Proceedings of the 35th European Safety and Reliability & the 33rd Society for Risk Analysis Europe Conference
Edited by Eirik Bjorheim Abrahamsen, Terje Aven, Frederic Bouder, Roger Flage, Marja Ylönen
©2025 ESREL SRA-E 2025 Organizers. *Published by* Research Publishing, Singapore.
doi: 10.3850/978-981-94-3281-3_ESREL-SRA-E2025-P9016-cd

Taking on the NASA and DNV Challenge 2025: Bayesian Calibration and Optimization under Hybrid Uncertainty

Nataly Manque[‖,¶], Lukas Fritsch[‡,†,¶], Chao Dang[‖,¶], Jan Grashorn[§,¶], Zhouzhou Song[‖], Marius Bittner[‡], Thomas Potthast[‡], Pei-Pei Li[‖], Marcos Valdebenito[‖], Matteo Broggi[*,‡], Matthias Faes[‖,#]

[‖]Chair for Reliability Engineering, TU Dortmund University, Dortmund 44227, Germany
[‡]Institute for Risk and Reliability, Leibniz University Hannover, Hannover 30167, Germany
[†]International Research Training Group 2657 (IRTG 2657): Computational Mechanics Techniques in High Dimensions, Leibniz University Hannover, Hannover 30167, Germany
[§]Chair of Engineering Materials and Building Preservation, Helmut-Schmidt-University, Hamburg 22159, Germany
[#]International Joint Research Center for Engineering Reliability and Stochastic Mechanics, Tongji University, Shanghai 200092, China

[*]: Corresponding author: broggi@irz.uni-hannover.de,
[¶]: These authors contributed equally to this work.

This paper addresses the NASA and DNV challenge on optimization under uncertainty, where participants were tasked with calibrating the uncertainty models of aleatory and epistemic parameters of an unknown system using a computational model and synthetic data, and identifying control parameters for different objectives. We present two approaches for model calibration, namely Bayesian optimization and sequential Bayesian updating. Additionally, a reliability-based optimization scheme based on a Bayesian approach and subset simulation is used to tackle a design optimization problem.

Keywords: Bayesian Model Updating, Bayesian Optimization, Approximate Bayesian Computation, Optimal Design under Hybrid Uncertainty, Subset Simulation

1. Introduction

Safety-critical systems frequently operate in harsh environments characterized by significant uncertainties. Organizations such as DNV and NASA are often challenged by limited quantitative data, complex and uncertain operating conditions, whether in the development of advanced engineering systems or ongoing monitoring and verification. Motivated by these challenges, NASA Langley Research Center and the DNV Group Research and Development have initiated the Challenge on Optimization under Uncertainty (see Agrell et al., 2024), which seeks to advance methodologies in uncertainty quantification (UQ) and design optimization.

The challenge framework is deliberately discipline-independent, allowing researchers to focus on two key aspects: constructing robust uncertainty models (UMs) and optimizing design parameters under mixed aleatory and epistemic uncertainties. Aleatory uncertainty, stemming from inherent variability, and epistemic uncertainty, due to incomplete knowledge, both play critical roles in accurately representing the behavior of complex systems. Participants are provided with a simulation model alongside (synthetic) real system data, which collectively serve as the basis for updating and calibrating these UMs. After obtaining the UMs for these aleatory and epistemic parameters, the optimal control parameters need to be specified according to different performance and risk requirements under mixed uncertainties.

In our response, we first present two approaches to calibrate the UMs of the aleatory and epistemic parameters. These models form the foundation for our subsequent design optimization. Then, a

Kriging surrogate model-based design optimization scheme is developed to tackle the second problem of the challenge.

2. Problem 1: Uncertainty quantification

Before diving into calibrating the UMs for the aleatory and epistemic parameters, a few general remarks are given. First, it is assumed that each time-varying response of the system (both computational model and synthetic real system) is stationary, which means time-dependency is not considered. Second, it is noteworthy that the data exhibits strong dependencies in the first three outputs $(y_{1,2,3})$ and the second three outputs $(y_{4,5,6})$. Therefore, the data can be split into these two groups and a separate distance measure can be defined for each group. Exemplary outputs are shown as bi-variate plots in Fig. 1. For $(y_{1,2,3})$, the data is concentrated in a three-dimensional manifold, whereas $(y_{4,5,6})$ are linearly dependent. For this reason, only y_4 and y_5 are shown on the figure. Changing the control parameters $\mathbf{X}_c$ affects the location and shape of the manifold, as well as the angle and magnitude of the vector that is spanned by $(y_{4,5,6})$. We note that $\mathbf{X}_e$ changes the angle when conditioned on a fixed value of $\mathbf{X}_c$.

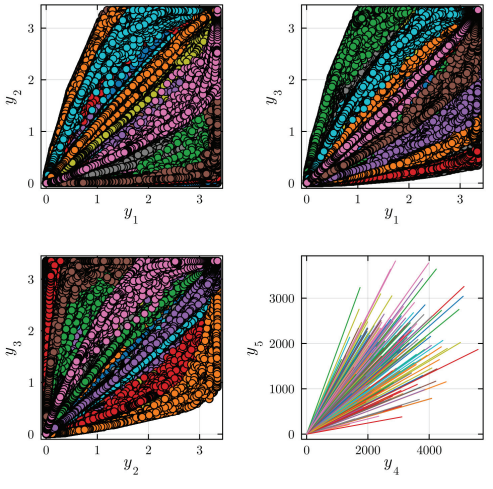

Fig. 1. Bi-variate outputs y for random samples of $\mathbf{X}_a$, $\mathbf{X}_e$. Colors indicate different values of $\mathbf{X}_c$.

2.1. *Data generation*

Since each team was only allowed to request $N = 10$ datasets in addition to the initial example dataset, it is crucial to carefully select the control variable $\mathbf{X}_c^{(i)}$ to maximize information gain for calibrating the UMs. In this study, three datasets are generated corresponding to the following control parameters:

$\mathbf{X}_c^{(0)} = [0.5330, 0.6660, 0.5000]$ (example)
$\mathbf{X}_c^{(1)} = [0.6063, 0.1242, 0.5856]$
$\mathbf{X}_c^{(2)} = [0.9775, 0.0038, 0.8015]$
$\mathbf{X}_c^{(3)} = [0.1201, 0.1715, 0.2152]$

The first control variable, $\mathbf{X}_c^{(1)}$, is chosen by maximizing the generalized first-order Sobol' indices with respect to $\mathbf{X}_a$, where $\mathbf{X}_a$ and $\mathbf{X}_e$ are assumed to be independent uniform random variables $U(0, 1)$, and the random seed ω is treated as a discrete uniform random variable $[0, 2^{32} - 1]$. $\mathbf{X}_c^{(2)}$ is chosen based on the maximum sensitivity of the outputs to the control parameters. Based on the first UM, samples are drawn for $\mathbf{X}_a$ and $\mathbf{X}_e$, then the output sensitivity is calculated by standardizing the outputs and subsequently computing the variance. $\mathbf{X}_c^{(3)}$ is identified similarly to $\mathbf{X}_c^{(1)}$ using the UMs of $\mathbf{X}_a$ and $\mathbf{X}_e$ calibrated by the data corresponding to $\mathbf{X}_c^{(1)}$.

2.2. *Subproblem 1.1: Uncertainty Models*

For the calibration of the UMs, two different approaches are presented. The first one is based on Bayesian Optimization (BO) and the second one uses sequential Bayesian updating.

2.2.1. *Uncertainty Model A: Bayesian optimization with multi-source data*

In this approach, we assume that $\mathbf{X}_a$ and $\mathbf{X}_e$ are mutually independent, with $\mathbf{X}_e$ following independent beta distributions and $\mathbf{X}_a$ following marginal beta distributions with a bivariate Gaussian copula. A probability distribution is assigned to the epistemic parameters to allow them being inferred in a Bayesian way.

Therefore, eleven parameters need to be inferred to determine the UMs of $\mathbf{X}_a$ and $\mathbf{X}_e$. The parameters include ten Beta distribution parameters for $X_{a1}, X_{a2}, X_{e1}, X_{e2}, X_{e3}$, as well as the Gaussian copula parameter ρ, which cap-

tures the correlation between X_{a1} and X_{a2}. These are collectively represented by $\boldsymbol{\theta} = [\alpha_{a1}, \beta_{a1}, \alpha_{a2}, \beta_{a2}, \alpha_{e1}, \beta_{e1}, \alpha_{e2}, \beta_{e2}, \alpha_{e3}, \beta_{e3}, \rho]$. To determine these parameters, we employ an optimization-based approach. The optimization objective is to minimize the weighted K-L divergence:

$$L(\boldsymbol{\theta}) = \sum_{i=1}^{3} \gamma_i D_{\mathrm{KL}}(P_{\mathrm{real}}^{(i)}(\boldsymbol{y}) \parallel P_{\mathrm{sim}}^{(i)}(\boldsymbol{y}|\boldsymbol{\theta})), \quad (1)$$

where γ_i are the weights of different K-L divergences and equal weights are used here. $P_{\mathrm{real}}^{(i)}$ and $P_{\mathrm{sim}}^{(i)}$ represent the multivariate PDF estimated from the histogram of real data $\mathcal{Y}^{(i)}$ and the simulated data $\mathbf{Y}^{(i)}$ generated by the computational model corresponding to the i-th control parameters $\mathbf{X}_c^{(i)}$, respectively. Then, the optimization formulation for inferring $\boldsymbol{\theta}$ is expressed as:

$$\min_{\boldsymbol{\theta}} \; L(\boldsymbol{\theta}), \text{ s.t. } \boldsymbol{\theta} \in [\boldsymbol{\theta}_L, \boldsymbol{\theta}_U] \quad (2)$$

where the lower bound of the parameters $\boldsymbol{\theta}_L$ is set as $[0.01, 0.01, 0.01, 0.01, 1, 1, 1, 1, 1, 1, -1]$ and $\boldsymbol{\theta}_U$ is set as $[50, 50, 50, 50, 5000, 5000, 5000, 5000, 5000, 5000, 1]$ here. To solve the optimization problem efficiently, Bayesian optimization (Jones et al., 1998) is used. Note that the random seed ω is treated as a discrete uniform random variable $[0, 2^{32} - 1]$.

After the Bayesian optimization, the solution of the optimization problem in Eq. (2) is given as: $\hat{\boldsymbol{\theta}} = [6.44, 18.87, 8.15, 22.40, 2388.11, 4867.03, 3843.83, 2253.14, 1988.92, 2154.10, 0.88]$. We denote this calibrated UM for aleatory and epistemic inputs as $\mathrm{UM_A}$. Fig. 2 shows the calibrated joint distribution $\mathbf{X}_a \sim f_a$, and Fig. 3 shows the calibrated $\mathbf{X}_e$. Using the modes of the final calibrated beta distributions, the best guess of the true value of $\mathbf{X}_e$ is $\mathbf{X}_e^* = [0.3291, 0.6305, 0.4801]$. From the results of Fig. 3, the 95% confidence intervals for $\mathbf{X}_e$ are $X_{e1} = [0.3184, 0.3400]$, $X_{e2} = [0.6183, 0.6425]$, and $X_{e3} = [0.4649, 0.4953]$.

The example dataset is used to validate the calibrated $\mathrm{UM_A}$. The comparison of the real and simulated data using the calibrated UM is shown in Fig. 4. The system response generally matches the data. Further, the comparison by fixing the

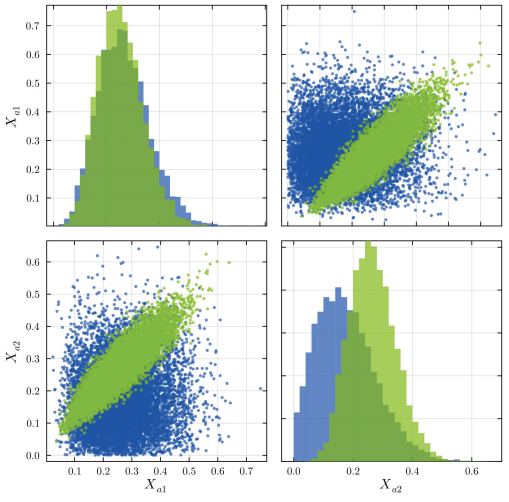

Fig. 2. Comparison of $\mathbf{X}_a \sim f_a$ calibrated in $\mathrm{UM_A}$ (green) and $\mathrm{UM_B}$ (blue).

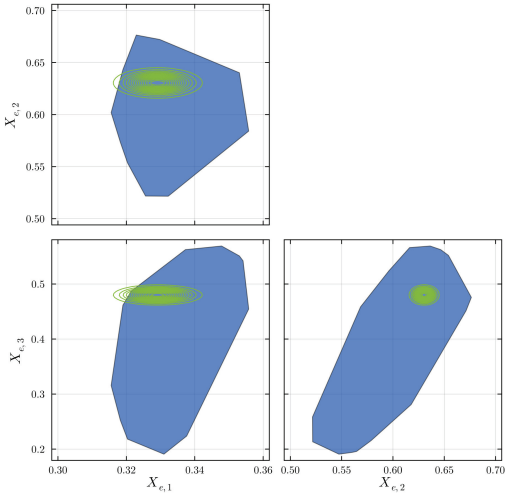

Fig. 3. Comparison of $\mathbf{X}_e \in E$ calibrated in $\mathrm{UM_A}$ (green contour) and $\mathrm{UM_B}$ (shaded blue) for $\mathbf{X}_e$.

epistemic parameters at $\mathbf{X}_e^*$ is shown in Fig. 5. Here, a better match can be observed.

2.2.2. *Uncertainty Model B: Sequential approximate Bayesian computation*

In the second approach, the calibration of the input model parameters is based on a sequential Transitional Markov Chain Monte Carlo (TMCMC) method (see Ching and Chen,

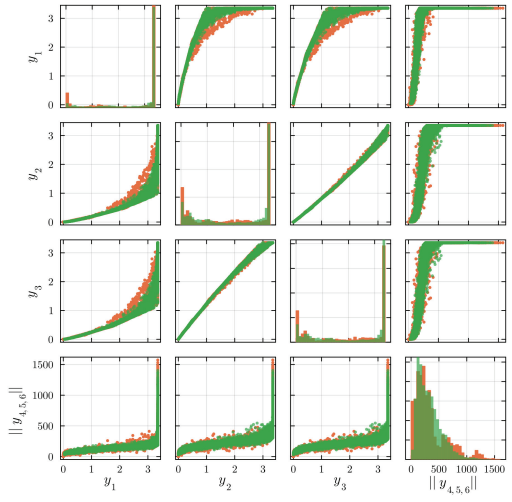

Fig. 4. Comparison of the model output using the calibrated $\mathrm{UM_A}$ varying $\mathbf{X}_e \in E$ (green) and data (orange).

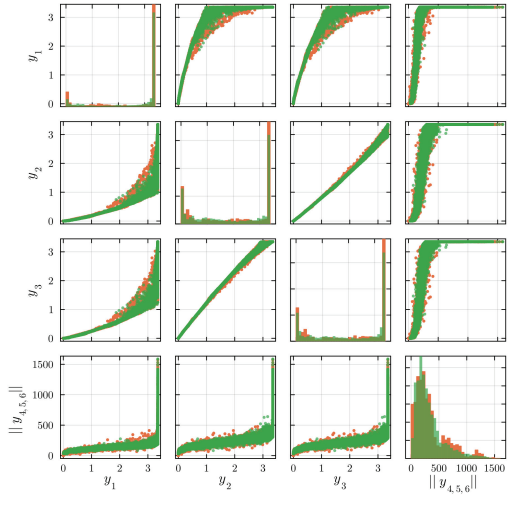

Fig. 5. Comparison of the model output using the calibrated $\mathrm{UM_A}$ fixing $\mathbf{X}_e$ at $\mathbf{X}_e^*$ (green) and data (orange).

2007). First, the epistemic parameters $\mathbf{X}_e$ and then the aleatory parameters $\mathbf{X}_a$ are sequentially updated using an implementation based on the `UncertaintyQuantification.jl` (Behrensdorf et al., 2025) `Julia` package. In this regard, both aleatory and epistemic parameter groups $\mathbf{X}_a$ and $\mathbf{X}_e$ are treated independently as an underlying assumption. This is based on the distinct effects that these parameters exhibit on the model output, as observed in prior analyses mentioned in the beginning of Section 2.

First, the epistemic input parameters $\boldsymbol{\theta}_e = \mathbf{X}_e$ are updated sequentially using $N = 5000$ samples starting with the control parameters $\mathbf{X}_c^{(0)}$ and the corresponding dataset. The linear relationship identified for the outputs $y_{4,5,6}$ (see Fig. 1) is used to calibrate $\mathbf{X}_e$. We infer the epistemic parameters for a fixed set of control parameters by identifying the direction of this line. We assume a uniform prior for all epistemic parameters $\mathbf{X}_e \sim U(0,1)$ in the a first updating sequence. To capture the information given in a dataset for the i-th control variable, $\mathcal{Y}_{4,5,6}^{(i)}$, Approximated Bayesian Computation (ABC) with a Gaussian likelihood function is used:

$$\mathcal{L} \propto \exp\left[-\left(d/\varepsilon\right)^2\right]. \tag{3}$$

Here, d describes the discrepancy between simulated output and real data and ε is a scale factor. In this regard, the model output $\mathbf{Y}_{4,5,6}^{(i)}$ and data $\mathcal{Y}_{4,5,6}^{(i)}$ are processed such that every realization of the output is interpreted as a position vector normalized by its length. Based on this, the Euclidean distance between both unit vectors $\hat{\mathbf{Y}}_{4,5,6}^{(i)}$ and $\hat{\mathcal{Y}}_{4,5,6}^{(i)}$ is used as the discrepancy metric in Eq. (3) along with $\varepsilon = 10^{-3}$. For the second and third updating sequences, the control parameters $\mathbf{X}_c^{(1)}$ and $\mathbf{X}_c^{(2)}$ are used along with the corresponding computational model outputs and datasets. We change the prior distribution to a Gaussian mixture model for both updating steps to approximate the resulting posterior input samples of $\boldsymbol{\theta}_e$ after the first updating sequence. TMCMC and the Gaussian ABC log-likelihood setup are again utilized. Through a three-step sequential updating process, we derive a hyper-rectangular set of $\mathbf{X}_e$, depicted by the convex hulls of the posterior samples in Fig. 3. This method yields the best estimate for the true epistemic parameter as $\mathbf{X}_e^* = [0.3349, 0.6083, 0.4234]$ using the maximum likelihood estimate.

Subsequently, the distributions for both aleatory parameters $\mathbf{X}_a$ are inferred from the modified output data $\mathcal{Y}_{1,2,3,||4,5,6||}$. By observing that the last three outputs align on a line, we combine them

using the Euclidean norm, thereby reducing the output dimensionality from six to four. The TM-CMC method is employed again, with $N = 500$ samples and a Gaussian ABC log-likelihood (with $\varepsilon = 10^{-1}$). During the aleatory updating, the epistemic parameter remains fixed at its prior MLE. To identify aleatory parameter distributions, we apply a double-loop approach as detailed by Bi et al. (2019). In this approach, both marginals are modeled as Beta distributions: $\mathbf{X}_{a_i} \sim \text{Beta}(\alpha_i, \beta_i)$. The distribution parameters α_i and β_i, considered as additional epistemic parameters, are initially assumed to follow $\alpha_i, \beta_i \sim U(0, 25)$. This setup forms a two-layer structure, with each (α_i, β_i) realization yielding a marginal PDF for X_{a_i}. Dependencies among $\mathbf{X}_a$ are captured by a Gaussian copula with $\rho \sim U(-1, 1)$. Thus, five hyperparameters are updated as $\boldsymbol{\theta}_a = [\alpha_1, \alpha_2, \beta_1, \beta_2, \rho]$. The discrepancy between the data $\mathcal{Y}_{1,2,3,||4,5,6||}$ and simulations $\mathbf{Y}_{1,2,3,||4,5,6||}$ is measured using the Bhattacharyya distance, employing a discrete version with the binning algorithm (Bi et al., 2019) and $N_{\text{bins}} = 20$.

In subsequent updates, the prior PDF of $\boldsymbol{\theta}_a$ is replaced by a Gaussian mixture model, fit with posterior samples from the first step. Output data for further updates is derived from the control parameters $\mathbf{X}_c^{(1)}$ and $\mathbf{X}_c^{(2)}$. This process results in the final uncertainty model for $\mathbf{X}_a$, shown in Fig. 2, with 100 realizations of the fitted parametric p-box, with each realization comprising 100 samples.

Together with the set of epistemic parameters presented in Fig. 3, this forms the second uncertainty model, denoted as UM_B. To validate the calibrated model, we compare the data generated by $\mathbf{X}_c^{(3)}$ with the outputs derived from $\mathbf{X}_e^*$ and the five parameters defining the joint distribution for $\mathbf{X}_a$. Fig. 6 illustrates a comparison between the data (orange) and the model response (blue).

2.2.3. *Comparison of the uncertainty models*

We note that generally, the BO approach identifies much narrower distributions than the sequential approach. This is mostly because the sequential approach uses hybrid uncertainties with parameterized beta distributions for $\mathbf{X}_a$, while the BO approach directly updates the parameter's distributions. Therefore, as is shown in Figs. 2 and 3, UM_A can be seen as subset of UM_B, at least for the aleatory parameters. Further, the marginal distributions of X_{a1} match well. The discrepancies in the epistemic parameters are larger, with UM_A again having narrower support than UM_B. Nevertheless, for the most likely value of $\mathbf{X}_e$, both methods show very similar results.

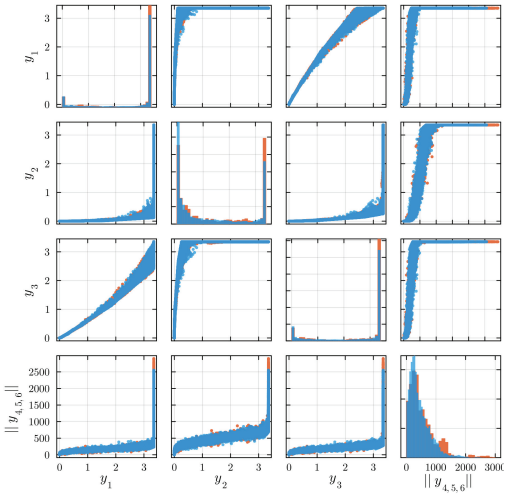

Fig. 6. Comparison of the model output using the calibrated UM_B fixing $\mathbf{X}_e$ at $\mathbf{X}_e^*$ (blue) and data (orange).

2.3. *Subproblem 1.3: Output Intervals*

Following the calibration of UMs, the next task is to find the tightest output bounds that suffice a confidence level α for all simulated outputs. We use $\mathbf{X}_a$ and $\mathbf{X}_e$ per UM_B as detailed in Sec. 2.2.1, with $\mathbf{X}_c^{(2)}$ listed in Sec. 2.1. We employ a double-loop sampling approach, as in Sec. 2.2.2. For configurations $i = 1, \ldots, n_{ae}$, we sample distribution parameters and realizations of $\mathbf{X}_e$, and for each i, generate $m = 1, \ldots, n$ samples. This produces an output tensor $\mathbf{Y}_{imtj}$, with i as the outer loop, m as the inner loop, t as time, and j as the output index ($j = 1, \ldots, n_y$). To obtain global output bounds within a prescribed probability α, we estimate confidence intervals simply by:

$$\begin{aligned} l(i,j) &= Q_{1-\alpha}(\min_t Y_{ij}) \\ u(i,j) &= Q_{\alpha}(\max_t Y_{ij}) \end{aligned} \tag{4}$$

where $Q_\alpha(x)$ is the α-quantile of x. The minimum and maximum outputs over time for each configuration i are determined using statistics from the inner loop samples. The following approach is then implemented to identify the bounds: Calculate an elementwise distance, normalize it, and apply a row-wise norm. Finally, the bounds are determined by the norm maximizing element i^*, i.e. $l_{\text{local}}(j) = l(i^*, j)$ for the lower-, and $u_{\text{local}}(j) = u(i^*, j)$ for the upper bound.

To obtain the results of different α values, 10^5 samples have been analyzed. For each of the estimated bounds, by design, the requirements on the confidence level are satisfied. Some results can be seen in Tab. 1. However, the results are varying for different control parameters, suggesting that the amount of information on the bounds was still not enough.

Table 1. Prediction intervals for $\mathbf{X}_c^{(2)}$.

n_y	$\alpha = 0.95$		$\alpha = 0.99$	
	l_{local}	u_{local}	l_{local}	u_{local}
1	0.0149	3.35	0	3.35
2	0.0046	3.35	0	3.35
3	0.0918	3.35	0	3.35
4	94.12	2288.30	9.113	4153.32
5	79.05	1922.01	7.654	3488.49
6	22.50	546.93	2.178	992.699

3. Problem 2: Design optimization

Building upon the uncertainty calibration and propagation, Problem 2 addresses optimizing the design with respect to the control variables $\mathbf{X}_c$ under diverse objectives. The goal is either to maximize performance, represented by $J(\mathbf{X}_e, \mathbf{X}_c)$, or to minimize the system's failure probability, $\text{pof}_{\text{sys}}(\mathbf{X}_e, \mathbf{X}_c)$. Given the high computational demands associated with reliability-based design optimization (RBDO), which escalate with hybrid uncertainties, we aim to mitigate these using Kriging surrogate models within the Bayesian optimization framework (Jones et al., 1998).

We use UM_B as baseline model for the optimization since the uncertainty bounds are wider than in UM_A. This increases the associated uncertainty and we expect to obtain conservative and robust results, which is crucial for safety-cricital systems. By sampling from UM_B for $\mathbf{X}_a$ and $\mathbf{X}_e$, and generating Sobol' samples for $\mathbf{X}_c$, we acquire outputs for $J(\mathbf{X}_e, \mathbf{X}_c)$ and $\text{pof}_{\text{sys}}(\mathbf{X}_e, \mathbf{X}_c)$. These samples are used to train the surrogate model for optimizing the design. In problem 2, we consider the following five objectives for the optimization:

I. Performance-based Design: Maximize J.
II. Reliability-based Design: Minimize pof_{sys}.
III. ϵ-Constrained Design: Maximize J subject to $\text{pof}_{\text{sys}} \leq \epsilon = 10^{-3}$.
IV. ϵ-Constrained Design: Maximize J subject to $\text{pof}_{\text{sys}} \leq \epsilon = 10^{-4}$.
V. Risk-based Design: Maximize J subject to $\mathbb{E}\left[h \mid h < 0\right] \geq -300$.

In the last objective $\mathbb{E}\left[h \mid h < 0\right]$ denotes the expectation of limit state function conditioned on values in the failure domain, $h < 0$. In contrast to the failure probability pof_{sys}, this quantity serves as a measure for the extent of overshoot of the threshold. Detailed descriptions of the different objectives are available in Agrell et al. (2024).

The results of Case I, the performance-based design, are summarized in Fig. 7. In this figure, the initial set of 128 points, used to construct the metamodel, as well as the additional 20 points added through Bayesian optimization, are displayed. Each point corresponds to a realization of $\mathbf{X}_c \in [0,\ 1]^3$, and the color in Fig. 7 represents the performance metric $\min_{\mathbf{X}_e \in E} J(\mathbf{X}_e, \mathbf{X}_c)$.

This performance metric is calculated via a double-loop Monte Carlo setup. In the first loop, $n_{ae} = 100$ realizations of $\mathbf{X}_e \in E$, along with samples of the parameterized p-boxes, are generated. In the second loop, for each realization of the p-box, $n = 1000$ samples of $\mathbf{X}_a \sim f_a$ are drawn. The optimized performance-based design, denoted as $\mathbf{X}_{c,\text{I}}^*$, is marked by an orange diamond in the figure. A comprehensive comparison of all objectives is presented in Tab. 2.

For the reliability-based design, we aim to identify control variables that minimize the failure probability, denoted as pof_{sys}. To derive a design that meets this criterion, represented by $\mathbf{X}_{c,\text{II}}^*$, we initially generate 64 samples of control variables

Table 2. Comparison of the identified designs and objectives.

i	Objective	Control Variable $\mathbf{X}^*_{c,i}$	Interval J	Interval pof_{sys}
I	Performance-based	$[0.6455, 0.3245, 0.4937]$	$[5.674,\ 9.097]$	$[1.50 \cdot 10^{-7},\ 7.39 \cdot 10^{-4}]$
II	Reliability-based	$[0.4466, 0.5879, 0.6190]$	$[5.467,\ 9.110]$	$[2.94 \cdot 10^{-12}, 1.38 \cdot 10^{-5}]$
III	ϵ-Constrained ($\epsilon = 10^{-3}$)	$[0.6455, 0.3245, 0.4937]$	$[5.674,\ 9.097]$	$[1.50 \cdot 10^{-7},\ 7.39 \cdot 10^{-4}]$
IV	ϵ-Constrained ($\epsilon = 10^{-4}$)	$[0.4466, 0.5879, 0.6190]$	$[5.467,\ 9.110]$	$[2.94 \cdot 10^{-12}, 1.38 \cdot 10^{-5}]$
V	Risk-based	$[0.3805, 0.1852, 0.5367]$	$[5.175,\ 8.621]$	$[6.76 \cdot 10^{-8},\ 5.89 \cdot 10^{-4}]$

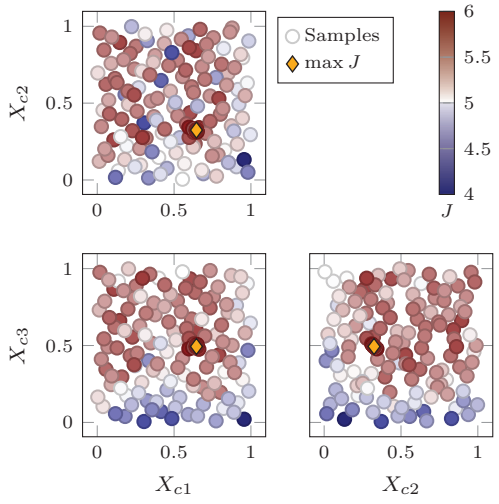

Fig. 7. Samples of $\mathbf{X}_c$ with performance-based design objective values $\min_{\mathbf{X}_e \in E} J(\mathbf{X}_e, \mathbf{X}_c)$ and optimization result from Bayesian optimization.

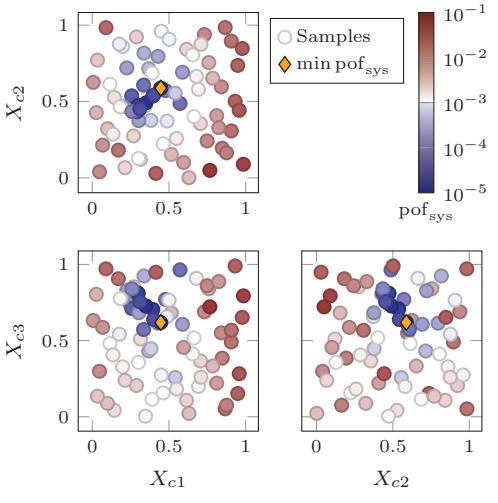

Fig. 8. Samples of $\mathbf{X}_c$ with reliability-based design objective $\max_{\mathbf{X}_e \in E} \text{pof}_{\text{sys}}(\mathbf{X}_e, \mathbf{X}_c)$ and optimization result from Bayesian optimization.

using Sobol's sampling. Subsequently, we estimate the imprecise failure probability for each control variable.

Subset simulation (Au and Beck, 2001) is employed to estimate the range of failure probabilities for each control variable, considering $n_{ae} = 100$ different realizations of $\mathbf{X}_e$. From this, we obtain the worst-case estimate, specifically $\max_{\mathbf{X}_e \in E} \text{pof}_{\text{sys}}$, which is depicted in Fig. 8. Additionally, 20 new samples are incorporated through Bayesian optimization to identify the global minimum.

The surrogate models constructed to find the respective extrema in the first two optimization problems are used for the constrained optimization in tasks III and IV. As seen in Tab. 2, the performance-based design $\mathbf{X}^*_{c,\text{I}}$ satisfies the constraint $\max \text{pof}_{\text{sys}} \leq 10^{-3}$. Hence, this global optimum also serves as the solution for constrained optimization. The constraint in problem IV, $\max \text{pof}_{\text{sys}} \leq 10^{-4}$, is not satisfied by $\mathbf{X}^*_{c,\text{I}}$. However, we find that the design $\mathbf{X}^*_{c,\text{II}}$ maximizes J under the ϵ-constraint. For the final objective, we calculate the expected value $\mathbb{E}[h \mid h < 0]$ from the data generated with the subset simulations employed in task II. We then construct a Kriging surrogate model for the expected value based on this data and use this model, along with the one for J, to perform the constrained optimization.

A comprehensive summary of the different objectives and results can be found in Tab. 2. In addition to the identified control variables, the table presents the ranges of the performance metric J and the failure probability pof_{sys}. All identified designs satisfy the requirement $\max \text{pof}_{\text{sys}} \leq 10^{-3}$. However, only one design satisfies the stricter

requirement $\max \text{pof}_{\text{sys}} \leq 10^{-4}$, namely the reliability-based design.

When comparing the ranges of J, it becomes clear that they are within the same order of magnitude. Notably, the reliability-based design represents a good compromise: while the worst-case J is only slightly smaller than the performance-based design, the failure probability is significantly reduced. Moreover, the risk-based design achieves the highest best-case J. In terms of performance, the risk-based design is the least favorable. However, this design limits the exceedance of the threshold of the limit state function to some extent. If minimizing that exceedance is the primary goal, the risk-based design $\mathbf{X}^*_{c,\text{V}}$ should be selected. Otherwise, we recommend the reliability-based design $\mathbf{X}^*_{c,\text{II}}$, which offers a good balance between low failure probability and performance.

A critical evaluation of the results and the process reveals that the findings are constrained by the precision of the utilized Kriging models. Using a more generalized Gaussian process approach (Rasmussen and Williams, 2005) could yield better results, since the variability of the outputs could be captured. Nevertheless, this approach was chosen to balance accuracy and computational feasibility.

4. Summary and conclusions

This work addresses the NASA and DNV 2025 challenge on optimization under uncertainty. We utilize Bayesian optimization and sequential ABC to calibrate UMs for aleatory and epistemic parameters. Despite using only three datasets, our approaches for epistemic uncertainty yielded similar outcomes. Our study was limited to a time-independent representation, and we expect that incorporating time-dependent calibration could improve data-simulation alignment. Future analyses with more data could enhance understanding and model accuracy. Further, due to the nature of the challenge, results are restricted by limited knowledge of the measurements. Some understanding of the noise in the data or the physics behind data generation surely would give more insight in how to approach the model calibration.

For the design optimization, we present three unique design configurations and discuss their suitability for different scenarios. Overall, we find that a reliability-based design offers the best compromise between safety and performance in most cases.

Acknowledgement

LF acknowledges the funding of the DFG within the framework of the International Research Training Group 2657 under Grant Number 433082294. CD is grateful for the financial support of the German Research Foundation (DFG) (Grant number 530326817). JG acknowledges funding by dtec.bw - Digitalization and Technology Research Center of the Bundeswehr. dtec.bw is funded by the European Union - NextGenerationEU. PL and ZS acknowledge the financial support of the Alexander von Humboldt Foundation. TP acknowledges the Collaborative Research Centre 1463 (SFB 1463) “Integrated Design and Operation Methodology for Offshore Megastructures” (subproject C01, grant number 434502799). MF acknowledges the support of the Henriette Herz Scouting program.

References

Agrell, C., L. G. Crespo, V. Flovik, E. Vanem, and S. Kenny (2024, November). The NASA and DNV Challenge on Optimization Under Uncertainty.

Au, S.-K. and J. L. Beck (2001, October). Estimation of small failure probabilities in high dimensions by subset simulation. *Probabilistic Engineering Mechanics 16*(4), 263–277.

Behrensdorf, J., A. Gray, A. Perin, J. Grashorn, M. Luttmann, M. Broggi, G. Agarwal, L. Fritsch, F. Mett, and L. Knipper (2025, February). FriesischScott/UncertaintyQuantification.jl: V0.12.0. Zenodo.

Bi, S., M. Broggi, and M. Beer (2019, February). The role of the Bhattacharyya distance in stochastic model updating. *Mechanical Systems and Signal Processing 117*, 437–452.

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

Jones, D. R., M. Schonlau, and W. J. Welch (1998). Efficient global optimization of expensive black-box functions. *Journal of Global Optimization 13*, 455–492.

Rasmussen, C. E. and C. K. I. Williams (2005). *Gaussian Processes for Machine Learning*. The MIT Press.