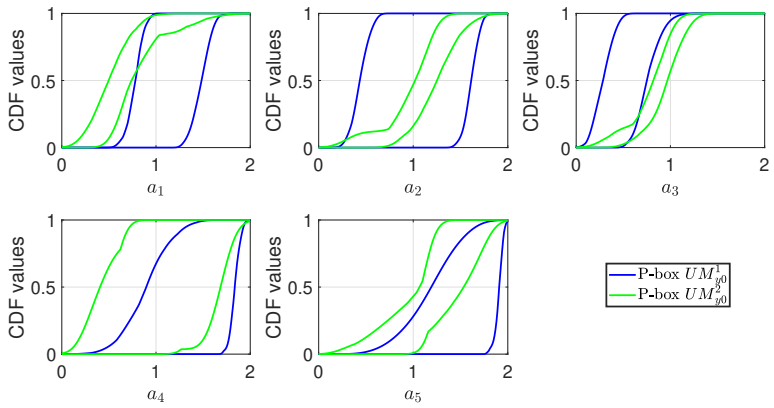

Figure 3: P-box for a_1 to a_5 obtained from the respective UMs.

To define the epistemic space, the same procedure is done on the resulting histograms of e_1 to e_4 obtained through $P(\boldsymbol{\Theta}|\boldsymbol{D}, M)$ given the respective $f_{\boldsymbol{a}}$. These histograms are illustrated in Figure 4 for the respective set-up. However, $L_c = 0.05$ level of confidence is considered in the case of $f^1_{\boldsymbol{a}}$ while $L_c = 0.025$ level of confidence is considered in the case of $f^5_{\boldsymbol{a}}$. The resulting intervals constitute the updated hyper-rectangular set E defined by each of the resulting UM to which numerical results are presented in Table 6. Let the UM determined from $P(\boldsymbol{\Theta}|\boldsymbol{D}, M)$ given $f^1_{\boldsymbol{a}}$ be denoted as UM^1_{y0}, while that given $f^5_{\boldsymbol{a}}$ be denoted as UM^2_{y0}.

To verify the calibration results, N_s samples are generated from the hyper-rectangle defined by the bounds of the respective distribution parameters of $f_{\boldsymbol{a}}$ and e_{i_e} according to the respective UMs. For each sample realization from this hyper-rectangle, 100 model outputs of $\hat{y}$ is obtained for $t \in [0, 5]$ s. This is done by generating 100 realizations of $\boldsymbol{a}$ from $f_{\boldsymbol{a}}$, given the distribution parameters from the hyper-rectangle sample, whilst keeping $\boldsymbol{e}$ fixed. This yields a $N_t \times 100 \times N_s$ array of data output of $\hat{y}$ for each UM whose results are plotted in Figure 5. From the figure, it can be observed that the model output bands of UM^1_{y0} (in blue) and UM^2_{y0} (in green) generally encompasses D_1 (in red) which indicates that the model calibration procedure, via Bayesian model updating,

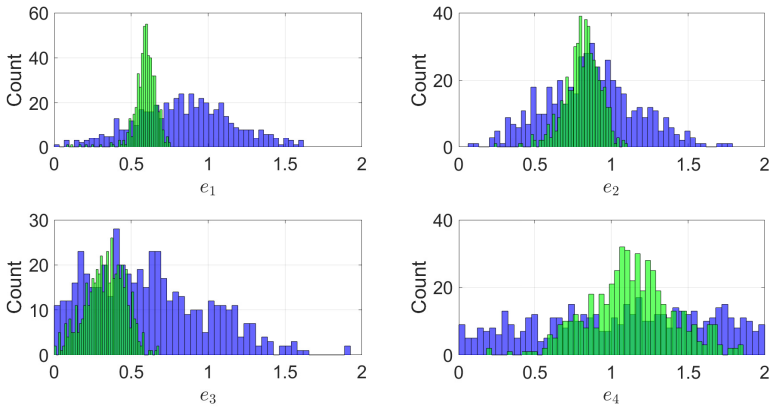

Figure 4: Histograms for e_1 to e_4 obtained from $P(\boldsymbol{\Theta}|\boldsymbol{D}, M)$ given $f_{\boldsymbol{a}}^1$ (in blue) and $f_{\boldsymbol{a}}^5$ (in green).

was done satisfactorily.

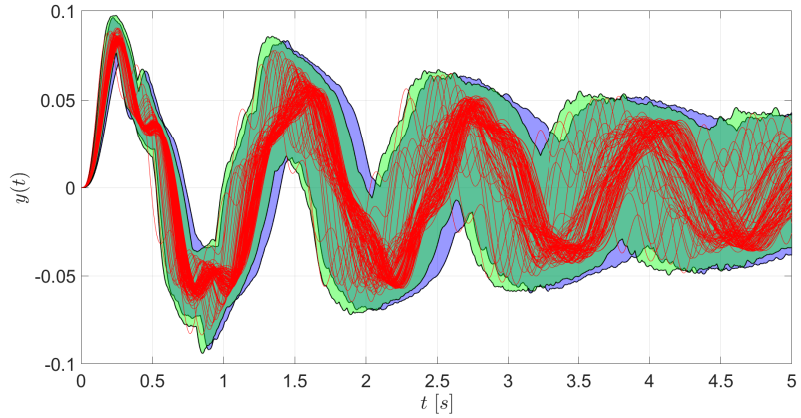

Figure 5: Output band from $\hat{y}$ according to UM_{y0}^1 (in blue) and UM_{y0}^2 (in green) along with the data sequence D_1 (in red) after calibration.

To further substantiate this, P-boxes of the calibrated model output of each UM are constructed at $t = \{0.5, 1.0, 2.0, 3.0, 4.0, 5.0\}$ s. Each P-box describes the extreme bounds of the distribution of the N_s ECDFs whereby each ECDF comprises of the 100 model output values at t. Figure 6 presents the resulting P-boxes from UM_{y0}^1 (in blue) and UM_{y0}^2 (in green) at each chosen t. From the

plots, it can be seen that the ECDF of D_1 (in red) at any given t is generally enclosed within the P-boxes. Furthermore, it can be observed that the shape of both P-boxes generally follow the shape profile of the ECDF of D_1 which indicates a good degree of fit by both UMs.

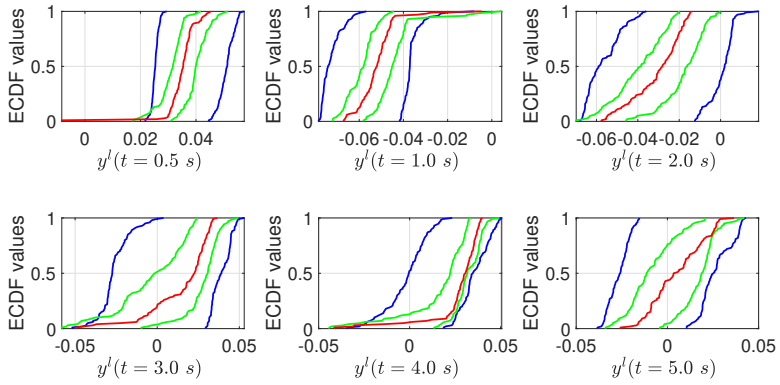

Figure 6: P-boxes of the model output from $\hat{y}$ obtained from UM^1_{y0} (in blue) and UM^2_{y0} (in green) at various time slices $t = \{0.5, 1.0, 2.0, 3.0, 4.0, 5.0\}\ s$. The red ECDF denotes the distribution of the data D_1.

2.5. Discussion

Figure 3 shows that the P-boxes obtained for a_1 to a_5 according to UM^1_{y0} is generally wider than those obtained using the second approach UM^2_{y0}. This indicates a higher degree of uncertainty on the true distribution of all a_{i_a} by UM^1_{y0} which makes it less informative in identifying the true $f_{\boldsymbol{a}}$ compared to UM^2_{y0}. In addition, the P-boxes for a_2 to a_5 obtained by UM^2_{y0} are generally enclosed within those of UM^1_{y0} which suggests that the true CDF defined by $f_{\boldsymbol{a}}$ could lie within the P-box defined by UM^2_{y0}.

The intervals obtained from the posterior distributions for e_1 to e_4 through the model based on Beta distributions is much wider compared to the UQ model obtained through the SRV based approach as shown in Figure 4. This further highlights the non-informative nature of UM^1_{y0}, especially for the case of e_3 and e_4. In addition, the posteriors obtained through the SRV based approach

show a much greater degree of update from the Uniform prior and that it is able to identify the epistemic parameters much more effectively as the peaks are more pronounced. This leads to the uncertainty bounds of $\boldsymbol{e}$ according to UM^2_{y0} being significantly narrower such that they are generally enclosed within that according to UM^1_{y0} as seen from Table 6.

Figure 5 shows that the output bands of $\hat{y}$ obtained from both UMs follow the trend defined by D_1. However, from Figure 6, it is observed that the P-boxes obtained by UM^2_{y0} have much tighter bounds compared to UM^1_{y0} whilst still enclosing the ECDF of D_1. This is attributed to the P-box of the $\boldsymbol{a}$, and the bounds on $\boldsymbol{e}$ being narrower for UM^2_{y0} than UM^1_{y0} which resulted in the former yielding a significantly better degree of fit over D_1 than the latter. From the results, it can be concluded that UM^1_{y0} is much more conservative compared to UM^2_{y0} in modelling the uncertainty of $f_{\boldsymbol{a}}$ and $\boldsymbol{e}$.

3. Task B: Uncertainty Reduction

The objective of this task is to identify the epistemic parameters which have more predictive capability and improve the UM. This is achieved by performing a sensitivity analysis for the epistemic model parameters and the subsequent refinement of the epistemic space.

3.1. Sensitivity Analysis

In this analysis, the epistemic uncertainties are ranked according to their ability to improve the predictive ability of the computational model of the subsystem. This predictive ability is quantified through the volume metric Ω defined as:

$$\Omega = \sum_{i_t=1}^{n_t} \rho_{i_t} \cdot \Delta_{i_t} \tag{12}$$

ρ_{i_t} is the area of the P-box at time-slice i_t, Δ_{i_t} is the time-step between time-slice $i_t - 1$ and i_t, and n_t is the total of time-slices used for the computation.

For the computation, we consider the 6 time-slices which were used for the illustration of the P-boxes in Figure 6.

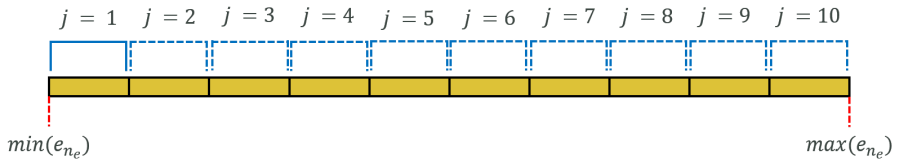

Figure 7: Illustration to the approach in identifying the maximum reduction of Ω from the pinching of e_{i_e}.

To rank the the epistemic parameters according to their respective sensitivity, an adaptive pinching method [29] is proposed to provide a non-empirical approach to determine the pinched bounds of a chosen epistemic parameter which yields the greatest reduction in the value of Ω. The procedure is as follows: For a given i_e, the uncertainty space of e_{i_e} is reduced by 90 %. This is done whilst keeping the uncertainty space of the remaining 3 epistemic model parameters untouched. For a given e_{i_e}, its bounds would first be divided into 10 equally-spaced units. Next, at iteration $j = 1$, a segment of bin length of 1 unit will be used to isolate the region of the epistemic space defined by the lower and upper bounds of the first bin. This isolated region serves as the reduced (or "pinched") space. From there, the corresponding realizations of $\{\boldsymbol{a}, \boldsymbol{e}\}$ from the UM, whose e_{i_e} value falls outside the bounds of the reduced epistemic space, is discarded. When this is done, the reduced volume Ω_p is computed again via Eq. (12). After this is done, the segment shifts by 1 unit to the right as illustrated in Figure 7 and this initiates iteration $j = 2$ where the above procedure is repeated all the way to iteration $j = 10$. This approach is done for e_1 to e_4. As an illustrative example, the results of the reduced volume Ω_p for the respective iteration j for each e_{i_e} according to UM^1_{y0} are presented in Figure 8. From the figure, the minimum value of Ω_p for each e_{i_e} is determined and the sensitivity index is computed:

$$S = 1 - \frac{\Omega_p}{\Omega_0} \tag{13}$$

where Ω_0 is the initial volume before pinching. This sensitivity metric would then be used to rank e_1 to e_4 to which the results according to the respective UMs are shown in Table 3.

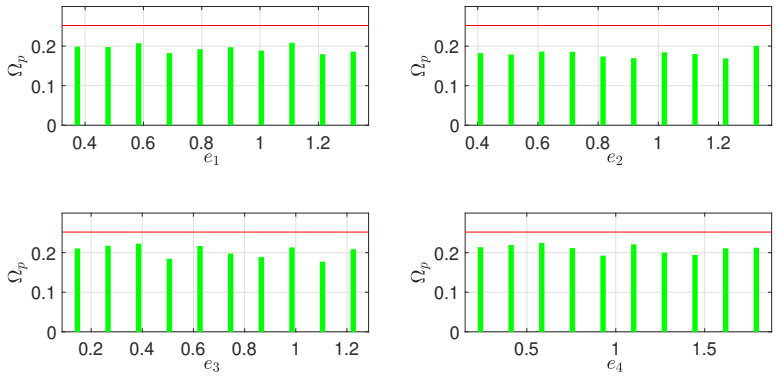

Figure 8: Results of Ω_p for different pinched intervals for e_1 to e_4 according to UM^1_{y0}. The red line denotes the initial volume $\Omega = 0.2521$, while the green bars represent the resulting Ω_p.

Rank	Uncertainty model UM^1_{y0}			Uncertainty model UM^2_{y0}		
	Parameter	Pinched bounds	S	Parameter	Pinched bounds	S
1	e_2	$[1.1729, 1.2748]$	0.3300	e_2	$[0.8677, 0.9117]$	0.4542
2	e_3	$[1.0449, 1.1647]$	0.2972	e_3	$[0.3595, 0.4072]$	0.3240
3	e_1	$[1.1619, 1.2670]$	0.2882	e_1	$[0.5715, 0.5987]$	0.2654
4	e_4	$[0.8425, 1.0148]$	0.2272	e_4	$[0.8242, 0.9320]$	0.2556

Table 3: The ranking order of the epistemic model parameters based on their respective sensitivity index for the respective UMs.

Based on the results provided in Table 3, it is observed that e_4 is ranked the lowest in sensitivity according to both UMs. This implies that it is impossible to improve the knowledge on e_4 with the available model and data, making impossible to extract or infer information on its true value, thus contributing the highest degree of non reducible epistemic uncertainty in the calibration of the UM. For this reason, the first refinement request to the challenge host was

made for the lower bound of e_4, given the heavier left tail as seen from its histogram obtained via the Distribution-free approach (in green) in Figure 4. Following this, a second round of sensitivity analysis was performed following the approach outlined above and accounting for the given refined bounds of e_4 and the results are presented in Table 4. From the results, e_3 is consistently ranked within the bottom 2 according to both UMs which suggests that e_3 is the least informative parameter after e_4. Hence, the second refinement request was made for the lower bound of e_3 given the lack of such information according to both UMs as illustrated in Figure 4. The resulting epistemic space, with the refined e_3 and e_4 bounds, constitutes the hyper-rectangle epistemic space denoted as E_1.

Rank	Uncertainty model UM^1_{y0}			Uncertainty model UM^2_{y0}		
	Parameter	Pinched bounds	S	Parameter	Pinched bounds	S
1	e_4	$[1.0224, 1.0575]$	0.5200	e_2	$[0.6036, 0.6476]$	0.5999
2	e_2	$[0.4652, 0.5625]$	0.5114	e_4	$[1.1276, 1.1632]$	0.4128
3	e_1	$[0.6333, 0.7367]$	0.5056	e_3	$[0.2166, 0.2642]$	0.4024
4	e_3	$[0.9227, 1.0393]$	0.4282	e_1	$[0.5715, 0.5988]$	0.3821

Table 4: The ranking order of the epistemic model parameters based on their respective sensitivity index for the respective UMs accounting for the refined bounds for e_4.

3.2. Updated Uncertainty Models

A second round of Bayesian model updating is performed with the bounds of the Uniform priors for the respective epistemic parameters defined by the hyper-rectangle E_1. The approach follows that outlined in Sections 2.3.1 and 2.3.2 from which UM^1_{y1} and UM^2_{y1} are obtained respectively. The corresponding numerical results of the updated bounds for each e_{i_e} according to UM^1_{y1} and UM^2_{y1} are presented in Table 6.

A sensitivity analysis was done again following the methodology presented in Section 3.1 and the results are summarized in Table 5. From the table, it can be observed that the sensitivity ranking of each e_{i_e} is the same as that in Table 3 for the respective UMs.

Rank	Uncertainty model UM^1_{y1}			Uncertainty model UM^2_{y1}		
	Parameter	Pinched bounds	S	Parameter	Pinched bounds	S
1	e_2	$[0.4447, 0.5340]$	0.4095	e_2	$[0.9514, 0.9726]$	0.3610
2	e_1	$[0.9500, 1.0413]$	0.2960	e_3	$[0.5675, 0.6173]$	0.3245
3	e_4	$[0.9274, 0.9585]$	0.2385	e_1	$[0.5550, 0.5725]$	0.2981
4	e_3	$[0.3333, 0.3590]$	0.2379	e_4	$[1.0225, 1.0584]$	0.2782

Table 5: The ranking order of the epistemic model parameters based on their respective sensitivity index for the respective UMs accounting for the refined space E_1.

3.3. Results and Discussion

The resulting model output of the response plot according to UM^1_{y1} and UM^2_{y1} are illustrated in Figure 9. From the figure, it can be observed that the response plots according to both UMs are well-fitted against D_1. However, such fitting is significantly tighter for the case of UM^2_{y1} as seen in Figure 9 and this is supported by Figure 10 where it can also be seen that the resulting P-boxes of the response plot across all chosen time-slices t are significantly narrower compared to UM^1_{y1} whilst enclosing the ECDF for D_1. This observation is consistent to that discussed in Section 2.5 and concludes that the response plot according to UM^2_{y1} is more representative of D_1.

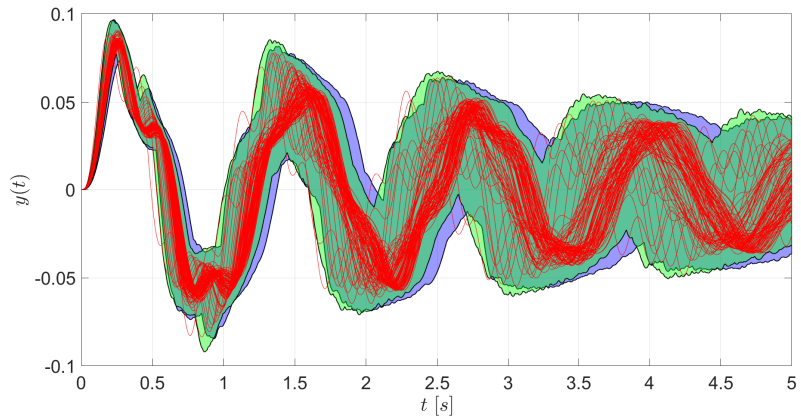

Figure 9: Output band from $\hat{y}$ according to UM^1_{y1} (in blue) and UM^2_{y1} (in green) along with the data sequence D_1 (in red) after calibration.

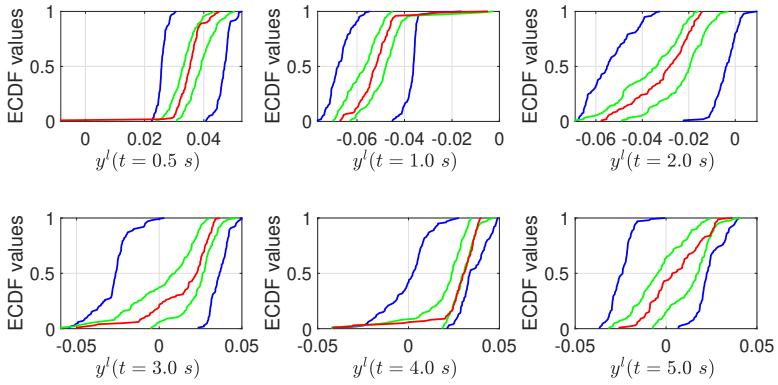

Figure 10: P-boxes of the model output from $\hat{y}$ obtained from UM^1_{y1} (in blue) and UM^2_{y1} (in green) at various time slices $t = \{0.5, 1.0, 2.0, 3.0, 4.0, 5.0\}\ s$. The red ECDF denotes the distribution of the data D_1.

Figure 11 illustrates the resulting P-boxes quantifying the uncertainty over the marginal distributions of $f_{\boldsymbol{a}}$ by the respective UMs. From the figure, it can be seen that the P-boxes according to UM^2_{y1} is significantly narrower and enclosed within that according to UM^1_{y1} which verifies that the true marginal distributions of $f_{\boldsymbol{a}}$ could lie within the P-boxes defined by UM^2_{y1}. The results by UM^1_{y1} once again highlights its conservative nature in its uncertainty over $f_{\boldsymbol{a}}$ compared to UM^2_{y1}.

Finally, results from Table 6 show that the uncertainty bounds over $\boldsymbol{e}$ according to UM^2_{y1} is significantly narrower and generally enclosed within that according to UM^1_{y1}. This result is supported by Figure 12 where it can be seen that the resulting histograms of the epistemic parameters obtained from $P(\boldsymbol{\Theta}|\boldsymbol{D}, M)$ given $f^5_{\boldsymbol{a}}$ are consistently narrower than that obtained from $P(\boldsymbol{\Theta}|\boldsymbol{D}, M)$ given $f^1_{\boldsymbol{a}}$. This verifies the results obtained by UM^2_{y1} which further highlights its informative nature over UM^1_{y1}. For this reason, UM^2_{y1} is chosen to address the subsequent tasks presented in this challenge.

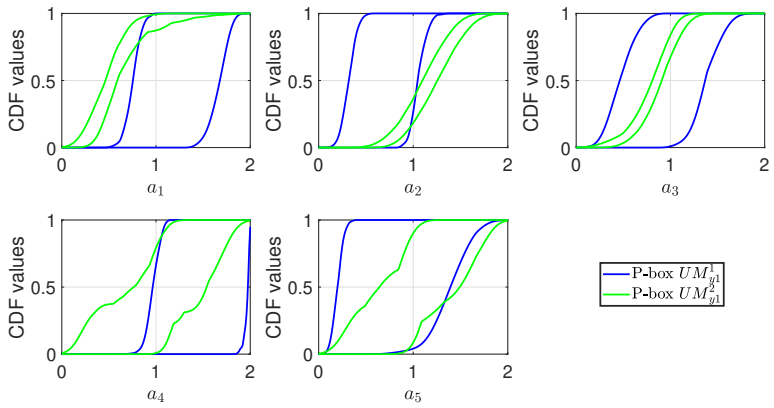

Figure 11: P-box for a_1 to a_5 obtained from the respective UMs.

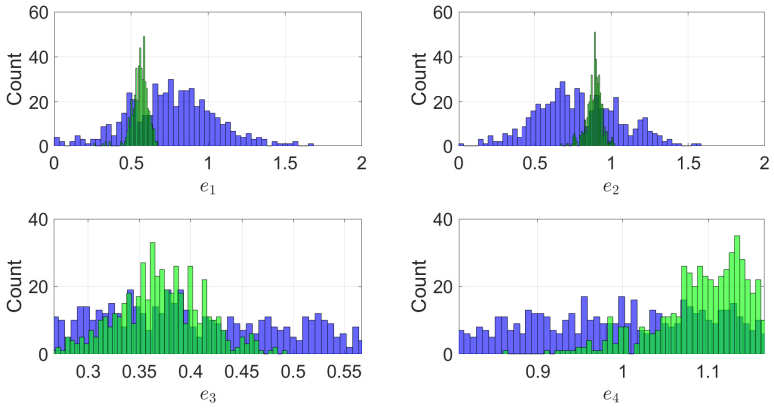

Figure 12: Histograms for e_1 to e_4 obtained from $P(\boldsymbol{\Theta}|\boldsymbol{D}, M)$ given $f^1_{\boldsymbol{a}}$ (in blue) and $f^5_{\boldsymbol{a}}$ (in green).

Uncertainty model	e_1	e_2	e_3	e_4
UM^1_{y0}	$[0.3182, 1.3787]$	$[0.3574, 1.3771]$	$[0.0827, 1.2870]$	$[0.1486, 1.8828]$
UM^2_{y0}	$[0.4351, 0.7082]$	$[0.5583, 1.0000]$	$[0.0721, 0.5511]$	$[0.6066, 1.6893]$
UM^1_{y1}	$[0.3097, 1.2306]$	$[0.3522, 1.2487]$	$[0.2819, 0.5400]$	$[0.8337, 1.1461]$
UM^2_{y1}	$[0.4674, 0.6433]$	$[0.7607, 0.9736]$	$[0.2865, 0.4583]$	$[0.9627, 1.1664]$

Table 6: Updated epistemic space E for e_1 to e_4 according to the respective UMs.

4. Task C: Reliability Analysis of Baseline Design

The objective of this task is to perform a reliability analysis on $\boldsymbol{\theta}_{base}$ according to UM^2_{y1} with respect to the individual requirements g_{i_g}, for $i_g = 1, 2, 3$. Requirements g_2 and g_3 are defined respectively as [3]:

$$g_2(\boldsymbol{a}, \boldsymbol{e}, \boldsymbol{\theta}) = \max_{t \in [2.5,5]s} |z_1(\boldsymbol{a}, \boldsymbol{e}, \boldsymbol{\theta}, t)| - 0.02 \tag{14}$$

$$g_3(\boldsymbol{a}, \boldsymbol{e}, \boldsymbol{\theta}) = \max_{t \in [0,5]s} |z_2(\boldsymbol{a}, \boldsymbol{e}, \boldsymbol{\theta}, t)| - 4 \tag{15}$$

where z_1 and z_2 are the time-dependent response output of the integrated system associated with the given $\boldsymbol{\theta}$. From there, the worst-case performance function w is defined:

$$w(\boldsymbol{a}, \boldsymbol{e}, \boldsymbol{\theta}) = \max_{i_g = 1,2,3} \; g_{i_g}(\boldsymbol{a}, \boldsymbol{e}, \boldsymbol{\theta}) \tag{16}$$

The system is defined to be system compliant for requirement i_g when $g_{i_g} < 0$ whereby: $g_1 < 0$ for the system to be stable; $g_2 < 0$ for the settling time of z_1 to be sufficiently fast; and $g_3 < 0$ for the energy consumption to be acceptable [3]. Conversely, requirement i_g is not satisfied when $g_{i_g} \geq 0$. Thus, for a fixed set of values of $\boldsymbol{\theta}$ and $\boldsymbol{e}$, the set of $\boldsymbol{a}$ points where $g_{i_g} < 0$ is regarded as the "safe" domain, while the complement set is the "failure" domain. From this, the imprecise failure probability R_{i_g} given requirement g_{i_g} is defined:

$$R_{i_g}(\boldsymbol{\theta}) = \left[\min_{\boldsymbol{e} \in E} \mathbb{P}(g_{i_g} \geq 0), \max_{\boldsymbol{e} \in E} \mathbb{P}(g_{i_g} \geq 0) \right] \tag{17}$$

through which the imprecise worst-case failure probability R is defined:

$$R(\boldsymbol{\theta}) = \left[\min_{\boldsymbol{e} \in E} \mathbb{P}(w \geq 0), \max_{\boldsymbol{e} \in E} \mathbb{P}(w \geq 0) \right] \tag{18}$$

and finally, the severity of each requirement violation s_{i_g} is defined:

$$s_{i_g}(\boldsymbol{\theta}) = \max_{\boldsymbol{e} \in E} \mathbb{E}\left[g_{i_g} | g_{i_g} \geq 0 \right] \cdot \mathbb{P}(g_{i_g} \geq 0) \tag{19}$$

where $\mathbb{P}(\bullet)$ is the probability operator, and $\mathbb{E}[\bullet|\bullet]$ is the conditional expectation.

4.1. Failure Probability and Severity Computation

In this work, the computation and analysis of the reliability metrics R_{i_g}, R, and s_{i_g} are done through Probability Bounds Analysis (PBA) with P-boxes [30].

To set up the P-box for each g_{i_g}, a Double-Loop Monte Carlo [31, 32] approach is used to generate $N_a \times N_e$ realizations of g_{i_g} from inputs $\{\boldsymbol{a}, \boldsymbol{e}\}$ defined by UM^2_{y1}. The outer-loop accounts for each of the N_e realizations of $\boldsymbol{e}$ obtained from the hyper-rectangle defined by E according to the UM, while the inner-loop accounts for the N_a realizations of $\boldsymbol{a}$ from $f_{\boldsymbol{a}}$. To ensure that the different failure domains are well-explored, especially small failure regions, and that the epistemic uncertainties well-represented, we set $N_a = 10000$ and $N_e = 500$. From there, the P-box is constructed from the bounds of the distribution of N_e ECDFs, each ECDF comprising of N_a values of g_{i_g}. As an illustration, the resulting P-boxes for g_1 to g_3 are illustrated in Figure 39.

From the P-box of a given g_{i_g}, its values computed at $g_{i_g} = 0$ has a lower and upper bound value denoted as $\underline{P}_{i_g}$ and $\overline{P}_{i_g}$ respectively. Following which, $R_{i_g}(\boldsymbol{\theta}_{base})$ can be approximated according to:

$$R_{i_g}(\boldsymbol{\theta}_{base}) \approx \left[1 - \overline{P}_{i_g}, 1 - \underline{P}_{i_g}\right] \tag{20}$$

To approximate $R(\boldsymbol{\theta}_{base})$, the $N_a \times N_e$ matrix of w is constructed by taking the element-wise maximum value between g_1, g_2, and g_3 as suggested in Eq. (16). A P-box is then constructed for w in similar fashion as g_{i_g} from which the resulting lower and upper bound values, denoted as $\underline{W}$ and $\overline{W}$ respectively, are obtained at $w = 0$. $R(\boldsymbol{\theta}_{base})$ can then be approximated according to:

$$R(\boldsymbol{\theta}_{base}) \approx \left[1 - \overline{W}, 1 - \underline{W}\right] \tag{21}$$

To approximate s_{i_g}, the numerical value of $\mathbb{E}(g_{i_g}|g_{i_g} \geq 0)$ needs to be approximated first. This can be done as follows: Considering a realization of the ECDF for g_{i_g} for a given $\boldsymbol{e}$, a numerical integration is performed to obtain the area of

the region above the ECDF plot between $g_{i_g} = 0$ to its maximum value $g_{i_g}^{max}$ for which the ECDF is defined. This is done according to:

$$\mathbb{E}(g_{i_g}|g_{i_g} \geq 0) \approx \sum_{k=1}^{N_a} \mathbf{1}_{(g_{i_g})_k \geq 0} \cdot (g_{i_g})_k \cdot \mathbb{P}(g_{i_g} \geq (g_{i_g})_k) \tag{22}$$

where $\mathbf{1}_{(g_{i_g})_k \geq 0}$ is the indicator function which gives the value 1 when $(g_{i_g})_k \geq 0$ and 0 otherwise. This value of $\mathbb{E}(g_{i_g}|g_{i_g} \geq 0)$ is then multiplied by $\mathbb{P}(g_{i_g} \geq 0)$ determined from its ECDF which gives the nominal severity index associated with the given $\boldsymbol{e}$. From which, the actual severity index s_{i_g} is computed according to Eq. (19).

The numerical results to the aforementioned reliability metrics are presented in Table 14 where it is observed that the failure probability with the highest upper-bound value is R_2 and, thus, contributes the most of the worst-case failure probability R. In addition, it is noted that the severity s_2 is the highest among all s_{i_g} which indicates that the failure $g_2 \geq 0$ is classified as a high-probability event with a large impact on the system relative to the failures $g_1 \geq 0$ and $g_3 \geq 0$. Further investigations to this are done in Section 4.3.

4.2. Sensitivity Analysis

In this analysis, the epistemic uncertainties are ranked according to the contraction of $R(\boldsymbol{\theta}_{base})$ resulting from their reduction. This will be done via the adaptive pinching [29] approach outlined in Section 3. The results of the reduced interval $R(\boldsymbol{\theta}_{base})$ for the respective iteration j for each e_{i_e} are illustrated in Figure 13. From the figure, the maximum reduction of $R(\boldsymbol{\theta}_{base})$ for each e_{i_e} is determined. Such information would then be used to rank e_1 to e_4 as shown in Table 7.

4.3. Identifying Different Transitions to Failure

From the N_a-by-N_e matrix of all g_{i_g}, the realizations of $\{\boldsymbol{a}, \boldsymbol{e}\}$ are classified into 7 distinct categories of failure: 1) $g_1 \geq 0$; 2) $g_2 \geq 0$; 3) $g_3 \geq 0$; 4) $g_1, g_2 \geq 0$;

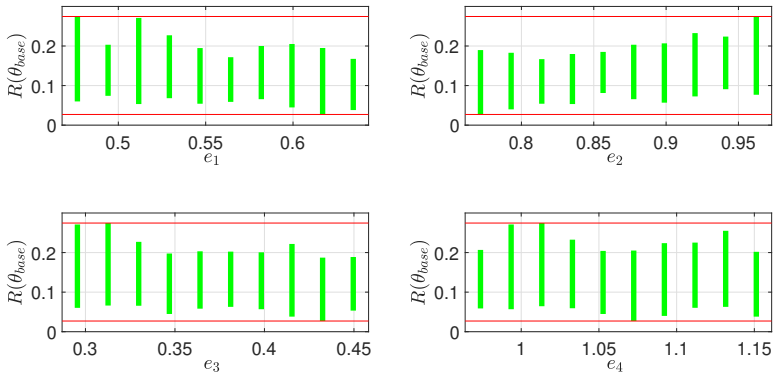

Figure 13: Results of the reduced $R(\boldsymbol{\theta}_{base})$ bounds for different pinched intervals for e_1 to e_4. The red lines represent the initial bounds of $[0.0294, 0.2721]$, while the green bars represent the reduced bounds.

Rank	Parameter	Pinched bounds	$R(\boldsymbol{\theta}_{base})$	
			Before pinching	After pinching
1	e_2	$[0.8458, 0.8670]$	$[0.0294, 0.2721]$	$[0.0812, 0.1851]$
2	e_1	$[0.5555, 0.5731]$	$[0.0294, 0.2721]$	$[0.0588, 0.1717]$
3	e_3	$[0.4411, 0.4582]$	$[0.0294, 0.2721]$	$[0.0533, 0.1888]$
4	e_4	$[0.9639, 0.9836]$	$[0.0294, 0.2721]$	$[0.0588, 0.2067]$

Table 7: The ranking order of the epistemic model parameters based on the maximum possible reduction in $R(\boldsymbol{\theta}_{base})$ interval according to UM^2_{y1}.

5) $g_1, g_3 \geq 0$; 6) $g_2, g_3 \geq 0$; and 7) $g_1, g_2, g_3 \geq 0$. The resulting statistics summarising the number of $\{\boldsymbol{a}, \boldsymbol{e}\}$ realizations in each failure category is presented in Table 12.

From the table, it can be seen that the most common failure type for $\boldsymbol{\theta}_{base}$ according to UM^2_{y1} is $g_2 \geq 0$ while the least likely failure type would be $g_1, g_2, g_3 \geq 0$. No failure of type $g_1, g_3 \geq 0$ has occurred. To provide a quantitative understanding of the characteristic of each failure and its severity, the response curves $z_1(t)$ and $z_2(t)$ are plotted for 25 representative sample sets of $\{\boldsymbol{a}, \boldsymbol{e}\}$ in each failure category and these are illustrated in Figures 14 and 15.

From the figures, it can be seen that the failure type of the worst severity is that of $g_1, g_2, g_3 \geq 0$ where it can be seen that the plots for $z_1(t)$ and $z_2(t)$ show an unstable behaviour with an increasing amplitude as time t increases. This results in the largest degree of deviation from the safety limits which causes the aforementioned failure type to contribute the most towards the severity values of s_1, s_2, and s_3. In addition, it can also be observed that the $z_2(t)$ plot for failure type $g_1, g_2 \geq 0$ also exhibit an increasing amplitude with time, although still within the safety limits and not as pronounced as that for $g_1, g_2, g_3 \geq 0$. Such unstable behaviour is due to the common failure of g_1 which concerns the stability of the system's behaviour. Hence, it is important to identify a new design point $\boldsymbol{\theta}_{new}$ in Section 5 such that the likelihood of occurrence of failure types $g_1, g_2 \geq 0$ and $g_1, g_2, g_3 \geq 0$ is as close to 0 as possible.

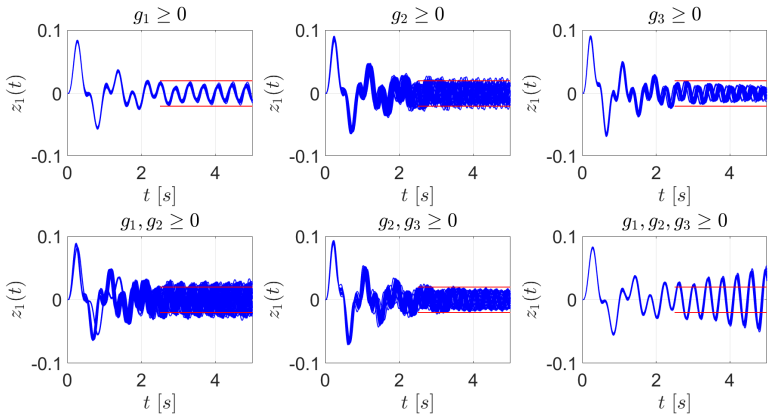

Figure 14: Response plot of $z_1(t)$ corresponding to 25 representative realizations of $\{\boldsymbol{a}, \boldsymbol{e}\}$ for the each failure type. The red lines denote the safety limits.

To identify the representative realizations of $\delta \in A \times E$ having a comparatively large likelihood near the failure domain, the methodology is as follows: For each i_g, realizations of $\{\boldsymbol{a}, \boldsymbol{e}\}$ from UM^2_{y1} corresponding to the top 500 numerically least negative matrix elements of g_{i_g} are identified. These realizations will be classified as those "near" the failure domain. Following this, the likelihood values of the identified $\{\boldsymbol{a}, \boldsymbol{e}\}$ are computed by calculating the PDF value

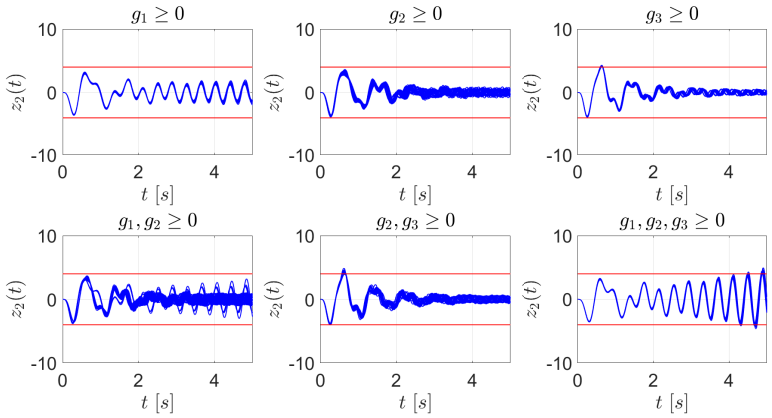

Figure 15: Response plot of $z_2(t)$ corresponding to 25 representative realizations of $\{\boldsymbol{a}, \boldsymbol{e}\}$ for the each failure type. The red lines denote the safety limits.

of the SDF for the corresponding realization of $\boldsymbol{a}$. This generates 500 values of likelihood values from which the sample set $\{\boldsymbol{a}, \boldsymbol{e}\}$ having the top 5-percentile likelihood values are identified. Let these sample sets be denoted as $\{\boldsymbol{a}, \boldsymbol{e}\}_{nf}^{g_{i_g}}$. This yields 25 sets of $\{\boldsymbol{a}, \boldsymbol{e}\}_{nf}^{g_{i_g}}$ remaining which will constitutes the realizations with comparatively large likelihood near the failure domain of g_{i_g}. This procedure is implemented for g_1 to g_3 and the resulting 25 sets of $\{\boldsymbol{a}, \boldsymbol{e}\}_{nf}^{g_{i_g}}$ identified for each g_{i_g} are presented as parallel plots in Figure 16 and whose corresponding response plots of $z_1(t)$ and $z_2(t)$ are illustrated in Figure 17.

From Figure 17, the following observations are made: Near the failure domain of g_1, the response plots $z_1(t)$ and $z_2(t)$ are all well-within the safe limits which indicates that the sample sets $\{\boldsymbol{a}, \boldsymbol{e}\}_{nf}^{g_1}$ all lie within the safe domain near the boundary of $g_1 \geq 0$ domain. Near the failure domain of g_2, the response plots $z_1(t)$ and $z_2(t)$ are all within safe limits as well. Based on the fast-decaying characteristics of the response plot for $z_2(t)$ and the stable behaviour of the plot for $z_1(t)$, it can be inferred that the requirement of g_1 is satisfied as well. However, it is noted that at approximately $t = \{2.7, 2.9\}$ s, the plots are extremely close to the lower and upper boundaries respectively which verifies that the realizations of $\{\boldsymbol{a}, \boldsymbol{e}\}_{nf}^{g_2}$ are in the safe domain near the boundary of $g_2 \geq 0$. Finally,

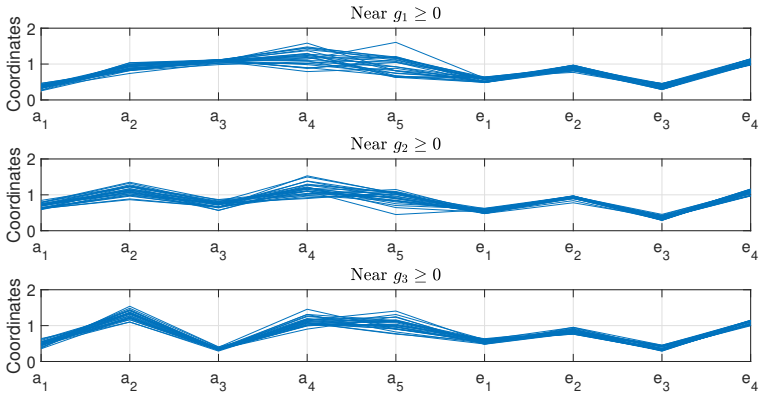

Figure 16: Parallel plots of $\{\boldsymbol{a}, \boldsymbol{e}\}_{nf}^{g_1}$, $\{\boldsymbol{a}, \boldsymbol{e}\}_{nf}^{g_2}$, and $\{\boldsymbol{a}, \boldsymbol{e}\}_{nf}^{g_3}$ from UM_{y1}^2.

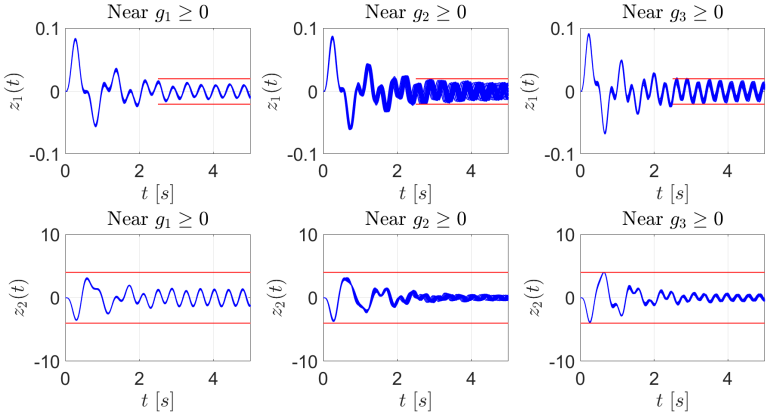

Figure 17: Response plot of $z_1(t)$ and $z_2(t)$ of $\{\boldsymbol{a}, \boldsymbol{e}\}_{nf}^{g_1}$, $\{\boldsymbol{a}, \boldsymbol{e}\}_{nf}^{g_2}$, and $\{\boldsymbol{a}, \boldsymbol{e}\}_{nf}^{g_3}$ from UM_{y1}^2 near the respective failure domains. The red lines denote the safety limits.

near the failure domain of g_3, it can be observed from the plot of $z_1(t)$ that the curves exceeded the safety limits at approximately $t = \{2.6, 2.9, 3.0\}$ s which indicates the failure of g_2. On the other hand, the plot of $z_2(t)$ are all within the safety limits although it can be observed that the plots at approximately $t = 0.6$ s are extremely close to the upper boundary. Due to the stable behavior of the plots of $z_1(t)$ and $z_2(t)$, it can be inferred that the requirement of g_1 is satisfied. Hence, it can be concluded that the realizations of $\{\boldsymbol{a}, \boldsymbol{e}\}_{nf}^{g_3}$ lie within

the domain of $g_2 \geq 0$ near the boundary of $g_2, g_3 \geq 0$.

5. Task D: Reliability-based Design Identification

The objective of this task is to identify a new design point $\boldsymbol{\theta}_{new}$ such that the likelihood of failure types $g_1, g_2 \geq 0$ and $g_1, g_2, g_3 \geq 0$ occurring is reduced to as close to 0 as possible given that such failures are responsible for the unstable behaviour of the system. To achieve this, $\boldsymbol{\theta}_{new}$ has to be optimised such that it satisfies the following conditions:

1. Minimise the upper-bound of R;
2. Reduce the worst-case severity metric $\tilde{s}$ defined as [14]:

$$\tilde{s}(\boldsymbol{\theta}) = \max_{\boldsymbol{e} \in E} \mathbb{E}[w | w \geq 0] \cdot \mathbb{P}(w \geq 0) \tag{23}$$

To perform the optimisation procedure, the generalised Non-intrusive Imprecise Stochastic Simulation (NISS) method is adopted to approximate a solution to $\boldsymbol{\theta}_{new}$ [33]. For the benefit of the readers, a description to the generalised NISS is provided in Section 5.1.

5.1. Generalised Non-intrusive Stochastic Simulation

The generalised NISS approach provides a surrogate model to compute R through the Random Sampling High-dimensional Model Representation (RS-HDMR) decomposition defined as:

$$\begin{aligned} R(\boldsymbol{e}, \theta) = R_0 + \sum_{i_e=1}^{4} R_{E_{i_e}}\left(e_{i_e}\right) + \sum_{1 \leq i_e < j_e \leq 4} R_{E_{i_e, j_e}}\left(\boldsymbol{e}_{i_e, j_e}\right) + \ldots \\ \sum_{i_\theta=1}^{9} R_{\tilde{\Theta}_{i_\theta}}\left(\theta_{i_\theta}\right) + \sum_{1 \leq i_\theta < j_\theta \leq 9} R_{\tilde{\Theta}_{i_\theta, j_\theta}}\left(\boldsymbol{\theta}_{i_\theta, j_\theta}\right) + \ldots \\ \sum_{1 \leq i_e \leq 4, 1 \leq i_\theta \leq 9} R_{E_{i_e} \tilde{\Theta}_{i_\theta}}\left(e_{i_e}, \theta_{i_\theta}\right) + \cdots + R_{E\tilde{\Theta}}(\boldsymbol{e}, \boldsymbol{\theta}) \end{aligned} \tag{24}$$

where the constant terms and the first 2 order component functions are defined respectively as:

$$R_0 = \mathbb{E}_{E\tilde{\Theta}}\left[R(\boldsymbol{e},\boldsymbol{\theta})\right]$$

$$R_{E_{i_e}}\left(e_{i_e}\right) = \mathbb{E}_{E_{-i_e}\tilde{\Theta}}\left[R(\boldsymbol{e},\boldsymbol{\theta})\right] - R_0$$

$$R_{E_{i_e,j_e}}\left(\boldsymbol{e}_{i_e,j_e}\right) = \mathbb{E}_{E_{-i_e,j_e}\tilde{\Theta}}\left[R(\boldsymbol{e},\boldsymbol{\theta})\right] - R_{E_{i_e}} - R_{E_{j_e}} - R_0$$

$$R_{\tilde{\Theta}_{i_\theta}}\left(\theta_{i_\theta}\right) = \mathbb{E}_{E\tilde{\Theta}_{-i_\theta}}\left[R(\boldsymbol{e},\boldsymbol{\theta})\right] - R_0$$

$$R_{\tilde{\Theta}_{i_e,j_e}}\left(\boldsymbol{\theta}_{i_\theta,j_\theta}\right) = \mathbb{E}_{E\tilde{\Theta}_{-i_e,j_e}}\left[R(\boldsymbol{e},\boldsymbol{\theta})\right] - R_{\tilde{\Theta}_{i_\theta}} - R_{\tilde{\Theta}_{j_\theta}} - R_0$$

$$R_{\tilde{\Theta}_{i_e,j_e}}\left(\boldsymbol{e}_{i_e,j_e}\right) = \mathbb{E}_{E_{-i_e},\tilde{\Theta}_{-i_\theta}}\left[R(\boldsymbol{e},\boldsymbol{\theta})\right] - R_{E_{i_e}} - R_{\tilde{\Theta}_{i_\theta}} - R_0$$

Here, $\mathbb{E}_{E\tilde{\Theta}}[\bullet]$ denotes the expectation operator as a function of $\boldsymbol{e}$ and $\boldsymbol{\theta}$, $\mathbb{E}_{E_{-i_e}\tilde{\Theta}}[\bullet]$ denotes the expectation operator as a function of $\boldsymbol{\theta}$ and the 3-dimensional vector $\boldsymbol{e}_{-i_e}$ which contains all elements of $\boldsymbol{e}$ except those of component e_{i_e}, $\mathbb{E}_{E\tilde{\Theta}_{-i_\theta}}[\bullet]$ denotes the expectation operator as a function of $\boldsymbol{e}$ and the $(n_\theta - 1)$-dimensional vector $\boldsymbol{\theta}_{-i_\theta}$ which contains all elements of $\boldsymbol{\theta}$ except those of component θ_{i_θ}, $\mathbb{E}_{E_{-i_e,j_e}\tilde{\Theta}}$ denotes the expectation operator as a function of $\boldsymbol{\theta}$ and all elements of $\boldsymbol{e}$ except e_{i_e,j_e}, $\mathbb{E}_{E\tilde{\Theta}_{-i_\theta,j_\theta}}$ denotes the expectation operator as a function of $\boldsymbol{e}$ and all elements of $\boldsymbol{\theta}$ except θ_{i_e,j_e}, and $\mathbb{E}_{E_{-i_e},\tilde{\Theta}_{-i_\theta}}$ denotes the expectation operator as a function of $\boldsymbol{e}_{-i_e}$ and $\boldsymbol{\theta}_{-i_\theta}$.

The above component functions are approximated numerically via Extended Monte Carlo Simulation (ECMS) [34, 35] of $N_{EMCS} = 50000$ realizations of the joint sample sets of $\{(\boldsymbol{a}_i, \boldsymbol{e}_i, \boldsymbol{\theta}_i)\}_{i=1,\ldots,N_{EMCS}}$. Realizations of $\boldsymbol{a}_i$ and $\boldsymbol{e}_i$ can be generated from UM^2_{y1} while those of $\boldsymbol{\theta}_i$ are generated from its target hyper-rectangular space $\tilde{\Theta}$ which will be discussed later in this section. From these realizations, a bootstrap scheme is implemented to compute the variance of each estimator, with the number of bootstrap replications set to be 20 in this work. This allows for the computation of the Sobol' sensitivity index [36, 37] for each of these component function as a by-product of the generalised NISS technique [33]. Given that the Sobol' indices measure the relative importance of each component function, the component functions with relatively small Sobol' index values can be neglected [36]. As such, based on initial analysis, it was found that we can just consider the first-order component functions in the optimisation

procedure. Readers can refer to [33] for more details to the generalised NISS technique.

It should be noted that the generalized NISS has been originally introduced for the efficient propagation of hybrid uncertainties by avoiding double loop Monte Carlo. A key aspect of generalized NISS is that it allows for an explicit formulation of the functional dependence of the probability of failure with respect to the epistemic parameters, one parameter at a time. In this work, the design variables are treated like epistemic parameters in the NISS framework, simplifying the optimization problem to a set of one-dimensional searches. However, no bounds are given to the design parameters limiting the optimization, thus the optimisation procedure is carried out for $\boldsymbol{\theta}_{new}$ in an iterative approach such that the first-order component function $R_{\tilde{\Theta}_{i_\theta}}$ is minimised which is assumed to also minimise the upper bound of R within the target design space. Based on this assumption, the procedure is undertaken as follows:

1. At iteration $j = 1$, the initial 9-dimensional hyper-rectangle of the design space $\tilde{\Theta}^j$ is defined such that component i_θ has bounds $\pm N_b^j = 5$ % of its nominal value: $[0.95, 1.05] \times (\theta_{base})_{i_\theta}$;
2. Identify the candidate optimal design point $\boldsymbol{\theta}_c^j$ such that $\boldsymbol{\theta}_c^j \in \tilde{\Theta}^j$;
3. If a unique local minimum value of $R_{\tilde{\Theta}_{i_\theta}}$ exists for a particular design parameter $(\theta_c^j)_{i_\theta}$, the parameter will be assigned that fixed value and will not be updated in subsequent iterations. Let the total number of such optimised design parameters up to iteration j be denoted by N_{op}^j, for $N_{op}^j = 0, \ldots, 9$;
4. Define the new design space $\tilde{\Theta}^{j+1}$ whose $(9 - N_{op}^j)$-dimensional hyper-rectangle is defined such that component i_θ, which is not optimised, has increased bounds of $\pm N_b^{j+1} = (j+1) \times N_b^{j=1}$ % of its nominal value;
5. Set $j = j + 1$ and repeat steps 2 to 4 until the local minimum of $R_{\tilde{\Theta}_{i_\theta}}$ is identified for all possible θ_{i_θ} for which an optimal point exists.

The entire recursive optimisation procedure involved 6 iterations in total and as a representative graphical illustration, Figure 18 presents the optimisation

result obtained at iteration $j = 3$. From the figure, it is observed that the local optimal points corresponding to the local minimum of $R_{\tilde{\Theta}_{i_\theta}}$ are identified for θ_3 and θ_7. In addition, it can also be seen that the remaining components of $\boldsymbol{\theta}$ tend to be increasing or decreasing monotonically in general. For these components, their bounds will be increased according to step 4 of the NISS optimisation procedure, and subsequently step 5, until the local optimal points are identified for all 9 components of $\boldsymbol{\theta}$.

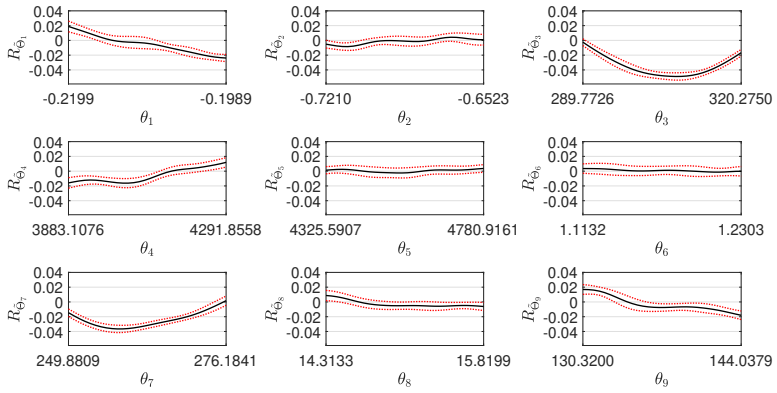

Figure 18: Results of the optimisation of each design parameter for $\boldsymbol{\theta}_{new}$ from the generalised NISS method at iteration $j = 3$. The red dotted lines denote the 95 % confidence interval bounds.

5.2. Reliability Analysis of New Design

The failure probability and severity computation on $\boldsymbol{\theta}_{new}$ is done according to UM^2_{y1} following the approach outlined in Section 4.1. The numerical results to the reliability metrics R_{i_g}, R, and s_{i_g} are presented in Table 14. From the table, it can be seen that the upper-bounds of the failure probabilities R_1 to R_3 as well as the severities s_1 to s_3 have been reduced tremendously which validates the improvement of the design $\boldsymbol{\theta}_{new}$ over $\boldsymbol{\theta}_{base}$. This can be seen from the 3-fold reduction in the upper-bound of R_1, a 2-fold reduction in R_2 and R, and a 10-fold reduction in $\tilde{s}$ between $\boldsymbol{\theta}_{base}$ and $\boldsymbol{\theta}_{new}$. Such results highlight the effectiveness of the optimisation procedure and the identified $\boldsymbol{\theta}_{new}$.

Next, the sensitivity analysis is performed on the epistemic parameters e_1 to e_4 where they are ranked according to the maximum possible reduction on $R(\boldsymbol{\theta}_{new})$ bounds. The methodology follows that outlined in Section 4.2 and the resulting illustrative plots from the analysis is shown in Figure 19. From which, the sensitivity ranking of e_1 to e_4, along with the corresponding reduced bounds of $R(\boldsymbol{\theta}_{new})$, are presented in Table 8. From the results, it can be observed that the ranking order is consistent with the results obtained in Section 4.2 (i.e. see Table 7).

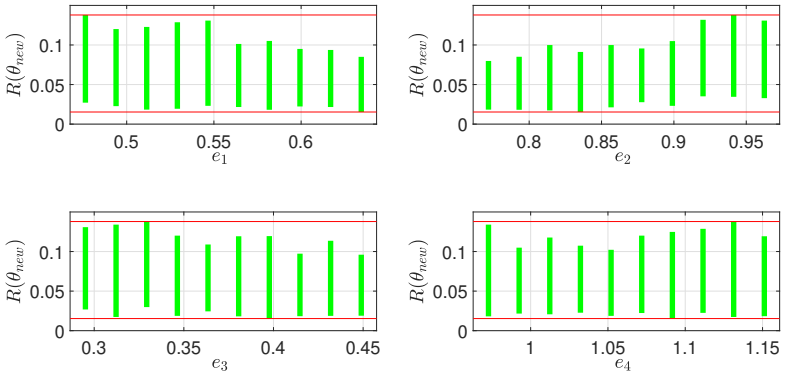

Figure 19: Results of the reduced $R(\boldsymbol{\theta}_{new})$ bounds for different pinched intervals for e_1 to e_4. The red lines represent the initial bounds of $[0.0058, 0.1418]$, while the green bars represent the reduced bounds.

Rank	Parameter	Pinched bounds	$R(\boldsymbol{\theta}_{new})$	
			Before pinching	After pinching
1	e_2	$[0.7613, 0.7825]$	$[0.0058, 0.1418]$	$[0.0183, 0.0798]$
2	e_1	$[0.6257, 0.6433]$	$[0.0058, 0.1418]$	$[0.0153, 0.0850]$
3	e_3	$[0.4404, 0.4575]$	$[0.0058, 0.1418]$	$[0.0189, 0.0960]$
4	e_4	$[0.9826, 1.0025]$	$[0.0058, 0.1418]$	$[0.0216, 0.1049]$

Table 8: The ranking order of the epistemic model parameters based on the maximum possible reduction in $R(\boldsymbol{\theta}_{new})$ interval according to UM^2_{y1}.

Finally, the failure analysis is performed on $\boldsymbol{\theta}_{new}$ following the methodology

outlined in Section 4.3. The resulting statistics summarising the number of realizations of $\{\boldsymbol{a}, \boldsymbol{e}\}$ in each failure category in presented in Table 12 where it can be observed that likelihood of occurrences for all failure types have been reduced significantly, most notably for failure types $g_3 \geq 0$, $g_1, g_3 \geq 0$, and $g_1, g_2, g_3 \geq 0$ which have been reduced to 0. Failure type $g_2 \geq 0$ still has the highest likelihood as it has the highest number of realizations among the different failure types as per $\boldsymbol{\theta}_{base}$. Following which, the sample sets $\{\boldsymbol{a}, \boldsymbol{e}\}_{nf}^{g_1}$, $\{\boldsymbol{a}, \boldsymbol{e}\}_{nf}^{g_2}$, and $\{\boldsymbol{a}, \boldsymbol{e}\}_{nf}^{g_3}$ are identified and the resulting parallel plots are presented in Figure 20 while the response plots of $z_1(t)$ and $z_2(t)$ are presented in Figure 21.

From Figure 21, the response plots of $z_1(t)$ for $\{\boldsymbol{a}, \boldsymbol{e}\}_{nf}^{g_1}$ exceeds the upper-bound of the safe boundary at approximately $t = 2.6\ s$ which indicates the presence of failure $g_2 \geq 0$. The plots for $z_2(t)$, on the other hand, are well-within the safety limits. This indicates that the sample set $\{\boldsymbol{a}, \boldsymbol{e}\}_{nf}^{g_1}$ lie within the domain of $g_2 \geq 0$ near the boundary of $g_1 \geq 0$. For $\{\boldsymbol{a}, \boldsymbol{e}\}_{nf}^{g_2}$, the response plots of $z_1(t)$ and $z_2(t)$ are within the safety limits although it can also be observed that the $z_1(t)$ plots are extremely close to both the upper and lower-bounds of the safe boundary which verifies that the identified $\{\boldsymbol{a}, \boldsymbol{e}\}_{nf}^{g_2}$ are near $g_2 \geq 0$. As seen from the stable behaviour of the plots, it can be inferred that the requirement of g_1 is satisfied. This indicates that the sample set $\{\boldsymbol{a}, \boldsymbol{e}\}_{nf}^{g_2}$ lie within the safe domain near the boundary of $g_2 \geq 0$. For $\{\boldsymbol{a}, \boldsymbol{e}\}_{nf}^{g_3}$, the response plots of $z_1(t)$ exceeds the upper and lower-bounds of the safe boundary at approximately $t = \{2.5, 2.7, 2.8, 3.1, 3.4\}\ s$ which indicates the presence of failure $g_2 \geq 0$. The plots for $z_2(t)$, on the other hand, are well-within the safety limits. Given the stable behaviour of the plots, it can be inferred that the requirement of g_1 is satisfied. This indicates that the sample set $\{\boldsymbol{a}, \boldsymbol{e}\}_{nf}^{g_3}$ lie within the domain of $g_2 \geq 0$ near the boundary of $g_2, g_3 \geq 0$.

A quantitative study is also done for the different failure types with non-zero likelihood. As per what was done in Section 4.3, the response plots of $z_1(t)$ and $z_2(t)$ for the 25 representative sample sets of $\{\boldsymbol{a}, \boldsymbol{e}\}$ for each failure type (5 for failure type $g_2, g_3 \geq 0$) are presented in Figures 22 and 23 respectively. From Figure 23, it can be seen that the unstable behaviour in $z_2(t)$ for failure type

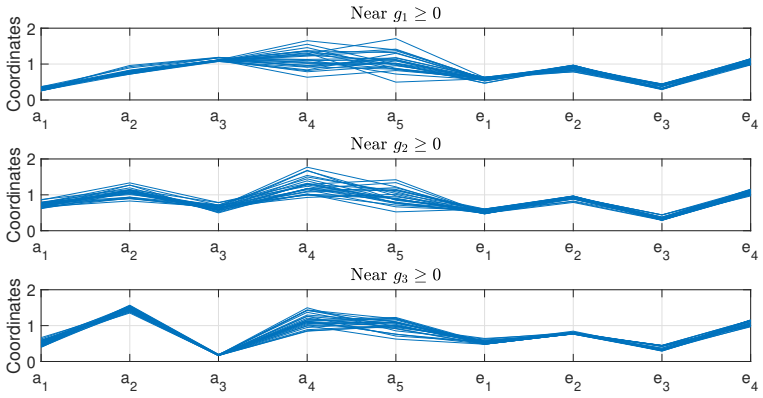

Figure 20: Parallel plots of $\{\boldsymbol{a}, \boldsymbol{e}\}_{nf}^{g_1}$, $\{\boldsymbol{a}, \boldsymbol{e}\}_{nf}^{g_2}$, and $\{\boldsymbol{a}, \boldsymbol{e}\}_{nf}^{g_3}$ from UM_{y1}^2.

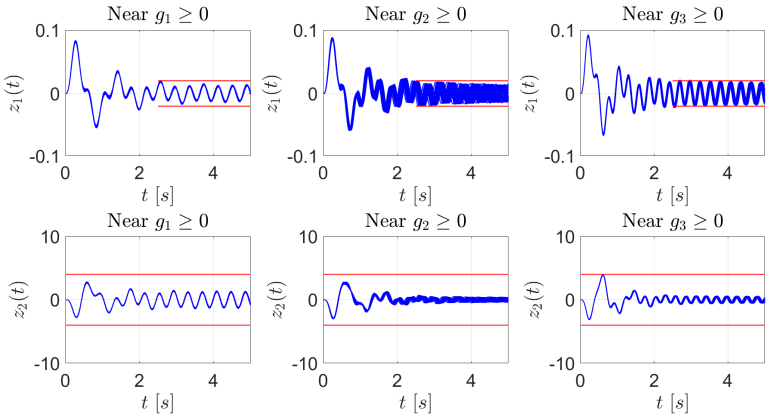

Figure 21: Response plot of $z_1(t)$ and $z_2(t)$ of $\{\boldsymbol{a}, \boldsymbol{e}\}_{nf}^{g_1}$, $\{\boldsymbol{a}, \boldsymbol{e}\}_{nf}^{g_2}$, and $\{\boldsymbol{a}, \boldsymbol{e}\}_{nf}^{g_3}$ from UM_{y1}^2 near the respective failure domains. The red lines denote the safety limits.

$g_1, g_2 \geq 0$ is still present along with those illustrating a stable behaviour. In addition, Figure 22 shows that the $z_1(t)$ response plots for failure type $g_1, g_2 \geq 0$ exceeds the safety limits to the largest extent relative to the other failure types which indicates that the failure type contributes the most to the severity s_1, s_2, and $\tilde{s}$. This motivates the need to identify $\boldsymbol{\theta}_{final}$ such that the number of realizations of $\{\boldsymbol{a}, \boldsymbol{e}\}$ corresponding to such unstable behaviour is minimised. The identification of $\boldsymbol{\theta}_{final}$ will be done in Section 6.3.

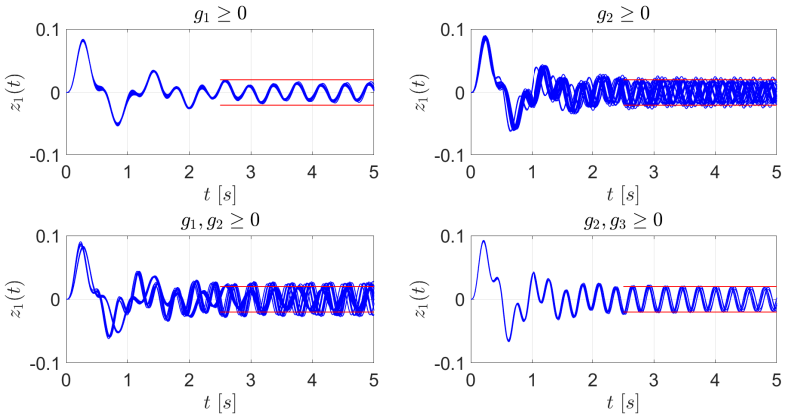

Figure 22: Response plot of $z_1(t)$ corresponding to 25 representative realizations of $\{\boldsymbol{a}, \boldsymbol{e}\}$ for the each failure type (5 for failure type $g_2, g_3 \geq 0$). The red lines denote the safety limits.

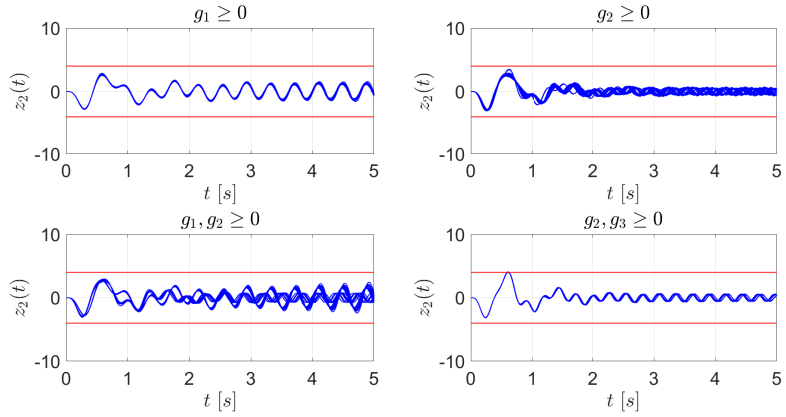

Figure 23: Response plot of $z_2(t)$ corresponding to 25 representative realizations of $\{\boldsymbol{a}, \boldsymbol{e}\}$ for the each failure type (5 for failure type $g_2, g_3 \geq 0$). The red lines denote the safety limits.

6. Task E: Model Update and Design Tuning

The objective of this task is to improve the current UM and identify an improved design point $\boldsymbol{\theta}_{final}$ based on the observations of $z_1(t)$ and $z_2(t)$ from the integrated system corresponding to $\boldsymbol{\theta}_{new}$.

6.1. Model Calibration

The model calibration is performed with the Black-box model function $\hat{z} = zfun(\boldsymbol{a}, \boldsymbol{e}, \boldsymbol{\theta}_{new}, t)$ using the new data sequence $D_2 = \{z^l(t)\}_{l=1,\ldots,100}$. This is done following the approach outlined in Section 2.3.2, but with bounds of the Uniform priors for the respective epistemic parameters defined by the hyper-rectangle E_1 and $P(\boldsymbol{D}|\boldsymbol{\Theta}, M)$ having configurations $n_b = 5$ and $N_{sim} = 500$ to reduce the computational costs. Let the calibrated UM be denoted as UM_{z0} whose resulting P-box representation of the marginal distributions of $f_{\boldsymbol{a}}$ is illustrated in Figure 32 and whose uncertainty bounds over $\boldsymbol{e}$ are presented in Table 10. To validate the model calibration results, the resulting model output bands of $\hat{z}$ from UM_{z0} (in blue) are illustrated in Figures 28 and 30.

6.1.1. Results and Discussion

In both figures, it is observed that the model output bands of UM_{z0} (in blue) generally encloses the data D_2 (in red) which indicates that the model calibration procedure, via Bayesian model updating, was done satisfactorily. However, from Figure 30, some unstable behaviour is observed in the response of $z_2(t)$. To investigate this, a scatterplot matrix of the aleatory samples is presented in Figure 24 where it is observed that the samples responsible for such unstable behaviour are predominantly located in the corners of the aleatory space as seen from the following 2-dimensional space: 1) a_1 vs a_2; 2) a_1 vs a_3; and 3) a_2 vs a_3. There are a total of 165 of such aleatory samples and in order reduce the number of samples from those regions, correlations need to be introduced such that a negative correlation exists between a_1 and a_2 (i.e. $C_{1,2}$), while a positive correlation exists between a_1 and a_3 (i.e. $C_{1,3}$) as well as between a_2 and a_3 (i.e. $C_{2,3}$). These correlations will be modelled using a Gaussian copula function [38].

To identify the correlation parameters $C_{1,2}$, $C_{1,3}$, and $C_{2,3}$, a second round of Bayesian model updating is performed on UM_{z0} with the inferred parameters being the aforementioned correlation parameters. The prior distributions and bounds for $C_{1,2}$, $C_{1,3}$, and $C_{2,3}$ are summarised in Table 9. This procedure

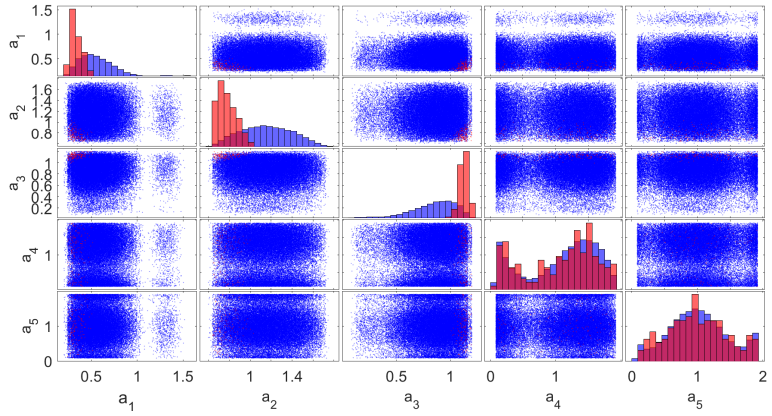

Figure 24: Scatterplot matrix of the aleatory samples from from UM^1_{z0} without considering correlations. The plots in blue contribute to the stable realizations while those in red contribute to the unstable realizations of $z_2(t)$.

is done keeping the P-box of $f_{\boldsymbol{a}}$ and the uncertainty bounds of $\boldsymbol{e}$ obtained in the previous round of Bayesian model updating as fixed models for $\boldsymbol{a}$ and $\boldsymbol{e}$. The set-up of $P(\boldsymbol{D}|\boldsymbol{\Theta}, M)$ follows that used to calibrate UM_{z0} initially. The resulting histograms for $C_{1,2}$, $C_{1,3}$, and $C_{2,3}$ are illustrated in Figure 25 from which the Most Probable Value (MPV) for the respective correlation parameters are obtained and presented in Table 9.

Parameter	$C_{1,2}$	$C_{1,3}$	$C_{2,3}$
Prior distribution parameters	$U[-1,0]$	$U[0,1]$	$U[0,1]$
MPV	-0.0427	0.2064	0.0316

Table 9: Summary of the prior distributions and the resulting MPVs for the respective correlation parameters.

To illustrate the effectiveness of introducing correlations, UM_{z0} is updated to account for $C_{1,2}$, $C_{1,3}$, and $C_{2,3}$ whose values correspond to the MPV as shown in Table 9. This is done following the set-up which was used to calibrate UM_{z0} prior to the introduction of correlations. The resulting scatterplot matrix is presented in Figure 26 which shows that the number of aleatory samples

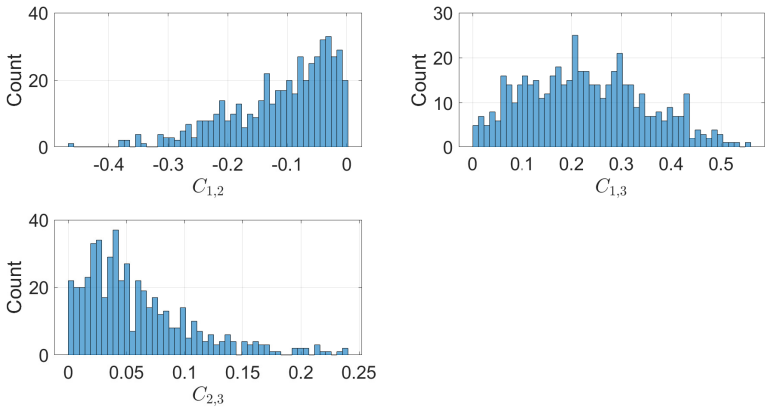

Figure 25: Histograms for $C_{1,2}$, $C_{1,3}$, and $C_{2,3}$ obtained from $P(\boldsymbol{\Theta}|\boldsymbol{D}, M)$.

corresponding to the unstable realizations of $z_2(t)$ has been reduced from 165 to 39.

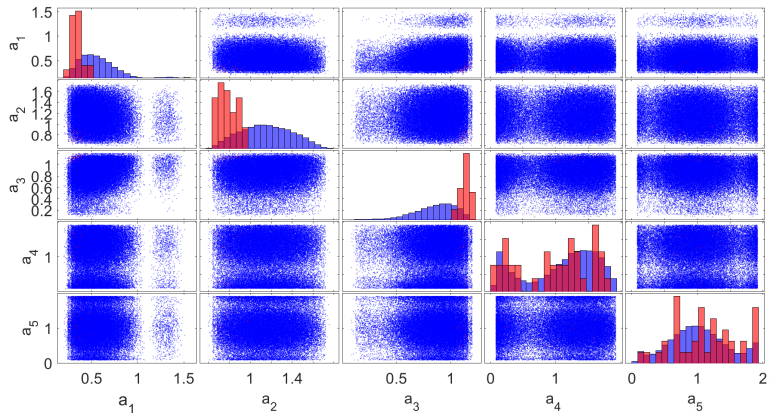

Figure 26: Scatterplot matrix of the aleatory samples from from UM^1_{z0} after considering correlations. The plots in blue contribute to the stable realizations while those in red contribute to the unstable realizations of $z_2(t)$.

6.2. Uncertainty Reduction

From the resulting bounds of the epistemic space according to UM_{z0} as shown in Table 10, it is observed that the uncertainty bounds for e_4 is the largest

among the epistemic parameters. This is attributed to the lack of information provided by e_4 to which evidence is provided by the sensitivity analysis done in Section 3.3 (i.e. see Table 5) which shows that e_4 is still the least sensitive parameter even after refinement. Due to substantial knowledge of its upper-bound, as indicated by the histograms of e_4 in Figures 12 and 33 which are truncated at the upper-bound, a request is made for the refinement of the lower-bound of e_4.

Based on the sensitivity analysis done in Sections 3.3, 4.2, and 5.2, the results show that e_2 has consistently been ranked the most sensitive epistemic parameter. In addition to this, Figure 27 shows the parallel plots of the samples according to UM_{z0} from which it can be seen that the lower-bound of the interval for e_2 is increased after correlation is introduced with the identified values of $C_{1,2}$, $C_{1,3}$, and $C_{2,3}$ (in blue) compared to the absence of correlation (in red). In fact, the change in interval of e_2 is the most substantial compared to the other epistemic parameters after correlation is introduced to the UM_{z0}. This further substantiates the evidence that e_2 is the most sensitive epistemic parameter whose refinement can further reduce the unstable realizations of $z_2(t)$. For these reasons, the final refinement request is made for the upper-bound of e_2 given that there is substantial knowledge in its lower-bound as observed from Figure 27. The resulting epistemic space, with the refined e_2 and e_4 bounds, constitutes the hyper-rectangle epistemic space denoted as E_2.

Following this, a third round of Bayesian model updating is performed with bounds of the Uniform priors for the respective epistemic parameters defined by the hyper-rectangle E_2 and the correlation parameters taking fixed values as defined in Table 9. The approach follows that outlined in Section 6.1 from which UM_{z0} was obtained, however, this time, accounting for the identified correlation parameters in the aleatory space. Let the refined UM be denoted as UM_{z1} whose uncertainty bounds over $\boldsymbol{e}$ are presented in Table 10. In addition, due to substantial information on the parameters of the SDF, their respective MPV values are used to define the final CDF representation of the marginal distributions of $f_{\boldsymbol{a}}$ which are illustrated in Figure 32. Finally, to validate the

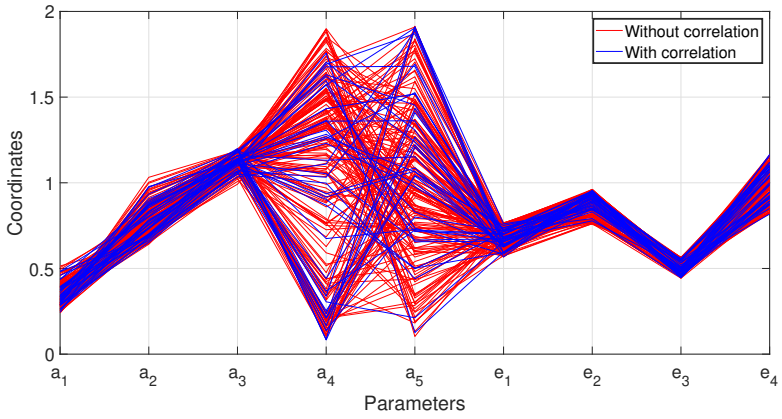

Figure 27: Parallel plots of the samples from UM_{z0} corresponding to the unstable realizations before considering correlations (in red) and after considering correlations (in blue).

model calibration results, the resulting model output bands of $\hat{z}$ from UM_{z1} (in green) are illustrated in Figures 28 and 30.

6.2.1. Results and Discussion

In both figures, it is observed that the model output bands of UM_{z1} (in green) generally encloses the data D_2 (in red) which indicates that the model calibration procedure was done satisfactorily. From Figure 30, it can also be observed that that there is no unstable response behaviour of $z_2(t)$ which indicates that the refinement procedure has removed all the remaining unstable realizations. Figures 29 and 31 illustrate the P-boxes of the model output of UM_{z0} (i.e. before and after accounting for correlations) and UM_{z1} at time-slices $t = \{0.5, 1.0, 2.0, 3.0, 4.0, 5.0\}$ s for $z_1(t)$ and $z_2(t)$ respectively. From the figures, it can be seen that the P-boxes obtained from the respective UMs generally enclose and follow the trend of the ECDF defined by D_2. This indicates a good degree of fit of the model outputs obtained by the UMs. Moreover, the P-box of the model output according to UM_{z1} (in green) is significantly narrower than UM_{z0}, with and without accounting for correlations, whilst still enclosing the ECDF of D_2 in general. This further highlights the effectiveness of the refinement procedure that is undertaken in this task and validates the

results obtained by UM_{z1}.

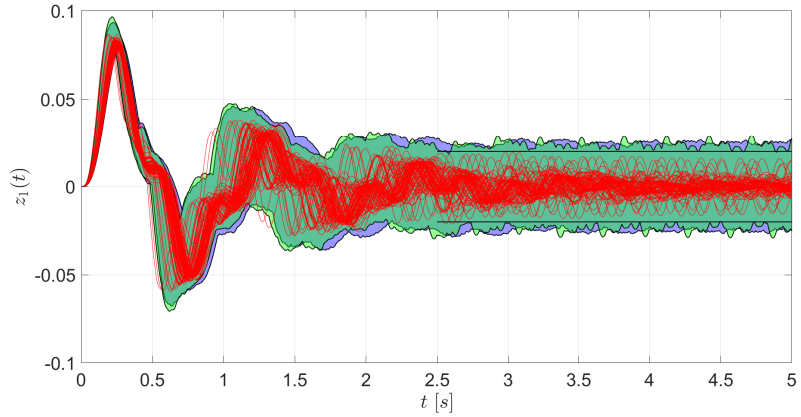

Figure 28: Output band from $\hat{z}$ according to UM_{z0} without correlations (in blue) and UM_{z1} (in green) along with the data sequence D_2 (in red) after calibration. The black lines denote the safety limits.

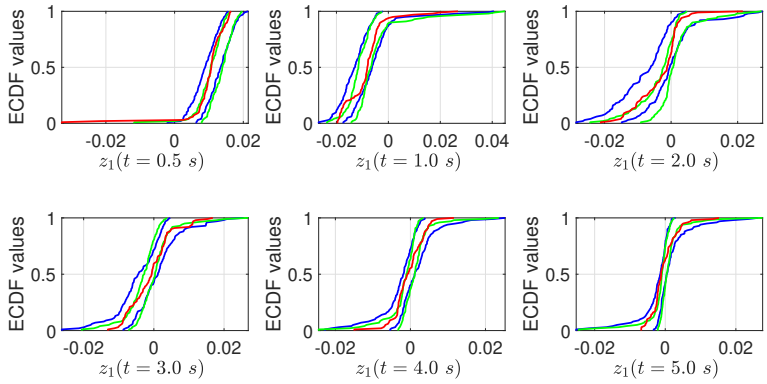

Figure 29: P-boxes of the model output from $\hat{z}$ obtained from UM_{z0} without correlations (in blue) and UM_{z1} (in green) at various time slices $t = \{0.5, 1.0, 2.0, 3.0, 4.0, 5.0\}\ s$. The red ECDF denotes the distribution of the data D_2.

Figure 32 illustrates the resulting P-box quantifying the uncertainty over the marginal distributions of $f_{\boldsymbol{a}}$ by UM_{z0} as well as the final CDF for $f_{\boldsymbol{a}}$ by UM_{z1}. From the figure, the P-boxes (in blue) generally encloses and show a good degree

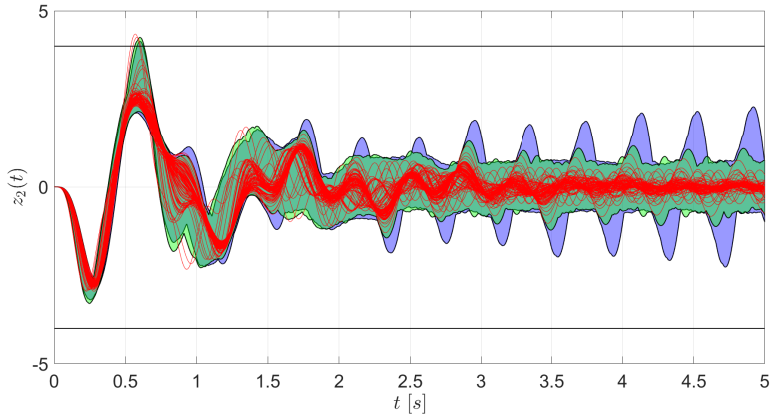

Figure 30: Output band from $\hat{z}$ according to UM_{z0} without correlations (in blue) and UM_{z1} (in green) along with the data sequence D_2 (in red) after calibration. The black lines denote the safety limits.

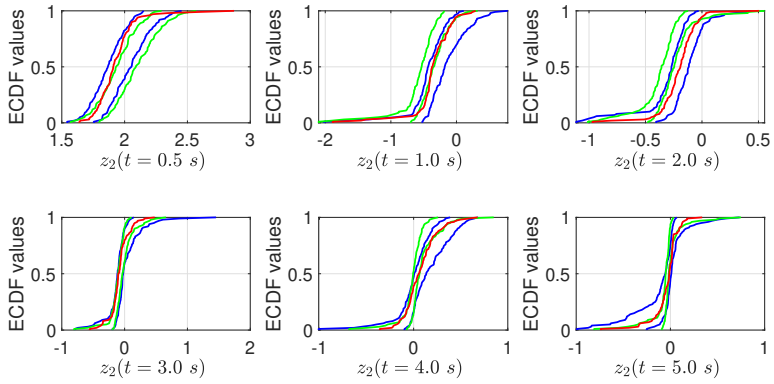

Figure 31: P-boxes of the model output from $\hat{z}$ obtained from UM_{z0} without correlations (in blue) and UM_{z1} (in green) at various time slices $t = \{0.5, 1.0, 2.0, 3.0, 4.0, 5.0\}\ s$. The red ECDF denotes the distribution of the data D_2.

of agreement with the final CDF (in green) which verifies the results of UM_{z1}. Such observation also provides good evidence that the final CDF illustrated in Figure 32 is a good representation of the true CDF of $f_{\boldsymbol{a}}$.

Finally, results from Table 10 show that the uncertainty bounds over $\boldsymbol{e}$ according to UM_{z1}, with the exception of e_3, is significantly narrower than those

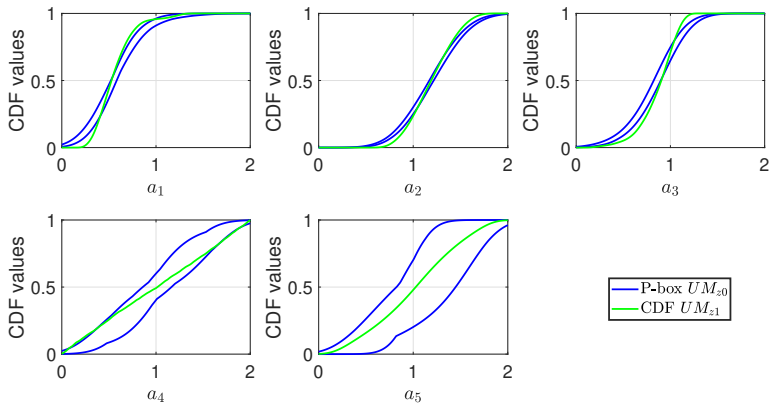

Figure 32: P-box for a_1 to a_5 obtained from the respective UMs.

according to UM_{z0}. For the case of e_1 and e_2, their bounds according to UM_{z1} are not enclosed within that defined by UM_{z0}. Such observation is supported by Figure 33 where it can be seen that there is little to no overlap between the histograms for e_1 and e_2 obtained before and after refinement. This is due to the fact that the initial bounds of e_2 defined by UM_{z0} lie entirely outside the refined bounds provided by the challenge hosts. As such, when the calibration was performed once again with such new information, the effect of such refinement is significant on e_1 which results in the epistemic parameter also having reduced bounds which largely outside the initial bounds defined by UM_{z0}. This observation also suggests a significant correlation between e_1 and e_2.

Uncertainty model	e_1	e_2	e_3	e_4
UM_{z0}	$[0.5961, 0.7319]$	$[0.7790, 0.9337]$	$[0.4777, 0.5670]$	$[0.8521, 1.1664]$
UM_{z1}	$[0.4384, 0.5795]$	$[0.5350, 0.5704]$	$[0.3353, 0.5670]$	$[0.9027, 0.9497]$

Table 10: Updated epistemic space E for e_1 to e_4 according to the respective UMs.

6.3. Identification and Reliability Analysis of Final Design

The objectives of this task are: 1) to identify the final design point $\boldsymbol{\theta}_{final}$; and 2) perform the reliability analysis on $\boldsymbol{\theta}_{final}$ according to UM_{z1}.

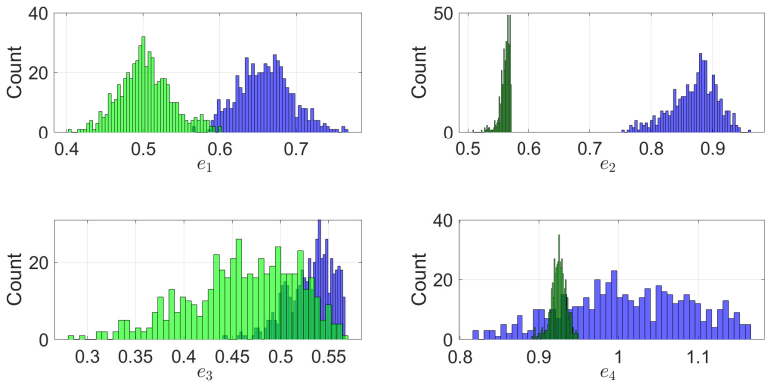

Figure 33: Histograms for e_1 to e_4 obtained from before refinement (in blue) and after refinement (in green).

The optimisation procedure to identify $\boldsymbol{\theta}_{final}$ follows the methodology outlined in Section 5.1 with $N_b^{j=1} = 3$ %.

The reliability analysis was performed following the methodology outlined in Section 4.1. The results for R_{i_g}, R, and s_{i_g} are presented in Table 14. From the table, it can be seen that the upper-bounds of the failure probabilities R_1 and R_3 have been reduced compared to $\boldsymbol{\theta}_{new}$ according to UM_{y1}^2, and that the severities s_1 to s_3 have all been reduced to almost 0. Although the upper-bound of R_2 shows a small increase from $\boldsymbol{\theta}_{new}$ according to UM_{z1}, the upper-bound of R and $\tilde{s}$ are reduced in the case of $\boldsymbol{\theta}_{final}$ according to UM_{z1}. Such results and observations highlight the effectiveness of the optimisation procedure and the identified $\boldsymbol{\theta}_{final}$.

Next, the sensitivity analysis is performed on the epistemic parameters e_1 to e_4 where they are ranked according to the maximum possible reduction on $R(\boldsymbol{\theta}_{final})$ bounds. The methodology follows that outlined in Section 4.2 and the resulting illustrative plots from the analysis is shown in Figure 34. From which, the sensitivity ranking of e_1 to e_4, along with the corresponding reduced bounds of $R(\boldsymbol{\theta}_{final})$, are presented in Table 11. From the results, it is observed that e_2 and e_1 have 2 of the lowest sensitivities which indicates that the bounds

obtained for these 2 parameters are sufficiently narrow such that no further information can be obtained on them by reducing their bounds any further.

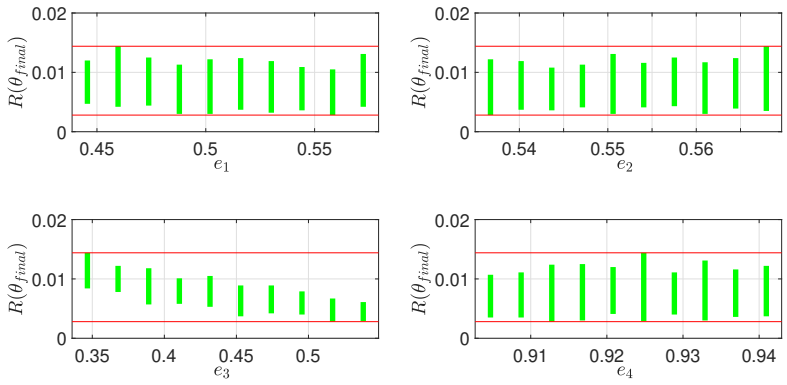

Figure 34: Results of the reduced $R(\boldsymbol{\theta}_{final})$ bounds for different pinched intervals for e_1 to e_4. The red lines represent the initial bounds of $[0.0028, 0.0144]$, while the green bars represent the reduced bounds.

Rank	Parameter	Pinched bounds	$R(\boldsymbol{\theta}_{final})$	
			Before pinching	After pinching
1	e_3	$[0.5274, 0.5487]$	$[0.0028, 0.0144]$	$[0.0029, 0.0061]$
2	e_4	$[0.9269, 0.9309]$	$[0.0028, 0.0144]$	$[0.0040, 0.0111]$
3	e_2	$[0.5419, 0.5454]$	$[0.0028, 0.0144]$	$[0.0036, 0.0108]$
4	e_1	$[0.5372, 0.5512]$	$[0.0028, 0.0144]$	$[0.0036, 0.0109]$

Table 11: The ranking order of the epistemic model parameters based on the resulting contractions of $R(\boldsymbol{\theta}_{final})$ from pinching according to UM_{z1}.

Finally, the failure analysis is performed on $\boldsymbol{\theta}_{final}$ following the methodology outlined in Section 4.3. The resulting statistics summarising the number of realizations of $\{\boldsymbol{a}, \boldsymbol{e}\}$ in each failure category in presented in Table 12 where it can be observed that likelihood of occurrences for all failure types have been reduced significantly from $\boldsymbol{\theta}_{new}$ according to UM^2_{y1}, except for failure types $g_3 \geq 0$ and $g_2, g_3 \geq 0$ where there is an increase in the number of sample realizations in these failure domains. Failure type $g_2 \geq 0$ still has the highest likelihood as

it has the highest number of realizations among the different failure types as per $\boldsymbol{\theta}_{base}$ and $\boldsymbol{\theta}_{new}$. Following which, the sample sets $\{\boldsymbol{a}, \boldsymbol{e}\}_{nf}^{g_1}$, $\{\boldsymbol{a}, \boldsymbol{e}\}_{nf}^{g_2}$, and $\{\boldsymbol{a}, \boldsymbol{e}\}_{nf}^{g_3}$ are identified and the resulting parallel plots are presented in Figure 35 while the response plots of $z_1(t)$ and $z_2(t)$ are presented in Figure 36.

From Figure 36, the response plots of $z_1(t)$ and $z_2(t)$ corresponding to $\{\boldsymbol{a}, \boldsymbol{e}\}_{nf}^{g_1}$ are all within the safety limits. This indicates that the sample set $\{\boldsymbol{a}, \boldsymbol{e}\}_{nf}^{g_1}$ lie within the safe domain near the boundary of $g_1 \geq 0$. For $\{\boldsymbol{a}, \boldsymbol{e}\}_{nf}^{g_2}$, the response plots of $z_1(t)$ and $z_2(t)$ are all within the safety limits as well which indicates that the sample set $\{\boldsymbol{a}, \boldsymbol{e}\}_{nf}^{g_2}$ lie within the safe domain near the boundary of $g_2 \geq 0$. For $\{\boldsymbol{a}, \boldsymbol{e}\}_{nf}^{g_3}$, the response plots of $z_1(t)$ exceeds the upper and lower-bounds of the safe boundary at approximately $t = \{2.5, 2.6, 2.9, 3.0, 3.2, 3.3, 3.4\}$ s which indicates the presence of failure $g_2 \geq 0$. The plots for $z_2(t)$, on the other hand, are well-within the safety limits. Given the stable behaviour of the plots, it can be inferred that the requirement of g_1 is satisfied which indicates that the sample set $\{\boldsymbol{a}, \boldsymbol{e}\}_{nf}^{g_3}$ lie within the domain of $g_2 \geq 0$ near the boundary of $g_2, g_3 \geq 0$.

Figures 37 and 38 present the response plots of $z_1(t)$ and $z_2(t)$ respectively for the 25 representative sample sets of $\{\boldsymbol{a}, \boldsymbol{e}\}$ for each failure type with non-zero likelihood. From Figure 38, it can be seen that the unstable behaviour in $z_2(t)$ that was previously present for failure type $g_1, g_2 \geq 0$ are no longer present as a result of the refinement procedure that was done in Section 6.2. In addition, Figure 37 shows that the $z_1(t)$ response plots for failure type $g_1, g_2 \geq 0$ still exceeds the safety limits to the largest extent compared to the other failure types.

6.4. Comparison of Design Points

The objective of this task is to present a quantitative comparison between the design points $\boldsymbol{\theta}_{base}$, $\boldsymbol{\theta}_{new}$, and $\boldsymbol{\theta}_{final}$ on the basis of the reliability metrics R_{i_g}, R, s_{i_g}, and $\tilde{s}$.

In Table 14, it can be seen that the reliability analysis results for the different $\boldsymbol{\theta}$ suggest that $\boldsymbol{\theta}_{final}$ is the most optimal design point. This is due to $\boldsymbol{\theta}_{final}$

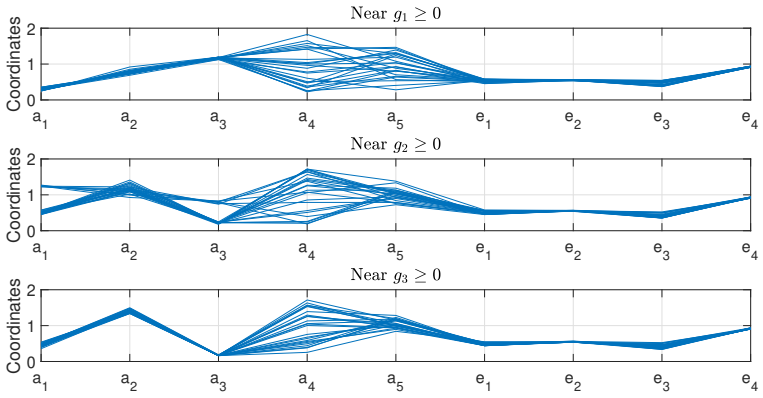

Figure 35: Parallel plots of $\{\boldsymbol{a}, \boldsymbol{e}\}_{nf}^{g_1}$, $\{\boldsymbol{a}, \boldsymbol{e}\}_{nf}^{g_2}$, and $\{\boldsymbol{a}, \boldsymbol{e}\}_{nf}^{g_3}$ from UM_{z1}.

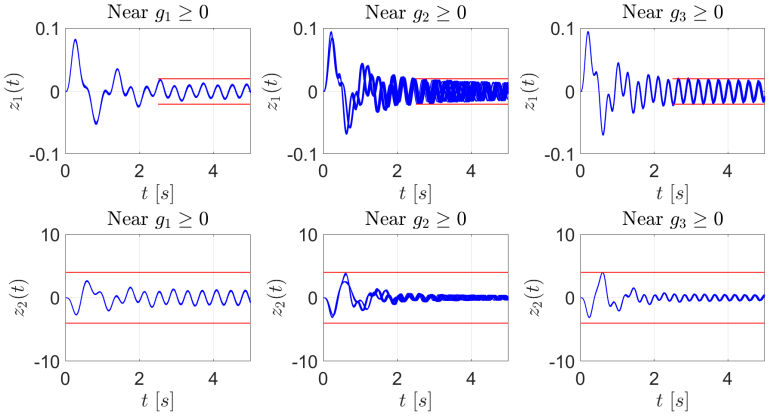

Figure 36: Response plot of $z_1(t)$ and $z_2(t)$ of $\{\boldsymbol{a}, \boldsymbol{e}\}_{nf}^{g_1}$, $\{\boldsymbol{a}, \boldsymbol{e}\}_{nf}^{g_2}$, and $\{\boldsymbol{a}, \boldsymbol{e}\}_{nf}^{g_3}$ from UM_{z1} near the respective failure domains. The red lines denote the safety limits.

having the lowest upper-bound failure probabilities and severities in general compared to $\boldsymbol{\theta}_{base}$ and $\boldsymbol{\theta}_{new}$ based on the reliability results according to UM_{y1}^2 and UM_{z1}. Such results are supported by Figures 39 and 40 which illustrate the resulting P-boxes for g_1 to g_3 according to the analysis for the different $\boldsymbol{\theta}$ by UM_{y1}^2 and UM_{z1} respectively. In both figures, it can be seen that the P-boxes for $\boldsymbol{\theta}_{final}$ are such that they mainly lie within the safe domain of the respective requirement and that the tails of the P-boxes do not extend as far into the

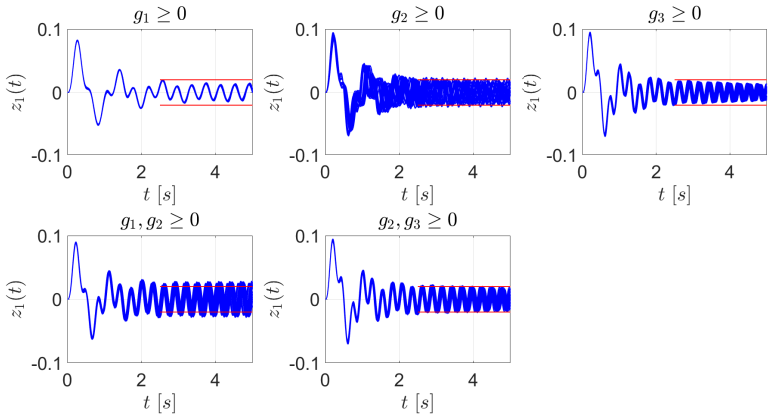

Figure 37: Response plot of $z_1(t)$ corresponding to 25 representative realizations of $\{\boldsymbol{a}, \boldsymbol{e}\}$ for the each failure type. The red lines denote the safety limits.

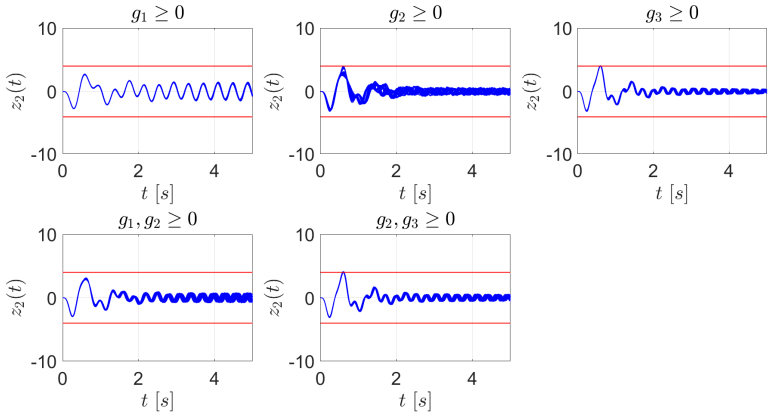

Figure 38: Response plot of $z_2(t)$ corresponding to 25 representative realizations of $\{\boldsymbol{a}, \boldsymbol{e}\}$ for the each failure type. The red lines denote the safety limits.

failure regions compared to $\boldsymbol{\theta}_{base}$ and $\boldsymbol{\theta}_{new}$.

A further analysis was also done to compare the number of realizations of $\{\boldsymbol{a}, \boldsymbol{e}\}$ for the different failure types according to UM^2_{y1} and UM_{z1} and the resulting statistics are provided in Table 12. For both UM^2_{y1} and UM_{z1}, the number of safe (i.e. no failure) realizations is the highest for $\boldsymbol{\theta}_{final}$ which further substantiates it being the most optimal design point for the system. While this

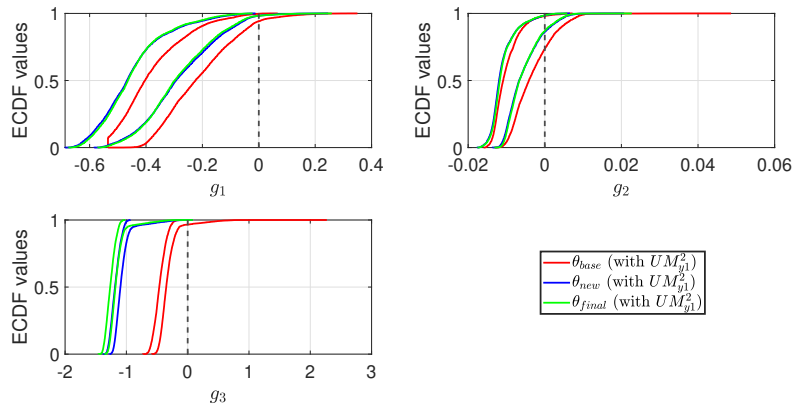

Figure 39: P-boxes obtained for g_1, g_2, and g_3 for different $\boldsymbol{\theta}$ according to UM^2_{y1}

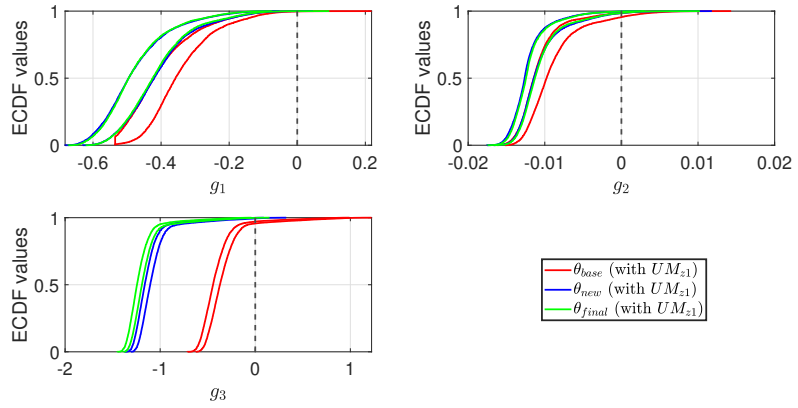

Figure 40: P-boxes obtained for g_1, g_2, and g_3 for different $\boldsymbol{\theta}$ according to UM_{z1}

is achieved, a drawback of $\boldsymbol{\theta}_{final}$ is the increase in realizations of failure types $g_1 \geq 0$ and $g_2 \geq 0$ from $\boldsymbol{\theta}_{new}$ according to UM_{z1} while for the case of UM^2_{y1}, such drawback comes in the form of a slight increase in realizations of failure type $g_1, g_2, g_3 \geq 0$ from $\boldsymbol{\theta}_{new}$.

6.5. Numerical Implementation and Computational Time

In addressing the tasks presented in this challenge, the adopted algorithms are mainly based on random sampling and stochastic algorithms. Thus, the ex-

Failure Type	$\boldsymbol{\theta}_{base}$		$\boldsymbol{\theta}_{new}$		$\boldsymbol{\theta}_{final}$	
	UM^2_{y1}	UM_{z1}	UM^2_{y1}	UM_{z1}	UM^2_{y1}	UM_{z1}
No Failure	4357960	4735066	4696943	4955517	4713788	4963836
$g_1 \geq 0$	54827	4002	3529	22	2485	36
$g_2 \geq 0$	490718	75237	290863	30231	275785	34004
$g_3 \geq 0$	16239	126310	0	7391	0	657
$g_1, g_2 \geq 0$	45802	1888	8660	187	7940	81
$g_1, g_3 \geq 0$	0	0	0	0	0	0
$g_2, g_3 \geq 0$	34256	57471	5	6652	0	1386
$g_1, g_2, g_3 \geq 0$	198	26	0	0	2	0
Total Samples	5×10^6					

Table 12: Statistics of the different failures based on analysis for different design points and UMs.

ecution time fluctuates significantly due to the inherent randomness depending on the uncontrolled conditions such as starting samples, the evolution of the samples, etc. However, it needs to be noted that the random seed has not been fixed to allow for generality of the implementation and solution. Therefore, we only provide the approximate timing as the performance indicator: 1) the TMCMC which takes between 5 to 8 hours of sampling time; 2) the adaptive pinching approach which involves less than a minute of computation time; 3) the Double-Loop Monte Carlo simulation which involves between 1.5 to 2 hours of simulation time; and 4) the NISS which takes only 3 minutes of simulation time.

It needs to be highlighted that the computational times stated for the respective tools are also dependent on the computational efficiency of the high-performance CPUs which are used. These estimated timings may differ between different CPUs of different specifications.

7. Conclusion

Different techniques have been presented for solving the NASA UQ challenge problem. Bayesian model updating technique has been used to calibrate

the uncertainty model by performing a stochastic update on both the distribution parameters as well as the epistemic parameters. 2 different uncertainty models have been analysed, each adopting a different choice of joint distribution function for the aleatory space: 1) Beta distribution; and 2) Staircase Density Functions. The use of the Staircase Density function provided more informative results on the bounds of the distribution parameters and the epistemic parameters from their respective posteriors and used in the subsequent problems.

An adaptive pinching analysis based on [26, 27] was proposed to perform the sensitivity analysis on the epistemic parameters, providing an efficient way of identifying the largest possible reduction of the proposed metric by the single pinched component of the epistemic space. In doing so, it allows for a systematic, non-empirical way to justify the pinched bounds and ensure that all regions of the individual components of the epistemic space are accounted for in the investigation of their respective effect on the 2 aforementioned quantities.

To perform the reliability analysis and compute the reliability metrics, Probability-boxes are constructed through the double-loop Monte Carlo approach. From which, the probability bounds analysis was performed on the resulting Probability-boxes to obtain necessary bounds on the respective failure probabilities and the worst-case failure probabilities. The approach does not assume a fixed distribution on each performance and considers only the extreme bounds of the probability values obtained.

Finally, to identify an optimal design point of the system, the Non-intrusive Stochastic Simulation technique [33] was used. The approach provides a systematic way to explore a defined hyper-rectangular space of the design point and identify the values of each design point parameter corresponding to the local minimum of the first-order component function. This allows for the optimised design point to be identified in accordance to the defined criteria.

A summary of key results are provided in Section 8.

8. Appendix

Uncertainty model	e_1	e_2	e_3	e_4
UM^1_{y0}	$[0.4149, 1.5271]$	$[0.2407, 1.5673]$	$[0.1506, 1.8001]$	$[0.1601, 1.9236]$
UM^2_{y0}	$[0.4351, 0.7082]$	$[0.5583, 1.0000]$	$[0.0721, 0.5511]$	$[0.6066, 1.6893]$
UM^1_{y1}	$[0.3730, 1.3457]$	$[0.1869, 1.1529]$	$[0.2997, 0.5570]$	$[0.8295, 1.1664]$
UM^2_{y1}	$[0.4674, 0.6433]$	$[0.7607, 0.9736]$	$[0.2865, 0.4583]$	$[0.9627, 1.1664]$
UM_{z0}	$[0.5961, 0.7319]$	$[0.7790, 0.9337]$	$[0.4777, 0.5670]$	$[0.8521, 1.1664]$
UM_{z1}	$[0.4384, 0.5795]$	$[0.5350, 0.5704]$	$[0.3353, 0.5670]$	$[0.9027, 0.9497]$

Table 13: Results to the epistemic spaces defined by the respective UMs.

Design point	$R_1(\boldsymbol{\theta})$	$R_2(\boldsymbol{\theta})$	$R_3(\boldsymbol{\theta})$	$R(\boldsymbol{\theta})$	$s_1(\boldsymbol{\theta})$	$s_2(\boldsymbol{\theta})$	$s_3(\boldsymbol{\theta})$	$\tilde{s}(\boldsymbol{\theta})$
$\boldsymbol{\theta}_{base}$ (with UM^2_{y1})	$[0.0028, 0.0580]$	$[0.0190, 0.2644]$	$[0.0000, 0.0343]$	$[0.0270, 0.2746]$	0.0413	0.1981	0.0411	0.3779
$\boldsymbol{\theta}_{base}$ (with UM_{z1})	$[0.0003, 0.0023]$	$[0.0137, 0.0445]$	$[0.0291, 0.0436]$	$[0.0386, 0.0699]$	2×10^{-6}	0.0007	0.1082	0.1669
$\boldsymbol{\theta}_{new}$ (with UM^2_{y1})	$[0.0000, 0.0176]$	$[0.0150, 0.1378]$	$[0.0000, 0.0001]$	$[0.0153, 0.1379]$	0.0009	0.0253	1×10^{-9}	0.0277
$\boldsymbol{\theta}_{new}$ (with UM_{z1})	$[0.0000, 0.0004]$	$[0.0025, 0.0141]$	$[0.0011, 0.0059]$	$[0.0040, 0.0154]$	3×10^{-9}	2×10^{-5}	0.0001	0.0001
$\boldsymbol{\theta}_{final}$ (with UM^2_{y1})	$[0.0000, 0.0171]$	$[0.0137, 0.1305]$	$[0.0000, 0.0001]$	$[0.0142, 0.1306]$	0.0010	0.0202	1×10^{-9}	0.0220
$\boldsymbol{\theta}_{final}$ (with UM_{z1})	$[0.0000, 0.0002]$	$[0.0024, 0.0143]$	$[0.0000, 0.0013]$	$[0.0028, 0.0144]$	2×10^{-9}	1×10^{-5}	4×10^{-7}	2×10^{-5}

Table 14: Reliability analysis results for the different design points $\boldsymbol{\theta}$ with UM^2_{y1} and UM_{z1}.

Refinement round	1	2	3	4
Refinement type	$\{e_4^-\}$	$\{e_3^-\}$	$\{e_4^-\}$	$\{e_2^+\}$

Table 15: Summary of the type of refinements requested for the respective refinement round. The superscripts "+" and "−" denote the refinement of the upper and lower-bound of the epistemic parameter respectively.

References

[1] W. L. Oberkampf, J. C. Helton, C. A. Joslyn, S. F. Wojtkiewicz, S. Ferson, Challenge problems: Uncertainty in system response given uncertain parameters, Reliability Engineering and System Safety 85 (2004) 11–19. doi:10.1016/j.ress.2004.03.002.

[2] C. J. Roy, W. L. Oberkampf, A comprehensive framework for verification, validation, and uncertainty quantification in scientific computing, Computer Methods in Applied Mechanics and Engineering 200 (2011) 2131–2144. doi:10.1016/j.cma.2011.03.016.

[3] L. G. Crespo, S. P. Kenny, The NASA Langley UQ Challenge on Optimization Under Uncertainty, Mechanical Systems and Signal Processing 152 (2021) 107405. doi:10.1016/j.ymssp.2020.107405.

[4] A. D. Kiureghian, O. Ditlevsen, Aleatory or epistemic? Does it matter?, Structural Safety 31 (2009) 105–112. doi:10.1016/j.strusafe.2008.06.020.

[5] Z. Meng, Y. Pang, Y. Pu, X. Wang, New hybrid reliability-based topology optimization method combining fuzzy and probabilistic models for handling epistemic and aleatory uncertainties, Computer Methods in Applied Mechanics and Engineering 363 (2020) 112886. doi:10.1016/j.cma.2020.112886.

[6] R. Rocchetta, M. Broggi, E. Patelli, Do we have enough data? Robust reliability via uncertainty quantification, Applied Mathematical Modelling 54 (2018) 710–721. doi:10.1016/j.apm.2017.10.020.

[7] L. G. Crespo, S. P. Kenny, D. P. Giesy, The NASA Langley Multidisciplinary Uncertainty Quantification Challenge, In the Proceedings of the 16th AIAA Non-Deterministic Approaches Conference 1. doi:10.2514/6.2014-13475.

[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 (2015) 140–169. doi:10.2514/1.i010273.

[9] J. L. Beck, L. S. Katafygiotis, Updating Models and Their Uncertainties. I: Bayesian Statistical Framework, Journal of Engineering Mechanics 124 (1998) 455–461. doi:10.1061/(asce)0733-9399(1998)124:4(455).

[10] L. S. Katafygiotis, J. L. Beck, Updating Models and Their Uncertainties. II: Model Identifiability, Journal of Engineering Mechanics 124 (1998) 463–467. doi:10.1061/(asce)0733-9399(1998)124:4(463).

[11] A. Lye, A. Cicirello, E. Patelli, Sampling methods for solving Bayesian model updating problems: A tutorial, Mechanical Systems and Signal Processing 159 (2021) 107760. doi:10.1016/j.ymssp.2021.107760.

[12] J. L. Beck, S. K. Au, Bayesian Updating of Structural Models and Reliability using Markov Chain Monte Carlo Simulation, Journal of Engineering Mechanics 128 (2002) 380–391. doi:10.1061/(ASCE)0733-9399(2002)128:4(380).

[13] J. Y. Ching, Y. C. Chen, Transitional Markov Chain Monte Carlo Method for Bayesian Model Updating, Model Class Selection, and Model Averaging, Journal of Engineering Mechanics 133 (7). doi:10.1061/(ASCE)0733-9399(2007)133:7(816).

[14] A. Gray, A. Wimbush, M. D. Angelis, P. O. Hristov, E. Miralles-Dolz, D. Calleja, R. Rocchetta, Bayesian Calibration and Probability Bounds Analysis Solution to the NASA 2020 UQ Challenge on Optimization Under Uncertainty, In Proceedings of the 30^{th} European Safety and Reliability Conference and the 15^{th} Probabilistic Safety Assessment and Management Conference 1. doi:10.3850/978-981-14-8593-0_5520-cd.

[15] 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. doi:10.1016/j.ymssp.2018.08.017.

[16] A. B. Abdessalem, N. Dervilis, D. Wagg, K. Worden, Model selection and parameter estimation in structural dynamics using approximate Bayesian computation, Mechanical Systems and Signal Processing 99 (2018) 306–325. doi:10.1016/j.ymssp.2017.06.017.

[17] B. M. Turner, T. V. Zandt, A tutorial on approximate Bayesian computation, Journal of Mathematical Psychology 56 (2) (2012) 69–85. doi:10.1016/j.jmp.2012.02.005.

[18] M. Heideman, D. Johnson, C. Burrus, Gauss and the history of the fast fourier transform, IEEE ASSP Magazine 1 (1984) 14–21. doi:10.1109/MASSP.1984.1162257.

[19] Y. Bai, Z. Huang, H. Lam, A Distributionally Robust Optimization Approach to the Nasa Langley Uncertainty Quantification Challenge, In Proceedings of the 30^{th} European Safety and Reliability Conference and the 15^{th} Probabilistic Safety Assessment and Management Conference 1. doi:10.3850/978-981-14-8593-0_5639-cd.

[20] V. M. Panaretos, Y. Zemel, Statistical aspects of wasserstein distances, Annual Review of Statistics and Its Application 6 (2019) 405–431. doi:10.1146/annurev-statistics-030718-104938.

[21] S. Ferson, W. Oberkampf, L. Ginzburg, Model validation and predictive capability for the thermal challenge problem, Computer Methods in Applied Mechanics and Engineering 197 (2008) 2408–2430. doi:10.1016/J.CMA.2007.07.030.

[22] L. G. Crespo, S. P. Kenny, D. P. Giesy, B. K. Stanford, Random variables with moment-matching staircase density functions, Applied Mathematical Modelling 64 (2018) 196–213. doi:10.1016/j.apm.2018.07.029.

[23] M. Kitahara, S. F. Bi, M. Broggi, M. Beer, Nonparametric Bayesian stochastic model updating with hybrid uncertainties, Mechanical Systems and Signal Processing 163 (2022) 108195. doi:10.1016/j.ymssp.2021.108195.

[24] M. Kitahara, S. F. Bi, M. Broggi, M. Beer, Bayesian Model Updating in Time Domain with Metamodel-based Reliability Method, ASCE-ASME Journal of Risk and Uncertainty in Engineering Systems, Part A: Civil Engineering 7. doi:10.1061/AJRUA6.0001149.

[25] A. W. Bowman, A. Azzalini, *Applied Smoothing Techniques for Data Analysis*, OUP Oxford, 1997, ISBN: 978-0191545696.

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

[27] S. Ferson, V. Kreinovick, L. Ginzburg, F. Sentz, Constructing Probability Boxes and Dempster-Shafer Structures, Sandia National Laboratories 4015 (2002) . doi:10.2172/809606.

[28] M. Beer, I. A. Kougioumtzoglou, E. Patelli, Emerging Concepts and Approaches for Efficient and Realistic Uncertainty Quantification, Maintenance and Safety of Aging Infrastructure (2014) 121–162. doi:10.1201/b17073-5.

[29] W. T. Tucker, S. Ferson, Sensitivity in risk analyses with uncertain numbers, Sandia National Laboratories 2801. doi:10.2172/886899.

[30] S. Ferson, W. T. Tucker, Sensitivity analysis using probability bounding, Reliability Engineering and System Safety 91 (10-11) (2006) 1435–1442. doi:10.1016/j.ress.2005.11.052.

[31] T. Ali, Modeling uncertainty in risk assessment using Double Monte Carlo method, International Journal of Engineering and Innovative Technology 1 (2012) 114–118.

[32] R. Rocchetta, M. Broggi, E. Patelli, Do we have enough data? Robust reliability via uncertainty quantification, Applied Mathematical Modelling 54 (2018) 710–721. doi:10.1016/j.apm.2017.10.020.

[33] J. Song, P. Wei, M. Valdebenito, S. Bi, M. Broggi, M. Beer, Z. Lei, Generalization of non-intrusive imprecise stochastic simulation for mixed uncertain variables, Mechanical Systems and Signal Processing 134 (2019) 106316. doi:10.1016/j.ymssp.2019.106316.

[34] P. Wei, Z. Lu, J. Song, Extended Monte Carlo Simulation for Parametric Global Sensitivity Analysis and Optimization, AIAA Journal 52 (2014) 867–878. doi:10.2514/1.j052726.

[35] K. Rezaie, M. Amalnik, A. Gereie, B. Ostadi, M. Shakhseniaee, Using extended Monte Carlo simulation method for the improvement of risk management: Consideration of relationships between uncertainties, Applied Mathematics and Computation 190 (2007) 1492–1501. doi:10.1016/j.amc.2007.02.038.

[36] I. Sobol', Global sensitivity indices for nonlinear mathematical models and their Monte Carlo estimates, Mathematics and Computers in Simulation 55 (2001) 271–280. doi:10.1016/s0378-4754(00)00270-6.

[37] A. Saltelli, M. Ratto, T. Andres, F. Campolongo, J. Cariboni, D. Gatelli, M. Saisana, S. Tarantola, *Global Sensitivity Analysis. The Primer*, Wiley, 2008, ISBN: 978-0470725177.

[38] R. B. Nelsen, *An Introduction to Copulas*, 2nd Edition, Springer-Verlag New York, 2006, ISBN: 978-0387286594.