Computational methods for system optimization under uncertainty

Computational Methods for System Optimization Under Uncertainty

Nicola Pedroni

NEMO Group, Dipartimento Energia, Politecnico di Torino, Italy. E-mail: nicola.pedroni@polito.it

In this paper, Subproblems A, B and C of the NASA Langley Uncertainty Quantification (UQ) Challenge on Optimization Under Uncertainty are addressed. Subproblem A deals with the model calibration and (aleatory and epistemic) uncertainty quantification of a subsystem, where a characterization of the parameters of the subsystem is sought by resorting to a limited number (100) of observations. Bayesian inversion is here proposed to address this task. Subproblem B requires the identification and ranking of those (epistemic) parameters that are more effective in improving the predictive ability of the computational model of the subsystem (and, thus, that deserve a refinement in their uncertainty model). Two approaches are here compared: the first is based on a sensitivity analysis within a factor prioritization setting, whereas the second employs the Energy Score (ES) as a multivariate generalization of the Continuous Rank Predictive Score (CRPS). Since the output of the subsystem is a function of time, both subproblems are addressed in the space defined by the orthonormal bases resulting from a Singular Value Decomposition (SVD) of the subsystem observations: in other words, a multivariate dynamic problem in the real domain is translated into a multivariate static problem in the SVD space. Finally, Subproblem C requires identifying the (epistemic) reliability (resp., failure probability) bounds of a given system design point. The issue is addressed by an efficient combination of: (i) Monte Carlo Simulation (MCS) to propagate the aleatory uncertainty described by probability distributions; and (ii) Genetic Algorithms (GAs) to solve the optimization problems related to the propagation of epistemic uncertainty by interval analysis.

Keywords: Singular Value Decomposition, Bayesian Inversion, Sensitivity Analysis, Factor Prioritization, Energy Score, Monte Carlo Simulation, Genetic Algorithms.

1. Introduction

In this work, we consider the mathematical (black-box) model of a subsystem, which includes nine inputs and one time-dependent output. The inputs are uncertain and divided into five purely aleatory variables (described by a possibly joint probability distribution) and four purely epistemic parameters (described by intervals). Within this context, we tackle issues (A), (B) and (C) raised by the NASA Langley Uncertainty Quantification (UQ) Challenge on Optimization Under Uncertainty (Crespo and Kenny (2020)).

In task A the Challengers provide a limited number (i.e., 100) of observations of the physical (sub)system (notice such observations are given in the form of discrete time histories). On this basis, an Uncertainty Model for the (nine) input variables/parameters to the subsystem model should be created. *Bayesian inversion*, carried out by Markov Chain Monte Carlo (MCMC) using the Affine Invariant Ensemble Algorithm (AIES) (Goodman and Weare (2010)), is here employed to address this task. Since the output of the subsystem is a function of time, the approach is applied in the space defined by the orthonormal bases resulting from a *Singular Value Decomposition* (SVD) of the subsystem observations: in other words, a multivariate dynamic problem in the real domain is translated into a multivariate static problem in the SVD space. In addition, the likelihood of the data is evaluated by Kernel Density Estimation (KDE) techniques in the SVD space (Wu et al. (2018)).

The task of uncertainty reduction (B) is tackled in two ways. In the first (namely, *sensitivity analysis* within a 'factor prioritization' setting) we rank the epistemic input parameters according to degree of *reduction* in the output epistemic uncertainty, which one could hope to obtain by refining their (epistemic) uncertainty models, i.e., by reducing the epistemic uncertainty range. A sensitivity index is introduced in analogy with variance-based Sobol indices (Pedroni and Zio (2015)): in this view, the most important epistemic parameters in the ranking are those that give rise to the *highest expected reduction* in the amount of epistemic uncertainty contained in the model output, when the corresponding parameter values are considered *constant* (i.e., when the amount of their epistemic uncertainty is reduced to zero). Notice that the 'amount' of epistemic uncertainty is here defined as the *volume* of the *convex hull* enveloping the realizations of the model output in the orthonormal SVD space (Faes et al. (2019)). In the second approach, the use of the *Energy Score* (ES) (computed in the orthonormal space) is proposed as a multivariate generalization of the Continuous Rank Predictive Score (CRPS) to assess the probabilistic

Proceedings of the 30th European Safety and Reliability Conference and the 15th Probabilistic Safety Assessment and Management Conference
Edited by Piero Baraldi, Francesco Di Maio and Enrico Zio
Published by Research Publishing, Singapore.
ISBN: 978-981-14-8593-0; doi:10.3850/978-981-14-8593-0

predictive capability of the subsystem model (Gneiting and Raftery (2007)). The idea is to rank the epistemic parameters according to their capability to *improve* the *predictive ability* of the model, i.e., to *decrease* the ES, when their epistemic uncertainty is *reduced*.

The task (C) of reliability analysis of baseline design (requiring the identification of the epistemic reliability – resp., failure probability – bounds of a given system design point) is addressed by an *efficient combination* of: (i) *Monte Carlo Simulation* (MCS) to propagate the aleatory uncertainty described by probability distributions; and (ii) *Genetic Algorithms* (GAs) to solve the optimization problems related to the propagation of epistemic uncertainty by interval analysis.

2. The Model of the Subsystem

The system of interest is modelled as a set of interconnected subsystems. The uncertain parameter δ is concentrated onto a single subsystem. This subsystem is modelled by the function $y(\boldsymbol{a}; \boldsymbol{e}; t)$, where $\boldsymbol{a}$ is a n_a-dimensional vector of aleatory variables (n_a = 4), $\boldsymbol{e}$ is a n_e-dimensional vector of epistemic parameters (n_e = 4) and t is time. Hence, the output of the subsystem is a function of time, which is given as a discrete time history, $y(t) = [y(0), \ldots, y(T)]$. The Uncertainty Model (UM) for $\boldsymbol{a}$ is denoted as $f_{\boldsymbol{a}}$, where $f_{\boldsymbol{a}}$ is a joint density supported in the set $A = [0, 2]^{na}$. In contrast, the UM for $\boldsymbol{e}$ is denoted as E, where E is a hyper-rectangular set included in $[0, 2]^{ne}$. Hence, the UM of is fully prescribed by the pair $\langle f_{\boldsymbol{a}}, E \rangle$ (Crespo and Kenny (2020)).

3. Approaches to the NASA Langley Uncertainty Quantification (UQ) Challenge on Optimization Under Uncertainty

3.1 *Subproblem A: Model Calibration & Uncertainty Quantification of a Subsystem*

Given the time-dependent (i.e., multivariate) nature of the function $y(\boldsymbol{a}, \boldsymbol{e}, t)$ (used to model the subsystem of interest), the calibration and uncertainty quantification of parameter δ (fully prescribed by the pair $<f_a, E>$) are carried out in the space of the orthogonal basis vectors resulting from Singular Value Decomposition (SVD) of the data $\boldsymbol{D}_1 = \{y^{(i)}(t)\}$, $i = 1, 2, \ldots, n_1 = 100$, $t = 0, 1, \ldots, T$ (Wu et al., 2018). In particular, the following steps are performed to pre-process the data before the model calibration and uncertainty quantification tasks:

1. Evaluate the (time-dependent) sample mean $m(t)$ of the dataset $\boldsymbol{D}_1$ as $m(t) = 1/n1 \cdot \sum_{i=1}^{n1} y^{(i)}(t)$.
2. Subtract the mean $m(t)$ from the available time-series $\{y^{(i)}(t)\}$ to obtain the *centred* data $\boldsymbol{D}_1{}^* = \{y^{*(i)}(t)\} = \{y^{(i)}(t)\} - m(t)$, $i = 1, 2, \ldots, n_1 = 100$, $t = 1, 2, \ldots, N_T$ (N_T = 5001 is the number of discrete time steps where the output y of the subsystem is evaluated).
3. Perform an SVD of the centered data $\{y^{*(i)}(t)\}$. If $\boldsymbol{D}_1{}^*$ is the (n_1 x N_T) centered matrix of the system realization, then it can be expressed as $\boldsymbol{U}^*\boldsymbol{S}^*\boldsymbol{V}'$, where: $\boldsymbol{U}$ is the (n_1 x n_1) matrix of left singular vectors; $\boldsymbol{S}$ is the (n_1 x N_T) diagonal matrix of the nonnegative singular values s_i, $i = 1, 2, \ldots, n_1$, in decreasing order; and $\boldsymbol{V}$ is the (N_T x N_T) matrix of right singular vectors: in other words, the columns of $\boldsymbol{V}$ contains N_T orthonormal N_T-dimensional (eigen)vectors $\boldsymbol{v}_t$, $t = 1, 2, \ldots, N_T$, constituting an orthonormal basis ß for $\boldsymbol{D}_1{}^*$.
4. In order to reduce the dimensionality of the problem (while accounting for the overall variability of the model output of interest), select a proper number $N_B < N_T$ of basis vectors to be retained in the analysis. In this work, n_B is selected so that at least $\varepsilon = 99\%$ of the variance associated to the $n_1 = 100$ observed time histories is explained. In details:

$$\sum_{i=1}^{n_B} \frac{s_i^2}{\sum_{j=1}^{n_1} s_j^2} \geq \varepsilon = 0.99, \qquad (1)$$

noting that $\sum_{j=1}^{n_1} s_j^2$ equals the overall variance of the entire dataset. In this case, the value of n_B turns out to be 10.
5. Project the centred dataset $\boldsymbol{D}_1{}^*$ onto the orthonormal basis defined by the n_B eigenvectors $\boldsymbol{v}_t$ corresponding to the n_B largest singular values s_t, $t = 1, 2, \ldots, n_B$, of matrix $\boldsymbol{S}$. In particular, the (n_1 x n_B) matrix $\boldsymbol{C} = \{c_{it}\}$ containing the projections of the centred data $\boldsymbol{D}_1{}^*$ onto the orthonormal basis ß = $\{\boldsymbol{v}_t, t = 1, 2, \ldots, n_B\}$, is obtained as:

$$\boldsymbol{C} = \boldsymbol{D_1}^* \cdot \boldsymbol{V}[1{:}\, n_B] =$$

$$\boldsymbol{D_1}^* \cdot \left[\boldsymbol{v}_1, \boldsymbol{v}_2, \ldots, \boldsymbol{v}_{n_B}\right]. \qquad (2)$$

The coefficients/projections $\{c_{it}: i = 1, 2, \ldots, n_1, t = 1, 2, \ldots, n_B\}$ are equivalently expressed as:

$$c_{it} = \sum_{k=1}^{N_T} y^{*(i)}(k) \cdot v_t(k) \ . \qquad (3)$$

The idea is to perform the calibration and uncertainty quantification tasks in the (static multivariate) *projected* space (i.e., in the space

defined by the orthonormal basis ß) rather than in the (dynamic multivariate) time domain.

The functional form of f_a (representing the Uncertainty Model-UM of the aleatory variables $\boldsymbol{a}$) is chosen as a multivariate Gaussian Mixture (GM). A GM model is a probabilistic model that assumes the data are drawn from a mixture of a n_G (multivariate) Gaussian distributions with unknown hyper-parameters. Such hyper-parameters and their corresponding weights are prescribed by minimizing a measure of the offset between the data and the prediction. The Probability Density Function (PDF) of a GM model is given by:

$$f_a^{GM}(\boldsymbol{a}|\boldsymbol{w},\boldsymbol{\mu},\boldsymbol{\Sigma}) = \qquad f_a^{GM}(\boldsymbol{a}|\boldsymbol{\varphi_a}) = \sum_{l=1}^{n_G} w_l \cdot \mathcal{N}(\boldsymbol{\mu}_l,\boldsymbol{\Sigma}_l), \tag{4}$$

where the l-th component of the mixture is characterized by a multivariate Normal distribution with weight $w_l \in [0, 1]$, means $\boldsymbol{\mu}_l$ and covariance matrix $\sum_l$ (notice that in *each* multivariate Gaussian distribution $\mathcal{N}(\boldsymbol{\mu}_l,\boldsymbol{\Sigma}_l)$ of the mixture, the covariance matrix $\sum_l$ is able to capture and describe *only linear dependences* between the aleatory variables $\boldsymbol{a}$). For the sake of compact notation, the ensemble of calibration hyper-parameters of the joint aleatory PDF f_a is indicated as $\boldsymbol{\varphi_a} = \{w_l, \boldsymbol{\mu}_l, \sum_l: l = 1, 2, \ldots, n_G\}$. This functional form has been chosen as the best possible because of its high flexibility: by adjusting the corresponding hyper-parameters, a GM can be forced to assume different (unimodal and multi-modal) shapes and to describe a large variety of dependence structures among the aleatory variables (Gaymann et al. (2018)).

The calibration approach adopted in this work (see Section 3.1.1 below) is based on the evaluation of the joint multivariate likelihood of the data in the projected space defined by the orthonormal basis ß, $L^{\mathrm{GM}}(\boldsymbol{C}|\boldsymbol{\Phi}) = L^{\mathrm{GM}}(\boldsymbol{C}|\boldsymbol{\varphi_a}, \boldsymbol{e}) = L^{\mathrm{GM}}(\boldsymbol{D}_1{*}|w_l, \boldsymbol{\mu}_l, \sum_l: l = 1, 2, \ldots, n_G; e_1, e_2, e_3, e_4)$, where $\boldsymbol{\Phi}$ is a vector containing the ensemble of *all* the calibration parameters (i.e., the hyper-parameters of the GM model and the epistemic parameters $\boldsymbol{e}$). The likelihood is here approximated by Monte Carlo Simulation (MCS) of the model $y(\boldsymbol{a}, \boldsymbol{e}, t)$, SVD of the corresponding response and finally Kernel Density Estimation (KDE) in the projected space ß. The detailed algorithm is as follows:

Inputs: $\boldsymbol{\Phi} = [\boldsymbol{\varphi_a}, \boldsymbol{e}]$, $\boldsymbol{C}$

Output: $L^{\mathrm{GM}}(\boldsymbol{C}|\boldsymbol{\Phi})$

1. Generate N_{like} realizations of $\boldsymbol{a}$ ($\boldsymbol{a}_q$, $q = 1, 2, \ldots, N_{like}$) by random sampling from the corresponding PDF $f_a^{GM}(\boldsymbol{a}|\boldsymbol{\varphi_a})$ (4). In this work, $N_{like} = 100000$.
2. For each $\boldsymbol{a}_q$ evaluate the model response $y(\boldsymbol{a}_q, \boldsymbol{e}, t)$, $q = 1, 2, \ldots, N_{like}$. Let $\boldsymbol{Y^{\Phi}}$ be the (N_{like} x N_T) matrix containing such responses.
3. Subtract the mean $m(t)$ of the *dataset* $\boldsymbol{D}_1$ ($1/n1 \cdot \sum_{i=1}^{n1} y^{(i)}(t)$) from the simulated time-series $y(\boldsymbol{a}_q, \boldsymbol{e}, t)$, $q = 1, 2, \ldots, N_{like}$, to obtain the model responses $\boldsymbol{Y^{*\Phi}}$, "*centered*" with respect to the *mean value* of the *real data* (i.e., with respect to the mean value of the "true" system response). Then, project $\boldsymbol{Y^{*\Phi}}$ onto the orthonormal basis ß = $\{\boldsymbol{v}_t, t = 1, 2, \ldots, n_B\}$ found above, in order to obtain the (N_{like} x n_B) matrix $\boldsymbol{H} = \{h_{qt}\}$ of the corresponding coefficients (projections) $\boldsymbol{H} = \boldsymbol{Y^{*\Phi}} \cdot \boldsymbol{V}[1: n_B] = \boldsymbol{Y^{*\Phi}} \cdot [\boldsymbol{v}_1, \boldsymbol{v}_2, \ldots, \boldsymbol{v}_{nB}]$.
4. Based on the (projected) model responses $\boldsymbol{H} = \{h_{qt}\}$ and relying on KDE techniques, estimate the likelihood $L^{\mathrm{GM}}(\boldsymbol{C}|\boldsymbol{\Phi})$ in the space defined by the orthonormal basis ß. In this paper, a *multivariate* product Gaussian kernel is employed, so that $L^{\mathrm{GM}}(\boldsymbol{C}|\boldsymbol{\Phi})$ becomes:

$$L^{GM}(\boldsymbol{C}|\boldsymbol{\Phi}) = \prod_{i=1}^{n_1} \frac{1}{N_{like} b_1 b_2 \ldots b_{nB}} \sum_{q=1}^{Nlike} \prod_{t=1}^{nB} K\left(\frac{c_{it} - h_{qt}}{b_t}\right). \tag{5}$$

where: b_t, $t = 1, 2, \ldots, n_B$, are the one-dimensional bandwidths of the kernel-smoothing windows, calculated applying the Silverman's rule to the N_{like} simulated output projections $\boldsymbol{H} = \{h_{qt}\}$; and $K(\cdot)$ is a one-dimensional Gaussian kernel function. It is worth noting that even if $L^{\mathrm{GM}}(\boldsymbol{C}|\boldsymbol{\Phi})$ (5) uses a product of one-dimensional kernels, this does *not* imply that the n_B variables are *independent* (this is obviously due to the fact that we employ a *sum* of N_{like} products of n_B one-dimensional kernels).

A final consideration is in order. Centering the simulated time-series $y(\boldsymbol{a}_q, \boldsymbol{e}, t)$ with respect to the mean $m(t)$ of the real data (step 3 above) guarantees that the likelihood $L^{\mathrm{GM}}(\boldsymbol{C}|\boldsymbol{\Phi})$ thereby generated drives the calibration process to match also the true mean of the system response (besides the 99% of the variance, as specified in (1)).

The likelihood $L^{\mathrm{GM}}(\boldsymbol{C}|\boldsymbol{\Phi})$ (5) is here employed to carry out model calibration and uncertainty quantification by *Bayesian inversion*, which is aimed at finding the joint posterior distribution of the ensemble of the calibration parameters $\boldsymbol{\Phi}$ (Section 3.1.1).

3.1.1 *Bayesian Model Calibration and Inverse Uncertainty Quantification*

A GM model with $n_{GM} = 2$ multivariate Gaussian distributions is chosen (notice that also the option with $n_{GM} = 3$ has been tested: in spite of the

dramatic increase in the number of calibration hyper-parameters, no significant improvement in the likelihood of the observations is registered). Uniform priors $\pi(\boldsymbol{\Phi})$ are assigned to the parameters $\boldsymbol{\Phi}$, which range in the following intervals (hyper-rectangles): the weights $\boldsymbol{w}$ in [0, 1], the Gaussian means $\boldsymbol{\mu}$ in $B = [0, 2]^{na}$, the Gaussian variances $\boldsymbol{\sigma}^2$ in $[0.0025, 16]^{na}$ and the Pearson correlation coefficients ρ_{ij}, $i = 1, 2, \ldots, n_a$-1, $j = i+1, \ldots, n_a$ (used to build the covariance matrices $\sum$), in [-1, 1]; the epistemic parameters $\boldsymbol{e}$ are defined in $E_0 = [0, 2]^{ne}$. The total number of decision variables is thus equal to 45.

Markov Chain Monte Carlo (MCMC) based on the Affine Invariant Ensemble Algorithm (AIES) (Goodman and Weare (2010)) (with N_{chain} = 100 chains and N_{MCMC} = 2000 samples per chain) is run to obtain the (joint) posterior distribution of $\boldsymbol{\Phi}$, i.e., $\pi(\boldsymbol{\Phi}|\boldsymbol{C}) = (L^{GM}(\boldsymbol{C}|\boldsymbol{\Phi})\cdot\pi(\boldsymbol{\Phi}))/\pi(\boldsymbol{C})$. The total number of model evaluations associated to the method is thus $N_{MCMC}\cdot N_{chain}$ = 200000. In Table 1 the Maximum A Posteriori (MAP) estimates of some calibration parameters ($\boldsymbol{w}_{MAP}$, $\boldsymbol{\mu}_{MAP}$, $\boldsymbol{\sigma}_{MAP}^2$) are reported. For the sake of brevity, we do not report the complete expressions of the correlation matrices $\sum_l$, l = 1, 2, associated to the two components of the GM model. However, in the following we provide an *overall* evaluation of the degree of dependency between the parameters $\boldsymbol{a}$ of $f_{\boldsymbol{a}}$ by means of a single *rank* correlation matrix $\boldsymbol{R}_{\boldsymbol{a},Spear}$ (in particular, based on the Spearman measure). The choice of a rank correlation coefficient is justified by its *nonparametric* nature and its *invariance* with respect to the marginals of the dependent random variables analyzed. Notice that the Spearman correlation between two variables is equal to the Pearson correlation between the *rank values* of those two variables; while Pearson's correlation assesses *linear relationships*, Spearman's correlation assesses *monotonic relationships* (whether *linear or not*). Notice that such dependency evaluation has been made by means of a sample of 100000 realizations of $\boldsymbol{a}$, generated from the PDF $f_{\boldsymbol{a}}$ with hyper-parameters fixed at their respective MAP values.

Table 1. MAP estimates for w_{MAP}, $\boldsymbol{\mu}_{MAP}$, and $\boldsymbol{\sigma}_{MAP}^2$ of the GM models resulting from the Bayesian calibration process.

	Parameter	MAP estimate
Gaussian Model 1	w_1	0.1880
	μ_1	0.2375
	μ_2	1.9817
	μ_3	0.6725
	μ_4	1.8965
	μ_5	1.7357
	σ_1^2	14.6103
	σ_2^2	9.2151
	σ_3^2	1.0544
	σ_4^2	1.5200
	σ_5^2	3.6629
Gaussian Model 2	w_2	0.8120
	μ_1	0.6492
	μ_2	1.6729
	μ_3	1.2502
	μ_4	0.8679
	μ_5	1.9241
	σ_1^2	7.0061
	σ_2^2	14.9051
	σ_3^2	1.0325
	σ_4^2	14.1371
	σ_5^2	13.0113

The Spearman *rank* correlation matrix $\boldsymbol{R}_{\boldsymbol{a},Spear}$ is as follows:

$$\boldsymbol{R}_{a,Spear} = \begin{bmatrix} 1 & 0.394 & 0.110 & -0.094 & 0.068 \\ 0.394 & 1 & -0.162 & 0.229 & -0.150 \\ 0.110 & -0.162 & 1 & 0.035 & 0.851 \\ -0.094 & 0.229 & 0.035 & 1 & -0.081 \\ 0.068 & -0.150 & 0.851 & -0.081 & 1 \end{bmatrix} \quad (6)$$

The sign of the Spearman correlation indicates the direction of association between variables a_i and a_j. If a_i tends to increase when a_j increases (resp., decreases), the Spearman correlation coefficient is positive (resp., negative). Also, the Spearman correlation increases in magnitude as a_i and a_j become closer to being perfectly monotonically related (see, e.g., a_3 and a_5, whose rank correlation coefficient is 0.851). It must be acknowledged that while $\boldsymbol{R}_{\boldsymbol{a},Spear}$ (6) represents a synthetic and easily interpretable measure of the *strength* of *correlations*, it may obviously fail to fully and accurately describe the possibly *complex* and *nonlinear* patterns of *dependence* between aleatory variables.

The uncertainty model E for the epistemic parameters $\boldsymbol{e}$ is chosen as follows: it is the hyper-rectangle $[\underline{e}_1, \overline{e}_1]$ x $[\underline{e}_2, \overline{e}_2]$ x $[\underline{e}_3, \overline{e}_3]$ x $[\underline{e}_4, \overline{e}_4]$ that envelops the (joint four-dimensional) 99% confidence interval of $\boldsymbol{e}$, resulting from the Bayesian calibration. The results are the following: $[\underline{e}_1, \overline{e}_1]$ = [0, 0.3719], $[\underline{e}_2, \overline{e}_2]$ = [0.1910, 0.7273], $[\underline{e}_3, \overline{e}_3]$ = [0, 0.8543] and $[\underline{e}_4, \overline{e}_4]$ = [0, 2]. The corresponding MAPs are 0.0704, 0.5176, 0.0411 and 1.9059, respectively.

3.2 Subproblem B: Uncertainty Reduction

The epistemic parameters should be ranked according to their ability to improve the *predictive capability* of the computational model of the subsystem. In other words, as specified by the challengers, the epistemic parameters leading to the largest *reduction* in the output's *spread* should be identified. Two approaches have been developed to address this task: the first is based on *sensitivity analysis* within a 'factor prioritization' setting in analogy with variance-based Sobol' indices (Section 3.2.1); the second relies on the evaluation of the energy score, i.e., a multivariate generalization of the *Continuous Rank Predictive Score* (CRPS), to measure the predictive capability of a stochastic computational model (Section 3.2.2).

3.2.1 *Sensitivity Analysis in a 'Factor Prioritization' Setting*

This approach is aimed at assessing how much less epistemic uncertainty the model output of interest (resp., higher predictive capability) would have if extra knowledge about an input were available. This can be done by comparing the epistemic uncertainty before and after '*pinching*' an input, i.e. replacing it with a value without (or with less) epistemic uncertainty. Quantifying this effect assesses the contribution by the input epistemic uncertainty to the overall epistemic uncertainty in the output of interest. The estimate of the value of information for an epistemically uncertain parameter/coefficient will depend on (i) how much epistemic uncertainty is present in the parameter, and (ii) how it affects the epistemic uncertainty in the result.

Let $U_{\boldsymbol{e}}$ be an indicator of the 'amount' of epistemic uncertainty contained in the output of the computational model $y(\boldsymbol{a}, \boldsymbol{e}, t)$. The subscript '$\boldsymbol{e}$' suggests that indicator $U_{\boldsymbol{e}}$ is computed over all the (epistemic) input parameters $\boldsymbol{e}$ (and over the corresponding space of variation E). The indicator $U_{\boldsymbol{e}}$ could be obviously measured in different ways (e.g., straightforwardly by the overall variance of y). In this paper, coherently with the approach presented in Section 3.1, the uncertainty in the subsystem model output y is measured by the overall *spread* of y in the projected space defined by the orthonormal bases $ß = \{\boldsymbol{v}_t, t = 1, 2, \ldots, n_B\}$. Such spread is here quantified by the *volume* $\mathcal{G}$ of the *convex hull* able to envelop the projections of $y(\boldsymbol{a}, \boldsymbol{e}, t)$ onto ß (i.e., $\mathcal{G}_{\boldsymbol{e}} = U_{\boldsymbol{e}}$). Obviously, the larger the volume of the convex hull, the larger the spread of the output y, i.e., the larger the (epistemic) uncertainty in y (Faes et al. (2019)). Notice that the computational burden related to the calculation of the convex hull volume in $n_B = 10$ dimensions may be prohibitive. Thus, the strategy proposed in (Faes et al. (2019)) is adopted: an approximation to the volume $\mathcal{G}$ is obtained using 45 2-dimensional projections of the full 10-dimensional convex hull (which reduces the computational time by more than two orders of magnitude).

We want to assess - by means of a sensitivity index $S_i(U_{\boldsymbol{e}})$ - the effect that a refinement of the uncertainty model of the generic epistemic input e_i (i.e., a reduction in its epistemic uncertainty) has on the amount of epistemic uncertainty $U_{\boldsymbol{e}}$ of the model output. In order to address this issue, a sensitivity index is introduced in analogy with variance-based Sobol' indices (Pedroni and Zio (2015)). Imagine that we fix e_i at a particular value e_i^* in $[\underline{e}_i, \overline{e}_i]$. Let $U_{\boldsymbol{e}}(e_i = e_i^*)$ be the resulting amount of epistemic uncertainty in $y(\boldsymbol{a}, \boldsymbol{e}_{-i}, t)$, taken over all parameters $\boldsymbol{a}$ and $\boldsymbol{e}_{-i}$ and keeping the epistemically-uncertain parameter e_i fixed at e_i^* (instead, all the other epistemically-uncertain coefficients $\boldsymbol{e}_{-i}$ are allowed to range in their corresponding space of variation E_{-i}). We would imagine that having frozen one potential source of epistemic uncertainty (e_i), the resulting indicator $U_{\boldsymbol{e}}(e_i = e_i^*)$ will be lower than the corresponding total (or unconditional) one $U_{\boldsymbol{e}}$. One could therefore conceive of using $U_{\boldsymbol{e}}(e_i = e_i^*)$ as a measure of the relative importance of e_i, reasoning that the smaller $U_{\boldsymbol{e}}(e_i = e_i^*)$, the greater the influence of e_i. However, notice that this approach makes the sensitivity measure dependent on the position of the point e_i^* for each input factor, which is impractical. Thus, we take the average of the measure $U_{\boldsymbol{e}}(e_i = e_i^*)$ over all the possible points e_i^* in $[\underline{e}_i, \overline{e}_i]$, which removes the dependence on e_i^*. The resulting indicator is then written synthetically as $E_{ei}[U_{\boldsymbol{e}}(e_i)]$ and represents the *expected* amount of epistemic uncertainty contained in the output y when the epistemically-uncertain parameter e_i is fixed to a constant value (i.e., when the amount of its epistemic uncertainty is reduced to zero). Obviously, the lower the value of $E_{ei}[U_{\boldsymbol{e}}(e_i)]$, the more important the corresponding parameter e_i: in other words, the most important parameter is that parameter which *on average*, once *fixed*, causes the greatest reduction in the epistemic uncertainty of y (as highlighted above, the consideration of "average sensitivities" is due to the fact that $U_{\boldsymbol{e}}(e_i = e_i^*)$ is in general strongly dependent on the position of the point e_i^*: this suggests the necessity to calculate the average of the measure $U_{\boldsymbol{e}}(e_i = e_i^*)$ over many possible points e_i^* in $[\underline{e}_i, \overline{e}_i]$, in order to obtain robust and reliable sensitivity rankings). Finally, the sensitivity $S_i(U_{\boldsymbol{e}})$ of the epistemic uncertainty of the output y to the epistemic uncertainty of parameter e_i can be synthesized with an expression like

$$S_i(U_{\boldsymbol{e}}) = 1 - \frac{E_{ei}[U_{\boldsymbol{e}}(e_i)]}{U_{\boldsymbol{e}}}. \quad (7)$$

Index (7) is an estimate of the *value* of *additional* empirical information about the input e_i in terms of the *fractional reduction* in epistemic uncertainty that might be achieved in y when the input parameter is replaced by a better estimate obtained from future empirical study. This 'pinching' procedure can be applied to each input quantity in turn and the results used to rank the inputs in terms of their sensitivities (i.e., in terms of their capability of reducing the output spread).

In this paper, the sensitivity index $S_i(U_e)$ (7) related to the generic parameter e_i is straightforwardly estimated as follows (Pedroni and Zio (2015)):

1. letting all the epistemically-uncertain parameters $\boldsymbol{e}$ range within the entire space of variation E, propagate the mixed aleatory and epistemic uncertainty from the inputs $\boldsymbol{a} \sim f_{\boldsymbol{a}}$ and $\boldsymbol{e}$, respectively, to the output of interest $y(\boldsymbol{a}, \boldsymbol{e}, t)$ and evaluate the resulting (total, unconditional) amount of epistemic uncertainty U_e.
2. select (deterministically or stochastically) N_e values e_i^k, $k = 1, 2, \ldots, N_e$, of the epistemically uncertain parameter e_i under analysis within its interval of variation $[\underline{e}_i, \overline{e}_i]$. These N_e realizations of epistemic uncertainty e_i^k, $k = 1, 2, \ldots, N_e$, should be chosen in such a way to evenly cover the corresponding interval $[\underline{e}_i, \overline{e}_i]$: in this paper, a sequence of 20 equally spaced points is adopted to this aim.
3. fixing the value of e_i to e_i^k, $k = 1, 2, \ldots, N_e$, and letting all the other epistemically uncertain parameters $\boldsymbol{e}_{-i}$ vary within E_{-i}, propagate the mixed aleatory and epistemic uncertainty from the inputs $\boldsymbol{a} \sim f_{\boldsymbol{a}}$ and $\boldsymbol{e}_{-i}$ to the output of interest $y(\boldsymbol{a}, \boldsymbol{e}, t)$ and evaluate the resulting (conditional) amount of epistemic uncertainty $U_e(e_i = e_i^k) = \mathcal{G}_{\boldsymbol{e}}\left(e_i = e_i^k\right)$ in y. The propagation of the mixed aleatory and epistemic uncertainty is carried out with $N_a = 10000$ samples.
4. estimate the index (7) as $S_i(U_e) = 1 - 1/N_e \cdot \sum_{k=1}^{N_e} \frac{U_e\left(e_i = e_i^k\right)}{U_e}$.

The total number of model evaluations required by the method is thus $N_a \cdot N_e \cdot n_e$ (= 800000 in this case). The values obtained for $S_i(U_e)$ are shown in Table 2, together with the corresponding epistemic parameter ranking. The author's choice is thus to refine parameters e_2 and e_3. The side of the epistemic interval to refine is determined as the one leading to the *largest contraction* of the interval, while still *including* the MAPs: in this case, the lower bound of e_2 should be increased, whereas the upper bound of e_3 should be reduced.

Table 2. Results of the sensitivity analysis in a factor prioritization setting.

Parameter (MAPs)	$S_i(U_e)$	Ranking
e_1 (0.0704)	0.0125	3
e_2 (0.5176)	0.1048	1
e_3 (0.0411)	0.0235	2
e_4 (1.9059)	$7.65 \cdot 10^{-6}$	4

3.2.2 *The Energy Score (ES) as a multivariate generalization of the Continuous Rank Predictive Score (CRPS)*

The Continuous Rank Predictive Score (CRPS) is arguably the most versatile scoring rule for probabilistic forecasts of a univariate scalar variable. It measures the *distance* between the CDF of the provided data (i.e., realizations/measurements of the real system of interest) and the CDF of the forecast data, i.e., data generated based on the predictive model. To assess probabilistic forecasts of a *multivariate* quantity (like the one of interest in the present Challenge), the use of the *Energy Score* (ES) (computed in the orthonormal space ß) is proposed (Gneiting and Raftery (2007)). The idea is to rank the epistemic parameters according to their capability to *improve* the *predictive ability* of the model (i.e., to *decrease* the ES), when their epistemic uncertainty is *reduced*.

Let $\boldsymbol{D}_1$ be the available dataset in the original measurement space and $\boldsymbol{C} = \{\boldsymbol{c}_i: i = 1, 2, \ldots, n_1 = 100\} = \{c_{it}: i = 1, 2, \ldots, n_1 = 100, t = 1, 2, \ldots, n_B\}$ the matrix containing the corresponding projections onto the orthonormal space ß. Let $y^{(q)}(\boldsymbol{a}, \boldsymbol{e}, t)$, $\boldsymbol{a} \sim f_{\boldsymbol{a}}$, $\boldsymbol{e} \in E$, $q = 1, 2, \ldots, N_s$, be a collection of N_s randomly generated realizations of the subsystem model, whose uncertainty is prescribed by the results of Subproblem A, and $\boldsymbol{H} = \{\boldsymbol{h}_q: q = 1, 2, \ldots, N_s\} = \{h_{qt}: q = 1, 2, \ldots, N_s, t = 1, 2, \ldots, n_B\}$ the matrix containing corresponding projections onto the orthonormal space ß. The evaluation of the predictive capability of model $y(\boldsymbol{a}, \boldsymbol{e}, t)$ (characterized by the UM $\boldsymbol{a} \sim f_{\boldsymbol{a}}$, $\boldsymbol{e} \in E$) with respect to the generic projected datum $\boldsymbol{c}_i$, $i = 1, 2, \ldots, n_1 = 100$, is:

$$ES(\langle f_{\boldsymbol{a}}, E\rangle, \boldsymbol{c}_i) = \frac{1}{N_s}\sum_{q=1}^{N_s}\left\|\boldsymbol{h}_q - \boldsymbol{c}_i\right\| - \frac{1}{2N_s^2}\sum_{q=1}^{N_s}\sum_{j=1}^{N_s}\left\|\boldsymbol{h}_q - \boldsymbol{h}_j\right\|. \quad (8)$$

where $\|\cdot\|$ is the Euclidean norm. It can be demonstrated that if the number n_B of dimensions equals 1, then Eq. (8) reduces to the well-known CRPS (Székely and Rizzo (2005)). Also, it is straightforward to notice that the smaller (8), the smaller the average distance between the (projected) model predictions $\boldsymbol{h}_q$ and the (projected) datum $\boldsymbol{c}_i$, i.e., the better the predictive capability of the model.

In order to have a global measure of the *overall* predictive capability of the model, an average of (8) over the entire dataset $\boldsymbol{C}$ is carried out:

$$ES(\langle f_{\boldsymbol{a}}, E\rangle) = \frac{1}{n_1}\sum_{i=1}^{n_1} ES(\langle f_{\boldsymbol{a}}, E\rangle, \boldsymbol{c}_i). \quad (9)$$

In order to assess the capability of the epistemic parameters $\boldsymbol{e}$ to *improve* the *predictive ability* of the model (i.e., to *decrease* the ES, when their epistemic uncertainty is *reduced*), a procedure similar to that outlined in the previous Section 3.2.1 is adopted. Each e_i is fixed at different values e_i^k, $k = 1, 2, \ldots, N_e = 20$, within its range of variation $[\underline{e}_i, \overline{e}_i]$. In correspondence of each e_i^k a value of the ES is computed as $ES\left(\langle f_{\boldsymbol{a}}, E_{-i}|e_i = e_i^k \rangle\right)$ according to (9) (for the sake of compact notation, let $ES\left(\langle f_{\boldsymbol{a}}, E_{-i}|e_i = e_i^k \rangle\right)$ be indicated as $ES\left(e_i = e_i^k\right)$. In this paper, the evaluation of (9) is obtained by resorting to N_s = 10000 randomly generated realizations of the subsystem model (for each epistemic value e_i^k). Then, the parameter characterized by the *highest* ability to improve the predictive capability of the computational model is the one with the *minimum* value of $ES\left(e_i = e_i^k\right)$ computed over its range $[\underline{e}_i, \overline{e}_i]$. In this view, the corresponding sensitivity indicator $S_i(ES)$ is computed as the ratio between the minimum ES obtained for the "reduced" epistemic model, i.e., $\min_{ei}\{ES(e_i)\}$, and the ES associated to the full epistemic model $ES(\langle f_{\boldsymbol{a}}, E\rangle)$:

$$S_i(ES) = \frac{\min_{ei}\{ES(e_i)\}}{ES(\langle f_{\boldsymbol{a}}, E\rangle)} \quad (10)$$

Obviously, the smaller (10), the higher the capability of parameter e_i of improving the predictive capability of the model. The total number of model evaluations required by the method is $N_s \cdot N_e \cdot n_e$ (= 800000 in this paper).

The values obtained for $S_i(ES)$ are shown in Table 3, together with the corresponding epistemic parameter ranking. In addition, the point values of the epistemic parameters, for which the ES is minimum, are indicated in parenthesis. As before, the author's choice is to refine parameters e_2 and e_3. The side of the epistemic interval to refine is determined as the one leading to the *largest contraction* of the interval, while *including* the point value leading to the minimum ES: coherently with the sensitivity based approach, the lower bound of e_2 should be increased, whereas the upper bound of e_3 should be reduced.

Table 3. Results of the ranking based on the ES.

Parameter	$S_i(ES)$	Ranking
e_1 (0.2349)	0.9900	3
e_2 (0.6036)	0.9436	1
e_3 (0.0674)	0.9802	2
e_4 (1.4737)	0.9999	4

3.3 Subproblem C: Reliability Analysis of Baseline Design

Given the refined epistemic space E_1 provided by the challengers, an updated epistemic box E is selected such that $E \subseteq E_1$. Subproblem (C) requires the evaluation of the (epistemic) range $R_i(\boldsymbol{\theta})$ of the failure probability $P[g_i(\boldsymbol{a}, \boldsymbol{e}, \boldsymbol{\theta}) \geq 0]$ for each individual stability requirement g_i, i = 1, 2, 3, given the baseline system design $\boldsymbol{\theta}$. Also, it is requested to calculate the range $R(\boldsymbol{\theta})$ of the failure probability $P[w(\boldsymbol{a}, \boldsymbol{e}, \boldsymbol{\theta}) \geq 0]$ for all requirements (Crespo and Kenny (2020)).

These problems are addressed by an efficient combination of: (i) Genetic Algorithms (GAs) to deeply search the epistemic space E and solve the optimization problems related to the propagation of epistemic uncertainty by interval analysis; (ii) Monte Carlo Simulation (MCS) to propagate aleatory uncertainty and estimate the failure probabilities (Pedroni and Zio (2015)). GAs are run for 100 generations with a population of 250 individuals; also, N_a = 100000 random samples are used to estimate the failure probabilities by MCS. The corresponding results are reported in Table 4.

Table 4. Results of the failure probability estimation.

Stability requirement	Failure probability range
g_1	[0.0577, 0.1459]
g_2	[0.0680, 0.2980]
g_3	[0.0548, 0.0827]
g_w	[0.1377, 0.3318]

The epistemic parameters are then ranked according to the *reduction* in the length $L[R(\boldsymbol{\theta})]$ of $R(\boldsymbol{\theta})$ that might result from their refinement. As before, each e_i is fixed at different values e_i^k, k = 1, 2, …, N_e = 20, within its range of variation $[\underline{e}_i, \overline{e}_i]$. In correspondence of each e_i^k the length $L[R(\boldsymbol{\theta}|e_i = e_i^k)]$ of the worst-case failure probability range is computed (for the sake of compact notation, let $L[R(\boldsymbol{\theta}|e_i = e_i^k)]$ be indicated as $L[R(\boldsymbol{\theta}|e_i)]$). The corresponding sensitivity indicator $S_i(L)$ for parameter e_i is then computed as:

$$S_i(L) = 1 - \frac{E_{ei}\left[L[R(\boldsymbol{\theta}|e_i)]\right]}{L[R(\boldsymbol{\theta})]}, \quad (11)$$

quantifying the *expected fractional contraction* in $R(\boldsymbol{\theta})$ that would result from a refinement in e_i. The values obtained for $S_i(L)$ are shown in Table 5, together with the corresponding epistemic parameter ranking.

Table 5. Epistemic parameters ranked according to the contraction of $R(\theta)$

Parameter	$S_i(L)$ (Ranking)
e_1	0.4726 (2)
e_2	0.6084 (1)
e_3	0.2725 (3)
e_4	0.0809 (4)

Finally, the severity $s_i(\boldsymbol{\theta})$ of each individual requirement violation is evaluated as the expected value of the requirement g_i *conditional* to failure: refer to (Crespo and Kenny (2020)) for mathematical details. As before, GAs and MCS are combined to solve this task. The results are shown in Table 6.

Table 6. Severity of requirement violation

Stability requirement	$s_i(\boldsymbol{\theta})$
g_1	0.02103
g_2	0.00247
g_3	0.08232

4. Conclusions

Subproblems A, B and C of the NASA Langley UQ Challenge on Optimization Under Uncertainty have been addressed. Model calibration has been addressed by Bayesian inversion in the orthonormal space identified by SVD of the available data: an uncertainty model has been proposed based on GM models. Epistemic uncertainty reduction has been driven by sensitivity analysis and by the ES, a multivariate generalization of the CRPS. Finally, system failure probability ranges have been obtained by GA-based optimization and MCS. In the future: (i) the robustness of some of the results obtained will be possibly increased by further simulations and comparison with other approaches; (ii) Subproblems D and E will be addressed in detail based on the current results.

References

Crespo, L.G. and S. Kenny (2020). The NASA Langley Challenge on Optimization Under Uncertainty. *Proceedings of the 30th European Safety and Reliability Conference and the 15th Probabilistic Safety Assessment and Management Conference 2020*.

Faes, M., M. Broggi, E. Patelli, Y. Govers, J. Mottershead, M. Beer, D. Moens (2019). A multivariate interval approach for inverse uncertainty quantification with limited experimental data. *Mechanical Systems and Signal Processing 118*, 534–548.

Gaymann, A., M. Pietropaoli, L.G. Crespo, S.P. Kenny, F. Montomoli (2018). Random Variable Estimation and Model Calibration in the Presence of Epistemic and Aleatory Uncertainties. *SAE Int. J. Mater. Manuf. 11(4)*, 453-466.

Gneiting, T. and A:E. Raftery (2007). Strictly Proper Scoring Rules, Prediction, and Estimation. *Journal of the American Statistical Association 102-477*, 359-378. Review Article.

Goodman, J. and J. Weare (2010). Ensemble samplers with affine invariance. *Communications in Applied Mathematics and Computational Science* 5(1), 65–80.

Pedroni, N. and E. Zio (2015). Hybrid Uncertainty and Sensitivity Analysis of the Model of a Twin-Jet Aircraft. *Journal of Aerospace Information Systems 12*, 73-96.

Székely, G.J. and M.L. Rizzo (2005). A New Test for Multivariate Normality. *Journal of Multivariate Analysis* 93, 58–80.

Wu, X., T. Kozlowski and H. Meidani (2018). Kriging-based inverse uncertainty quantification of nuclear fuel performance code BISON fission gas release model using time series measurement data. *Reliability Engineering and System Safety 169*, 422–436.