Mechanical Systems and Signal Processing 164 (2022) 108266

Contents lists available at ScienceDirect

Mechanical Systems and Signal Processing

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

Risk-based functional black-box optimization Contribution to the NASA Langley UQ challenge on optimization under uncertainty

Christian Agrell [a,b,*,1], Simen Eldevik [a,1], Odin Gramstad [a,1], Andreas Hafver [a,1]

[a] DNV Group Research and Development, Norway
[b] Department of Mathematics, University of Oslo, Norway

ARTICLE INFO

Communicated by J.E. Mottershead

Keywords:
Model calibration
Functional data
Epistemic uncertainty
Robust optimization

ABSTRACT

This paper presents an approach to solve the 2019/2020 NASA Langley UQ challenge problem on optimization under uncertainty. We define an uncertainty model (UM) as a pair $\langle f_{a|e}, E\rangle$, where $f_{a|e}$ is a probability density over a for each $e \in E$, and proceed to infer $f_{a|e}$ in a Bayesian fashion. Special attention is given to dimensionality reduction of the functional (time-series) data, to obtain a finite dimensional representation suitable for robust Bayesian inversion. Reliability analysis is performed using $f_{a|e}$, whereas for design optimization we approximate $f_{a|e}$ using truncated Gaussians and a Gaussian copula. We apply an unscented transform (UT) in the standard normal space to estimate moments of the limit state, which is numerically very efficient. Design optimization is performed with this procedure to obtain negligible failure probability in g_1 and g_3 and acceptable failure probability and severity in g_2.

1. Introduction and notation

This paper[2] presents an approach to solve the NASA Langley UQ challenge problem on optimization under uncertainty [2]. We recall just the main problem setup and notation herein, and refer to [2] for a complete description.

In the first part of the problem, the goal is to establish a probability distribution for an aleatory random variable $a \in A = [0,2]^5$, that will depend on another variable $e \in E_0 = [0,2]^4$ with epistemic uncertainty. Inference is based on a set of time series, $D = \{y_i(t)\}_{i=1}^N$, that are the result of a functional mapping $Y : A \times E_0 \to y(t)$ where $y(t) : [0,5] \to \mathbb{R}$. The time series data D correspond to $y_i(t) = Y(a_i, e_{true})(t)$ for some fixed (but unknown) $e_{true} \in E_0$ and $N = 100$ i.i.d. samples a_i of $a \in A$.

We assume that the random variable a can be represented by a joint density f_a, that we wish to infer given the set of observations D. A numerical model of the data generating process, $\hat{Y}(a,e) \approx Y(a,e)$, is provided, and will be the basis for inference on the aleatory random variable a, as well as finding plausible values of the epistemic variable e. As we do not know the true value e_{true} used to generate the data, we will find a set $E \subseteq E_0$ of possible candidates for e, together with a family of probability distributions $\{f_{a|e}\}$, parametrized by $e \in E$ (see Tables 1–3).

* Corresponding author at: Department of Mathematics, University of Oslo, Norway.
E-mail address: chrisagr@math.uio.no (C. Agrell).

[1] All authors contributed equally.

[2] This is an updated and extended version of the paper [1] which was prepared for the 30th European Safety and Reliability (ESREL) 2020 Conference.

https://doi.org/10.1016/j.ymssp.2021.108266
Received 21 January 2021; Received in revised form 28 June 2021; Accepted 21 July 2021
Available online 31 July 2021

Table 1
Variables.

Variable	Domain	Description
a	$A = [0,2]^5$	Aleatory input variable
u	$\mathbb{R}^5$	Standard normal variable
e	$E_0 = [0,2]^4$	Epistemic input variable
θ	$\mathbb{R}^9$	Control variable

Table 2
Physical system.

$y(t) = Y(a,e)(t)$	Subsystem
$\hat{y}(t) = \hat{Y}(a,e)(t)$	Numerical model
$z(t) = (z_1, z_2)(t) = Z(a,e,\theta)(t)$	Integrated system
$\hat{z}(t) = \hat{Z}(a,e,\theta)(t)$	Numerical model
$g(a,e,\theta) = (g_1, g_2, g_3)(a,e,\theta)$	Limit-state
$w(a,e,\theta) = \max_{i=1,2,3} g_i(a,e,\theta)$	Limit-state

Table 3
Uncertainty model.

$\langle f_{a\|e}, E \rangle$	Uncertainty model
$f_{a\|e}$	Probability density of a given (fixed) e
$\hat{f}^{\alpha}_{a\|e}$	Parametric approximation of $f_{a\|e}$ with parameter α
E	Hyper-rectangular set
$\zeta(a,e)$	Parameter mapping
f_ζ	Probability density of ζ

The pair $\langle f_{a|e}, E \rangle$ is referred to as the uncertainty model (UM) of (a,e). Once $\langle f_{a|e}, E \rangle$ has been established, the remaining tasks are related to structural reliability analysis (SRA) of a larger physical system, where the mechanics given by $Y(a,e)$ is a subcomponent. This includes estimation of failure probabilities for a set of provided limit-state functions, sensitivity analysis and design optimization. We leave further details on the specific tasks to the relevant subsections below.

Our notation is aligned with the problem description, with some few additions. In particular, we may write $y(a,e,t)$ as $Y(a,e)(t)$ to clarify when we are working with the functional mapping Y vs a given function of time y. We also write $\hat{Y}$ and $\hat{y}$ to emphasize when Y and y are computed using the provided numerical model, and similarly for the integrated system $z(a,e,\theta,t)$. An overview of the notation used is given below, and the following sections, Section 2 to Section 7, correspond to respective subproblems in the UQ challenge. The final results are collected in Section 8, and we end with some concluding remarks in Section 9.

2. (A) model calibration & uncertainty quantification of a subsystem

Given a numerical model of the physical subsystem, $\hat{Y}(a,e) \approx Y(a,e)$, we seek to characterize the parameters (a,e) from a limited set of observations $\{y_i(t)\}$. An observation y_i corresponds to $Y(a_i, e_{true})$ where e_{true} is fixed (but unknown) and a_i are i.i.d. samples from some (unknown) distribution f_a. Our approach here is based on fitting a distribution to the observations, $y_1(t), \ldots, y_{100}(t)$, from which the conditional distribution $f_{a|e}$ can be determined.

2.1. Dimensionality reduction

As we wish to fit a density to the observations, we first need to compress the functional data to a finite-dimensional (preferably low-dimensional) representation. Various alternatives were considered, from the naive approach of evaluating $y(t)$ at a finite set of times $\{t_1, \ldots, t_N\}$, to more sophisticated function approximation techniques. The Karhunen–Loéve transform as used in functional principal component analysis [3] was first considered. This approach is based on finding an orthonormal eigenbasis of L^2 (all square integrable functions) from the estimated covariance function corresponding to the observations. By projecting each observation $y_i(t)$ onto the subspace spanned by the first 10 eigenfunctions, we observed that the residual (the projection onto the complement space), had negligible L^2 norm.

As both observations $y_i(t)$ and samples from $\hat{Y}(a,e)(t)$ consistently showed two or three distinct frequency components when performing a Fast Fourier Transform (FFT), it was also deemed appropriate to consider damped complex exponentials as the function basis. With this approach we write

$$y_{exp}(t) = \sum_{i=1}^{k} B_i \cos(\omega_i t + \phi_i) e^{d_i t}, \tag{1}$$

where the coefficients $(B_i, \omega_i, \phi_i, d_i)$ can be estimated efficiently using Prony's method (see for instance [4]). Unlike the eigenfunctions used in the Karhunen–Loéve transform, the damped complex exponentials do not form an orthonormal basis. But we noticed

that any of the observations $y_i(t)$, as well as any function $\hat{y}(t)$ computed using the numerical model $\hat{Y}(a,e)$ for some $(a,e) \in A \times E_0$, could be represented as the sum of $k = 3$ damped complex exponentials, up to machine precision.

Given some parametric function approximation, we define the parameter mapping

$$\zeta(a,e) : A \times E_0 \to \mathbb{R}^{n_\zeta}. \tag{2}$$

We will write $\zeta(a,e)$ or $\zeta(y)$ interchangeably, depending on whether we consider a specific function $y(t)$ or variables (a,e), for which $\zeta(a,e) = \zeta(Y(a,e))$. If $\zeta(a,e)$ is the parameter vector corresponding to the damped complex exponentials, then $\zeta(a,e)$ is given by (some of) the $4k$ parameters needed in Eq. (1). In the case where the Karhunen–Loéve transform is used to represent $Y(a,e)(t)$, we could let $\zeta(a,e)$ correspond to the projection coefficients (FPCA scores) of the first n_ζ eigenfunctions with the largest eigenvalues.

The reason for introducing $\zeta(a,e)$ is that, if we can fit a probability density $f_\zeta(\zeta)$ to ζ, then we can establish a distribution $f_{a|e}(a)$ as

$$f_{a|e}(a) = \frac{1}{C} f_\zeta(\zeta(a,e)), \tag{3}$$

where C is a normalizing constant that assures that $f_{a|e}$ integrates to 1. Eq. (3) comes from a Bayesian formulation, assuming that $e = e_{true}$ and a uniform prior on a (with constant density that goes into C), and where for a fixed e, the density f_ζ defines a (improper) likelihood for any $\zeta(a,e)$ which we assign to a.

The reason why we define $f_{a|e}(a)$ as in (3), by first fitting a probability density $f_\zeta(\zeta)$ to the (transformed) observations, is because we find it easier to determine the appropriate amount of regularization needed to avoid overfitting. Note that this approach to inversion is not the same as computing the distribution corresponding to a transformation of f_ζ through ζ^{-1} (if ζ was bijective), which would include the Jacobian of ζ in the right hand side of (3).

Because of this, together with the fact that the number of observations is limited, we need some means of ensuring that the distribution $f_{a|e}(a)$ is conservative, in the sense that it does not assign negligible probability to values of (a,e) that could be plausible. For this, some qualitative judgment is usually needed. With our approach, we address this by fitting a high-entropy distribution to f_ζ.

We observed that both the eigenfunction approximation and the damped complex exponentials could provide a reasonable dimensionality reduction, from the relevant space of functions $y(t)$ to a set of 10-12 parameters. We chose to go with the damped complex exponentials, as it turned out to provide lossless compression of the functional data. We also found that it was more straightforward to fit a density f_ζ to $\zeta(a,e)$ using this approach.

2.2. Model discrepancy

The problem of assessing model discrepancy without controlled experiments, i.e. when only observed output of the physical system is available, is generally ill-posed. This is due to problems with identifiability, and additional assumptions on the accuracy of the numerical model, as well as other sources of uncertainty in the true physical process, are generally needed. See for instance [5] and the discussions therein.

We will assume that there are no other sources of uncertainty in the physical subsystem besides input uncertainty. That is, variability in the observed $y_i's$ is due to variability in aleatory input a alone. Similarly, we would like to assume zero model discrepancy as well.[3] i.e. $\hat{Y}(a,e) = Y(a,e)$ for all $(a,e) \in A \times E_0$. However, we found that we needed to assume some model discrepancy in order for inference on (a,e) to be possible.

If we want to fit a probability density to ζ, and use this to create a distribution on the input a assuming zero error in $\hat{Y}(a,e)$, then we must first verify that the parameter vector of the observations, $\zeta(y_i)$, are within the range of $\zeta(a,e)$ (i.e. each y_i can be reproduced from $\hat{Y}$). Otherwise, the resulting distribution $f_{a|e}$ could assign zero probability to almost all $a \in A$.

In practice, we will not make use of all components of ζ when fitting a distribution, in order to impose some regularization. But before we do this, it is useful to use the complete ζ to investigate if there is any model discrepancy. As noted in Section 2.1, if we let the parameter mapping be defined as $\zeta(a,e) = [\zeta_1, \zeta_2, \zeta_3]$, where $\zeta_i = [B_i, \omega_i, \phi_i, d_i]$ are the parameters of the ith wave component in Eq. (1), then ζ provides a bijection between the range of the numerical model, $\hat{Y}(A, E_0)$, and a subset of $\mathbb{R}^{12}$. We can therefore investigate whether the observations y_i are within the range of $\hat{Y}$ through ζ. We found that this was not the case, as it turns out that $\zeta(y_i)$ falls outside $\zeta(A, E_0)$ in the subspace spanned by (d_1, ω_1), see Fig. 1. If we exclude the parameter d_1 from ζ, the set $\zeta(A, E_0)$ will include all transformed observations $\zeta(y_i)$. In practice, the exclusion of d_1 corresponds to the assumption that the response from the true physical subsystem, $y = Y(a,e)$, is a bit less "damped", compared to what would be expected were the response to agree with the numerical model $\hat{Y}(a,e)$. Of course, this may be an assumption that is not appropriate, and generally one would assess such an assumption based on knowledge related to the physical phenomenon and information regarding what kind of model discrepancy (or observational noise/error) to expect.

Fig. 2 shows effectively what the assumed model discrepancy looks like under this assumption. Here, to give one example, one of the observations (# 71) was fitted using (1), and we vary d_1 within the relevant range from Fig. 1. Any amount of damping that keeps the time series within the blue shaded area is considered negligible in terms of model discrepancy. In Fig. 3 we see some examples of functions $\hat{y}(t)$ that are *not* equivalent with this observation, which illustrates that the assumed model discrepancy is rather small. We note that the assumed model discrepancy is probably negligible for all practical purposes, but it is necessary for inference on the model input (a,e) to be possible.

[3] In reality it would be natural to make use of some model uncertainty, either estimated or assumed, when the UM we aim to establish will be used for SRA of a safety-critical system.

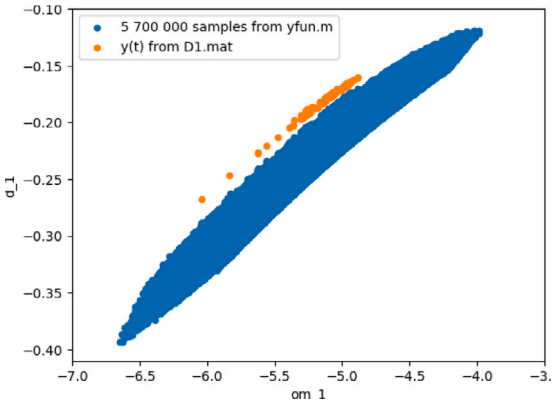


Fig. 1. Frequency (ω_1) vs damping (d_1) coefficients of the first wave component in Eq. (1), corresponding to observations (y_i) and output from the numerical model ($\hat{y}$).

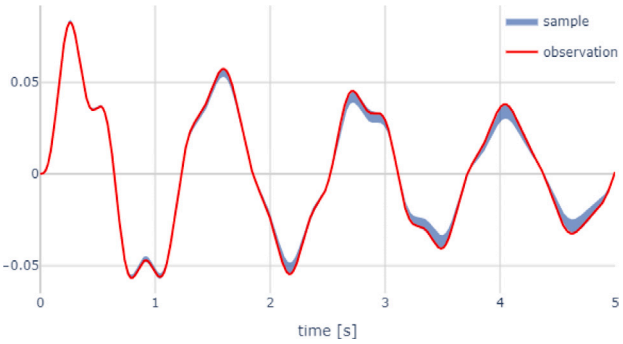


Fig. 2. Example of assumed model discrepancy. Any function $\hat{y}(t)$ is interpreted as equivalent to the observation $y_{i=71}$ (red) if it is more damped (compressed in the y-direction) than $y_{i=71}$, as long as $\hat{y}(t)$ falls within the blue shaded area.

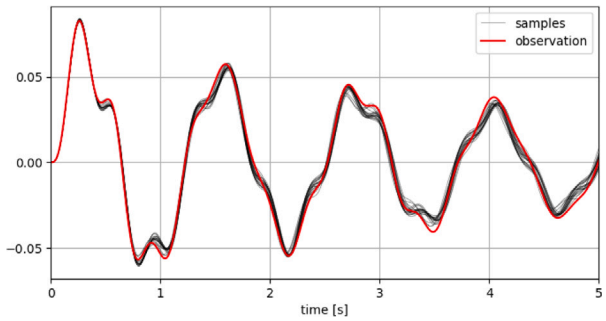


Fig. 3. Examples of functions $\hat{y}(t)$ that do *not* agree with the observation $y_{i=71}$ (red), with respect to the assumed model discrepancy. There are relevant features of the time series that make the samples and the observation significantly different.

Remark 2.1. The results presented in this paper may be sensitive to the assumed model discrepancy. If the criterion illustrated in Figs. 2 and 3 is too strict, i.e. larger deviations between the data generating process $Y(a,e)$ and the computer model $\hat{Y}(a,e)$ is expected, the resulting UM may be overly optimistic. That is, as inputs (a,e) are deemed less plausible, the volume of E and the entropy of $f_{a|e}$ are reduced.

This assumed model discrepancy is based on: (1) no model error has been specified for $\hat{Y}(a,e)$, and (2) with the relaxed assumption on damping of the time series data, it is for each observation y_i possible to find inputs (a,e) such that $\hat{Y}(a,e)$ agrees with y_i.

2.3. The uncertainty model $\langle f_{a|e}, E \rangle$

We will establish an uncertainty model $\langle f_{a|e}, E \rangle$ based on the parameter mapping (ζ) given by the damped complex exponentials with $k = 3$. As discussed in Section 2.2, the component d_1 is excluded to account for some model discrepancy, and 3 other parameters

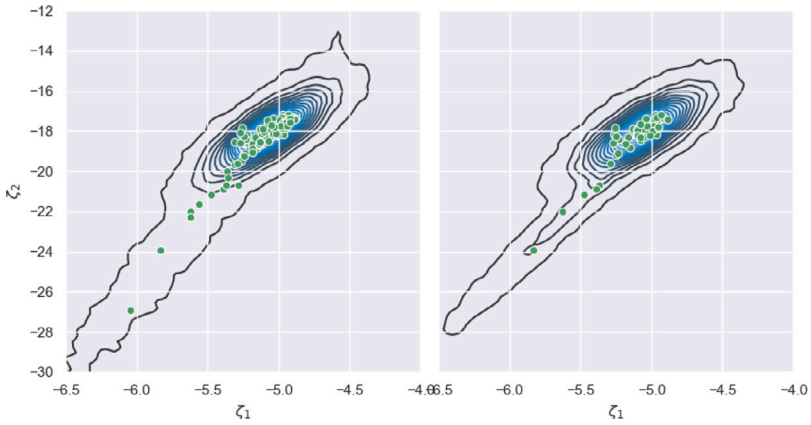


Fig. 4. Marginal of two components of ζ, where f_ζ is obtained from the 100 observations y_i (left), and a subset of 50 observations (right).

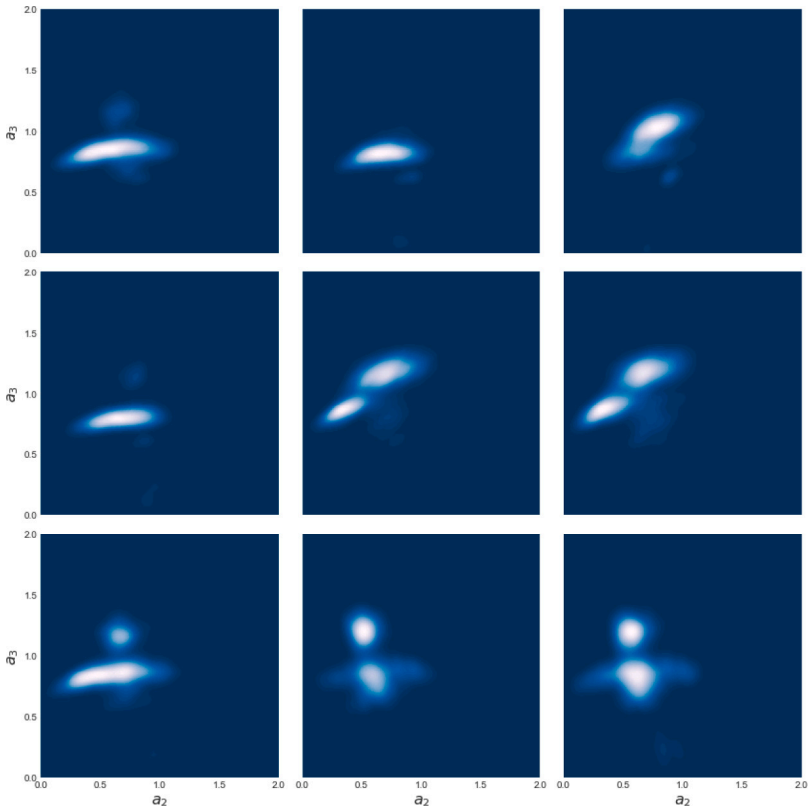


Fig. 5. Marginal of $f_{a|e}$ in the (a_2, a_3)-subspace for 9 different values of e.

are excluded after principal component analysis (PCA) where we found that 99% of the variance could be described by the final 8 parameters. From the final parameter vector $\zeta : A \times E_0 \to \mathbb{R}^8$, we fit a mixture of two Gaussians to the 100 observations y_i under the transformation $\zeta(y_i)$.

Fig. 4 shows the marginal distribution corresponding to two of the components of ζ. Some conservatism was included by increasing the variance of the fitted distribution, in order to account for the limited number of observations. Starting from a maximum likelihood fit, the variance was increased by a constant to produce a more conservative (higher entropy) distribution. Hence, the corresponding distribution $f_{a|e}$ will likely underfit the data, which is intentional. The method used to fit a distribution to ζ is fairly robust to the number of observations. By sampling subsets of size 50 out of the total 100 observations, we find that the resulting distributions fitted to $\{\zeta(y_i)\}$ are fairly consistent as illustrated in Fig. 4.

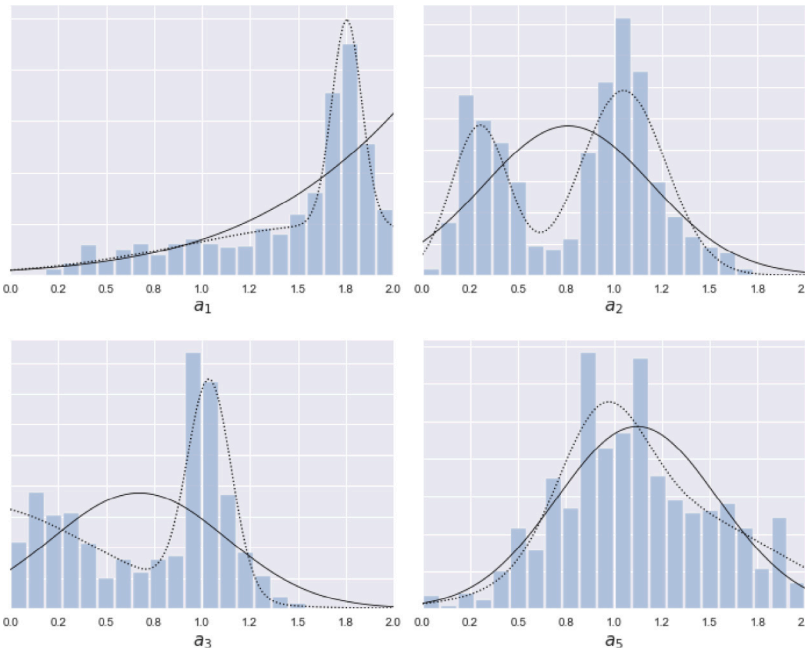


Fig. 6. Parametric distributions fitted to marginals of $f_{a|e}$. The dotted curve shows a mixture of two truncated normal distributions versus the one single truncated normal.

Using the distribution f_ζ, we can assess the likelihood $f_\zeta(\zeta(a,e))$ for any $(a,e) \in A \times E_0$. For a point $e \in E_0$ to be plausible, it should be possible to find some $a \in A$ such that $f_\zeta(\zeta(a,e))$ is large. That is, there must exist some $a \in A$ such that $\zeta(a,e)$ lies within the main bulk of the distribution shown in Fig. 4. In fact, there must exist some $a_i \in A$ such that $\zeta(a_i,e)$ corresponds with $\zeta(y_i)$, for all observations y_i, in order for e to be plausible. However, such a strategy for finding the set E would require computing $\zeta(a,e)$ for all $a \in A$. As a practical alternative, we generate a large set of samples $(a,e)_i \in A \times E_0$, and filter out all samples with likelihood ≈ 0. We then estimate the set $\hat{E}$ of plausible e-values from the remaining samples, and determine E as the smallest hyper-rectangular set containing $\hat{E}$. This is the initial strategy we use when E is determined for the first time. For further uncertainty reduction we will use a refined strategy discussed in Section 3.1.

From the fitted distribution f_ζ, we obtain the non-parametric distribution $f_{a|e}$ as in Eq. (3). An illustration is given in Fig. 5 where some 2d marginals of $f_{a|e}$ are plotted for a few different values of e. It will also be useful to establish a parametric approximation to $f_{a|e}$, and for this we use a multivariate truncated Gaussian over (a_1, a_2, a_3, a_5), with a_4 uniform on the interval $[0,2]$. We write the parametric approximation as

$$\hat{f}^{\alpha}_{a|e} \approx f_{a|e}, \tag{4}$$

with distribution parameter α. We define the parametric approximation in terms of the marginals

$$\begin{aligned} &a_i \sim TN(\mu_i, \sigma_i, 0, 2) \text{ for } i = 1,2,3,5, \\ &a_4 \sim U([0,2]), \end{aligned} \tag{5}$$

and a Gaussian copula specified by a 5×5 correlation matrix $R = [\rho_{i,j}]$, where $\rho_{i,j}$ is the Spearman rank correlation coefficient between a_i and a_j. Here $a_i \sim TN(\cdot)$ denotes that a_i has a univariate normal distribution, conditioned on the event $a_i \in [0,2]$, and α is the vector of all parameters, μ_i, σ_i and $\rho_{i,j}$. Note that all of these parameters depend on e, i.e. $\alpha = \alpha(e)$, and we will estimate α based on samples (MCMC) from $f_{a|e}$.

Fig. 6 shows an example of the parametric distribution for a given value of e, using a maximum likelihood estimate of α. It turns out that for many values of $e \in E$, a mixture of two truncated Gaussians would provide a better fit. Other alternatives for the marginals could also be considered. However, we only intend to use the parametric distribution for initial estimates of failure probabilities, to help with importance sampling, and to approximate moments of $g(a,e,\theta)$ (in particular, as a crude approximation of how the variability of g changes with e and θ). And for this purpose, this simple model seems sufficient.

From the correlation matrix R we observe that there is usually some correlation between a_1, a_2 and a_3. Fig. 7 shows 2D marginals of some of these pairs (a_i, a_j). Note that the value of e used to generate the plots in Figs. 6 and 7 may be far from the correct one, but cannot be ruled out by the observations $\{y_i\}$ alone.

In reliability analysis, it is often useful to work with the random variables in the standard normal space. From the selected parametric distribution, we can perform the Nataf-type of transformation

$$a \xrightarrow{T_e} u \sim N(0, I),$$

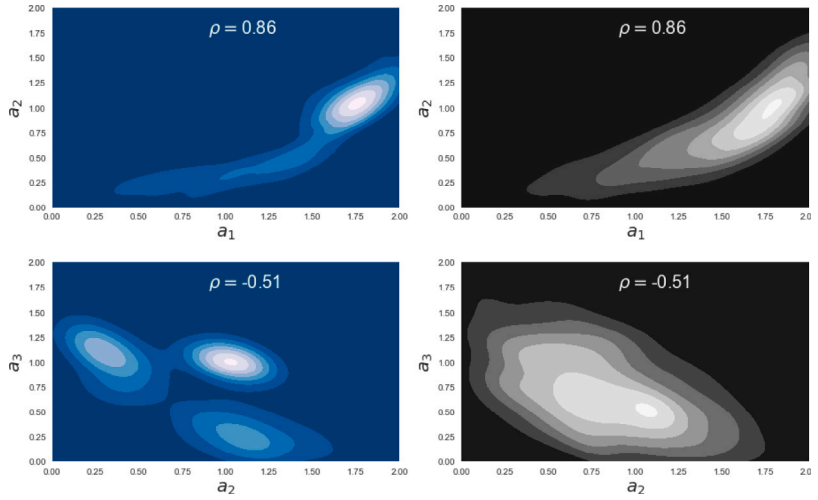


Fig. 7. 2D marginal distributions of (a_1, a_2) and (a_1, a_3), from the non-parametric distribution $f_{a|e}$ (left) and parametric approximation (right), for a fixed value of e.

which has the property that, for any $e \in E_0$ and $a \sim \hat{f}^{\alpha}_{a|e}$, $u = T_e(a)$ is n^a-dimensional standard normal. The transformation is obtained by first letting $z_i = \Phi^{-1}(F_i(a_i))$ where Φ is the standard normal CDF and F_i is the CDF of a_i. Then $z = (z_1, \dots, z_{n_a}) \sim N(0, R_0)$ with covariance matrix

$$[R_0]_{i,j} = 2 \sin\left(\frac{\pi \rho_{i,j}}{6}\right).$$

See for instance [6] for details. The standard normal variable is then obtained by $u = L_0^{-1} z$ where $L_0 L_0^T = R_0$.

Remark 2.2. Different values of e have been used to generate the plots in Fig. 5, Fig. 6 and Fig. 7, in order to illustrate some different versions of $f_{a|e}$ and $\hat{f}^{\alpha}_{a|e}$. These are not necessarily plausible values of e.

3. (B) uncertainty reduction

3.1. Ranking of epistemic parameters

To rank the epistemic parameters according to their ability to improve the predictive ability of $\hat{Y}$, we study how $f_{a|e}$ changes with respect to $e \in E$. To measure the effect of one component of $e = (e_1, e_2, e_3, e_4)$, say e_1, we estimate the expected Kullback–Leibler divergence

$$\mathbb{E}[\mathcal{KL}(f_{a|e} \parallel f_{a|e'})] = \mathbb{E}\left[\int_A f_{a|e}(a) \log \frac{f_{a|e}(a)}{f_{a|e'}(a)}\, da\right].$$

Here e' represents a small perturbation of e_1, $e_1' = e_1 + 0.1$, and the expectation is taken over the other parameters (e_2, e_3, e_4) that are assumed uniform within the bounds set by E.

From Fig. 8 we conclude that e_3 is more influential than e_2 and e_4, in the sense that small perturbations of e_3 has a larger effect on the related distribution $f_{a|e}$. The same is true for e_1, if e_1 is large in the first place. When deciding on the uncertainty reductions to make, we must also take into account that we may only request to increase or decrease the bounds of the initial set $E_0 = [0, 2]^4$. We observed that $e \in E \Rightarrow e_1 \ll 2$, and so decreasing the upper bound on e_1 might not provide any new information. Hence, we combine the information from Fig. 8 with the initial uncertainty reduction described in Section 2.3, where we assess the plausability of e. When also considering which e_i is close to the border of E_0, we decided to request uncertainty reduction on the lower bound on e_3 and e_4 ($\{e_3^-, e_4^-\}$).

3.2. First UM update

Our initial method for defining a set E of plausible e-values was based on the simple procedure described in Section 2.3, where we generate samples in $A \times E_0$ and filter out those with negligible likelihood. Now, when the bounding intervals have been reduced by this method, together with additional refinement provided by the request $\{e_3^-, e_4^-\}$, we switch to a more detailed method for further refinement of $E \to E_1$. The first updated UM is then $\langle f_{a|e}, E_1 \rangle$.

We may assess whether any $e \in E$ is plausible by generating samples $\{a^{(j)}\}$ from $f_{a|e}$, computing $\hat{y}_j = \hat{Y}(a^{(j)}, e)$, and comparing the set $\{\hat{y}_j\}$ with the provided observations $\{y_i\}$. In practice we will work in the reduced space, comparing $\zeta(a^{(j)}, e)$ against $\{\zeta_i\} = \{\zeta(y_i)\}$.

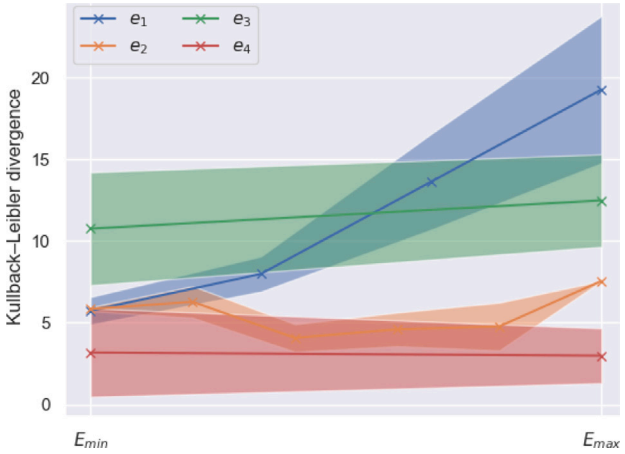


Fig. 8. Kullback–Leibler divergence between respective distributions $f_{a|e}$ for $e \in E$, when component e_i is perturbed by $e_i \to e_i + 0.1$. The figure shows mean and mean ± 1 standard deviation, when the remaining components, e_j for $j \neq i$, are uniform within E.

There are different ways of defining *plausability* of e, and we decided to go with a simple approach, where we compare the median of the probability density of 10^3 samples of $\zeta(a^{(j)}, e)$, with the density of the observations $\{\zeta_i\}$ given by (3). We use this approach, because of its simplicity, and because it was easy to find a suitable threshold that could be used to determine when a value of e was implausible. This procedure is more refined than the one presented in Section 2.3, where we filtered out the values of e with likelihood ≈ 0. But we note that the initial filtering was necessary, as there would be numerical issues with sampling from $f_{a|e}$ for the extremely unlikely values of e.

Our strategy for reducing the size of E is based on this procedure, where we check whether $e^{(k)}$ is plausible for a representative set $\{e^{(k)}\} \subset E$, and let E_1 be the smallest hyper-rectangular set containing all plausible $e^{(k)}$. The representative set $\{e^{(k)}\}$ is given as the union of the following two sets:

- a Latin Hypercube (LHS) sample within E,
- samples gathered form a Bayesian Optimization (BO) targeted at finding the most plausible $e \in E$.

Bayesian optimization (see for instance [7]) is a method for finding the maximum of a function using a small number of function evaluations, and we use it here as a technique to locate the most plausible candidates in E. We use this procedure as evaluating the plausability of a single value of e is based on sampling from $f_{a|e}$ and many computations of $\zeta(\cdot)$, which takes a bit of time.

Remark 3.1. The updated set E is determined from a threshold on the likelihood of the epistemic variable e. Initially we consider a threshold that is not overly conservative when we perform the model calibration. This makes it easier to find a design θ which is both acceptable with respect to improbable values of e, and closer to optimal with respect to the more realistic values of e. In Section 7 we discuss this in more detail, and we will also verify that the final design is acceptable even for a more conservative (larger) set E.

4. (C) reliability analysis of baseline design

The reliability analysis is performed for multiple designs, θ_{baseline}, θ_{new}, θ_{final} and $\theta_{\hat{r}\%\text{risk}}$, and the final results are collected in Section 8.

4.1. Numerical procedure for estimation of failure probabilities

To estimate failure probabilities, we will work with the density $f_{a|e}$ given in Eq. (3). For some of the subproblems we address, it is not necessary to compute the normalizing constant C, and we define the proportional density

$$\dot{f}_{a|e}(a) = f_{\zeta}(\zeta(a, e)), \tag{6}$$

such that $f_{a|e}(a) = \dot{f}_{a|e}(a)/C$. In particular, from $\dot{f}_{a|e}(a)$ we may estimate moments of $f_{a|e}(a)$ using importance sampling, or generate samples using Markov chain Monte Carlo (MCMC).

Given a limit state function $g_i(a, e, \theta)$, the simplest way to estimate the failure probability p_f is by crude Monte Carlo (MC) sampling,

$$\hat{p}_{f,MC} = \frac{1}{n} \sum_{j=1}^{n} \mathbb{1}_{\{g_i(a_j, e, \theta) \geq 0\}}, \tag{7}$$

where $a_1, \dots, a_n$ are i.i.d. samples from $f_{a|e}$. The unbiased sample variance can be estimated by

$$\widehat{var}(\hat{p}_{f,MC}) = \frac{\hat{p}_{f,MC}(1-\hat{p}_{f,MC})}{n}. \tag{8}$$

A common technique to reduce the sample variance is by importance sampling, where we make use of some importance distribution $q(a)$, with the property that $q(a) \neq 0$ whenever $f_{a|e}(a) \neq 0$. The importance sampling estimate and the estimated sample variance are given by

$$\begin{aligned} \hat{p}_{f,IS} &= \frac{1}{Cn}\sum_{j=1}^{n} \mathbb{1}_{\{g_i(a_j,e,\theta)\geq 0\}} \frac{\dot{f}_{a|e}(a_j)}{q(a_j)}, \\ \widehat{var}(\hat{p}_{f,IS}) &= \frac{1}{C^2 n}\cdot \sum_{j=1}^{n}\left(\mathbb{1}_{\{g_i(a_j,e,\theta)\geq 0\}} \frac{\dot{f}_{a|e}(a_j)}{q(a_j)} - C\hat{p}_{f,IS}\right)^2, \end{aligned} \tag{9}$$

where a_j are sampled from $q(a)$.

We will use two different strategies for failure probability estimation, depending on whether the failure probability is large (> 0.01) or small (< 0.01). For large failure probabilities, we will rely on the MC estimate Eq. (7), where n is chosen such that the relative error is acceptable. To generate the samples a_j we make use of the Affine Invariant Markov chain Monte Carlo (MCMC) Ensemble sampler provided by [8].

For smaller failure probabilities, sampling directly from $f_{a|e}$ will be inefficient. Here we use the importance sampling estimate Eq. (9), where we select $q(a)$ from the design point of a FORM analysis. This is achieved by first running a FORM analysis using the parametric approximation $\hat{f}^{\alpha}_{a|e}$. FORM is based on obtaining a linear approximation to the limit-state at a point a^* called the *design point*. The design point is the point on the limit state, $g_i = 0$, with largest probability density. See for instance [9] for further details. We define $q(a)$ as a modified version of $\hat{f}^{\alpha}_{a|e}$, where the distribution is shifted such that $\mathbb{E}[q] = a^*$ (it may also be useful to increase the variance slightly). The distribution $q(a)$ obtained in this way should then be able to produce samples a_j where both $g_i(a_j, e, \theta) \geq 0$ and $f_{a|e}(a_j)$ is large, which is needed to reduce the variance of the estimated failure probability. When the failure probability for g_1, g_2 and g_3 have been estimated, the samples used for each individual limit state can also be used for w in order to reduce the total number of function evaluations of $g(\cdot)$.

To estimate the range of the failure probability for e in some set E, we compute the failure probability for a finite subset $\{e^{(k)}\} \subset E$ as described in Section 3.2. In order to capture the relevant ranges of failure probabilities, we found that around 100 e-values, $|\{e^{(k)}\}| \approx 100$, seemed sufficient.

4.2. Ranking of epistemic uncertainties

To rank the epistemic uncertainties according to the contraction of the failure probability $p_{f,w}(e,\theta) = \mathbb{P}(w(a,e,\theta) \geq 0)$, we estimate the change in the minimum and maximum of $\{p_{f,w}(e,\theta)\}_{e\in E}$ when we replace E with a reduced set E'. Again we make use of the finite subset $\{e^{(k)}\} \subset E$ for which $p_{f,w}(e,\theta)$ is already available. The results are given in Section 8, where we let E' be the set given by increasing the lower bounds and decreasing the upper bounds of E with 25%.

5. (D) reliability-based design

5.1. Optimality criterion

As we do not have any information on the criticality of each failure mode, i.e. whether some limit state g_i should be seen as more important than the others, we will seek a design θ where $g_i(a,e,\theta)$ is as "small as possible" for all $i = 1,2,3$. Our optimality criterion will be based on the characteristic values

$$c_i(e,\theta) = \mathbb{E}\left[g_i(a,e,\theta)\right] + 2\ \mathrm{Std}(g_i(a,e,\theta)), \tag{10}$$

where the expectation and standard deviation (Std) are taken with respect to $a \sim f_{a|e}$. From the characteristic values c_i, we define the following loss function:

$$L(e,\theta) = \sum_{i=1}^{3} \exp[\gamma_i c_i(e,\theta)]. \tag{11}$$

Here $\gamma_i > 0$ are constants needed to bring each c_i to the same scale, as the output of each g_i are of different orders of magnitude. In our implementation we have used $\gamma_1 = 10$, $\gamma_2 = 200$ and $\gamma_3 = 1$.

Given some set E of plausible e-values, we define the reliability-optimal design as a solution θ^* of the optimization problem

$$\theta^* \in \arg\min_{\theta}\{\max_{e\in E} L(e,\theta)\}. \tag{12}$$

We will make use of an approximation to Eq. (11) described below, and the final results are collected in Section 8.

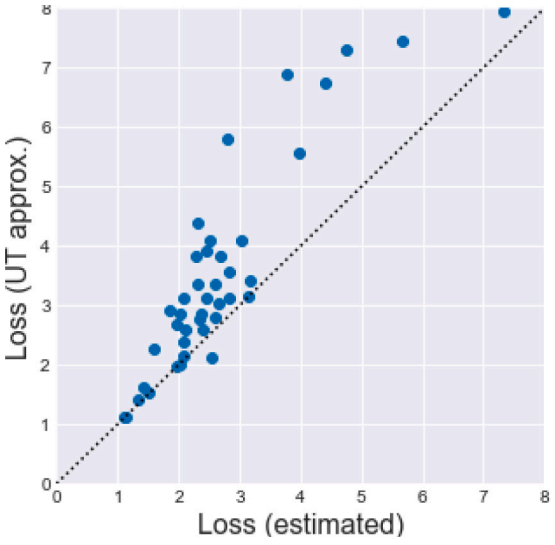


Fig. 9. UT approximation of the loss function Eq. (11) vs estimated (MCMC) loss.

5.2. Numerical UT approximation

For numerical efficiency, we introduce an approximation of the loss function Eq. (11) based on the Unscented Transform (UT). UT is an efficient method for estimating the mean and covariance of a random variable after nonlinear transformation. In short, given some random variable x we define a set of *weighted sigma-points* ($\{w_i, x_i\}$), such that if ($\{w_i, x_i\}$) were considered as a discrete probability distribution, then its mean and covariance would coincide with x. For any nonlinear transformation $y = h(x)$, if x is discrete we may compute the mean and covariance of y exactly. The UT approximation is the result of such computation, when we approximate x with ($\{w_i, x_i\}$). For details see e.g. [10].

To select the set of sigma-points in UT, we make use of the method developed by [11], which produces a set of $2n_a + 1$ sigma-points and weights in $\mathbb{R}^{n_a}$. These points are generated under the assumption that x follows a n_a-dimensional standard normal distribution. The sigma-points corresponding to the parametric distribution $\hat{f}^{\alpha}_{a|e}$ can then be obtained by the transformation discussed in Section 2.3. From here, the characteristic values c_i in Eq. (10) can then be estimated by just 11 evaluations of the limit state.

Fig. 9 shows the UT approximation of the loss $L(e, \theta)$ for some different values of θ and e. With the goal of minimizing the loss estimated using samples from $f_{a|e}$, the UT approximation seems like a viable proxy, and the optimization problem in Eq. (12) can then be solved by standard tools for numerical optimization.

6. (E) model update and design tuning

Here we rely on the same procedure for dimensionality reduction as discussed in Section 2.1, in order to refine the UM based on samples from the integrated system $\{z^{(i)}\}$, together with the numerical model $\hat{Z}(a, e, \theta)$. Hence, we obtain an updated density ζ fitted to two sets of parameter vectors, corresponding to the complex exponential representation of the two datasets $\{y^{(i)}\}$ and $\{z^{(i)}\}$. We found it difficult to assess potential model discrepancy in $\hat{Z}(a, e, \theta)$. We suspect that there is more discrepancy in $\hat{Z}(a, e, \theta)$ than in $\hat{Y}(a, e)$, but chose to update the UM under the assumption that $\hat{Z}(a, e, \theta)$ still provides an accurate model of the data generating process. However, if this is not the case then our second refined UM may be too optimistic (see Remark 2.1). Based on the same type of assessment as in Section 3.1, we decided to again request the uncertainty reduction $\{e_3^-, e_4^-\}$.

We refer to the second updated UM as $\langle f_{a|e}, E_2 \rangle$, where $E_2 \subset E_1$ and $f_{a|e}$ is given by Eq. (3) where f_ζ is the updated based on $\{z^{(i)}\}$.

7. (F) risk-based design

As discussed in Remark 3.1, the set E_2 used to find θ_{final} is based on a threshold on the likelihood of the epistemic variable e which is not overly conservative. Now, we will introduce a set $E_{0\%\text{risk}}$ where $e \notin E_{0\%\text{risk}}$ is assumed impossible. We say that $e \notin E_{0\%\text{risk}}$ if there is at least one observation y_i which cannot be explained by $\hat{Y}(a, e)$, for any $a \in A$. (But note that the comments in Remark 2.1 still apply).

When we define smaller sets $E_{r\%\text{risk}} \subset E_{0\%\text{risk}}$, corresponding to neglecting a portion of "r% risk", we let $E_{r\%\text{risk}}$ be the set containing the $(1 - r)$% values of e with largest likelihood. In practice we use a finite sets of points $\{e_i\}$ to represent these sets, and so the shape of $E_{r\%\text{risk}}$ is not necessarily rectangular. But we can still associate to each set $E_{r\%\text{risk}}$ the smallest hyper-rectangular set containing it, as illustrated in Fig. 10. The set E_2 defined in Section 6, which is used in the optimization of θ_{final}, is the set

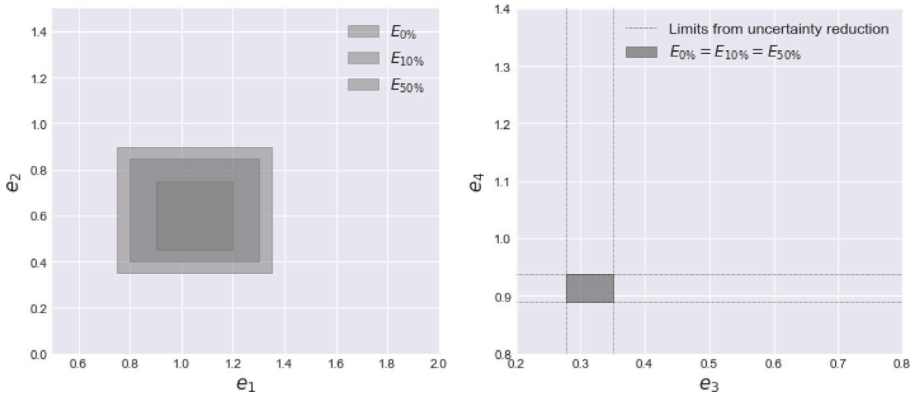


Fig. 10. Illustration of the risk-based sets $E_{r\%\text{risk}}$. The set E_2 used in the optimization of θ_{final} is $E_2 = E_{50\%\text{risk}}$. The projection onto e_3–e_4 is the same for all three sets.

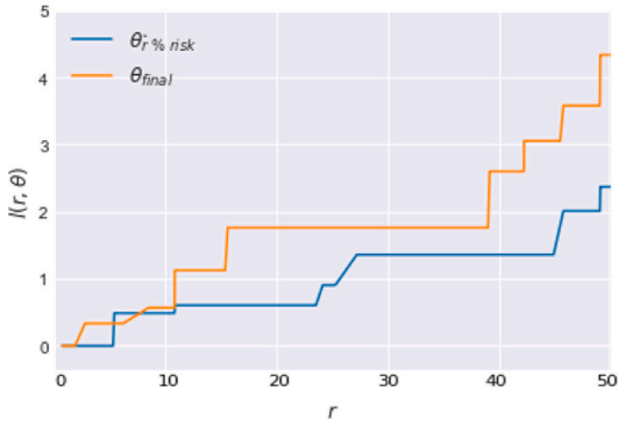


Fig. 11. The gain (13) for θ_{final} and $\theta_{\hat{r}\%\text{risk}}$.

corresponding to $r = 50\%$ risk. Now we will also find an optimal design $\theta_{\hat{r}\%\text{risk}}$ for $\hat{r} = 5\%$, corresponding to $E_{5\%\text{risk}}$. Both θ_{final} and $\theta_{\hat{r}\%\text{risk}}$ will then be evaluated with respect to the conservative scenario $E_{0\%\text{risk}}$ and the more optimistic scenario $E_{50\%\text{risk}}$ in Section 8.

To quantify the gain $I(r,\theta)$ resulting from taking the risk r, we let

$$I(r,\theta) = \max_{e\in E_{0\%}} L(e,\theta) - \max_{e\in E_{r\%}} L(e,\theta), \tag{13}$$

where $L(e,\theta)$ is the loss function defined in (11). The evaluation of the design $\theta_{\hat{r}\%\text{risk}}$ that maximizes $I(\hat{r},\theta)$ for $\hat{r} = 5\%$ is presented in Section 8, Table 6, and the gain $I(r,\theta)$ for both θ_{final} and $\theta_{\hat{r}\%\text{risk}}$ is shown for a range of values r in Fig. 11.

Here we note that a 5% reduction of the epistemic space has very little effect, and optimization over $E_{5\%\text{risk}}$ is practically the same as optimization over $E_{0\%\text{risk}}$. We may therefore consider $\theta_{\hat{r}\%\text{risk}}$ as the design optimized for the worst-case scenario, which we can compare against the more optimistic θ_{final} corresponding to $r = 50\%$. In Fig. 11 we would expect that the curve for θ_{final} had the steepest slope, i.e. that there is more to be gained by reducing the epistemic set E for θ_{final} than $\theta_{\hat{r}\%\text{risk}}$, but in practice it is difficult to determine if there is any significant difference in the two designs at all. This is also reflected in the evaluation of failure probabilities in Section 8.

8. Final results

8.1. Failure probability and severity

We planned to make use of crude MC for initial computation of failure probabilities that are not too small, and switch to importance sampling after UM refinement and design optimization where more accurate estimation is needed. However, after the optimization and UM refinement, the failure probabilities are no longer computable. This happens when it is not possible to find any $(a,e) \in A \times E$ where $g_i \geq 0$. And this seems to be the case for the second refined UM and θ_{final}, as we were not able to find any $(a,e) \in A \times E_2$ where $g_i(a,e) \geq 0$ for $i = 1$ and $i = 3$ through numerical global optimization (maximization). We could of course tune

Table 4
Failure probability — First refined UM.

θ	Limit-state	p_f min	p_f max	Severity
θ_{baseline}	g_1	–	0.145	0.030
	g_2	0.010	0.715	0.003
	g_3	–	0.507	0.246
	w	0.048	0.727	NA
θ_{new}	g_1	–	0.045	$1.8 \cdot 10^{-4}$
	g_2	–	0.522	$1.1 \cdot 10^{-3}$
	g_3	–	–	–
	w	–	0.522	NA

*MC estimates. Missing p_f values (–) can be assumed $< 10^{-3}$.

Table 5
Failure probability — Second refined UM.

θ	Limit-state	p_f min	p_f max	Severity
θ_{baseline}	g_1	–	0.224	$2.1 \cdot 10^{-2}$
	g_2	0.019	0.162	$4.5 \cdot 10^{-4}$
	g_3	–	0.021	$1.1 \cdot 10^{-2}$
	w	0.026	0.341	NA
θ_{new}	g_1	–	0.001	$3.9 \cdot 10^{-6}$
	g_2	0.006	0.030	$2.9 \cdot 10^{-5}$
	g_3	–	–	–
	w	0.006	0.030	NA
θ_{final}	g_1	**0**	**0**	**0**
	g_2	–	0.008	$1.6 \cdot 10^{-5}$
	g_3	**0**	**0**	**0**
	w	–	0.008	NA

*MC estimates. Missing p_f values (–) can be assumed $< 10^{-3}$, and **0** indicates that $\max_{a\in A, e\in E_2} g_i < 0$.

Table 6
Failure probability — Risk-based evaluation.

θ	Limit-state	Full E		Reduced E	
		p_f max	Severity	p_f max	Severity
$\theta_{p\%\text{risk}}$	g_1	–	–	–	–
	g_2	0.055	$1.2 \cdot 10^{-4}$	0.021	$3.8 \cdot 10^{-5}$
	g_3	–	–	–	–
	w	0.055	NA	0.021	NA
θ_{final}	g_1	–	–	**0**	**0**
	g_2	0.061	$1.6 \cdot 10^{-4}$	0.008	$1.6 \cdot 10^{-5}$
	g_3	–	–	**0**	**0**
	w	0.061	NA	0.008	NA

*MC estimates. Missing p_f values (–) can be assumed $< 10^{-3}$, and **0** indicates that $\max_{a\in A, e\in E} g_i < 0$. Full E and reduced E correspond to $E_{0\%}$ and $E_{50\%}$ in Fig. 10.

the optimization of θ_{final} further, putting less wight on g_1 and g_3, but chose to with the current alternative under the assumption that g_1 and g_3 are the most critical failure modes.

The range of the failure probability, $p_f(e, \theta)$, for each of the limit-states g_1, g_2, g_3 and w, is shown in Tables 4 and 5 for the first and second refined UM's respectively. The tables also include the severity of each individual requirement violation.

8.2. Ranking of epistemic uncertainties

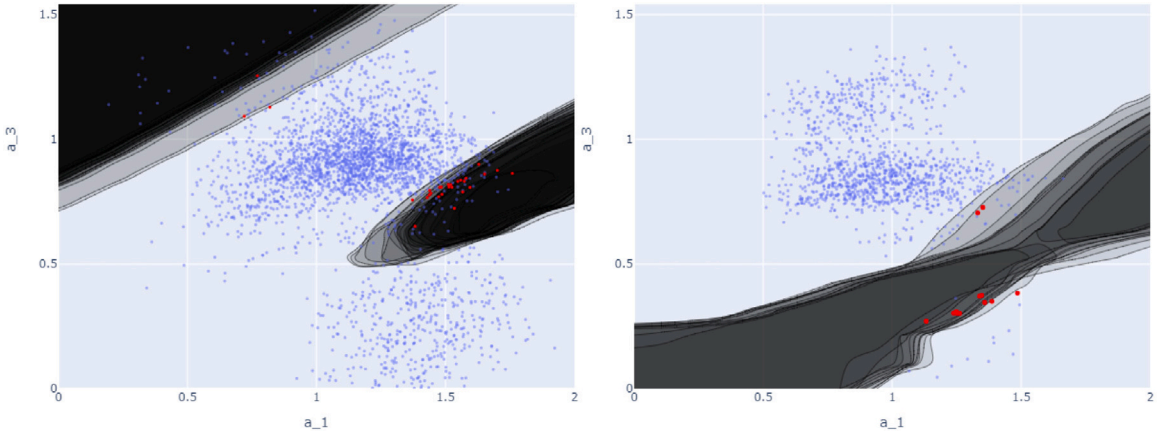
The resulting ranking as described in Section 4.2 is given in Table 7.

8.3. Transition to failure

Based on a set of plausible $e \in E_1$, the design point a^* (most probable point on $g_i = 0$) has been estimated based on MCMC. Fig. 12(a) shows an example of the failure regions for a_1, a_3. The black lines represent the transition boundary $g_1 = 0$ for plausible e's, while the shaded area represent the failure regions where $g_1 > 0$. a_1 and a_3 has been varied across the entire possible range $[0, 2]$ while a_2, a_4, a_5 has been kept fixed at each e's design point a^* plotted as red dots. The blue dots are samples from $f_{a|e}$ for different values of e. Note that, even though some of the blue dots are located inside the failure domains, this does not necessarily

Table 7
e-ranking — First refined UM.

θ	#1	#2	#3	#4
	First refined UM			
θ_{baseline}	e_3	e_4	e_1	e_2
θ_{new}	e_3	e_1	e_2	e_4
	Second refined UM			
θ_{baseline}	e_1	e_2	e_4	e_3
θ_{new}	e_3	e_4	e_1	e_2
θ_{final}	e_4	e_3	e_1	e_2



(a) $g_1(a_1, a_3, \theta_{\text{baseline}}|a^*, e)$ for $e \in E_1$ (b) $g_2(a_1, a_3, \theta_{\text{final}}|a^*, e)$ for $e \in E_2$

Fig. 12. Transition to failure.

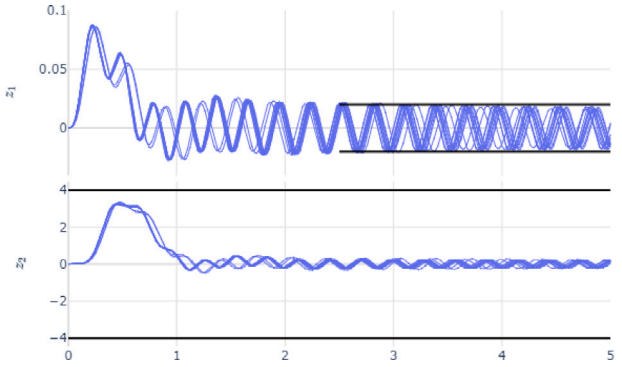


Fig. 13. $\hat{Z}(a^*, e, \theta_{\text{final}})$ for $e \in E_2$.

mean that these are failure points, it is just their projection down in a_1, a_3. But it gives an indication of where the main mass of the distribution lies with respect to the failure regions.

The failure boundaries in Fig. 12(a) are quite co-located in the a_1, a_3 space, which shows that for the e-box E_1 the failure region is reasonably stable in this parameter space. Other parameter spaces do not have a similar clear failure region.

Fig. 12(b) shows a similar plot of the g_2 failure region and transition boundary for plausible $e \in E_2$ in the a_1, a_3 parameter space, and Fig. 13 shows the corresponding time responses of the integrated system.

9. Concluding remarks

We conclude by summing up some of the key lessons learned through the challenge. A central part of the problem is dealing with functional (time-series) data, which makes inference challenging, especially without underlying knowledge about what the data represents. In particular, the nonparametric route can give challenges with MCMC, and ensuring that this works as correctly can

be time consuming. With some assumptions regarding the input distribution (if correct), we could make use of a more numerically stable alternative.

The combination of epistemic and aleatory uncertainty also makes the Bayesian inference challenging. Especially with a set of observations that is relatively small, combined with a nonlinear mapping that makes the true input distribution unidentifiable. That is, there could be different values of the epistemic variable e, combined with quite different distributions over the aleatory variable a, that could explain the data equally well. But for the objectives posed in this challenge, inferring the correct distribution of a is not really important, as long as we can find designs that are robust to any of the plausible distributions $f_{a|e}$.

Our initial attempt was to try an entropy-based approach based on information geometry. Here we would let the aleatory distribution be the distribution with maximum entropy that satisfies some constraints set by the data. This is possible, but numerically challenging, and we did not find a way to develop a numerically stable method that could handle time-series data. As discussed in Section 2.1, we did FPCA (Functional Principal Component Analysis) on the time-series data, which gives a set of uncorrelated features that can be used for dimensionality reduction. Interestingly, these features still had a very complicated dependency structure, and the theoretical motivation for using FPCA was then not so relevant.

We define the different epistemic sets E using a threshold on the likelihood (plausibility) of the epistemic variable e, and this threshold value is something that we set manually. We chose a non-conservative set ($\theta_{50\%\text{risk}}$ in Section 7) to use in the initial design optimization, followed by a verification of the design using a larger conservative set ($\theta_{0\%\text{risk}}$) of epistemic values. The idea behind this, is that there may be many designs that are acceptable with respect to a large set E, and out of these we want to select one that has good performance with respect to the most likely values of e. We considered a more generic methodology, by establishing an criterion that gives a suitable balance between acceptable performance in the worst-case scenario and optimality in the high-likelihood scenarios. Alternatively by considering Pareto optimality with respect to a criterion for acceptable worst-case performance. But we found it difficult to come up with a meaningful criterion, especially without knowledge about the true physical system.

As discussed in Section 2.2, it is important that model discrepancy is handled appropriately. For inference to be possible, we need to decide when two time series are "basically the same". The choice of metric will have to be made based on some assumptions, which could have a large effect on subsequent analysis. Since we assume noiseless observations, we did not want to impose too much regularization, as this might "wash out" the information given in the relatively small set of observations. As a result, we had to spend some time investigating model discrepancy. In the end, in order to assume zero observational error, we needed to assume a small model discrepancy in order for any form of inference to be possible. As we comment in Section 6 and Remark 2.1 in Section 2.2, our final results may be sensitive to this assumption. But in a real-world scenario, involving either observational noise, model discrepancy, or both, this part of the challenge would of course be treated quite differently.

Declaration of competing interest

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

Acknowledgments

This work has been supported by grant 276282 from the Norwegian Research Council and DNV Group Research and Development, Norway. Finally we would like to thank Luis G. Crespo and Sean P. Kenny from the NASA Langley Research Center for providing an interesting and relevant problem to the research community and for hosting the UQ challenge.

References

[1] C. Agrell, S. Eldevik, O. Gramstad, A. Hafver, Contribution to the NASA langley UQ challenge on optimization under uncertainty, in: ESREL 2020, 2020.
[2] L.G. Crespo, S.P. Kenny, The NASA langley challenge on optimization under uncertainty, Mech. Syst. Signal Process. 152 (2021) 107405.
[3] J.-L. Wang, J.-M. Chiou, H.-G. Müller, Functional data analysis, Annu. Rev. Stat. Appl. 3 (1) (2016) 257–295.
[4] J.F. Hauer, C.J. Demeure, L.L. Scharf, Initial results in Prony analysis of power system response signals, IEEE Trans. Power Syst. 5 (1) (1990) 80–89.
[5] M.C. Kennedy, A. O'Hagan, Bayesian Calibration of computer models, J. R. Stat. Soc. Ser. B Stat. Methodol. 63 (3) (2001) 425–464.
[6] R. Lebrun, A. Dutfoy, An innovating analysis of the Nataf transformation from the copula viewpoint, Probab. Eng. Mech. 24 (2009) 312–320.
[7] E. Brochu, V.M. Cora, N. de Freitas, A tutorial on Bayesian optimization of expensive cost functions, with application to active user modeling and hierarchical reinforcement learning, 2010, arXiv:1012.2599.
[8] D. Foreman-Mackey, D.W. Hogg, D. Lang, J. Goodman, Emcee: The MCMC hammer, PASP 125 (2013) 306–312.
[9] H. Madsen, S. Krenk, N. Lind, Methods of Structural Safety, in: Dover Civil and Mechanical Engineering Series, Dover Publications, 2006.
[10] S. Julier, J. Uhlmann, Unscented filtering and nonlinear estimation, Proc. IEEE 92 (2004) 401–422.
[11] R. van der Merwe, E. Wan, Sigma-point Kalman filters for probabilistic inference in dynamic state-space models, in: Proceedings of the Workshop on Advances in Machine Learning, 2003.